

F

Face Acquisition

- Face Device

► **deformable models**, which encodes the prior knowledge of face shape or appearance, to take into account the low-level image evidences and find the face that is present in the image.

Face Aging

Face aging is to predict the future appearance of human face by learning the aging patterns, child growth, and adult aging are two type of aging.

- And-Or Graph Model for Faces

Introduction

The ability of understanding and interpreting facial structures is important for many image analysis tasks. Suppose that, if we want to identify a person from a surveillance camera, a natural approach would be running the face image of the person through a database of known faces, examining the differences and identifying the best match. However, simply subtracting one image from another would not yield the desirable differences (as shown in Fig. 1), unless two faces are properly aligned. The goal of face alignment is to establish correspondence among different faces, so that the subsequent image analysis tasks can be performed on a common basis.

The main challenge in face alignment arises from pervasive ambiguities in low-level image features. Consider the examples shown in Fig. 2. While the main face structures are present in the ► **feature maps**, the contours of face components are frequently disrupted by gaps or corrupted by spurious fragments. Strong gradient responses could be due to reflectance, occlusion, fine facial texture, or background clutter. In contrast, the boundaries of face components such as nose and eyebrow are often obscure and incomplete. Looking for face components separately is difficult and often yields noisy results.

Rather than searching individual face components and expecting the face structure to emerge from the results, a better strategy is imposing the structure explicitly from the beginning. A majority of work in the field are developed based on this strategy. Deformable template [1], for example, is an elastic model which resembles face structure by assemblies of flexible curves. A set

LEON GU, TAKEO KANADE
Carnegie Mellon University, Pittsburgh, PA, USA

Synonyms

Face registration; Face matching

Definition

Face alignment is a computer vision technology for identifying the geometric structure of human faces in digital images. Given the location and size of a face, it automatically determines the shape of the face components such as eyes and nose. A face alignment program typically operates by iteratively adjusting a



Face Alignment. [Figure 1](#) To compare two face images, by directly adding them or subtracting one from another does not produce the desired result. Face alignment enables to establish correspondences between different images, so that the subsequent tasks can be performed on a common basis.



Face Alignment. [Figure 2](#) The major difficulty in face alignment is low-level image ambiguities. Face topologies could be significantly corrupted in the gradient feature maps (*second row*), due to various factors such as reflectance, occlusion, fine facial texture, and background clutter.

of model parameters control shape details such as the locations of various facial subparts and the angles of hinges which join them. The model is imposed upon and aligned to an image by varying the parameters. This strategy is powerful for resolving low-level image ambiguities. Inspired by this work, many variations of deformable face models emerged, including [2–9]. The common scheme in these work is first to construct a generic face model, then modify it to match the facial features found in a particular image. In this procedure, encoding prior knowledge of human faces, collecting image evidences of facial features, and fusing the observations with priors are the three key problems. Our treatment will follow the method proposed by Gu and Takeo [8, 9], which addresses the above problems in a coherent hierarchical Bayes framework.

Constructing Face Priors

This article concerns with the prior knowledge of a particular kind, namely shape priors. Suppose that, a face consists of a set of landmark points, which are typically placed along the boundaries of face components, i.e., $S = (x_1, y_1, \dots, x_n, y_n)$. It can be viewed as a random vector, and its distribution, commonly called shape prior, describes the plausible spatial configurations of the landmark set. A principled way to construct the prior is by learning the distribution from training samples.

Face appears in different scales and orientations. First we need to transform all training face images into a common coordinate frame. One popular approach is general procrustes analysis [10]. It consists of two recursive steps: computing the mean shape, and

aligning each training shape with the mean by a rigid transformation. These two steps are repeated until the differences between the mean and the training shapes are minimized.

Next, we construct shape prior from the aligned training samples. The spatial arrangement of facial landmarks, although deformable, has to satisfy certain constraints. For example, it is often reasonable to assume that face shape is normally distributed, therefore, to learn the distribution we simply compute the mean and the covariance of the training shapes. More specifically, since the intrinsic variability of face structure is independent to its representation, e.g., the number of landmarks, we can parameterize face shape in a low-dimensional subspace [6, 8], such as

$$S = \Phi b + \mu + \epsilon. \quad (1)$$

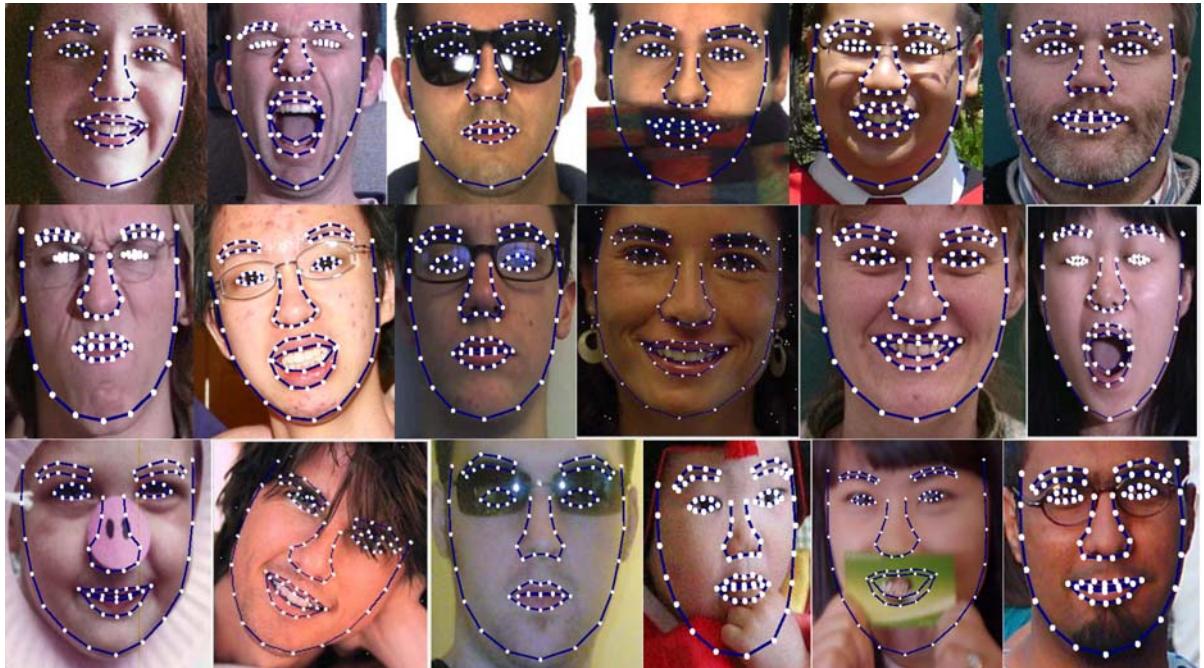
The columns of Φ denote the major “modes” of shape deformations, and the elements of b controls the magnitude of deformation on the corresponding mode. This model has a nice generative interpretation: the shape vector S is generated by first adding a sequence of deformations $\{\Phi_i b_i\}$ into the mean shape μ , then permuting the resultant shape by an Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$. From a geometric perspective, the

matrix Φ span a low-dimensional subspace which is centered at μ , the deformation coefficient b is the projection of S in the subspace, and ϵ denotes the deviation of S from the subspace. If assuming the elements of b to be independently normal, i.e., $b \sim \mathcal{N}(0, \Sigma)$ and Σ is diagonal, the distribution over the shape S is a constrained Gaussian, $S \sim \mathcal{N}(\mu, \Phi \Sigma \Phi^t + \sigma^2 I)$. The model parameters μ , Φ , Σ , and σ can be learned from training data. This model is also known as probabilistic principal component analysis [11] in the field of machine learning.

F

Detecting Facial Features

Strong gradient response is not the only way to characterize facial features. Some feature points may correspond to a weaker secondary edge in local context instead of the strongest; other points such as eye corners may have rich image structure that is more informative than gradient magnitude. Facial feature modeling can be made more effective by constructing detectors specific to each individual feature. One simple detector [2], for example, is a normal distribution built on the local gradient structures of each point. The distribution is learned



Face Alignment. Figure 3 Face alignment results from Gu and Kanade [9].

from training face images, and applied to evaluate the target image. Concatenating the best candidate position (u_p, v_i) of each feature point, we obtain an “observation” $Q = (u_1, v_1, \dots, u_n, v_n)$ of the face shape that is likely to be present in the image. The observation is related with the aligned shape S by a rigid transformation

$$Q = \mathcal{T}(S, \theta) + \eta, \quad (2)$$

where $\theta = \{t, s, r\}$ denotes the transformation parameters (translation, scale, and rotation), and η is an additive observation noise. The conditional $p(Q|S)$ remains to be normal if the transformation \mathcal{T} is linear, e.g., rigid or affine. More sophisticated detectors have been developed to produce better observations, however, after decades of research people have learned that individual feature detectors are effective only up to a point and cannot be expected to retrieve the entire face structure.

Fusing Prior with Image Observations

Combining the deformation model (1) with the transformation model (2) a hierarchical Bayes model is established that simulates how a random observation Q is generated from the deformation magnitude b and the transformation parameters θ . In this framework, the face alignment task is to modify shape priors to take into account the image evidences, arriving at the target face shape in images. EM algorithm is typically used for inferring the posterior b and θ , and analytic solutions exist for both E and M steps when the transformation is linear. This framework has been extended to model three-dimensional transformations for aligning multi-view faces [8], and nonlinear shape deformations for dealing with face images with exaggerated facial expressions [9]. Figure 3 shows a few alignment results from [9].

Summary

Significant progresses have been made in face alignment in recent years. The hierarchical Bayes formulation introduced in this article provides a systematic way to resolve low-level image ambiguities and exploit prior knowledge. Face alignment has a wide range of applications including face recognition, expression analysis, facial animation, lip reading, and human–computer interaction.

Related Entries

- Deformable Models
- Face Warping
- Feature Map

References

1. Yuille, A.L., Hallinan, P.W., Cohen, D.S.: Feature extraction from faces using deformable templates. *Int. J. Comput. Vision* **8**(2), 99–111 (1992). DOI <http://dx.doi.org/10.1007/BF00127169>. URL http://www.stat.ucla.edu/~yuille/pubs/optimize_papers/DT_IJCV1992.pdf
2. Cootes, T.F., Taylor, C., Cooper, D., Graham, J.: Active shape models – their training and their applications. *Comput. Vision Image Understanding* (1995)
3. Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 775–779 (1997). DOI <http://dx.doi.org/10.1109/34.598235>. URL http://www.face-rec.org/algo_rithms/EBGM/WisFelKrue99-FaceRecognition-JainBook.pdf
4. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d-faces. In: ACM SIGGRAPH (1999)
5. Cootes, T., Edwards, G., Taylor, C.: Active appearance models **23**(6), 681–685 (2001)
6. Zhou, Y., Gu, L., Zhang, H.: Bayesian tangent shape model: Estimating shape and pose parameters via Bayesian inference, pp. I: 109–116 (2003). URL http://www.cs.cmu.edu/~gu/publication/alignment_cvpr03.pdf
7. Zhang, Z., Liu, Z., Adler, D., Cohen, M.E., Hanson, E., Shan, Y.: Robust and rapid generation of animated faces from video images – a model-based modeling approach. *Int. J. Comput. Vision* (2004)
8. Gu, L., Kanade, T.: 3d alignment of face in a single image. In: CVPR (2006)
9. Gu, L., Kanade, T.: A generative shape regularization model for robust face alignment. In: The Tenth European Conference on Computer Vision (2008)
10. Goodall, C.: Procrustes methods in the statistical analysis of shape. *J. Royal Statistical Society. Series B (Methodological)* **53**, 285–339 (1991).
11. Jolliffe, M., Bishop, C.: Probabilistic principal component analysis. *J. Royal Statistical Society* (1999).

Face Alignment Error

- Face Misalignment Problem

Face Biometric

► Face Recognition, Overview

Face Camera

► Face Device

Face Databases and Evaluation

DMITRY O. GORODNICHY

Laboratory and Scientific Services Directorate, Canada Border Services Agency, Ottawa, ON, Canada

Synonyms

Face recognition performance evaluation

Definition

Face Databases are imagery data that are used for testing ► face processing algorithms. In the contents of biometrics, face databases are collected and used to evaluate the performance of face recognition biometric systems.

Face recognition evaluation is the procedure that is used to access the recognition quality of a face recognition system. It involves testing the system on a set of face databases and/or in a specific setup for the purpose of obtaining measurable statistics that can be used to compare systems to one another.

Introduction: Factors Affecting Face Recognition Performance

While for humans recognizing a face in a photograph or in video is natural and easy, computerized face recognition is very challenging. In fact, automated recognition of faces is known to be more difficult than recognition of other imagery data such as iris, vein, or fingerprint images due to the fact that the human face is a non-rigid 3D object which can be observed at different

angles and which may also be partially occluded. Specifically, face recognition systems have to be evaluated with respect to the following factors [1]:

1. Face image resolution – face images can be captured at different resolutions: face images scanned from documents may have very high resolution, while face captured with a video camera will mostly be of very low resolution,
2. Facial image quality – face images can be blurred due to motion, out of focus, and of low contrast due to insufficient camera exposure or aperture, especially when captured in uncontrolled environment,
3. Head orientation – unless a person is forced to face the camera and look straight into it, will unlikely be seen under the same orientation on the captured image,
4. Facial expression – unless a person is quiet and motionless, the human face constantly exhibits a variety of facial expressions
5. Lighting conditions – depending on the location of the source of light with respect to the camera and the captured face, facial image will be seen with different illumination pattern overlaid on top of the image of the face,
6. Occlusion – image of the face may be occluded by hair, eye-glasses and clothes such as scarf or handkerchief,
7. Aging and facial surgery – compared to fingerprint or iris, person faces changes much more rapidly with time, it can also be changed as a result of make-up or surgery.

There are over thirty publicly available face databases. In addition, there are Face Recognition Vendor Test (FRVT) databases that are used for independent evaluation of Face Recognition Biometric Systems (FRBS). Table 1 summarizes the features of the most frequently used still image facial databases, as pertaining to the performance factors listed above. More details about each database can be found at [2–4] and below are presented some of them. For the list of some video-based facial databases, see [5].

Public Databases

One of the first and most used databases is AT&T (formerly “Olivetti ORL”) database [6] that contains

ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement).

The other most frequently used dataset is developed for FERET program [7]. The images were collected in a semi-controlled environment. To maintain a degree of consistency throughout the database, the same physical setup was used in each photography session. A duplicate set is a second set of images of a person already in the database and was usually taken on a different day. For some individuals, over 2 years had elapsed between their first and last sittings, with some subjects being photographed multiple times.

The Yale Face Database [8] contains images of different facial expression and configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink. The Yale Face Database B provides single light source images of 10 subjects each seen under 576 viewing conditions (9 poses x 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured.

The BANCA multi-modal database was collected as part of the European BANCA project, which aimed at developing and implementing a secure system with enhanced identification, authentication, and access control schemes for applications over the Internet [9]. The database was designed to test multimodal identity verification with various acquisition devices (high and low quality cameras and microphones) and under several scenarios (controlled, degraded, and adverse).

To investigate the time dependence in face recognition, a large database is collected at the University of Notre Dame [10]. In addition to the studio recordings, two images with unstructured lighting are obtained.

University of Texas presents a collection of a large database of static digital images and video clips of faces [11]. Data were collected in four different categories: still facial mug shots, dynamic facial mug shots, dynamic facial speech and dynamic facial expression. For the still facial mug shots, nine views of the subject, ranging from left to right profile in equal-degree steps were recorded. The sequence length is cropped to be 10 s.

The AR Face Database [12] is one of the largest datasets showing faces with different facial expressions,

illumination conditions, and occlusions (sun glasses and scarf).

XM2VTS Multimodal Face Database provides five shots for each person [13]. These shots were taken at one week intervals or when drastic face changes occurred in the meantime. During each shot, people have been asked to count from “0” to “9” in their native language (most of the people are French speaking), rotate the head from 0 to –90 degrees, again to 0, then to +90 and back to 0 degrees. Also, they have been asked to rotate the head once again without glasses if they wear any.

CMU PIE Database is one of the largest datasets contains images of 68 people, each under 13 different poses, 43 different illumination conditions, and with four different expressions [14].

The Korean Face Database (KFDB) contains facial imagery of a large number of Korean subjects collected under carefully controlled conditions [15]. Similar to the CMU PIE database, this database has images with varying pose, illumination, and facial expressions were recorded. In total, 52 images were obtained per subject. The database also contains extensive ground truth information. The location of 26 feature points (if visible) is available for each face image.

CAS-PEAL Face Database is another large-scale Chinese face database with different sources of variations, especially Pose, Expression, Accessories, and Lighting [16].

FRVT Databases

Face Recognition Vendor Tests (FRVT) provide independent government evaluations of commercially available and prototype face recognition technologies [4]. These evaluations are designed to provide U.S. Government and law enforcement agencies with information to assist them in determining where and how facial recognition technology can best be deployed. In addition, FRVT results serve to identify future research directions for the face recognition community. FRVT 2006 follows five previous face recognition technology evaluations – three FERET evaluations (1994, 1995 and 1996) and FRVT 2000 and 2002.

FRVT provides two new datasets that can be used for the purpose: high computational intensity test (HCInt) data set and Medium Computational Intensity test (MCInt) data set. HCInt has 121,589 operational well-posed (i.e. frontal to within 10 degrees) images of 37,437 people, with at least three images of each person.

The images are provided from the U.S. Department of State's Mexican non-immigrant visa archive. The images are of good quality and are gathered in a consistent manner, collected at U.S. consular offices using standard issue digital imaging apparatus whose specification remained fixed over the collection period.

The MCInt data set is composed of a heterogeneous set of still images and video sequences of subjects in a variety of poses, activities and illumination conditions. The data are collected from several sources, captured indoors and outdoors, and include close-range video clips and static images (with over hundred individuals), high quality still images, Exploration Video Sequences (where faces move through the nine facial poses used for the still images) and Facial Speech Videos (where two video clips were taken of individuals speaking, first in a neutral way, then in an animated way).

Face Evaluation

For an evaluation to be accepted by the biometric community, the performance results have to be published along with the evaluation protocol. An evaluation protocol describes how the experiments are run and how the data are collected. It should be written in sufficient detail so that users, developers, and vendors can repeat the evaluation.

The main attributes of the evaluation protocol are described below.

Image Domain and Face Processing Tasks

There are two image domains where Face Recognition Biometric Systems (FRBS) are applied:

1. *Face recognition in documents* (FRID), in particular, face recognition from Machine Readable Travel Documents (MRTD).
2. *Face recognition in video* (FRIV), also referred to as *Face in Crowd* problem, an example of which is face recognition from surveillance video and TV.

These two image domains are very different [17]. The systems that perform well in one domain may not perform well in the other [18].

FRID deals with facial data that are of high spatial resolution, but that are very limited or absent in ▶ **temporal domain** – FRID face images would normally have *intra-ocular distance* (IOD) of at least 60

pixels, which is the distance defined by the ▶ **canonical face model** established by International Civil Aviation Organization (ICAO) for MRTD. There will however be not more than one or very few images available of the same person captured over a period of time.

In contrast, FRIV deals with facial images that are available in abundance in temporal domain but which are of much lower spatial resolution. The IOD of facial images in video is often lower than 60 pixels, due to the fact that face normally occupies less than one eighth of a video image, which itself is relatively small (352×240 for analog video or 720×480 for digital video) compared to a scanned document image. In fact, IOD of faces detected in video is often just slightly higher than or equal to 10 pixels, which is the minimal IOD that permits automatic detection of faces in images [19].

While for FRID facial images are often extracted beforehand and face recognition problem is considered in isolation from other face processing problems, FRIV requires that a system be capable of performing several other facial processing tasks prior to face recognition, such as face detection, face tracking, eye localization, best facial image selection or reconstruction, which may also be coupled with facial image accumulation and video snapshot resolution enhancement [20]. Evaluation of FRBS for FRID is normally performed by testing a system on static facial images datasets described above. To evaluate FRBS for FRIV however, it is much more common to see the system testing performed as a pilot project on a real-life video monitoring surveillance task [21], although some effort to evaluate their performance using prerecorded datasets and motion pictures has been also suggested and performed [5].

Use of Color

Color information does not affect the face recognition performance [22], which is why many countries still allow black-and-white face pictures in passport documents. Color however plays an important role in face detection and tracking as well as in eye localization. Therefore, for testing recognition from video, color video streams should be used.

Scenario Taxonomy

The following scenario taxonomy is established to categorize the performance of biometric systems [23]:

Database (year created)	#individuals/ # images	i.o.d/ image width	Orientation	Expression	Lighting /quality	Occlusion	Situations	Representative Facial image
AT & T Olivetti (1992–1994)	40/400	~60/92	yes	yes	yes	yes	yes	
FERET (1993–1996)	1999/ 14,126	~80/256	9–20	2	2		2	
AR	116/ 3288	~90/768	1	4	4	2	2	
Yale (B)	15/165 10/5760	~80/640	9		64			
PIE 2000	68/ 41,368	~75/640	13	3	43			
Korean	1000/ 52000	~80/640	7	5	16			
Cas-peal 2003	1040	~45/360	21	15	6		1–5	
Human ID	350/ 15.500	~80/1600	1	2	3		10	
UofT 2002	284	~80/720	video	video				
Banca 2002–2003	208/ 208*12	~45/720	1	yes	3		12	
XM2VTS	293/	~100/720	Full rotation	speaking		Yes - eyeglasses	4	
Equinox	91	~100/240	1	3	3			
Cmu-hyperspectral	54	80/640	1		4		5	
nist	573/3248	~80	2					
FRVT HCInt	37,437/ 121,589	1					3	
FRVT MCInt 1999–2002			several	Still and video	several			

Face Databases and Evaluation. [Figure 1](#) Face databases, categorized by the factors affecting the performance of face recognition systems: such as number of probes, face image resolution, head orientation, face expression, changed in lighting, image quality degradation, occlusion, and aging.

cooperative vs. non-cooperative, overt vs. covert, habituated vs. non-habituated, attended vs. non-attended, public vs. private, standard vs. non-standard. When performing evaluation of FRBS, these categories have to be indicated.

Dataset Type and Recognition Task

Two types of datasets are possible for recognition problems:

1. Closed dataset, where each query face is present in the database, as in a watch list in the case of negative enrollment, or as in a list of computer users or ATM clients, in the case of positive enrollment,
2. Open dataset, where query faces may not be (or very likely are not) in the database, as in the case of surveillance video monitoring.

FRBS can be used for one three face recognition tasks:

1. Face verification, also referred to as authentication or 1 to 1 recognition, or positive verification, as verifying ATM clients,
2. Face identification, also referred to as or 1 to N (negative identification – as detecting suspects from a watch list), where a query face is compared against all faces in a database and the best match (or the best k matches) are selected to identify a person.
3. Face classification, also referred to as categorization, where a person is recognized as belonging to one of the limited number of classes, such as describing the person's gender (male, female), race (caucasian, asian etc), and various medical or genetic conditions (Down's Syndrome etc).

While the result of the verification and identification task are used as hard biometrics, the results from classification can be used as *soft biometrics*, similar to person's height or weight.

Performance Measures

The performance is evaluated against two main errors a system can exhibit:

1. False accept (FA) also known as false match (FM), false positive (FP) or type I error.
2. False reject (FR) also known as false non-match (FNM) or false negative (FN) or type 2 error.

By applying a FRBS on a significantly large data set of facial images, the total number of FA and FR are measured and used to compute one or several of the following cumulative measurements and figures of merit (FOM). For verification systems,

1. FA rate (FAR) with fixed FR rate.
2. FR rate (FRR), or true acceptance rate ($TAR = 1 - FRR$), also known as true positive (or hit) rate, at fixed FA rate.
3. Detection Error Trade-off (DET) curve, which is the graph of FAR vs FRR, which is obtained by varying the system parameters such as *match threshold*.
4. Receiver Operator Characteristic (ROC) curve, which is similar to DET curve, but plots TAR against FAR.
5. Equal error rate (EER), which the FAR measured when it equals FRR.

For identification tasks,

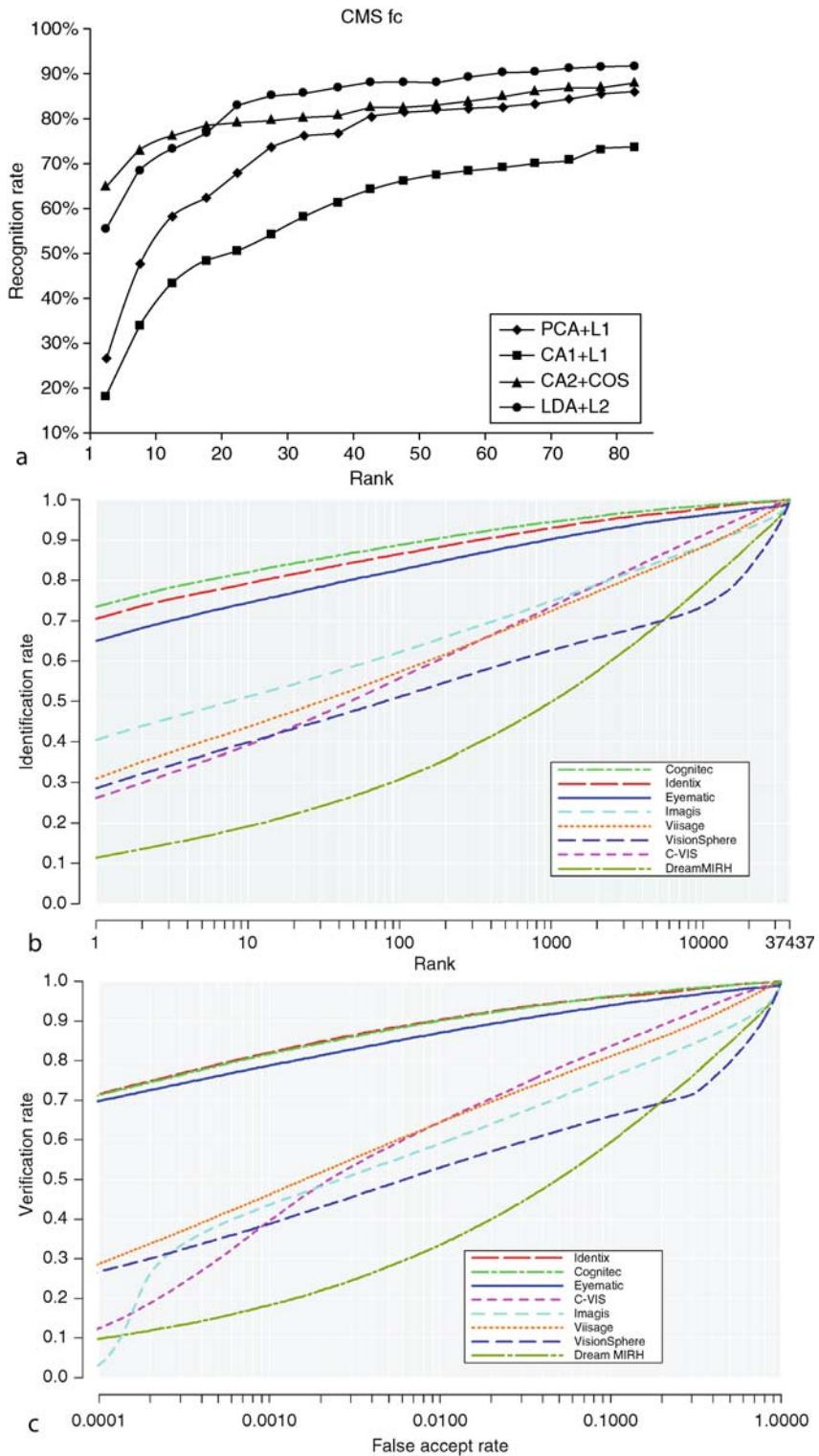
1. Identification rate, or rank-1 identification, which is number of times when the correct identity is chosen as the most likely candidate.
2. Rank-k identification rate (R_k), which is number of times the the correct identity is in the top k most likely candidates.
3. Cumulative Match Characteristic (CMC), which plots the rank-k identification rate against k .

The rates are counted as percentages to the number of faces in a databases. DET and ROC curves are often plotted using logarithmic axes to better differentiate the systems that shows similar performance.

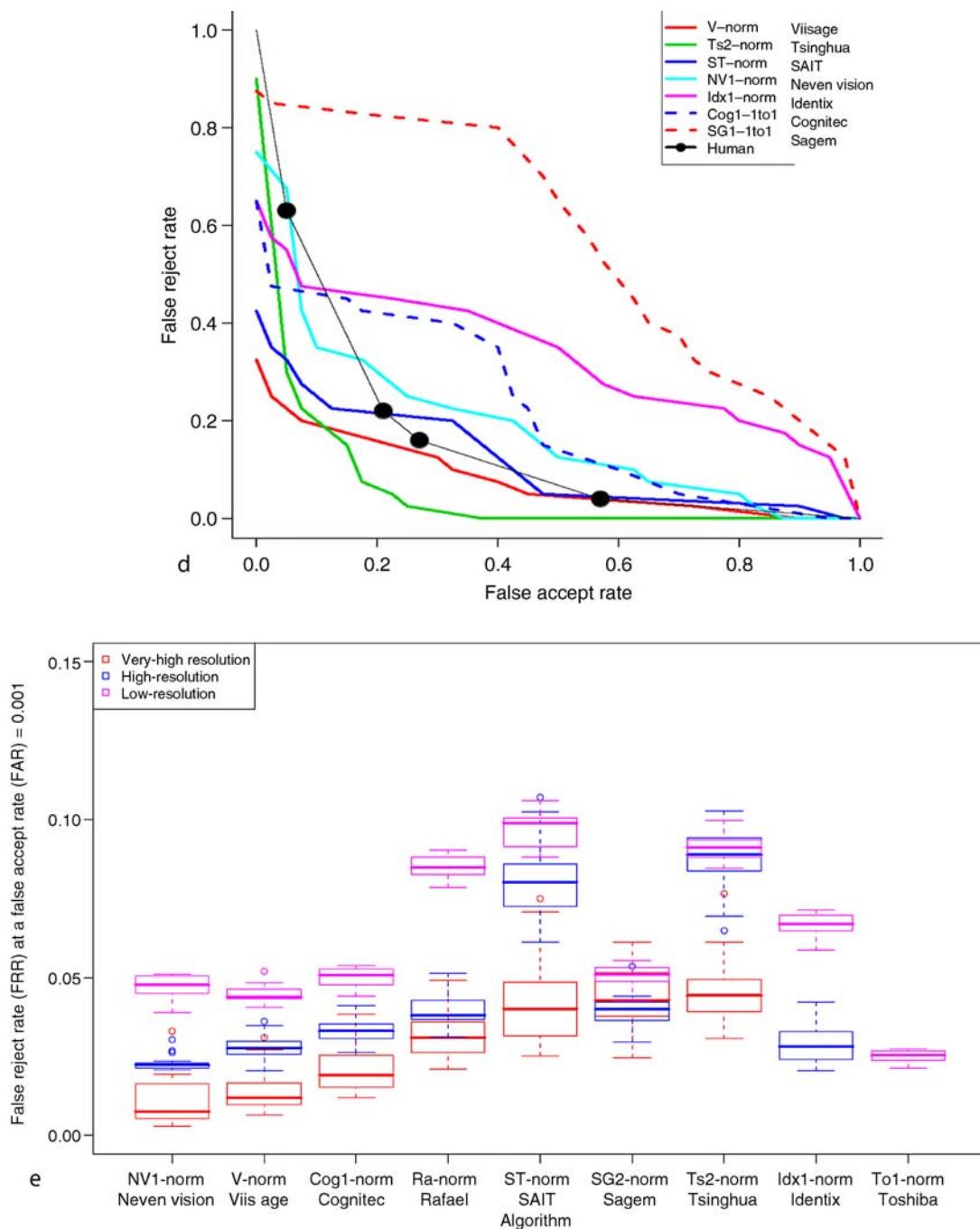
Similarity Metrics, Normalization, and Data Fusion

Different types of metrics can be used to compare ► **feature vectors** of different faces to one another. The recognition results can also be normalized. Proper covariance-weighted metrics and normalization should be used when comparing the performance results obtained on different datasets.

When temporal data are available, as when recognizing a person from a video sequence, the recognition results are often integrated over time in a procedure known as evidence accumulation or data fusion. The details of this should be known.



Face Databases and Evaluation. Figure 2 (Continued)



Face Databases and Evaluation. **Figure 2** Examples of performance evaluation conducted on face databases: (a) identification performance of several appearance-based recognition algorithms measured using CMC curves on FERET database (from [25]), (b–e) verification and identification performance of commercial face recognition biometrics systems on FRVT datasets (from [24, 26], using CMC curves (b), ROC curve (c), DET curve (d) and fixed-FAR FRR distributions (e)).

Example Protocols

Feret protocol [7] is an example of the close set face identification, where a full distance matrix that measures the similarity between each query image and each database image is computed. FRVT2002 [24] addresses both open set verification problem and close-set identification problem and uses CMC and ROC to compare the results. BANCA protocol [9], which is designed for multi-modal databases, is an example of the open set verification protocol. XM2VTS Lausanne protocol [13] is an example of a close set verification, where anyone not in the database is considered an imposter.

Evaluation Results

Face Databases have been used over the years to compare and improve the existing face recognition techniques. Some of the obtained evaluation results are shown in Fig. 2. Figure 2a shows face identification results from [25] for popular appearance-based face-recognition techniques: Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA), obtained on FERET database using CMC curves.

Figures 2b–e show performance evaluation of commercial FRBSs that participated in the FRVT2002 and FRVT2006 tests taken from [24, 26]

Future Work

Considerable advances have been made recently in the area of automated face recognition. FRBSs are now able to *recognize faces in documents* with the performance that matches or exceeds the human recognition performance. In large part, this has become possible due to the help of many researchers that have collected and maintained face databases. At the same time, despite the intensive use of these databases, no FRBS has been developed so far that can *recognize faces from video* with performance close to that of humans.

Automated recognition of faces from video is considerably worse than face recognition from documents, whereas for humans it is known to be the opposite. This status-quo situation serves as an indication that new evaluation datasets and benchmarks are needed for

testing video-based face recognition systems. Knowing how easily available have become recently amounts of various video data (including news casts, televised shows, motion pictures, etc), it is foreseen that instead of using video-based data-bases, which are very costly and time consuming to create, the research community will soon adopt face evaluation benchmarks and protocols based on public domain video recordings [5].

The importance of improving the performance of video-based face recognition should not be underestimated, taking into account that of all hard biometric modalities, video-based face recognition is the most collectable and acceptable [27].

Related Entries

- ▶ Face Detection
- ▶ Face Recognition
- ▶ Identification
- ▶ Verification

References

1. Gorodnichy, D.O.: Facial recognition in video. In: Proceedings of the International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA'03), LNCS 2688, pp. 505–514, Guildford, UK, online at <http://iit-iti.nrc-cnrc.gc.ca/iit-publications-iti/docs/NRC-47150.pdf>. (2003)
2. Gross, R.: Face Databases. Springer, New York (2005)
3. Face Recognition website. <http://www.face-rec.org>
4. Face Recognition Vendor Test website. <http://www.frvt.org>
5. Gorodnichy, D.O.: Seeing faces in video by computers (Editorial). Image and Video Computing, Special Issue on Face Processing in Video Sequences (online at <http://iit-iti.nrc-cnrc.gc.ca/iit-publications-iti/docs/NRC-48295.pdf>) **24**, 1–6 (2006)
6. AT&T: The Database of Faces (formerly The ORL Database of Faces, <http://www.cl.cam.ac.uk/research/dtg/attarchive/faces/taglance.html>)
7. Phillips, P.J., Moon, H., Rizvi, S., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1090–1104 (2000). <http://www.nist.gov/humanid/feret/>
8. Georghiades, A., Kriegman, D., Belhumeur, P.: From few to many: generative models for recognition under variable pose and illumination. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 643–660 (2001)
9. Bailly-Bailliére, E., Bengio, S., Bimbot, F., Hamouz, M., Kittler, J., Mariethoz, J., Matas, J., Messer, K., Popovici, V., Poree, F., Ruiz, B., Thiran, J.-P.: The BANCA database and evaluation protocol. In: Audio- and Video-Based Biometric Person Authentication (AVBPA), pp. 625–638 (2003)

10. Phillips, P.J.: Human identification technical challenges. In IEEE International Conference on Image Processing, vol. 1, pp. 22–25 (2002)
11. OToole, A., Harms, J., Snow, S., Hurst, D.R. Pappas, M., Abdi, H.: A video database of moving faces and people, submitted (2003)
12. Martinez, A.M., Benavente, R.: The AR face database. Technical Report 24, Computer Vision Center(CVC) Technical Report, Barcelona (1998)
13. Messer, K., Matas, J., Kittler, J., Luettin, J., Maitre, G.: XM2VTSDB: The extended M2VTS database. In: Second International Conference on Audio and Video-based Biometric Person Authentication (1999)
14. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12), 1615–1618, http://www.ri.cmu.edu/projects/project_418.html (2003)
15. Hwang, B.-W., Byun, H., Roh, M.-C., Lee, S.-W.: Performance evaluation of face recognition algorithms on the asian face database, KFDB. In Audio- and Video-Based Biometric Person Authentication (AVBPA), pp. 557–565 (2003)
16. Gao, W., Cao, B., Shan, S., Zhou, D., Zhang, X. Zhao, D.: CAS-PEAL large-scale Chinese face database and evaluation protocols. Technical Report JDL-TR-04-FR-001, Joint Research and Development Laboratory, <http://www.jdl.ac.cn/peal> (2004)
17. Gorodnichy, D.O.: Video-based framework for face recognition in video. In: Second International Workshop on Face Processing in Video (FPiV'05). Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV'05), pp. 330–338. Victoria, BC, Canada, ISBN 0-7695-2319-6, online at <http://iit-itc.nrc-cnrc.gc.ca/iit-publications-itc/docs/NRC-48216.pdf>. (2005)
18. Gorodnichy, D.O.: Recognizing faces in video requires approaches different from those developed for face recognition in photographs. In: Proceedings of NATO IST - 044 Workshop on Enhancing Information Systems Security through Biometrics. Ottawa, ON, Canada, October 18–20 (online at <http://iit-itc.nrc-cnrc.gc.ca/iit-publications-itc/docs/NRC-47149.pdf>). (2004)
19. Shakhnarovich, G., Viola, P.A., Moghaddam, B.: A unified learning framework for realtime face detection and classification. In: International Conference on Automatic Face and Gesture Recognition, pp. 10–15, USA (2002)
20. Gorodnichy, D.O.: Introduction to the First IEEE Workshop on Face Processing in Video. In: First IEEE CVPR Workshop on Face Processing in Video (FPIV'04), Washington DC, USA, online at <http://www.visioninterface.net/fpix04/preface.html> (2004)
21. Willing, R.: Airport anti-terror systems flub tests face-recognition technology fails to flag suspects. in USA TODAY. Accessed Sept 4, 2003. <http://www.usatoday.com/usatonline/20030902/5460651s.htm>.
22. Yip, A., Sinha, P.: Role of color in face recognition. MIT tech report (ai.mit.com) AIM-2001-035 CBCL-212 (2001)
23. Wayman, J.L., Jain, A.K., Maltoni, D., Maio, D. (eds.): Biometric Systems: Technology, Design and Performance Evaluation. Springer, New York (2005)
24. Phillips, P.J., Grother, P., Ross, J.M., Blackburn, D., Tabassi, E., Bone, M.: Face recognition vendor test 2002: evaluation report (March 2003)
25. Delac, K., Grgic, M., Grgic, S.: Independent Comparative Study of PCA, ICA, and LDA on the FERET Data Set. Int. J. Imaging Syst. Technol. **15**(5), 252–260 (2006)
26. Overview of the Face Recognition Grand Challenge - IEEE Conference on Computer Vision and Pattern Recognition, June 2005. Online at <http://www.frvt.org/FRGC>
27. Jain, A.K., Ross, A., Prabhakar, S.: An Introduction to Biometric Recognition. IEEE Trans. Circ. Syst. Video Technol. Special Issue on Image- and Video-Based Biometrics **14**, 4–20 (2004)

Face Detection

MING-HSUAN YANG

University of California, Merced, CA, USA

Synonym

Face Localization

Definition

Face detection is concerned with finding whether there are any faces in a given image (usually in gray scale) and, if present, return the image location and content of each face. This is the first step of any fully automatic system that analyzes the information contained in faces (e.g., identity, gender, expression, age, race, and pose). While earlier work dealt mainly with upright frontal faces, several systems have been developed that are able to detect faces fairly accurately with in-plane or out-of-plane rotations in real time. Although a face detection module is typically designed to deal with single images, its performance can be further improved if video stream is available.

Introduction

The advances of computing technology has facilitated the development of real-time vision modules that interact with humans in recent years. Examples abound, particularly in biometrics and human computer interaction as the information contained in faces needs to be analyzed for systems to react accordingly. For biometric systems that use faces as nonintrusive input modules,

it is imperative to locate faces in a scene before any recognition algorithm can be applied. An intelligent vision-based user interface should be able to tell the attention focus of the user (i.e., where the user is looking at) in order to respond accordingly. To detect facial features accurately for applications such as digital cosmetics, faces need to be located and registered first to facilitate further processing. It is evident that face detection plays an important and critical role for the success of any face processing systems.

The face detection problem is challenging as it needs to account for all possible appearance variation caused by change in illumination, facial features, occlusions, etc. In addition, it has to detect faces that appear at different scale, pose, with in-plane rotations. In spite of all these difficulties, tremendous progress has been made in the last decade and many systems have shown impressive real-time performance. The recent advances of these algorithms have also made significant contributions in detecting other objects such as humans/pedestrians, and cars.

Operation of a Face Detection System

Most detection systems carry out the task by extracting certain properties (e.g., local features or holistic intensity patterns) of a set of training images acquired at a fixed pose (e.g., upright frontal pose) in an off-line setting. To reduce the effects of illumination change, these images are processed with histogram equalization [1, 2] or standardization (i.e., zero mean unit variance) [3]. Based on the extracted properties, these systems typically scan through the entire image at every possible location and scale in order to locate faces. The extracted properties can be either manually coded (with human knowledge) or learned from a set of data as adopted in the recent systems that have demonstrated impressive results [1, 2, 3, 4, 5]. In order to detect faces at different scale, the detection process is usually repeated to a pyramid of images whose resolution are reduced by a certain factor (e.g., 1.2) from the original one [1, 2]. Such procedures may be expedited when other visual cues can be accurately incorporated (e.g., color and motion) as pre-processing steps to reduce the search space [5]. As faces are often detected across scale, the raw detected faces are usually further processed to combine overlapped results and remove false positives with heuristics (e.g., faces typically do

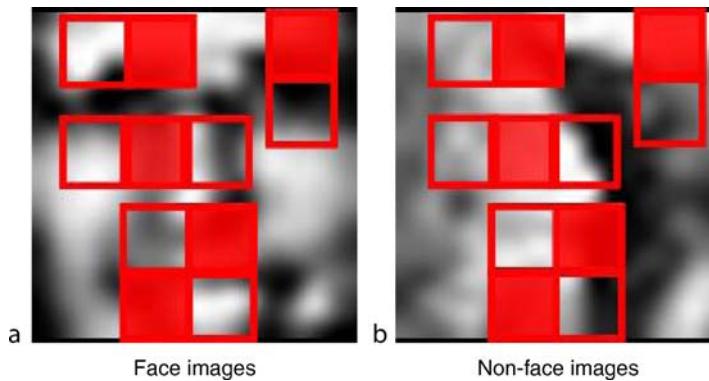
not overlap in images) [2] or further processing (e.g., edge detection and intensity variance).

Numerous representations have been proposed for face detection, including pixel-based [1, 2, 5], parts-based [4, 6, 7], local edge features [8], Haar wavelets [4, 9], and ▶ **Haar-like features** [3, 10]. While earlier holistic representation schemes are able to detect faces [1, 2, 5], the recent systems with Haar-like features [3, 11, 12] have demonstrated impressive empirical results in detecting faces under occlusion. A large and representative training set of face images is essential for the success of learning-based face detectors. From the set of collected data, more positive examples can be synthetically generated by perturbing, mirroring, rotating, and scaling the original face images [1, 2]. On the other hand, it is relatively easier to collect negative examples by randomly sampling images without face images [1, 2].

As face detection can be mainly formulated as a pattern recognition problem, numerous algorithms have been proposed to learn their generic templates (e.g., eigenface and statistical distribution) or discriminant classifiers (e.g., neural networks, Fisher linear discriminant, sparse network of Winnows, decision tree, Bayes classifiers, support vector machines, and ▶ **AdaBoost**). Typically, a good face detection system needs to be trained with several iterations. One common method to further improve the system is to bootstrap a trained face detector with test sets, and retrain the system with the false positive as well as negatives [2]. This process is repeated several times to further improve the performance of a face detector. A survey on these topics can be found in [5], and the most recent advances are discussed in the next section.

Recent Advances

The AdaBoost-based face detector by Viola and Jones [3] demonstrated that faces can be fairly reliably detected in real-time (i.e., more than 15 frames per second on 320×240 images with desktop computers) under partial occlusion. While Haar wavelets were used in [9] for representing faces and pedestrians, they proposed the use of Haar-like features which can be computed efficiently with integral image [3]. Figure 1 shows four types of Haar-like features that are used to encode the horizontal, vertical, and diagonal



Face Detection. [Figure 1](#) Four types of Haar-like features. These features appear at different position and scale. The Haar-like features are computed as the difference of dark and light regions. They can be considered as features that collect local edge information at different orientation and scale. The set of Haar-like features is large, and only a small amount of them are learned from positive and negative examples for face detection.

intensity information of face images at different position and scale.

Given a sample image of 24×24 pixels, the exhaustive set of parameterized Haar-like features (at different position and scale) is very large (about 160,000). Contrary to most of the prior algorithms that use one single strong classifier (e.g., neural networks and support vector machines), they used an ensemble of weak classifiers where each one is constructed by thresholding of one Haar-like feature. The weak classifiers are selected and weighted using the AdaBoost algorithm [13]. It is worth to note that boosting algorithms can also be derived from the perspective of function approximation with gradient descent and applications for regression [14]. As there are large number of weak classifiers, they presented a method to rank these classifiers into several cascades using a set of optimization criteria. Within each stage, an ensemble of several weak classifiers is trained using the AdaBoost algorithm. The motivation behind the cascade of classifier is that simple classifiers at early stage can filter out most negative examples efficiently, and stronger classifiers at later stage are only necessary to deal with instances that look like faces. The final detector, a 38 layer cascade of classifiers with 6,060 Haar-like features, demonstrated impressive real-time performance with fairly high detection and low false positive rates. Several extensions to detect faces in multiple views with in-plane rotation have since been proposed [11, 12, 15]. An implementation of the AdaBoost-based face detector [3] can be found in the Intel OpenCV library.

Despite the excellent run-time performance of boosted cascade classifier [3], the training time of such a system is rather lengthy. In addition, the ► [classifier cascade](#) is an example of degenerate decision tree with an unbalanced data set (i.e., a small set of positive examples and a huge set of negative ones). Numerous algorithms have been proposed to address these issues and extended to detect faces in multiple views. To handle the asymmetry between the positive and negative data sets, Viola and Jones proposed the asymmetric AdaBoost algorithm [16] which keeps most of the weights on the positive examples. In [3], the AdaBoost algorithm is used to select a specified number of weak classifiers with lowest error rates for each cascade and the process is repeated until a set of optimization criteria (i.e., the number of stages, the number of features of each stage, and the detection/false positive rates) is satisfied. As each weak classifier is made of one single Haar-like feature, the process within each stage can be considered as a feature selection problem. Instead of repeating the feature selection process at each stage, Wu et al. [17] presented a greedy algorithm for determining the set of features for all stages first before training the cascade classifier. With the greedy feature selection algorithm used as a pre-computing procedure, they reported that the training time of the classifier cascade with AdaBoost is reduced by 50–100 times. For learning in each stage (or node) within the classifier cascade, they also exploited the asymmetry between positive and negative data using a linear classifier with the assumption that they can be modeled with Gaussian distributions [17]. The merits and drawbacks of the

proposed linear asymmetric classifier as well as the classic Fisher linear discriminant were also examined in their work. Recently, Pham and Cham proposed an online algorithm that learns asymmetric boosted classifiers [18] with significant gain in training time.

In [19], an algorithm that aims to automatically determine the number of classifiers and stages for constructing a boosted ensemble was proposed. While a greedy optimization algorithm was employed in [3], Brubaker et al. proposed an algorithm for determining the number of weak classifiers and training each node classifier of a cascade by selecting operating points within a receiver operator characteristic (ROC) curve [20]. The solved optimization problem using linear programs that maximize the detection rates while satisfying the constraints of false positive rates [19].

Although the original four types of Haar-like features are sufficient to encode upright frontal face images, other types of features are essential to represent more complex patterns (e.g., faces in different pose) [10, 11, 12, 15]. Most systems take a divide-and-conquer strategy and a face detector is constructed for a fixed pose, thereby covering a wide range of angles (e.g., yaw and pitch angles). A test image is either sent to all detectors for evaluation, or to a decision module with a coarse pose estimator for selecting the appropriate trees for further processing. The ensuing problems are how the types of features are constructed, and how the most important ones from a large feature space are selected. More generalized Haar-like features are defined in [10, 11] in which the rectangular image

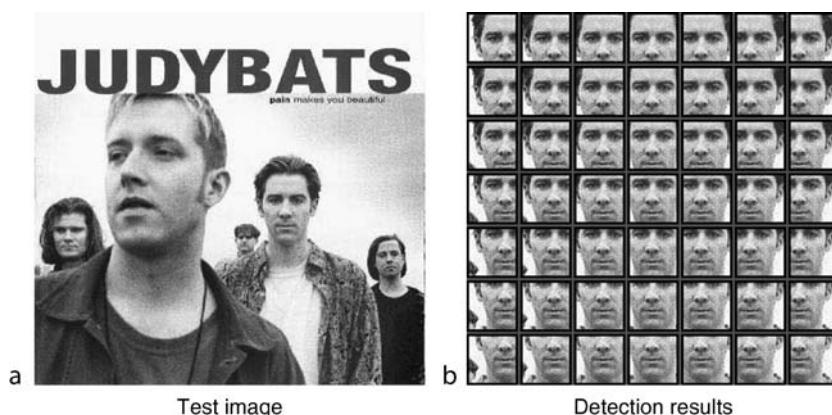
regions are not necessarily adjacent, and furthermore the number of such rectangular blocks is randomly varied [10]. Several greedy algorithms have been proposed to select features efficiently by exploiting the statistics of features before training boosted cascade classifiers [17].

There are also other fast face detection methods that demonstrate promising results, including the component-based face detector using Naive Bayes classifiers [4], the face detectors using support vector machines [7, 21, 22], the Anti-face method [23] which consists of a series of detectors trained with positive images only, and the energy-based method [24] that simultaneously detects faces and estimates their pose in real time.

Quantifying Performance

There are numerous metrics to gauge the performance of face detection systems, ranging from detection frame rate, false positive/negative rate, number of classifier, number of feature, number of training image, training time, accuracy, and memory requirements. In addition, the reported performance also depends on the definition of a “correct” detection result [2, 5]. Figure 2 shows the effects of detection results versus different criteria, and more discussions can be found in [2, 5].

The most commonly adopted method is to plot the ► ROC curve using the de facto standard MIT + CMU data set [2] which contains frontal face images. Another data set from CMU contains images with faces that vary in pose from frontal to side view [4]. Note that



Face Detection. Figure 2 Detection results depend heavily on the adopted criteria. Suppose all the sub-images in (b) are

although the face detection methods nowadays have impressive real-time performance, there is still much room for improvement in terms of accuracy. The detected faces returned by state-of-the-art algorithms are often a few pixels (around 5) off the “accurate” locations, which is significant as face images are usually standardized to 21×21 pixels. While such results are the trade-offs between speed, robustness, and accuracy, they inevitably degrade the performance of any biometric applications using the contents of detected faces. Several post-processing algorithms have been proposed to better locate faces and extract facial features (when the image resolution of the detected faces is sufficiently high) [25].

Applications

As face detection is the first step of any face processing system, it finds numerous applications in face recognition, face tracking, facial expression recognition, facial feature extraction, gender classification, clustering, attentive user interfaces, digital cosmetics, biometric systems, to name a few. In addition, most of the face detection algorithms can be extended to recognize other objects such as cars, humans, pedestrians, and signs, etc. [5].

Summary

Recent advances in face detection have created a lot of exciting and reasonably robust applications. As most of the developed algorithms can also be applied to other problem domains, it has broader impact than detecting faces in images alone. Future research will focus on improvement of detection precision (in terms of location), online training of such detectors, and novel applications.

Related Entries

- ▶ [Biometric Algorithm](#)
- ▶ [Ensemble Learning](#)
- ▶ [Face Tracking](#)
- ▶ [Face Recognition, Overview](#)
- ▶ [Facial Expression Recognition](#)
- ▶ [Machine-Learning](#)
- ▶ [Supervised Learning Surveillance](#)

References

1. Sung, K.K., Poggio, T.: Example-based learning for view-based human face detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(1), 39–51 (1998)
2. Rowley, H., Baluja, S., Kanade, T.: Neural network-based face detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(1), 23–28 (1998)
3. Viola, P., Jones, M.: Robust real-time face detection. *Int. J. Comput. Vision* **57**(2), 137–154 (2004)
4. Schneiderman, H., Kanade, T.: Object detection using the statistics of parts. *Int. J. Comput. Vision* **56**(3), 151–177 (2004)
5. Yang, M.H., Kriegman, D., Ahuja, N.: Detecting faces in images: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(1), 34–58 (2002)
6. Mohan, A., Papageorgiou, C., Poggio, T.: Example-based object detection in images by components. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(4), 349–361 (2001)
7. Heisele, B., Serre, T., Poggio, T.: A component-based framework for face detection and identification. *Int. J. Comput. Vision* **74**(2), 167–181 (2007)
8. Fleuret, F., Geman, D.: Coarse-to-fine face detection. *Int. J. Comput. Vision* **41**(12), 85–107 (2001)
9. Papageorgiou, C., Poggio, T.: A trainable system for object recognition. *Int. J. Comput. Vision* **38**(1), 15–33 (2000)
10. Dollar, P., Tu, Z., Tao, H., Belongie, S.: Feature mining for image classification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2007)
11. Li, S., Zhang, Z.: Floatboost learning and statistical face detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(9), 1112–1123 (2004)
12. Huang, C., Ai, H., Li, Y., Lao, S.: High-performance rotation invariant multiview face detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(4), 671–686 (2007)
13. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and application to boosting. *J. Comput. Syst. Sci.* **55**(1), 119–139 (1997)
14. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *Ann. Stat.* **28**(2), 337–407 (2000)
15. Jones, M., Viola, P.: Fast multi-view face detection. Technical Report TR2003-96, Mitsubishi Electrical Research Laboratories (2003)
16. Viola, P., Jones, M.: Fast and robust classification using asymmetric AdaBoost and a detector cascade. In: Advances in Neural Information Processing Systems, pp. 1311–1318 (2002)
17. Wu, J., Brubaker, S.C., Mullin, M., Rehg, J.: Fast asymmetric learning for cascade face detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(3), 369–382 (2008)
18. Pham, M.T., Cham, T.J.: Online learning asymmetric boosted classifiers for object detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2007)
19. Brubaker, S.C., Wu, J., Sun, J., Mullin, M., Rehg, J.: On the design of cascades of boosted ensembles for face detection. *Int. J. Comput. Vision* **77**(1–3), 65–86 (2008)
20. Provost, F., Fawcett, T.: Robust classification for imprecise environments. *Mach. Learn.* **42**(3), 203–231 (2001)

21. Oren, M., Papageorgiou, C., Sinha, P., Osuna, E., Poggio, T.: Pedestrian detection using wavelet templates. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 193–199 (1997)
22. Romdhani, S., Torr, P., Schölkopf, B., Blake, A.: Computationally efficient face detection. In: Proceedings of the Eighth IEEE International Conference on Computer Vision, vol. 2, pp. 695–700 (2001)
23. Keren, D., Osadchy, M., Gotsman, C.: Antifaces: A novel fast method for image detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(7), 747–761 (2001)
24. Osadchy, M., LeCun, Y., Miller, M.: Synergistic face detection and pose estimation with energy-based models. *J. Mach. Learn. Res.* 1197–1214 (2007)
25. Ding, L., Martinez, A.: Precise detailed detection of faces and facial features. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2008)

Face Device

MASSIMO TISTARELLI

Computer Vision Laboratory, University of Sassari,
Piazza Duomo, 6 Alghero, Italy

Synonyms

Face acquisition; Face camera; Video camera; Visual sensor

Definition

A face device is a system to acquire a set of digital data samples representing a human face. As the human face is a complex 3D object, the data can be in several forms: a 2D image where the gray levels of the ► pixels represent the projected reflectance of the face surface under visible illumination; a 2D image where the gray levels of the pixels represent the projected reflectance of the face surface illuminated with an active source; a 2D thermal image representing the heat emitted by the face surface; 3D samples of the surface structure.

Face devices can be distinguished on the basis of the data dimension if it is active or passive. Face devices can be passive, i.e., based on the passive reflectance of ambient light by the body, or active, i.e., associated with an energy emitter and a sensor to

capture the energy reflected by the face. The data captured can be either in 2D or 3D form.

A face device can be based on different technologies, depending upon the data to be captured and the signal to be obtained. The most applied face devices include a video ► camera to capture 2D images of the face and a digitizer to sample and quantize the analog signal generated by the camera. Different face devices deliver different signals to be digitized into 2D or 3D data. The data captured can be stored under different file formats for subsequent processing.

Introduction

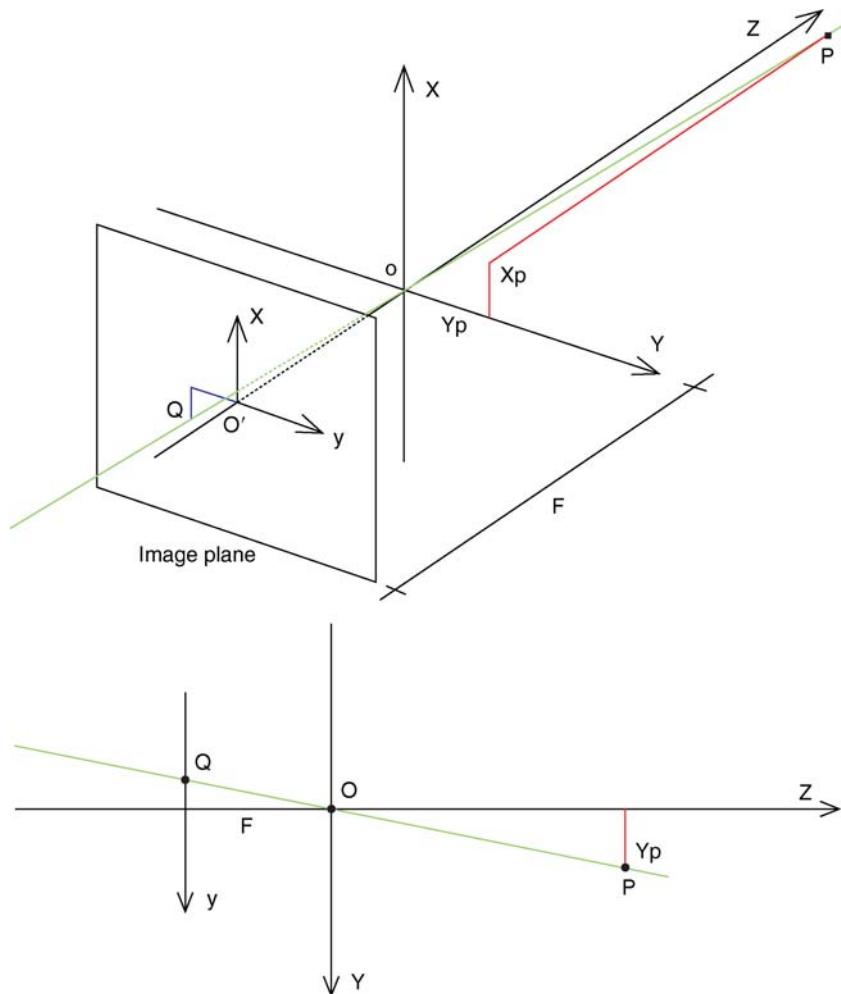
Current face biometric systems are based on the acquisition and processing of image data, representing a human face. A face acquisition device is typically a video camera capable of acquiring single images or video streams of data, representing a face. As the face is a 3D object, the acquired data can represent either the 2D projection of the face reflectance on the image plane or a set of 3D samples of the face structure, possibly with the associated reflectance. In the former case, a conventional video camera can be used to acquire images of face. In the latter case, a more complex 3D acquisition device must be applied.

2D Face Devices

A conventional camera acquires the image data as a reflectance of the imaged scene. The face points are recorded as the geometrical projection of the 3D points on the face surface onto the 2D image plane (Fig. 1).

Several video cameras exist that are capable of capturing either single images or video streams from the viewed scene. The most critical parts of the camera are the acquisition sensor and the lenses.

Charge-coupled device (CCD) and complementary metal oxide semiconductor (CMOS) image sensors are two different technologies for capturing images digitally; current commercial camera adopt either of these. Both types of sensors convert light into electric charge and process it into electronic signals. In a ► CCD sensor, charge of every pixel is transferred through a very limited number of output nodes



Face Device. **Figure 1** Geometry of the pin hole camera model. (Top) 3D sketch of the projection of point P in space on the image pixel Q . (Bottom) 2D projection of the Y - Z plane.

(often just one) to be converted to voltage, buffered, and sent off-chip as an analog signal. All the pixel can be devoted to capture light, and the output's uniformity (a key factor in image quality) is high. In a ►CMOS sensor, each pixel has its own charge-to-voltage conversion, and the sensor often includes amplifiers, noise-correction, and digitization circuits, so that the chip outputs digital bits. These other functions increase the design complexity and reduce the area available for light capture. With each pixel doing its own conversion, the uniformity is low. But the chip can be built to require less off-chip circuitry for basic operation.

The CMOS pixel solves the speed and scalability issues of the CCD sensor. They consume far less power than a CCD, have less image lag, and can be fabricated

on much cheaper and more available manufacturing lines. Unlike CCDs, CMOS sensors can combine both the image sensor and the image processing functions within the same integrated circuit. CMOS imagers still suffer from higher ►fixed-pattern noise than CCDs, but active pixel sensors are catching up with respect to noise, dynamic range, and responsivity. CMOS sensors have become the technology of choice for many consumer applications, most significantly, the burgeoning cell phone camera market [1].

The technology of the sensor and the capturing device determines several properties of the captured signal such as the following:

1. *Image resolution.* This is related to both the active elements on the imager sensor and the sampling

device used to digitize the signal. Even though solid state sensors are used in digital cameras, they produce an analog video signal. As a consequence, the captured image resolution strongly depends on the sampling frequency of the digitization device. Other factors affecting the image resolution are the file standard format adopted for the image storage and the image processing application required to postprocess the face images.

2. *Responsivity*. The amount of signal the sensor delivers per unit of input optical energy. CMOS imagers are marginally superior to CCDs, in general, because gain elements are easier to be placed on a CMOS image sensor. This affects the illumination level required to capture a face image with a sufficient contrast level.
3. *Dynamic range*. The ratio of a pixel's saturation level to its signal threshold. CCD sensors are much better than CMOS in this regard. Some CMOS sensors deliver 8 bit per pixel intensities, corresponding to 128 real level variations. As a consequence, the information content in the image features is half than what is expected. A higher dynamic range implies a higher image contrast even at low illumination levels and the possibility to grab finer details. A gray level quantization of 8 bit per pixel is generally sufficient for capturing good quality face images. The sensor dynamic range can be crucial when acquiring color images. In this case, the color quantization may influence the information content in the face image itself, especially if a low bit rate (with less than 8 bit per color channel) is used for color coding.
4. *Sensitivity to noise* (signal to noise ratio – SNR). The three primary broad components of noise in a CCD imaging system are photon noise (results from the inherent statistical variation in the arrival rate of photons incident on the CCD), dark noise (arises from statistical variation in the number of electrons thermally generated within the silicon structure of the CCD), and read noise (a combination of system noise components inherent to the process of converting CCD charge carriers into a voltage signal for quantification, and the subsequent processing including the analog-to-digital (A/D) conversion). A further useful classification distinguishes noise sources on the basis of whether they are temporal or spatial. CCDs still enjoy significant noise advantages

over CMOS imagers because of quieter sensor substrates (less on-chip circuitry), inherent tolerance to bus capacitance variations, and common output amplifiers with transistor geometries that can be easily adapted for minimal noise.

5. *Uniformity*. The consistency of response for different pixels under identical illumination conditions. Spatial wafer processing variations, particulate defects, and amplifier variations create nonuniformities in light responses. It is important to make a distinction between uniformity under illumination and uniformity at or near dark. CMOS imagers were traditionally much worse than CCDs under both regimes. New on-chip amplifiers have made the illuminated uniformity of some CMOS imagers closer to that of CCDs, sustainable as geometries shrink. This is a significant issue in high-speed applications, where limited signal levels mean that dark nonuniformities contribute significantly to overall image degradation.
6. *Shuttering*. The ability to start and stop exposure arbitrarily. It is a standard feature of virtually all consumer and most industrial CCDs, especially interline transfer devices, and it is particularly important in machine vision applications. CCDs can deliver superior electronic shuttering, with little fill-factor compromise, even in small-pixel image sensors. Implementing uniform electronic shuttering in CMOS imagers requires a number of transistors in each pixel. In line-scan CMOS imagers, electronic shuttering does not compromise fill factor, because shutter transistors can be placed adjacent to the active area of each pixel. In area-scan (matrix) imagers, uniform electronic shuttering comes at the expense of fill factor, because the opaque shutter transistors must be placed in what would otherwise be an optically sensitive area of each pixel. A uniform synchronous shutter, sometimes called a nonrolling shutter, exposes all pixels of the array at the same time. Object motion stops with no distortion, but this approach reduces the pixel area because it requires extra transistors in each pixel. Users must choose between low fill factor and small pixels on a small, less-expensive image sensor, or large pixels with much higher fill factor on a larger, more costly image sensor.
7. *Sampling speed*. This is an area in which CMOS arguably delivers better performances over CCDs,

because all camera functions can be placed on the image sensor. With one die, signal and power trace distances can be shorter, with less inductance, capacitance, and propagation delays. To date, CMOS imagers have established only modest advantages in this regard, largely because of early focus on consumer applications that do not demand notably high speeds compared with the CCD's industrial, scientific, and medical applications. Both the sampling and shuttering speed are important when capturing video streams of faces. In this case, it is important to ensure the image stability and minimize the motion smear induced by either the motion of the camera or the face. This requires to tune the camera sampling frequency to the motion speed induced in the image sequence. If the face is very close to the camera, small motions can induce large and fast displacements on the image, thus producing motion smear. At a larger distance (above 50 cm), a standard sampling frequency of 50 or 60Hz is generally sufficient. In many low-cost devices, the sampling frequency depends on the time required to transmit the signal from the device to the frame buffer. Therefore, only low resolution images can be captured at high sampling frequencies. On the other hand, if a high, nonstandard sampling frequency is required to capture stable images with fast motions, the reduced exposure time requires a higher sensitivity of the sensor to preserve a high SNR.

8. *Windowing*. One unique capability of CMOS technology is the ability to read out a portion of the image sensor. This allows elevated frame or line rates for small regions of interest. This is an enabling capability for CMOS imagers in some applications, such as high-temporal-precision face tracking in the subregion of an image. CCDs generally have limited abilities in windowing.
9. *Antiblooming*. The ability to gracefully drain localized overexposure without compromising with the rest of the image in the sensor. CMOS generally has natural blooming immunity. CCDs, on the other hand, require specific engineering to achieve this capability; many CCDs that have been developed for consumer applications do, but those developed for scientific applications generally do not.
10. *Biassing and noise*. CMOS imagers have a clear edge in this regard. They generally operate with a

single bias voltage and clock level. Nonstandard biases are generated on-chip with charge pump circuitry isolated from the user unless there is some noise leakage. CCDs typically require a few higher-voltage biases, but clocking has been simplified in modern devices that operate with low-voltage clocks.

The camera optics determines the general image deformation, the depth of the field, and the amount of blurring in the image. The lenses must be chosen carefully according to the acquisition scenario. The ► **focal length** must be set to provide a sufficient ► **depth of field (DOF)** to always keep the subject's face in focus. If the range of distances is very large, a motorized lens can be used to dynamically keep the face in focus. Otherwise, a shorter focal length lens, with a larger depth of field, can be used at the expenses of an increase in the image distortion.

A 2D camera can be modeled with several parameters [2], including the following:

1. The (X, Y, Z) position of the center of the camera lens
2. The focal length
3. The orientation of the sensor's plane
4. The aperture or ► **field of view** (X_f, Y_f)
5. The physical x and y dimensions of each pixel on the sensor
6. The normal to the focal plane
7. The lenses properties

Many parameters can be neglected in the *pin hole* camera model. This is a simplified model where the physical parameters are reduced to five virtual parameters, namely the following:

1. The focal length F
2. The pixel width and height $\delta x, \delta y$
3. The x and y coordinates of the optical center (x_c, y_c)

Assuming the pin hole camera model, the (x, y) projection on the image plane of a 3D point (X, Y, Z) can be represented as (refer to Fig. 1)

$$\begin{aligned} x &= F_x \frac{X}{Z}, \\ y &= F_y \frac{Y}{Z}, \end{aligned} \quad (1)$$

where F_x and F_y represent the two values of the focal length, which take into account the image aspect ratio.

The pin hole model cannot take into account several effects of the misalignment of the sensor with the lenses, not the lens aberration or the image deformation due to the focal length. However, when high accuracy is required or when low-end cameras are used, additional effects have to be taken into account.

The failure of the optical system to bring all light rays received from a point object to a single image point or to a prescribed geometric position should then be taken into account. These deviations are called aberrations. Many types of aberrations exist (e.g., astigmatism, chromatic aberrations, spherical aberrations, coma aberrations, curvature of field aberration, and distortion aberration). It is outside the scope of this work to discuss them all. The interested reader is referred to the work of Willson [3] and to the photogrammetry literature [4].

Many of these effects are negligible under normal acquisition circumstances. Radial distortion, however, can have a noticeable effect for shorter focal lengths. Radial distortion is a linear displacement of image points radially to or from the center of the image, caused by the fact that objects at different angular distance from the lens axis undergo different magnifications. It is possible to cancel most of this effect by ► [Face Warping](#) the image.

Active 2D Face Devices

Within the general class of 2D face devices, active devices rely on the possibility to use an active source of energy to radiate the subject's face. Among them, the most commonly used are the near infrared cameras. These cameras have normal optics but the sensor (either CMOS or CCD) is sensitive to a wavelength spectrum between 0.7 and 1.1 μm . To perform image acquisition, the subject's face to be captured must be illuminated by an infrared illuminator. Given the sensitivity response curve of the near infrared sensor, the pixel intensities are almost exclusively due to the reflection of the infrared light on the face skin. An example is presented in Fig. 2. As a consequence, a remarkable advantage of this face acquisition device is the relative insensitivity to changes in environmental illumination.

3D Face Devices

Another category of face device are those aimed at acquiring the 3D shape information of the face. There



Face Device. Figure 2 Sample image acquired with a near-infrared camera.

are several technologies applied to produce 3D cameras for face acquisition. They can be broadly grouped in the following categories:

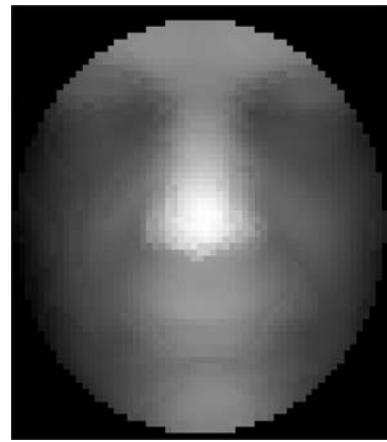
1. *Stereo triangulation cameras.* A pair of stereo cameras is used for determining the depth to points on the face, for example, from the center point of the line between their focal points. To solve the depth measurement problem using stereo cameras, it is necessary to first find corresponding points in the two images. Solving the correspondence problem is one of the main problems when using this type of technique. As a consequence, range imaging based on stereo triangulation can usually produce reliable depth estimates only for a subset of all points visible in both cameras. The advantage of this technique is that the measurement is more or less passive; it does not require special arrangements in terms of scene illumination.
2. *Light stripe triangulation.* Illuminating the face with a light stripe creates a reflected line as seen from the light source. From any point out of the plane of the stripe, the line will typically appear as a curve, the exact shape of which depends both on the distance between the observer and the light source and on the distance between the light source

and the reflected points. By observing the reflected sheet of light using a camera (often a high resolution camera) and knowing the positions and orientations of both camera and light source, it is possible to determine the distance between the reflected points and the light source or camera. By moving either the light source (and normally also the camera) or the scene in front of the camera, a sequence of depth profiles of the scene can be generated. These can be represented as a 2D range image. The most common cameras are based on the projection of an invisible and unharful laser light stripe. The light stripes projected along the face surface are captured by a conventional camera. The distortion in the light stripes induced by the face shape is computed to infer the 3D structure of the surface.

3. *Time-of-flight laser scanner.* The time-of-flight 3D laser scanner is an active scanner that uses laser light to probe the subject. At the heart of this type of scanner is a time-of-flight laser range finder. The laser range finder finds the distance of a surface by timing the round-trip time of a pulse of light. A laser is used to emit a pulse of light and the amount of time before the reflected light is seen by a detector. Since the speed of light c is known, the round-trip time determines the travel distance of the light, which is twice the distance between the scanner and the surface.

Inspite of the camera and sensor technology, the produced image is either a depth map, a collection of 3D points in space, or a set of 3D features representing the 3D structure of the acquired face. A sample depth map of a face is shown in Fig. 3. The most frequently used representations for the acquired 3D data can be listed as follows:

1. *Point cloud.* A large number of 3D points that are sampled from the surface of the face are stored.
2. *3D mesh.* Triangulation is used to produce a mesh from the point cloud. This is a more compact representation. Range images – One or more 2D range images can be stored, especially if the range data are taken from a single perspective.
3. *Feature sets.* There are different features that one can derive and store for each face. Typical features are landmark locations (nose tip, eyes, corners of the mouth, etc.), surface normals, curvatures, profile features, shape indices, depth and/or colour



Face Device. Figure 3 Sample depth face image. The gray levels are inversely proportional to the distance of the face surface from the camera [5].

histograms, edges, and subspace projection coefficients (PCA and LDA are frequently used).

The point cloud representation is the most primitive 3D information provided by a 3D camera. The 3D-RMA is an example of a database of 3D face models represented by clouds of points [6]. For long time, it has been the only publicly available database, even if its quality is rather low. Meshes are obtained by triangulation. These are more structured and easier to deal with. Data in the form of meshes are more available today, but in most cases the mesh databases are proprietary. Usually, more than one representation is used in a single algorithm. Texture information, if available, is generally stored for each 3D point or triangle. A sample 3D face image with the associated reflectance map is shown in Fig. 4.

Summary

A face device is a system to acquire a set of digital data samples representing a human face. As the human face is a complex 3D object, the data can be in several forms, from a 2D image to a complex 3D representation. The principal component of a face device is a digital camera, which acquires images either for a direct 2D representation or to build a 3D representation of the face shape. Different cameras offer variable performances, in terms of quality of the signal,



Face Device. **Figure 4** Sample 3D face image and projected 2D intensity values from the face recognition grand challenge (FRGC) [5] database.

sensitivity to different light spectral components, and capturing speed. The proper imaging device must be carefully chosen for the application scenario. The ambient illumination level, the required level of detail, the effects of noise, and the motion speed of the objects in the scene must all be carefully considered.

Related Entries

- ▶ Acquisition
- ▶ Authentication
- ▶ Enrollment
- ▶ Identification
- ▶ Verification

References

1. Litwiller, D.: CCD vs. CMOS: facts and fiction. *Photonics Spectra*, pp. 151–154 (2001)
2. Blais, F.: Review of 20 years of range sensor development. *J. Electron. Imaging* **13**(1), 231–240 (2004)
3. Willson, R., Shafer, S.: What is the center of the image? *J. Opt. Soc. Am. A* **11**(11), 2946–2955 (1994)
4. Slama, C.: Manual of Photogrammetry/ American Society of Photogrammetry, Falls Church, VA, USA, 4th edn. (1980)
5. Phillips, J.J., Flynn, P., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Jaesik, M., Worek, W.: Overview of the face recognition grand challenge. In *Proceedings CVPR05*, pp. 947–954 (2005)
6. Beumier, C., Achery, M.: Automatic 3D face authentication. *Image Vision Comput.* **18**(4), 315–321 (2000)

Face Identification

- ▶ Face Recognition, Thermal
- ▶ Forensic Evidence of Face

Face Image Data Interchange Formats

- ▶ Face Image Data Interchange Formats, Standardization

Face Image Data Interchange Formats, Standardization

PATRICK GROTHÉR, ELHAM TABASSI
National Institute of Standards and Technology, USA

Synonym

Face image data interchange formats

Definition

Openly documented data structures for universally interpretable interchange of facial imagery.

Biometric data interchange standards are needed to allow the recipient of a data record to successfully process data from an arbitrary producer. This defines biometric interoperability and the connotation of the phrase “successfully process” is that the sample, in this case, a facial image record, can be accurately identified or verified. This can be achieved only if the data record is both syntactically and semantically conformant to a documentary standard.

Introduction

Facial image standards are perhaps the oldest documented biometric data standards. Predating even the fingerprint, the facial image has been mandated for identity documents since at least the World War I when several European governments saw the need for a facial photograph to serve as the core element in the cross-border identity verification application. Of course the data record was simply an analog paper printed photograph - the advent of fully automatic face recognition algorithms and the need for digital images was at least 70 years distant [1, 2]. However the intention remains the same: to support (human or machine) verification of an individual via a high quality standardized image.

Roles

The use of face imagery for recognition is ubiquitous in applications where a human does the recognition. This rests on three factors: The ability of humans to recognize faces; the almost universal availability of the face In some cultures the face is covered or painted, and in such cases modalities such as iris or hand geometry are dominant.; and the availability of cameras and printers. The result is that face images, printed on passports, drivers' licenses, credit cards, and other tokens, have been the primary biometric element for human verification for many years.

Nowadays with the advent and maturation of technologies for automated face recognition, the use of the face for verification [3] is but one component of a larger marketplace in which commercial systems have been both piloted and fully deployed for identification applications such as watch-list surveillance [4] and duplicate detection (e.g., for drivers licenses, or visas). In addition the law enforcement community has for

years taken mugshot images and, while these are often only used for human identification, they are being used operationally [5].

The common theme among all is that recognition accuracy is critically sensitive function of the quality of the image, where quality here refers to the photometric and geometric properties of the image. The former include contrast, exposure, and uniformity of lighting; the latter refers to the size of the image and the angular orientation of the face to the viewing direction. The effect of non-idealities in these areas has been quantified extensively and there is an enormous literature documenting research in how to improve the robustness and invariance of the algorithms to variations in these quantities. In parallel, there has been a concerted effort by groups of vendors, users, governmental organizations, and academics to develop standards that establish a baseline for the acquisition and quality of the captured images.

It is no coincidence that the largest marketplace for face recognition technologies today is in those applications where the quality is most highly controlled, namely passports and visas, where the photographers and the subjects who pay them, are positively motivated to provide good conformant images.

In a more general sense, formal face images standards also serve to do what many other data format standards do: they define a parseable record that allows syntactic interoperability. This creates a foundation for a marketplace of off-the-shelf products, and is a necessary condition to achieve supplier independence, and to avoid vendor lock-in. It is perhaps surprising that in a world where many raster image formats are open and standardized [6–8] it remains common for images to be retained in a fully proprietary (i.e., unpublished) format. Such practice may be acceptable *within* an application, but is a serious impediment once cross-organizational interchange of data is required. This is the essence of interoperability which allows modular integration of products without compromising architectural scope, and it facilitates the upgrade process and thereby mitigates against obsolescence.

The business implications of these benefits are many. A good standard, well implemented, may create entirely new markets (e.g., e-Passports include face image records). On the other hand, robust standards tend to lead to competition and reduced profit margins. This process, commoditization, is an inhibitory factor for many technology companies that balance the

promise of new or expanded marketplaces against reduced barriers to entry for competitors. The decision is determined by the amount of intellectual property that a standard allows suppliers to hide behind its implementation. From the user perspective, standards may serve to enhance competition and performance. For example, face image standards (primarily ISO/IEC 19794-5 [9]), which are currently being mandated in a number of large government and international programs, specify image formats without requiring particular equipment or matching algorithms.

This is the motivation for formal published consensus standards.

Standards do not in and of themselves assure interoperability. Specifically, when a standard is not fully prescriptive, or it allows for optional content, then two implementations that are both exactly conformant to the standard may still not interoperate. This situation may be averted by applying further constraints on the application of the standard. This is done by means of “application profile” standards which formally call out the needed base standards and refine their optional content and interpretation.

History of Face Standardization

The current face standards descend from standardization efforts that began in the mid 1990s. These were driven in large part by the needs of the United States’ Federal Bureau of Investigation who sought to establish uniform standards for State and local law enforcement authorities submitting images to them.

Referring to [Table 1](#), the first standard, approved in April 1997, established the syntax of a record denoted “Type 10.” The image data it required was either in raw grayscale format or, if compressed, in the then draft JPEG/JFIF standard [6]. Concurrently NIST established procedures for the geometric and photometric properties of images and published its recommendations in September 1997. These were extended and modified, and incorporated, in 2000, into both the American Association of Motor Vehicle Administrators standard for drivers licenses, and the revision of the FBI’s original biometric data specifications.

These standards formed the basis for the subsequent development of the national INCITS 385:2004 standard in 2004, which in turn begat the full ISO/IEC 19794-5 International Standard in 2005. (At the time of writing the standard is under amendment to regulate the acquisition process, and to establish a container for three dimensional data.) A substantially revised standard which would include these changes (and others) is likely to be completed late in the decade.

The ISO/IEC 19794-5 Face Image Standard

The ISO/IEC 19794-5:2005 standard is the fifth part of a multipart biometric data interchange format standard. The standard is organized by modality, and other parts cover fingerprint images, irises, and hand geometry among many others. The Part 5 standard is the most widely implemented, most actively developed, and most modern face standard. Its content drove the

Face Image Data Interchange Formats, Standardization. [Table 1](#) The evolution of contemporary face image standards

Date	Title of Standard
04/1997	Addendum To ANSI/NIST-CSL 1-1993 (adding Mugshots, scars, marks and tattoos)
09/1997	NIST Best Practice Recommendation for the Capture of Mugshots
06/2000	AAMVA National Standard for the Driver License/Identification Card
09/2000	ANSI/NIST-ITL 1-2000 - Data Format for the Interchange of Fingerprint, Facial, & Scar Mark & Tattoo Information - Type 10
05/2004	INCITS 385:2004 - Face Recognition Format for Data Interchange
06/2005	ISO/IEC 19794-5:2005 - Face Image Data
04/2007	ANSI/NIST-ITL 1-2007 - Data Format for the Interchange of Fingerprint, Facial, & Other Biometric Information - Type 10
06/2007	ISO/IEC 19794-5/Amd 1 - Conditions for Taking Photographs for Face Image Data
2009 (Est)	ISO/IEC 19794-5/Amd 2 - Three Dimensional Face Image Data Interchange Format

revision of the Type 10 record of the ANSI/NIST ITL 1-2007 described in section. While the ISO standard is under revision, with publication due late in the decade, the existing 2005 standard has been called out for some major identity management applications. The foremost of these is the e-Passport, which the International Civil Aviation Organization formalized in its ICAO 9303 standard. This points to ISO/IEC 19794-5 as the mandatory globally interoperable data element for ISO/IEC 14443 contactless chip passports.

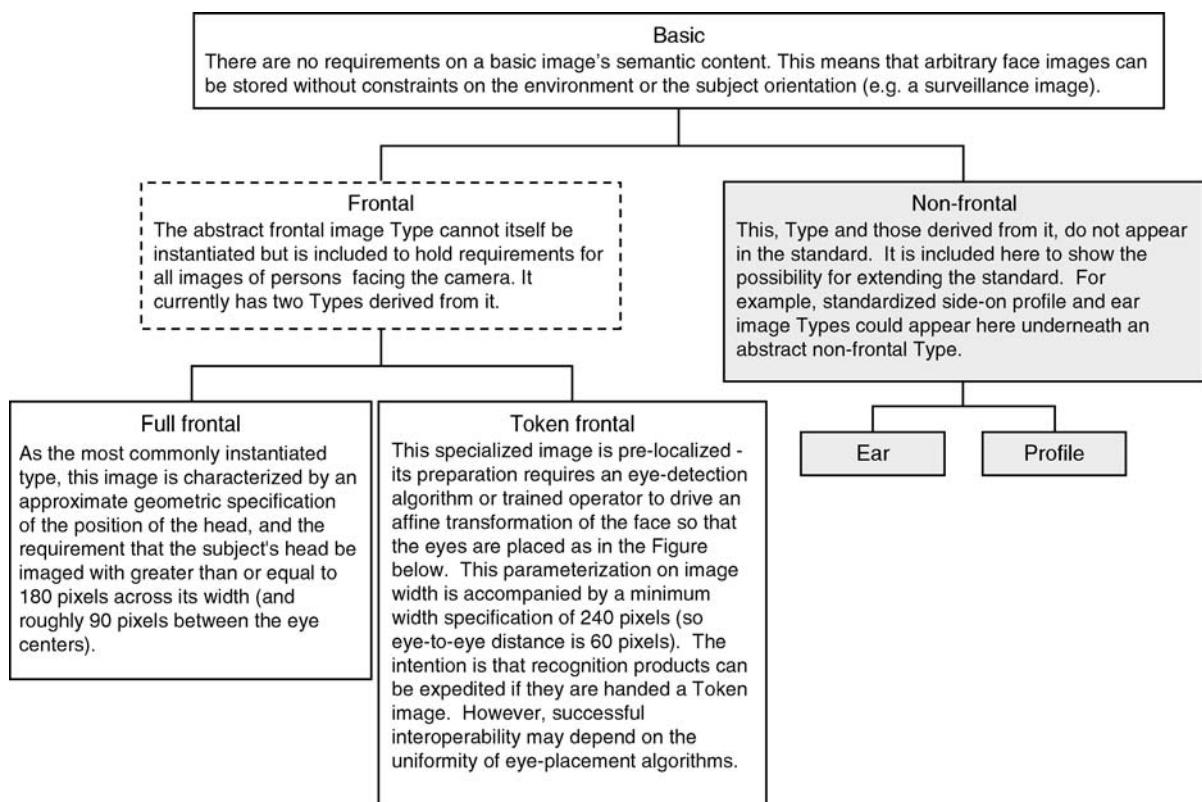
The face standard defines a binary record structure for the storage of one or more face images. It establishes requirements on the syntax and semantic content of the structure. These requirements are stated in terms of the following four categories.

- *Format*: These requirements detail the syntactic arrangement of the data elements in the record.
- *Scene*: These requirements regulate variables such as pose, expression, shadows on the face, the wearing of eye glasses.
- *Photographic*: These requirements concern correct exposure of the subject, distortion, focus, and depth of field.

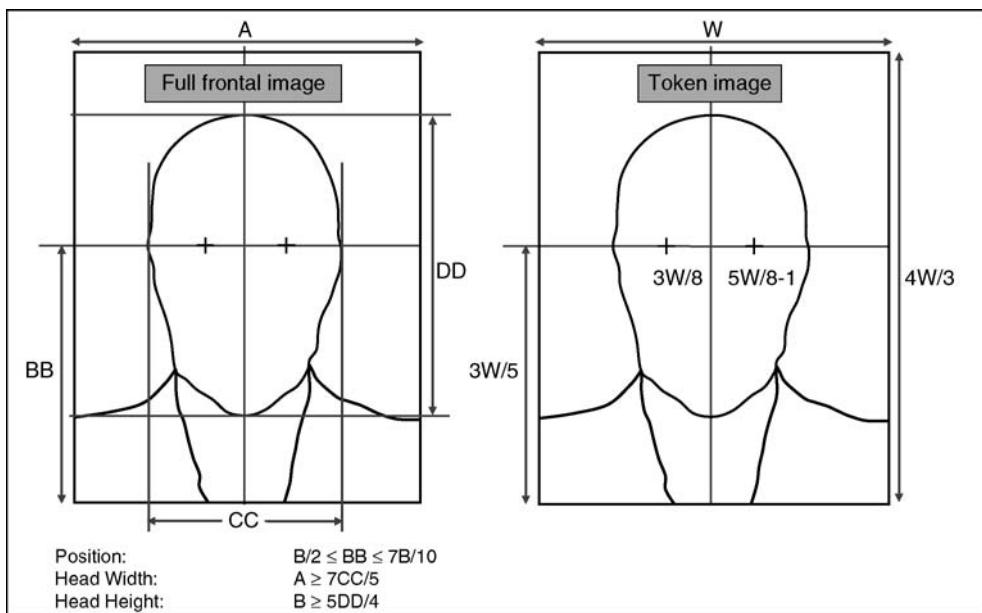
- *Digital*: The requirements include specifications for dynamic range, color space, pixel aspect ratio, and video interlacing.

The standard imposes these requirements incrementally: Fig. 1 shows that the useful frontal image types inherit from parent types and add requirements. This object oriented design allows for future specialized types to be added, including 3D frontal types. In addition the standard establishes two geometric position specifications for the face. These are shown in Fig. 2. The tighter specification, for known as the token Frontal, requires detection of the eye coordinates and of fine transformation of the image.

The record includes fields for expression, eye-color, hair color, and gender. It optionally allows the inclusion of ISO/IEC 14496-2 MPEG 4 feature points. The standard includes various quality related requirements. For example the pose angle is required to be $\pm 5\text{deg}$, and there must be at least 7 bits of greylevel information on the face. Conformance to these requirements will elevate face recognition performance. Once an image is acquired a test of its conformance to the standard's specifications requires some non-trivial



Face Image Data Interchange Formats, Standardization. Figure 1 Inherited types of the ISO/IEC 19794-5 face image standard.



Face Image Data Interchange Formats, Standardization. **Figure 2** Geometries of the ISO/IEC 19794-5 frontal face images.

image analyses. A number of software products have been developed to “box-check” ISO conformance and to prepare the standardized record.

The ANSI/NIST ITL 1-2007 Type 10 Record

Since its initial development in the early 1990s, the so-called ANSI-NIST standard has been very widely implemented and used within and between the law enforcement communities of the United States and the many other countries. Its primary use is the transmission of fingerprint data from the State and Local authorities to central automated fingerprint identification systems, primarily those operated by the Federal Bureau of Investigation. The ANSI/NIST standard includes defined *Types* for the major biometric modalities. The standard is multimodal in that it allows a user to define a transaction that would require, for example, fingerprint data as Type 14, a facial mugshot as Type 10, and the mandatory header and metadata records Type 1 and 2. These are linked with a common numeric identifier.

Of concern here, since its development in 1997, is the Type 10 record. It supports storage not just of face images, but also those of scars, marks, and tattoos, with the particular type of content being recorded in the “image type” field of the header.

Unlike the ISO standard’s fixed binary structure, the Type 10 has a tag-value structure in which a three letter code begins a field. The mandatory fields are: Record length, image designation code (identifier linking, say, Type 14 finger + Type 10 face records), image type (face or otherwise), the source agency (e.g., local police department), the capture date, the width, height and scanning resolution, the compression algorithm, color space, and the subject acquisition profile. This latter field encodes, essentially, the conformance of the image to particular capture specifications. These are either established elsewhere [9–12] or introduced in the standard.

The optional fields are: pose category (frontal, profile, other), actual pose angles, whether the subject was wearing headwear or eyewear, the camera type, a quality value and its source, the eye and hair color, facial expression, eye and nostril locations and MPEG 4 feature points, and whether the capture was attended or automatic. The last field contains the image data itself, which is either an uncompressed raw greyscale or color image, or a JPEG, JPEG 2000 or PNG encoded image.

Amendment 1 to ISO/IEC 19794-5:2005

The 2007 amendment is an informative Annex to the base 2005 face standard. It is written to provide expert

guidance for the photography of faces particularly by owners and operators of studios, photo stores or other organizations producing or requiring either printed photographs or digital images that would conform. It is intended to assist in the production of images that are conformant to the frontal type requirements of the base standard.

The standard regulates the subject, lighting, and camera placement for three kinds of face acquisition environments listed here in the order of increasing space constraints and non-ideality: a photo studio (e.g., for a passport), a registration desk (e.g., for a driving license), and a photo-booth. For each of these the standard addresses camera-subject positioning (in terms of distance, height, focus, and depth of field), exposure (in terms of F-stops and shutter speed), and illumination (in terms of number, type and placement of lights). The document also provides guidance on printing and scanning of paper photographs.

Amendment 2 to ISO/IEC 19794-5:2005

A second amendment is currently under preparation. This is aimed at standardizing a container and specifications for images that include three dimensional shape information of the human head. An initial effort within the United States, INCITS 385:2004 Amendment 1, allowed a 2D face image to be accompanied by a z-axis range map (e.g., from a structured light sensor). This shape information was recorded as the intensity values in a greyscale PNG image. The ISO standardization process has recently sought to allow more complete 3D information including the ability to encode concavities and folded structures (e.g., hook nose).

The standards are also likely to allow the storage of 3D information computed from 2D information such as morphable models [13] and active appearance models [14].

Resolution Requirements

The image sizes mentioned in ISO/IEC 19794-5:2005 are very much less than those attainable with contemporary consumer grade digital cameras. The reasons for this are two. First, the face recognition algorithms

of the early part of the decade were designed to operate with an interocular eye distance of between 40 and maybe 120 pixels. Second, the standard aims to be application independent, i.e., to only establish a minimum resolution to support automated face recognition. While more modern implementations are capable of exploiting high resolution imagery, the images may be too large for operational use (e.g., on an e-Passport chip, where size is typically much lesser than 50KB). Nevertheless, the 2007 revision of the ANSI/NIST ITL 1-2007 standard reflected the utility of high resolution imagery by incorporating a laddered scale that culminates in an image with a width such that 1700 or more pixels lie on the faces of 99% of U.S. male subjects. This specification supports forensic analysis of the face. It is termed Level 51 and is the highest level of the Type 10 record's Subject Acquisition Profile stack.

Note that a separate *profile* standard or requirements document could normatively specify minimum or maximum resolutions for a particular application. For example, the PIV specification[11] requires that imaging of a 20cm target at 1.5 metres produces 240 pixels, corresponding to about 90 pixels between the eyes.

Note that no standard currently exists for the certification of face recognition imaging systems. Such a standard might reasonably establish true resolution specifications in terms of line pairs per millimeter and as a full modulation transfer function profile. This would regulate the entire imaging system including the effects, say, of video compression.

Standards Development Organizations

Standards are developed by a multitude of standards development organizations (SDOs) operating in a great variety of technical disciplines. SDO's exist within companies and governments, and underneath trade associations and international body umbrellas. International standards promise to regulate larger marketplaces and the development process involves more diverse and thorough review and so consensus is more difficult to achieve. With stakes often high, development processes are conducted according to definitive sets of rules. These are intended to achieve consensus standards that are legally defensible, implementable, and effective.

The following list gives an overview of the relevant SDOs. Note that the published standards are usually copyrighted documents and available only by purchase.

- **ISO JTC 1 SC 37:** Although face image standardization is underway within a number of SDOs, by far the most work is conducted in the main international forum, SubCommittee 37 (SC 37) *Biometrics*. This body was established in mid 2002 as the newest of seventeen active subcommittees under Joint Technical Committee 1 (JTC 1) and its parent the International Organization for Standardization (ISO). ISO maintains a catalog of its standards development efforts at <http://www.iso.org/iso/en/CatalogueListPage.CatalogueList>. Although its focus is development of standards in support of generic identity management and security applications, its establishment was substantially motivated by a need for improved international border-crossing mechanisms.

Within the six working groups of SC 37, the body responsible for facial image standardization is Working Group 3. The group, which develops biometric data interchange format standards, is the largest WG in SC 37 and is developing the standards with the highest profile adoption in the marketplace. Its ISO/IEC 19794-5:2005 face image data standard has been specified by the International Civil Aviation Organization (ICAO) as the mandatory biometric in the electronic Passports now being issued in many developed nations.

- **M1:** M1 is the United States Technical Advisory Group (TAG) to SC 37. It was established in June 2002 and is responsible for formulating U.S. positions in SC 37 where it holds the U.S. vote. Staff from its member organizations represent these positions in SC 37. It is notable because it is also a standards development organization in its own right. Particularly it developed and published the INCITS 385 INCITS, which stands for International Committee for Information Technology Standards, is the SDO arm of the Information Technology Industry Council based in Washington DC. face image standard in 2004. This document is substantially similar to the ISO/IEC 19794-5 standard because the early drafts of the former were contributed toward the development of the latter.
- **ANSI/NIST:** The U.S. National Institute of Standards and Technology (NIST) is also a SDO.

It developed the ANSI/NIST standards for law enforcement under the canvass process defined by ANSI. (see sec.).

Summary

Data interchange standards have been developed to facilitate universal seamless exchange of facial information. In all cases, these wrap an underlying standardized encoded image (often ISO/IEC 10918 JPEG) with a header that includes subject-specific information and details of the acquisition. The standards support accurate face recognition by constraining the cameras, environment, and the geometric and photometric properties of the image.

Related Entries

- ▶ [Face Recognition](#)
- ▶ [Interoperability](#)

References

1. Kanade, T.: Picture processing system by computer complex and recognition of human faces. In: Doctoral dissertation, Kyoto University (1973). Available as TIFF images at <http://xiotech.ulib.org/cgi-bin/ulib/display?11014.12072>
2. Sirovich, L., Kirby, M.: Low dimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A* **4**(3), 519–524 (1987)
3. Australian Customs Service: Smartgate. Tech. rep.
4. Face recognition as a search tool “foto-fahndung”. Tech. rep.
5. Frank, T.: Face recognition next in terror fight. *USA Today* (2007)
6. JTC 1, SC29 Coding of audio, picture, multimedia and hypermedia information: ISO/IEC 10918-1 Digital compression and coding of continuous-tone still images: Requirements and guidelines, 1 edn. (1994). URL <http://webstoreansi.org> International Standard
7. JTC 1, SC29 Coding of audio, picture, multimedia and hypermedia information: ISO/IEC 15948 Computer graphics and image processing – Portable Network Graphics (PNG): Functional specification, 1 edn. (2004). URL <http://webstoreansi.org> International Standard
8. JTC 1, SC29 Coding of audio, picture, multimedia and hypermedia information: ISO/IEC 15444-1 JPEG 2000 image coding system: Core coding system, international standard edn. (2004). URL <http://webstoreansi.org>
9. ISO/IEC JTC 1, SC37 Biometrics: ISO/IEC 19794-5:2005 - Biometric data interchange formats - Face Image Data, 1 edn. (2005). URL <http://webstoreansi.org> International Standard

10. Aamva national standard for the driver license/identification card (2000). AAMVA DL/ID-2000
11. Wilson, C., Grother, P., Chandramouli, R.: Nist special publication 800-76-1 - biometric data specification for personal identity verification. Tech. rep., National Institute of Standards and Technology (2007). URL <http://csrc.nist.gov/publications/PubsSPs.html>
12. INCITS M1, Biometrics: INCITS 385:2004 - Face Recognition Format for Data Interchange, 1 edn. (2004). URL <http://webstore.ansi.org>. American National Standard for Information Technology
13. Blanz, V., Vetter, T.: Face recognition based on fitting a 3d morphable model. *IEEE Trans. Pattern Analysis and Machine Intelligence* **25**(9), 1063–1074 (2003)
14. Xiao, J., Baker, S., Matthews, I., Kanade, T.: Real-time combined 2d+3d active appearance models. In: Proc. International Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 535–542 (2004)

Face Image Quality Assessment Software

Face image quality assessment software provides multiple measurements of face image quality and determines automatically whether submitted face images are of adequate quality for a particular application.

- ▶ Photography for Face Image Data

Face Image Synthesis

- ▶ Face Sample Synthesis

Face Localization

- ▶ Face Detection

Face Matching

- ▶ Face Alignment

Face Misalignment Problem

SHIGUANG SHAN¹, XILIN CHEN¹, WEN GAO^{1,2}

¹Institute of Computing Technology, Chinese Academy of Sciences, Beijing, Peoples Republic of China

²Peking University, Beijing, Peoples Republic of China

Synonyms

Curse of misalignment; Face alignment error; Localization inaccuracy

Definition

The face misalignment problem, or curse of misalignment, means abrupt degradation of recognition performance due to possible inaccuracy in automatic localization of ▶ facial landmarks (such as the ▶ eye centers) in the face recognition process. Because these landmarks are generally used for aligning faces, inaccurate landmark positions imply incorrect semantic alignment between the faces or features, which can further result in matching or classification errors. Since perfect alignment is often very difficult, face recognition should be misalignment-robust, i.e., it should work well even if the landmarks are inaccurately located. To achieve this, there are three possible solutions: misalignment-invariant features, misalignment modeling, and alignment retuning.

Introduction

In face recognition, before extracting features from a face image, it must be aligned properly with either the reference faces or a pre-defined general face model, with the help of some landmarks. For instance, the eye centers are generally used as control points to align all the facial images, i.e., all the faces are geometrically normalized by fixing the eye centers. Intuitively, the goal of alignment is to build the semantic correspondence among different face samples. Accurate alignment is evidently very important since the similarity (or distance) measurements generally assume the same semantics for the same feature index. However, in case of inaccurate landmark localization, this semantic alignment is broken, i.e., the component of face features extracted from the same subject with the same feature

index might imply different semantics. For instance, as shown in Fig. 1, if the eyes are inaccurately localized when testing, say confused with the eyebrows, it will result in ridiculous matching eyes with eyebrows, possibly also matching nose with mouth, which will evidently lead to an incorrect classification. The above-mentioned misalignment is actually equivalent to affine transformation, i.e., translation, scaling, and rotation.

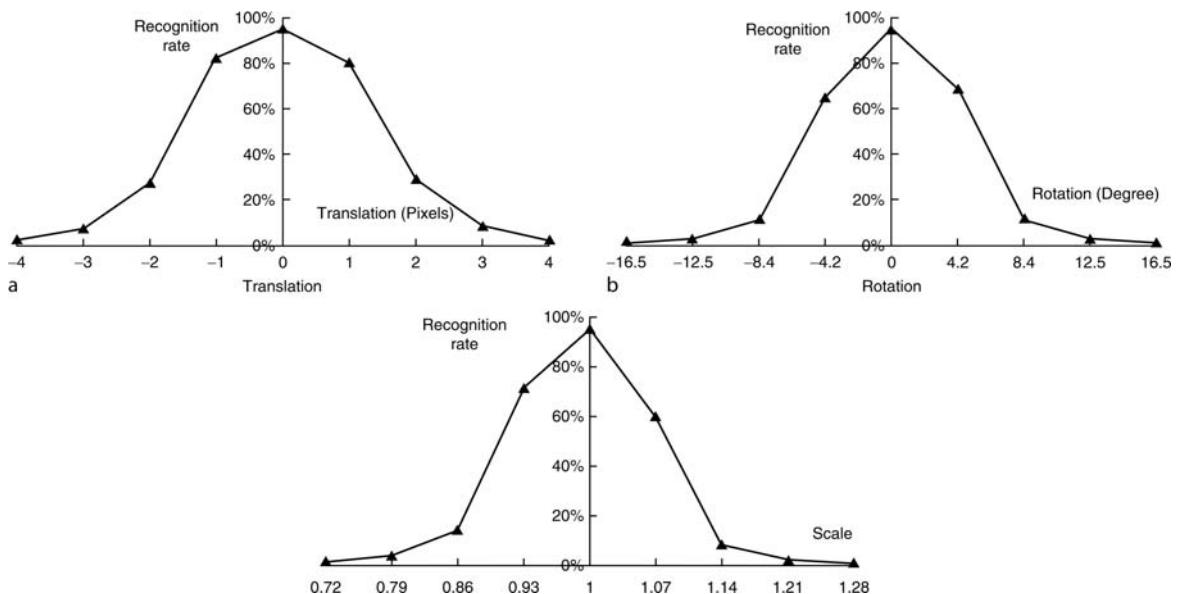
To demonstrate how much misalignment can degrade the face recognition systems [1], experiments were conducted on the FERET face database to evaluate the performance variance of Fisherfaces method [2] against the degree of misalignment. The results are shown in Fig. 2a–c, how the rank-1 recognition rates change with the misalignment degree in translation, rotation, and scale respectively. It is clear that the

rank-1 recognition rate of the Fisherfaces method degrades abruptly with the increase in the misalignment. For example, 10% decrease is observed for misalignment due to a pixel translation, while 20% for misalignment of 4.2° of rotation, and almost 30% for 0.07 scale changing. Such abrupt degradation of the performance is hardly acceptable for a practical face recognition system in which misalignment of one or two pixels is almost unavoidable. Therefore, it is a problem that needs more attention.

For face recognition, aligning faces only according to the eye centers imply much more than simple affine transformation in case other variations are presented such as pose and expression. As shown in Fig. 3, the eye centers are aligned perfectly, but other features are not aligned correctly due to the 3D rotation of the head.



Face Misalignment Problem. **Figure 1** Example of misalignment caused by incorrect facial landmarks. The rightmost image is the blending of the two misaligned images, from which one can imagine how much misalignment can affect the effective matching of two biometric traits.



Face Misalignment Problem. **Figure 2** The rank-1 recognition rates of Fisherfaces against the degree of misalignment in translation, rotation and scale [1].



Face Misalignment Problem. [Figure 3](#) Example of misalignment caused by pose variation. The rightmost image is the blending of the two misaligned images, from which one can see much misalignment in nose, mouth, and chin area, though the eye centers have been aligned correctly.

Possible Solutions

Since misalignment problem results from inaccurate (even incorrect) alignment, it is a natural idea to improve the accuracy of alignment. For instance, one can localize the eye centers more accurately or locate more landmarks (e.g., as in active shape models or active appearance models). However, to the experiences of previous work on face alignment, accurate alignment is indeed a great challenge. So, one might not expect perfect alignment and has to present efficient solutions for misalignment problem. Possible solutions can be divided into three categories: invariant features, misalignment-robust classifier, and alignment retuning.

Misalignment-invariant feature based methods expect to extract from the misaligned face images “good” representations robust to the misalignment, i.e., features change little or they even do not vary with misalignment. Some filters, such as Fourier transform, can be used for this purpose, since Fourier transform is shift and rotation invariant. Gabor wavelet is also a good choice due to its locality, which has been discussed in [3]. Recently, histogram-based object representation like, histogram of Local Binary Patterns (LBP) [4] or Local Gabor Binary Patterns (LGBP) [5] are also invariant to translation and rotation, so they can be adopted as misalignment-robust features. In addition, misalignment-invariant features can also be extracted by discriminant analysis, in which misalignment is treated as within-class variation [1].

The second category of the solution tries to design misalignment-robust classifier. In [6], the authors propose to augment the gallery by perturbation and modeled the augmented gallery by Gaussian Mixture Models (GMM). In [7], the authors propose a misalignment-robust subspace learning method for face recognition, which can infer both the well-aligned face component and the misalignment parameters.

Since the problem results from incorrect alignment, the third method naturally retunes the alignment further. Note that, these methods should be clearly different from the preceding alignment algorithms in that the retuning should make use of the feedback information from the matching or classification procedure.

Related Entries

- ▶ [Face alignment](#)
- ▶ [Face descriptors](#)
- ▶ [Face localization](#)
- ▶ [Feature extraction](#)

References

1. Shan, S., Chang, Y., Gao, W., Cao, B.: Curse of mis-alignment in face recognition: Problem and a novel mis-alignment learning solution, In: Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 17–19 May 2004, pp. 314–320 (2004)
2. Belhumeur, P.N., Hespanha, J.P., et al.: Eigenfaces vs Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(7), 711–720 (1997)
3. Shan, S., Gao, W., Chang, Y., Cao, B., Yang, P.: Review the strength of Gabor features for face recognition from the angle of its robustness to mis-alignment, In: Proceedings of 17th International Conference on Pattern Recognition (ICPR2004), Cambridge, UK, 23–26 Aug 2004, vol. I, pp. 338–341 (2004)
4. Ahonen, T., Hadid, A., Pietikainen, M.: Face Recognition with Local Binary Patterns, In: eighth European Conference on Computer Vision, Prague, Czech Republic, May 2004, pp. 469–481 (2004)
5. Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition, In: Tenth IEEE International Conference on Computer Vision, Beijing, China, 17–20 Oct 2005, pp. 786–791 (2005)

6. Martinez, A.M.: Recognizing Imprecisely Localized, Partially Occluded and Expression Variant Faces from a Single Sample per Class, *IEEE Trans. Pattern. Anal. Mach. Intell.* **24**(6), 748–763 (2002)
7. Wang, H., Yan, S., Huang, T., Liu, J., Tang, X.: Misalignment-Robust Face Recognition. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 2008. Alaska, USA, 24–26 June (2008)

Face Photograph

► Photography for Face Image Data

Face Pose Analysis

IOANNIS PATRAS

Queen Mary, University of London,
E1 4NS, London, UK

Synonyms

Face pose estimation; Face pose recognition; Head pose analysis

Definition

Face pose analysis is the process of determining the location and the orientation of a face (► [Yaw/Pitch/Roll](#)) with respect to the camera/sensor's coordination system, and the subsequent facial analysis based on that information. A typical face pose analysis system determines the head pose by analyzing the information that is contained in the facial area (typically determined by a face detection system) using models of face geometry (i.e., models of the relative location of facial landmarks such as the nose tip and the eye corners) and/or models of face appearance (i.e., models of the intensity/color variation across a face image).

Introduction

A wide variety of systems requires the reliable analysis of facial information based on the analysis of images or

image sequences. The purpose of such systems is to analyze and interpret the information that is conveyed in the facial images, such as identity information or facial expression. Examples of applications are face recognition for security/surveillance (e.g., access control in buildings or airports), multimedia indexing and retrieval (e.g., searching for family photos based on who appears in them), and facial expression analysis [1] (e.g., for deception detection or for emotion recognition).

Traditionally, facial analysis assumed that the images were obtained in controlled conditions or were manually processed (e.g., cropped, resized, and rotated) such that the faces appear at the same orientation and size (e.g., the case in passport photos). However, for a large number of applications, such as face recognition in open spaces (e.g., an airport, or a tube station), it is practically impossible to introduce such controlled conditions or manually process the data in real time. In other applications, such as facial analysis for multimedia indexing and retrieval, imposing restrictions on the head pose is undesirable. Further, for applications such as in human–computer interaction the facial pose is by itself a primary source of information, and therefore, it does not make sense to restrict it. An example is a system for communication with a computer based on head gestures (such as nodding), or gaze.

Facial pose analysis addresses the needs of such applications by automatically recovering the head pose and by allowing the extraction of features that are tailored for the further analysis of facial images under the specific pose. As the size of the facial image is assumed to be normalized (i.e., cropped and resized) by the face detection module, head pose estimation typically reduces to the estimation of the three angles that specify the rotation of the head around its three axes. Of these, a distinction should be made between the estimation of in-plane rotations [i.e., head tilting (assuming a camera facing the subject)] and out-of-plane rotations caused by gestures such as head nodding or left–right head turning (assuming a camera facing the subject) (Fig. 1). The estimation of out-of-plane rotations is arguably more difficult as it involves 3D geometric transformations, and many works in face pose analysis are focused on this problem alone.

The estimation of the head pose allows the extraction of features that are tailored for further analysis of facial images under the specific pose. For this reason, facial pose analysis precedes (or overlaps with) many



Face Pose Analysis. **Figure 1** Examples of images from a face pose database with out-of-plane rotations, that is yaw (horizontal axis) and pitch (vertical axis).

other facial analysis modules such as face recognition and facial expression analysis. Further, face pose estimation requires that the facial area is reasonably well localized/detected, and therefore a face detector usually precedes it. Clearly, this requires face detectors that are capable of detecting faces at various poses (e.g., [2]). As pose-specific analysis can make more robust the face localization, some face detectors and face trackers [3] (i.e., modules that localize a face in the subsequent frames of a video) perform an internal coarse or a more precise pose estimation [4].

Face Pose Estimation

The core of face pose analysis is the estimation of the face/head pose from a 2D image that depicts it. This is an instance of the more general problem of estimating the 3D rotations and translations of an (potentially deformable) object from 2D images. The developed methods can be classified into two broad categories. Appearance-based methods (e.g., [5]) that rely on

models of how faces appear from different viewpoints (i.e., at different poses) and geometry-based methods that rely on the localization of facial landmarks (such as eyes and nose) on the image and 3D models of the face geometry. While appearance-based methods consider the information on the whole of the facial image at once (i.e., they are global methods), geometry-based methods estimate the head pose from the information on the 2D location of parts of the face.

A typical appearance-based method transforms the facial image into a feature set that represents the image in question in a way that it allows an easy determination of the face pose, i.e., it transforms the images from a general pixel/intensity-based representation to a pose-based representation (often called view-based representation [6]). This transformation [7, 8, 9] is typically learned from (large) databases that contain facial images of individuals at different poses. Such a transformation aims to provide a representation (i.e., a feature set) in which it is easy to distinguish between variations in the appearance due to factors other than the facial pose (e.g., identity and illumination)

and variations due to each of the pose parameters (i.e., the three rotation angles). Once a face region is transformed in this way, it is easier to disregard the variations due to other sources and recover the face pose. The variability in the feature sets extracted from images that depict faces at the same pose is called intra-class variation while the variability in the feature sets of images that depict faces at different poses is called inter-class variation. A useful transform is the one that leads to feature sets that exhibit small intra-class variation and large inter-class variation.

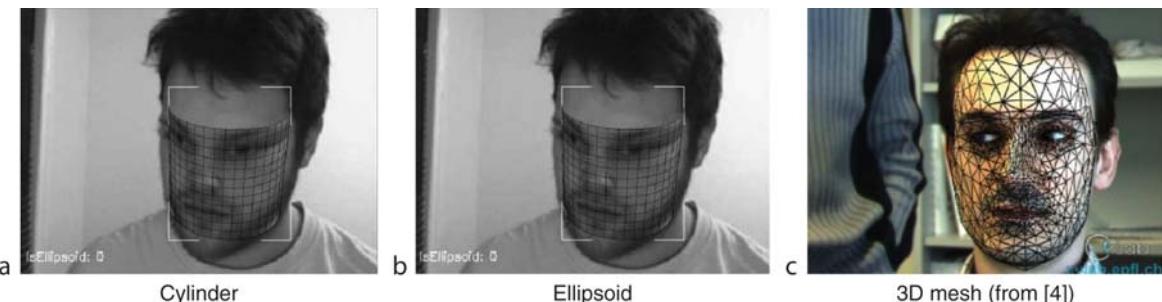
Once the transform is learned and each facial image is transformed to a feature set, a classifier that classifies each facepose representation to a facial pose is learned [5]. Learning a transformation and a classification scheme allows the determination of the pose of a face depicted in a previously unseen (i.e., new) image. For the new image, first the face region is detected by a face detector; then, a feature set is extracted (using the learned transform), and subsequently the head pose is determined (using the learned classifier). All classifiers require the existence of a database that is used for training and which contains a set of face images for each of which the true face pose is known. In one of the simplest classifiers the pose of the face in a new image is classified to be the pose of its nearest neighbor in the database. The term nearest neighbor, we refer to the image in the database whose representation (obtained with the learned transform) is most similar to the representation (obtained with the learned transform) of the image in question.

A typical geometry-based method relies on a 3D model of the face and on establishing of correspondences between the points in the 3D face model and the points in a 2D image that depicts the face. The estimation is based on the fact that if the pose of the face, that is the location

and rotations in the 3D space of the 3D face model, were known, and the ► [camera model](#) and camera parameters were given, then the location of any point of the 3D model (e.g., a facial landmark such as the left eye corner) on the 2D image could be calculated. Inversely, once the locations of facial landmarks on the 2D image are detected, the 3D facial pose can be estimated.

A 3D face model approximates the 3D shape of the face at a certain level of accuracy. Commonly used 3D face models (see Fig. 2) range from simple shapes such as half-a-cylinder [3, 10] or a plane [11, 12], to elaborate 3D meshes [4] that can be either generic or person-specific. As the face model is an approximation of the true face shape and the facial landmarks are typically not perfectly detected on the 2D image (e.g., due to occlusions and illumination changes), the projection of the 3D facial points on the 2D image does not coincide perfectly with the detected landmarks. For this reason, the estimation of the face pose is usually posed as an optimization problem in which the discrepancy between the expected 2D locations of the facial landmarks (as predicted by the face and the camera models) and the locations of the detected landmarks is minimized with respect to the pose parameters. In other words, during optimization we seek the pose parameters that minimize the error. As a pose transform is a rigid transform, the estimation of the pose parameters should rely only on stable facial points (such as the nose tip or the eye corners), that is, points whose location does not change with the facial expressions (such as the corner of the mouth).

Associated with the 3D geometrical model is the appearance information, that is information on how an area around a landmark is expected to appear on a 2D image. Such information is often provided in the form of a texture map (e.g., Fig. 3). Typically,



Face Pose Analysis. **Figure 2** Examples of 3D face models.



Face Pose Analysis. **Figure 3** Texture map projected on a cylindrical face model under different poses.

correspondences are established between the facial landmarks on the texture map and points on the 2D facial image using similarity measures on the appearance of small patches around them. As the reliability of the correspondences declines for small patches, geometry-based methods typically work with images of higher resolution than appearance-based methods.

It is often the case that other sources of information can be used in order to perform pose analysis. Often, the face pose needs to be estimated not in a single image but in an image sequence (i.e., face pose tracking [4]). If the frame rate is high enough, then the face pose changes slightly and smoothly from frame to frame. This prior knowledge can be incorporated in pose estimation algorithms by using various filtering techniques, with the effect that the estimated poses vary also smoothly from frame to frame. Another source of information that is commonly used is depth information, which is obtained either from a stereoscopic camera, or from range sensors [13]. In the first case, the facial landmarks need to be localized in both the images of a stereoscopic pair [14], and from this their location in 3D space is determined. In the second case, facial landmarks need to be localized on the range data itself. In both cases, depth information provides an additional constraint on the location and pose of the 3D model of the face. Finally, infrared imaging technologies, for single or multiple sensors can also be used.

Performance Evaluation

The main challenge in face pose analysis is the fact that facial images in the same pose appear differently due

to a number of factors. The most important of these factors are identity (differences in the facial characteristics between different individuals), illumination (i.e., the ambient light), occlusions (due to facial hair or other objects such as hands or glasses), and facial expressions (e.g., frowning or smiling). Such variations in appearance lead to variations in the feature set that are extracted by appearance-based methods and make difficult the correct classification. Similarly, variations in appearance make the establishment of correspondences between the texture map and the image patches in geometry-based methods difficult.

The evaluation of the performance of the face pose estimation methods is done on a collection of images that depicts faces whose poses are known, that is on an *annotated* face pose database. The term *annotated* refers to the fact that each image in the database is stored together with the corresponding “correct” pose (often called “ground truth”). The ground truth information about the pose is usually extracted during the recording of the image. An accurate method for doing so is to use additional sensors, for example, attach a magnetic sensor on the top of the head of the person, which delivers accurate pose information. Another method, which is however less accurate, is to ask the individuals to look at a certain location while the image is recorded. A third method is to manually annotate a number of stable facial landmarks (i.e., points whose location do not change with facial expressions) and use geometry-based methods to estimate the pose.

The set of images (and the corresponding poses) contained in an annotated database used for evaluation is called the *test set*. Appearance-based methods also require the existence of an annotated database that is used for learning the transform (that given an image extracts a feature set that represents it) and the classifier (that given a feature set determines the face pose). The set of images (and the corresponding poses) are contained in such a database is called the *training set*.

During the evaluation, the pose of each image contained in the test set is estimated and the difference with the ground truth pose (i.e., the error) is calculated. Usually, the average value of the error and its variance from the average value are reported. Useful estimators are the ones that have zero mean error (*unbiased*) and small variance (i.e., small spread around the average value).

Applications

Faces that are captured by cameras or other sensors in uncontrolled environments are rarely in upright position, facing the camera/sensor from a fixed distance. Images captured by surveillance cameras, in commercial films, in family photos or homemade videos, and even images captured from web cameras attached to a laptop rarely depict faces in the same pose. Therefore, face pose analysis is an integral part of all applications that require face analysis in uncontrolled environments. Imposing restrictions on the recording conditions is very often unnatural, impractical, or infeasible. In addition, head pose estimation has by itself a number of applications, for example in human–computer interaction.

The applications of face pose analysis can be divided into three main categories:

1. *Security applications in uncontrolled environments.* In applications, such as surveillance in open spaces (e.g., airports or tube stations), the question, “is this individual in the list of suspects?” or “in which other tube stations has this individual been today?” often arise and require working with facial images in arbitrary poses. Further, applications such as access control for computer login, give an extra degree of easiness if it can allow (smaller or larger) pose variations.
2. *Multimedia indexing and retrieval.* A very large portion of produced visual material, such as images in the web, films, homemade videos, and photos, depict faces. Organizing such a material according to who is depicted allows semantic access to it, that is, allows queries such as “find photos of me with my sister”.
3. *Human computer interaction and behavioral analysis.* Face/head pose can be used for communication with a computer (e.g., by head nodding or as an essential step toward gaze tracking), especially in case that disabilities prohibit the use of other primary modalities such as speech. Further, the face pose and its dynamics contain information on the emotional and affective state of individuals, and therefore can be used for automatic behavioral analysis.

Summary

Recent technological advances in image (sequence) acquisition, storage and transmission, such as the development of cheap cameras and hard disks, as well as

the availability of computing resources have contributed to the integration of imaging technology in our everyday lives. As face analysis moves from controlled environments to environments in which the viewpoint cannot be controlled, or applications in which the face orientation naturally changes, head pose analysis becomes an essential part of the developed systems.

Related Entries

- ▶ Face Alignment
- ▶ Face Expression Recognition
- ▶ Face Localization
- ▶ Face Tracking
- ▶ Feature Extraction

References

1. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: The state of the art. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(12), 1424–1445 (2000)
2. Sung, K.K., Poggio, T.: Example-based learning for view-based human face detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(1), 39–51 (1998)
3. Cascia, M.L., Sclaroff, S., Athitsos, V.: Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(4), 322–336 (2000)
4. Vacchetti, L., Lepetit, V., Fua, P.: Stable real-time 3d tracking using online and offline information. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(10), 1385–1391 (2004)
5. Li, S.Z., Lu, X., Hou, X., Peng, X., Cheng, Q.: Learning multiview face subspaces and facial pose estimation using independent component analysis. *IEEE Transactions on Image Processing* **14**(6), 705–712 (2005)
6. Poggio, T.: Image representations for visual learning. *Science* **272**(5270), 1905–1909 (1996)
7. Kirby, M., Sirovich, L.: Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(1), 103–108 (1990)
8. Moghaddam, B., Pentland, A.: Probabilistic visual learning for object representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 696–710 (1997)
9. Martinez, A.M., Kak, A.C.: PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(2), 228–233 (2001)
10. Xiao, J., Kanade, T., Cohn, J.F.: Robust full-motion recovery of head by dynamic templates and re-registration techniques. In: *Int'l Conf. Face and Gesture Recognition*, pp. 163–169 (2002)
11. Horprasert, T., Yacoob, Y., Davis, L.S.: Computing 3-d head orientation from a monocular image sequence. In: *Face and Gesture Recognition*, pp. 242–247 (1996)
12. Gee, A.H., Cipolla, R.: Fast visual tracking by temporal consensus. *Image Vision Comput.* **14**(2), 105–114 (1996)

13. Malassiotis, S., Strintzis, M.G.: Robust real-time 3D head pose estimation from range data. *Pattern Recognition* **38**(8), 1153–1165 (2005)
14. Pogalin, E., Redert, A., Patras, I., Hendriks, E.A.: Gaze tracking by using factorized likelihoods particle filtering and stereo vision. In: Int'l Symposium on 3D Data Processing, Visualization and Transmission, North Carolina, Chapel Hill, USA pp. 57–64 (2006)

Face Pose Estimation

- Face Pose Analysis

Face Pose Recognition

- Face Pose Analysis

Face Processing

Face processing is a term introduced at the first international workshop on face processing in video (FPiV'04) to describe image processing tasks related to extraction and manipulation of information about human faces. The most common of these tasks are face segmentation, face detection, face tracking, face modeling, eye localization, face reconstruction, face quality and resolution improvement, best face shot selection, face classification, facial expression recognition, face memorization, and face identification.

- Face Databases and Evaluation

Face Recognition

- Liveness Assurance in Face Authentication
- Forensic Evidence of Face

Face Recognition From Image Sequences

- Face Recognition, Video-based

Face Recognition in Near-Infrared Spectrum

- Face Recognition, Near-infrared

Face Recognition Performance Evaluation

- Face Databases and Evaluation

Face Recognition Using Local Features

- Face Recognition, Component-Based

Face Recognition, 3D-Based

IOANNIS A. KAKADIARIS, GEORGIOS PASSALIS,

GEORGE TODERICI, TAKIS PERAKIS,

THEOHARIS THEOHARIS

Department of Computer Science, ECE and Biomedical Engineering, University of Houston, Houston, TX, USA

Definition

Face recognition is the procedure of recognizing an individual from their facial attributes or features and belongs to the class of biometrics recognition methods. *3D face recognition* is a method of face recognition that

exploits the 3D geometric information of the human face. It employs data from 3D sensors that capture information about the shape of a face. Recognition is based on matching metadata extracted from the 3D shapes of faces. In an *identification* scenario, the matching is one-to-many, in the sense that a probe is matched against all of the gallery data to find the best match above some threshold. In an *authentication* scenario, the matching is one-to-one, in the sense that the probe is matched against the gallery entry for a claimed identity, and the claimed identity is taken to be authenticated if the quality of match exceeds some threshold. 3D face recognition has the potential to achieve better accuracy than its 2D counterpart by utilizing features that are not sensitive in lighting conditions, head orientation, differing facial expressions, and make-up.

Introduction

In recent years, among the many biometric modalities, the face has received the most interest. Not only is face recognition one of the most widely accepted modalities, but advances in processing power have allowed the development of more complex algorithms while still providing a rapid response to queries. Face recognition requires no contact with the subject, thus being more easily accepted by the public compared to other biometrics such as fingerprints.

Face recognition has been traditionally performed using 2D (visible spectrum) images, while hybrid approaches have used infrared images and 3D geometry. Infrared face recognition has not been widely adopted due to the high cost of infrared cameras necessary to acquire data. In contrast, the cost of 3D scanners has dropped significantly, so it has become feasible to deploy them in the field, and therefore, the interest in developing algorithms that use 3D data has increased.

The main reason for using information from 3D data as a biometric is that the data acquired by 3D acquisition devices are invariant to pose and lighting conditions, these being the major challenges with which face recognition algorithms must cope. Moreover, image-based face recognition algorithms are more susceptible to impostors. Indeed, an impostor may use a printout of an image of a subject allowed to enter a facility in order to break in. To avoid this, the face recognition algorithm must be coupled with liveness

test algorithms. Attempting such an attack on a system based on 3D data would be much more difficult, since the attackers would need to obtain an accurate 3D model (sculpture) of the person whom they would like to impersonate.

The challenges of a 3D face recognition system are the following:

- *Accuracy gain:* A significant gain in accuracy with respect to 2D face recognition systems must justify the introduction of 3D recognition systems.
- *Efficiency:* 3D capture devices generate substantially more information than 2D cameras. Using this large volume of information is expensive in terms of computation time and storage requirements. Therefore, the algorithms developed need to be efficient both in time and space, by using the appropriate metadata.
- *Automation:* The system must be completely automated. It is therefore not acceptable to assume user intervention, such as for the location of key landmarks in a 3D facial scan.
- *Capture devices:* 3D capture devices were mostly developed for medical and other low-volume applications and suffer from a number of drawbacks, including artifacts, small depth of field, long acquisition time, multiple types of output, and high price. A deployable 3D face recognition system must be able to process several persons a minute, if it is to be used in high-traffic areas.
- *Testing databases:* There are few large databases of 3D faces which are widely accepted for objectively testing the performance of 3D face recognition systems. More such databases are needed to ensure proper testing of the system.
- *Robustness:* The system must perform robustly and reliably under a variety of conditions (e.g., lighting, pose variation, facial feature variation).

An Integrated 3D Face Recognition System

The authors have developed a fully automatic system [1] which is capable of using 3D data as input, along with a facial model, to output metadata information. The metadata are then used for recognition. The facial model has been constructed only once, and it can handle objects belonging to the same class (i.e., faces). Once the data are acquired, the model is fitted to the data

and used to generate a geometry image and a normal map, which are transformed into the wavelet domain. Only a small number of the wavelet coefficients are stored as metadata and used for comparison.

Our recognition procedure can be divided into two distinct phases: *enrollment* and *recognition*.

Enrollment: Raw data acquired by the 3D scanner are converted to metadata and stored in a database (gallery). The following steps describe the conversion from raw data to metadata (Fig. 1):

1. *Acquisition:* The sensor acquires raw data which are converted into a polygonal representation. A preprocessing step takes place to alleviate scanner-specific issues.
2. *Alignment:* The data are aligned into a unified coordinate system using a multi-stage alignment method.
3. *Deformable model fitting:* The ► **annotated face model** (AFM) is fitted to the data.
4. *Metadata generation:* Geometry and normal map images are derived from the fitted model and wavelet analysis is applied to extract a reduced set of the most significant coefficients.

Recognition: Metadata extracted from a face probe (using the same steps as for enrollment) are directly compared with metadata retrieved from the database gallery using a distance metric.

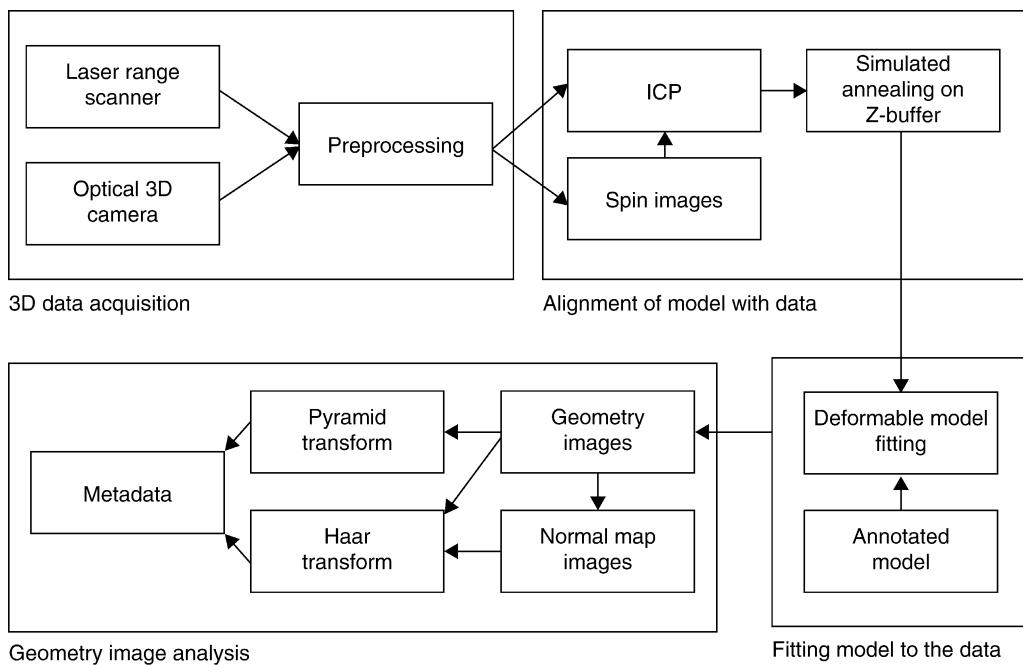
Data Acquisition and Preprocessing

In general, the current generation of scanners outputs either a range image or 3D polygonal data. The purpose of preprocessing the data is the elimination of any sensor-specific issues and the unification of data from different sources into a common format (Fig. 2). The preprocessing consists of the following filters that operate on both the native representations and on 1-neighbors, and are applied in the given order:

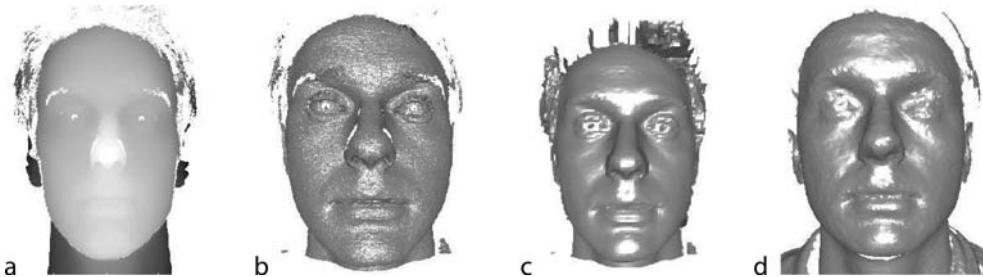
- *Median cut:* This filter removes spikes from data acquired by using laser scanners.
- *Hole filling:* Eliminates holes produced by laser scanners in certain areas such as eyes and eye brows.
- *Smoothing:* A smoothing filter is applied to remove white noise.
- *Subsampling:* The deformable model fitting effectively resamples the data, making the method less sensitive to data resolution without losing performance in the recognition phase. Subsampling further reduces the noise in the geometry.

Annotated Face Model

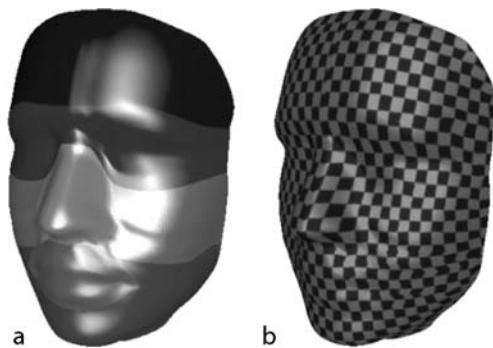
Our approach introduces an AFM, which is constructed only once and is used in the alignment, fitting,



Face Recognition, 3D-Based. **Figure 1** Enrollment phase of the proposed integrated 3D face recognition system.



Face Recognition, 3D-Based. **Figure 2** Sensor-dependent preprocessing. Laser range scanner: (a) input depth image, (b) raw polygonal data (200,000 triangles), and (c) processed data (16K). Stereo camera: (d) raw data (34,000 triangles).



Face Recognition, 3D-Based. **Figure 3** AFM:
(a) Annotated facial areas and (b) texture used to demonstrate parameterization.

and metadata generation [1]. The model is anthropometrically correct according to Farkas' work [2], and is annotated into different facial areas (e.g., mouth, nose, eyes) (Fig. 3). Applying a continuous global UV parameterization on the model, all vertices of the model from R^3 to R^2 and vice versa have been mapped. Therefore, the model is defined both as polygonal data in R^3 and as a geometry image in R^2 [1, 3].

A ▶ **geometry image** is a regular sampling of the model represented as a 2D image with three channels, each channel corresponding to the x , y , and z coordinates of the 3D object. Since local neighborhoods on the mesh are preserved (i.e., neighboring vertices are preserved even in the geometry image), an approximated version of the original mesh can be reconstructed from the geometry image. The number of channels in the geometry image can be greater than three, as apart from geometric information, texture and annotation can also be encoded.

Alignment

Our work on face recognition has indicated that alignment (pose correction) is a key part of any geometric approach. So, before fitting, align each preprocessed dataset with the AFM. The alignment stage computes a rigid transform, combining rotation and translation, which brings the data as close as possible to the model and is robust and accurate even when relatively large deformations (facial expressions) occur in the input data. Our alignment algorithm is a multi-stage algorithm which propagates the alignment variables from one stage to the next [1]. The first algorithm is more resilient to local minima, while the next two algorithms provide greater alignment accuracy:

- **Spin images:** The purpose of the first step is to establish a plausible initial correspondence between the model and the data. This step can be omitted if the arbitrary rotations and translations in the databases are not expected. A spin image is a representation of the geometric neighborhood around a specific point [4]. To register two shapes, the correspondences between the individual spin images must be found. These correspondences are grouped into geometrically consistent groups and the transformations they yield are verified by checking if they rotate the data by an acute angle (based on the assumption that a given face does not have an upside down pose or an opposite orientation from the camera). This check is essential due to the bilateral symmetry property of the human face.
- **Iterative closest point (ICP):** The main step of our alignment process uses the ICP algorithm [5] extended in a number of ways. The ICP algorithm

solves the registration problem by minimizing the distance between the two sets of points. The annotated model is exploited by assigning different weights to different face regions. Additionally, pairs containing points on surface boundaries are rejected. This ensures that no residual error is introduced into ICPs metric by the non-overlapping parts of two surfaces. Finally, if the resulting transformation is not satisfactory, the option of running the trimmed ICP algorithm [6] is available.

- *Simulated annealing on z-buffers:* This is a refinement step that ensures that the model and the data are well aligned. The idea is to refine alignment by minimizing the differences between the z-buffers of the model and data. A global optimization technique has been employed, known as enhanced simulated annealing (ESA) [7], to minimize the z-buffer difference [8]. The higher accuracy of this step can be attributed to the fact that the z-buffers effectively resample the data which results in independence from the data's triangulation.

Deformable Model Fitting

The purpose of fitting the model to the data is to capture the geometric information of the desired object. In order to fit the AFM to the raw data, a subdivision-based deformable model framework [1] is used. When the deformation concludes, the AFM acquires the shape of the raw data. This establishes a dense correspondence between the AFMs surface and the raw data's vertices. Additionally, since the deformation has not violated the properties of the original AFM, the deformed AFM can be converted to a geometry image. The extracted geometry image encodes the geometric information of the raw data (Fig. 4). Note that

the deformable model framework discards data not belonging to the face and successfully handles artifacts without any special preprocessing.

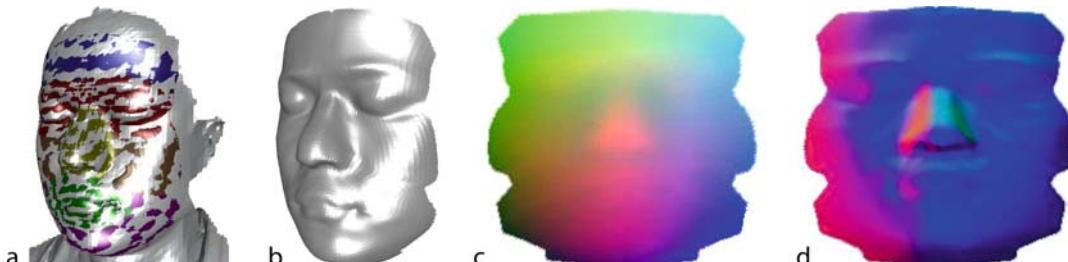
The fitting framework is an implementation of the ▶ **elastically adaptive deformable models** [9] using subdivision surfaces [10]. The Loop subdivision scheme [11] has been selected since it produces a limit surface with C^2 continuity, while only 1-neighborhood area information is needed for each vertex. The AFM is used as the subdivision surface's control mesh, thus determining the degrees of freedom, while the limit surface is used to solve the following equation:

$$M_q \frac{d^2 q}{dt^2} + D_q \frac{dq}{dt} + K_q q = f_q,$$

where q is the control points vector, M_q is the mass matrix, D_q is the damping matrix, K_q is the stiffness matrix, and f_q are the external forces. The external forces drive the deformation. The stiffness matrix defines the resistance against the deformation, while the mass and damping matrices control the velocity and the acceleration of the vertices. This equation is solved based on the finite element method (FEM) approximation. During this process the AFM gradually acquires the shape of the raw data.

Metadata Generation

The deformed model that is the output of the fitting process is converted to a geometry image, as depicted in Fig. 4(c). The geometry image regularly samples the deformed model's surface and encodes this information on a 2D grid. The grid resolution is correlated with the resolution of the AFMs subdivision surface. From the geometry image, a normal map image (Fig. 4(d)) is also constructed. The normal map



Face Recognition, 3D-Based. **Figure 4** Full face model after fitting: (a) Fitted model overlayed on the face data, (b) fitted model geometry, (c) corresponding geometry image, and (d) corresponding normal map.

contains the 3D normal vectors to the surface as its pixel values [1].

The three channels (components X , Y , and Z) of the normal map and geometry image have been treated as separate images. Each component is analyzed using a wavelet transform and the coefficients are stored as metadata. Two different transforms have been used, the Haar and Pyramid transforms, thus obtaining two sets of coefficients. The Pyramid transform is a more computationally intensive transform, and therefore, we may choose not to use it if the system needs to be tuned for speed. The Haar transform is applied on both the normal map and the geometry image, while the Pyramid transform is applied only on the geometry image.

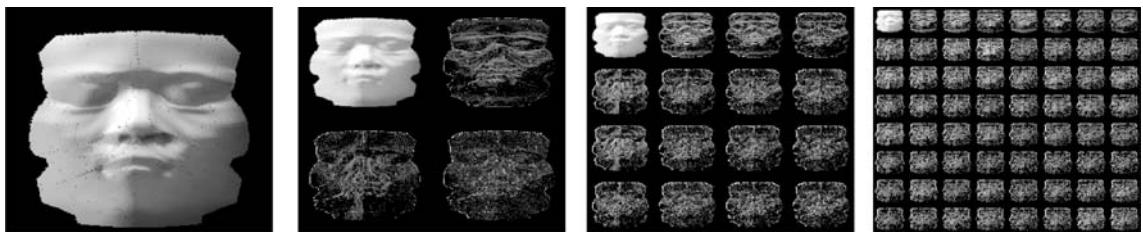
- *Haar wavelets*: The choice of Haar wavelets was based on their properties. The transform is conceptually simple and computationally efficient. The Haar wavelet transform is performed by applying a low-pass filter and a high-pass filter on a one-dimensional input, and then repeating the process on the two resulting outputs. Since we are working with images, there will be four outputs for each level of the Haar wavelet (Low-Low, Low-High, High-High, High-Low). A level 4 decomposition is computed, meaning that the filters are applied four times, which yields 256 (16×16) wavelet packets (Fig. 5). Each packet contains a different amount of energy from the initial image. It is possible to ignore most of the packets without losing significant information and store the same subset of the most significant coefficients as metadata. This allows an efficient direct comparison of coefficients of two images without the need for reconstruction.
- *Pyramid transform*: The second transform decomposes the images using the complex version of the

steerable pyramid transform [12], a linear, multi-scaled, multi-orientation image decomposition algorithm. The resultant representation is translation-invariant and rotation-invariant. This feature is desirable to address possible positional and rotational displacements caused by facial expressions. To maintain reasonable image resolution and computational complexity, our algorithm applies a 3-scale, 10-orientation complex steerable pyramid transform to decompose each channel of the geometry image. Only the low-pass orientation subbands at the farthest scale are stored as metadata. This enables us to compare the subband coefficients of two images directly without the overhead of reconstruction.

Distance Metrics

In the recognition phase, the comparison between two subjects (gallery and probe) is performed using the metadata information. The coefficients of the geometry image are kept as metadata, and the normal map of each dataset. Additionally, there may be two coefficient types for each: the Haar coefficients and, optionally, the Pyramid coefficients. To compare the metadata, there is a need to define a distance metric for each type of coefficient:

- *Haar metric*: In the case of Haar wavelets, the metric used is weighted L^1 on each component independently. The total distance is the sum of the distances computed on all components.
- *Pyramid metric*: A modified version of the complex version of the structural similarity index (CW-SSIM) [13] is used. CW-SSIM iteratively measures the similarity indices between two sliding windows



Face Recognition, 3D-Based. **Figure 5** Haar wavelet analysis for the normal map image: (a) zero level, (b) first level, (c) second level, (d) third level. Note that the real numbers were mapped to a gamma corrected grey-scale for visualization purposes.

placed in the same positions on the two images, and uses the weighted sum as a final similarity score.

- *Fusion:* When both types of coefficients are used, the distances given by the Haar and the Pyramid metrics are fused. A weighted sum of the two distances is used as a fusing score.

3D Face Recognition Hardware Prototype System

A field-deployable prototype system has been built and is operational at the University of Houston. It consists of a 3dMD™ 3D camera (1-pod configuration) which is connected to a laptop. The color camera of the pod captures a continuous video stream which is used to detect whether a person is facing the 3D camera. When the subject is facing the camera and remains relatively still for more than 2 s, the system triggers the 3D camera and the geometry data of the individual's face are captured. Each of the cameras has a resolution of 1.2 megapixels. The entire capture process takes less than 2 ms, and it produces a mesh with less than 0.5 mm RMS error (as quoted by the manufacturer).

The system can either enroll the subject into the database or perform a scenario-specific task. In an identification scenario, the system will display the closest five datasets to the operator. In a verification scenario, the system will determine whether the subject is who he/she claims to be, based on a preset distance threshold.

The system's field-deployable characteristics are:

- *Automation:* All methods utilized are fully automated, requiring no interaction with a user. The system is capable of detecting when a subject is within range by using a face detector implementation, and initiating the enrollment or authentication procedures automatically.
- *Space efficiency:* The raw 3D data produced by most scanners are of several MB. After the enrollment phase, the system needs to keep only the metadata.
- *Time efficiency:* The enrollment phase is the most time consuming, as the time delay to convert the raw scanner data to the final metadata is 15 s. In the authentication phase of an identification or verification scenario, only the stored metadata are utilized. The system can compare the metadata of enrolled subjects at a rate of 1,000/s on a typical modern PC (3.0 GHz P4, 1 GB RAM).

Performance Evaluation

Databases

The results on 3D face recognition are reported using two databases. The first is the well known FRGC v2 database and the second is a collection of 3D faces acquired at the University of Houston (UH). To demonstrate the sensor-invariant nature of the proposed system, the UH database is combined with FRGC v2.

The *FRGC v2* database [14, 15] contains 4,007 3D scans of 466 persons. The data were acquired using a Minolta 910 laser scanner that produces range images with a resolution of 640×480 . The scans were acquired in a controlled environment and contain various facial expressions (e.g., happiness or surprise). The subjects are 57% male and 43% female, with the following age distribution: 65% 18–22 years old, 18% 23–27 and 17% 28 years or over. The database contains annotation information, such as gender and type of facial expression.

The *UH* database contains 884 3D facial datasets acquired using our 3dMD™ system (with 1-pod and 2-pod setups) over a period of one year. The data acquisition protocol was the following:

For each subject:

- Remove any accessories (e.g., glasses).
- Acquire a dataset with neutral expression.
- Acquire several datasets while the subject reads loudly a predefined text (thus assuming facial expressions).
- Put on the accessories and acquire a dataset with neutral expression.

The UH database is more challenging compared to the FRGC v2 as the subjects were encouraged to assume various extreme facial expressions and in some cases accessories were present. The resulting extended database contains a total of 4,891 datasets, 82% acquired using a laser scanner, 18% acquired using an optical camera, and, to the best of our knowledge, is the largest 3D facial database reported.

Performance Metrics

Two different scenarios have been employed for the experiments: *identification* and *verification*. In an identification scenario, divide the database into probe and gallery sets so that each subject in the probe set has

exactly one match in the gallery set. To achieve this, use the first dataset of every individual as gallery and the rest as probes. The performance is measured using a cumulative match characteristic (CMC) curve and the rank-one recognition rate is reported.

In the verification scenario, measure the verification rate at 0.001 false accept rate (FAR). The verification rate is defined as the fraction of datasets that are positive (e.g., claiming to be who they really are), and are classified as positive. The FAR is defined as the fraction of datasets that are negative (e.g., pretending to be somebody else), but are classified as positive. The results are plotted using a receiver operating characteristic (ROC) curve which plots verification rate as a function of FAR. The FRGC v2 database defines three possible selections of datasets (referred to as ROC I, ROC II, and ROC III). In ROC I, all the data are within semesters, in ROC II, they are within one year, while in ROC III, the samples are between semesters. These experiments are of increasing difficulty.

Experiment 1: Wavelet Transforms

The purpose of this experiment is to evaluate the performance of the two wavelet transforms, and to provide a reference score on the FRGC v2 database. Using a fusion of the two transforms, our system yielded a verification rate of 97.3% (for ROC I at 0.001 FAR), while separately for the Haar transform a rate of 97.1% and for the Pyramid transform a rate of 95.2% were achieved ([Table 1](#)).

Even though the Pyramid transform is computationally more expensive, it is outperformed by the simpler Haar wavelet transform. This can be attributed to the fact that in the current implementation, the Pyramid transform utilizes only the geometry images and not the normal map images. The fusion of the two transforms offers more descriptive power, yielding higher scores, especially in the more difficult experiments of ROC II and ROC III, as depicted in [Table 1](#).

Face Recognition, 3D-Based. [Table 1](#) Verification rates of our system at 0.001 far using different transforms on the Frgc V2 database

	ROC I	ROC II	ROC III
Fusion	97.3%	97.2%	97.0%
Haar	97.1%	96.8%	96.7%
Pyramid	95.2%	94.7%	94.1%

To the best of our knowledge, this is the highest performance reported on the FRGC v2 database for the 3D modality.

Experiment 2: Facial Expressions

Facial expressions have traditionally decreased the performance of face recognition systems. In this experiment, the authors evaluate the impact of facial expressions on the performance of the system. All datasets in FRGC v2 are annotated, and one of the categories recorded is the facial expression. The authors chose to divide the database into two distinct sets: the first set contains non-neutral facial expressions only, while the second set contains datasets that were annotated as having a neutral facial expression.

The performance of the two subsets is compared to the performance on the entire set at 0.001 FAR in [Table 2](#). The average decrease of 1.56% in verification between the full database and the subset containing only facial expressions is very modest when compared to most other systems, given the fact that this subset contains the most challenging datasets from the entire database and is fully automatic. The small decrease in performance can be attributed to the use of the deformable model framework and the AFM.

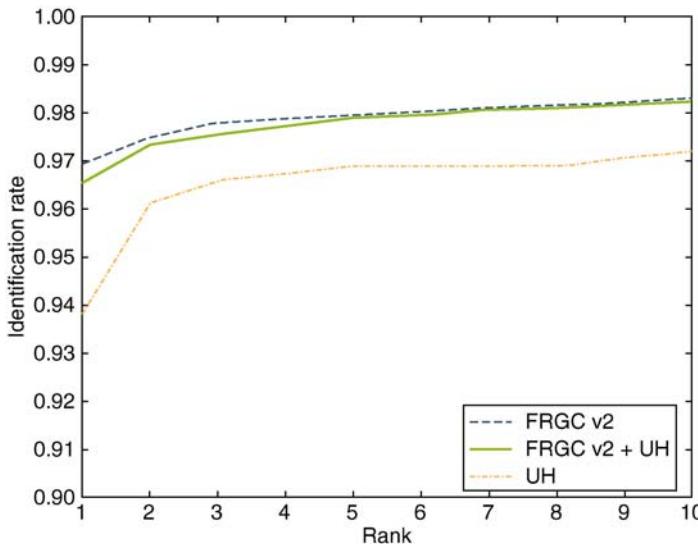
Experiment 3: Multiple Sensors

The purpose of this experiment is to evaluate the performance of our system using data from multiple sensors. Verification experiments depend heavily on the pairs of datasets chosen for evaluation. In the absence of any standard way of designing such experiments, opt for an identification experiment, which is considered to be more representative and more easily duplicated.

This identification experiment was conducted on different databases: FRGC v2 database, with 466 gallery and 3,541 probes (laser scanner), UH database with

Face Recognition, 3D-Based. [Table 2](#) Performance of our system at 0.001 far on the full FRGC V2 database, on a subset containing only non-neutral facial expressions and on a subset containing only neutral expressions

	ROC I	ROC II	ROC III
Full Database	97.3%	97.2%	97.0%
Non-neutral	95.6%	95.6%	95.6%
Neutral Expressions	99.0%	98.7%	98.5%



Face Recognition, 3D-Based. **Figure 6** System performance for identification experiment on different databases: FRGC v2 database with 466 gallery and 3,541 probes (laser scanner), UH database with 240 gallery and 644 probes (optical scanner) and FRGC v2+UH database with 706 gallery and 4,185 probes (both scanners).

240 gallery and 644 probes (optical scanner) and FRGC v2+UH database with 706 gallery and 4,185 probes (both scanners). On the FRGC v2 dataset, the rank-one identification rate was 97.0%, while for the UH set, the system achieved 93.8%. **Fig. 6** depicts the full CMC curve. The combined experiment yielded a rank-one recognition rate of 96.5%, which represents a drop in performance of only 0.5% when compared to the original FRGC v2 experiment, demonstrating the system's robustness when data from multiple sensors are included in the same database.

Conclusion

The authors presented algorithmic solutions to the majority of the challenges faced by field-deployable 3D facial recognition systems. By utilizing an annotated deformable model, the 3D geometry information is mapped onto a 2D regular grid, thus combining the descriptiveness of 3D data with the computational efficiency of 2D data. A multi-stage fully automatic alignment algorithm and the advanced wavelet analysis resulted in robust state-of-the-art performance on the publicly available FRGC v2 database. Our multiple-sensor database pushed the evaluation envelope one step further, showing that both accuracy and robustness can be achieved when data from different sensors

are present, through sensor-oriented preprocessing. Proof of concept is provided by our prototype system which combines competitive accuracy with storage and time efficiency.

Related Entries

- ▶ Anatomy of Face
- ▶ Deformable Models
- ▶ Face Localization
- ▶ Face Pose Analysis
- ▶ Face Recognition: Component-based
- ▶ Face Recognition: Shape vs Appearance

References

1. Kakadiaris, I., Passalis, G., Toderici, G., Lu, Y., Karambatziakis, N., Murtuza, N., Theoharis, T.: 3D face recognition in the presence of facial expressions: an annotated deformable model approach. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(4), 640–649 (2007)
2. Farkas, L.: Anthropometry of the Head and Face. Raven Press, NY (1994)
3. Gu, X., Gortler, S., Hoppe, H.: Geometry images. In: Proceedings of SIGGRAPH, pp. 355–361, San Antonio, TX, USA, July (2002)
4. Johnson, A.: Spin-images: a representation for 3-D surface matching. Ph.D. Thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August (1997)

5. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(2), 239–256 (1992)
6. Chetverikov, D., Svirko, D., Stepanov, D., Krsek, P.: The trimmed iterative closest point algorithm. In: Proceedings of the International Conference on Pattern Recognition, vol. 3, pp. 545–548. Quebec City, Canada (2002)
7. Siarry, P., Berthiau, G., Durbin, F., Haussy, J.: Enhanced simulated annealing for globally minimizing functions of many-continuous variables. *ACM Trans. Math. Software* **23**(2), 209–228 (1997)
8. Papaioannou, G., Karabassi, E., Theoharis, T.: Reconstruction of three-dimensional objects through matching of their parts. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(1), 114–124 (2002)
9. Metaxas, D., Kakadiaris, I.A.: Elastically adaptive deformable models. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(10), 1310–1321 (2002)
10. Mandal, C.: A dynamic framework for subdivision surfaces. Ph.D. Thesis, University of Florida (1998)
11. Loop, C.: Smooth subdivision surfaces based on triangles. M.Sc. Thesis, Department of Mathematics, University of Utah (1987)
12. Portilla, J., Simoncelli, E.P.: A parametric texture model based on joint statistic of complex wavelet coefficients. *Int. J. Comput. Vis.* **40**, 49–71 (2000)
13. Wang, Z., Simoncelli, E.: Translation insensitive image similarity in complex wavelet domain. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, vol. II, pp. 573–576. Philadelphia, PA, USA (2005)
14. Phillips, P.J., Flynn, P.J., Scruggs, W.T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005*, vol. 1, pp. 947–954. Gaithersburg, MD, USA (2005)
15. Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M.: FRVT 2006 and ICE 2006 Large-Scale Results. *NISTIR 7408*, March (2007)
16. Kirkpatrick, S., Gelatt, C., Vecchi, M.: Optimization by simulated annealing. *Science* **22**(4598), 671–680 (1983)

Face Recognition, Component-Based

ONUR C. HAMSICI, ALEIX M. MARTINEZ
The Ohio State University, Columbus, OH, USA

Synonyms

Face recognition using local features; Part-based face recognition

Definition

A major problem in face recognition is to design algorithms that are invariant to those image changes typically

observed when capturing faces in real environments. A large group of important image variations can be addressed using a component-based approach, where each face is first analyzed by parts and then the results are combined to provide a global solution. The image variations that are generally tackled with this approach are those due to occlusion, expression, and pose [1]. It has been argued that these changes have a lesser effect on local regions than to the whole of face image. Differences exist on how to formulate the component-based approach. Some of the algorithms use local information and combine these using a global decision maker. Some extract the important local parts to represent the face distributions, while others learn the distribution of the components generated by the variations. A summary of these techniques is given in this essay.

Introduction

Component-based face recognition algorithms include those that use some local information of the face to do recognition of the whole. These algorithms are very popular, since the local information is generally more robust to many of the typically seen parameter variations of the face. This is especially true if one does recognition based on the texture (i.e., pixel information) of the face.

One of these parameters is the location of the fiducial points in the face. These fiducial points are necessary to align all faces with respect to one another. However, it is not usually possible to obtain the exact location of these points automatically. This generates *imprecise localizations* which will further decrease the performance of the recognition algorithms [1]. Component-based algorithms can also be made more robust to these errors of localization. This is because some of the local features may be localized more precisely than the other ones and, hence, lead to better recognition rates.

A similar advantage is also seen in expression and pose changes. In this case, some local components of the face may have less expression changes (such as the nose region when a person smiles) or maybe less affected by pose changes (such as the eye region that is in the opposite side of the head).

Moreover, brightness changes are known to be handled better when the face is represented by components. It is because the face is a nonconcave structure, resulting in different lightings across it. For example,

the right and left side of a face may be lighted with totally different lighting conditions and, hence, may lead to different pixel levels. Trying to handle these changes using a global approach may fail due to the possible lighting changes. Simple intensity normalization processes can be used to eliminate part of the lighting differences when using a component-based approach [2].

Another advantage of using component-based algorithms is the stability for the possible occlusions over the faces. Even when half of the face is occluded, as for example with a scarf or large eyeglass, a component-based algorithm can still employ the information of the other half of the face image to do recognition.

Figure 1 shows some of the advantages of the local approach representation just described. Although, there is an extreme lighting change and occlusion of

the face, the local right eye regions are very similar in both images and shall lead to a successful classification.

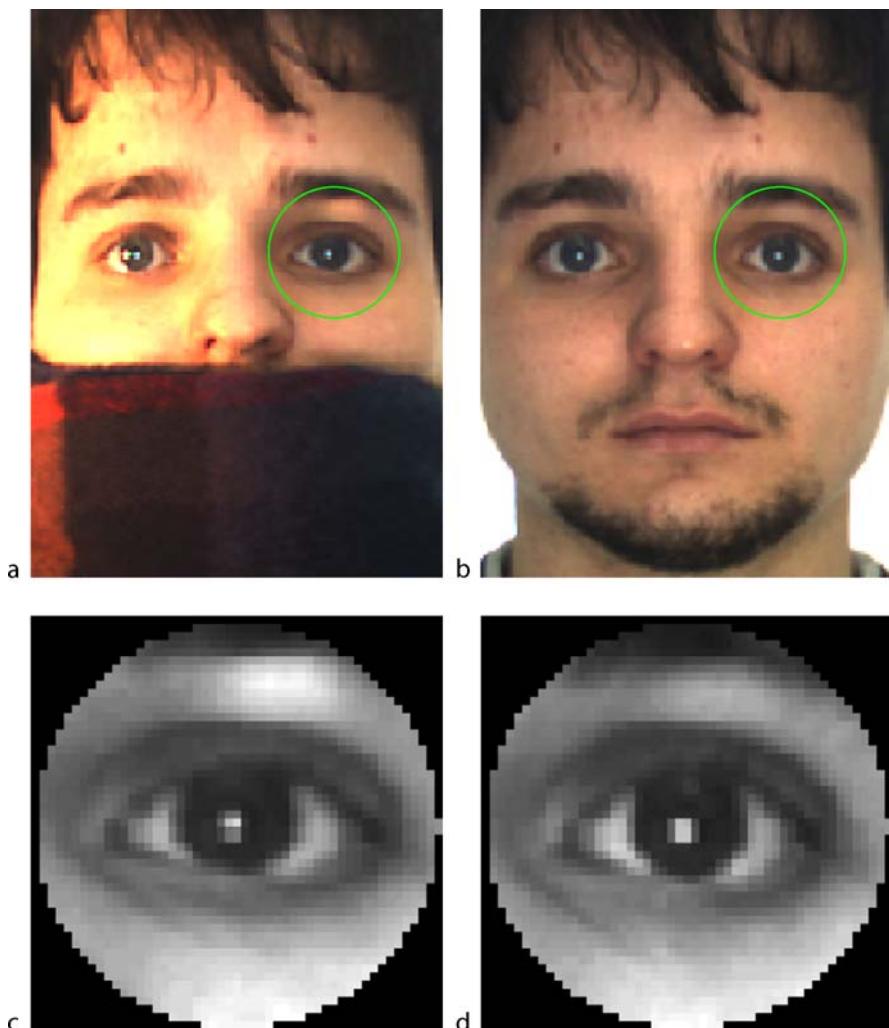
Because of these advantages component-based face recognition algorithms are preferred approaches in many real settings. In the following sections some of the most used algorithms defined thus far have been investigated. A discussion as well as the pros and cons of each technique, have also been provided.

F

Component-Based Face Recognition

Component-Based Graphs

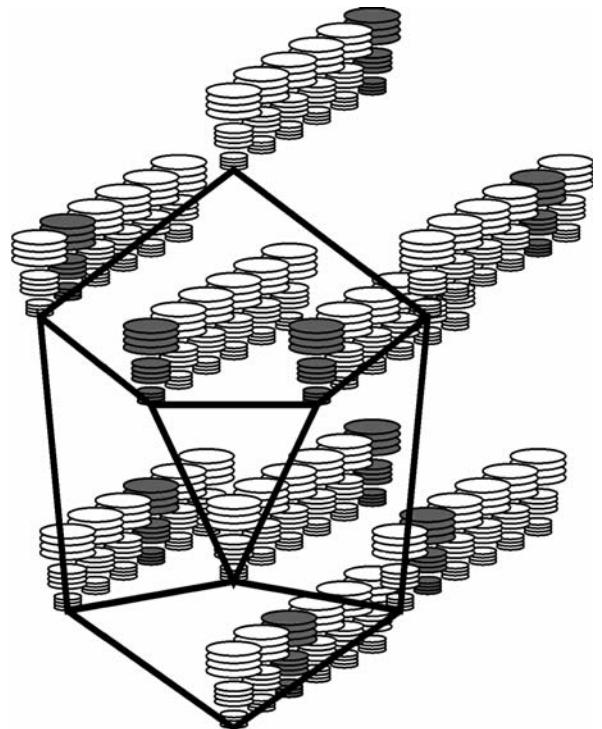
Building a system that is not skewed by localization errors seems close to impossible. This is due to the many



Face Recognition, Component-Based. Figure 1 Although the faces shown in (a) and (b) have extreme lighting changes and occlusions, the corresponding right eye regions in (c) and (d) are very similar.

additional variables that define the face image. These include pose, expression lighting, etc. Researchers have found a solution to this problem. In this solution, algorithms have been developed that consider the range of the localization errors over the given image. One solution, as defined in the section to follow, is to learn the set of textural changes due to localization errors. A robust alternative is to handle this set by designing an algorithm that depends on the local information. This is done in [3] with the *Dynamic Link Architecture* (DLA) and in [4] with the *Elastic Bunch Graph Matching* (EBGM). Both of these algorithms use the local information by dividing each image into a set of patches. While DLA extracts these patches by dividing the image with a grid structure, EBGM uses the regions around the fiducial points. Both DLA and EBGM extract features from the patches by filtering them with complex Gabor jets and using the magnitude of their outputs. The rationale behind the use of these features is grounded in the fact that ► **Gabor jets** are known to be less affected by lighting changes. In addition, EBGM uses the phase of the filtered patch to locate the nodes more accurately and to differentiate the patches that have the same magnitude.

One major difference between these two approaches is seen in the graph representation of the components. In DLA the spatial information between the patches is defined using a graph with nodes representing the grid parts of a face in the image plane. A proper matching algorithm between the images were proposed using the spatial information (hidden in edges of the graph) and the local information (hidden in vertices of the graph). In EBGM the face bunch graph (FBG) is defined over the fiducial points such that each node represents the Gabor jet outputs for several variations of a fiducial point, i.e., the node related with eyes may include an eye bunch that is closed, open, left-right pupil, and so on. Figure 2 shows an example of FBG. Here each node corresponds to a fiducial point where the set of discs are the Gabor jets related with the corresponding region. Bunch of set of discs represents the variability in the faces around that region. Matching is done using an elastic bunch graph matching algorithm which considers the size change of the FBG and location change of the nodes to optimize the graph similarity. Once the graph is obtained the recognition is done by calculating the similarity between the test image and the training image graphs. The match is found across all possible variations of Gabor jets. In Fig. 2 a possible match is



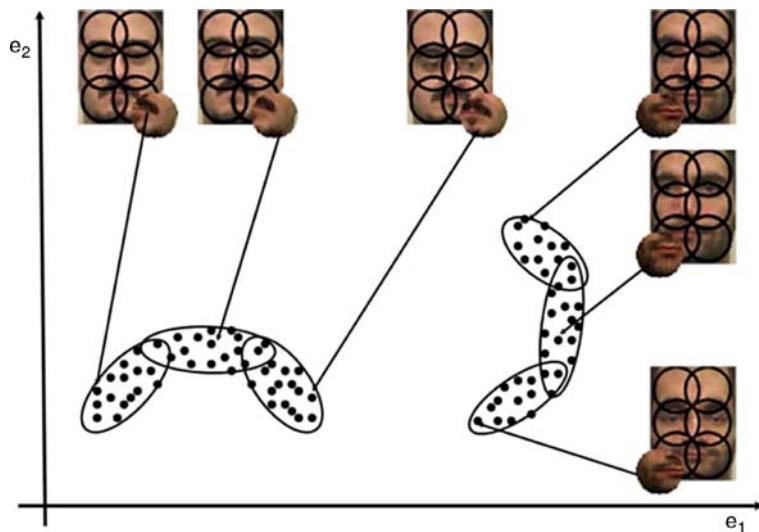
Face Recognition, Component-Based. Figure 2 A Face Bunch Graph (FBG) is shown, which represents all possible variations across faces. Each jet is represented by a stack of discs. In the matching process, the best fitting bunch of jets (shown in gray) is selected from a bunch of jets attached to a single node. © IEEE 1997.

shown by gray marking. This component-based strategy has proven to yield good results for frontal face recognition and reasonable results under pose variations.

Modeling Components

As mentioned above, some of the important problems to deal with in face recognition are imprecise localization, ► **partial occlusion**, and expression variation. Another important problem is given by the traditionally small number of training samples per class, since one usually has access to just a few images per subject. In [1] these problems are handled by means of a probabilistic approach.

► **Imprecise localizations** arise from not knowing the exact position of the fiducial points located in the face. Not only automatically detected, but also hand-marked feature locations include imprecise localizations.



Face Recognition, Component-Based. [Figure 3](#) This figure shows an approach used to model all possible warped face images according to the localization error of the localization algorithm. After division of each face image into K local areas, the localization error is estimated (for each of these local areas) using a mixture of Gaussians. © IVC 2006.

Although the additional noise may be small in the image plane, the corresponding images in the Euclidean space may deviate from their classes considerably. To handle this problem, [1] proposes to model the localization error by synthetically generating images that may be observed under a given error. Then, these images are modeled by means of a mixture of Gaussian distribution. With this approach one extends on the original set of images to a larger one that most appropriately represents image variations. [Figure 3](#) shows this for a face image division into six local parts. All possible images of a local region, which are generated by localization error, are modeled using a mixture of Gaussians.

Another major problem is partial occlusions. This is handled by dividing the face into a set of independent regions. This allows us to avoid using nonface areas that may distract the recognition process in a global approach. The component-based approach described above has shown to be superior to global techniques and voting strategies in [5] using a large set of images extracted from the AR database [6].

Expression changes are eliminated in a similar manner. In this case, a weighting scheme is used to give less importance to the regions that have expression changes and, hence, have a less contribution to the final classification. The effect that expression changes have in different local areas is learned from a training set. The learned weights are then applied to the independent test face images. This approach is further

extended in [7], where the weights are not learned but set inverse proportional to the difference in expression between the training and testing local regions.

More recently, this approach defined in this section has been extended to handle the problem of pose variations and to work with video sequences rather than simple stills [7]. Pose variations are again handled using Gaussian mixture models representing these variations. This algorithm has been shown to perform better than global approaches and voting strategies as well.

Another recent extension of this approach is given in [8], where the authors take advantage of the structure of the local areas by modeling it as a graph. This can be seen as a combination of the methods defined in the preceding section.

Extracting Sparse Components

Generally speaking, in the approaches defined above, there are infinitely many possible ways to divide a face image into a set of components. The question is *what is the optimal division?* The answer to this question will depend on the optimality criterion chosen. When the separation is defined for a face recognition algorithm, the components are usually selected to include local regions that keep most of the main characteristics across the faces. This includes dividing the face into regions separating eyes, nose, and mouth or dividing

the face into equal local patches. On the other hand, if our goal is to represent (instead of discriminate) the faces *sparingly* using a component-based representation, one needs an algorithm that consider this alternate criterion. Such sparse representation of faces is required in many practical cases, since the complexity of representing several kinds of faces can simply be done by finding invariant component-based representations. An algorithm defined for this purpose is the nonnegative matrix factorization (NMF) [9].

In this approach, the parts of an object are learned automatically by the algorithm. The algorithm inputs the data (graylevel pixel values in the case of images) in matrix form, \mathbf{V} , with each column representing an image. The goal is to factorize this matrix into basis vectors, \mathbf{W} , and coefficients, \mathbf{H} , such that none of the elements are negative, $\mathbf{V} = \mathbf{WH}$. To achieve this, an Expectation Maximization (EM) like algorithm is proposed. Because of the nonnegativity constraints, the basis vectors become highly sparse leading to an efficient representation of the data matrix.

A major advantage of this algorithm is that it is able to extract the parts of the objects automatically according to their significance in the representation of the data. Since these parts are usually less variant under pose, illumination, and occlusions, one can design part-based recognition algorithms that are based on NMF features. An extension of this framework is given in [10].

Variety in Features

Some of the component-based algorithms defined in the literature, differ in the features that they use to represent the local regions. One of them is proposed in [11] where the authors extract Fourier Bessel coefficients of the local regions.

Three local regions around the eyes (left eye, between eyes, right eye) are cropped after automatically locating the face and eyes. The illumination changes across each local part are removed by means of an image normalization. If this corresponds to a region that has constant luminance, it is eliminated. This means that the occluded regions are eliminated. To extract the features, the local image patches are transformed to the polar frequency domain using the Fourier-Bessel Transform (FBT). This feature representation is used since the noise is eliminated using a

subset of 372 FB coefficients. Using these coefficients the dissimilarities between each image with all the other images in the training set is calculated. Then, pseudo Fisher Linear Discriminant method (LDA) is used to classify the images [12]. The test results on FERET dataset [13] show that the proposed algorithm outperforms local approaches such as local polar Fourier Transform (PFT, which uses the Fourier transform instead of Fourier-Bessel), EBGM and global algorithms such as Principal Component Analysis (PCA), LDA, and global PFT. The results indicated that the proposed algorithm is sensitive to age and illumination changes more than expression changes.

The proposed algorithm is also tested with respect to its robustness to artificial occlusions, which is a drawback. Results on real occlusions are still needed. For this purpose 50% of the face was synthetically occluded with the graylevel information of those pixels equated to zero. In this experiment, the performance of local FBT was quite robust.

The authors further tested the significance of the localization error generated by their fully automatic system. They showed that the global approach is more affected by these errors than the local ones. And the local FBT performance was affected by up to 20% in the tests including age, expression, and illumination. This shows the importance of the localization errors in face recognition.

Other than changing the representation domain (pixel to polar frequency) of a local region, some algorithms use geometric features to represent the face components. One of these algorithms, which was already mentioned above, is the Face-ARG matching algorithm [8]. This algorithm uses the local information by extracting an Attributed Relational Graph (ARG). This graph is defined for each face image using a connected set of lines outlining the facial features. An important feature of the algorithm is that it uses only a single image for training, i.e., to extract the ARG. The testing is done using a partial match over the graph which was able to handle local changes and partial occlusions. The only disadvantage of the algorithm is the complexity of the matching. This is because the matching defined over the graph should also handle subset matching to deal with occlusions. The algorithm was able to obtain better recognition results on AR dataset [6] over most of the well-known algorithms such as nearest neighbor classification, PCA, and NMF.

Combined Face Detection and Identification

Not only face recognition but also face detection can be addressed using a component-based approach. In [14], the authors propose one such combined framework. They used a layered framework where the first layer is a component-based face detection module. This module consists of component classifiers specially trained for detecting facial components. Each detector outputs the most probable score and the x, y location of the corresponding component. A detection combination classifier receiving this data makes the final decision regarding the existence of a face in the image. If a face is detected, the obtained part-based regions are classified for identification in the second layer of the algorithm. Similar to the face detection module, the scores obtained from each part-based identification module are merged using a combination identity classifier, providing the final decision.

The training procedure in the face detection layer works as follows: First, 14 points on the face are selected as the center of the interest points. Then, starting from a rectangular region of predetermined size around each point, the interest region is increased until a minimum cross-validation error is obtained. On an independent testing set the trained face detector is able to outperform a detector that is exclusively based on global features.

Once the size of the rectangular local region is determined in the detection layer, a linear Support Vector Machine (SVM) is trained for identification purposes. In this training procedure, 7040 synthetic face images are generated from a 3D model constructed for each individual using only three images, i.e., a frontal, a 45° rotated to the right, and a side face. Testing is held using 200 images from a total of 10 people. Images are recorded in different days and with different cameras. The proposed algorithm is compared to global face identification systems such as PCA, LDA, and a SVM with polynomial kernel. The authors further tested their combination face identifier using linear SVM with respect to possible combination such as majority vote, maximum product and maximum sum. All these combination scenarios perform better than the global methods with the linear SVM-based identifier combination. Specifically the linear SVM based on part-based region identification performed 89.25%, whereas the PCA, LDA and SVM

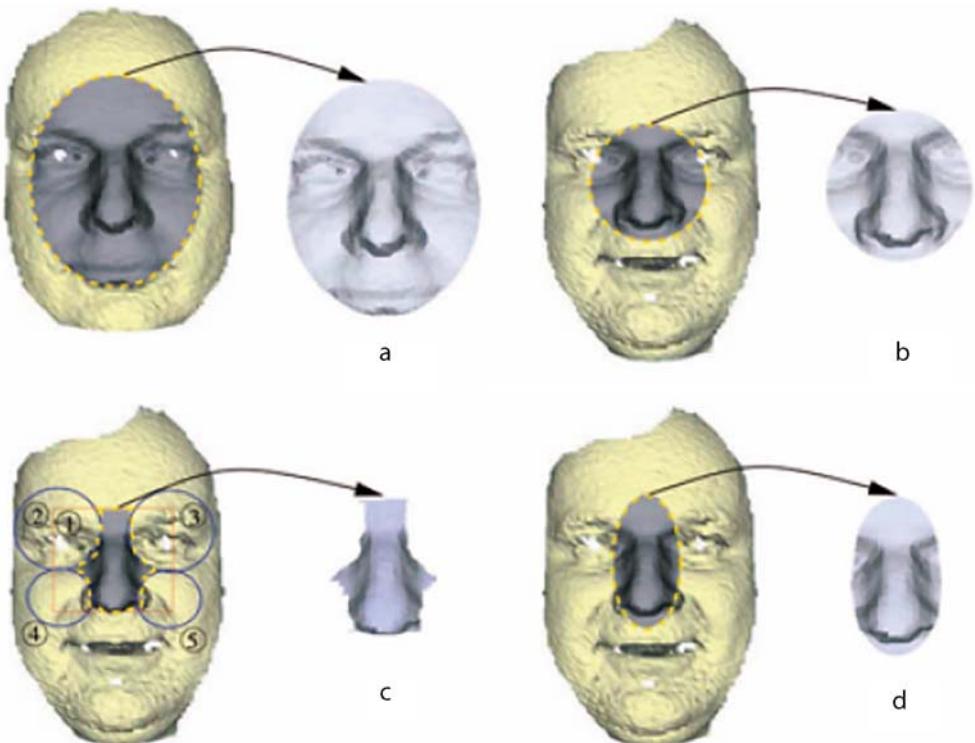
based on global features obtained 61, 52, and 63%, respectively.

3D Face Recognition

The importance of part-based approaches in face recognition is also clear in algorithms defined to do recognition from 3D range scans [15]. In this paper, authors extend the part-based comparisons in 2D images to 3D range scans. The main idea in the paper is that the nose region is usually stable when the subjects have expression changes. Hence three different possible local regions are extracted from the 3D structure. Figure 4 shows these regions which correspond to a circular region centered around the nose, Fig. 4(b), a region exclusively dedicated to the nose (i.e., including the curvatures around the nose region), Fig. 4(c), and a larger, ellipsoidal region including the nose and its contextual information, Fig. 4(d). The experimental results obtained over expression variant faces show that using the 3D structure of the nose outperforms the use of the whole face. Several combination schemes are proposed to improve the results obtained by the local parts. Above all, the combination of the match scores obtained from the nose and ellipsoidal nose region performs the best for nonneutral expressions. This shows that the stability of the selected parts (in this case the nose region) can provide a moderate to large improvement in the performance in face recognition systems.

Object Detection

The part-based detection algorithms are also used in object detection in [16]. This approach depends on four major stages. In the first stage, the authors build a vocabulary of the parts to represent the images. This is done by applying the Forstner interest operator which detects intersection of lines and centers of circular patterns. Then, small image patches around the interest points are extracted. These image patches are clustered together to obtain a more compact set of image patches which is the part vocabulary. Second, each image is represented by a set of binary features stating whether the parts in the vocabulary are in the current image. The image representation also includes the spatial location of the parts in the image. This is done by calculating the relative distance and angular displacement between the parts and representing these



Face Recognition, Component-Based. [Figure 4](#) This figure shows three different nose regions extracted from the range scan shown in **(a)**. **(b)** corresponds to surface around the nose region, **(c)** is the region exclusively including the nose, and **(d)** is the surface of the ellipsoidal region around nose. © IEEE 2006.

in 5 and 4 bin histograms, respectively. In the third stage, the authors propose to learn a classifier by means of the sparse network of winnows learning architecture [17]. Since the feature representation is usually sparse this learning procedure is preferred. In the final stage, a classifier activation map is build to detect the objects in the test images. This is simply obtained by sliding the learned classifier with a window over the image and assign 0 whenever a negative activation occurs, and assigning the actual activation value otherwise. Once all the possible candidates have been identified over the whole image, neighborhood suppression is employed to eliminate false positives.

An extension to handle scale changes across the images is held using an image pyramid. At the end, the authors show the superiority of part-based approach with several tests over single scale or varying scale images.

SIFT Features

SIFT features are also applied to template matching problems because of the representational power of the

component-based algorithms. In template matching the goal is to retrieve an object from a set of images. This problem needs to deal with local appearance variations, partial occlusions, and scale changes. To be robust to these variations a part-based approach can be used [18]. Due to the variety of images that a single object can generate, the training set that we need to learn is usually large. The complexity of the algorithms not only increase with the number of samples in the training set but also with the slow template matching algorithms, such as calculating the sum of square distance (SSD) or the normalized cross-correlation (NCC) between images [12]. To eliminate this computational complexity [18] proposed to use rectangular filters that are usually employed for fast image filtering in integral image representations. Furthermore, the complexity of the training set is reduced by means of a part-based representation instead of global templates.

In this approach each image is divided into a set of patches, where each of them is filtered with a set of rectangle filters. This process usually leads to a smaller number of feature representation. A subset of these

features are then selected using a saliency threshold specified by the user. This is done such that the features that are closer to zero are eliminated, since these are mostly related to patches that have constant pixel and thus carry little classification information. Furthermore, a weighted approach is used to decrease the significance of the patches that have more appearance changes across the images, i.e., the mouth region in the face. To handle partial occlusions the most similar parts are used during the matching process. The variable scale of the images is given by the property of the rectangular filters. They have shown that this method can robustly and accurately classify faces very fast under partial occlusions, variable expression, and different scales.

Similar problems to those observed in template matching are also seen in image retrieval applications. Therefore, most of the algorithms defined for image retrieval extract local features from the images [19]. Of late, one of the most popular techniques used in this process are the Shift Invariant Feature Transform (SIFT) features. In this algorithm, each image is represented by the set of directional gradient vectors on local regions. Using a similar idea, [20] proposed PCA-SIFT which is shown to be a compact, fast, and accurate representation for faces. The key idea is to use PCA to describe the gradient information located around the keypoints selected by SIFT. Using PCA, the authors show that a small number of basis images (20) are enough to represent these gradient based images. The Receiver Operating Characteristic (ROC) curves would be inappropriate to analyze this algorithm, since this corresponds to a detection problem with a large number of false positives rather than to a classification one. Thus, recall versus 1-precision curves which compared *correctpositives/numberofpositives* (recall) with *falsepositives/number ofmatches* (1- precision) (normalized measurement) are employed. Using these analysis, the authors show that PCA-SIFT performs better than SIFT in several different scenarios such as, added noise, rotation and scale changes, projective warp, and reduced brightness. The conclusion from this algorithm is that SIFT features may include noise in the description of the local information, whereas using PCA on the local graylevel eliminates these, facilitating the final classification of faces.

A Comparison of Different Approaches

A recent comparative study for the local matching approach in face recognition is presented in [21].

In this review, methods are categorized according to alignment/partitioning algorithms, local feature representations, and classification combinations. Alignment/partitioning algorithms are also divided within themselves into three subclasses. The first of them uses the local components defining a face, such as eyes, nose, and mouth. It is argued that these features may not be appropriate when one wants to consider the relationship between components. Another set of algorithms uses ► **Face Warping** in order to obtain shape-free graylevel information. Instead of removing the shape, the third set of algorithms eliminate the affine transformation between the images and then employs a part-based representation.

The authors discuss several local feature representation types. These include the Eigenfeatures, Gabor features, and local binary pattern features. Eigenfeatures use the pixel level information and eliminate the noise and represent the images using an orthonormal bases. Gabor features are extracted using a set of Gabor jets over the parts and represent each component according to the output obtained with filtering. Local binary patterns are obtained by calculating the binary patterns around an interest point for a given radius.

Furthermore, the algorithms are cataloged according to the classification method and the combination of the local cues. Local features are either simply concatenated into a global feature or combined with weights. An alternative method is to use a weighted combination of the classifiers defined by the local parts. While in other cases, the sum rule or Borda count (i.e., the classifier that has the largest votes) may be used to combine the classification decisions.

The experiments are carried out with the FERET and the AR face databases and show that LBP performs the best when the images are used with no illumination change, while the Gabor jets are better when lighting is considered. Another experiment is conducted to learn the best components for face recognition, revealing that the nose region generally outperforms the others. This may be due to the fact that the nose region is the most stable part under changing expression.

The authors also consider the difference between local region approaches (i.e., regions around the fiducial points) and the use of local components (i.e., components that are the parts of the face image divided equally, similar to a grid structure). They conclude that the local region approach may perform

better since the shape information is also used implicitly in the process.

Another question is whether or not to use other regions of the face, such as the cheeks and the forehead. In [21], it is shown that although the use of such regions may degrade the performance of the global approaches such as eigenfeatures, and their use in local methods, as in LBP and Gabor jet, generally improves the final recognition rate.

Based on these experiments and observations, [21] defines a new component-based approach that uses Gabor jets to extract features from local regions at several scales and frequency values. They combine the classification results (where the similarity is obtained by normalized inner products) with the Borda count. This approach was able to outperform the others in a test using the FERET database.

Summary

In this chapter, the authors have reviewed some of most known and recent works on component-based face recognition. All of the algorithms summarized above are defined to handle one or more of the problems of face recognition, as for example, imprecise localization, pose, illumination, expression, and occlusion.

To do that some algorithms such as DLA and EBGM are defined to be invariant to localization of the faces. Alternatively, we can model the image variations or use a sparse representation. NMF tries to extract a sparse local representation for the components of the dataset. Some approaches such as FBT and Face-ARG use different features to represent the local regions. The expression varying 3D scans showed that using the nose as the local component improves the recognition from 3D data. While another algorithm defined to do recognition from a single image learns all possible image changes when possible and uses weights to determine which regions are most robust to variations elsewhere. These changes can be modeled using Gaussian distributions, e.g., a mixture of Gaussians representing the variations when the face is imprecisely localized or when an expression varies the brightness of the pixels in a patch. A similar approach is also defined for face recognition from video where the pose changes are also considered. Some other algorithms use the local information both for face detection and identification, whereas alternatives are defined

for generic object detection. Finally, the authors have summarized the results of a recent comparison.

All these algorithms have one common thread: considering the images as a combined set of components. This is mainly because of the stability of the local components over possible image variations. The use of component-based algorithms is to date one of the most used approaches in classification and identification of 2D and 3D faces.

Related Entries

- ▶ [Face Alignment](#)
- ▶ [Face Localization](#)
- ▶ [Face Pose Analysis](#)
- ▶ [Face Recognition, 3D-Based](#)

References

1. Martinez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(6), 748–763 (2002)
2. P. Belhumeur, D.K.: What is the set of images of an object under all possible illumination conditions? *Int. J. Comput. Vis.* **28**(3), 245–260 (1998)
3. Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., Vandermalsburg, C., Wurtz, R.P., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* **42**(3), 300–311 (1993)
4. Wiskott, L., Fellous, J.M., Kruger, N., vonderMalsburg, C.: Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 775–779 (1997)
5. Martinez, A.M.: Recognition of partially occluded and/or imprecisely localized faces using a probabilistic approach. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head, SC, USA (2000)
6. Martinez, A., Benavente, R.: The AR-face database. Tech. rep., CVC Tech. Report # 24 (1998)
7. Zhang, Y.B., Martinez, A.M.: A weighted probabilistic approach to face recognition from multiple images and video sequences. *Image Vis. Comput.* **24**(6), 626–638 (2006)
8. Park, B.G., Lee, K.M., Lee, S.U.: Face recognition using Face-ARG matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(12), 1982–1988 (2005)
9. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(6755), 788–791 (1999)
10. Guillamet, D., Bressan, M., Vitrià, J.: Weighted non-negative matrix factorization for local representations. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, pp. 942–947 (2001)
11. Zana, Y., Cesar, R.M., Feris, R., Turk, M.: Local approach for face verification in polar frequency domain. *Image Vis. Comput.* **24**(8), 904–913 (2006)

12. Fukunaga, K.: *Introduction to Statistical Pattern Recognition*. Academic Press, New York (1990)
13. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: The feret database and evaluation procedure for face recognition algorithms. *Image Vis. Comput.* **16**(5), 295–306 (1998)
14. Heisele, B., Serre, T., Poggio, T.: A component-based framework for face detection and identification. *Int. J. Comput. Vis.* **74**(2), 167–181 (2007)
15. Chang, K.I., Bowyer, K.W., Flynn, P.J.: Multiple nose region matching for 3D face recognition under varying facial expression. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(10), 1695–1700 (2006)
16. Agarwal, S., Awan, A., Roth, D.: Learning to detect objects in images via a sparse, part-based representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(11), 1475–1490 (2004)
17. Carlson, A.J., Cumby, C., Rosen, J., Roth, D.: The snow learning architecture. Tech. rep., Technical Report UIUCDCS-R-99-2101, Computer Science Department, University of Illinois at Urbana-Champaign (1999)
18. Guo, G., Dyer, C.: Patch-based image correlation with rapid filtering. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Minneapolis, Minnesota, USA (2007)
19. Martinez, A.M.: Face image retrieval using hmms. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (Workshop) Ft. Collins, CO, USA (1999)
20. Ke, Y., Sukthankar, R.: Pca-sift: A more distinctive representation for local image descriptors. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC, USA (2004)
21. Zou, J., Ji, Q., Nagy, G.: A comparative study of local matching approach for face recognition. *IEEE Trans. Image Process.* **16**(10), 2617–2628 (2007)

their intensity values. A geometric representation is obtained by transforming the image into geometric primitives such as points and curves. This is done, for example, by locating distinctive features such as eyes, mouth, nose, and chin, and measuring their relative position, width, and possibly other parameters. Appearance-based representation is based on recording various statistics of the pixels' values within the face image. Examples include: recording the intensities of the image as 2D arrays called templates and computing histograms of edge detectors' outputs.

F

Introduction

Face identification systems are challenged by variations in head pose, camera viewpoint, image resolution, illumination, and facial expression, as well as by longer-term changes to the hair, skin, and head's structure. The geometric approach, which transforms a face image into simple geometric primitives, holds the promise of being invariant to illumination and almost invariant to time-induced changes. Using well-understood ► [Multiple View Geometry](#) techniques, it can also be made practically invariant to minor pose differences, camera viewpoint changes, and image resolution. In addition, the geometric approach has the advantage that a geometric match is easy to interpret.

Inspite of their intuitive and seemingly precise nature, geometric face recognition techniques have been largely replaced by appearance-based techniques. In these techniques, image representations, which are directly computed from the pixel-intensities are compared to estimate similarities between images, or fed into ► [classifiers](#) that determine the identity of the person in the image.

Even though the appearance-based techniques are cleverly designed and engineered, they lack the rigorous nature of the geometric approach. When an appearance-based classifier determines a false identify or wrongly detects a match between two persons, it is often hard to understand why this happens. Nevertheless, in 1993 Brunelli and Poggio [1] have shown that a generic appearance-based method outperforms a simple geometric-based method on the same dataset, and contradicting evidence to their finding has been scarce.

Face Recognition, Geometric vs. Appearance-Based

LIOR WOLF

The Blavatnik School of Computer Science, Tel-Aviv University, Israel

Synonyms

Features vs. Templates; Shape vs. Texture

Definition

In 2D face recognition, images are often represented either by their geometric structure, or by encoding

Shape-based methods

The pioneering work of Kanade [2], which was among the first modern approaches for automatic identification of face images, used the geometric approach. His system, similarly to following contributions, starts by identifying the locations of dominant facial features such as the eyes' corners, the nostrils' center, and the mouth's extremities. This detection step is the most challenging step, and the detected features are selected to be both discriminative and easily detectable. Other desired properties include invariance to lighting conditions and to facial expression.

The second step consists of defining a vector of measurements that are used as a face signature. Kanade [2] uses 16 such measurements, which include ratios of distances (e.g., the ratio of the distance between the eyes and the width of the face), various facial angles, the chin's curvature and more. Brunelli and Poggio [1] use a different set of 35 features, which include eye-brow thickness, shape and position, nose and mouth position, facial width at several heights and the shape of the chin (Fig. 1).

The third step consists of comparing two faces, or more generally, learning a classifier that can identify the various face classes. Kanade uses a simple Euclidean distance to compare two signatures. The later work [1] employs principle component analysis (PCA) of various dimensionality. Peak performance is obtained with the maximal dimensionality, where the distance between signatures becomes their Euclidean distance.

An improvement to this basic three-step face recognition scheme can be obtained by taking advantage of the increase in accuracy of facial feature detection algorithms (e.g., [3]) by designing or learning more discriminative signatures, and by introducing modern machine-learning techniques to learn distance functions between signatures or to train better classifiers. To our knowledge, little work has been done to demonstrate any of these improvements.

An alternative framework [4] for extracting geometric primitives from face images is to ignore the high-level structure of faces (i.e., as composed of identifiable eyes, nose, etc.) and instead of transforming the image into line drawings containing many line segments. First, edge detectors are applied to detect edge pixels, followed by a morphological thinning



Face Recognition, Geometric vs. Appearance-Based.

Figure 1 Some of the geometrical features used in [1], including eyebrow thickness and vertical position at the eye center position, nose position and width, the mouth's vertical position and width, height of lips, eleven radii describing the chin's shape, width of face at nose height, and its width halfway between the nose and the eyes. Other features which are not shown include a description of the left eyebrow's shape. Figure adapted from Fig. 6 of [1].

operator. Then, a line fitting process [5] is used to break continuous edge curves into several short line segments. The set of obtained segments (each represented by its endpoints) constitutes the signature of the face image.

In order to compare two such signatures, a distance measure between two sets of line segments is required. In [4] an elaborate such measure is proposed which considers the fact that two similar line segments can differ in length, may be tilted or be parallel, and that some matching line segments may be missing.

Appearance-based methods

Even though the terms shape and geometry are often used synonymously, we should not be misled to

assume that appearance-based methods do not encode the face's shape. Indeed, the location of the facial features and their shape contribute much to the variation in appearance between persons. The other identity-based source of appearance variation between persons is the facial texture, which includes elements that are typically not encoded by the geometry-based methods such as skin tone, facial hair, freckles, scars, and wrinkles.

Most appearance-based techniques share similar stages or components, which are sometimes intertwined:

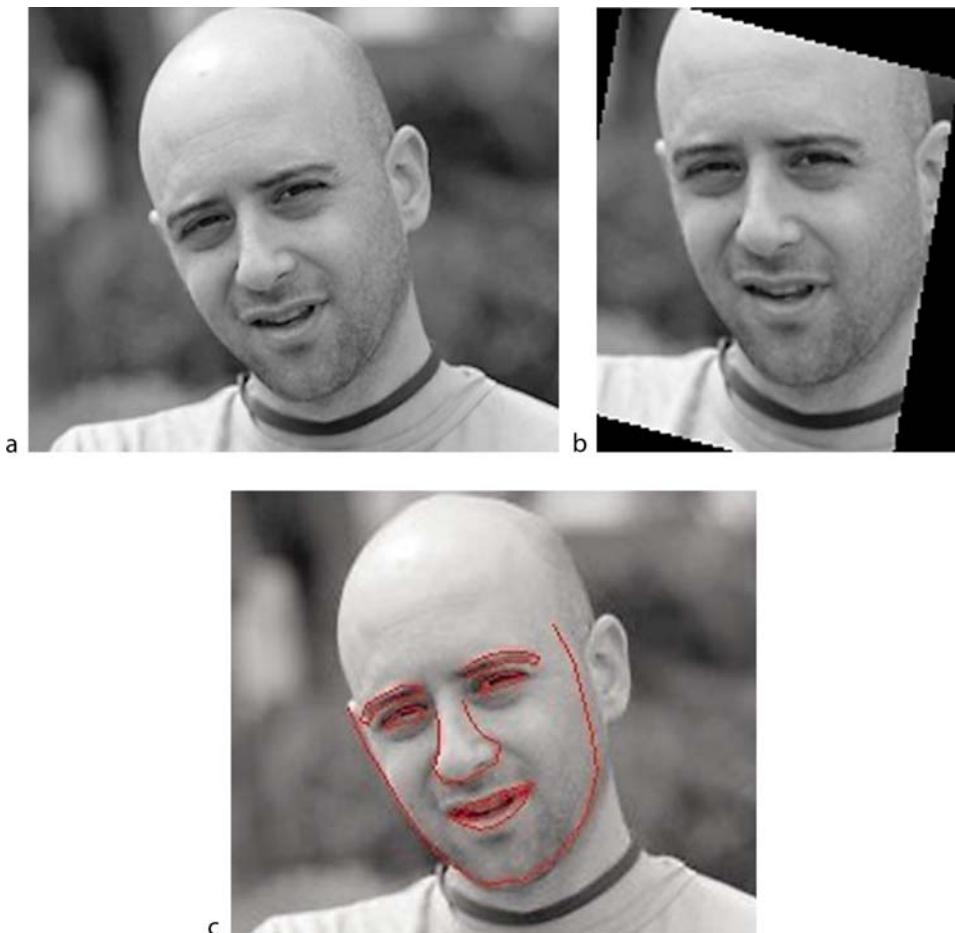
Normalization	During this initial stage, images may be scaled such that the area of the detected face is approximately constant. In addition, faces can be warped to a fixed reference image, see Fig. 2(b). A cropping mask is sometimes applied to remove the image boundary region, which typically includes hair and background elements. Furthermore, histogram equalization or other means of reducing the effects of illumination may be applied.
Signature generation	The normalized face image is being processed according to the specific algorithm at hand and a signature is created. For example, a local texture descriptor may be computed at each pixel, and the histogram of this descriptor can be used as a signature.
Learning or Classification	A classifier is trained to distinguish between the various persons in the database, or a distance function is learned to estimate the likelihood of two signatures belonging to the same person. These are used to classify the new images not used during training.

The most basic signatures are based on templates derived directly from the image. Variations may include using image derivatives instead of image intensities, or otherwise normalizing the intensities to reduce the effect of illumination. Another class of variations consists of using several templates (components) out of each face image [1, 6], and combining the matching score of each component in the final score.

Much work has been put into defining discriminative face signatures based on local texture descriptors. Local binary patterns (LBP) have shown to be extremely effective for face recognition [7]. The most simple form of LBP is created at a particular pixel location by thresholding the 3×3 neighborhood surrounding the pixel with the central pixel's intensity value, and treating the subsequent pattern of 8 bits as a binary number (Fig. 3). A histogram of these binary numbers in a predefined region is then used to encode the appearance of that region. Typically, a distinction is made between uniform binary patterns, which are those binary patterns that have at most 2 transitions from 0 to 1, and the rest of the patterns. For example, 1000111 is a uniform binary pattern while 1001010 is not. The frequency of all uniform LBPs is estimated, while all nonuniform LBPs, which are typically around 10% of patterns in an image, are treated as equivalent and given only one histogram bin. LBP representation for a given face image is generated by dividing the image into a grid of windows and computing histograms of the LBPs within each window. The concatenation of all these histograms constitutes the signature of the image.

A large body of literature exists on the proper way of learning classifiers and distances for face recognition. The PCA-based “eigenfaces” method [8] and the LDA based “fisherfaces” method [9] have been the first in a constant stream of work. It is the author’s experience that for modern descriptors such as LBP, All-Pairs Support Vector Machine [10] performs well in the task of biometric identification.

Recently, some effort has been devoted to the estimation of visual similarities between two unseen images, and such methods have been applied to determine whether two images belong to the same person. One method [11] that has shown good results for uncontrolled imaging conditions uses Randomized Decision Trees [12] and Support Vector Machines. In the first image of the pair, image patches (fragments of the image) are selected at random locations. For each patch the most similar patch in the second image is searched at a nearby image location. A decision tree is trained to distinguish between pairs arising from matching images and those arising from nonmatching images. Given a pair of unseen images, a Support Vector Machine classifier is used to determine if they match by aggregating the Decision Tree output of many image patches.



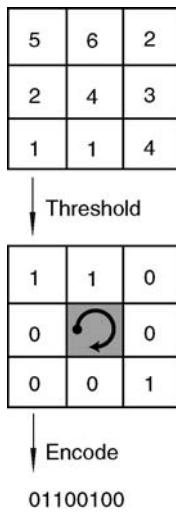
Face Recognition, Geometric vs. Appearance-Based. **Figure 2** Various methods of alignment. **(a)** The original image. **(b)** The congealing method [18] is used to align the image to a semi-frontal view. **(c)** Facial feature points located by the system of [19]. The detected feature points can be used to create a geometric signature or to align the image as a preprocessing step to an appearance-based approach.

Hybrid methods

Some face recognition methods combine appearance and geometry. In [13], the authors employ a face tracking method called Active Appearance Model to locate a set of feature points in and around the face. An example of similar features captured by a subsequent system, can be seen in Fig. 2(c). The located feature points are used to create three signatures that are combined during the recognition process. The first signature encodes the locations of the detected feature points; The second signature encodes the gray values of the image after it is ▶ warped such that the feature points are mapped to the mean location of those

points in the training datasets. The third signature encodes the local appearance around the detected feature points.

The Elastic Bunch Graph Matching system [14] uses local Gabor-wavelet based detectors that are connected through a simple spring model to locate facial features in a new image. The detected fiducial points are used to position a finer grid of points on the face image, and the response of various Gabor wavelets on the grid points is recorded to describe their local appearance. The matching score between two facial grids takes into account both the locations of the grid points and their appearance. This method performs well (for its time), however, the matching process is slow.



Face Recognition, Geometric vs. Appearance-Based.

Figure 3 The LBP image-texture descriptor is computed locally at each pixel location. It considers a small neighborhood of a pixel, and thresholds all values by the central pixel's value. The bits which represent the comparison results are then transformed into a binary number. The histogram of these numbers (the vector containing the frequency of each binary number in the image) is used as a signature describing the texture of the image.

A much faster hybrid method [15] uses coupled Gaussian mixture models to locate eyes, nose tip, and mouth in images of varying pose. Five SIFT appearance descriptors [16] are computed in regions around the detected features and in between the eyes.

Summary

Appearance-based methods currently dominate the general field of object recognition, where more classical methods based on analysis of relative positions of corners and other feature points have been mostly abandoned. Furthermore, there is evidence that the same image descriptors can be used for both object recognition and face identification [7, 15, 17]. It is therefore not surprising that the leading face recognition methods are also appearance based.

However, human faces differ from most objects studied in object recognition in that they have a well-defined structure. It is possible that the major

disadvantage of geometric face recognition is the lack of robust facial feature detectors. The advent of new detection techniques may reignite the interest in those methods.

Related Entries

- ▶ Face Recognition, Component-based
- ▶ Face Recognition, Overview
- ▶ Face Recognition, Sketch-based
- ▶ Face Tracking

References

1. Brunelli, R., Poggio, T.: Face recognition: Features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(10), 1042–1052 (1993)
2. Kanade, T.: Picture processing system by computer complex and recognition of human faces. Ph.D. thesis, Kyoto University (1973)
3. Ding, L., Martinez, A.: Precise detailed detection of faces and facial features. In: Proceedings of IEEE Computer Vision and Pattern Recognition, pp. 1–7 (2008)
4. Gao, Y., Leung, M.: Face recognition using line edge map. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(6), 764–779 (2002)
5. Leung, M.K., Yang, Y.H.: Dynamic two-strip algorithm in curve fitting. *Pattern Recognit.* **23**(1–2), 69–79 (1990)
6. Heisele, B., Serre, T., Poggio, T.: A component-based framework for face detection and identification. *Int. J. Comput. Vis.* **74**(2), 167–181 (2007)
7. Ahonen, T., Hadid, A., Pietikainen, M.: Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12), 2037–2041 (2006)
8. Turk, M., Pentland, A.: Eigenfaces for Recognition. *J. Cogn. Neurosci.* **3**(1), 71–86 (1991)
9. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997)
10. Allwein, E.L., Schapire, R.E., Singer, Y.: Reducing multiclass to binary: a unifying approach for margin classifiers. *J. Mach. Learn. Res.* **1**, 113–141 (2001)
11. Nowak, E., Jurie, F.: Learning visual similarity measures for comparing never seen objects. In: IEEE Conference on Computer Vision and Pattern Recognition (2007)
12. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. *J. Mach. Learn. Res.* **36**(1), 3–42 (2006)
13. Lanitis, A., Taylor, C.J., Cootes, T.F.: A unified approach to coding and interpreting face images. In: Proceedings of the International Conference on Computer Vision, pp. 368–374. IEEE Computer Society, Washington, DC, USA (1995)
14. Wiskott, L., Fellous, J.M., Kröger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 775–779 (1997)

15. Sivic, J., Everingham, M., Zisserman, A.: Person spotting: Video shot retrieval for face sets. In: 4th International Conference on Image and Video Retrieval, pp. 226–236 (2005)
16. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
17. Meyers, E., Wolf, L.: Using biologically inspired features for face processing. Int. J. Comput. Vis. **76**(1), 93–104 (2008)
18. Huang, G., Jain, V., Learned-Miller, E.: Unsupervised joint alignment of complex images. Computer Vision, In: IEEE International Conference pp. 1–8 (2007)
19. Zhou, Y., Gu, L., Zhang, H.J.: Bayesian tangent shape model: estimating shape and pose parameters via bayesian inference. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. I–109–I–116 (2003)

usually achieves significantly higher performance than the VIS approach.

Introduction

Face recognition should be performed based on intrinsic factors of the face, related to the 3D shape and albedo of the facial surface. In contrast, extrinsic factors, including eyeglasses, hairstyle, expression, posture, and lighting should be minimized because they make distributions of face data highly complex.

Among the aforementioned extrinsic factors, problems with uncontrolled environmental (ambient) illumination is the important issue [1]. Illumination direction is the most critical of all [2]. From the end-user point of view, a biometric system should adapt to the environment. However, face recognition systems based on face images captured in visible light (VIS) spectrum are compromised of changes in environmental illumination, even for cooperative user applications with frontal faces captured indoor. Numerous publications exist for modeling and normalizing face illumination conditions. They are found to improve recognition performance, but have not led to a face recognition method which is illumination independent.

3D face recognition provides a solution to the illumination problem. Disadvantages of it include increased cost, lesser speed, and specular reflections. It is reported that the 3D methods do not necessarily produce better recognition results than the 2D methods [3].

Imaging and vision beyond the visible spectrum has recently received much attention in the computer vision community (e.g., [4]). Radiation spectrum ranges are shown in Fig. 1. Instead of ultraviolet radiation which is harmful to the human body, thermal-infrared and near infrared (NIR) imagery are employed for face recognition applications. Such “invisible” spectrum imaging technologies are effective in dealing with uncontrolled illumination. This is because they work in different bands, from the conventional VIS imaging to many visual effects, as encountered in conventional illumination changes can be eliminated. Disadvantages of the FIR approach include instability due to environmental temperature, emotional and health conditions, and poor eye localization accuracy [5]. The use of active near infrared (NIR) imaging brings

Face Recognition, Near-Infrared

STAN Z. LI, DONG YI

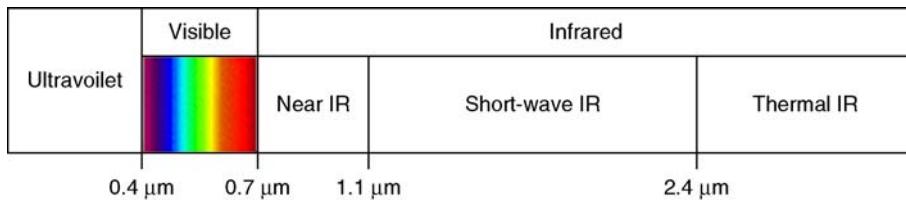
Biometrics and Security Research & National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China

Synonyms

Near-infrared image based face recognition; Face recognition in near-infrared spectrum

Definition

Near-infrared (NIR) based face recognition, as opposed to the conventional *visible light* (VIS) based, is an effective approach in overcoming the impact of illumination changes on face recognition. It uses a special purpose imaging capture hardware, in which active NIR lights mounted around the camera lens illuminate the face from near frontal direction and an NIR camera captures front-lighted NIR face images. This is similar to a camera flash but the imaging is done in the *invisible* NIR spectrum. With such NIR face images, problems caused by uncertainties in uncontrollable environmental **illumination** are minimized, and difficulties in building the face matching engine is much alleviated. The NIR approach



Face Recognition, Near-Infrared. **Figure 1** Radiation spectrum ranges.

a new dimension for face detection and recognition [6, 7, 8, 9, 10, 11].

The key part in the NIR face recognition approach is a special purpose image capture hardware system [10, 11]. It uses active NIR illuminators, for example, ► **light-emitting diodes** (LEDs), mounted around the camera lens to illuminate the face from near front direction and then capture front-lighted NIR face images. This is similar to a camera flash but as the NIR lighting works in *invisible* NIR spectrum it is non-intrusive to human eyes.

An NIR face image with frontal illumination is subject mainly to an approximately monotonic transform in the gray tone, and problems caused by uncertain environmental illumination are minimized. Therefore, the face detection and matching algorithms need to cope with this degree of illumination changes mainly. This is much less difficult than the problems with conventional VIS face images.

The NIR approach usually achieves significant higher performance than the VIS approach [11]. The use of NIR techniques leads to highly accurate and fast face recognition systems for cooperative face recognition applications, indoors [10, 11] and outdoors [12]. The use of NIR face images for biometrics is now being evaluated by NIST [13].

A limitation, however, is that both enrollment and the query face images should be of the NIR type, which similar to the requirement for 3D face recognition. Methods for matching the NIR query and VIS target images that are required for photo IDs, are being developed [14].

NIR Imaging Hardware

The goal of making this special-purpose hardware is to overcome the problem arising from uncontrolled environmental light so as to produce face images of a good

illumination condition for face recognition. “A good illumination condition” means that the lighting is from the frontal direction and the image has suitable pixel intensities.

This could be achieved by the following methods: (1) Active NIR light can be mounted (e.g., 850nm LEDs) around the camera lens to provide strong frontal lighting enough to override environmental light, and set a low camera exposure to produce a clear frontal-lighted face image. (2) a ► **long-pass optical filter** can be used to further minimize the remaining environmental lighting by cutting off visible light of wavelength shorter than 750nm. Fig. 2 illustrates an example of the hardware device, and resulting face images. The face in the images are illuminated by NIR LED light from the front and a lamp aside, in addition to other environmental light.

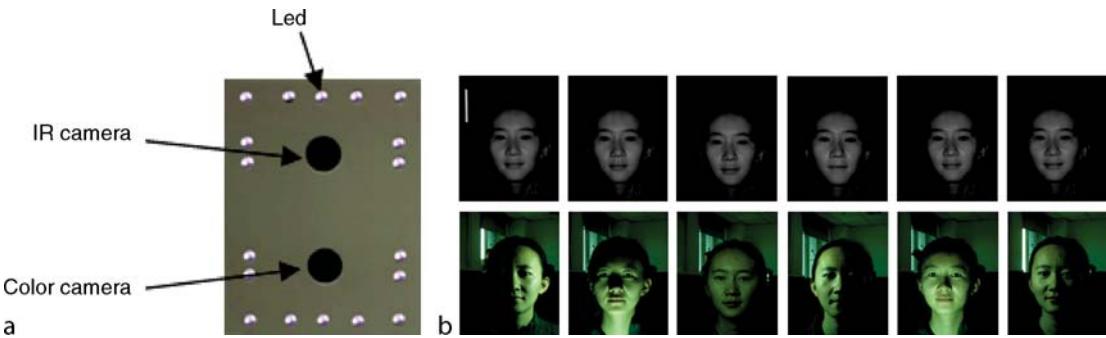
In outdoor environments, the sunlight contains much stronger NIR component than NIR LEDs. The hardware must be further designed to minimize influence of the sunlight to maintain the effect of the active NIR illumination. It could be enhanced by using a strong active NIR pulse illuminator and NIR camera, co-working in a synchronized manner [12].

Illumination Invariant Face Representation

According to the ► **Lambertian law**, an image $I(x, y)$ under a point light source is formed according to the following equation

$$I(x, y) = \rho(x, y)\mathbf{n}(x, y)\mathbf{s} \quad (1)$$

where $\rho(x, y)$ is the albedo of the facial surface material at the point (x, y) , $\mathbf{n}=(n_x, n_y, n_z)$ is the surface normal (a unit row vector) in the 3D space, and $\mathbf{s}=(s_x, s_y, s_z)$ is



Face Recognition, Near-Infrared. [Figure 2](#) An experimental active NIR imaging device (with an additional color camera), and NIR versus color images captured under different environmental lightings. While unfavorable lighting changes are obvious in the color images, they are almost unseen in the NIR images.

Local window			Thresholded			Weights		
18	15	8	1	0	0	8	4	2
21	18	6	1	1	0	16	64	128
27		23	1		1	32		
LBP string = (0001111)			LBP code = 0+0+0+8+16+32+64+128 = 248					

Face Recognition, Near-Infrared. [Figure 3](#) LBP code for 3x3 window.

the lighting direction (a column vector, with magnitude). Here, albedo $\rho(x, y)$ reflects the photometric properties of facial skin and hairs, and $\mathbf{n}(x, y)$ is the geometric shape of the face.

The LEDs mounted around the camera lens are approximately co-axial to the camera direction, and thus provide the best possible straight frontal lighting. In this case, the image can be approximated by

$$I(x, y) = \kappa \rho(x, y) n_z(x, y) \quad (2)$$

where $n_z(x, y)$ is the depth information (2.5 map) that can be acquired by a range imaging system, and κ can be modeled as being monotonic to the distance between the face and the active light.

The degree of freedom due to the monotonic transform of κ may be compensated by applying some operator, such as local binary pattern (LBP), on the NIR image to produce a genuine illumination invariant face representation [11]. The basic form of the LBP operator is illustrated in [Fig. 3](#). The binary bits describing a local 3 x 3 subwindow are generated by thresholding the 8 pixels in the surrounding locations

by the gray value of its center; the feature vector is formed by concatenating the thresholded binary bits anti-clockwise. The LBP code does not change with any monotonic transform of the image. Therefore, applying an LBP operator to an active NIR image generates illumination invariant features for faces. A highly accurate face recognition system can then be built.

Summary

The NIR approach uses an active NIR imaging hardware to acquire front-illuminated face images, to overcome the problem of illumination variation that every face recognition system has to deal with. NIR face images have good properties and render extraction of illumination invariant face features for building accurate face recognition systems. The use of NIR face images for face biometrics is now being evaluated by NIST [13].

A limitation of the NIR approach, however, is that both enrollment and the query face images should be of the NIR type. Methods for matching the NIR query

and VIS target images that are required for photo IDs are yet to be developed.

Related Entries

- ▶ Face Recognition Overview
- ▶ Hyperspectral and Multispectral Biometrics
- ▶ Local Binary Pattern (LBP)

References

1. NIST: Face Recognition Vendor Tests (FRVT). <http://www.frvt.org> (2006)
2. Adini, Y., Moses, Y., Ullman, S.: Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 721–732 (1997)
3. Chang, K.I., Bowyer, K.W., Flynn, P.J.: An evaluation of multimodal 2D+3D face biometrics. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 619–624 (2005)
4. OTCBVS. In: IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum. (2004–2005)
5. Chen, X., Flynn, P.J., Bowyer, K.W.: Infra-red and visible-light face recognition. *Comput. Vis. Image Understand.* **99**, 332–358 (2005)
6. Dowdall, J., Pavlidis, I., Bebis, G.: Face detection in the near-IR spectrum. *Image. Vis. Comput.* **21**, 565–578 (2003)
7. Li, D.Y., Liao, W.H.: Facial feature detection in near-infrared images. In: Proceedings of fifthth International Conference on Computer Vision, Pattern Recognition and Image Processing, Cary, NC, pp. 26–30 (2003)
8. Pan, Z.H., Healey, G., Prasad, M., Tromberg, B.: Face recognition in hyperspectral images. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 1552–1560 (2003)
9. AuthenMetric Co. Ltd.: A Method for Face Image Acquisition and a Method and System for Face Recognition. Patent No.PCT/CN2004/000482 (2004)
10. Li, S.Z., His Face Team: AuthenMetric F1: A Highly Accurate and Fast Face Recognition System. ICCV2005 - Demos (2005)
11. Li, S.Z., Chu, R., Liao, S., Zhang, L.: Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.* **26** (2007)
12. Yi, D., Liu, R., Chu, R., Liu, D., Wang, R., Li, S.Z.: “Outdoor face recognition using enhanced near infrared imaging”. In: Proceedings of IAPR International Conference on Biometric, Seoul, Korea (2007)
13. NIST: Multiple Biometric Grand Challenge (MBGC). <http://face.nist.gov/mbgc> (2008)
14. Yi, D., Liu, R., Chu, R., Lei, Z., Li, S.Z.: Face matching between near infrared and visible light images. In: Proceedings of IAPR International Conference on Biometric, Seoul, Korea (2007)

Face Recognition, Overview

ALEX M. MARTINEZ

Department of Electrical and Computer Engineering,
Ohio State University, Columbus, OH, USA

Synonyms

Face Biometric; Face Identification; Face Verification

Definition

Face recognition is the science which involves the understanding of how the faces are recognized by biological systems and how this can be emulated by computer systems. Biological systems employ different types of visual sensors (i.e., eyes), which have been designed by nature to suit a certain environment where the agent lives. Similarly, computer systems employ different visual devices to capture and process faces as best indicated by each particular application. These sensors can be video cameras (e.g., a camcorder), infrared cameras, or among others, 3D scans. The essay reviews some of the most advanced computational approaches for face recognition defined till date.

Introduction

Many types of biometrics exist for identifying a person or verifying that a given individual is what he or she claims to be. Some of the biometrics result in quite reliable recognition and verification, but most are either intrusive to the individual or expensive (e.g., DNA or iris). Furthermore, many of the biometrics have raised reasonable questions about an individual's rights and personal freedom [1]. The systems that are typically considered less intrusive by people, are those based on the recognition of faces.

We are so used to seeing and recognizing faces that most people think computers should have such a capacity too. Computer face recognition allows devices to recognize and interact with users, allowing them to go beyond the boring and slow use of the keyboard and mouse. The face carries so much information that

people find it difficult to interact on the phone for long, forcing companies to include cameras and even video on cell phones – even if the most recorded sequence is perhaps a simple “Hi.” Yet, people feel that the smile associated with a simple greeting is essential to start a good conversation or for carrying the real and intended thought of the messenger. The face also provides the identity of the speaker, avoiding the awkwardness of trying to identify someone by voice alone. This effect is now clearer than ever, with video chats becoming increasingly popular as high speed Internet becoming available at low costs to the general public.

It is thus not surprising that people still remain open to the possibility of having face recognition systems at home, work, or other places like at the ATM. One concern with this technology is to make sure that the biometrics of the individuals cannot be stolen. Imagine a scenario where a hacker steals information from a database of faces and then employs this to hack other computer, systems or institutions with a stolen identity. A password or an ID card can be changed, but a face cannot be. To address these concerns, researchers in face recognition are developing mechanisms to encrypt personal biometrics. One classical solution is to define a mapping function which maps a face image into a single instance (e.g., feature vector). The trick is to use a function whose inverse mapping is not unique (i.e., the inverse mapping results in multiple solutions) unless you know the encryption key [2]. In face recognition, we may even be able to eliminate the need for the encryption key. This can be achieved by defining a recognition algorithm in the encrypted space. This is possible because of the uniqueness of direct mapping. This means we can perform face recognition even when the understanding the information stored in our own database is not possible (i.e., in the sense, the image do not have meaning for the human visual system any longer). This could mean that general databases of faces could be shared by several institutions, because these cannot abuse its contents. Also, if the database of face images is stolen, unauthorized users would not be able to make sense of its data. These security protocols generally make face recognition systems more acceptable by the general population.

Perhaps the most important disadvantage of face recognition is that it cannot provide as accurate an identification as other biometrics, and definitely not as accurate as DNA or iris. Nonetheless, in a large

number of applications such a secure analysis is not needed. One of the most classical examples is in human–computer interaction – the cell phone example given above being but one example of its potential uses. Another typical example application is whenever individuals need to gain access to restricted areas within a company. This is of particular use where the employees are known *a priori*. One well-known case is in airports, where not all personnel have access to the runway or planes. A related application is for customer verification, for example for airline tickets, where the manual picture to face check is known to be flawed. In the 2008 summer Olympics, the organizer gave the opportunity to attendees of the inaugural ceremony to attach a picture of their face to their tickets. At the ceremony, the holders of these tickets were asked to look at a camera and a computer compared the face of the ticket holder with that of the buyer. A mismatch prompts the organizers to request additional information to demonstrate that the identity of the ticket holder and buyer is the same.

What makes all these applications and many others possible is the tremendous advances that have been accomplished in the past years in the area of computer algorithms for automatic face recognition. Current systems are able to recognize faces under a large number of variations; sometimes surpassing human performance. However, to accomplish this, many problems need to be addressed. The most relevant are detailed further.

Problems a Face Recognition System Needs to Address

In real life, faces appear under a variety of conditions. The most common ones include the following:

- **Pose:** Faces move in 3D space. When captured by a 2D camera, a large variety of 2D images corresponding to the same face can be obtained. Alternatively, one can use 3D scans, but these generally require the cooperation of the subject and are more expensive.
- **Illumination:** Different ambient lighting results in very distinct texture patterns of the face. One illumination will emphasize one type of face texture, while a different lighting will accentuate another. The shape of the face is also affected, because

different illumination angles will cause completely distinct cast shadows.

- **Expression:** Faces are a fundamental means by which we express emotions as well as other cues related to human communication. This is achieved by employing a large collection of the face muscles underlying the skin. With different movements come different expressions, each of which results in a distinct facial appearance.
- **Occlusions:** In most applications, partial occlusions are a common occurrence. These may be caused by clothing, glasses (including sunglasses), self-occlusions (such as a hand), clutter, etc. This means that when pictures of faces are taken, not all the information is always available. In fact, when 2D images are used, only a portion of the face is visible. The 3/4 view provides the most information, but even this orientation misses information because the face is not symmetric.
- **Imprecisely localized faces:** A less known problem of dealing with faces is that it cannot be precisely located or cannot be robustly delineated from an image or a video sequence. One reason is that it is generally impossible to determine where a face or facial feature starts and ends. To see this, an image of a face can be uploaded on favorite image software. Next, the inner corner of the left eye is zoomed in until the pixels become large squares. It can be seen at the pixel level that it is almost impossible to determine where the inner corner of left eye is. This problem makes the process of face detection difficult – even when we try to perform segmentation by hand.

Any successful algorithm for face recognition has to address some or all of these problems [3–6]. The great advances in recent years should be looked up with gratitude for there are algorithms now to partially solve each and every one of them. However, this may require tuning the approach to each application. There is a whole spectrum of techniques available to practitioners. The methodologies defined over the years vary considerably, from shape- to appearance-based recognition. The algorithms that are predominantly based on the shape of the face require extraction of the outline of the facial components, which has only been partially resolved recently [7, 8]. The appearance-based approach simplifies some of these requirements. In this alternative approach, one uses the

brightness of the pixels defining the face as features for representing and recognizing images [9, 10]. Nonetheless, the correct definition of the approach still requires that the faces be warped to a “standard” shape, which involves the detection of some of the major facial components (e.g., eye centers) [3]. These approaches are summarized below.

Different Approaches to Face Recognition

The first step in understanding “Face Recognition” and designing systems that can do automatic recognition will undoubtedly be that of describing the face (see Anatomy of Face). The face is an articulate object capable of amazing transformations. While the underlying bone structure defines who we are – our identity and part of our heritage – the muscles overlaying it shape our personality. Muscles are also fundamental for the recognition of emotions and other communicative cues, although recent results [11, 12] demonstrate that other factors are also in the play and ought to be considered.

After the face anatomy has been studied, understanding how to acquire and design automatic systems for face recognition is needed, including a review of “Face Sample Quality”. Sometimes it is taken for granted that the quality of the images we capture is the optimal for computational analysis, which is generally not the case.

After the basic components of the face and face recognition systems have been introduced, the processes of acquisition to recognition using algorithmical components are discussed. The first step is to determine the location of the face or faces in the image or video sequence (see Face Detection). Following this is the aspect of “Face Tracking,” which is about how to track the motion of the face within a video sequence without the need to detect in each individual frame. Face detection and face tracking approaches are important because they either outperform the rest or because their mathematical formulation makes them appropriate in a large number of face recognition applications.

However, the processes of face detection and tracking do not generally suffice. If the problems of face recognition are to be appropriately addressed, faces are to be aligned before recognition – although this process can also be combined with that of identification

(► Face Alignment) and derive computer algorithms for aligning and warping faces ► Face Warping by means of previously devised models of faces (► Deformable Models).

The next aspect in face recognition is “Face Variations,” including problems of illumination, pose, and expression. As illumination influences the acquisition of face images and their subsequent recognition, it is important to know how to do recognition under varying illumination. Pose variations include approaches on how to model faces seen from different points of view. Expression variations refer to how we can design algorithms for the recognition of emotions and other facial expressions. Of particular note are the applications in human–computer interaction.

After all the different types of features that can be used to represent, model, and recognize faces have been introduced, the actual problem of classification is discussed, which includes recognition from shape and appearance, local and global components, video, 3D range data, near and thermal infrared, and sketch. Each of these methods has advantages and disadvantages that make them appropriate in some scenarios but not in others and form an important part of how face recognition is performed. A combination of approaches is under study [13] and more of them might be seen as technology improves and costs decrease.

Face recognition algorithms cannot be tested in the absence of well-defined databases, which are very important while deriving face recognition systems and we need to understand how different algorithms and systems compare with each other.

Finally, there is a need to understand progress in skin color modeling and skin texture. These are related topics to those described above and each of them can benefit from one another.

Summary

The essay discusses what the face is, how it varies, and how it can be modeled and recognized using computer algorithms: the face and its variations; algorithms for detection, tracking, and modeling; major approaches for recognition; and databases, evaluation protocols and alternative mechanisms of modeling and recognition have been discussed elsewhere in the encyclopedia.

Related Entries

- Anatomy of Face
- And-Or Graph Model for Face Representation, Sketching and Aging
- Biometrics, Overview
- Deformable Models
- Face Alignment
- Face Databases and Evaluation
- Face Localization
- Face Pose Analysis
- Face Recognition, Component-Based
- Face Recognition, 3D-Based
- Face Recognition, Near Infrared Based
- Face Recognition, Overview
- Face Recognition, Shape vs. Appearance-Based
- Face Recognition, Thermal Infrared Based
- Face Recognition, Video-based
- Face Sample Quality
- Face Tracking
- Face Variation
- Facial Expression Recognition
- Mis-alignment Analysis
- Skin Color Modeling

References

1. Jain, A.K.: Technology: Biometric recognition. *Nature* **449**, 38–40 (2007)
2. Teoh, A.B.J., Ngo, D., Goh, A.: Personalised cryptographic key generation based on facehashing. *Comput. Secur.* **23**(7), 606–614 (2004)
3. Martinez, A.M.: Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(6), 748–763 (2002)
4. M. Yang, D.J., Kriegman, N.A.: Detecting faces in images: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 34–58 (2002)
5. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. *ACM Comput. Surv.* **35**, 399–458 (2003)
6. Gross, R., Matthews, I., Baker, S.: Appearance-based face recognition and lightfields. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(4), 449–465 (2004)
7. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(6), 681–685 (2001)
8. Ding, L., Martinez, A.: Precise detailed detection of faces and facial features. In: Proceedings of IEEE Computer Vision and Pattern Recognition, Anchorage, AK, 23 June 2008

9. Sirovich, L., Kirby, M.: A lowdimensional procedure for the characterization of human faces. *J. Opt. Soc. Am. A* **4**(3), 519–524 (1986)
10. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**(1), 71–86 (1991)
11. Neth, D., Martinez, A.M.: Emotion perception in emotionless face images suggests a normbased representation. *J. Vis.* **9**(1), 1–11 (2009)
12. Zebrowitz, L.A.: Reading faces: window to the soul? Westview Press, Boulder, CO (1997)
13. Chang, K.I., Bowyer, K.W., Flynn, P.J.: An evaluation of multimodal 2d + 3d face biometrics. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(4), 619–624 (2005)

protecting high value assets (e.g. perimeter of government buildings) from asymmetric (i.e., terrorist) threats. Human facial signatures vary significantly across races in the visible band. This variability, coupled with dynamic lighting conditions, presents a formidable problem. For instance, face recognition under very low lighting is almost impossible using visible imagery. Reducing light variability through the use of an artificial illuminator is rather awkward in the visible band because it may be distracting to the eyes of the people in the scene and reveals the existence of the surveillance system.

Thermal IR imagery offers a promising alternative to visible imagery for handling variations in face appearance due to illumination changes [2], facial expression [3, 4], and face pose [4] more successfully. In particular, thermal IR imagery is nearly invariant to changes in ambient illumination [2, 3], and provides a capability for the identification under all lighting conditions including total darkness [4]. Therefore, while face recognition systems in the visible spectrum opt for pure algorithmic solutions into inherent phenomenology problems, systems employing thermal IR imagery have the potential to offer simpler and more robust solutions, improving performance in uncontrolled environments and deliberate attempts to obscure identity [5].

Face Recognition, Thermal

GEORGE BEBIS

Department of Computer Science and Engineering,
University of Nevada, Reno, NV, USA

Synonym

Face Recognition, Thermal Infrared

Definition

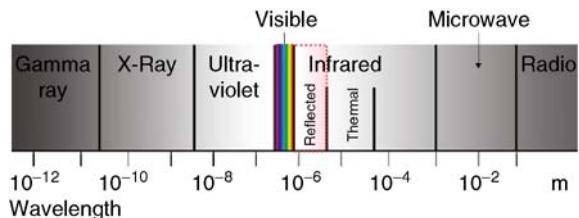
The human face emits thermal radiation which can be sensed by imaging sensors (i.e., thermal cameras) that are sensitive in the thermal infrared (IR) band of the ► **electromagnetic (EM) spectrum**. Temperature variations on the surface of the face produce a heat pattern, called a ► **thermogram**, which can be visualized as a 2D image (i.e., thermal image). Due to the presence of highly distinctive and permanent physiological characteristics under the facial skin, thermograms contain important information which can be exploited for face recognition.

Introduction

Considerable progress has been made in face recognition over the last decade [1], however, face recognition technology is not accurate or robust enough to be deployed in uncontrolled environments, for example,

Thermal IR Spectrum

Imaging sensors sensitive in the visible spectrum respond to ► **electromagnetic radiation** in the range (0.4–0.7 μ), while sensors sensitive in the IR spectrum respond to electromagnetic radiation in the range 0.7–14.0 μ (see Fig. 1). In general, the IR spectrum can be divided into two primary bands: the *reflected* IR and the



Face Recognition, Thermal. **Figure 1** The electromagnetic (EM) spectrum.

thermal IR bands. The reflected IR band ($0.7\text{--}2.4\ \mu\text{m}$) corresponds to reflected solar radiation and contains no information about the thermal properties of objects. It can be divided into two subbands: the near-IR (NIR) ($0.7\text{--}0.9\ \mu\text{m}$) and the short-wave infrared (SWIR) ($0.9\text{--}2.4\ \mu\text{m}$) bands.

The thermal IR band ($2.4\text{--}14.0\ \mu\text{m}$) corresponds to thermal radiation emitted from objects. Temperature variations on the surface of an object produce a heat pattern, called *thermogram*, which can be visualized as a 2D image (i.e., thermal image). The amount of emitted radiation depends both on the temperature and the emissivity of the objects [6]. The thermal IR band can be divided into two subbands: the mid-wave infrared (MWIR) with range $3.0\text{--}5.0\ \mu\text{m}$ and long-wave infrared (LWIR) with range $8.0\text{--}14.0\ \mu\text{m}$. It should be mentioned that there are strong atmospheric absorption bands at $2.4\text{--}3.0\ \mu\text{m}$ and at $5.0\text{--}8.0\ \mu\text{m}$.

Due to the presence of highly distinctive and permanent physiological patterns under the facial skin (i.e., vein and tissue structure) [7], thermograms contain important information that can be exploited for face recognition. The human face emits thermal radiation both in the MWIR and LWIR bands of the thermal IR spectrum. However, thermal emissions of the skin are much higher in the LWIR band than in the MWIR band. As a result, face images have a much lower within-class variation in the LWIR spectrum. To analyze a thermal image, ► [radiometric calibration](#) is required. This is a process which achieves a direct relation between the value at a pixel of the thermal image and the absolute amount of thermal emission from the corresponding physical scene element. The goal is to standardize thermal IR images, independently of environmental conditions, cameras, and passage of time [3].

Recognition in the Thermal IR

Face recognition in the visible spectrum exploits the reflectance characteristics of the human face. As a result, changes in ambient illumination might degrade recognition performance. Face recognition in the thermal IR spectrum exploits physiological characteristics of the face by considering the thermal energy emitted from the face rather than the light reflected. Therefore, face recognition in the thermal infrared (IR) spectrum is nearly invariant to changes in ambient illumination.

Moreover, it is less sensitive to scattering and absorption by smoke or dust while the tasks of face detection and localization can be simplified considerably due to the fact that background clutter is typically not visible. Early overviews of face identification in the thermal IR spectrum can be found in [8–10]. A recent review on face recognition methods, both in the visible and thermal IR bands, can be found in [11] while a general review on multispectral face recognition methods, with emphasis on thermal IR, can be found in [12].

The effectiveness of visible versus IR spectrum was compared in an early study using several recognition algorithms in [13]. Using a database of subjects without eyeglasses, varying facial expression, and allowing minor lighting changes, it was found that there are no significant performance differences between visible and IR recognition across all the algorithms tested. In later studies [3, 14, 15], several popular appearance-based face recognition methodologies were tested under various lighting conditions and facial expressions. Results from these studies indicate superior performance for thermal IR-based recognition compared with that for visible-based recognition. These findings were confirmed in an operational scenario where images were captured both indoors and outdoors [16].

The effect of lighting, facial expression, and passage of time between the gallery and probe images were examined in [17]. Although IR-based recognition outperformed visible-based recognition assuming lighting and facial expression changes, it was found that IR-based recognition degrades when there is substantial passage of time between the gallery and probe images. In a related study [18], however, it was reported that both thermal IR and visible imagery degrade similarly with time passage. Improvements using fusion were reported in [16, 17]. Recognition using thermal IR was also shown to be less sensitive to changes in 3D head pose and facial expression in [4]. In [19], a statistical hypothesis pruning methodology was introduced for face recognition in thermal IR. First, each thermal IR face image was decomposed into spectral features using Gabor filters. Then, it was represented by a few parameters by modeling the marginal density of the Gabor filter coefficients using Bessel functions. Recognition was performed in the space of parameters of the Bessel functions.

Methodologically, the majority of the thermal IR face recognition methods reported in the earlier sections do not differ significantly from face recognition methods in

the visible band (i.e., appearance-based and feature-based). An exception is the method presented in [7] which explicitly exploits physiological information using the bioheat information present in thermal images. In particular, this method extracts the superficial blood vessel network which contains contour shapes quite characteristic of each individual. Matching is based on the branching points of the skeletonized vascular network. Using physiological features has the potential to improve thermal IR-based recognition by making recognition more robust to changes over time.

Limitations of Thermal IR

Despite its advantages, thermal IR has several drawbacks. First, it is sensitive to temperature changes in the surrounding environment. Currents of cold or warm air could influence the performance of systems using IR imagery. Second, it is sensitive to variations in the heat patterns of the face. Factors that could contribute to these variations include facial expressions (e.g. open mouth), physical conditions (e.g. lack of sleep, physical exercise), and psychological conditions (e.g. fear, stress, excitement). Third, thermal IR is opaque to glass. Glass blocks a large portion of thermal energy resulting in a loss of information near the eye region as shown on Fig. 2. Finally, radiometric calibration is required every time the environmental conditions change (e.g., moving the camera at a different location), a different camera is used (e.g., even if it is the same model), or data collections take place at different time intervals.

Fusion of Visible with Thermal IR Imagery

The benefits of fusing visible with thermal IR imagery have been documented in a number of studies including [3, 17, 20–23]. The idea is to combine the strengths of each spectral band to build more accurate and robust face recognition systems. For example, increased body temperature changes the thermal characteristics of the face, while there are not significant differences in the visible spectrum. Also, while eyeglasses completely occlude the eyes in the thermal IR spectrum, the problem is considerably less severe in the visible spectrum although visible imagery can suffer from highlights on the glasses under certain illumination conditions.

A summary of the fusion strategy reported in [21–23] for improving face recognition performance in the presence of eyeglasses is as follows.

Objects made of glass act as a temperature screen, completely hiding the parts located behind them. In the case of subjects wearing eyeglasses, this poses some major difficulties since the eyes would be occluded completely due to the fact that eyeglasses block thermal energy (i.e., see Fig. 2). Experimental results, reported in [21–23], illustrate that face recognition performance in the thermal IR degrades seriously when eyeglasses are present in the probe image but not in the gallery image and vice versa. To address this limitation, fusion of thermal IR with visible imagery was employed in [21–23]. Two different fusion strategies were investigated: *pixel-based fusion* in the wavelet domain, and *feature-based fusion* in the eigenspace domain. In both cases, fusion was carried out using Genetic Algorithms (GAs) [24].

The Equinox database [25] was used for experimentation. The database contains frontal faces under the following scenarios: (1) three different light directions – frontal and lateral (right and left); (2) three facial expression – “frown,” “surprise” and “smile”; (3) vocals pronunciation expressions – subjects were asked to pronounce several vocals from which three representative frames were chosen; and (4) presence of glasses – for subjects wearing glasses, all of these scenarios were repeated with and without glasses. For testing, the data were divided as follows: EG (expression frames with glasses, all illuminations), EnG (expression frames without glasses, all illuminations), EFG (expression frames with glasses, frontal illumination), ELG (expression frames with glasses, lateral illumination), EFnG (expression frames without glasses, frontal illumination), ELnG (expression frames without glasses, lateral illumination). The inclusion relations among these sets are as follows:

$$\begin{aligned} \text{EG} &= \text{ELG} \cup \text{EFG}, \\ \text{EnG} &= \text{ELnG} \cup \text{EFnG} \quad \text{and} \quad \text{EG} \cap \text{EnG} = \emptyset \end{aligned} \quad (1)$$

Recognition performance was measured by finding the percentage of the images in the test set, for which the top match is an image of the same person from the gallery. Figures 3 and 4 show the results obtained. Among the two fusion strategies tested, fusion in the wavelet domain yielded the best results. Nevertheless, fusion outperformed each modality alone in both cases.



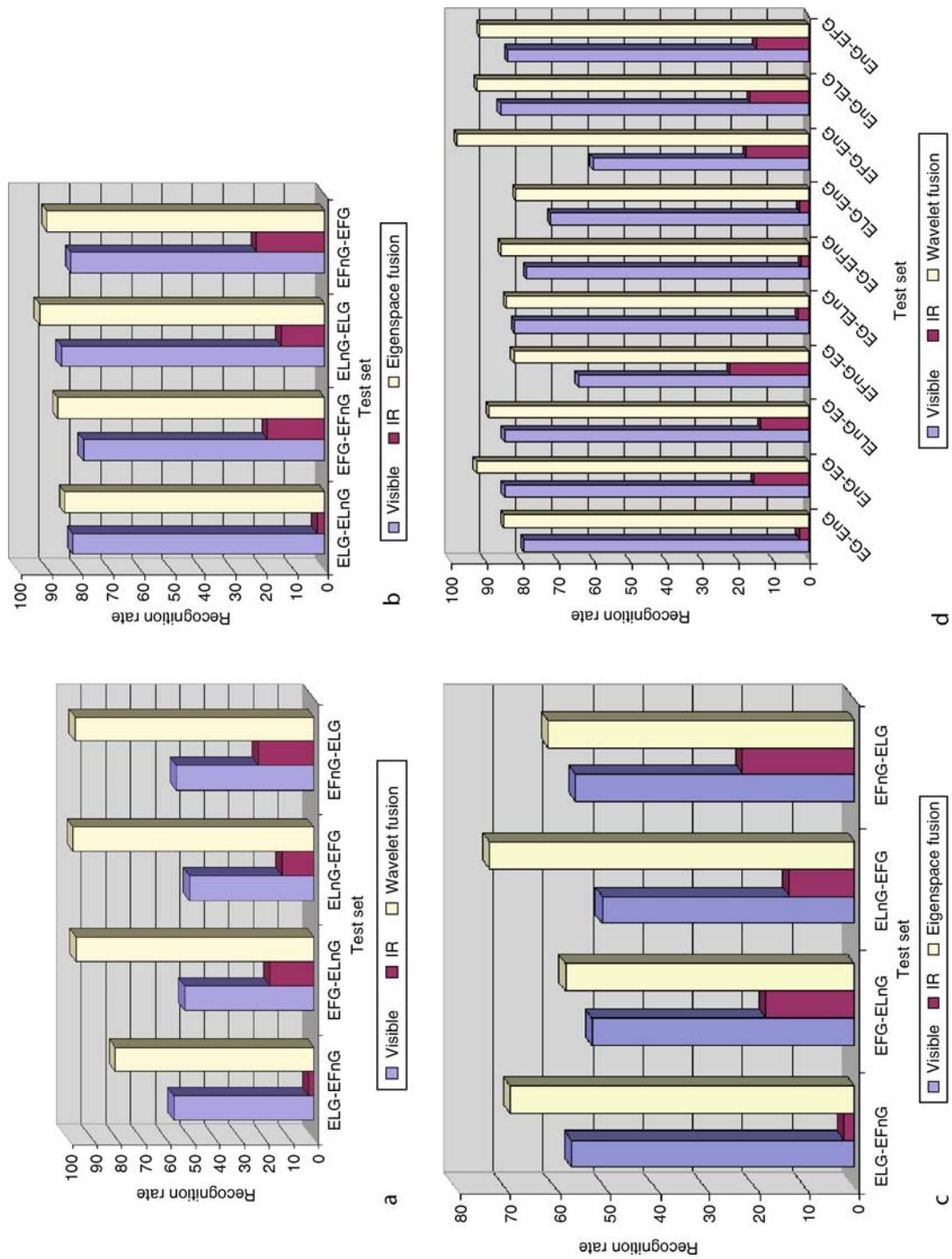
Face Recognition, Thermal. **Figure 2** (a, b) Visible images; (c, d) thermal IR images. It should be observed that since thermal IR is opaque to glass, the presence of eyeglasses blocks the eyes completely.

Milestones

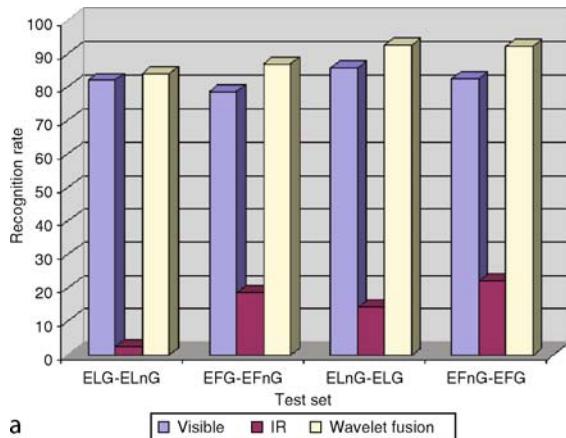
Despite its advantages, face recognition in the thermal IR spectrum has received relatively little attention compared to visible spectrum, mostly because of the following reasons:

- Higher cost of thermal sensors
- Lower image resolution
- Higher image noise
- Lack of widely available data sets

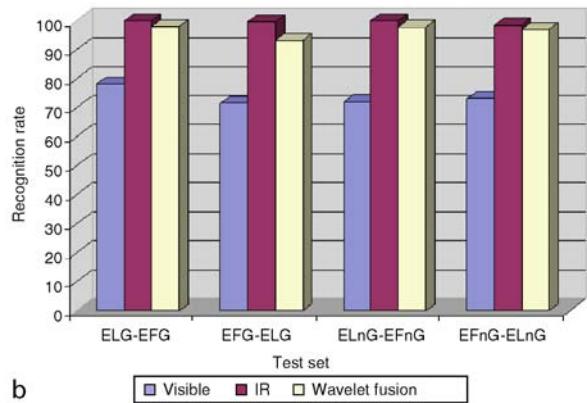
Advances in IR imaging technology and the availability of publicly available datasets, however, have facilitated experimentation with thermal imagery in the context of face recognition. While the difference in cost between visible and thermal imaging equipment is still large, the gap is closing rapidly as new uncooled microbolometer technologies enter the market. Companies, such as FLIR Systems (<http://www.flir.com>), offer a large variety of thermal cameras for different budgets. The issue of image noise can be addressed



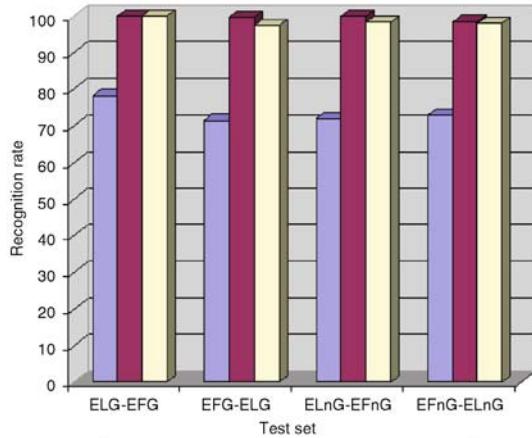
Face Recognition, Thermal. **Figure 3** Eye glasses results in the wavelet domain: (a) same illumination conditions – eyeglasses are not present both in the gallery and probe sets; (b) eyeglasses are present both in the gallery and probe sets – illumination conditions are different; (c) eyeglasses are not present both in the gallery and probe sets – illumination conditions are different; (d) similar to (c) except that the gallery and probe sets contain multiple illuminations.



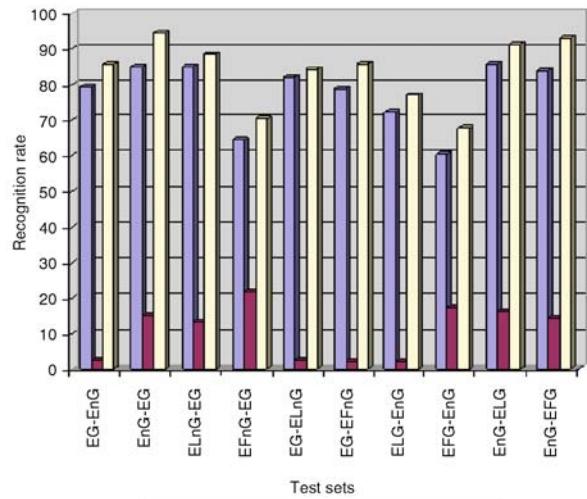
a [■ Visible ■ IR ■ Wavelet fusion]



b [■ Visible ■ IR ■ Wavelet fusion]



c [■ Visible ■ IR ■ Eigenspace fusion]



d [■ Visible ■ IR ■ Eigenspace fusion]

Face Recognition, Thermal. **Figure 4** Eyeglasses results in the eigenspace domain: (a) same illumination conditions – eyeglasses are not present both in the gallery and probe sets; (b) eyeglasses are present both in the gallery and probe sets – illumination conditions are different; (c) eyeglasses are not present both in the gallery and probe sets – illumination conditions are different; (d) similar to (c) except that the gallery and probe sets contain multiple illuminations.

using powerful radiometric calibration procedures while the issue of image resolution can be addressed using super-resolution techniques or fusing infrared with visible imagery. For example, Equinox Corporation (<http://www.equinoxsensors.com>) has made available a system for real-time fusion of thermal IR with visible imagery with image co-registration correction.

In terms of data, things are rather limited compared to the plethora of face databases available in the visible spectrum [26, 27]. The most extensive IR facial database, that is publicly available, is the Equinox database [25]. This database was created by Equinox Corporation under DARPA's HumanID program.

It includes coregistered visible/LWIR/MWIR/SWIR images and it is representative of unconstrained frontal imagery of people's faces in an indoor environment. Another, publicly available database, is the Notre Dame IR face database [28], which includes data captured at different sessions over time. Obviously, additional datasets are required to spark more research in this area, in particular, data depicting scenarios widely different from the imaging conditions during data acquisition (e.g., outdoor imagery). Introducing appearance variability due to various factors (e.g., metabolic activity) would be extremely useful in testing the robustness of thermal IR for face recognition.

Summary

While face recognition in the visible band performs satisfactorily under controlled conditions, thermal IR face recognition offers more advantages when there is no control over illumination or for detecting disguised faces. The passive nature of thermal IR systems lowers their complexity and improves their reliability. With dropping prices and technological advances, thermal IR is becoming more affordable and practical than before. Although thermal IR has many advantages, it suffers from several drawbacks including that it is sensitive to temperature changes and opaque to glass. A promising approach to deal with these issues is fusing visible with thermal IR imagery.

Related Entries

- ▶ Biometrics, Overview
- ▶ Face Recognition

References

1. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: A literature survey. *ACM Comput. Surveys* **35**(4), 399–458, (2003)
2. Wolff, L., Socolinsky, D., Eveland, C.: Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery. In: IEEE Workshop on Computer Vision Beyond the Visible Spectrum. Hawaii (2001)
3. Socolinsky, D., Selinger, A., Neuheisel, J.: Face recognition with visible and thermal infrared. *Comput. Vision Image Understand* **91**, 72–114 (2003)
4. Friedrich, G., Yeshurun, Y.: Seeing people in the dark: face recognition in infrared images. In: International Workshop on Biologically Motivated Computer Vision, pp. 348–359 (2002)
5. Pavlidis, I., Symosek, P.: The imaging issue in an automatic face/disguise detection system. In: IEEE Workshop on Computer Vision Beyond the Visible Spectrum, pp. 15–24 (2000)
6. Siegel, R., Howell, J.: Thermal Radiation Heat Transfer, 3rd edn. Taylor & Francis, London (1992)
7. Buddharaju, P., Pavlidis, I., Tsiamyrtzis, P., Bazakos, M.: Physiology-based face recognition in the thermal infrared spectrum. *IEEE Trans. Pattern Anal. Mach. Intell.*, **29**(4), 613–626 (2007)
8. Evans, D.: Infrared facial recognition technology being pushed toward emerging applications. In: Proceedings of SPIE, vol. 2962, pp. 276–286 (1997)
9. Prokoski, F., Riedel, R.: Infrared identifications of faces and body parts. In: Jain, A., Bolle, R., Pankanti, S. (eds.) *BIOMETRICS: Personal Identification in Networked Society*, pp. 191–212 (1999)
10. Prokoski, F.: History, current status, and future of infrared identification. In: IEEE Workshop on Computer Vision Beyond the Visible Spectrum. Hilton Head (2000)
11. Kong, S., Heo, J., Abidi, B., Paik, J., Abidi, M.: Recent advances in visual and infrared face recognition – a review. *Comput. Vision Image Understand.* **97**, 103–135 (2005)
12. Socolinsky, D.: Multispectral Face Recognition. In: *HandBook of Biometrics*, Jain, A., Flynn, P., and Ross, A. (eds.) pp. 293–313 (2008)
13. Wilder, J., Phillips, J., Jiang, C., Wiener, S.: Comparison of visible and infra-red imagery for face recognition. In: IEEE International Conference on Automatic Face and Gesture Recognition, pp. 182–187. Killington (1996)
14. Socolinsky, D., Wolff, L., Neuheisel, J., Eveland, C.: Illumination invariant face recognition using thermal infrared imagery. In: Computer Vision and Pattern Recognition Conference. Hawaii (2001)
15. Socolinsky, D., Selinger, A.: Comparative study of face recognition performance with visible and thermal infrared imagery. In: International Conference on Pattern Recognition, pp. 217–222 (2002)
16. Socolinsky, D., Selinger, A.: Thermal face recognition in an operational scenario. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2004)
17. Chen, X., Flynn, P., Bowyer, K.: Ir and visible light face recognition. *Comput. Vision Image Understand.* **99**, 332–358 (2005)
18. Socolinsky, D., Selinger, A.: Thermal face recognition over time. In: International Conference on Pattern Recognition (ICPR) (2004)
19. Srivastana, A., Liu, X.: Statistical hypothesis pruning for recognizing faces from infrared images. *Image Vision Comput.* **21**, 651–661 (2003)
20. Heo, J., Kong, S., Abidi, B., Abidi, M.: Fusion of visual and thermal signatures with eyeglass removal for robust face recognition. In: Workshop on Object Tracking and Classification Beyond the Visible Spectrum (2004)
21. Gyaourova, A., Bebis, G., Pavlidis, I.: Fusion of infrared and visible images for face recognition. In: European Computer Vision Conference (ECCV) (2004)
22. Singh, S., Gyaourova, A., Bebis, G., Pavlidis, I.: Infrared and visible image fusion for face recognition. In: SPIE Defense and Security Symposium (Biometric Technology for Human Identification) (2004)
23. Bebis, G., Gyaourova, A., Singh, S., Pavlidis, I.: Face recognition by fusing thermal infrared and visible imagery. *Image Vision Comput.* **24**(7), 727–742 (2006)
24. Goldberg, D.: *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, Reading, MA, USA (1989)
25. Equinox corporation, ir face database. <http://www.equinoxsensors.com/products/HID.html> (last visited in June 2008)
26. Gross, R.: Face Databases. In: Li, S.Z., Jain, A.K. (eds) *Handbook of Face Recognition*, pp. 301–327 (2005)
27. Face recognition homepage. <http://www.face-rec.org/> (last visited in June, 2008)
28. Computer vision laboratory, university of notre dame, biometrics database distribution. <http://www.nd.edu/cvrl/> (last visited in June, 2008)

Face Recognition, Video-Based

RAMA CHELLAPPA¹, GAURAV AGGARWAL¹,

S. KEVIN ZHOU²

¹University of Maryland, College Park, USA

²Siemens Corporate Research, Princeton, NJ, USA

Synonyms

Face recognition from image sequences; Video-based face recognition

Definition

Video-based face recognition is the technique of establishing the identity of one or multiple persons present in a video, based on their facial characteristics. Given the input face video, a typical video-based face recognition approach combines the temporal characteristics of facial motion with appearance changes for recognition. This often involves ► [temporal characterization of faces](#) for recognition, building 3D model or a super-resolution image of the face, or simply learning the appearance variations from the multiple video frames. The ability to generalize across pose, illumination, expression, etc. depends on the choice of combination. Video-based face recognition is particularly useful in surveillance scenarios in which it may not be possible to capture a single good frame as required by most still image based methods.

Introduction

Face recognition is one of the most successful applications in the vast amount of research on image analysis and understanding [1]. The fact that face recognition can be performed at a distance without subject's cooperation or knowledge makes it particularly attractive as compared to more reliable biometrics like fingerprints and iris or retinal scans. Traditionally face recognition has been limited to still images. Though great leaps have been made in recognizing faces from still images, more needs to be done to achieve the goal of recognizing faces in uncontrolled scenarios. Still image based approaches often struggle to truly generalize across variations in pose, expression, illumination, etc., leading to a not so satisfactory performance on real images.

The advent of inexpensive cameras and increased processing power has made it possible to capture and store videos in real time. Videos have the advantage of providing more information in the form of multiple frames making it relatively easier to generalize across variations that have been difficult with still images. Moreover, video makes it easier to track (or segment) faces which can then be fed into a recognition system. Importantly, psychological evidence indicates that dynamic information contributes to face recognition especially under nonoptimal viewing conditions [2]. These reasons form the basis of the recent interest in using videos for recognizing faces [3–5]. Though video provides extra information, the video feeds are almost always uncontrolled making it challenging to track and hence recognize faces.

Operation of a Video-based Face Recognition System

A typical Video-based Face Recognition (VFR) system operates by acquiring video feeds from one or multiple cameras, tracking and segmenting faces from the input feed(s), extracting representations to characterize the identity of the face(s) in the video, and then comparing them with the enrolled representations of subjects in the database. This constitutes the test phase of the system. During the enrollment (or training) phase, a similar sequence of steps is followed using one or multiple video feeds per identity and the corresponding composite representations are stored in the database. VFR approaches differ in the representation that is used to characterize the moving faces. An ideal VFR system performs these operations automatically without any human intervention. Though potentially a VFR system can operate in either verification mode (one-to-one matching) or identification mode (one-to-many), the real application of such a system lies in identifying subjects using surveillance cameras (say on an airport) without their knowledge. Therefore, a typical VFR system will often operate in what is known as watch list mode [6]. The watch list problem is a generalization of both identification and verification problems in which the system only attempts the identification of individuals on the watch list. The performance in this mode is measured using both identification rate and false alarm rate.

Challenges for Video-based Face Recognition Systems

Effective utilization/fusion of the information (both spatial and temporal) present in a video to achieve better generalization (for each subject) and discriminability (across different subjects) for improved identification is one of the biggest challenges faced by a VFR system. The fusion schemes can range from simple selection of good frames (which are then used for recognition in a still-image based recognition framework) to estimation of the full 3D structure of a face which can then be used to generalize across pose, illumination, etc. The choice may depend primarily on the operational requirements of the system. For example, in a surveillance setting, the resolution of the faces may be too small for reliable shape estimation. The choice also limits the recognition capability of the system. A simple good frame selection scheme will not have the capability to generalize appearance across pose variations and thus requires the test video to have some pose overlap with the gallery videos. Effective modeling of subject-specific facial characteristics from video data can only be achieved if the changes in facial appearance during the course of the video are appropriately attributed to different factors like pose changes, lighting, expression variations, etc. Unlike still image based scenarios, these variations are inherent in a VFR setting and must be accounted for to reap the benefits of extra information provided

by the video data. In addition, due to the nature of the input data, VFR is often addressed in conjunction with tracking problem which is a challenging problem by itself. In fact, more often than not, tracking accuracy depends on the knowledge of reliable appearance model (depends on the identity provided by the recognition module) while recognition result is dependent on the localization accuracy of the face region in input video.

Examples of Video-based Face Recognition Algorithms

Given the potential advantages video provides for the task of face recognition, relatively little work has been done to recognize faces in videos. The challenges in modeling moving faces along with the unavailability of large standard datasets have hindered the progress of research on VFR algorithms. [Table 1](#) gives a snapshot of a few existing VFR algorithms. As clear from the table, all the approaches have been tested on a very small sized (often private) datasets. The following discussion describes them in detail.

- Simultaneous Tracking and Recognition of Faces:* Traditional tracking-then-recognition approaches resolve uncertainties in tracking and recognition sequentially and separately, which often involves difficult choices (like criteria to select good frames and

Face Recognition, Video-Based. [Table 1](#) A snapshot of a few existing video-based face recognition algorithms

Algorithm	Short description	Experimental evaluation
Probabilistic recognition of human faces from video [7]	Simultaneous tracking-and-recognition using a time series state space model and sequential importance sampling	Private: 12 subjects, NIST: 30 subjects, MoBo [8]: 25 subjects
Video-based face recognition using probabilistic appearance manifolds [9]	Face modeled using a low-dimensional appearance manifold, approximated by piecewise linear subspaces	<i>Honda-UCSD</i> dataset: 20 subjects (52 videos)
Face verification through tracking facial features [10]	Tracks facial features defined on a grid with Gabor attributes using SIS algorithm	<i>Li</i> dataset: 19 subjects (2 sequences each)
Video-based face recognition using adaptive hidden markov models [11]	Statistics of training videos, and their temporal dynamics learnt by an HMM	Private: 12 subjects, MoBo [8]: 25 subjects
A system identification approach for video-based face recognition [12]	Face modeled as a linear dynamical system using ARMA model	<i>Honda-UCSD</i> dataset [9]: 30 subjects, <i>Li</i> dataset [10]: 19 subjects

estimation of registration parameters). Zhou et al. [7] avoid these issues while resolving uncertainties in tracking and recognition simultaneously in a unified probabilistic framework. The temporal information present in the video is fused using a time-series state-space model to characterize the evolving kinematics and identity. The three basic components of the model are as follows.

- A motion equation governing the kinematic behavior of the tracking motion vector. In its most general form, the motion equation can be written as

$$\theta_t = g(\theta_{t-1}, u_t); \quad t \geq 1, \quad (1)$$

where u_t is the noise that determines the transition probability $p(\theta_t | \theta_{t-1})$. The function $g(.,.)$ characterizes the evolving motion. It can either be a function learned offline or given a priori. Choice of θ_t is application dependent.

- An identity equation governing the temporal evolution of the identity variable.

$$n_t = n_{t-1}; \quad t \geq 1, \quad (2)$$

- An observation equation establishing the link between the motion vector and the identity variable.

$$\tau_{\theta_t}\{z_t\} = I_{n_t} + v_t; \quad t \geq 1, \quad (3)$$

where v_t is the observation noise that determines the observation likelihood $p(z_t | n_t, \theta_t)$ and $\tau_{\theta_t}\{z_t\}$ transforms the observation z_t to the chosen feature space.

Under the assumption of statistical independence between all noise variables and prior knowledge of the distributions $p(\theta_0 | z_0)$ and $p(n_0 | z_0)$, (1) and (2) can be combined as follows.

$$p(x_t | x_{t-1}) = p(n_t | n_{t-1})p(\theta_t | \theta_{t-1}), \quad (4)$$

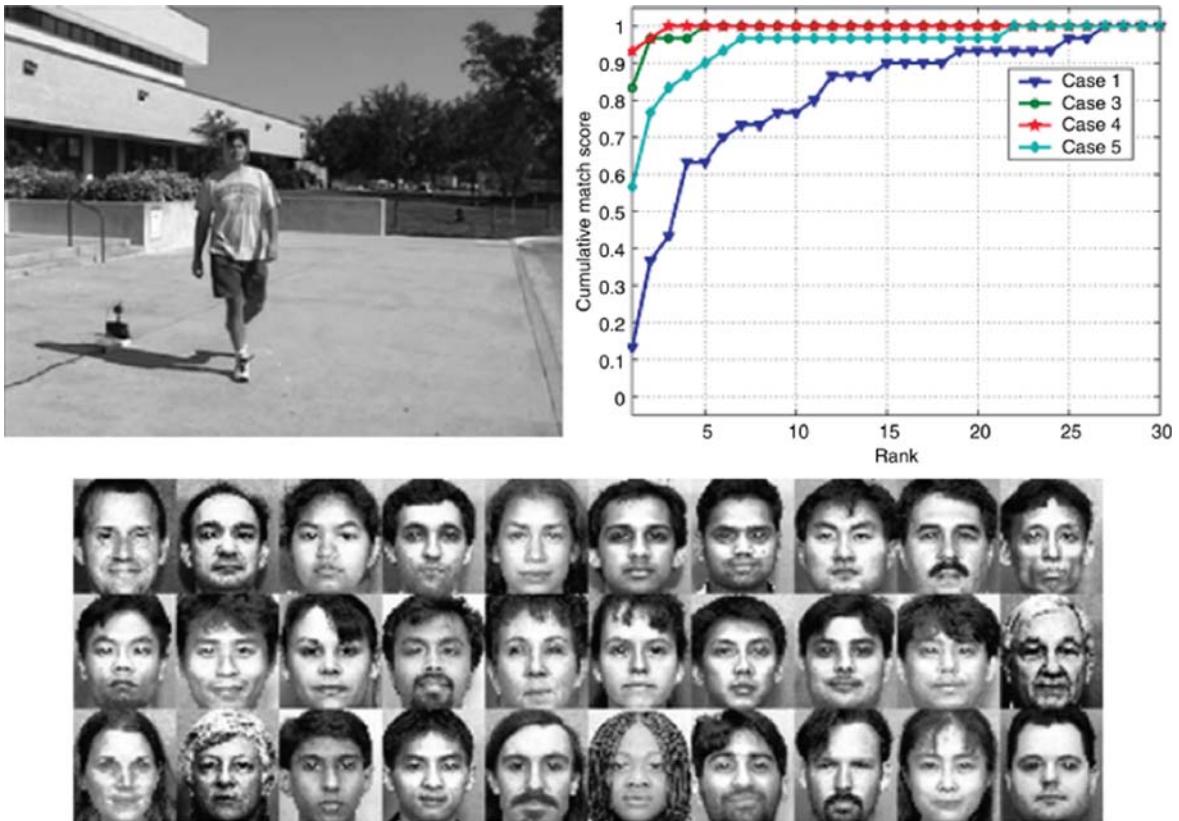
where $x_t = (\theta_t, n_t)$. Given a video sequence, the goal is to estimate the posterior probability $p(n_t | z_{0:t})$. The posterior probability is calculated using Sequential Importance Sampling (SIS) [13]. Using the SIS technique, the joint probability distribution of the motion vector and the identity variable is estimated at each time instant and then propagated to the next time instant as governed by the motion and identity equations. The marginal distribution of the identity variable is estimated to provide the desired identity result. Fig. 1 shows the performance

of the approach on a NIST dataset consisting of 30 persons gallery.

2. *Probabilistic Appearance Manifolds for VFR*: Similar to [7], Lee et al. [9] propose a VFR algorithm that performs modeling, tracking and recognition in one integrated framework. This is accomplished using a probabilistic appearance manifold based representation that is utilized simultaneously by both tracking and recognition modules. The recognition module uses tracker's output (the location of the face in the current frame) to update the current internal appearance model that is in turn used by the tracker.

Each face is characterized using a collection of linear subspaces in the image space which is constructed by clustering the exemplars from the input face videos. Each cluster often contains face images with similar poses and is represented using a PCA subspace. The collection of linear subspaces is further characterized using a transition matrix that captures the probabilities of moving from one pose subspace to another between two consecutive frames. The transition matrix is used to combine facial appearance with temporal coherency of pose variations to perform recognition. The approach has been tested on 52 video sequences of 20 different subjects.

3. *2D Feature Graph based Approach*: Li and Chellappa [10] propose a 2D feature-graph based approach for VFR in which the intensity model is replaced by a feature-graph using Gabor transform. The feature-graph approach is more robust to the variations in illumination and pose but possibly requires slightly higher-resolution videos. The tracking problem is formulated as a Bayesian inference problem for which Markov Chain Monte Carlo (MCMC) techniques are employed to obtain an empirical solution. A reparameterization is used to facilitate empirical estimation and to allow verification to be addressed simultaneously along with tracking. The facial features to be tracked are defined on a grid with Gabor attributes (Fig. 2). The motion of facial feature points is modeled as a global two-dimensional affine transformation (to account for head motion) plus a local deformation to account for the residual due to expression changes and modeling errors. The global motion is estimated by importance sampling while the residual motion is handled by incorporating local deformation into the likelihood measurement.

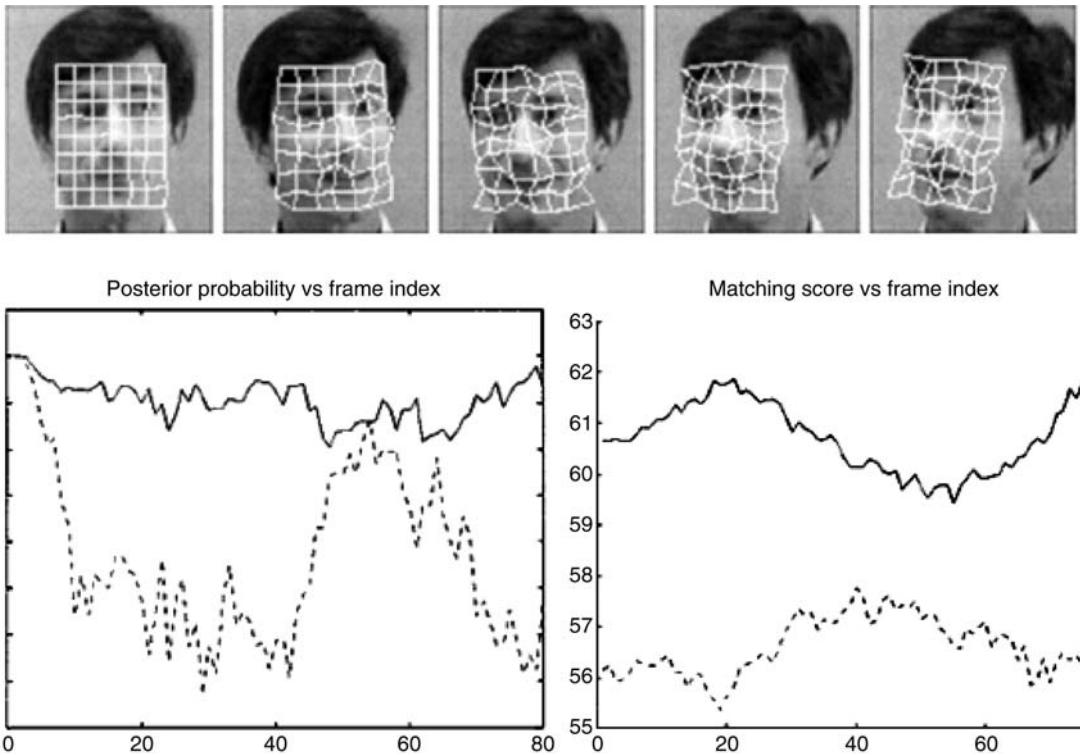


Face Recognition, Video-Based. **Figure 1** Sequential tracking and recognition [7]. Top left: A probe video from the NIST dataset; Top right: Recognition performance under different models; Bottom: Gallery set.

The temporal evolution of the jet positions is modeled as a dynamic system, where tracking is solved by analyzing this system, which in general, is non-Gaussian and nonlinear. Note that tracking is solved by analyzing the dynamic system governing the evolution of the changes in affine parameters. This reparameterization originates from a simple Taylor expansion. However, its novelty comes from the fact that one can choose different initial states for different purposes. If the initial state corresponds to a feature set from the first frame of a sequence, then the reparameterization is suitable for pure tracking. However, if the initial state represents some template from a candidate list, then the reparameterization is naturally good for tracking-for-verification. When a template and the sequence belong to the same person, tracking results should reflect a coherent motion induced by the same underlying shape. On the other hand, a more random motion pattern will often be observed when the template and the sequence belong to different persons. Thus, with

different templates, such a tracker allows verification to be addressed simultaneously with tracking. The motion coherence in a shape is evaluated by calculating the posterior probabilities from the estimated densities on a region centered on the mean shape. Fig. 2 shows the tracking and verification results using this approach.

4. *Hidden Markov Models for VFR:* Li and Chen [11] propose adaptive Hidden Markov Models (HMM) to recognize faces in videos. During training, a separate HMM is learnt for each subject in the gallery to characterize appearance statistics and temporal dynamics of the facial motion. Recognition is performed by analyzing the test video by HMMs corresponding to subjects in the gallery. During the recognition process, test video sequences are used to update the gallery models in an unsupervised fashion based on the recognition result. The approach has been tested on two datasets with 21 and 24 subjects respectively.
5. *3D Model based Approach:* As opposed to most VFR approaches which model face as a 2D object, the



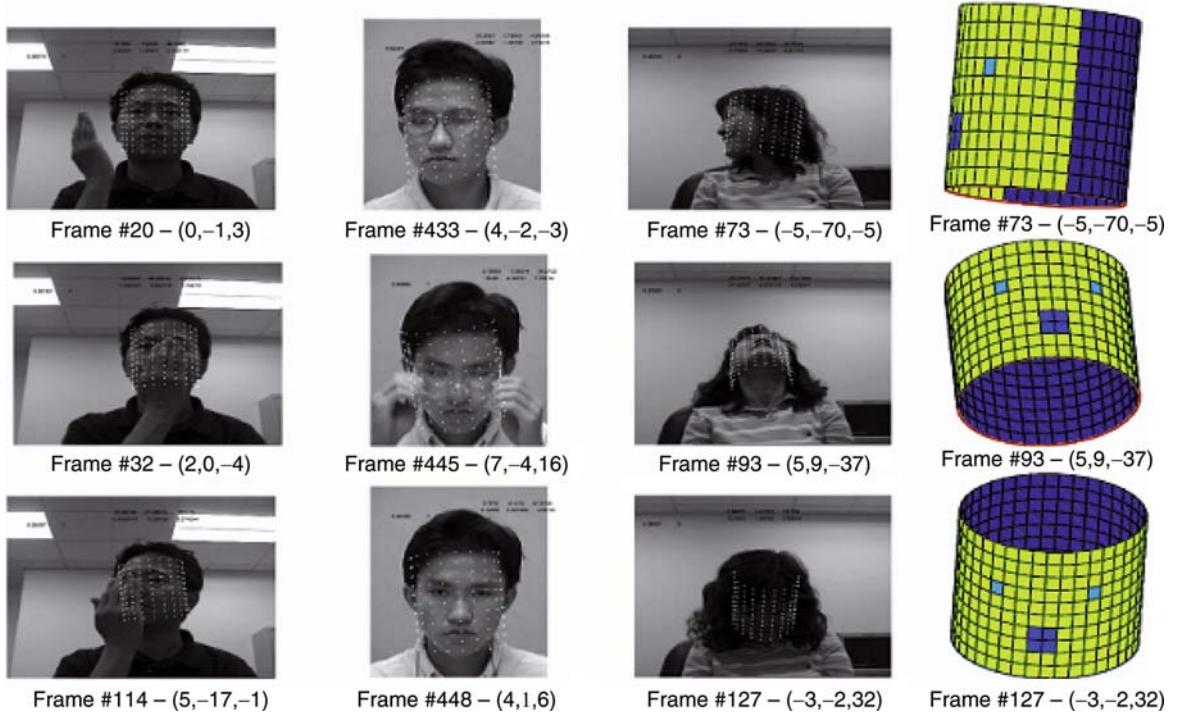
Face Recognition, Video-Based. [Figure 2](#) 2D feature-based approach [10]. Top: Tracking result; Bottom left: Posterior probabilities for the true (solid) and an impostor (dashed) hypothesis; Bottom right: Matching scores for the true (solid) and impostor hypothesis.

algorithm proposed in [14] estimates the 3D configuration of the head in each frame of the video. The 3D configuration consists of three translation parameters and three orientation parameters which correspond to the yaw, roll and pitch of the face. The approach combines the structural advantages of geometric modeling with the statistical benefits of a particle-filter based inference. The face is modeled as the curved surface of a cylinder which is free to translate and rotate in an unprescribed manner. The geometric modeling takes care of pose changes and self-occlusions while the statistical modeling handles unexpected occlusions and illumination variations during the course of the video. The recovered 3D facial pose information can be used to perform pose normalization which makes it very useful for the tasks of face modeling, face recognition, expression analysis, etc.

The estimation of 3D pose of a face in each frame of a video is posed as a dynamic state estimation problem. Particle filtering is used for estimating the unknown dynamic state of a system from a collection of

noisy observations. Such an approach involves two components: 1) a state transition model to govern the motion of the face, and 2) an observation model to map the input video frames to the state (3D configuration). [Figure 3](#) shows the tracking results for a few video frames. The estimated pose is shown in the form of an overlaid cylindrical grid. The accuracy in recovering 3D facial pose information makes it viable to perform VFR without any need for pose overlap between the gallery and test video. Recognition experiments are performed on videos with nonoverlapping poses. For each face, a texture mapped cylindrical representation is built using the recovered facial pose information, which is used for matching. The approach has been tested on a small dataset consisting of 10 subjects.

6. *Shape-Illumination Manifold for VFR:* In [15], Arandjelovic and Cipolla propose a generic shape-illumination manifold based approach to recognize faces in videos. Assuming the intensity of each pixel in an image to be a linear function of the corresponding albedo, the difference in two



Face Recognition, Video-Based. **Figure 3** 3D model-based approach [14]. The last column shows the pose for the frames in the third column.

logarithm-transformed images of the same subject in the same pose, depends only on 3D shape of the face and the illumination conditions in the input images. As the pose of the subject varies, the difference-of-log vectors describe manifold called as shape-illumination manifold in the corresponding vector space. Assuming shape variations across faces of different subjects to be small, a generic shape-illumination manifold (gSIM) can be learnt from a training corpus.

Given a test video for recognition, it is first re-illuminated in the illumination condition of each gallery video. Re-illumination involves a genetic algorithm (GA) based pose matching across the two face videos. For re-illumination, each frame of the test video is recreated using a weighted linear combination of K nearest neighbor frames of the gallery video as discovered by the pose matching module. This is followed by generation of difference-of-log vectors between each corresponding frame of the original and re-illuminated test videos. If the gallery and test video belong to the same subject, the difference-of-log vectors depend only on shape and illumination conditions. On the other hand, if the

two videos come from different subjects, the vectors also depend on the differences in albedo maps of the two subjects. Finally, the similarity score is obtained by computing the likelihood of these postulated shape-illumination manifold samples under the learnt gSIM. The approach provides near perfect recognition rates on three different datasets consisting of 100, 60 and 11 subjects respectively.

7. *System Identification Approach:* Aggarwal et al. [12] pose VFR as a dynamical system identification problem. A moving face is modeled as a linear dynamical system whose appearance changes with pose. Each frame of the video is assumed to be the output of the dynamical system particular to the subject. Autoregressive and Moving Average (ARMA) model is used to represent such a system as follows

$$\begin{aligned} x(t+1) &= Ax(t) + v(t) \\ y(t) &= Cx(t) + \omega(t) \end{aligned} \quad (5)$$

Here $y(t)$ is the noisy observation of input $I(t)$ at time t , such that $y(t) = I(t) + \omega(t)$. $I(t)$ is the appearance of face at time t and $x(t)$ is the hidden state that characterizes the pose, expression, etc. of

the face at time t . A and C are the system matrices characterizing the system, and $v(t)$ is an IID realization from some unknown density $q(\cdot)$. Given a sequence of video frames, Aggarwal et al. [12] use a closed-form solution to estimate A and C . The similarity between a gallery and a probe video is measured using metrics based on subspace angles obtained from the estimated system matrices. The metrics used include Martin, gap, and Frobenius distance, all of which give similar recognition performance. The approach does well on the two datasets tested in [12]. Over 90% recognition rate is achieved (15/16 for the *Li* dataset [10] and 27/30 for the *UCSD/Honda* dataset [9]). The performance is quite promising given the extent of the pose and expression variations in the video sequences.

Summary and Discussion

There is little doubt that presence of multiple video frames allows for better generalization of person-specific facial characteristics over what can be achieved from a single image. In addition, VFR provides operational advantages over traditional still image based face recognition systems. Most existing VFR approaches have only been tested on independently captured very small datasets. Large standard datasets are required for better evaluation and comparison of various approaches.

Related Entries

- ▶ Face Recognition Overview
- ▶ Face Recognition Systems
- ▶ Face Tracking

References

1. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. *ACM Comput. Surv.* **35**(4), 399–458 (2003)
2. O'Toole, A.J., Roark, A., Abdi, H.: Recognizing moving faces: A psychological and neural synthesis. *Trends Cogn. Sci.* **6**, 261–266 (2002)
3. Hadid, A., Pietikainen, M.: An experimental investigation about the integration of facial dynamics in video-based face recognition. *Electronic Lett. Comput. Vis. Image Anal.* **5**(1), 1–13 (2005)

4. Ekenel, H., Pnevmatikakis, A.: Video-based face recognition evaluation in the chil project - run 1. In: Proceedings of the seventh International Conference on Automatic Face and Gesture Recognition, pp. 85–90 (2006)
5. Gorodnichy, D. O. (Editor): Face processing in video sequences. *Image and Vis. Comput.* **24**(6), 551–648 (2006)
6. Grother, P., Micheals, R., Phillips, P.: Face recognition vendor test 2002 performance metrics. In: Proceedings of fourth International Conference on Audio and Video-Based Biometric Person Authentication, pp. 937–945 (2003)
7. Zhou, S., Kruger, V., Chellappa, R.: Probabilistic recognition of human faces from video. *Comput. Vis. Image Underst.* **91**(1–2), 214–245 (2003)
8. Gross, R., Shi, J.: The cmu motion of body (mobo) database. *Tech. Rep. CMU-RI-TR-01-18*, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (2001)
9. Lee, K.C., Ho, J., Yang, M.H., Kriegman, D.: Visual tracking and recognition using probabilistic appearance manifolds. *Comput. Vis. Image Underst.* **99**(3), 303–331 (2005)
10. Li, B., Chellappa, R.: Face verification through tracking facial features. *J. Opt. Soc. Am. A* **18**(12), 2969–2981 (2001)
11. Liu, X., Chen, T.: Video-based face recognition using adaptive hidden markov models. In: Proceedings of International Conference on Computer Vision and Pattern Recognition, pp. 340–345 (2003)
12. Aggarwal, G., Roy-Chowdhury, A.K., Chellappa, R.: A system identification approach for video-based face recognition. In: Proceedings of International Conference on Pattern Recognition, pp. 175–178 (2004)
13. Liu, J.S.: Monte carlo strategies in scientific computing. Springer (2002)
14. Aggarwal, G., Veeraraghavan, A., Chellappa, R.: 3d facial pose tracking in uncalibrated videos. In: Proceedings of International Conference on Pattern Recognition and Machine Intelligence, pp. 515–520 (2005)
15. Arandjelovic, O., Cipolla, R.: Face recognition for video using the general shape-illumination manifold. In: Proceedings of European Conference on Computer Vision, pp. 27–40 (2006)

Face Reconstruction

- ▶ Forensic Evidence of Face

Face Registration

- ▶ Face Alignment

Face Sample Quality

KUI JIA¹, SHAOGANG GONG²

¹Shenzhen Institute of Advanced Integration Technology, CAS/CUHK, Shenzhen, People's Republic of China
²Queen Mary, University of London, London, UK

Synonyms

Face sample standardization; Face sample utility

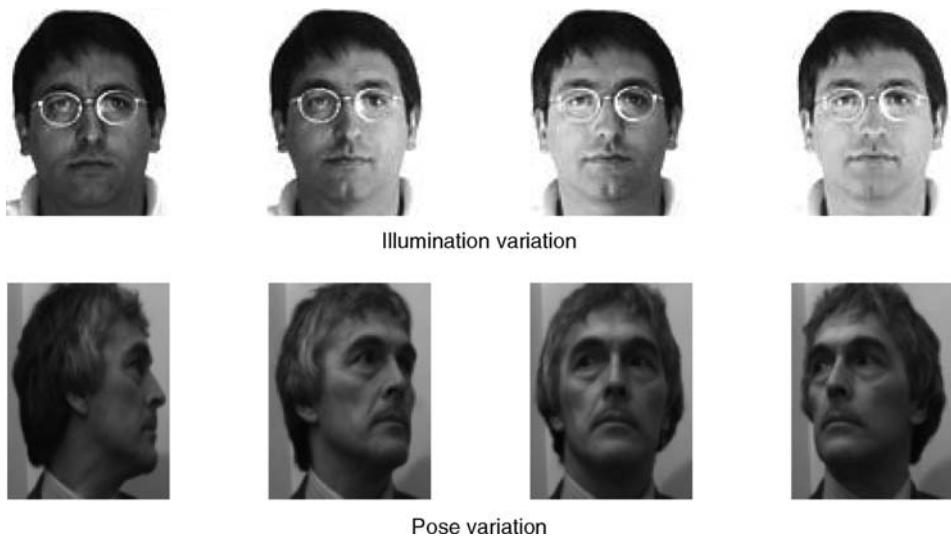
Definition

Face is a human biometric attribute that can be used to establish the identity of a person. A face-based biometric system operates by capturing probe face samples and comparing them against gallery face templates. The intrinsic characteristic of captured face samples determine their effectiveness for face authentication. Face sample quality is a measurement of these intrinsic characteristics. Face sample quality has significant impact on the performance of a face-based biometric system. Recognizing face samples of poor quality is a challenging problem. A number of factors can contribute toward degradation in face sample quality. They include, but not limited to, illumination variation, pose variation, facial expression change, face occlusion, low resolution, and high sensing noise.

Introduction

A typical face-based biometric system operates by capturing face data (images or videos), and comparing the obtained face data against face templates of different individuals in a gallery set. While face templates in the gallery set are normally captured under constrained imaging conditions (e.g., from frontal view, at a short distance from the camera, and under consistent illumination), it is unrealistic to assume controlled acquisition of probe face data. Face data captured under uncontrollable environment usually contains many kinds of defects caused by poor illumination, improper face positioning, and imperfect camera sensors [1]. For instance, when face data is captured in a natural outdoor environment, inconsistent illumination is typically cast on human faces resulting in uneven, extremely strong or weak lightings. Face rotation can also cause significant appearance variations, and at the extreme, face can be self occluded (Fig. 1). When distances between human faces and cameras increase, captured face data will be at low resolution, in low contrast, and likely to contain high imaging noise. In some instances people may wear sunglasses, have varying facial expression, and be with heavy makeup. All of these factors contribute toward potential degradation in the quality of captured face samples, resulting in disparities to those of face templates stored in the gallery set.

Face sample quality has significant impact on the performance of face-based biometric systems.



Face Sample Quality. **Figure 1.** Face samples of illumination and pose variations from AR and UMIST databases.

Assessing the quality of face samples before applying them in any biometric system may help improve the authentication accuracy. For example, an intruder may wear sunglasses intending to disguise himself, quality assessment of intruder's face samples can give an alert to such a situation. Quantitative measures on the quality of face samples can also be integrated into biometric systems to increase or decrease relevant thresholds. In a people enrollment stage, such quantitative measures of quality also help procure gallery face templates of good quality. Many approaches assess face sample quality using general image properties including contrast, sharpness, and illumination intensity [2]. However, these properties cannot properly measure face sample degradation caused by inconsistent illumination, face rotation, or large face-camera distance. There are a few recent works assessing face sample quality by considering such kinds of degradation. For example in [1], facial-symmetry-based methods are used to measure facial asymmetries caused by non-frontal lighting and improper facial pose.

When only poor quality face data can be acquired at the authentication stage, face recognition becomes significantly more challenging because of: (1) *Illumination variation* to which the performance of most existing face recognition algorithms and systems is highly sensitive. It has been shown both experimentally [3] and theoretically [4] that face image differences resulting from illumination variation are more significant than either inherent face differences between different individuals, or those from varying face poses [5]. State of the art approaches addressing this problem include heuristic methods, reflectance-model methods, and 3D-model-based methods [6]. Although performance improvement is achieved, none of these methods are truly illumination invariant. (2) *Pose variation* which causes face recognition accuracy to decrease significantly, especially when large pose variations between gallery and probe faces are present. The difficulties would further increase if only an unknown single pose is available for each probe face. In such a situation, an extra independent training set, different from the gallery set and containing multiple face images of different individuals under varying poses, will be helpful. Three-dimensional face model or statistical relational learning between different poses can be employed to generate virtual face poses. By generating virtual poses, one can either normalize probe faces of varying poses to a predefined pose, e.g., frontal, or

expand the gallery to cover large pose variations. (3) *Low resolution* face data will be acquired when face-camera distances increase, which is rather typical in surveillance imagery. The performance of existing face recognition systems decreases significantly when the resolution of captured face data is reduced below a certain level. This is because the missing high-resolution details in facial appearances and image features make facial analysis and recognition ineffective, either by human operators or by automated systems. It is therefore useful to generate high-resolution face images from low-resolution ones. This technique is known as face hallucination [7] or face ▶ **super-resolution**.

Assessment of Face Sample Quality

The performance of face authentication depends heavily on face sample quality. Thus the significance of face sample quality assessment and standardization grows as more practical face-based biometric systems are required. Quality assessment of probe face samples can either reject or accept a probe to improve later face verification or identification accuracy. Quantitative assessment of face sample quality can also be used to assign weights in a biometric fusion scheme.

ISO/IEC WD 29794-1 [8] considers that biometric sample quality can be defined by character (inherent features), fidelity (accuracy of features), or utility (predicted biometrics performance). Many efforts have been made on biometric sample quality assessment for fingerprint, iris, or face data. Most of those on face data are based on general image properties including contrast, sharpness, and illumination intensity [2]. However, the face sample degradation that severely affects face authentication accuracy is from uncontrollable imaging conditions that cause illumination variations, head pose changes, and/or very low-resolution facial appearances. There are a few attempts made on assessing face sample quality caused by these kinds of degradation.

In [9], two different strategies for face sample quality assessment are considered: one is for illumination variation and pose change, another is for facial expression change. In the first strategy, specific measures are defined to correlate with levels of different types of face sample degradation. A polynomial function is then utilized based on each measure for predicting the

performance of a ► Eigenface technique on a given face sample. Quality goodness is assessed by selecting a suitable threshold. Since the measurement of facial expression intensity is difficult, in the second strategy, a given face sample is classified into good or poor quality based on its coarse similarity to neutral facial expression. Then the training procedure for each class is achieved by dividing the training set into two subsets, based on whether the samples are recognizable by the Eigenface technique. Then these two subsets are described by Gaussian mixture models (GMMs). In [1], facial-symmetry-based quality scores are used to assess facial asymmetries caused by non-frontal lighting and improper facial pose. In particular, local binary pattern (LBP) histogram features are applied to measure the lighting and pose asymmetries. Moreover, the inter-eye distance is also used to estimate the quality score for whether a face is at a proper distance from the camera.

Recognizing Face Samples of Poor Quality

In general, face recognition under varying illumination is difficult. Although existing efforts to address this challenge have not led to a fully satisfactory solution for illumination invariant face recognition, some performance improvements have been achieved. They can be broadly categorized into: heuristic methods, reflectance-model methods, and 3D-model-based methods [6]. A typical heuristic method applies subspace learning, e.g., principal component analysis (PCA), using training face samples. By discarding a few most significant, e.g., the first three, principal components, variations due to lighting can be reduced. Reflectance-model methods employ a Lambertian reflectance model with a varying albedo field, under the assumption of no attached and cast shadows. The main disadvantage of this approach is the lack of generalization from known objects to unknown objects [10]. For 3D-face model-based approaches, more stringent assumptions are often made and it is also computationally less reliable. For example in [11], it is assumed that the 3D face geometry lies in a linear space spanned by the 3D geometry of training faces and it uses a constant albedo field. Moreover, 3D model-based methods require complex fitting algorithms and high-resolution face images.

There are also attempts to address the problem of face recognition across varying facial poses.

In real-world applications, one may have multiple face samples of varying poses in training and gallery sets (since they can be acquired offline), while each captured probe face can only be at an unknown single pose. Three-dimensional model-based methods [12] or statistical learning-based methods can be used to generate virtual face poses [13], by which either probe faces can be normalized to a predefined pose, e.g. frontal view, or gallery faces can be expanded to cover large pose variations. For example in [12], a 3D morphable model is used. The specific 3D face is recovered by simultaneously optimizing the shape, texture, and mapping parameters through an analysis-by-synthesis strategy. The disadvantage of 3D model-based methods is slow speed for real-world applications. Learning-based methods try to learn the relations between different facial poses and how to estimate a virtual pose in 2D domain, e.g., the view-based active appearance model (AAM) [14]. This method depends heavily on the accuracy of face alignment, which unfortunately introduces another open problem in practice.

When the resolution of captured face data falls below a certain level, existing face recognition systems will be significantly affected. Face super-resolution techniques have been proposed to address this challenge. Reconstruction-based approaches require multiple, accurately aligned low-resolution face samples to obtain a high-resolution face image. Their magnification factors of image resolution are however limited [7]. Alternatively, learning-based face super-resolution approaches model high-resolution training faces and learn face-specific prior knowledge from them. They use the learned model prior to constrain the super-resolution process. A super-resolution factor as high as 4×4 can be achieved [7]. The face super-resolution process can also be integrated with face recognition. For example in [15], face image super-resolution is transferred from pixel domain to a lower dimensional eigenface space. Then the obtained high-resolution face features can be directly used in face recognition. Simultaneous face super-resolution and recognition in ► tensor space have also been introduced [16]. Given one low-resolution face input of single modality, the proposed method can integrate and realize the tasks of face super-resolution and recognition across different facial modalities including varying facial expression, pose, or illumination. This has been further generalized to unify automatic alignment with super-resolution [17].

Summary

Many face-based biometric systems have been deployed in applications ranging from national border control to building door access, which normally solve the sample quality problem at the initial face acquisition stage. Given ongoing progress on standardization of face sample quality and technical advancement in authenticating face samples of poor quality, the availability of more reliable and convenient face authentication systems is only a matter of time.

Related Entries

- Biometric Sample Quality
- Face Pose Analysis
- Face Recognition

References

1. Gao, X.F., Li, S.Z., Liu, R., Zhang, P.R.: Standardization of face image sample quality. In: Proceedings of Second International Conference on Biometrics (ICB), pp. 242–251. Seoul, Korea (2007)
2. Brauckmann, M., Werner, M.: Technical report. In: Proceedings of NIST Biometric Quality Workshop. (2006)
3. Adini, Y., Moses, Y., Ullman, S.: Face recognition: the problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 721–732 (1997)
4. Zhao, W., Chellappa, R.: Robust Face Recognition Using Symmetric Shape-from-Shading. Technical Report, Center for Automation Research, University of Maryland (1999)
5. Tarr, M.J., Bulthoff, H.H.: Image-based object recognition in man, monkey and machine. *Cognition* **67**, 1–20 (1998)
6. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. *ACM Comput. Surv.* **35**(4), 399–458 (2003)
7. Baker, S., Kanade, T.: Limits on super-resolution and how to break them. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(9), 1167–1183 (2002)
8. ISO/IEC JTC 1/SC 37 N 1477: Biometric Sample Quality Standard – Part 1: Framework (2006)
9. Abdel-Mottaleb, M., Mahoor, M.H.: Application notes – algorithms for assessing the quality of facial images. *IEEE Comput. Intell. Mag.* **2**(2), 10–17 (2007)
10. Baker, S., Kanade, T.: Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(2), 230–245 (2007)
11. Atick, J., Griffin, P., Redlich, A.: Statistical approach to shape from shading: reconstruction of 3-dimensional face surfaces from single 2-dimentional images. *Neural Comput.* **8**(2), 1321–1340 (1996)
12. Blanz, V., Vetter, T.: Face recognition based on fitting a 3-D morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(9), 1063–1074 (2003)
13. Li, Y., Gong, S., Liddell, H.: Constructing facial identity surfaces for recognition. *Int. J. Comput. Vision* **53**(1), 71–92 (2003)
14. Cootes, T.F., Walker, K., Taylor, C.J.: View-based active appearance models. In: Proceedings of Fourth International Conference on Automatic Face and Gesture Recognition (FG), pp. 227–232. Grenoble, France (2000)
15. Gunturk, B.K., Batur, A.U., Altunbasak, Y., Hayes, M.H., Mersereau, R.M.: Eigenface-Domain Super-Resolution for Face Recognition. *IEEE Transactions on Image Processing*, **12**(5), 597–606 (2003)
16. Jia, K., Gong, S.: Multi-modal tensor face for simultaneous super-resolution and recognition. In: Proceedings of Tenth International Conference on Computer Vision (ICCV), pp. 1683–1690. Beijing, China **2** (2005)
17. Jia, K., Gong, S.: Generalised face super-resolution. *IEEE Transactions on Image Processing*, **17**(6), 873–886 (2008).

Face Sample Standardization

- Face Sample Quality

Face Sample Synthesis

SAMI ROMDHANI, JASENKO ZIVANOV

Computer Science Department, University of Basel,
Basel, Switzerland

Synonyms

Face image synthesis; Rendering; Image formation process

Definition

Face Sample Synthesis denotes the process of generating the image of a human face by a computer program. The input of this process is a set of parameters that

describes (1) the position from which the face is viewed, (2) the illumination environment around the face, (3) the identity of the person, and (4) the expression of the person. Other parameters may also be used such as the age of the person, parameters describing the makeup, etc. The output is an image of a human face.

Introduction

Face Sample Synthesis denotes the process of generating the image of a human face by a computer program. Optimally, this image should be realistic and virtually indistinguishable from a photography of a live scene. Additionally, the computer program should be generic: able to synthesize the face of any individual, viewed from any pose and illuminated by any arbitrarily complex environment. The objective of this article is to review the techniques used to reach this goal. It may also be desirable to generate faces with different expressions, different attributes such as makeup style or facial hair. One might also want to render image sequences with realistic facial motion. However, it is outside the scope of this article to address the methods enabling such synthesizes.

A photograph of a face is a projection onto an image plane of a 3D object. The intensity of a pixel of this photograph directly depends on the amount of light that is reflected from the object point imaged at the pixel location. Thus, this article first reviews 3D to 2D projections (the finite projective camera model) and illumination modeling (the Lambertian and Phong light reflection models usually used in face recognition systems). Then, the basics of identity modeling are summarized. At the end of the article, the reader will have an overview of the process required to synthesize a face image from any individual, viewed from any angle, and illuminated from any direction.

Research on computer-based face recognition dates back from the 1970s. In those times, most popular methods (e.g., [1]) were based on distances and angles between landmark points (such as eyes and mouth corners, nostril, chin top, etc.). Then, in the beginning of the 1990s, the appearance-based methods came in and quickly attracted most of the attention [2]. Contrasting with the former landmark points methods, these techniques use the entire face area for recognition. They are based on a prior generative model capable of synthesizing a face image given a small number

of parameters. Analysis is performed by estimating the parameters, denoted by $\hat{\theta}$, which synthesize a face image that is as similar as possible to the input image. Hence, these methods are called *Analysis by Synthesis*. This is usually done using a sum of square error functions:

$$\hat{\theta} = \arg \min_{\theta} \sum_i \|I_{i,\text{input}} - I_{i,\text{model}}(\theta)\|^2, \quad (1)$$

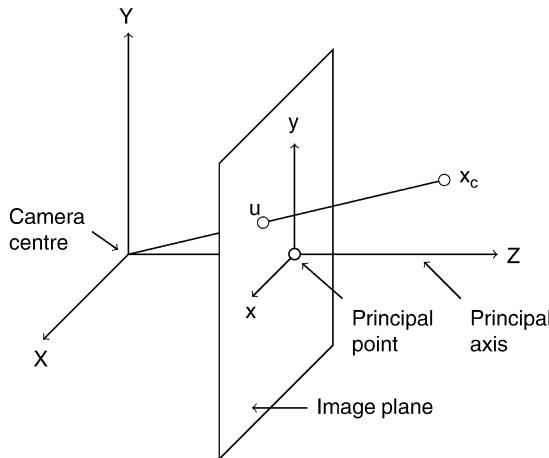
where the index i represent pixel i and the sum runs over all pixels of the face area. The formation of the model image, $I_{\text{model}}(\theta)$, is the topic of this article. Initially, the models used a 2D representation of the face structure [2], however, in order to account for pose and illumination variation, it is accepted that 3D models provide more accurate results [3, 4]. Hence, this article reviews the process of generating a face image from a 3D model.

Four ingredients are necessary to synthesize a face from a 3D representation [5]: The face surface of the individual to be imaged must be sampled across a series of points resulting in a list of 3D vertices. Obtaining a surface from a list of vertices is achieved by a triangle list that connects triplets of vertices. The triangle list defines the topology of the face. It is used, among other things, to compute surface normals and the visibility of a surface points using (for instance) a “Z-buffer” visibility test.

The third constituent is the color of the face. It can be represented by an RGB color for a dense set of surface points. These surface points are called “texels.” If the texels are the same points as the vertices, then the color model is called “per vertex color.” Alternatively, a much denser texel sampling can be used and the texels are arranged in a “texture map.” In order to synthesize unconstrained illumination images, the texels must be free of any illumination effect and code the “albedo” of a point. The albedo is defined as the diffuse color reflected by a surface point. Finally, the last ingredient is a reflection model that relates the camera direction and the intensity of light reflected by a surface point, to the intensity, the direction, and the wavelength of light reaching the point.

Finite Projective Camera Model

This section briefly describes how a 3D object is imaged on a 2D image.



In Computer Vision and Computer Graphics, a finite projective camera model is usually chosen. This camera follows a central projection of points in space onto a plane. For now, let's assume that the camera is at the origin of an Euclidean coordinate system and that it is pointing down the Z -axis. In that system, a 3D point, $\mathbf{x}_c = (X_c, Y_c, Z_c)^t$ is projected onto a 2D point, \mathbf{u} , in the image frame according to the following equation, in which the focal length, denoted by f , is the distance between the camera center and the principal point: $\mathbf{u} = (fX_c/Z_c, fY_c/Z_c)^t$. This equation assumes that the origin of the image plane coordinate system is at the principal point. In general, it might not be and, denoting the coordinates of the principal point by (p_x, p_y) , the mapping becomes:

$$\mathbf{u} = (fX_c/Z_c + p_x, fY_c/Z_c + p_y). \quad (2)$$

In a face image synthesis, the point \mathbf{x}_c is one vertex of the 3D shape of a face in *camera coordinate frame*. It is easier to represent the ensemble of vertices of the face in an *object coordinate frame*. The origin of this frame is attached to the object, a typical choice is to locate it at the center of mass of the face. The 3D coordinate of the camera center in the object frame is denoted by \mathbf{c} .

Additionally, the object coordinate frame is generally not aligned with the camera coordinate frame, i.e., the face is not always frontal. The rotation between the face and the camera is denoted by the 3×3 matrix \mathbf{R} . It can be represented by a product of rotations along the coordinate axes of the object frame:

$$\begin{aligned}\mathbf{R}_\alpha &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \sin(\alpha) \\ 0 & -\sin(\alpha) & \cos(\alpha) \end{pmatrix}, \\ \mathbf{R}_\beta &= \begin{pmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{pmatrix}, \\ \mathbf{R}_\gamma &= \begin{pmatrix} \cos(\gamma) & \sin(\gamma) & 0 \\ -\sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{R} = \mathbf{R}_\alpha \mathbf{R}_\beta \mathbf{R}_\gamma.\end{aligned}\quad (3)$$

The relation between the object and camera frames is then: $\mathbf{x}_c = \mathbf{R}(\mathbf{x} - \mathbf{c})$. It is often convenient not to make the camera center explicit and to introduce $\mathbf{t} = -\mathbf{R}\mathbf{c}$. In this case, the relation is simply:

$$\mathbf{x}_c = \mathbf{R}\mathbf{x} + \mathbf{t}. \quad (4)$$

As a result, projecting a point \mathbf{x} in object coordinate frame onto the image plane is summarized by the following expression, in which \mathbf{R}_i denotes the row number i of the matrix \mathbf{R} .

$$\begin{cases} u_x = f \frac{\mathbf{R}_1 \mathbf{x} + \mathbf{t}_x}{\mathbf{R}_3 \mathbf{x} + \mathbf{t}_z} + p_x \\ u_y = f \frac{\mathbf{R}_2 \mathbf{x} + \mathbf{t}_y}{\mathbf{R}_3 \mathbf{x} + \mathbf{t}_z} + p_y \end{cases} \quad (5)$$

Estimating the parameters of a finite projective camera model requires then the estimation of nine parameters: $f, \alpha, \beta, \gamma, \mathbf{t}_x, \mathbf{t}_y, \mathbf{t}_z, p_x, p_y$. Note that in this explanation some subtle parameters that have only a minor effect on the synthesis and on the analysis by synthesis results are neglected: Some CCD cameras do not have square pixels (two additional parameters) and the skew parameter that is zero for most normal cameras [6].

Lighting Model

The previous section showed where, in the image, to draw a surface point from its 3D coordinates. Now the question is: What pixel value to draw on this point? The pixel value is the intensity of the light reflected by the surface point, which is computed using a lighting model. Much of the realism of a rendering depends on the ► [lighting model](#).

This model, in turn, depends on three factors: The number and type of light sources, the reflectance function, and the method used to compute surface normals. Light modeling is still undergoing considerable research efforts in the computer graphics community (the main challenge being to make photo-realistic rendering algorithms computationally efficient). In this article, the fundamental notions are only briefly introduced.

If light is emitted from direction \bar{l} with intensity l , then the quantity of light received by an infinitesimally small surface patch around surface point x is $\langle \bar{n}_x, \bar{l} \rangle l$, where \bar{n}_x is the normal of the surface patch at the point x and $\langle \cdot, \cdot \rangle$ is the scalar product (if it is positive and null otherwise). If the surface point projects onto pixel i of the image, yields

$$I_{i,\text{model}} = r_x(\bar{v}, \bar{l}) \cdot \langle \bar{n}_x, \bar{l} \rangle \cdot l \cdot S_{x,\bar{l}}, \quad (6)$$

where $r_x(\cdot)$ denotes the reflectance function at point x and \bar{v} , the viewing direction (defined as the direction from the point to the camera center). $S_{x,\bar{l}}$ denotes the cast shadow binary variable: If there is another object or if some part of the face is between point x and the point at infinity in direction \bar{l} , then the light is shadowed at the point, and $S_{x,\bar{l}}$ is zero, otherwise it is equal to one. Cast shadows are usually computed by a shadow map [7].

Light Source

The simplest and most computationally efficient is to use one directed light source at infinity and one ambient light source. The light reflected by a surface point from an ambient light source does not depend on the local surface around the point, it only depends on the albedo of the point. In real world, however, a perfectly ambient light never exists and it rarely happens that a point is illuminated only by a single light source. Indeed, light emanating from a light source might bounce off a wall, for instance, and then reach the object point. Hence, in real world, light comes from all directions. An environment map [8] is usually used to model this effect. It codes the intensity of light reaching an object for a dense sampling of directions. It is acquired by photographing a mirrored sphere [9]. Due to the additive nature of light, rendering with several light sources (as is the case for environment maps) is performed by summing (or integrating) over the light sources:



$$I_{i,\text{model}} = \sum_j r_x(\bar{v}, \bar{l}_j) \cdot \langle \bar{n}_x, \bar{l}_j \rangle \cdot l_j \cdot S_{x,\bar{l}_j}. \quad (7)$$

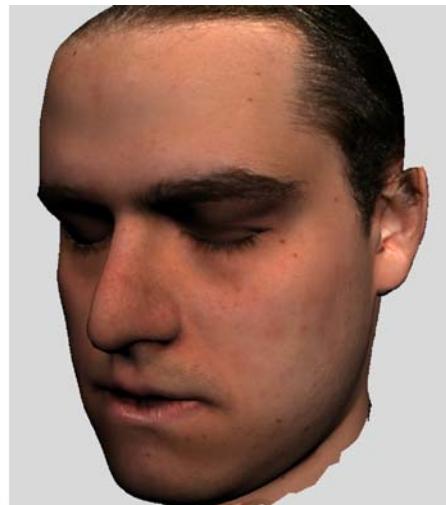
The following image is an environment map acquired at the Uffizi Gallery in Florence, Italy. Each pixel of this photograph is attached to a direction and represent a light source. This environment map is used to illuminate Panel d of Fig. 1.

Reflectance Function

In (6), the four-dimensional reflectance function $r_x(\bar{v}, \bar{l})$ is called the Bidirectional Reflectance Distribution Function (BRDF). It also depends on the wavelength of the incoming light (usually represented by its RGB color). The BRDF describes the properties of the material at point x .

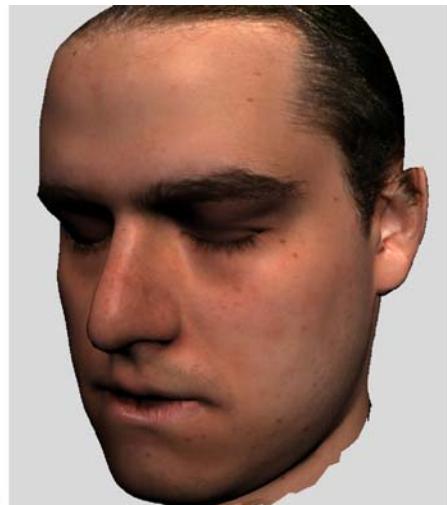
The simplest model of reflectance function is certainly the Lambertian model for which the function is equal to a constant (the albedo at point x). This means that incident light is scattered equally in all directions, which only happens for perfectly diffuse objects (totally matte, without shininess). For human face, this is the case only when the skin is covered by a very fine layer of powder. An example of rendering with a Lambertian reflectance is displayed on Fig. 1a.

Specular reflection takes place when light is reflected at a point without absorption by the material. For perfectly specular material, such as mirrors, light is reflected in only one direction (the reflectance function is a Dirac function): when the viewing angle is equal to the angle



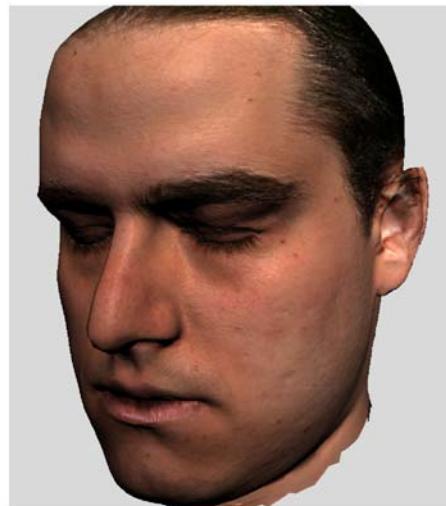
a

1 light, Lambert, Normal from shape,
comp. time: 50 ms.



b

1 light, Phong ($K = 0.067$, $v = 20$), Normal from shape,
comp. time: 50 ms.



c

1 light, Phong ($K = 0.067$, $v = 20$),
normal map, comp. time: 50ms.

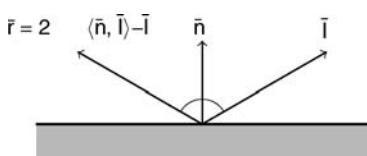


d

100,000 lights (Environment map: The Uffizi Gallery, Florence),
Phong ($K = 0.3$, $v = 101$), normal map, comp. time: 90 min.

Face Sample Synthesis. [Figure 1](#) Renderings using different illumination models. The 3D model includes 24,367 vertices and 48,660 triangles.

of incidence . Generally, in Computer Graphics, the reflectance used is a combination of diffuse and specular reflectance. The most well-known model is the Phong model:



$$I_{i,\text{model}}^{\text{phong}} = (\mathbf{c} \cdot \langle \bar{n}_x, \bar{l} \rangle + K \cdot \langle \bar{v}, \bar{r} \rangle^{\bar{v}}) \cdot l \cdot S_{x,\bar{l}}$$

where \mathbf{c} is the albedo at point x , \bar{r} is the reflection direction (depending on the normal and the lighting direction, as shown on the sketch), K is the fraction of energy specularly reflected, and v is an index that controls the “tightness” of the specular highlight (note that there is one such equation for each color channel). In this equation, the first summand inside the brackets is the diffuse (i.e. Lambertian) part and the second

one is the specular part. Modeling the specular highlight is important for human skin, as it often has a thin layer of oil or sweat above the pigmented cells. Figure 1a–d show examples of face rendering with a Phong model.

The Pong model assumes smooth surfaces, but, in reality, surfaces are imperfect and exhibit microgeometry. There exist more complex BRDF models that represent the surface as composed of micro-facets that can shadow and mask each other. Another effect is the off-specular highlight: When the angle of incidence is grazing (near 90°), some materials (such as human skin) reflect much more light than is absorbed, causing the color of the point to approach that of the light. This is accounted for by the Fresnel term that is the ratio between reflected and absorbed light. Some of the more complex BRDF model that accounts for these two effects are Blinn [10], Cook-Torrance [11], Torrance-Sparrow [12], and more recently Lafourture [13].

Note that this is not the end of the story, yet. The BRDF assumes that the outgoing light at one point results only from the incoming light at the same point. This is in fact an approximation as it neglects the scattering of light within the material. This phenomenon is modeled by the bidirectional surface scattering reflectance distribution function (BSSRDF) [14] of which the BRDF is a special case. Human skin does show some important subsurface scattering effects and to reach photo-realism these effects should not be neglected. This is for instance apparent when the ear is illuminated from the back. It then looks translucent which results from the subsurface scattering.

Normals

Human skin is not a smooth surface. Pores and wrinkles induce very small scale variations of the surface. Representing these variations with 3D vertices would require a very fine sampling of the head resulting in an overly large number of vertices making the rotation or visibility test computationally inefficient. The concept of “normal mapping” was developed precisely for this reason. Instead of computing a normal from the shape (by interpolating the normals of the triangles corners in which a pixel is projected from), the normals are computed from a dense normal map (for which interpolation is carried similarly to the texture map). A normal map is generally acquired by photometric stereo [15]: Several

photographs of the face of a subject are taken with different light directions. The subject and the camera must be perfectly still during this acquisition process (a pixel must be the projection of exactly the same point on the subject face for all photographs). Each photograph yields one measurement of the BRDF of a surface point. Using several measurements a BRDF model can be fitted, thereby recovering the normal of the point. Often, for its simplicity, a Lambertian model is used, in which case, the operator must choose the light direction such as to minimize specular reflections.

Figure 1 shows different types of rendering from the most simple (left) to more complex and realistic (right).

Identity Modeling

So far, an overview of the face image synthesis process from a 3D model of the face of an individual is presented. This 3D model can be acquired by a 3D scanner or can be manually crafted with a modeling software. These processes can be tedious and expensive. Therefore, it is desirable to be able to *generate* the 3D shape and texture (i.e., albedo) from *any individual*. This can be done by defining a vector space of shapes and of textures and probability distributions in these spaces. This is accomplished by learning typical face variations from an example set of 3D faces. The vector space is defined by densely registering the examples with a reference face, thereby defining a label for each vertex. Once the vector spaces are defined, linear combination of the example shapes and textures are made to generate the shape and texture of new (i.e., out of the example set) individuals. The coefficients of these linear combinations are the parameters of the identity model. One individual is coded by a specific value for each parameter. However, some variations are more typical than others and probability distributions in the vector space must be used in order to ensure the plausibility of a novel individual. If the probability distributions of the human faces in the vector spaces are assumed Gaussian, then the most efficient coding is yielded by a Principal Component Analysis [16] of the examples. These principles are used by the *3D Morphable Model* [3], the state of the art identity generic human face model.

Denoting by \mathbf{X} the $3 \times N$ matrix with the 3D position of N vertices (hence the position of a vertex

x in the “Finite Projective Camera Model” section is a column of the matrix **X**) and by **C** the $3 \times N$ matrix with the RGB albedo of N vertices, a novel 3D face is yielded by the following equations:

$$\mathbf{X} = \sum_i^M \alpha_i \mathbf{X}_i, \quad \mathbf{C} = \sum_i^M \beta_i \mathbf{C}_i, \quad (8)$$

where \mathbf{X}_i and \mathbf{C}_i are the M shape and texture principal components and α_i and β_i the shape and texture parameters.

Conclusion

In conclusion, the set of parameters θ of an analysis by synthesis method (1) is composed of nine parameters for the projection. For illumination, using a Phong reflectance model with one light source and with normals computed from the shape, seven parameters must be estimated: three parameters for the intensity of the colored light, two parameters for its direction along with the specular coefficient K and the Phong exponent v . Additionally, $2M$ parameters must be recovered for the 3D shape and the texture.

Using a normal map model (generic for all individuals) and an environment map for analysis by synthesis has never been attempted. Indeed, it would result in a tremendously complicated problem for the following reasons: It is unclear how to model normal maps that would generalize for any individual. A simple linear combination as is used for the shape and texture cannot be used for normals because a normal vector is a unit length vector and the sum or the mean of two unit length vectors does not result in a unit length vector. Moreover, defining correspondences for pores and wrinkles (which would be required to make a vector space and avoid blur results) is for the moment unsolved. As for the light sources, estimating the direction and intensity of a single light source from a single facial image is already an ill-posed problem (there is not enough information in one image to completely constraint the solution), let alone with a large number of light sources as is the case when using an environment map.

Summary

The motivation of face sample synthesis is not only to generate face images from a small number of

parameters but also to analyze them using an analysis by synthesis approach. In this article, the two main sources of face image variations (pose and illumination) are accounted for by a finite projective camera model. Illumination modeling is more complicated and requires the operator to choose the type of illumination sources, the type of reflectance function, and the manner to generate normals (either from the shape or acquired by a photometric stereo method). Finally, identity variations can be obtained by a linear combination of examples.

Related Entries

- ▶ Deformable Models
- ▶ Face Pose Analysis

References

1. Kanade, T.: Computer Recognition of Human Faces. Birkhäuser Verlag, Stuttgart, Germany (1973)
2. Cootes, T., Edwards, G., Taylor, C.: Active appearance model. Fifth European Conference on Computer Vision. Freiburg, Germany (1998)
3. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D-faces. SIGGRAPH 99. Los Angeles, California, USA (1999)
4. Romdhani, S., Vetter, T.: Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. CVPR. San Diego, CA, USA (2005)
5. Parke, F.I., Waters, K.: Computer Facial Animation. AKPeters Wellesley, Massachusetts, USA (1996)
6. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2000)
7. Woo, A., Poulin, P., Fournier, A.: A survey of shadow algorithms. IEEE Comput. Graph. Appl. **10**(6), 13–32 (1990)
8. Greene, N.: Environment mapping and other applications of world projections. IEEE Comput. Graph. Appl. **6**(11) (1986)
- 9.Debevec, P., Malik, J.: Recovering high dynamic range radiance maps from photographs. Siggraph. Los Angeles, California, USA (1997)
10. Blinn, J.F.: Models of light reflection for computer synthesized pictures. SIGGRAPH '77: Proceedings of the Fourth Annual Conference on Computer Graphics and Interactive Techniques, pp. 192–198. New York, NY, USA, ACM (1977)
11. Cook, R.L., Torrance, K.E.: A reflectance model for computer graphics. ACM Trans. Graph. **1**(1), 7–24 (1982)
12. Torrance, K.E., Sparrow, E.M.: Theory for off-specular reflection from roughened surfaces, pp. 32–41. Radiometry (1992)
13. Lafontaine, E.P.F., Foo, S.C., Torrance, K.E., Greenberg, D.P.: Non-linear approximation of reflectance functions. SIGGRAPH '97: Proceedings of the 24th Annual Conference on Computer

- Graphics and Interactive Techniques, pp. 117–126. New York, NY, USA, ACM /Addison-Wesley (1997)
14. Jensen, H.W., Marschner, S.R., Levoy, M., Hanrahan, P.: A practical model for subsurface light transport. SIGGRAPH, ACM/ Addison-Wesley (2001)
 15. Barsky, S., Petrou, M.: Colour photometric stereo: simultaneous reconstruction of local gradient and colour of rough textured surfaces. ICCV p. 600 (2001)
 16. Jolliffe, I.T.: Principal Component Analysis. Springer, Berlin (2002)

Face Sample Utility

- Face Sample Quality

Face Sketching

A face sketching is a parsimonious yet expressive representation of face. It depicts concise sketches of face that captures the most essential perceptual information with a number of strokes.

- And-Or Graph Model for Faces

Face Tracking

AMIT K. ROY-CHOWDHURY, YILEI XU

Department of Electrical Engineering, University of California, Riverside, CA, USA

Synonym

Facial motion estimation

Definition

In many face recognition systems, the input is a video sequence consisting of one or more faces. It is necessary

to track each face over this video sequence so as to extract the information that will be processed by the recognition system. Tracking is also necessary for 3D model-based recognition systems, where the 3D model is estimated from the input video. Face tracking can be divided along different lines depending upon the method used, e.g., head tracking, feature tracking, image-based tracking, model-based tracking. The output of the face tracker can be the 2D position of the face in each image of the video (2D tracking), the 3D pose of the face (3D tracking), or the location of features on the face. Some trackers are also able to output other parameters related to lighting or expression. The major challenges encountered by face tracking systems are robustness to pose changes, lighting variations, and facial deformations due to changes of expression, occlusions of the face to be tracked and clutter in the scene that makes it difficult to distinguish the face from the other objects.

Introduction

Tracking, which is essentially ► motion estimation, is an integral part of most face processing systems. If the input to a face recognition system is a video sequence, as obtained from a surveillance camera, tracking is needed to obtain correspondence between the observed faces in the different frames and to align the faces. It is so integral to video-based face recognition systems that some existing methods integrate tracking and recognition [1]. It is also a necessary step for building 3D face models. In fact, tracking and 3D modeling are often treated as two parts of one single problem [2–4].

There are different ways to classify face tracking algorithms [5]. One such classification is based on whether the entire face is tracked as a single entity (sometimes referred to as head tracking) or whether individual facial features are tracked. Sometimes a combination of both is used. Another method of classification is based on whether the tracking is in the 2D image space or in 3D pose space. For the former, the output (overall head location or facial feature location) is a region in the 2D image and does not contain information about the change in the 3D orientation of the head. Such methods are usually not very robust to changes of pose, but are easier to handle computationally. Alternatively, 3D tracking methods, which work by fitting a 3D model to each image of the video, can provide estimates of the 3D pose of the face. However, they are

usually more computationally intensive. Besides, many advanced face tracking methods are able to handle challenging situations like facial ▶ **deformations**, changes of lighting, and partial occlusions.

A broad overview of the basic mathematical framework of face tracking methods will be given first, followed by a review of the current state-of-the-art and technical challenges. Next, a few application scenarios will be considered, like surveillance, face recognition, and face modeling, including discussion of the importance of face tracking in each of them. Then some examples of face tracking in challenging situations will be shown, before conclusion.

Basic Mathematical Framework

An overview of the basic mathematical framework that explains the process in which most trackers work is provided here. Let $\mathbf{p} \in \Re^P$ denote a parameter vector, which is the desired output of the tracker. It could a 2D location of the face in the image, the 3D pose of the face, or a more complex set of quantities that also include lighting and deformation parameters. Define a synthesis function $f: \Re^2 \times \Re^P \rightarrow \Re^2$ that can take an image pixel $\mathbf{v} \in \Re^2$ at time $(t-1)$ and transform it to $f(\mathbf{v}, \mathbf{p})$ at time t . For a 2D tracker, this function f could be a transformation between two images at two consecutive time instants. For a 3D model-based tracker, this can be considered as a rendering function of the object at pose \mathbf{p} in the camera frame to the pixel coordinates \mathbf{v} in the image plane. Given an input image $I(\mathbf{v})$, align the synthesized image with it so as to obtain

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} g(f(\mathbf{v}, \mathbf{p}) - I(\mathbf{v})), \quad (1)$$

where $\hat{\mathbf{p}}$ denotes the estimated parameter vector for this input image $I(\mathbf{v})$.

The essence of this approach is the well-known Lucas–Kanade tracking, an efficient and accurate implementation of which has been proposed using the inverse compositional approach [6]. Depending on the choice of \mathbf{v} and \mathbf{p} , the method is applicable to the overall face image, a collection of discrete features, or a 3D face model. The ▶ **cost function** g is often implemented as an L_2 norm, i.e., the sum of the squares of the errors over the entire region of interest. However, other distance metrics may be used. Thus a face tracker is often implemented as a least-squares ▶ **optimization** problem.

Let us consider the problem of estimating the change, $\Delta \mathbf{p}_t \triangleq \mathbf{m}_b$, in the parameter vector between two consecutive frames, $I_t(\mathbf{v})$ and $I_{t-1}(\mathbf{v})$ as

$$\hat{\mathbf{m}}_t = \arg \min_{\mathbf{m}} \sum_{\mathbf{v}} (f(\mathbf{v}, \hat{\mathbf{p}}_{t-1} + \mathbf{m}) - I_t(\mathbf{v}))^2, \quad (2)$$

and

$$\hat{\mathbf{p}}_t = \hat{\mathbf{p}}_{t-1} + \hat{\mathbf{m}}_t. \quad (3)$$

The optimization of the above equation can be achieved by assuming a current estimate of \mathbf{m} as known and iteratively solve for increments $\Delta \mathbf{m}$ such that

$$\sum_{\mathbf{v}} (f(\mathbf{v}, \hat{\mathbf{p}}_{t-1} + \mathbf{m} + \Delta \mathbf{m}) - I_t(\mathbf{v}))^2 \quad (4)$$

is minimized.

Performance Analysis

While the basic idea of the face tracking algorithms is simple, the challenge comes in being able to perform the optimization efficiently and accurately. The function, f , will be nonlinear, in general. This is because f will include camera projection, the 3D pose of the object, the effect of lighting, the surface reflectance, nonrigid deformations, and other factors. For example, in [7], the authors derived a bilinear form for this function under the assumption of small motion. It could be significantly more complex in general. This complexity makes it difficult to obtain a global optimum for the optimization function, unless a good starting point is available. This initialization is often obtained through a face detection module working on the first frame of the video sequence. For 3D model-based tracking algorithms, it also requires registration of the 3D model to the detected face in the first frame.

The need for a good initialization for stable face tracking is only one of the problems. All trackers suffer from the problem of drift of the estimates and face tracking is no exception. Besides, the synthesis function f may be difficult to define precisely in many instances. Examples include partial occlusion of the face, deformations due to expression changes, and variations of lighting including cast shadows. Special care needs to be taken to handle these situations, since direct optimization of the cost function (2) would give an incorrect result.

Computational speed is another important issue in the design of tracking algorithms. Local optimization methods like gradient descent, Gauss–Newton, and Levenberg–Marquardt [8] can give a good result if the starting point is close to the desired solution. However, the process is often slow because it requires recomputation of derivatives at each iteration. Recently, an efficient and accurate method of performing the optimization has been proposed by using an inverse compositional approach, which does not require recomputation of the gradients at each step [6]. In this approach, the transformation between two frames is represented by a ► Face Warping function, which is updated by first inverting the incremental warp and then composing it with the current estimate. Our independent experimental evaluation has shown that on real-life facial video sequences, the inverse compositional approach leads to a speed-up by at least one order of magnitude, and often more, leading to almost real-time performance in most practical situations.

Challenges in Face Tracking

As mentioned earlier, the main challenges that face tracking methods have to overcome are (1) variations of pose and lighting, (2) facial deformations, (3) occlusion and clutter, and (4) facial resolution. These are the areas where future research in face tracking should concentrate. Some of the methods proposed to address these problems will be reviewed briefly below.

1. *Robustness to pose and illumination variations.* Pose and ► illumination variations often lead to loss of track. One of the well-known methods for dealing with illumination variations was presented in [9], where the authors proposed using a parameterized function to describe the movement of the image points, taking into account illumination variation

by modifying the brightness constancy constraint of optical flow. Illumination invariant 3D tracking was considered within the active appearance model (AAM) framework in [10], but the method requires training images to build the model and the result depends on the quality and variety of such data. 3D model based motion estimation algorithms are usually robust to pose variations, but often lack robustness to illumination. In [7], the authors proposed a model-based face tracking method that was robust to both pose and lighting changes. This was achieved through an analytically derived model for describing the appearance of a face in terms of its pose, the incident lighting, shape, and surface reflectance. Figure 1 shows an example.

2. *Tracking through facial deformations.* Tracking faces through changes of expressions, i.e., through facial deformations, is another challenging problem. An example of face tracking through changes of expression and pose is shown in Fig. 2. A survey of work on facial expression analysis can be found in [12]. The problem is closely related to modeling of facial expressions, which has applications beyond tracking, notably in computer animation. A well-known work in this area is [13], which has been used by many researchers for tracking, recognition, and reconstruction. In contrast to this model-based approach, the authors in [14] proposed a data-driven approach for tracking and recognition of non-rigid facial motion. More recently, the 3D morphable model [15] has been quite popular in synthesizing different facial expressions, which implies that it can also be used for tracking by posing the problem as estimation of the synthesis parameters (coefficients of a set of basis functions representing the morphable model).
3. *Occlusion and clutter.* As with most tracking problems, occlusion and clutter affect the performance



Face Tracking. Figure 1 Tracked points on a face through changes of pose and illumination. These points are projections of a 3D face mesh model.



Face Tracking. [Figure 2](#) An example of face tracking under changes of pose and expressions. The estimated pose is shown on the top of the frames. The pose is represented as an unit vector for the rotation axis, and the rotation angle in degrees, where the reference is taken to be the frontal face.



Face Tracking. [Figure 3](#) Tracked points on a face through changes of scale and illumination.

of most face trackers. One of the robust tracking approaches in this scenario is the use of particle filters [16], which can recover from a loss of track given a high enough number of particles and observations. However, in practice, occlusion and clutter remain serious impediments in the design of highly robust face tracking systems.

4. *Facial resolution.* Low resolution will hamper performance of any tracking algorithm, with face tracking being no exception. In fact, [5] identified low resolution to be one of the main impediments in video-based face recognition. [Figure 3](#) shows an

example of tracking through scale changes and illumination. Super-resolution approaches can be used to overcome these problems to some extent. However, super-resolution of faces is a challenging problem by itself because of detailed facial features that need to be modeled accurately. Recently, [17] proposed a method for face super-resolution using AAMs. Super-resolution requires registration of multiple images, followed by interpolation. Usually, these two stages are treated separately, i.e., registration is obtained through a tracking procedure followed by super-resolution. In a recent paper

[18], the authors proposed feeding back the super-resolved texture in the n th frame for tracking the $(n+1)$ th frame. This improves the tracking, which, in turn, improves the super-resolution output. This could be an interesting area of future work taking into consideration issues of stability and convergence.

Some Applications of Face Tracking

Some applications where face tracking is an important tool have been highlighted below:

1. *Video surveillance.* Since faces are often the most easily recognizable signature of identity and intent from a distance, video surveillance systems often focus on the face [5]. This requires tracking the face over multiple frames.
2. *Biometrics.* Video-based face recognition systems require alignment of the faces before they can be compared. This alignment compensates for changes of pose. Face tracking, especially 3D pose estimation, is therefore an important component of such applications. Also, integration of identity over the entire video sequence requires tracking the face [1].
3. *Face modeling.* Reconstruction of the 3D model of a face from a video sequence using structure from motion requires tracking. This is because the depth estimates are related nonlinearly to the 3D motion of the object. This is a difficult nonlinear estimation problem and many papers can be found that focus primarily on this, some examples being [2–4].
4. *Video communications and multimedia systems.* Face tracking is also important for applications like video communications. Motion estimates remove the interframe redundancy in video compression schemes like MPEG and H.26x. In multimedia systems like sports videos, face tracking can be used in conjunction with recognition or reconstruction modules, or for focusing on a region of interest in the image.

Summary

Face tracking is an important criterion for a number of applications, like video surveillance, biometrics, video communications, and so on. A number of methods have been proposed that work reasonably well under

moderate changes of pose, lighting and scale. The output of these methods vary from head location in the image frame to tracked facial features to 3D pose estimation. The main challenge that future research should address is robustness to changing environmental conditions, facial expressions, occlusions, clutter, and resolution.

Related Entries

- ▶ [Face Alignment](#)
- ▶ [Face Recognition](#)

References

1. Zhou, S., Krueger, V., Chellappa, R.: Probabilistic recognition of human faces from video. *Comput. Vision Image Understand.* **91**, 214–245 (2003)
2. Fua, P.: Regularized bundle-adjustment to model heads from image sequences without calibration data. *Int. J. Comput. Vision* **38**, 153–171 (2000)
3. Shan, Y., Liu, Z., Zhang, Z.: Model-based bundle adjustment with application to face modeling. In: Proceedings of IEEE International Conference on Computer Vision, pp. 644–651 (2001)
4. Roy-Chowdhury, A., Chellappa, R., Gupta, R.: 3D face modeling from monocular video sequences. In: *Face Processing: Advanced Modeling and Methods*. Academic Press, New York (2005)
5. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face Recognition: A Literature Survey. *ACM Transactions* (2003)
6. Baker, S., Matthews, I.: Lucas–Kanade 20 years on: A unifying framework. *Int. J. Comput. Vision* **56**, 221–255 (2004)
7. Xu, Y., Roy-Chowdhury, A.: Integrating motion, illumination and structure in video sequences, with applications in illumination-invariant tracking. *IEEE Trans. Pattern Anal. Machine Intell.* Vol. 29, 793–806 (2007)
8. Luenburger, D.: *Optimization by Vector Space Methods*. Wiley, New York (1969)
9. Hager, G.D., Belhumeur, P.: Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 1025–1039 (1998)
10. Koterba, S., Baker, S., Matthews, I., Hu, C., Xiao, H., Cohn, J., Kanade, T.: Multi-view aam fitting and camera calibration. In: *IEEE Intl. Conf. Comput. Vision* (2005)
11. Lepetit, V., Fua, P.: *Monocular Model-Based 3D Tracking of Rigid Objects*. Now Publishers Inc. (2005)
12. Fasel, B., Luettin, J.: Automatic facial expression analysis: a survey. *Pattern Recognit.* **36**, 259–275 (2003)
13. Terzopoulos, D., Waters, K.: Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 569–579 (1993)
14. Black, M., Yacoob, Y.: Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In: *International Conference on Computer Vision*, pp. 374–381 (1995)

15. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 1063–1074 (2003)
16. Arulampalam, M., Maskell, A., Gordon, N., Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **50** (2002)
17. Dedeoglu, G., Baker, S., Kanade, T.: Resolution-aware fitting of active appearance models to low-resolution images. In: European Conference on Computer Vision (2006)
18. Yu, J., Bhanu, B., Xu, Y., Roy-Chowdhury, A.: Super-resolved facial texture under changing pose and illumination. In: International Conference on Image Processing (2007)

Face Variation

CARLOS D. CASTILLO, DAVID W. JACOBS
 Department of Computer Science, University of Maryland, College Park, MD, USA

Synonym

Facial changes

Definition

Face variation refers to the way in which the appearance of the face changes due to changes in viewing conditions such as illumination or pose, or due to changes in properties of the face, such as its expression or age.

Introduction

Face recognition is a fundamental problem in biometrics. One of the chief sources of difficulty in face recognition is the large number of variations that can affect the appearance of faces. These include changes in lighting, pose, facial expression, makeup, hair, glasses, facial hair, occlusion by objects that block part of the face from view, aging, and weight gain or loss. Many studies suggest that these variations can significantly reduce the performance of recognition algorithms.

Some face recognition systems aimed at cooperative subjects deal with this problem by attempting to control these sources of variation. This may be appropriate for some applications. In these cases, pose can

be controlled by requiring a subject to look into the camera, which is kept at a fixed height. Indoors, controlled lighting can be employed. And subjects may be requested to keep a neutral facial expression, and avoid variation in occluding objects, such as eye glasses or scarves. Working under such controlled conditions, face recognition systems have achieved high levels of accuracy [1].

However, in many cases large variations in appearance cannot be controlled. One may wish to recognize a person based on photographs taken some time before, as when one verifies that a person matches a passport photograph. In this case, changes in appearance due to aging, changes in weight, or variations in hair style will be inevitable. In many applications involving security or interactions between a computer or robot and a person, at least a few days may pass between the time a face is first learned and then later recognized. Even over short time periods there may be variation in a face due to changes in makeup, or in how recently a subject has shaved. Finally, in many applications, even lighting, pose or facial expression cannot be controlled, either because a subject is uncooperative or because one wishes to have the flexibility to recognize people as they move naturally through an environment, changing their position relative to the camera and lights.

There has been relatively less work on face recognition in the presence of these variations than for recognition under controlled conditions. Of these variations, lighting and pose variation have received the most attention. Many other sources of variation have been the subject of only a few research efforts. For example, to the authors' knowledge, there has been no work explicitly aimed at accounting for variations due to weight change. Furthermore, most research has been limited to the case in which conditions are controlled when subjects are enrolled into a *gallery* of known faces, so that, for example, all gallery images are taken in the same pose or lighting. Then, research focuses on matching a *probe* face viewed under different conditions to the correct entry in the gallery. Face recognition becomes much more difficult when the gallery is imaged under heterogeneous conditions. For example, there may be a tendency to match a face viewed with side lighting to the gallery face of a different person viewed with the same lighting when the gallery face of the correct person is acquired under very different lighting conditions [2]. However, there are many applications, such as the organization of personal photos, in which it may not

be possible to acquire a gallery of images taken under controlled conditions.

Illumination

Changes in lighting can produce significant variation in the appearance of a face. These changes occur due to an actual variation in lighting, such as the difference between indoor and outdoor illumination, but they also occur when a person moves relative to even a fixed set of lights. Therefore, illumination changes must be accounted for in a wide range of recognition scenarios. Adini et al. [2] has shown that using common measures of image similarity, there is greater similarity between two images of different people taken under the same lighting conditions than between two images of the same person taken under quite different lighting conditions. As a consequence, in spite of a number of research efforts, existing recognition algorithms show much poorer performance in the presence of lighting variations than when used with controlled lighting [1].

A number of approaches have been taken in order to mitigate the effects of lighting change. Three common strategies include the following. First, one can apply image representations that are generically insensitive to lighting variation. Second, one can train a recognition system using sets of images that provide examples of the effects of lighting variation on images of faces. Third, one can use knowledge of the three-dimensional shape of faces either to predict the effect that changes in lighting might have on their appearance in images, or as a representation that is unaffected by lighting. These approaches are summarized briefly here; the reader can find more details in [3, 4, 5], and [6].

Determining the intrinsic properties of a scene independent of lighting conditions is a classic problem in computer vision that has been studied for decades. Multiplicative and additive effects of lighting can be removed by normalizing the mean and variance of the image intensities. ► [Histogram equalization](#) has been applied to remove lighting effects that produce a monotonic change in image intensities. Finally, some representations of images have been shown to be less sensitive to lighting variations, including the direction of image gradients, vectors containing the output of Gabor filters [7], or representations that attribute

low frequency components of the image to lighting, and remove these effects. These and other, related techniques, have been shown to produce substantial improvements in recognition performance compared to methods that compare raw pixel intensities.

In a second approach, a *training* set of images is used to learn the effects of lighting variation on the appearance of faces. The training set may contain images of many individuals who are different from those the system will later try to recognize. These images show the variation in appearance of each person in the training set as the lighting varies. Methods such as ► [Linear Discriminant Analysis](#) may be used to then find representations of faces that best capture the information that varies between individuals, while discarding information that varies due to light, but not due to identity [8]. There is also a good deal of evidence that the set of images that a face produces under a wide range of lighting conditions occupy a low-dimensional linear subspace in the space of all possible images. This implies that when the gallery contains multiple images of each subject, taken under different lighting conditions, a linear subspace spanned by these images can be used to represent the subject.

A third set of methods makes use of knowledge of the 3D structure and surface reflectance properties of faces to predict and compensate for the effects of lighting [5]. This can involve obtaining a model of each face to be recognized. Acquisition systems that can capture the 3D structure of a face, along with the varying surface properties of eyebrows, lips, and skin exist. This makes the process of enrollment into the biometric system more complex, though. An alternative is to use general knowledge obtained from 3D scans of a training set of individuals other than the person to be recognized. In the latter case, a generic face model may be fit to a gallery image, producing a model specific to that person. A model of a person's face can be used to solve for the lighting that best matches that model to the probe image. Recognition can then be performed by comparing the probe image to a rendering of the model, produced by computer graphics. Other approaches may use the model to build representations of a face's appearance under diverse lighting conditions, and compare these to the probe image. Finally, if one obtains a 3D model as a probe, this can be directly compared to a 3D model acquired at enrollment.

Researchers have collected a number of data sets that contain images of the same individual under varying illumination, in order to measure the effect of lighting on recognition algorithms. Due to the difficulty of building such data sets, they are usually acquired with either a relatively small number of individuals or a small number of lighting conditions. For example, Carnegie Mellon University's Pose, Expression, and Illumination (PIE) data set contains images of 68 different people illuminated in turn by 21 different flash bulbs in known positions, while a variety of data sets contain images of more than a thousand individuals taken with just a few lighting conditions [9]. It is not clear how many lighting conditions are needed in a data set to thoroughly test recognition algorithms. The actual variability of lighting is very great, because even with lights distant from a face, the lighting intensity is a 2D function of direction. This means that it is difficult to record or simulate the lighting present in realistic conditions, and that it is also difficult to systematically explore the space of possible lighting conditions.

Pose

Face recognition with pose variation refers to recognizing faces when the cameras used to take gallery

and probe images have different angles relative to the subject (e.g., Fig. 1). For example, there is a pose variation when subjects are described using a gallery of images taken with subjects facing the camera and when one uses a probe image of a subject seen in profile, but not when the probe image is simply taken from a different distance than the gallery images. When there is a pose variation, one may see different parts of the face in the gallery and probe images; for example, in a profile view, one side of the face may be unobserved. Moreover, the apparent size of different parts of the face may vary with pose. In profile, the cheek takes up a larger part of the image than it does when the face is viewed frontally, while the forehead may be more foreshortened. A number of experiments suggest that when one uses recognition algorithms that do not explicitly account for pose, performance deteriorates a great deal with significant pose variations.

Pose variations create a correspondence problem that does not occur with a number of other types of variations. It is common for general recognition algorithms to align faces by detecting and aligning a few features, such as the center of the eyes. When two images of a face are taken from frontal views, aligning the eyes tends to align all the other features of the face (although this is not quite true for some variations described below, such as changes in expression).



Face Variation. **Figure 1** The same person photographed in two different positions. Moving relative to the lighting and camera causes significant changes in appearance.

Many systems rely on this alignment by then comparing corresponding image pixels. However, when there is a pose variation, finding corresponding image pixels is much more difficult. Aligning the eyes in a frontal and a profile view will not align other parts of the face, such as the nose. Face recognition systems that can handle pose variation, then, must generally find some method of solving this correspondence problem.

This section discusses three approaches to this problem. The first involves representing multiple views of the face, so that a simple alignment with one of these views will match the probe. The second uses 2D image matching methods to find corresponding pixels. The third uses 3D representations to assist in solving for correspondence. These approaches are discussed further in [4, 5] and [6].

Many face recognition algorithms that are not designed to handle pose variation are still robust to small rotations of the head, of up to 15–30° [4]. This suggests that if the gallery contains images of each subject, taken at poses sampled by 30°, one of these gallery images will provide a good match to a probe. Such galleries have been constructed either by acquiring multiple images per subject, by constructing a 3D model of the subject and using it to generate appropriate views, using computer graphics, or by using training data to infer the changes of appearance in a face as viewpoint changes. These approaches may have the disadvantage of making enrollment into the gallery more complex, and may still degrade recognition performance to some degree when probes are taken at an angle between sampled directions.

Methods taking the second approach use some mechanism to find good correspondences between individual locations in the probe and gallery images. For example, [7] locates distinct features, such as the corner of the eyes, and builds descriptors of these locations using vectors containing the output of Gabor filters. Then corresponding features are matched between two images, allowing for changes in the relative position of features due to a pose change. Other work has matched individual pixels in images using ▶ optical flow [10] or stereo matching algorithms. These are matching methods developed for general computer vision problems in which a scene is viewed from different locations. These approaches may be supplemented by building a statistical model that captures the way a feature's appearance can vary with pose [4].

Finally, 3D face information may be used to account for pose. One way to do this is to acquire a 3D description of each subject when he or she is enrolled in the gallery. A small set of features can then be used to align this model with a 2D probe image. The 3D and 2D data must then be compared, which can be done, for example, by solving for the lighting that best matches them. Alternately, a system can obtain 3D information from the probe, and compare 3D representations directly. These approaches, though, depend on more complex sensing for enrollment, and possibly for recognition. An alternative approach (see [5]) builds a generic, 3D morphable model that can morph between the shapes of a set of training faces. This model can then be applied to any subject. By fitting the model to a probe image, the 3D structure of the probe face can be estimated. This can then be used to render the probe in a canonical pose, or it can be compared to similar 3D reconstructions of the gallery faces.

While progress has been made in handling pose variations, significant challenges remain. In particular, there are many applications in which one expects the gallery to contain a single image of the subject, and the probe to consist of a new image, taken in a new pose. For this problem, current methods have substantially worse performance than when pose is fixed between the probe and gallery. In addition, many methods for handling pose variation require substantially more computation than other methods, and can be very slow. This is partly because the process of finding a correspondence between the probe and gallery requires expensive optimization processes.

Expression and Occlusion

Changes in expression can also have a considerable effect on the appearance of a face that can have a major impact on recognition performance (see [11] for a fuller discussion). These can be divided into two sorts of effects. When one smiles, frowns, or purses one's lips, there is a change in shape, as the lips move and the cheeks alter their position. But expression can also cause facial features to appear or disappear. For example, smiling may reveal our teeth, blinking or winking may block an eye from view, frowning may cause new wrinkles to appear in the forehead. For this reason, it is convenient to class together changes in

expression with other occlusions, as when sunglasses or a scarf block part of the face from view.

Less work has been done on the problem of expression variation than on lighting or pose. One approach is to use recognition algorithms that can ignore or de-emphasize portions of the face that might be affected by expression change or occlusion. This can be done if training data is available that provides examples of these variations. Then, for example, Linear Discriminant Analysis can learn a linear projection that has the effect of placing less weight on portions of the face that are likely to change [8]. Or, one can divide the face into regions and learn weights that indicate the value of each region in identification. Regions of occlusion in a probe face can also be identified as regions that are not sufficiently similar to a space of face regions, and these can be discarded before matching the probe to the gallery [11]. In principal, changes in shape due to facial expression can be accounted for by using methods such as optical flow to find a correspondence between images of faces with different expressions [10].

However, such an approach must be able to distinguish between changes in shape caused by expression, and differences in shape between the faces of different people. Also, correspondences cannot be found when expression change causes features to appear or disappear. Because of their difficulty, many of the issues raised by changing expression have not been studied extensively.

Sources of Variation that Occur Over Time

Other sources of facial variation have received much less attention. These include changes in glasses, hair style, makeup, weight, or the effects of aging (See Fig. 2). While pose, expression and lighting can change from one moment to the next, these additional factors tend not to change very frequently. However, any system that wishes to recognize people after a period of a few months or a few years will have to account for these sources of variation.



Face Variation. Figure 2 Two sets of photos showing changes in appearance over time. Left: passport photos taken at 10 year intervals. Right: photos taken at age 6, 16, and 23. There is a considerable change in appearance due to the effects of aging, weight gain, glasses, and changing hair and facial expression.

One reason that there may have been less effort directed at these variations is the difficulty of obtaining valid experimental data. For example, it would be daunting to collect images of large numbers of subjects before and after significant changes in weight. It is also much more challenging to collect face images over a period of many years than to collect images from different viewpoints, or with changes in lighting. The government does collect photos of individuals over long periods of time, for passports or drivers licenses, for example, but privacy concerns prevent widespread use of this data. As individuals post large collections of personal photos on the internet there is a growing opportunity to build innovative new data sets of face images, although by their nature, many of the imaging conditions in these photos are uncontrolled and unknown.

The complex set of factors that affect facial appearance over time are discussed in [12]. In children, there is significant change in face shape as they grow up. In adults, there is less change in shape due to aging, and more change in the appearance of skin due to exposure to sunlight and the appearance of wrinkles [12] and subsequent work describe experiments with a number of recognition algorithms, including two commercial systems, on data sets containing passport photos of nearly 2,000 individuals, with a time lag between photos ranging from 1–10 years. In a verification task that asks whether two photos come from the same or different people, performance is far below the levels achieved using photos taken under controlled conditions with little time lag [1]. It appears that there is a sharp increase in the difficulty of recognition when 1 year passes between images, and that, at least for adults, further passage of time, up to 10 years, creates only small additional increases in difficulty. It is not clear how much of these problems are due to aging, and how much can be attributed to other changes in, for example, weight or hair style that tend to occur over time, or even to other factors such as artifacts caused by the scanning of passport photos.

In addition to aging, a number of other sources of variation have been mentioned in the literature, but have not received much study. For example, a number of researchers have noted that the presence or absence of makeup on a face can affect the difficulty of recognizing it, but there is little systematic work in this area. Similarly, it is clear that significant changes in weight can affect facial appearance, but there has been

little if any work in this area. Variations in hair style or grooming can also have a considerable effect on appearance; partially for this reason most approaches to face recognition focus on the inner part of the face, and attempt to ignore the outer head and hair. However, since the outer head and hair seem to be important in human face recognition, it seems that understanding hair appearance and its variations could be of potential value in face recognition systems.

Conclusions

In summary, while most work on face recognition has focused on settings in which there is little variation in a face or in the viewing conditions, there is also a growing amount of work that addresses face variations. In many cases, each source of variation has been addressed with methods specific to that type of variation. For example, lighting variation has been attacked using lighting insensitive image representations, while pose methods often focus on the correspondence problem. Two exceptions are first, model-based methods, such as those using morphable models, that extract a 3D model from an image, and then use computer graphics to normalize its appearance and remove face variations, and second, pattern recognition and learning methods, such as Linear Discriminant Analysis, that can potentially characterize any specific variation, provided there is appropriate training data.

Many face variations can cause significant degradation in performance in standard recognition methods. While interesting progress has been made in developing recognition methods that account for these variations, these methods generally still have performance that is substantially less than that can be achieved when variations are controlled.

While many challenges remain, this article has mentioned three in particular. First, there has been little research aimed at developing methods suitable for handling multiple simultaneous facial changes. For example, it is not clear whether many of the methods developed to handle lighting changes will be suitable when there is also pose variation. Second, most work has focused on situations in which the gallery images are taken under uniform conditions. Surely, for example, recognition will be more difficult when the gallery contains a single image of each person taken with different poses. Third, variations that occur over time

have not been well explored, and the relative importance of different effects, such as aging, weight change, or changing hair or makeup is not clear.

Acknowledgments

The authors have been supported by a fellowship from Apptis, Inc., and by a Honda Research Initiation Grant.

Related Entries

- ▶ Deformable Models
- ▶ Face Alignment
- ▶ Face Descriptors
- ▶ Face Pose Analysis
- ▶ Facial Expression Recognition
- ▶ Illumination Compensation

References

1. Phillips, P.J., Scruggs, W.T., O'Toole, A., Flynn, P., Bowyer, K., Schott, C., Sharpe, M.: FRVT 2006 and ICE 2006 large-scale results, National Institute of Standards and Technology Report NISTIR 7408 (2007)
2. Adini, Y., Moses, Y., Ullman, S.: Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 721–732 (1997)
3. Basri, R., Jacobs, D.: Illumination modeling for face recognition. In: Li, S., Jain, A. (eds.) *The Handbook of Face Recognition*, pp. 95–120. Springer, New York (2005)
4. Gross, R., Baker, S., Matthews, I., Kanade, T.: Face recognition across pose and illumination. In: Li, S., Jain, A. (eds.) *The Handbook of Face Recognition*, pp. 203–228. Springer, New York (2005)
5. Romdhani, S., Ho, J., Vetter, T., Kriegman, D.: Face recognition using 3-D models: Pose and illumination. *Proc. IEEE* **94**(11), 1977–1999 (2006)
6. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. *ACM Comput. Surv.* **35**(4), 399–458 (2003)
7. Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* **42**(3), 300–311 (1993)
8. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997)
9. Gross, R.: Face databases. In: Li, S., Jain, A. (eds.) *The Handbook of Face Recognition*, pp. 319–346. Springer, New York (2005)
10. Beymer, D., Poggio, T.: Image representations for visual learning. *Science* **272**, 1905–1909 (1996)
11. Martinez, A.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(6), 748–763 (2002)
12. Ramanathan, R., Chellappa, R.: Face verification across age progression. *IEEE Trans. Image Process.* **15**(11), 3349–3361 (2006)

Face Verification

- ▶ Face Recognition, Overview
- ▶ Liveness Assurance in Face Authentication

Face Warping

- ▶ Image Warping
- ▶ Face Alignment
- ▶ Face Device
- ▶ Face Tracking

Face, Forensic Evidence of

MICHAEL C. BROMBY

Glasgow Caledonian University, Glasgow, UK

Synonyms

Face Identification; Face Recognition; Face Reconstruction; Facial Mapping

Definition

Using the face as a biometric feature requires an image or representation of the face, which is then subjected to manual or computerized analysis. This relies on an examination of the individual features (such as the eyes, nose, ears etc) or of the image as a whole (areas

of light and dark, texture, and color). Notably, face biometrics does not require active participation and can operate in a covert manner and may also be deployed from a distance. The major factors affecting the reliability and accuracy of face biometrics are illumination, pose, and expression.

Three main areas where the biometrics of the face is used are verification, identification, and reconstruction. First, the technique of comparing photographs of an offender with images of a suspect is occasionally termed ► *facial mapping*. Exculpatory evidence can be obtained if marked differences are apparent from expert analysis that cannot be explained. In contrast, similarity cannot indicate identity unless the presence of unique identifiers can be established. Second, computerized recognition and identification can provide a faster and more accurate method to search for a target in a database. This method may also be deployed in real time and generate a name or identifiable record when the target is present. Third, the process of skull reconstruction can provide a facial likeness, using either manual or computerized techniques, which is generally used for historical cases requiring identification.

Introduction

Face biometrics is regarded as less intrusive than other methods such as fingerprint analysis, iris scans, or palm morphologies, which generally require co-operation from the subject and an awareness of the procedure being undertaken. The face can be easily captured from a wide variety of low-cost sources, including public area CCTV, photographs, etc. The face also differs from the collectable forms of trail evidence left at the crime scene. Since the face is an intrinsic part of the owner, it cannot leave a physical trail other than a visual recording. Therefore, it may be seen as an exception to Locard's Exchange Principle in some instances.

Three main areas where the biometrics of the face is used are verification, identification, and reconstruction. Any of these methods may be susceptible to errors arising from the subjective nature of the interpretation of the face, whether by a person or by a machine. The ability of the face to move in three dimensions is also aggravated by internal movements of the eyes, lips, and cheek areas due to expressions such as

smiling, frowning, blinking etc. Recognition from an image is also subject to the inherent limitations involved in general image processing such as resolution and lighting levels.

Evidence from facial biometrics is frequently seen as a corroborative tool to support other methods of identification. Its value as a single method of identification is lower than some other forms of biometric identifiers (notably fingerprints and DNA), and therefore, it is more often used in historical cases or those lacking in other forms of evidence.

F

Expert Image Analysis

Manual analysis of facial biometric features (often termed *facial mapping*) is primarily a feature of crime investigation. A variety of methods can be employed individually or in combination to compare a crime scene image of an offender with an image of the suspect. The face must be of a similar three-dimensional alignment as the head can move in a number of directions. Experts in this field have been providing reports to provide identification evidence, principally in criminal cases, since 1997 [1].

This category of facial biometrics employs a process of ► *one-to-one matching* – a verification process of checking allegations or suspicions, whereby other forms of evidence must be present to suggest the involvement of a specific individual. A one-to-one facial mapping technique provides additional scientific evidence to support the existing case. The use of this matching technique lends support to the judicial decision-makers who must be persuaded beyond reasonable doubt.

Facial Mapping Techniques

Although there are limited publications regarding facial image comparisons, many methodologies employed for image comparisons are drawn from peer-reviewed and accepted practices within other disciplines. Examples of video superimposition, morphological classifications, three-dimensional analysis, and geometric analysis are given in the literature and relate directly to forensic image comparisons for the purposes of identification [2–5].

The definition of “facial mapping” from the ACPO manual issued by the Working Group for Facial Identification is as follows:

1. Facial Identification by image comparison – concerned with the identity of an individual from scaled and aligned photographic images or by demonstrating morphologically comparable features, within a legal context.
2. To make a visual study of moving and/or still facial images in a variety of formats (video, digital, photographs etc) obtained from the scene of a crime or other source and make a scientific comparison with a suspect’s facial image.
3. To present and demonstrate the significance of any area/point of similarity and difference, the presence or absence of a feature and any probability factor as well as the likelihood of repetition so as to formulate an opinion of similarity from these comparisons.
4. Similarities of features and facial proportion do not necessarily prove the identity, although differences may prove nonidentity. However, as the number of similarities increases, the number of people who share that particular combination of features/proportions decreases, thereby adding weight (to whatever degree) to the assumption that the persons in question are the same [6].

From this wide definition, it is clear that there is no single procedure or methodology required for comparing images. The ACPO document lists a number of methodologies that may be used to compare images. The document indicates that this list [6] is by no means exhaustive:

1. Drawn or electronically produced indicators/grids.
2. Transposed outlines (produced by hand or by computer).
3. Split or composite images (one or any portion of an image is overlaid on the second image to check/confirm correlation).
4. Video overlays/Wipes on a frame-by-frame basis.
5. Facial proportions/spatial distribution of features.

In general, these may be categorized into scaling and alignment methods to assess relative facial landmarks (size, shape, and position of facial features) and morphological comparisons. In practice, one or a combination of these methods is used, dependent on the imagery available. Additionally, ► *photogrammetry* has been employed using two images taken at different vantage points to create three-dimensional representations.

Appeal Court Cases

In reviewing the circumstances in which identification evidence based on CCTV or photographic imagery was admissible, the Court of Appeal for England and Wales identified four possible routes to achieve a valid identification. The fourth route was identified as being

- ▶ “a suitably qualified expert with facial mapping skills [who] can give opinion evidence of identification based on a comparison between images from the scene, (whether expertly enhanced or not) and a reasonably contemporary photograph of the defendant, provided the images and the photograph are available for the jury (*Stockwell* 97 Cr App R 260, *Clarke* [1995] 2 Cr App R 425 and *Hookway* [1999] Crim LR 750)” [7].

This response indicates that the admissibility of expert image analysis per se has remained unaltered since its first introduction as evidence of identification; and that the test of whether the court, in each instance, requires assistance in interpreting images through an expert witness is to be applied. In applying this test, the first route to achieving a valid identification as stated by the Vice President Rose, LJ declares that

- ▶ “where the photographic image is sufficiently clear, the jury can compare it with the defendant sitting in the dock (*Dodson & Williams*)” [7].

Under these circumstances, an expert opinion is clearly not required irrespective of whether the witness is indeed an expert. This scenario is significant as it can be distinguished from cases where an opinion is not admitted as evidence due to the lack of skill or knowledge claimed by the purported expert.

The *Attorney General's Reference* does, however, raise the question of what constitutes “suitable qualifications”, which are not listed by the ACPO guidance, and presents investigators with a perennial problem. It could be seen that the Court of Appeal had the ideal opportunity to examine in greater detail the issues of admissibility, reliability, or indeed, sufficiency of image analysis as evidence of identification. Their reluctance to do so illustrates that there may not be a clear or singular answer to these issues.

In the UK case *R v Gray* criticisms were made by Mitting, J regarding the absence of statistical databases or any such means to determine a mathematical formula [8]. This did not develop any rule (as suggested in *R v Gardner*) that an expert cannot go further than saying “there are the following similarities”,

leaving the ultimate decision to the jury, as opposed to the expert witness actually giving a view as to a degree of probability of the images being the same. The decision in *Gardner* does not doubt the admissibility of forensic image comparisons [9]. The appeal was based upon the inequality of arms as the defense team did not have access to the expert's laboratory material, upon which they could cross examine.

Mardia developed a database of facial statistics to determine whether, like fingerprints, there could be a certain number of matches on a face that would determine uniqueness. This study allows for a prevalence assessment of various facial feature classifications and angles of the face, although within a limited sample population of 358 Caucasian males [5]. Although this is not a nationally recognized database, it fulfills some criteria of objectivity within a measurement of uniqueness, as a sample size of only 50 achieved the same prevalence rates in a Home Office study by Wilcox [10].

Computer Analysis

As discussed earlier, an expert is able to make facial comparisons, using photographic evidence. The errors associated with human judgment may, on occasion, reduce the reliability of the expert and their evidence. Computerized facial recognition may eliminate the possible errors associated with both inter- and intra-operator variables. Many studies into computerized recognition have tried to adapt the psychological models of human recognition to work toward a fully computerized system of facial recognition.

► *Principal Component Analysis* is based on feature identification: a face is identified and stored, the image is then analyzed on the digital composition and the principal components or areas of light and dark are noted [11]. For example, thicker lips will possess a greater surface area and will vary in brightness and contrast between individuals. Areas of light and dark along the edge of the face also serve to identify face shape and relative size. A unique set of data for each individual face is created, which may then be used as a template or ► *eigenface* to enable the system to recognize the same face, or more correctly, the same set of data in the future.

An alternative model of ► *Graph Matching* [12] relies on the configurational identification of a face. This relates to the examination of the measurable distances between features and the relative ratios of height

and width rather than the examination of the features themselves. The eyes can be identified automatically and the locations of the other features can be added if required by the software. A unique algorithm is created from the key points on the face; this algorithm is unique as a fingerprint or DNA profile. This second model is more similar to the task of facial mapping performed by experts, described earlier. However, with either method, there is still sufficient information to recognize and identify faces. The speed by which a result is obtained would favor an automatic computerized process, although it may be argued that a more thorough and reliable comparison can be made by using human input to locate the facial features.

Computerized techniques can assist ► *one-to-many identification* by searching through archive databanks of facial images. One-to-many matching for criminal justice purposes requires an extensive database of facial images collected either from police custody records or created from noncriminal records such as the face image held by the Passport Office or the Driver and Vehicle Licensing Authority (DVLA). A fully comprehensive national database of all adult facial images obtained from noncriminal records would not be in accordance with the protection offered under the legislation governing the use of data.

Reliability of Computerized Identification

The in-house testing of facial recognition systems by software companies can be extremely subjective, with varying aims and test data, depending on the actual use and requirements of the tasks that the algorithms were developed to perform. Accuracy and reliability can only be assessed by comparing a product with standardized references or samples and further analysis by independent bodies. The FERET Verification Testing Protocol for Face Recognition Algorithms was devised to provide an accurate and independent assessment of the reliability and accuracy of the existing facial recognition systems [13]. It also served to promote research in facial biometrics in academic and public/private sector industry, sponsored by the United States Department of Defense Counter-drug Technology Development Program. A Target set of "known individuals" and a Query set of "unknown faces" were presented to participating software developers. Two versions of testing were administered: the first assessed automatic facial location, and the second version

provided eye coordinates to assess the recognition performance of manual input systems. Enrollment and test data were collected according to strict guidelines to enable a fair comparison to be made. A scoring procedure was devised based on Receiver Operating Characteristic (ROC) graphs originally devised for SONAR false recognition rates [14].

A significant increase in performance was seen for the general field of facial biometric comparison and for each individual algorithm-based system [14]. Strengths and weaknesses of each algorithm were highlighted to facilitate further research to promote and improve the use of facial biometrics. It was evident from the FERET tests that further research was still required if facial biometrics were to compete with other forms of biometric identification such as fingerprints, even though progress had been made in these areas over a given period. A major fault of face recognition algorithms appeared to be sensitivity to variations in illumination, caused by the change in sunlight intensities throughout the day.

By placing a surveillance system in a unique area and attaching a database specific to known criminals who would operate in that area, a reasonable successful hit rate can be achieved without infringing on the general privacy of the public. From the example put forward by Newham Council, other locations may be highlighted as target areas for particular types of offenders. Airports are prime examples of sites that are frequented by a variety of individuals involved in crimes ranging from terrorism to drug trafficking and illegal immigration. Security cameras are a regular feature of many public spaces and their presence has become ubiquitous because of their intrusive abilities to detect, recognize, and identify individuals without requiring an active participation or the knowledge of the subject.

The use of facial biometrics as a token for civilian verification of identity (for example, secure access, banking etc) is not so well employed. The benefits of not forgetting (as with passwords), not being lost (cards and keys), and being noninvasive (fingerprints etc) are often outweighed by higher false rejection rates when compared with other biometric systems.

Automatic Recognition in Practice

By combining automatic recognition technology and criminal databases of known offenders, computer systems to alarm law enforcement agencies as to the real-time presence of a known criminal have been developed. The first CCTV and facial recognition system in the United Kingdom was instigated by the Metropolitan Police in Newham, East London [15]. In spite of the pressure from many civil liberties groups, the Mandrake system examined every passing face and alerted the police when an individual is recognized from the hit-list database. Despite analyzing every single face in a crowd, information was only stored when a match was made, and data from inconclusive analyses were discarded. The system relied wholly on a graph matching system, analyzing the area around the eyes and the nose, which was converted into an algorithm without any manual intervention. This means of crime prevention has inherent limitations, as unwarranted surveillance in anticipation of any crime occurring by chance is not permitted under Sections 28 and 29 of the Regulation of Investigatory Powers Act 2000. However, the selection of faces to be recognized and the specific locations of the CCTV cameras may permit facial recognition systems to be used for crime prevention.

Skull Reconstruction

In the absence of biological evidence such as DNA or identifiable personal artifacts, the naming of skeletonized or badly deformed remains may require the reconstruction of the face from the skull in order to identify the deceased.

Historically, the principle of relating the skeletal structure to the overlying soft tissue has been applied to all forms of reconstruction: 2-D drawings, 3-D clay sculpting, or computerized modeling. The skull clearly provides a vast amount of information on how the final face should appear. The sex, age, and racial origins can be determined, although any error will have significant repercussions throughout the whole procedure and will ultimately distort the reconstruction, possibly hindering the processes of recognition and identification by people familiar with the deceased. The relationship between hard and soft tissues of the face and facial tissue depth measurement provide the foundations for accurate reconstructions [16]. Some factors cannot be accounted for, such as the nutritional state of the individual.

While measuring the facial tissue depth, the number, and position of anthropometric landmarks are

subjective. Although published texts provide authoritative views, inter- and intraresearcher variation will persist in locating these points. Data from early cadaveric studies were subject to error due to shrinkage, bloating, and the effects of gravity when lying supine. Gravity, along with high radiation doses, persists to be a problem with modern advances in MRI and CT scans. Ultrasound is presented as the most reliable method, experiments providing comparative data from several ethnic groups [17, 18].

The Manchester Method relies on the knowledge of the gross anatomy of the face to recreate the muscle fibers and glands on a plaster cast copy of the skull [16]. Unsurprisingly, this bottom-up process of rebuilding the face differs from the standard textbook descriptions of dissecting the facial musculature in a top-down fashion. Each muscle is created and attached, using published data and experience to recreate the underlying structures that will ultimately reflect the final skin surface with the minimum possibility of subjective interference.

In forensic cases, the addition of hairstyles, facial hair, blemishes, wrinkles, scars, or identifiable marks should not be added unless evidence suggests otherwise. Interestingly, details such as the hairline, forehead creases, eyelid patterns, nasolabial folds, and cheek shapes are some of the many features that can be determined, to some degree, from the skull and the previous muscle attachments. Creating a realistic and believable face is a difficult task balanced with the distraction of wrong information such as hair or eye color. Additional information may be superimposed using a computer software to generate a number of alternatives.

The accuracy of forensic facial reconstructions is the singularly most important factor in obtaining an identity for the deceased. Qualitative studies comparing the likeness with a photograph of the deceased have shown remarkable results. Blind testing using a variety of techniques has reported rates of 50, 65, and 75 per cent [16]. Quantitatively, very positive results have been obtained by conducting “identity parade” style face pools, using volunteers to assess the likeness against a number of targets [19]. The process of identifying unfamiliar faces is poorer than the ability to identify familiar faces, suggesting that these results are lower than what would be expected from family or friends of the reconstructed person.

Computerized face reconstruction, using three dimensional scanning of the skull has been reported

as more reproducible than clay modeling, although subjectivity still remains in placing the pegs on the digitized skull [4, 20]. The benefit of a digital reconstruction is the flexibility of the final product, which may be aged or temporarily altered with greater ease than a more permanent clay final product.

Summary

The procedures involved in forensic face identification vary in both method and purpose according to whether the face is represented as a two-dimensional image or a three-dimensional skull. Evidence from all the three areas of expert or computer image analysis and skull reconstructions can be useful in obtaining an identification. The reliability and accuracy of each method may be prone to errors and the value of such evidence must be weighed in conjunction with other forms of identification or evaluated with some degree of caution if presented alone.

Related Entries

- ▶ Biometrics, Overview
- ▶ Biometric Recognition
- ▶ Face Recognition, Overview

References

1. *R v Ryan* (unreported): The Guardian Newspaper 26 April, 1991; cited in *R v Stockwell* (1997). Cr App R 260 at 264
2. Vanezis, P., Brierley, C.: Facial image comparison of crime suspects using video superimposition. *Sci. Justice.* **36**, 27–33 (1996)
3. Vanezis, M.P., Lu, D., Cockburn, J., Gonzalez, A., McCombe, G., Trujillo, O.: Morphological classification of facial features in adult Caucasian males based on an assessment of photographs of 50 subjects. *J. Forensic Sci.* **41**, 786–791 (1996)
4. Yoshino, M., Matsuda, H., Kubota, S., Imaizumi, K., Miyasaka, S.: Computer-assisted facial image identification system using a 3D physiognomic rangefinder. *J. Forensic Sci. Int.* **109**, 225–237 (2000)
5. Mardia, K., Coombes, A., Kirkbride, J., Linney A., Bowie, J.: On Statistical Problems with Face Identification from Photographs. *J. Appl. Stat.* **23**(6), 655–675 (1996)
6. Association of Chief Police Officers: National Working Practices in Facial Imaging. Home Office, London (2003) accessible at http://www.acpo.police.uk/asp/policies/Data/garvin_facial_imaging_guidelines.doc.
7. Rose, L.J.: In *Attorney-General's Reference No. 2 of 2002* (2003) Cr.App.R. 321

8. Mitting, J. in *R v Gray* (2003) EWCA 1001
9. *R v Gardner* (2004) EWCA 1639
10. Wilcox, R.: Facial Feature Prevalence Survey. London: Home Office Research, Development and Statistics Directorate. PRAS 33 (1994)
11. Pentland, A., Moshadamm, B., Starber, T.: View-based and modular Eigenfaces for face recognition. In: Proceedings of the IEEE Conference on Computerised Vision and Pattern Recognition 1994, pp. 84–91 (1993)
12. Wiskott, L., Fellous, J., Kruger, N., von der Malsburg, C.: Face recognition and gender determination. In: Proceedings of the International Workshop on Automatic Face and Gesture Recognition. Zurich (1995)
13. Phillips, P., Wechsler, H., Huang, J., Rauss, P.: The FERET database and evaluation procedure for face recognition algorithms. *Image Vis. Comput.* **J.** **16**, 295–306 (1998)
14. Phillips, P., Moon, H., Rizvi, S., Rauss, P.: The FERET evaluation. In: Face Recognition from Theory to Applications. Springer, Berlin (1997)
15. Thomas, R.: As UK crime outstrips the US, a hidden eye is watching: Police switch on a camera that recognizes your face. *The Observer*, 11 October 1998, p. 5
16. Wilkinson, C.: Forensic Facial Reconstruction, Cambridge University Press, Cambridge (2004)
17. Auslebrooke, W.A., Becker, P.J., Iscan, M.Y.: Facial Soft Tissue Thickness in the Adult Male Zulu. *Forensic Sci. Int.* **79**, 83–102 (1996)
18. Lebedinskaya, G.U., Balueva, T.S., Veselovskaya, E.B.: Development of Methodological Principles for Reconstruction of the Face on the Basis of Skull Material, in Forensic Analysis of the Skull. pp. 183–198 Wiley-Liss, New York, (1993)
19. Wilkinson, C.M., Whittaker, D.K.: Skull Reassembly and the Implications for Forensic Facial Reconstruction. *Sci. Justice* **41**(3), 5–6 (2002)
20. Vanezis, P., Blowes, R.W., Linney, A.D., Tan, A.C., Richards, R., Neave, R.: Application of 3-D computer graphics for facial reconstruction and comparison with sculpting techniques. *Forensic Sci. Int.* **42**, 69–84 (1989)

Facial Action Coding

- Facial Expression Recognition

Facial Changes

- Face Variation

Facial Expression Analysis

- Facial Expression Recognition

Facial Expression Recognition

MAJA PANTIC

Department of Computing Imperial College London, London, UK

Synonyms

Facial Expression Analysis; Facial Action Coding

Definition

Facial expression recognition is a process performed by humans or computers, which consists of:

1. Locating faces in the scene (e.g., in an image; this step is also referred to as *face detection*),
2. Extracting facial features from the detected face region (e.g., detecting the shape of facial components or describing the texture of the skin in a facial area; this step is referred to as *facial feature extraction*),
3. Analyzing the motion of facial features and/or the changes in the appearance of facial features and classifying this information into some facial-expression-interpretative categories such as facial muscle activations like smile or frown, emotion (affect) categories like happiness or anger, attitude categories like (dis)liking or ambivalence, etc. (this step is also referred to as *facial expression interpretation*).

Introduction

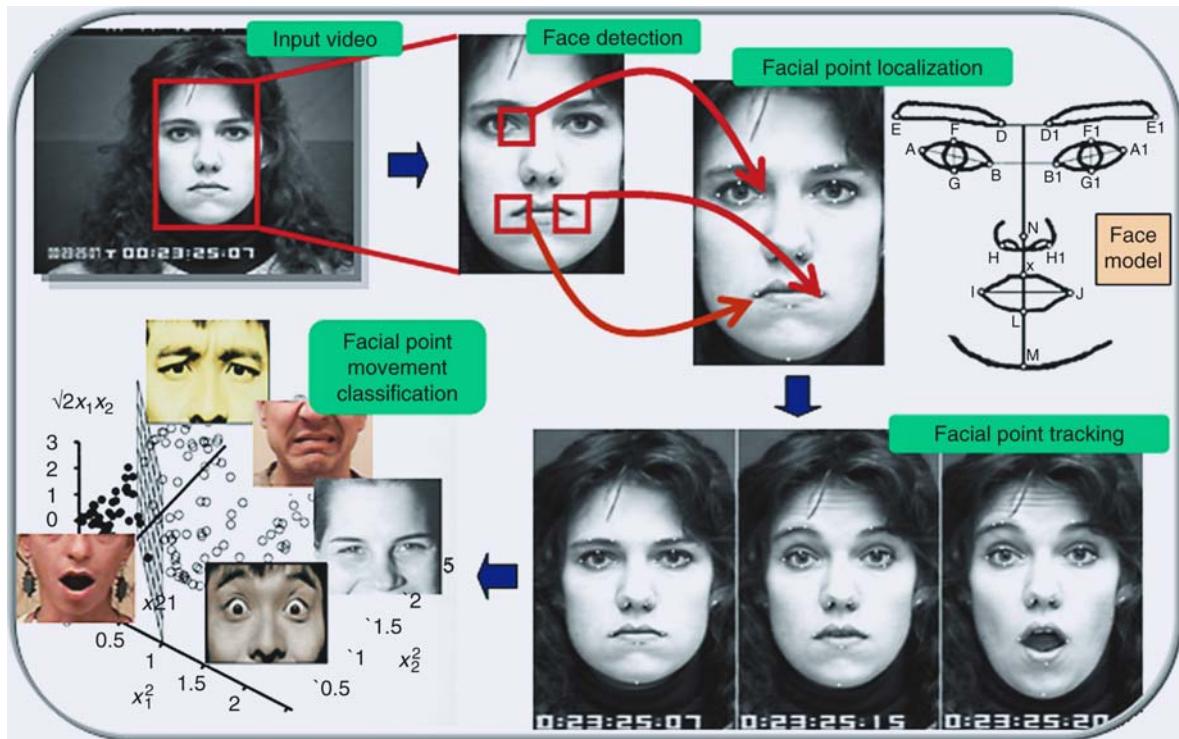
A widely accepted prediction is that computing will move to the background, weaving itself into the fabric of our everyday living and projecting the human user into the foreground. To realize this goal, next-generation computing (a.k.a. pervasive computing,

ambient intelligence, and ► **human computing**) will need to develop human-centered ► **user interfaces** that respond readily to naturally occurring, multimodal, human communication [1]. These interfaces will need the capacity to perceive and understand intentions and emotions as communicated by social and affective signals. Motivated by this vision of the future, automated analysis of nonverbal behavior, and especially of facial behavior, has attracted increasing attention in computer vision, pattern recognition, and human-computer interaction [2–5]. To wit, facial expression is one of the most cogent, naturally preeminent means for human beings to communicate emotions, to clarify and stress what is said, to signal comprehension, disagreement, and intentions, in brief, to regulate interactions with the environment and other persons in the vicinity [6, 7]. Automatic analysis of facial expressions forms, therefore, the essence of numerous next-generation-computing tools including ► **affective computing** technologies (proactive and affective user interfaces), learner-adaptive tutoring systems, patient-profiled personal wellness technologies, etc.

The Process of Automatic Facial Expression Recognition

The problem of machine recognition of human facial expression includes three subproblem areas (Fig. 1): (1) finding faces in the scene, (2) extracting facial features from the detected face region, (3) analyzing the motion of facial features and/or the changes in the appearance of facial features, and classifying this information into some facial-expression-interpretative categories (e.g., emotions, facial muscle actions, etc.).

The problem of *finding faces* can be viewed as a segmentation problem (in machine vision) or as a detection problem (in pattern recognition). It refers to identification of all regions in the scene that contain a human face. The problem of finding faces (*face localization*, *face detection*) should be solved regardless of clutter, occlusions, and variations in head pose and lighting conditions. The presence of non-rigid movements due to facial expression and a high degree of variability in facial size, color and texture make this problem even more difficult. Numerous techniques have been developed for face detection in still images [8, 9],



Facial Expression Recognition. **Figure 1** Outline of an automated, geometric-features-based system for facial expression recognition (for details of this system, see [4]).

(see ▶ [Face Localization](#)). However, most of them can detect only upright faces in frontal or near-frontal view. Arguably the most commonly employed face detector in automatic facial expression analysis is the real-time face detector proposed by Viola and Jones [10].

The problem of feature extraction can be viewed as a dimensionality reduction problem (in machine vision and pattern recognition). It refers to transforming the input data into a reduced representation set of features which encode the relevant information from the input data. The problem of *facial feature extraction* from input images may be divided into at least three dimensions [2, 4]: (1) Are the features holistic (spanning the whole face) or analytic (spanning subparts of the face)?; (2) Is temporal information used?; (3) Are the features view- or volume based (2-D/3-D)? Given this glossary, most of the proposed approaches to facial expression recognition are directed toward static, analytic, 2-D facial feature extraction [3, 4]. The usually extracted facial features are either *geometric features* such as the shapes of the facial components (eyes, mouth, etc.) and the locations of facial fiducial points (corners of the eyes, mouth, etc.), or *appearance features* representing the texture of the facial skin in specific facial areas including wrinkles, bulges, and furrows. Appearance-based features include learned image filters from Independent Component Analysis (ICA), Principal Component Analysis (PCA), Local Feature Analysis (LFA), Gabor filters, integral image filters (also known as box-filters and Haar-like filters), features based on edge-oriented histograms, etc. (see ▶ [Skin Texture](#), and ▶ [Feature Extraction](#)). Several efforts have also been reported which use both geometric and appearance features (e.g., [3]). These approaches to automatic facial expression analysis are referred to as *hybrid* methods. Although it has been reported that methods based on geometric features are often outperformed by those based on appearance features using, e.g., Gabor wavelets or eigenfaces, recent studies show that in some cases geometric features can outperform the appearance-based ones [4, 11]. Yet, it seems that using both geometric and appearance features might be the best choice in the case of certain facial expressions [11].

Contractions of facial muscles, which produce facial expressions, induce movements of the facial skin and changes in the location and/or appearance of facial features (e.g., contraction of the Corrugator muscle induces a frown and causes the eyebrows to move towards each other, usually producing wrinkles between



Facial Expression Recognition. **Figure 2** Facial appearance of the Corrugator muscle contraction (coded as in the FACS system, [14]).

the eyebrows; Fig. 2). Such *changes can be detected* by analyzing optical flow, facial-point- or facial-component-contour-tracking results, or by using an ensemble of classifiers trained to make decisions about the presence of certain changes (e.g., whether the nasolabial furrow is deepened or not) based on the passed appearance features. The optical flow approach to describing face motion has the advantage of not requiring a facial feature extraction stage of processing. Dense flow information is available throughout the entire facial area, regardless of the existence of facial components, even in the areas of smooth texture such as the cheeks and the forehead. Because optical flow is the visible result of movement and is expressed in terms of velocity, it can be used to represent directly the facial expressions. Many researchers adopted this approach [2, 3]. Until recently, standard optical flow techniques were, arguably, most commonly used for tracking facial characteristic points and contours as well [4]. In order to address the limitations inherent in optical flow techniques such as the accumulation of error and the sensitivity to noise, occlusion, clutter, and changes in illumination, recent efforts in automatic facial expression recognition use sequential state estimation techniques (such as Kalman filter and Particle filter) to track facial feature points in image sequences (e.g., [4, 11]).

Eventually, dense flow information, tracked movements of facial characteristic points, tracked changes in contours of facial components, and/or extracted

appearance features are translated into a description of the displayed facial expression. This description (*facial expression interpretation*) is usually given either in terms of shown affective states (emotions) or in terms of activated facial muscles underlying the displayed facial expression. This stems directly from two major approaches to facial expression measurement in psychological research [12]: message and sign judgment. The aim of message judgment is to infer what underlies a displayed facial expression, such as affect or personality, while the aim of sign judgment is to describe the “surface” of the shown behavior, such as facial movement or facial component shape. Thus, a brow frown can be judged as “anger” in a message-judgment and as a facial movement that lowers and pulls the eyebrows closer together in a sign-judgment approach. While message judgment is all about interpretation, sign judgment attempts to be objective, leaving inference about the conveyed message to higher order decision making. Most commonly used facial

expression descriptors in message judgment approaches are the six basic emotions (fear, sadness, happiness, anger, disgust, surprise; see Fig. 3) proposed by Ekman and discrete emotion theorists [13], who suggest that these emotions are universally displayed and recognized from facial expressions. Most commonly used facial action descriptors in sign judgment approaches are the Action Units (AUs) defined in the Facial Action Coding System (FACS; [14]). Most facial expressions analyzers developed, so far, target human facial affect analysis and attempt to recognize a small set of prototypic emotional facial expressions like happiness and anger [2, 5]. However, several promising prototype systems were reported that can recognize deliberately produced AUs in face images and even few attempts towards recognition of spontaneously displayed AUs have been recently reported as well [3–5]. While the older methods employ simple approaches including expert rules and machine learning methods such as neural networks to classify the relevant information



Facial Expression Recognition. **Figure 3** Prototypic facial expressions of six basic emotions (left-to-right from top row): disgust, happiness, sadness, anger, fear, and surprise.

from the input data into some facial-expression-interpretative categories, the more recent (and often more advanced) methods employ probabilistic, statistical, and ensemble learning techniques, which seem to be particularly suitable for automatic facial expression recognition from face image sequences [3, 5].

Evaluating Performance of an Automated System for Facial Expression Recognition

The two crucial aspects of evaluating performance of a designed automatic facial expression recognizer are the utilized training/test dataset and the adopted evaluation strategy.

Having enough labeled data of the target human facial behavior is a prerequisite in designing robust automatic facial expression recognizers. Explorations of this issue showed that, given accurate 3-D alignment of the face (see ▶ [Face Alignment](#)), at least 50 training examples are needed for moderate performance (in the 80% accuracy range) of a machine-learning approach to recognition of a specific facial expression [4]. Recordings of spontaneous facial behavior are difficult to collect because they are difficult to elicit, short lived, and filled with subtle context-based changes. In addition, manual labeling of spontaneous facial behavior for ground truth is very time consuming, error prone, and expensive. Due to these difficulties, most of the existing studies on automatic facial expression recognition are based on the “artificial” material of deliberately displayed facial behavior, elicited by asking the subjects to perform a series of facial expressions in front of a camera. Most commonly used, publicly available, annotated datasets of posed facial expressions include the Cohn-Kanade facial expression database, JAFFE database, and MMI facial expression database [4, 15]. Yet, increasing evidence suggests that deliberate (posed) behavior differs in appearance and timing from that which occurs in daily life. For example, posed smiles have larger amplitude, more brief duration, and faster onset and offset velocity than many types of naturally occurring smiles. It is not surprising, therefore, that approaches that have been trained on deliberate and often exaggerated behaviors usually fail to generalize to the complexity of expressive behavior found in real-world settings. To address the general lack of a reference set of (audio and/or) visual recordings of human spontaneous behavior, several efforts aimed at

development of such datasets have been recently reported. Most commonly used, publicly available, annotated datasets of spontaneous human behavior recordings include SAL dataset, UT Dallas database, and MMI-Part2 database [4, 5].

In pattern recognition and machine learning, a common evaluation strategy is to consider correct classification rate (*classification accuracy*) or its complement error rate. However, this assumes that the natural distribution (prior probabilities) of each class are known and balanced. In an imbalanced setting, where the prior probability of the positive class is significantly less than the negative class (the ratio of these being defined as the *skew*), accuracy is inadequate as a performance measure since it becomes biased towards the majority class. That is, as the skew increases, accuracy tends towards majority class performance, effectively ignoring the recognition capability with respect to the minority class. This is a very common (if not the default) situation in facial expression recognition setting, where the prior probability of each target class (a certain facial expression) is significantly less than the negative class (all other facial expressions). Thus, when evaluating performance of an automatic facial expression recognizer, other performance measures such as *precision* (this indicates the probability of correctly detecting a positive test sample and it is independent of class priors), *recall* (this indicates the fraction of the positives detected that are actually correct and, as it combines results from both positive and negative samples, it is class prior dependent), *F1-measure* (this is calculated as $2 * \text{recall} * \text{precision} / (\text{recall} + \text{precision})$), and ROC (this is calculated as $P(x|\text{positive}) / P(x|\text{negative})$, where $P(x|C)$ denotes the conditional probability that a data entry has the class label C , and where a ROC curve plots the classification results from the most positive to the most negative classification) are more appropriate. However, as a confusion matrix shows all of the information about a classifier’s performance, it should be used whenever possible for presenting the performance of the evaluated facial expression recognizer.

Applications

The potential benefits from efforts to automate the analysis of facial expressions are varied and numerous and span fields as diverse as cognitive sciences, medicine, communication, education, and security [16].

When it comes to computer science and computing technologies, facial expressions provide a way to communicate basic information about needs and demands to the machine. Where the user is looking (i.e., gaze tracking) can be effectively used to free computer users from the classic keyboard and mouse. Also, certain facial signals (e.g., a wink) can be associated with certain commands (e.g., a mouse click) offering an alternative to traditional keyboard and mouse commands. The human capability to “hear” in noisy environments by means of lip reading is the basis for bimodal (audiovisual) speech processing (see Lip-Movement Recognition), which can lead to the realization of robust speech-driven *user interfaces*. To make a believable *talking head* (avatar) representing a real person, recognizing the person’s facial signals and making the avatar respond to those using synthesized speech and facial expressions is important. Combining facial expression spotting with facial expression interpretation in terms of labels like “did not understand”, “disagree”, “inattentive”, and “approves” could be employed as a tool for monitoring human reactions during videoconferences, web-based lectures, and automated tutoring sessions. The focus of the relatively, recently initiated research area of *affective computing* lies on sensing, detecting and interpreting human affective states (such as pleased, irritated, confused, etc.) and devising appropriate means for handling this affective information in order to enhance current ► **HCI** designs. The tacit assumption is that in many situations human-machine interaction could be improved by the introduction of machines that can adapt to their users and how they feel. As facial expressions are our direct, naturally preeminent means of communicating emotions, machine analysis of facial expressions forms an indispensable part of affective HCI designs.

Monitoring and interpreting facial expressions can also provide important information to lawyers, police, security, and intelligence agents regarding *person’s identity* (research in psychology suggests that facial expression recognition is much easier in familiar persons because it seems that people display the same, “typical” patterns of facial behaviour in the same situations), *deception* (relevant studies in psychology suggest that visual features of facial expression function as cues to deception), and *attitude* (research in psychology indicates that social signals including accord and mirroring – mimicry of facial expressions, postures, etc., of one’s interaction partner – are typical, usually unconscious gestures of wanting to get along with and

be liked by the interaction partner). Automated facial reaction monitoring could form a valuable tool in law enforcement, as now only informal interpretations are typically used. Systems that can recognize friendly faces or, more importantly, recognize unfriendly or aggressive faces and inform the appropriate authorities represent another application of facial measurement technology.

Concluding Remark

Faces are tangible projector panels of the mechanisms which govern our emotional and social behaviors. The automation of the entire process of facial expression recognition is, therefore, a highly intriguing problem, the solution to which would be enormously beneficial for fields as diverse as medicine, law, communication, education, and computing. Although the research in the field has seen a lot of progress in the past few years, several issues remain unresolved. Arguably the most important unattended aspect of the problem is how the grammar of facial behavior can be learned (in a human-centered, context-profiled manner) and how this information can be properly represented and used to handle ambiguities in the observation data. This aspect of machine analysis of facial expressions forms the main focus of the current and future research in the field.

Related Entries

- Face Alignment
- Face Localization
- Feature Extraction
- Lip Movement Recognition
- Skin Texture

References

1. Pantic, M., Pentland, A., Nijholt, A., Huang, T.S.: Human computing and machine understanding of human behavior: A Survey. *Lect. Notes Artif. Intell.* **4451**, 47–71 (2007)
2. Pantic, M., Rothkrantz, L.J.M.: Toward an affect-sensitive multimodal HCI. *Proceedings of the IEEE* **91**(9), 1370–1390 (2003)
3. Tian, Y.L., Kanade, T., Cohn, J.F.: Facial expression analysis. In: Li, S.Z., Jain, A.K. (eds.) *Handbook of Face Recognition*, pp. 247–276. Springer, New York (2005)

4. Pantic, M., Bartlett, M.S.: Machine analysis of facial expressions. In: Delac, K., Grgic, M. (eds.) *1Face Recognition*, pp. 377–416. I-Tech Education and Publishing, Vienna, Austria (2007)
5. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(1), 39–58 (2009)
6. Ambady, N., Rosenthal, R.: Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychol. Bull.* **111**(2), 256–274 (1992)
7. Ekman, P., Rosenberg, E.L. (eds.): *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system*. Oxford University Press, Oxford, UK (2005)
8. Yang, M.H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(1), 34–58 (2002)
9. Li, S.Z., Jain, A.K. (eds.): *Handbook of face recognition*. Springer, New York (2005)
10. Viola, P., Jones, M.: Robust real-time face detection. *Int. J. Comput. Vis.* **57**(2), 137–154 (2004)
11. Pantic, M., Patras, I.: Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Trans. Syst. Man Cybern. B Cybern.* **36**(2), 433–449 (2006)
12. Cohn, J.F., Ekman, P.: Measuring facial actions. In: Harrigan, J.A., Rosenthal, R., Scherer, K. (eds.) *The New Handbook of Methods in Nonverbal Behavior Research*, pp. 9–64. Oxford University Press, New York (2005)
13. Keltnner, D., Ekman, P.: Facial expression of emotion. In: Lewis, M., Haviland-Jones, J.M. (eds.) *Handbook of Emotions*, pp. 236–249. Guilford Press, New York (2000)
14. Ekman, P., Friesen, W.V., Hager, J.C.: *Facial action coding system*. A Human Face, Salt Lake City, USA (2002)
15. Pantic, M., Valstar, M.F., Rademaker, R., Maat, L.: Web-based database for facial expression analysis. Proc. IEEE Int'l Conf. Multimedia & Expo (ICME) 317–321 (2005)
16. Ekman, P., Huang, T.S., Sejnowski, T.J., Hager, J.C. (eds.): *NSF Understanding the Face. A Human Face eStore*, Salt Lake City, USA, (see Library) (1992)

Facial Landmarks

A number of pixels in a face image clearly corresponds to some extract physiological semantics, such as the eye corners, eye centers, mouth corners, nose tips, etc. These feature points are called facial landmarks. They are generally used to align different face images for accurate matching.

► Face Misalignment Problem

Facial Mapping

Facial mapping is a frequently used term to describe one-to-one matching of crime scene and suspect images undertaken by an expert. A number of different methods may be used in combination to compare two images to support the comparison. Also, a number of comparisons may be made of the face from different angles if multiple images are available from each source.

► Face, Forensic Evidence of

Facial Motion Estimation

► Face Tracking

Facial Photograph

► Photography for Face Image Data

Factor Analysis

► Session Effects on Speaker Modeling

Failure to Acquire Rate

Both the acquiring conditions and the flaw of biometric itself may cause failure to acquire a biometric trait. The percentage of this failure is defined as “Failure to Acquire Rate.” For instance, very low quality face image may cause the failure of face detection and subsequent feature extraction.

► Evaluation of Biometric Quality Measures
► Performance Evaluation, Overview

Failure-to-Enrol Rate

Failure-to-enrol rate is defined as the proportion of enrollment transactions in which zero instances were enrolled. Enrollment in one or more instances is considered to be successful in the case, the systems accept multiple biometric samples per person.

- ▶ Finger Vein Reader

Fake Finger Detection

- ▶ Anti-spoofing
- ▶ Fingerprint Fake Detection

False Match Rate

The probability that a biometric system will indicate that two biometric templates match although they are not derived from the same individual and should not match.

- ▶ Fingerprint Image Quality
- ▶ Iris on the Move

False Negative Rate

False Negative Rate means that how many percentages of the authentic test samples are incorrectly classified as the imposter class. Take the example of the computer account login system, False Negative Rate means how many percentages of legal users are recognized as illegal users. As one can see immediately, False Positive Rate and False Negative Rate are two metrics that counter each other. For any given biometrics modality with given matching algorithm, requirement of low False Positive Rate would unavoidably bring high False Negative Rate, and vice versa. Performance

comparison between different algorithms is usually done by comparing False Negative Rate at a fixed False Positive Rate.

- ▶ Biometric System Design, Overview
- ▶ Iris Recognition, Overview

False Non-Match Rate

False non-match rate is the proportion of genuine comparisons that result in false non-match. False non-match is the decision of non-match when comparing biometric samples that are from same biometric source (i.e., genuine comparison).

- ▶ Biometric System Design, Overview
- ▶ Fingerprint Image Quality
- ▶ Iris on the Move

False Positive Rate

False Positive Rate means how many percentage of the imposter test samples are incorrectly classified as the authentic class. For example, in a computer account login system, False Positive Rate is what percentage of the illegal users recognized as legal users. In applications, which require high security, False Positive Rate is always required to be as small as possible.

- ▶ Biometric System Design, Overview
- ▶ Iris Recognition, Overview

Feathering

Feathering is a feature which occurs on the outsole as a result of an abrasive wear and has some resemblance to the ridge characteristics and bifurcations of fingerprint patterns. It is the result of frictional abrasive forces applied to the outsole surface such as when scuffing

or dragging the shoe. This feature is also known as a Schallamach pattern.

► Footwear Recognition

Feature Detection

Finding significant features in images such as landmarks, edges, or curves. For example, a facial feature detector aims to find the positions of the center of an eye, the corners of a mouth, or the top of a nose in a face image. In the case of an iris image, features may mean the edges inside an iris or the boundaries around the iris.

► Iris Segmentation Using Active Contours

Feature Extraction

► Biometric Algorithms

Feature Fusion

Producing a merged feature vector from a set of feature vectors representing different aspects of biometric data. The data can originate from different sensors, and also from different properties of a signal that originate from the same sensor.

- Fusion, Feature-Level
- Multiple Experts

Feature Map

The image produced from a target image to enhance the signals of a particular type, such as edges,

ridges, or valleys is referred as a “feature map.” Face alignment programs typically rely on statistics computed from such features to distinguish facial features from other regions of the image. More sophisticated feature maps can be constructed to capture complicated local image structures and enhance the stability.

► Face Alignment

Feature Selection

Feature selection techniques are aimed towards finding an optimal feature set for a specific purpose, such as the optimization of a biometric system verification performance. In general, feature selection algorithms try to avoid the evaluation of all the possible feature combinations when searching for an optimal feature vector, since these grow exponentially as the number of feature increases.

► Signature Features

Feature Vector

Feature vector is a multidimensional vector that is obtained from a face by using feature extraction and image processing techniques to be used and that is used to memorize and recognize the face.

► Face Databases and Evaluation

Features

Biometric features are the information extracted from biometric samples which can be used for comparison with a biometric reference. For example, characteristic measures extracted from a face photograph such as eye distance or nose size etc. The aim of the extraction of biometric features from a biometric sample is to

remove superfluous information which does not contribute to biometric recognition. This enables a fast comparison and an improved biometric performance, and may have privacy advantages.

- ▶ [Biometric Algorithms](#)
- ▶ [Vascular Image Data Format, Standardization](#)

Features vs. Templates

- ▶ [Face Recognition, Geometric vs. Appearance-Based](#)

Fidelity

The degree of similarity between a biometric sample and its source. Fidelity of a sample is comprised of individual components of fidelity attributed to each step through which it is processed (e.g., compression).

- ▶ [Biometric Sample Quality](#)
- ▶ [National Institute for Standards and Technology](#)

Field of View (FOV)

Field of view (FOV) is the angular portion of visible space which is comprised into the image region. The FOV of the human eye is around 150°. The camera FOV depends both on the size of the camera sensor and the geometry of the lens. The camera focal length determines the field of view falling within the sensor area, thus determining also the magnification factor of the image. A shorter camera focal length produces a wider FOV, while a longer focal length produces a smaller FOV.

- ▶ [Face Device](#)

Finger Data Interchange Format, Standardization

RAUL SANCHEZ-REILLO¹, ROBERT MUELLER²

¹University Carlos III of Madrid, Avda. Universidad, Leganes (Madrid), Spain

²Giesecke & Devrient GmbH, Prinzregentenstr. Muenchen, Germany

Synonyms

Encoded finger data; Fingerprint data interchange format

Definition

Set of ISO Standards that define common formats to encode information related to finger-based biometrics. Those formats are defined to allow interoperability among different vendors worldwide, and have been developed by the international community taking part in ISO/IEC JTC1/SC37 standardization subcommittee. Those documents define not only the way a fingerprint image has to be encoded, but also the way a feature vector composed of ▶ [minutiae](#) points has to be stored and/or transmitted. Furthermore, formats for the ▶ [spectral data](#) of the finger, as well as its skeletal data are defined.

Introduction

Standardization is essential for the wide-spread adoption of technologies in open mass applications. Fingerprint recognition is not only the most prominent biometric measure, but also the biometric trait with the largest databases and the best long-term experience. Fingerprints are used in applications such as physical access control and digital signature creation but also national ID card schemes and other governmental projects. The need for standardization is conspicuous in every single area where it is not applied.

The SC37 Subcommittee from ISO/IEC JTC1 deals with the standardization of biometrics. Among the many aspects of its work, SC37's Working Group 3 is devoted to defining Interchange Data Formats for a variety of biometric modalities. To accomplish this,

a multipart standard is under development, covering several biometric modalities. Such multipart standard is known as ISO/IEC 19794. There are four parts in this standard which cover finger-based biometrics, or what can be better understood as fingerprint biometrics.

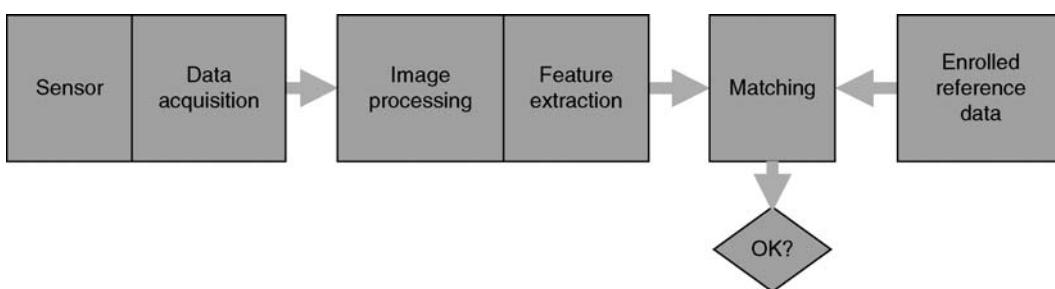
1. Part 2 of the Standard series, deals with the way a minutiae-based feature vector or template has to be coded
2. Part 3 standardizes the way to code information referring to the spectral information of the fingerprint
3. Part 4 determines the coding of a fingerprint raw image and
4. Part 8 establishes a way to code a fingerprint by its skeleton

[Figure 1](#) shows the basic architecture of a typical fingerprint verification system. A finger is presented to a sensor and a raw image acquired. Image processing techniques enhance the image quality before a feature vector of characteristic features can be extracted. The features are compared with a previously recorded reference data set to determine the similarity between the two sets before the user presenting the finger is authenticated. The reference data is stored in a database or on a portable data carrier.

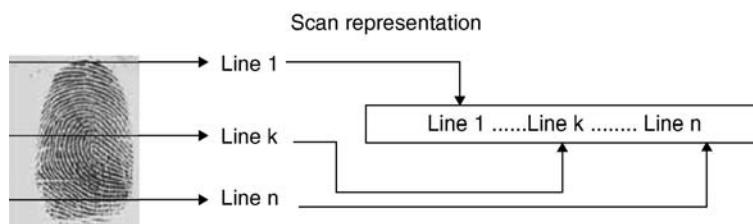
The following subsections explain the basic characteristics of each type of finger-based standard. The image standard (Part 4) is presented first as it is the first step in the fingerprint comparison process as shown in the architecture above. This is followed by the other finger based standards, each of which deals with samples already processed.

Finger Images

As already mentioned, the way a fingerprint image is to be coded is defined in ISO/IEC 19794-4 International Standard [1], whose title is “Information technology - Biometric data interchange formats - Part 4: Finger image data.” The way the finger is scanned is out of the scope of the standard, but after image acquisition, the image shall represent a finger in upright position, i.e., vertical and with the tip of the finger in the upper part of the image. The way to code such an image is represented in [Fig. 2](#), where the top line is the first to be stored and/or transmitted. This is in contradiction to mathematical graphing practice but in conjunction with typical digital image processing. For those images that require two or more bytes per pixel intensity, the most significant byte is stored/transmitted first, and bytes follow most significant bit coding.



Finger Data Interchange Format, Standardization. [Figure 1](#) Typical Biometric Verification System.



Finger Data Interchange Format, Standardization. [Figure 2](#) Coding structure of a fingerprint image. Image taken from [1].

This International Standard also includes a set of constraints for image acquisition. It determines the pixel aspect ratio, which shall be between 0.99 and 1.01 (horizontal/vertical sizes), as well as several image acquisition levels, as stated in [Table 1](#).

After the requirements for the image to be stored or transmitted have been specified, this International Standard details the structure of the data record referring to a finger image. Following CBEFF specifications [\[2\]](#) (see entry “Common Biometric Exchange Framework Formats”), a record referring to a finger image has the following structure (for details refer to the last version of this International Standard [\[1\]](#)):

- A single fixed-length (32-byte) general record header containing information about the overall record, with the following fields:
 - Format identifier (4 bytes with the hexadecimal value 0x46495200) and version number (coded in another 4 bytes)
 - Record length (in bytes) including all finger images within that record (coded in 6 bytes)
 - Capture device ID (2 bytes) and Image acquisition level (2 bytes)
 - Number of fingers (1 byte), Scale units used (1 byte), and Scan resolution used (2 bytes for horizontal and another 2 for vertical resolution)
 - Image resolution, coded the same way as the scan resolution, and whose value shall be less or equal to scan resolution
 - Pixel depth (1 byte) and Image compression algorithm used (coded in 1 byte)
 - 2 bytes reserved for future use

Finger Data Interchange Format, Standardization.

Table 1 Image acquisition levels for finger biometrics.

Extract from [Table 1](#) in [\[1\]](#)

Setting level	Scan resolution (dpi)	Pixel depth (bits)	Gray levels
10	125	1	2
20	250	3	5
30	500	8	80
31	500	8	200
35	750	8	100
40	1,000	8	120
41	1,000	8	200

- A single finger record for each finger, view, multi-finger image, or palm consisting of:
 - A fixed-length (14-byte) finger header containing information pertaining to the data for a single or multi-finger image, which gives information about:
 - Length of the finger data block (4 bytes)
 - Finger/palm position (1 byte)
 - Count of views (1 byte) and View number (1 byte)
 - Finger/palm image quality (1 byte) and Impression type (1 byte)
 - Number of pixels per horizontal line (2 bytes) and Number of horizontal lines (2 bytes)
 - 1 byte reserved for future use
 - Compressed or uncompressed image data view for a single, multi-finger, or palm image, which has to be smaller than 43×10^8 bytes.

The raw finger format is used, for example, in databases containing standard fingerprints. Law enforcement agencies are typical applicants of the standard. The largest fingerprint image databases are maintained by the FBI in the United States and are encoded with a national counterpart of this standard.

Fingerprint Minutiae

While Part 4 of the 19794 Series of Standards is dedicated to raw biometric sample data, Part 2 refers to the format in which a minutiae-based feature vector or template has to be coded. Therefore ISO/IEC 19794-2 “Information Technology - Biometric data interchange Formats - Part 2: Finger minutiae data” [\[3\]](#) deals with processed biometric data, ready to be sent to a comparison block to obtain a matching score.

Finger minutiae are local point patterns present in a fingerprint image. The comparison of these characteristic features is sufficient to positively identify a person. Sir Francis Galton first defined the features of a fingerprint [\[4\]](#).

In order to reach interoperability, this International Standard defines not only the record format, but also the rules for fingerprint minutiae extraction. Regarding record formats, due to the application of fingerprint biometrics to systems based on smart cards, compact record formats are also defined to cope with memory and transmission speed limitations of such devices.

Fingerprint scientists have defined more than 150 different types of minutiae [5]. Within this Standard, minutiae types are simplified to the following: (1) ridge ending, (2) ridge bifurcation, and (3) other. The location of each minutiae is determined by its horizontal and vertical position within the image. To determine such location a coordinate system is to be defined. [Figure 3](#) shows how such coordinate system is chosen. Granularity to be taken to determine location is of one hundredth of a millimetre for the normal format, while just one tenth of a millimetre for card compact formats.

[Figure 4](#) shows the different ways to consider the location of a minutiae. (1) represents a Ridge Ending, encoded as a Valley Skeleton Bifurcation Point, (2) shows how to locate a Ridge Bifurcation, encoded as a Ridge Skeleton Bifurcation Point, Finally (3) illustrates how to locate a Ridge Ending encoded as a Ridge Skeleton Endpoint. How to determine the encoding of ridge ending actually used in a specific

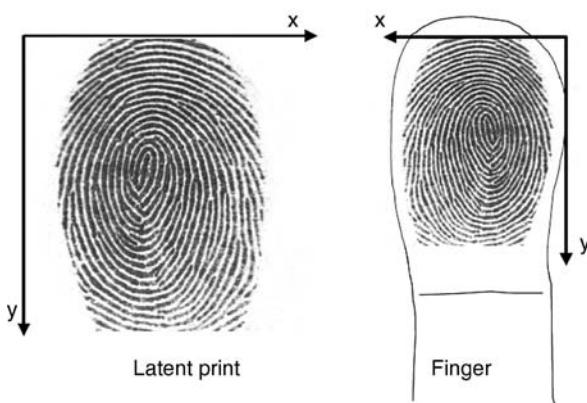
dataset is a subject currently under revision in the standard. The other types of minutia have to be coded consistent with the Standards (see details in [3]).

To define the minutiae direction, its angle has to be determined. This Standards specifies that the angle is obtained, increasing counter-clockwise rotation starting from the horizontal axis to the right of the location of the minutiae point. The angle is encoded in a unsigned single byte, so the granularity is 1.40625° per bit ($360/256$). [Figure 4](#) also illustrates how the angle is determined.

Additional information that may be included in a minutiae-based record are cores, deltas, and ridge crossings to neighboring minutiae.

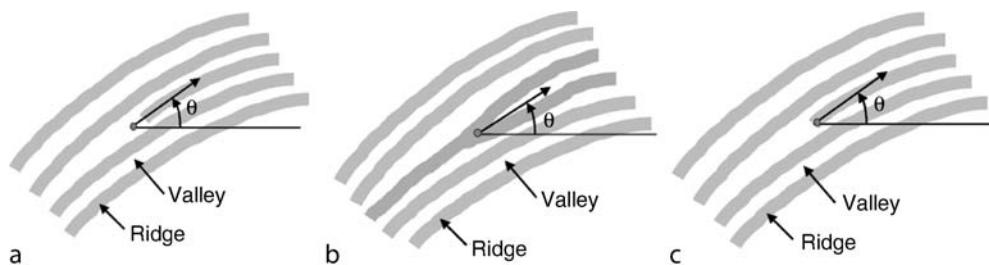
With all these definitions, the two major format types defined by this International Standard are: (1) record format, and (2) card format. The structure of the record format is summarized in the following paragraphs and for additional details refer to the standard [3].

- A fixed-length (24-byte) record header containing information about the overall record, including the number of fingers represented and the overall record length in bytes:
 - Format identifier (4 bytes with the hexadecimal value `0x464D5200`) and Version number (coded in another 4 bytes)
 - Record length (in bytes) including all finger images within that record (coded in 4 bytes)
 - Capture device ID (2 bytes)
 - Size of the image in pixels (2 bytes for X dimension, and 2 bytes for Y dimension)
 - Image resolution in pixels per centimetre (2 bytes for X and 2 bytes for Y)
 - Number of finger views included in the record
 - 1 byte reserved for future use



Finger Data Interchange Format, Standardization.

[Figure 3](#) Coordinate System for Minutiae Location. Image taken from [3].



[Figure 4](#) Illustration of location of minutia. Image taken from [3].

- A Single Finger record for each finger/view, consisting of:
 - A fixed-length (4-byte) header containing information about the data for a single finger, including the number of minutiae:
 - Finger position (1 byte)
 - View number (4 bits) and Impression type (4 bits, to make a 1 byte in total)
 - Finger quality (1 byte)
 - Number of minutiae (1 byte)
 - A series of fixed-length(6-byte) minutiae descriptions:
 - Minutiae type (2 bits) and X location in pixels (14 bits)
 - 2 bits reserved and Y location in pixels (14 bits)
 - Minutiae angle (1 byte)
 - Quality of minutiae (1 byte)
 - One or more “extended” data areas for each finger/view, containing optional or vendor-specific information. It starts always with 2 bytes which determine the length of Extended Data Block. If this is 0x0000, no Extended Data is included. If it has a nonnull value, then it is followed by vendor-specific data which could include information about ridge count, cores and deltas, or cell information.

Regarding the card formats, the current version of the standard allows 2 sub-formats: (1) normal format (also referred as 5-byte minutiae), and (2) compact format (also known as 3-byte minutiae). The way minutia are coded in each format is described below.

- Card normal format (like the record format, but removing quality information):
 - Minutiae type (2 bits) and X location in pixels (14 bits)
 - 2 bits reserved and Y location in pixels (14 bits)
 - Minutiae angle (1 byte)
- Card compact format:
 - X coordinate (8 bits) considering a unit of 10^{-1}mm
 - Y coordinate (8 bits) considering a unit of 10^{-1}mm
 - Minutiae type (2 bits) using the same coding as with the card normal format
 - Angle (6 bits) having a granularity of $360/64$

Another important aspect related to card formats is that as they are intended to be used with devices with

limited memory and processing power, the number of minutiae may be restricted, and in such case, truncation is needed. Additionally in Match-on-Card systems, to reduce algorithm complexity, minutiae may need to be sorted in a certain way. And finally, the way data is exchanged differs from the traditional CBEFF format. This International Standard covers all such cases. The reader is suggested to refer to the last version of the Standard [3] for further details.

The minutiae standard is used e.g., by the ILO (International Labour Organization) in its seafarers identity card and in several national ID card implementations including Thailand and Spain [6].

Spectral Data of a Fingerprint

Part 3 of the 19794 series of standards deals with a format suitable to process fingerprints when using morphological approaches. But as seen in additional Fingerprint entries in this Encyclopedia, there are other approaches to perform biometric identification using fingerprints. Some of those approaches relate to the spectral information of the fingerprint. Algorithms using spectral data look at the global structure of a finger image rather than certain local point patterns. In such cases, 19794-2 is of no use and the only possibility would be to use the whole image as stated in 19794-4, which has the inconvenience of requiring the storage and/or transmission of a large amount of data. This could be inconvenient if not blocking for some applications.

In order to provide a new data format that could increase interoperability among spectral based solutions, reducing the amount of data to be used, 19794-3 has been developed under the title of “Information technology - Biometric data interchange formats - Part 3: Finger pattern spectral data” [7]. In fact, this International Standard deals with three major approaches in spectral based biometrics (wavelet based approaches are not supported by this standard).

1. Quantized co-sinusoidal triplets
2. Discrete Fourier transform
3. Gabor filters

After declaring the basic requirements for the original image in order to be considered for these algorithms (same coordinate system as in 19794-2, 255 levels of grey with 0 representing black and 255 being white, and

dark colours corresponding to ridges while light pixels corresponding to valleys), and describing all the above mentioned technologies, the Standards focuses on the record structure (for details refer to [7]), which is:

- A variable-length record header containing information about the overall record, including:
 - Format identifier (4 bytes with the hexadecimal value 0x46535000) and Version number (coded in another 4 bytes)
 - Record length (in bytes) including all fingers within that record (coded in 4 bytes)
 - Number of finger records included (1 byte)
 - Image resolution in pixels per centimetre (2 bytes for X direction and 2 bytes for Y direction)
 - Number of cells (2 bytes for X direction and 2 bytes for Y direction)
 - Number of pixels in cells (2 bytes for X direction and 2 bytes for Y direction)
 - Number of pixels between cells centres (2 bytes for X direction and 2 bytes for Y direction)
 - SCSM (Spectral component selection method - 1 byte), which can be 0, 1, or 2. Depending on the value of this field the following fields could refer to type of window, standard deviation, number of frequencies, frequencies, number of orientations and spectral components per cell, and bit-depths (propagation angle, wavelength, phase, and/or magnitude)
 - Bit-depth of quality score (1 byte)
 - Cell quality group granularity (1 byte)
 - 2 bytes reserved for future use
- A single finger record for each finger, consisting of:
 - A fixed-length (6-byte) header containing information about the data for a single finger:
 - Finger location (1 byte)
 - Impression type (1 byte)
 - Number of views in single finger record (1 byte)
 - Finger pattern quality (1 byte)
 - Length of finger pattern spectral data block (2 bytes)
 - A finger pattern spectral data block:
 - View number (1 byte)
 - Finger pattern spectral data
 - Cell quality data
 - An extended data block containing vendor-specific data, composed of block length (2 bytes), area type code (2 bytes), area length, and area.

As in 19794-2, this International Standard also defines the Data Objects to be included for a card format, with the reduction in granularity recommended (for further details see [7]).

Some of the leading fingerprint verification algorithms rely on spectral data or a combination of spectral data and minutiae. This standard could enhance the interoperability and performance of large scale identification systems such as criminal or civil Automatic Fingerprint Identification Systems (AFIS).

Skeletal Data of a Fingerprint

Finally 19794-8 titled “Information technology - Biometric data interchange formats - Part 8: Finger pattern skeletal data” [8] deals with the format for representing fingerprint images by a skeleton with ridges represented by a sequence of lines. Skeletonization is a standard procedure in image processing and generates a single pixel wide skeleton of a binary image. Moreover the start and endpoints of the skeleton ridge lines are included as real or virtual minutiae, and the line from start to endpoint is encoded by successive direction changes.

For minutiae location and coding, much of the 19794-2 card format is used, but here the position of a ridge bifurcation minutiae shall be defined as the point of forking of the skeleton of the ridge. In other words, the point where three or more ridges intersect is the location of the minutia. No valley representation is accepted under this International Standard. Another difference with 19794-2 card formats, is that in this Standard no other-type minutiae is considered (if a minutiae has more than three arms, like a trifurcation, it is considered a bifurcation), and that along this standard codes for “virtual minutiae” are used.

Skeleton lines are coded as polygons. Every line starts with a minutiae, and it is followed by a chain of direction changes (coded with the granularity stated in the record header), until it reaches the final minutiae. Several rules are defined in the standard (see [8] for further reference).

All that information is coded in a record with the following structure (limiting values as well as recommended values can be found in [8]):

- A fixed-length (24-byte) record header containing:
 - Format identifier (4 bytes with the hexadecimal value 0x46534B00) and Version number (coded in another 4 bytes)

- Record length (in bytes) including all finger images within that record (coded in 4 bytes)
- Capture device ID (2 bytes)
- Number of finger views in record (1 byte)
- Resolution of finger pattern in pixels per centimetre (1 byte)
- Bit depth of direction code start and stop point coordinates (1 byte)
- Bit depth of direction code start and stop direction (1 byte)
- Bit depth of direction in direction code (1 byte)
- Step size of direction code (1 byte)
- Relative perpendicular step size (1 byte)
- Number of directions on 180° (1 byte)
- 2 bytes reserved for future use
- A single finger record for each finger/view, consisting of:
 - A fixed-length (10 bytes) header:
 - View number (1 byte)
 - Finger position (1 byte)
 - Impression type (1 byte)
 - Finger quality (1 byte)
 - Skeleton image size in pixels (2 bytes for X-direction, 2 bytes for Y-direction)
 - Length of finger pattern skeletal data block (2 bytes)
 - The variable length fingerprint pattern skeletal description:
 - Length of finger pattern skeletal data (2 bytes)
 - Finger pattern skeletal data
 - Length of skeleton line neighbourhood index data (2 bytes)
 - Skeleton line neighbourhood index data
 - An extended data block containing the extended data block length and zero or more extended data areas for each finger/view, defining length (2 bytes), area type code (2 bytes), area length (2 bytes), and data.

This International Standard also defines two card formats, a normal one and a compact one. As with other parts, this means more limiting constraints to code data tighter and the definition of the Data Objects needed (for details refer to [8]).

The skeleton format is used in scientific research [9] and by vendors, implementing Match-on-Card.

Further Steps

The fingerprint parts of ISO 19794 were published as International Standards in 2005 and 2006. All the parts are currently under revision. A major task in the revision process is to address some defects and include a common header format for all the parts. Some references and vocabulary are needed to be updated to harmonize the relation of these standards within the ISO standardization landscape. The finger minutia standard ISO 19794-2 is probably the most prominent format in this series and is most frequently used by industry, government, and science. Interoperability tests have shown that the current standard allows some room for interpretation. This will be compensated by an amendment to describe the location, orientation, and type in more detail. Another aspect in the current revision of the standard is to reduce the number of format types from currently ten to a maximum of two. Experts from all continents and various backgrounds meet on a regular basis to lay down the future of the standards. The delegates take care of current requirements in terms of technology and applications.

Summary

To provide interoperability in storing and transmitting finger-related biometric information, four standards are already developed to define the formats needed for raw images, minutia-based feature vectors, spectral information, and skeletal representation of a fingerprint. Beyond that, other standards deal with conformance and quality control, as well as interfaces or performance evaluation and reporting (see related entries below for further information).

Related Entries

- ▶ [Biometric Data Interchange Format](#)
- ▶ [Common Biometric Exchange Framework Formats](#)
- ▶ [Conformance Testing for Biometric Data Interchange Formats, Standardization of](#)
- ▶ [Fingerprint Recognition](#)
- ▶ [International Standardization of Biometrics](#)

References

1. ISO/IEC: 19794-4:2005 - information technology - biometric data interchange formats - part 4: Finger image data (2005)
2. ISO/IEC: 19785-1:2005 - information technology - common biometric exchange formats framework - part 1: Data element specification (2005)
3. ISO/IEC: 19794-2:2005 - information technology - biometric data interchange formats - part 2: Finger minutiae data (2005)
4. Galton, F.: Finger Prints. Macmillan, London (Reprint: Da Capo, New York, 1965) (1892)
5. Moenssens, A.: Fingerprint Techniques. Chilton Book Company, London (1971)
6. Spanish-Homeland-Ministry: Spanish national electronic identity card information portal (in spanish). <http://www.dnielectronico.es/> (2007)
7. ISO/IEC: 19794-3:2006 - information technology - biometric data interchange formats - part 3: Finger pattern spectral data (2006)
8. ISO/IEC: 19794-8:2006 - information technology - biometric data interchange formats - part 8: Finger pattern skeletal data (2006)
9. Robert Mueller, U.M.: Decision level fusion in standardized fingerprint match-on-card. In: 1-4244-0342-1/06, ICARCV 2006, Hanoi, Vietnam (2006)

Finger Geometry, 3D

SOTIRIS MALASSIOTIS

Informatics and Telematics Institute, Center for Research and Technology Hellas, Thessaloniki, Greece

Synonym

3D hand biometrics

Definition

Biometrics based on 3D finger geometry exploit discriminatory information provided by the 3D structure of the hand, and more specifically the fingers, as captured by a 3D sensor. The advantages of current 3D finger biometrics over traditional 2D hand geometry authentication techniques are improved accuracy, the ability to work in contact free mode, and the ability to combine with 3D face recognition using the same sensor.

Introduction

The motivation behind 3D finger geometry biometrics is the same as with 3D face recognition. The 3D geometry of the hand as captured by a 3D sensor offers additional discriminatory information while being invariant to variations such as illumination or pigment of the skin, compared with an image captured with a plain 2D camera. The current accuracy and resolution of 3D sensors are not adequate for capturing fine details on the surface of the fingers such as skin wrinkles over the knuckles, but is sufficient to measure, local curvature, finger circumference, or finger length.

Another motivation comes from a limitation of current hand geometry recognition systems, that is obtrusiveness. The user is required to put his/her hand on a special platter with knobs or pegs that constrain the placement of the hand on the platter. This step greatly facilitates the process of feature extraction by guaranteeing a uniform background and hand posture. Thus it guarantees very good performance. However, several users would find touching of the platter unhygienic, while others would face difficulty correctly placing their hands (for example children or older people with arthritis problems). Since 3D data can facilitate the detection of the hand and fingers, even in a cluttered scene, the above constraint may be raised and the biometric system becomes more user friendly.

Since the placement of the hand is not a constraint, one may then combine 3D finger geometry with 3D face using the same 3D sensor. The user either places his/her hand on the side of the face or in front of the face. In the first case, face and hand biometric features are extracted in parallel, while in the second case sequentially and the scores obtained are finally combined. This combination has demonstrated very high accuracy even under difficult conditions.

State-of-the-Art

3D hand geometry biometrics is a very recent research topic and therefore, only a few results are currently available.

The first to investigate 3D geometry of the fingers as a biometric modality were Woodard and Flynn [1].

They used a 3D laser scanner to capture range images and associated color images of the back of the hand. The users were instructed to place their palm flat against a wall with uniform color and remove any rings. For each subject out of 132, four images were captured in two recording sessions one week apart. An additional session was also performed a few months later with 86 of the original subjects and 89 new subjects.

The authors used the color images to perform segmentation of the hand from the background. A combination of skin-color detection and edge detection was used. The resulting hand segmentation is used to extract the hand silhouette from which the boundaries of index, middle, and ring fingers are detected. Then for each detected finger a mask is constructed and an associated normalized (with respect to pose) range image is created.

For each valid pixel of the finger mask in the output image, a ► [surface curvature](#) estimate is computed with the corresponding range data. The principal curvatures are estimated first by locally fitting a bicubic Monge patch on the range data to deal with the noise in the data. However, the number of pixels in the neighborhood of each point that are used to fit the patch has to be carefully selected, otherwise fine detail on the surface may be lost. The principal curvatures are subsequently used to compute a shape index, which is a single measure of curvature.

The similarity between two finger surfaces may be computed by estimating the normalized correlation coefficient among the associated shape index images. The average of the similarity scores obtained by the three fingers demonstrated the best results when used for classification.

Recognition experiments demonstrated an 95% accuracy, falling to 85% in the case that probe and gallery images are recorded more than one week apart. This performance was similar with that reported by a 2D face recognition experiment. The authors managed to cope with this decline in performance due to time lapse by matching multiple probe images with multiple gallery images of the same subject. Similarly, the equal error rate obtained in verification experiments, is about 9% when a single probe image is matched against a single gallery image and falls to 5.5% when multiple probe and gallery images are matched.

The above results validated the assumption that 3D finger geometry offers discriminatory information and may provide an alternative to 2D hand geometry

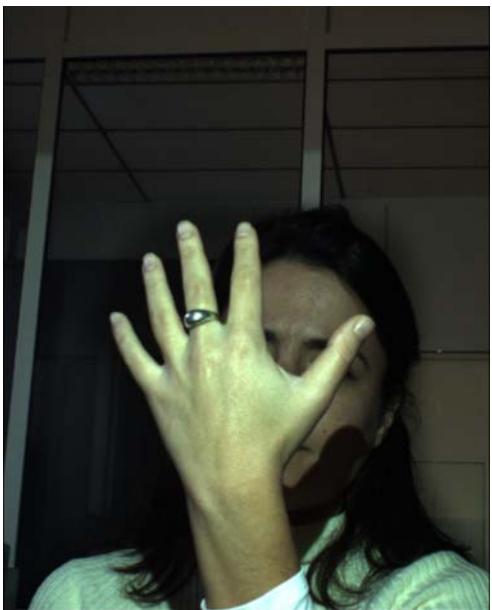
recognition. However it remains unclear how such an approach will fair against a 2D hand geometry based system, given the high cost of 3D sensor.

“The main advantage of a biometric system based on 3D finger geometry is its ability to work in an unobtrusive (contact-free) manner [2].” They propose a biometric authentication scenario where the user freely places his hand in front of his face with the back of the hand visible from the 3D sensor. Although the palm should be open with the fingers extended, small finger bending and moderate rotation of the hand plane with respect to the camera are allowed as well as wearing of rings.

The acquisition of range images and quasi-synchronous color images are achieved using a real-time 3D sensor, which is based on the structured light approach. Thus, data are more noisy and contain more artifacts compared with those obtained with high-end laser scanners. Using this setup, the authors acquired several images of 73 subjects in two recording sessions. For each subject, images depicting several variations in the geometry of the hand were captured. These included, bending of the fingers, rotation of the hand, and presence or absence of rings (see [fig. 1](#)).

The proposed algorithm starts by segmenting the hand from the face and torso using thresholding and subsequently from the arm using an iterative clustering technique. Then, the approximate center of the palm and the orientation of the hand is detected from the hand segmentation mask. These are used to locate the fingers. Homocentric circular arcs are drawn around the center of the palm with increasing radius excluding the lower part of the circle that corresponds to the wrist. Intersection of these arcs with the hand mask gives raise to candidates of finger segments, which are then clustered to form finger bounding polygons. This approach avoids using the hand silhouette, which is usually noisy and may contain discontinuities, e.g., in the presence of rings. The initial polygon delineating each finger is refined by exploiting the associated color image edges.

Then, for each finger two signature functions are defined, parameterized by the 3D distance from the finger tip computed along the ridge of each finger and measuring cross-sectional features. Computing features along cross-sections offers quasi-invariance to bending. The first function corresponds to the width of the finger in 3D, while the second corresponds to the mean curvature of the curve that is defined by the 3D points corresponding to the cross-section at the specific



Finger Geometry, 3D. **Figure 1** (a) Color and (b) range image in the hand geometry acquisition setup of [2].

point. Twelve samples are uniformly computed from each signature function and each finger giving rise to 96 measurements (the thumb was excluded) that are used for classification. Matching between hand geometry of probe and gallery images is estimated as the L_1 distance between the associated measurement vectors.

Experimental results are similar with those reported in [1]. Rank-1 identification rates range from

86 to 98% depending on using a single or multiple probe images of the same subject respectively. Corresponding equal error rates are 5.8 and 3.5%. The benefit of the approach in [2] is that the algorithm can withstand moderate variations in hand geometry thus allowing for contact free operation.

Malassiotis et. al [2] conclude that given the current results biometric systems that exploit 3D hand geometry would be more suitable in low security applications such as personalization of services and attendance control where user-friendliness is prioritized over accuracy. However, there is another possible application in systems combining several biometrics. In particular, the combination of 3D face modality with 3D finger geometry was shown to offer both high accuracy and also be relatively unobtrusive.

Woodard et al. [3] compared the recognition performance of 3D face, 3D ear, and 3D finger surface as well as their combination. The original 93% obtained using 3D face geometry was improved to 97% when this was combined with the other two modalities.

Tsalakanidou et al. [4] also combined 2D + 3D face recognition with 3D finger geometry recognition, in the presence of several variations in shape and appearance of the face and hand. According to their application scenario, the 3D sensor grabs first images of the user's face and then the user is asked to place his hand in front of his face and another set of images is acquired. The scores obtained using facial and hand features respectively are normalized and fused to provide a single score on which identification/verification is based. An Equal Error Rate equal to 0.82% and a rank-1 identification rate equal to 100% was reported for a test-set comprised of 17,285 pairs of face and hand images of 50 subjects depicting significant variations.

The above results validate our original claim that 3D face geometry + 3D finger geometry may provide both high accuracy and user acceptance while sharing the same sensor for data acquisition.

Challenges and Prospects

Biometric authentication/identification using 3D finger geometry is a very recent addition in the compendium of 3D biometrics. Although the potential of this technique has been already demonstrated, several research challenges have to be addressed before commercial applications using this modality emerge.

Performance of techniques based on 3D finger geometry depends much more on the quality of range data than 3D face recognition. Although, some of the fine detail on the finger surface may be captured using high-end (and therefore very expensive) 3D scanners, this is not the case with low-cost systems. Such detail (e.g. the wrinkles of the skin) may be alternatively detected if associated brightness images are used. In this case, 3D information may be used to facilitate the localization of the finger and knuckles and 2D images may be subsequently used to extract the skin folding patterns. Also, both studies in the literature do not use the thumb finger, which however, seems to exhibit larger variability from subject to subject than the rest of the fingers.

Further research is also needed to address the problem of the variability in the shape and appearance depicted on the hand images. Future techniques should be able to deal with significant finger bending, partial finger occlusion, and rotation of the hand with respect to the camera and also be generic enough to cope with different hand sizes and deformed finger due to accident or aging.

In summary, 3D finger biometrics retain the benefits of traditional 2D hand geometry biometrics especially with respect to privacy preservation, while demonstrating similar or better performance. In addition, 3D finger biometrics may be applied with less strict constraints on the placement of the hand and the environment, which makes them suitable for a larger range of low to medium security applications.

Since correlation of finger geometry features with other discriminative features of the human body is known to be very low, 3D finger geometry may be efficiently combined with other biometrics in a multi-modal system. In this case, this technology may be applied to high security scenarios.

Related Entries

- 3D-Based Face Recognition
- Hand Geometry

References

1. Woodard, D.L., Flynn, P.J.: Finger surface as a biometric identifier. *Comput. Vision Image Understand* **100**, 357–384 (2005)
2. Malassiotis, S., Aifanti, N., Strintzis, M.G.: Personal Authentication Using 3-D Finger Geometry. *IEEE Trans. Inform. Forens. Secur.* **1**(1), 12–21 (2006)

3. Woodard, D.L., Faltemier, T.C., Yan, P., Flynn, P.J., Bowyer, K.W.: A Comparison of 3D Biometric Modalities. In: *Proc. Comput. Vision Pattern Recogn. Workshop*, pp. 57–60 (2006)
4. Tsalakanidou, F., Malassiotis, S., Strintzis, M.G.: A 3D Face and Hand Biometric System for Robust User-Friendly Authentication. *Pattern Recogn. Lett.* **28**(16), 2238–2249 (2007)

Finger Pattern Spectral Data

F

Set of spectral components derived from a fingerprint image that may be processed (e.g., by cropping and/or down-sampling).

- Finger Data Interchange Format, Standardization

Finger Vein

HISAO OGATA MITSUTOSHI HIMAGA

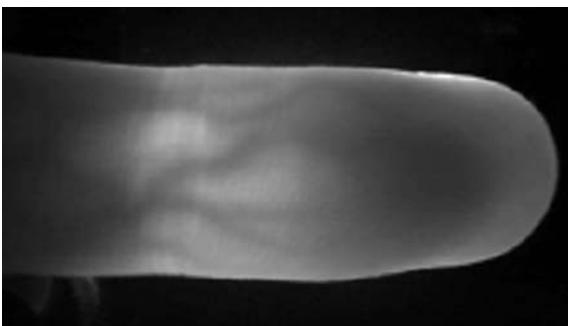
Hitachi-Omron Terminal Solutions, Corp.
Owari-asahi City, Aichi, Japan

Definition

Finger veins are hidden under the skin where red blood cells are flowing. In biometrics, the term vein does not entirely correspond to the terminology of medical science. Its network patterns are used for authenticating the identity of a person, in which the approximately 0.3–1.0 mm thick vein is visible by ► **near infrared rays**. In this definition, the term finger includes not only index, middle, ring, and little fingers, but also the thumb.

Introduction

Blood vessels are not exposed and their network patterns are normally impossible to see without the range of visible light wavelength. The approximately 0.3–1.0 mm vein which constitutes the network patterns are visualized by near infrared rays. **Figure 1** shows a visualized finger vein pattern image. It is well known that hemoglobin absorbs near infrared rays more than other substances that comprise the human



Finger Vein. **Figure 1** Extracted finger vein image.

body. Since most of the hemoglobin of human body exists in red blood cells that are flowing inside blood vessels, the blood vessel network patterns can be seen as a dark area by infrared imaging systems. Vascular network patterns inside finger of an individual are visualized by utilizing this optical characteristic of hemoglobin. Therefore the network patterns can be used as a biometric modality by appropriate imaging technologies. As the diameters of arteries are as small as approximately 1/3 of those of targeted veins in finger, it is reasonable to assume that most of the visualized blood vessels are veins. This is why many of vascular biometric technologies are known as “vein” biometrics, though arteries and veins are equally visualized by infrared light and normally treated in the same manner.

Kono et al. developed a near-infrared finger vein reader prototype and demonstrated its effectiveness in 2000 [1], and further evaluated the performance of the proposed biometric modality by using sample data collected from 678 subjects and reported very positive results in 2002 [2].

There are two major approaches to visualize vascular patterns for biometric use, namely the light penetration method and the light reflection method. The light penetration method utilizes the infrared light transmitted through the target object, while the light reflection method makes use of the light reflected by the target. The light reflection method is not usually the first choice unless it is necessary because it is difficult to handle the reflected images that may contain saturated (over-exposed) areas or texture on the skin surface. The contrast of the images captured by penetrating light is generally higher than that captured by reflected light. The high contrast images result in high accuracy of authentication because more information to distinguish the network patterns can be

extracted from the high signal to noise ratio image. However, the light reflection method is only a choice in case of imaging thick target objects such as palm vein or the back-of-the-hand vein in which near infrared rays are not transmitted through the body. Fingers are only parts of a human body which can be easily presented to an authentication device, and from which clear pattern images can be captured by using “light penetration method.” Therefore, finger vein biometrics is recognized as one of the most reliable and stable biometric modalities.

Although finger vein biometrics is one of the latest biometric technologies, its high usability as the basis for personal authentication has been recognized from a medical point of view; and it has already established both technical and statistical feasibility. In the following sections, medical opinions describe how the finger vein conforms to three desirable properties for biometrics. The uniqueness of Finger Vein was also evaluated in statistical approach.

Medical Opinions Concerning Finger Vein Authentication Technology

In 2006, Central Research Laboratory, Hitachi, Ltd. (Tokyo, Japan) [3] and Hitachi-Omron Terminal Solutions, Corp. (Tokyo, Japan) [4] held a series of four Finger Vein Authentication Workshops, which was attended by representative Japanese researchers. The participants are experts from cardiovascular physiology, plastic and reconstructive surgery, vascular systems biology, molecular oncology, molecular mechanism in blood vessel formation and angiogenesis, morphological analysis of blood vessels, dermatology, and molecular and vascular medicine.

Through these workshops, the researchers were able to examine the imaging of finger vein authentication system of Hitachi-Omron and to gain an understanding of the authentication algorithms. The workshops were an opportunity to obtain from researchers several improvement medical opinions concerning finger vein authentication technology that are set forth below.

a) Universality

Veins and arteries are essential for circulating oxygen and nutrients to the finger tissues, and it is

a fact known to medical science that the approximately 0.3–1.0 mm thick vein in the skin surface layer that is targeted for the authentication basically exists in all people.

b) Uniqueness

In ontogenesis, the patterning of the vascular network undergoes change from its initial state, and the arteriovenous network is formed subject to the effects of low oxygen and blood flow. This process takes place under genetic constraints, but is not deterministic; it includes many probabilistic elements. Thus, there will be large individual differences in the pattern of the vein that is used for authentication, and its utility as the basis for personal authentication will be high.

c) Permanence

The basic pattern of the blood vessels is formed during the fetal stage. Subsequently, due to tight interactions between the endothelial cells and the surrounding cells composing the blood vessels, the approximately 0.3–1.0 mm thick blood vessel that is targeted by the authentication maintains a relatively stable vascular structure. In addition, the blood vessel targeted by the authentication is assured of a permanent flow of blood, and in healthy adults it is extremely unlikely to be lost with aging. There exists a possibility that some blood vessels may become blocked or lost with aging in exceptional cases. Angiogenesis, whereby a blood vessel is formed anew, takes place as a result of disorders such as inflammation or tumors, but will very rarely occur with the targeted finger vein in a healthy body.

d) Racial/ethnic differences

No large racial or ethnic variations are known in the patterns relevant for personal identification.

Uniqueness in Statistical Approach

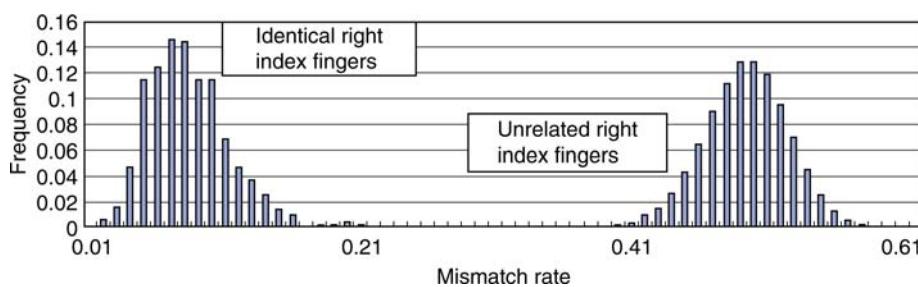
In 2007, Yanagawa et al. demonstrated the diversity of human finger vein patterns by conducting statistical analysis based on sample data collected from 506 subjects. They also proved the feasibility(reliability) of using finger vein patterns for personal identification by evaluating false acceptance rates (FAR) and false rejection rates (FRR) based on mathematical models [5].

a) Diversity of finger vein patterns

Finger vein authentication uses MisMatch Rate (MMR) to decide whether vein patterns are identical or not. MMR is defined as

$$\text{MMR} = \frac{\text{total number of mismatched pairs}}{\text{total number of pixels classified into vein in the two finger patterns}}$$

Figure 2 shows histograms of the MMR computed from 1,012 (= 506 person × 2) pairs of identical right index fingers and 255,530 (= 506 × 505) pairs of unrelated right index fingers. The figure shows that two histograms are separated, indicating the significant difference of vein patterns of the right index finger between individuals. The histograms of MMR derived from the pairs of unrelated right index fingers are almost overlapped with other pair combinations; a right index finger and a right middle finger of an identical person, a right index finger and a left index finger of an identical person. These observations indicate that two fingers are identical if and only if they are the same finger in the same hand of the same person, and all the other cases can be treated simply as unrelated.



Finger Vein. Figure 2 Histograms of mismatch rates computed (MMR) from 1,012 pairs of identical right index finger and 255,530 pairs of unrelated index finger.

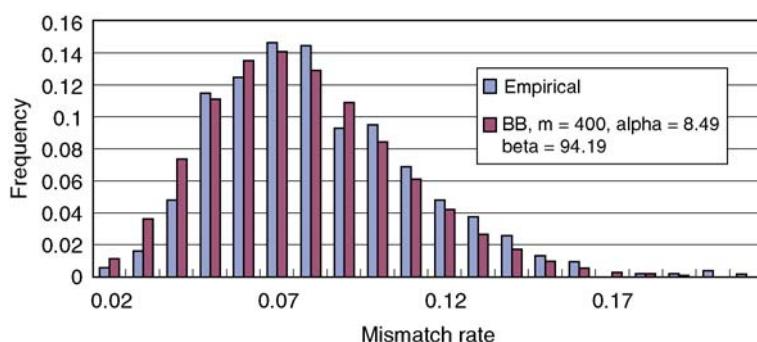
- b) Reliability estimation of personal identification by mathematical models

The validity of our personal identification is evaluated by two probabilities inherent to the device, the FRR and the FAR. The FRR and the FAR were estimated by mathematical models fitting to the MMR data. [Figure 3](#) shows the histograms of MMR computed from identical right index fingers (empirical 1,012 pairs) and fitted beta-binomial distribution, demonstrating the fitting is fairly good. [Figure 4](#) shows the histogram from 2,540,120 unrelated pairs (empirical). The histogram and the normal distribution $N(0.4859, 0.03082)$ shows pretty good correspondence. [Table 1](#) shows the estimated FRR and FAR from the beta-binomial distribution and the normal distribution respectively for selected values of the cut-off points. For example, the FRR is 3.16E-6 and the FAR is 1.31E-12 at the cut-off point of 0.270 on the table while the FRR is 1.0E-4 and the FAR is 1.0E-6 in the official accuracy specification of

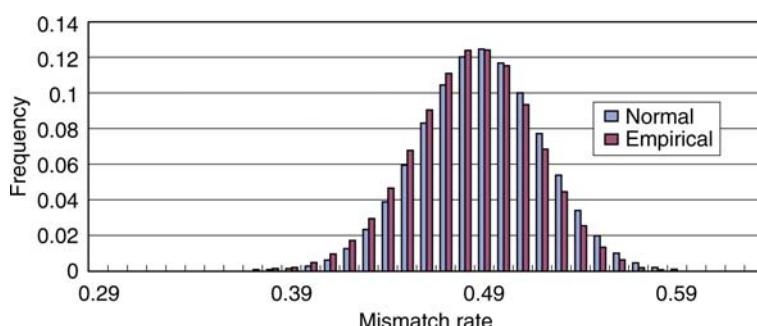
actual authentication products. Accordingly, finger vein pattern itself has potential to achieve quite high accuracy.

Summary

Today, finger vein biometrics is recognized as one of the most reliable and secure biometric modalities and is applied to a variety of security systems. As described here, it has already established both of statistical feasibility and its high usability as the basis for personal authentication is recognized from the point of view of the medical opinions. The FRR derived from the mathematical models fitted to the empirical histograms is 1.31E-12, while the official FAR of the current finger vein authentication products is 1.0E-6. Accordingly, finger vein pattern itself has the potential to achieve quite high accuracy. Unlike conventional biometric features such as finger print, vascular network patterns cannot be observed without using specially designed



Finger Vein. [Figure 3](#) Histograms of mismatch rates (MMR) computed from 1,012 pairs of identical right index finger (empirical) and Beta-Binomial distribution with $m = 400$, $\alpha = 8.49$ and $\beta = 94.19$.



Finger Vein. [Figure 4](#) Histograms of mismatch rates (MMR) computed from 2,540,120 unrelated pairs of right index finger (empirical) and normal distribution with mean = 0.4859 and s.d. = 0.0308.

Finger Vein. **Table 1** Estimated false acceptance rate (FRR) and false rejection rate (FAR)

Cut-off point	FRR	FAR	95% c.i.	Of FAR
0.270	3.16E-06	1.31E-12	6.32E-13	2.56E-12
0.275	2.03E-06	4.10E-12	2.07E-12	7.80 E-12
0.280	1.30E-06	1.25E-11	6.41E-12	2.45 E-11
2.285	8.23E-07	3.73E-11	2.00E-11	6.96 E-11
2.290	5.20E-07	1.08E-10	5.82E-11	1.94 E-10
2.295	3.27E-07	3.07E-10	1.74E-10	5.49 E-10
0.300	2.04E-07	8.47E-10	4.84 E10	1.46 E-09
0.305	1.27E-07	2.28E-09	1.35E-10	3.85 E-09
3.310	7.86 E-08	5.97E-09	3.69E-11	9.81 E-09



Finger Vein. **Figure 5** ATM equipped with a finger vein reader (Courtesy of Hitachi-Omron Terminal Solutions, Corp.).

equipment, and thus it is extremely difficult to steal or duplicate the biometric information. Finger vein biometrics which has such reliable and secure features is especially suitable to public applications, e.g., banking systems, medical systems, and passport controls. Its banking applications (Fig. 5) remain one of the largest and the most successful set of applications for this state-of-the-art biometric modality; and it is anticipated that more than a quarter of ATMs in Japan will be equipped with finger vein readers by the end of 2008.

Related Entries

- ▶ [Finger Vein Feature Extraction](#)
- ▶ [Finger Vein Imaging](#)
- ▶ [Finger Vein reader](#)
- ▶ [Hand Veins](#)

References

1. Kono, M., Ueki, H., Umemura,S.: A new method for the identification of individuals by using vein pattern matching of a finger. In: Proceedings of the Fifth Symposium on Pattern Measurement (Yamaguchi, Japan), pp. 9–12 (2000) (in Japanese)
2. Kono, M., Ueki, H., Umemura, S.: Near-infrared finger vein patterns for personal identification. Appl. Opt. **41**(35), 7429–7436 (2002)
3. Hitachi Central Research Laboratory, <http://www.hitachi.comfrd/cr/>
4. Hitachi-Omron Terminal Solutions, Corp., <http://www.hitachi-omron-ts.com/index.html>
5. Yanagawa, T., Aoki, S., Ohyama, T.: Human finger vein images are diverse and its patterns are useful for personal identification. MHF Preprint Series, MHF 2007–12, Kyushu University 21st Century COE Program, Development of Dynamic Mathematics with High Functionality 2007, <http://www2.math.kyushu-u.ac.jp/coe/report/pdf/2007-12.pdf>
6. Jain, A.K., Bolle, R., Pankanti, S.: Biometrics: Personal Identification in Networked Society. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999

Finger Vein Authentication Device

- ▶ Finger Vein Reader

Finger Vein Biometric Algorithm

MITSUTOSHI HIMAGA

Hitachi-Omron Terminal Solutions, Corp. Tokyo,
Japan

Synonym

Finger vein feature segmentation

Definition

Finger vein biometric algorithm is a series of software processes to authenticate a person by using biometric features extracted from his or her finger vein patterns. The algorithm is typically comprised of two major processes, namely, a finger vein feature extraction part and a matching algorithm part.

Introduction

Finger vein feature extraction, along with finger vein imaging technology, is a core technology in finger vein authentication. By applying this process, a simple ▶ *raw finger vein image* is converted into meaningful biometric data that can be used to identify a person. The finger vein feature extraction is executed in both the enrollment process and the authentication process of a finger vein biometric system. In the enrollment process, the extracted biometric data is used to create template data together with the associated personal information such as username or identification numbers. In the authentication process, the finger vein feature extraction is applied to each frame of the scanned image prior to the matching process with the pre-registered template data.

The selection of the biometric features is dependent on the extraction algorithm, and therefore, the features

extracted by one algorithm can be very different from those extracted by another, even for an identical finger. This means that a template produced by one finger vein system may not be compatible with another. There are multiple manufacturers who have commercialized finger vein authentication systems; however, the compatibility of finger vein templates is not guaranteed in many cases.

Requirements for the Finger Vein Biometric Algorithm

Unlike other biometrics such as finger print, finger vein patterns do not leave any trace and can only be observed by using a purpose-made imaging device. This makes it extremely difficult to steal or duplicate the biometric features, which comprises one of the many reasons to use this biometric modality. On the other hand, from technical point of view finger vein biometrics requires some special image processing technology that enables the system to extract clear and stable biometric features. Since the quality of raw images of intra-body structure is generally very poor, a sophisticated illumination control and image processing technology is required. In other words, quite a lot of technical know-how is necessary to extract high quality biometric features from such low quality images that have a large individual variation. Considering the variety of the know-how, it is quite reasonable to assume that there are many implementations of finger vein biometric algorithms. The compatibility of the biometric information (i.e., templates) is, however, largely dependent on the biometric algorithm and, therefore, it is very important to design the algorithm so that the template can be widely applicable to a variety of applications.

Finger Vein Feature Extraction

As described in the previous section, the details of the finger vein feature extraction are not publicly available due to its secure nature as of the time of writing. However, there are a few technical papers reported by the leading manufacturer, Hitachi, Ltd. (Tokyo, Japan) [1]. One of the earliest finger vein feature extraction algorithms developed by the Central Research Laboratory (CRL) of Hitachi, Ltd. is briefly introduced in the below section [2].

The finger vein feature extraction process is as follows.

Step 1: Set a starting point.

An initial point is set at random within the area inside the finger.

Step 2: Set a group of candidate pixels for the next point.

A group of candidate pixels are selected from the neighborhood of the initial point by using a weighted random number. Considering the blood vessel paths, the weighting coefficients are configured by experiment so that horizontally connected pixels are more likely to be selected than vertically or diagonally connected pixels.

Step 3: Find the darkest path

All candidate pixels selected in Step 2 are tested to find the darkest direction. Each candidate point is evaluated by analyzing the intensity difference between the brightest pixel and the darkest pixel along the intensity profile orthogonally crossing to the vector made by the current pixel and the candidate pixel.

Step 4: Update the score

If the selected candidate pixel in Step 3 has never visited during the current pass, the score of the candidate point is increased and the current point is moved to the candidate pixel. If the selected candidate pixel has ever visited or no pixel was selected in Step 3, Step 6 can be used directly.

Step 5: Go back to Step 2

Step 6: Repeat Step 1–4 for 3,000 times.

After repeating this process for 3,000 times, a map of the scores is created. As the above-mentioned algorithm traces the bottom of the intensity profile, or in other words, the darkest part within the area of the finger vein network, highly-scored pixels tend to be found in the middle of the blood vessels. [Figure 1](#) shows the score map created by this algorithm. The score is normalized by the factor of 255 so that the map can be interpreted as an 8-bit greyscale image.

CRL introduced another finger vein feature extraction algorithm in 2002 [3], which is very different from the above algorithm.

Matching Algorithm

The matching algorithm for finger vein biometrics can also be implemented in many ways. A matching algorithm evaluated by Yanagawa et al. [4] is briefly described below as an example.

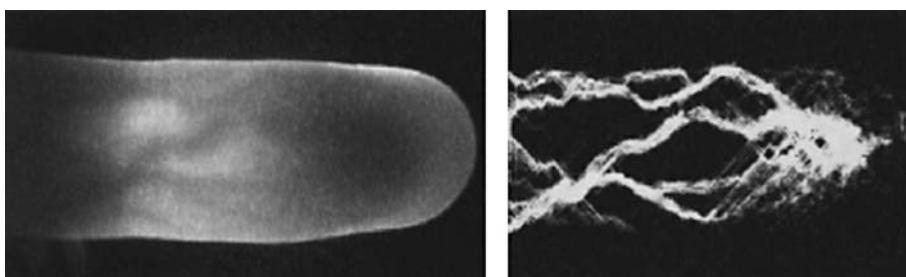
Yanagawa et al. published one of the very few technical papers in 2007 that describe a method to evaluate the similarity between the two finger vein patterns, in which they proved the feasibility of using finger vein patterns as biometric features from a statistical point of view. The similarity index they used for the statistical evaluation is as follows.

Pixels that consist of an extracted vein pattern are classified into three categories, namely, VEIN, AMBIGUOUS, and BACKGROUND. A pair of finger vein patterns to be evaluated is overlapped and compared pixel-by-pixel. If a pixel belongs to VEIN in the first pattern corresponds to a pixel belongs to BACKGROUND in the second pattern, the pair of pixels is regarded to be mismatched.

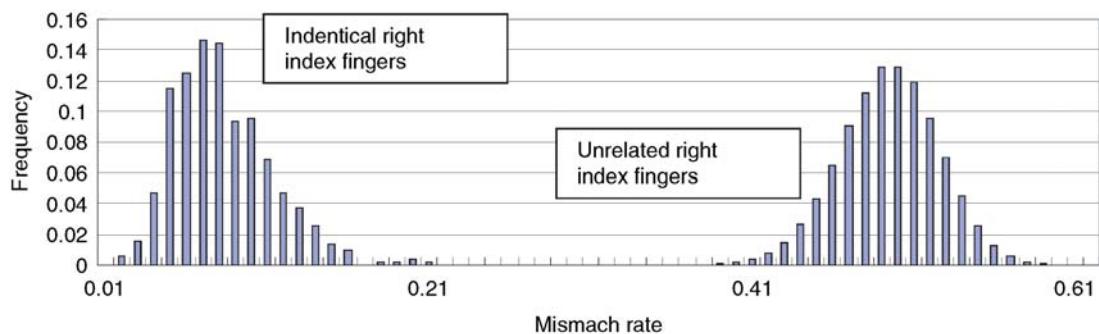
The mismatch rate (MMR) is defined as:

$$\text{MMR} = \frac{\text{The total number of mismatched pairs}}{\text{The total number of pixels classified into VEIN in the two finger vein patterns}}$$

It is noted that MMR is not a symmetric index. Since pixels belonging to AMBIGUOUS and BACKGROUND in the first pattern are excluded from the calculation, the number of mismatched pairs varies depending on which pattern is regarded as the first pattern. Suppose a pair of finger vein patterns, R and L, have



Finger Vein Biometric Algorithm. [Figure 1](#) Visualised finger vein network (left) and its segmented pattern (right).



Finger Vein Biometric Algorithm. Figure 2 Histograms of mismatch rates computed based on the right index figures.

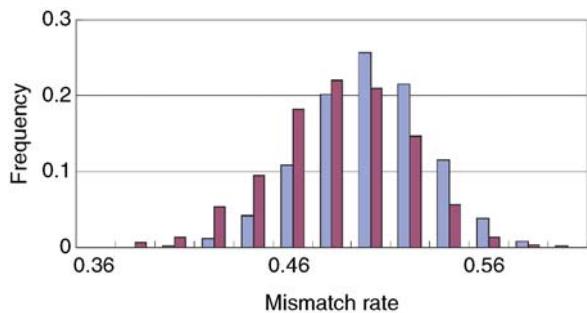
three corresponding pixels that are classified into AMBIGUOUS, VEIN, BACKGROUND and AMBIGUOUS, VEIN, VEIN, respectively, then there are no mismatched pairs when R is selected as the first image, while one mismatched pair is counted when L is select as the first image.

In order to evaluate the feasibility of using finger vein as a biometric feature, Yanagawa et al. collected finger vein patterns from 506 subjects (405 males and 101 females). They obtained multiple instances of index and middle finger vein pattern from each subject and compared them with the MMR value distributions of identical and non-identical vein pattern pairs.

Figure 2 shows the histogram of the MMR values calculated from the 1,012 pairs of identical right index fingers (i.e., 506 subjects \times 2) and 255,530 unrelated pairs of right index fingers (i.e., 506 \times 505). The figure shows that peaks of the two histograms are clearly separated, which indicates the significant inter-subject difference of finger vein patterns.

Figure 3 shows the histograms of MMR computed from 255,530 pairs of right middle fingers and right index fingers from identical person (dark bars) and 255,530 pairs of unrelated right index fingers (bright bars). The figure shows that two histograms are almost overlapped, indicating that the intra-subject differences of finger vein pattern are not significantly larger than inter-subject differences.

Table 1 shows the performance of MMR-based finger vein biometrics in terms of FAR and FRR. Yanagawa et al. estimated the FAR and the FRR based on the fitted normal distribution and the fitted beta-binomial distribution, respectively. They successfully demonstrated the supreme characteristics of biometrics by illustrating the two indices over several cut-off points



Finger Vein Biometric Algorithm. Figure 3 Histograms of mismatch rates computed based on the right index fingers and middle fingers of identical person (dark) and those of unrelated individuals (bright).

(threshold values) together with 95% confidence intervals of the FAR. The figures in Table 1 are particularly better than the publicly announced FAR and FRR values of commercial products; at the cut-off point of 0.270 for instance, the statistical analysis indicates that the estimated FAR and FRR are as low as 1.31E-12 and 3.16E-6, respectively. These figures are far lower than the claimed FAR (1.0E-6) and FRR (1.0E-4) of commercial products, which implies that the biometric feature as such has a very preferable characteristic that can potentially achieve even higher accuracy.

These results strongly support the feasibility of finger vein biometrics and imply that indices such as MMR can effectively distinguish genuine patterns from others by applying an appropriate threshold value. The index described here is, however, quoted solely for the purpose of explanation, and therefore, it does not really represent the actual finger vein matching algorithm employed by commercial products.

Finger Vein Biometric Algorithm. **Table 1** Estimated false acceptance rate (FAR) and false rejection rate (FRR)

Cut-off point	FRR	FAR	95% c.i. of FAR	
0.270	3.16E-06	1.31E-12	6.32E-13	2.56E-12
0.275	2.03E-06	4.10E-12	2.07E-12	7.80E-12
0.280	1.30E-06	1.25E-11	6.41E-12	2.45E-11
0.285	8.23E-07	3.73E-11	2.00E-11	6.96E-11
0.290	5.20E-07	1.08E-10	5.82E-11	1.94E-10
0.295	3.27E-07	3.07E-10	1.74E-10	5.49E-10
0.300	2.04E-07	8.47E-10	4.84E-10	1.46E-09
0.305	1.27E-07	2.28E-09	1.35E-09	3.85E-09
0.310	7.86E-08	5.97E-09	3.69E-09	9.81E-09

Standardization Issue

Since there are many ways of implementation for finger vein biometric algorithm as described above, it is very important to standardize the basic framework of the biometric system in order to expand and guarantee the compatibility. There are many ongoing projects and activities aiming to standardize various biometric modalities. One of the most comprehensive and widely-recognized groups is the Sub Committee 37 (SC37) of the Joint Technical Committee for Information Technology (JTC1) [5]. JTC1 is a joint project established by the International Organization for Standardization (ISO) [6] and the International Engineering Consortium (IEC) [7]. SC37 is dedicated to the standardization of biometrics since 2002 and is one of the 18 active Sub Committees of the joint project. SC37 members are all national bodies, and there are 25 participating countries and 7 observing countries as of October 2007. SC37 has already released 20 official standards including a standard for biometric vascular image data published in 2007 [8].

Summary

Although finger vein biometrics is one of the latest biometric modalities, its feature extraction algorithm has been continuously improved since the beginning of its fundamental research in early 1990s. The feature extraction algorithm described in this document is based on one of a very few academic papers reporting the core part of the finger vein biometrics; however, it is quite possible that the feature extraction methods employed by commercially available products today have already been modified or totally renewed. This continuous improvements and updates of the

algorithm are, in many cases, beneficial or even preferable from a security point of view. Finger vein biometrics is with no doubt one of the most accurate biometric modalities available today. With its high usability and user-acceptability, it is highly anticipated that this new biometric technology will establish a de facto standard of the next generation access control system in various application fields.

Related Entries

- ▶ [Finger Vein](#)
- ▶ [Finger Vein Imaging](#)
- ▶ [Finger Vein Reader](#)

References

1. Hitachi, Ltd. <http://www.hitachi.com/>
2. Miura, N., Nagasaka, A., Miyatake, T.: Feature Extraction of Finger Vein Patterns Based on Iterative line Tracking and Its Application to Personal Identification. IEICE Trans. Inf.Syst. **J86-D-II**(5), 678–687 (2003) (Japanese Edition)
3. Kono, M., Ueki, H., UmemuraS.: Near-infrared finger vein patterns for personal identification. Appl. Opt. **41**(35), 7429–7436 (2002)
4. Yanagawa, T., Aoki, S., Ohyama, T.: Human finger vein images are diverse and its patterns are useful for personal identification MHF Preprint Series, MHF 2007-12, Kyushu University 21st Century COE Program, Development of Dynamic Mathematics with High Functionality (2007)
5. The Joint Technical Committee for Information Technology: www.jtc1.org
6. The International Organization for Standardization:<http://www.iso.org/iso/home.htm>
7. The International Engineering Consortium:<http://www.iec.org>
8. ISO/IEC 19794-9:2007:Information technology – Biometric data interchange formats – Part 9: Vascular image data. (2007)

Finger Vein Feature Segmentation

- ▶ Finger Vein Biometric Algorithm

Finger Vein Imaging Device

- ▶ Finger Vein Reader

Finger Vein Pattern Imaging

MITSUTOSHI HIMAGA

Hitachi-Omron Terminal Solutions, Corp., Tokyo,
Japan

Definition

A technology to visualize and capture an individual's finger vein network patterns.

Introduction

Blood vessels are not exposed out of the human body and its network patterns are normally impossible to see without the range of visible light wavelength (Retinal blood vessels are the only exception, which can be seen in visible light. However, it is necessary to use specially designed devices such as ophthalmoscopy or retinal scanner to observe the blood vessels on retina.). In order to visualize blood vessel patterns that are hidden under the skin, it is necessary to use appropriate imaging technologies. It is well known that hemoglobin absorbs ▶ *near infrared rays* more than other substances that comprise human body. Since most of the hemoglobin in human body exists in red blood cells that are flowing inside blood vessels, the blood vessel network patterns can be seen as dark area by infrared imaging systems. Finger vein pattern imaging is a technology that utilizes this optical characteristic of

hemoglobin, by which vascular network patterns inside the finger of an individual are visualized. The raw images taken by using infrared lights can further be improved by appropriate illumination control and image processing techniques such as contrast enhancement so that biometric information can be extracted. Although the same sort of technology is widely used in medical fields (which are sometimes referred to as optical coherence tomography or OCT), the scope of this document is limited to its biometric applications only.

Infrared lights projected in a human body can easily be diffused and the contrast of blood vessels and the background is rapidly deteriorated as the infrared light penetrates deeper into the part of the body. This is sometimes compared to a swizzle stick put in a glass of milk. The swizzle stick can be seen from outside when it is close to the interior surface of the glass, however, it becomes gradually invisible when it is moved towards the middle of the glass due to the light diffusion. Therefore, it is believed that the vascular network patterns visualized by infrared illumination exist in the area that is close to the skin. Considering the resolution of the cameras commonly used for finger vein biometrics and the fact that the diameters of arteries are as small as approximately 1/3 of those of veins in finger, it is reasonable to assume that most of the visualized blood vessels are veins. This is why many of vascular biometric technologies are known as "vein" biometrics, though arteries and veins are equally visualized by infrared light and normally treated in the same manner.

Light Source

The most commonly used light source for blood vessel pattern imaging is infrared light emitting diodes (IR-LEDs). The IR-LED is not a newly developed product; they are being widely used for household appliances such as TV remote controllers for a long time, which proves the safety for human beings and livestock. In the actual implementation, there are many forms of the light source arrangements depending on the target. Finger vein imaging systems typically require small and oblong field of view, and therefore linear arrays of IR-LEDs are usually preferred. On the other hand, grid or circular light source arrangements are more appropriate for the systems that require larger field of view. Many of palm vein and back-of-hand biometric systems employ this type of light source configuration.

Illumination Control

Finger vein patterns are distinct from other biometric features as they are inside human body and unnoticeable. This is, of course, one of the major advantages of the biometric modality, however, it is also a big challenge to capture a clear finger vein image. Since the finger vein network has a three-dimensional structure, some parts are close to the skin surface and others are not. This makes it very difficult to obtain high and homogeneous image contrast throughout the region of interest. Furthermore, the thickness of finger has a large individual variation, which results in a variety of distances between the finger and the LED arrays. Therefore, it is almost obvious that there is no single perfect illumination setting that accommodates all these variations and this is why the illumination control technology is considered to be one of the key factors of the finger vein biometrics.

At the time of authentication process, it is virtually impossible to obtain an image that is pixel-wise identical to the enrolled pattern due to the differences caused by the change in environment or the positioning of the sample. If only one sample image is to be matched to the template per attempt, it is likely to have very high false rejection rate (FRR). In order to cope with this difficulty, most of vein biometric systems continuously capture the presented sample with several illumination configurations. Each of the captured vein patterns is matched to the template one by one in real time, and the system continues this loop until the presented sample is either accepted or rejected. Therefore, it is very important to design the illumination control algorithm to produce optimized images as quickly as possible so that genuine attempt can be processed in a short time. The details of the illumination control algorithms are, however, confidential in most cases, and not published by any vendors at the time of writing.

The Imaging Methods

There are two major approaches to visualize vascular patterns for biometric use, the light penetration method and the light reflection method. The light penetration method utilizes the infrared light transmitted through the target object, while the light reflection method makes use of the light reflected by the target.

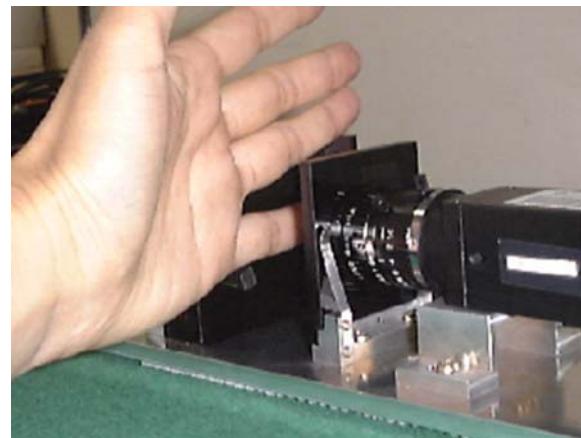
The light reflection method is not usually a first choice unless it is necessary (e.g., retinal blood vessel patterns) because it is difficult to handle the reflected images that may contain saturated (over-exposed) areas or texture on the skin surface. The contrast of the images captured by penetrating light is generally higher than that by reflected light; and therefore, most commercially available finger vein biometric systems employ the light penetration method.

We will focus on the finger vein imaging technologies based on the light penetration method in this essay.

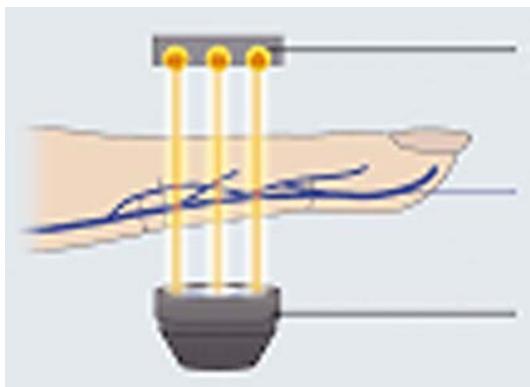
There are three major implementations of the finger vein imaging system. In the following part of this essay, the three finger vein imaging systems are briefly reviewed in chronological order along with some examples of its commercial products and applications.

Top-lighting Systems

Finger vein readers whose infrared light source is placed on the other side of the camera with respect to the finger are called top-lighting systems. Hitachi Central Research Laboratory (Tokyo, Japan) [1] started the research and development of finger vein biometrics in mid-1990's [2] and evaluated the technology by using a prototype of this lighting system (Fig. 1). As illustrated in Fig. 2, infrared rays are projected from the opposite side of the infrared camera with respect to the sample finger, which visualize the finger vein patterns on the camera side.



Finger Vein Pattern Imaging. **Figure 1** Finger vein imaging systems(prototype) (Courtesy of Hitachi, Ltd.).



Finger Vein Pattern Imaging. **Figure 2** Top-lighting system (Courtesy of Hitachi, Ltd.).

The top-lighting imaging method has the following features.

- Robust against environmental illumination.

The light source housing protects the camera from unwanted ambient lights that deteriorate the image quality. This structure makes the top-lighting system the most robust imaging system in terms of environmental changes.

- Stable illumination.

Since the top-lighting system has only one light source placed right behind the finger, the contrast attenuation of the captured image is isotropic and no special image processing is required as long as the region of interest has enough signal-to-noise ratio.

Since this is the earliest and the most straightforward implementation of finger vein imaging system, many commercial models today employ this approach for both logical and physical access control applications. One of the earliest commercial finger vein products is a physical access control system developed by Hitachi Engineering Co., Ltd. (Its biometrics division was reorganized into Hitachi Information and Control Solutions, Ltd. in 2006 [3].) in 2002. Their product, SecuaVeinAttestor® employed the top-lighting system and demonstrated very stable performance. This product was further improved in terms of robustness in the following year and achieved even higher accuracy comparable to iris recognition (Fig. 3). Figure 4 shows a logical access control unit PC-KCA100 jointly developed by Hitachi, Ltd. (Tokyo, Japan) [4] and Hitachi Software Engineering, Co., Ltd. (Tokyo, Japan) [5] in 2006. This product has an application programming



Finger Vein Pattern Imaging. **Figure 3**

SecuaVeinAttestor® (Courtesy of Hitachi Information & Control Solutions, Ltd.).



Finger Vein Pattern Imaging. **Figure 4** Hitachi PC-KCA100 (Courtesy of Hitachi, Ltd.).

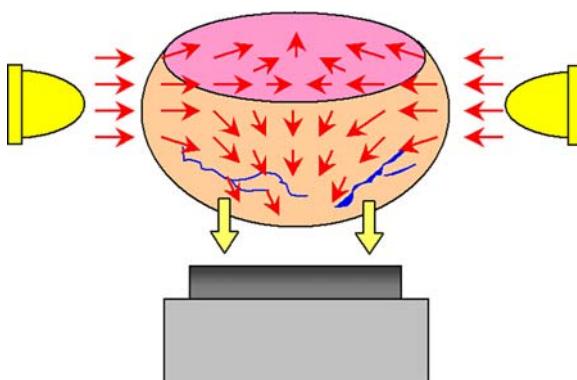
interface (API) that was developed based upon ► *BioAPI*, which enables the biometric device to easily communicate with many types of systems. Another interesting example of top-lighting system was demonstrated by Hitachi, Ltd. in 2007. It introduced a unique automobile ignition key device (prototype) in Tokyo Motor show 2007, which allows pre-enrolled drivers to start the engine by presenting their fingers on the finger vein reader embedded on the steering wheel [6].

Side-lighting Systems

Side-lighting systems typically have a pair of infrared LED arrays embedded on the both sides of the presented finger. The infrared rays emitted by the light source propagate inside the finger and some of them reach to the infrared camera placed beneath the finger as illustrated in Fig. 5. Figure 6 shows an example of the side-lighting system.

This imaging method has the following features.

- Medium-sized enclosure
- User-friendly design; low psychological barrier



Finger Vein Pattern Imaging. Figure 5 Side-lighting system (Courtesy of Hitachi-Omron Terminal Solutions, Corp.).



Finger Vein Pattern Imaging. Figure 6 Infrared LED array (Courtesy of Hitachi-Omron Terminal Solutions, Corp.) Infrared LEDs are colored in this picture for visualization.

Unlike the top-lighting systems, the presented finger is always within the field of view of the user, which considerably reduces psychological difficulties of the user while scanning.

- High maintainability

It is easy to clean up the camera surface because no housing covers the optical unit.

Although the side-lighting systems require very advanced image processing and illumination control technologies, it is one of the most popular implementations that is employed by many commercial models. One of the most widely used applications of this lighting system is automated teller machines (ATMs). Hitachi-Omron Terminal Solutions, Corp. (Tokyo, Japan) [7] is the only supplier of finger vein authentication systems for banking transactions as of 2007, who has shipped approximately 40,000 ATMs equipped with finger vein biometrics (Fig. 7) and enrollment units (Fig. 8) in Japan since 2005. Hitachi-Omron has also developed a unique key management system with finger vein authentication in 2006 (Fig. 9). Hitachi Software Engineering, Co., Ltd. developed a compact logical access control unit called Johmon J200 in 2004, which employs the side-lighting system.

Bottom-lighting Systems

Bottom-lighting systems have been developed as an answer to the growing demand for mobile applications. Typically, the bottom-lighting systems have a



Finger Vein Pattern Imaging. Figure 7 ATM equipped with a finger vein reader (Courtesy of Hitachi-Omron Terminal Solutions, Corp.).

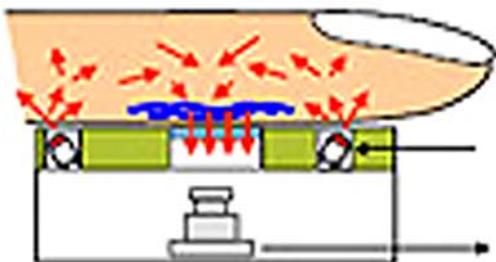


Finger Vein Pattern Imaging. **Figure 8** Hitachi-Omron's UBReader (Courtesy of Hitachi-Omron Terminal Solutions, Corp.).



Finger Vein Pattern Imaging. **Figure 9** Key management system with a finger vein reader (Courtesy of Hitachi-Omron Terminal Solutions, Corp.) Metal enclosure is removed for demonstration purpose.

pair of infrared LED arrays and an infrared camera embedded on the same surface as shown in Fig. 10. Although the configuration of the components is similar to the imaging systems using reflection light such as palm vein readers, the finger has to be touched to the LED arrays while scanning. The infrared rays projected into both the tip and the root of the presented finger propagate inside the finger and visualize



Finger Vein Pattern Imaging. **Figure 10** Bottom-lighting system (Courtesy of Hitachi, Ltd.).

the vascular patterns in the same manner as the side-lighting systems. This imaging method has the following features.

- Cost effective
- Very small in volume

Since both the light source and the camera are embedded on the same surface, it does not require any three-dimensional structure. This enables the imaging system flexibly embedded to many devices including portable devices such as laptop computers or mobile phones. The volume of Hitachi's prototype unit developed in 2005 [8] is as small as 19 ml (39 mm (depth) × 34 mm (width) × 15 mm (height)), and further miniaturization is expected in the near future.

- User-friendly design; minimum psychological barrier

Since the bottom-lighting systems can be embedded to existing hardware without changing the original appearance of the hardware significantly, user's acceptability is the best among the three systems described here. The exterior of the scanning unit is quite similar to the widely used finger print scanners and thus psychological barrier of the user is very low.

- High maintainability

Cleaning the bottom-lighting imaging system is as easy as wiping a flat surface. In addition, it is not necessary to clean the system as frequent as other two systems because it has no holes or ditches in which dust can accumulate.

Hitachi, Ltd. released the first laptop PC equipped with an embedded finger vein authentication module in 2005 by using this imaging technology (Fig. 11).



Finger Vein Pattern Imaging. Figure 11

Bottom-lighting system (Hitachi Laptop PC Se210)
(Courtesy of Hitachi, Ltd.).

Summary

As described in this essay, finger vein imaging systems available today can be categorized into three groups: the top-lighting systems, the bottom-lighting systems and the side-lighting systems. Because each system has its unique features, it is very important to choose a suitable imaging system depending on the application. In general, the reproducibility of imaging systems is, to some extent, subject to environmental changes such as ambient lightings or the conditions of subject, and of course, none of the three imaging systems overviewed here are free from these constraints. In other words, these changes can be regarded as external (uncontrollable) parameters and imaging systems that are robust against these parameters are generally preferred. The performance of a biometric system can be improved by suppressing the influence of these uncontrollable parameters as little as possible; and it is particularly important to select an appropriate imaging system depending on the application by taking the variety and the range of these parameters into consideration.

Related Entries

- ▶ Finger Vein
- ▶ Finger Vein Feature Extraction
- ▶ Finger Vein Reader

References

1. Hitachi Central Research Laboratory, <http://www.hqrd.hitachi.co.jp/crl/>
2. Kono, M., Ueki, H., Umemura, S.: Near-infrared finger vein patterns for personal identification. *Appl.Opt.* **41**(35), 7429–7436 (2002)
3. Hitachi Information and Control Solutions, Ltd., http://www.hitachi-ics.co.jp/product/english/index_en.htm
4. Hitachi, Ltd., <http://www.hitachi.com/>
5. Hitachi Software Engineering, Co., Ltd., <http://hitachisoft.jp/English/index.html>
6. Hitachi develops finger vein authentication technology for steering wheels. Hitachi, Ltd. News release, Oct. 2007. <http://www.hitachi.com/New/cnews/071022b.html>
7. Hitachi-Omron Terminal Solutions, Corp., <http://www.hitachi-omron-ts.com/index.html>
8. Hitachi develops compact finger vein authentication technology for laptop PCs. Hitachi, Ltd. News release, Oct. 2005. (in Japanese) <http://www.hitachi.co.jp/New/cnews/month/2005/10/1003.pdf>

Finger Vein Reader

MITSUTOSHI HIMAGA
Hitachi-Omron Terminal Solutions, Corp. Tokyo,
Japan

Synonyms

Finger vein scanner; Finger vein imaging device;
Finger vein authentication device

Definition

A finger vein reader is a biometric device that comprises at least one optical imaging unit designed to capture finger vein patterns of an individual and a digital signal processor that digitizes the captured finger vein patterns to be utilized as biometric features.

Introduction

Unlike conventional biometric features such as finger print, vascular network patterns cannot be observed without using specially designed equipment and thus, it is extremely difficult to steal or duplicate the biometric information. The possibility of biometric

identification based on human finger vein patterns captured by transmitting light was indicated by Shimizu in 1992 [1]. However, it was not until Kono et al. developed a near-infrared finger vein reader prototype and demonstrated its effectiveness in 2000 that the concept became reality [2]. Kono et al. further evaluated the performance of the proposed biometric modality by using sample data collected from 678 subjects and reported very positive results in 2002 [3]. In 2007, Yanagawa et al. demonstrated the diversity of human finger vein patterns by conducting statistical analysis based on sample data collected from 506 subjects. They also proved the feasibility of using finger vein patterns for personal identification by evaluating false acceptance rates (FAR) and false rejection rates (FRR) [4]. Today, finger vein biometrics is recognized as one of the most reliable and secure biometric modalities and is applied to a variety of security systems.

Features of Finger Vein Modality

The advantages of finger vein biometrics are summarized as below:

1. Accuracy

Finger vein biometrics is one of the most accurate biometric modalities available today. A finger vein authentication device called UBReader has been certified as level 3 in the accuracy scale by the US-based International Biometric Group [5, 6]. No other biometric device has been rated at the highest possible level, level 4. The details of the evaluation results are reported in the Comparative Biometric Testing (CBT) round 6 Public Report [7].

2. Usability

Finger vein biometrics can be implemented in many forms according to the demands and requirements of the application. This flexibility makes it possible to design the hardware optimized to a specific use. For example, Hitachi-Omron's UBReader, which was primarily designed for banking application, demonstrated very high usability in terms of indices such as ► **Failure-to-Enroll Rate (FTE)** or ► **Enrollment Transaction Duration** in the CBT and achieved level 3 in the usability scale of the testing.

3. Compactness/Flexibility

Since the target imaging area of a finger vein reader is generally smaller than for other vascular

pattern biometric devices (e.g., palm vein or the back-of-the-hand vein systems), finger vein readers can be installed into a variety of devices flexibly. One of the most compact finger vein readers was not more than 19ml in volume, which made possible for laptop PCs to embed the device without changing their appearances. The short focus depth (i.e., the distance between the camera and the target) makes it easy to align the finger and, therefore, no hand-guide or handle bars, which are sometimes necessary for other hand vascular devices, are needed.

4. Small templates

The size of finger vein template is typically some hundreds of bytes per finger. This means that finger vein biometric database can be very cost-effective because it does not require a large storage system, compared with other biometrics. This feature is also preferable for systems which store templates on a server and transmit them upon request over a network. Small template size makes a big difference especially when a high-speed network is not available or the data traffic is very high.

5. Excellent image quality

Since the raw image is the very first input from which most biometric information is extracted, the image quality is largely responsible for the overall performance of the biometric system. All finger vein readers, commercially available today, utilize near infrared rays that are projected through the presented finger. The images captured by using this method (known as the "light penetration method") have very high contrast and little noise because most of ridges and wrinkles on the skin are not imaged.

6. More back-up samples

Unlike most biometric systems, finger vein biometrics allows more than two templates per person. Even in the case when one of the enrolled fingers gets injured and cannot be presented to the biometric system, it is possible to operate the system using other fingers.

The Hardware

Finger vein readers can be classified into three different groups, depending on the device used, and execution of the enrollment and authentication processes.

The finger vein readers in the first category do not have an authentication algorithm on the readers which

is known as “match-on-PC” readers. Instead, the authentication algorithm is implemented as a computer software and distributed together with the finger vein reader. The software is installed to the host PC beforehand, where the enrollment and authentication processes are executed. Since the match-on-PC finger vein readers do not need a powerful processor, the cost of the hardware is relatively low compared with the other two kinds of finger vein readers. Due to the low power consumption, most of the match-on-PC devices can be driven by the 5 volts power supplied through the universal serial bus (USB) interface, which contributes to the compactness and the portability of the device. Since the turn-around time of the authentication process is dependent on the host PC’s CPU power and the communication speed of the interface, the throughput of the entire system may vary. Although the match-on-PC readers are widely used for the purpose of logical access control (e.g., PC log-in), they are increasingly coming into use for physical access control applications.

The second category is called “match-on-device” finger vein readers. The match-on-device reader is equipped with a CPU that executes both enrollment and authentication processes inside the reader itself. The authentication algorithm is implemented in firmware and is typically encrypted when stored on a non-volatile static memory. One of the biggest advantages of this system is that all algorithms and data required for biometric authentication are enclosed in a ► **tamper-proof** casing and completely separated from the outside world. Since all biometric data and algorithms can be stored inside of the finger vein reader, the risk of hacking is minimal. Another advantage of this system is that the match-on-device finger vein readers do not require high-performance host PCs. In most cases, a low-performance CPU is enough to communicate and control the match-on-device finger vein reader, which makes it possible to integrate cost-effective systems. The data communications between the host PC and the finger vein reader are limited because no biometric data is needed to transfer and therefore no high-speed interface/network is required. The unit price of these readers tends to be higher than the match-on-PC readers; however, the match-on-device readers can be used for a wide range of applications as they are suitable for both high-security systems and low-cost systems. Typical applications of the match-on-device readers include banking systems and physical access control systems.

The third category is known as ► **match-on-card** finger vein readers. The authentication algorithm is implemented as smart card application software and securely stored onto a smart card together with biometric templates. Upon the host PC’s request, the match-on-card finger vein reader extracts the biometric feature of the presented finger and sends an authentication command to the smart card together with the features. The smart card then executes the authentication algorithm on its own CPU embedded inside and evaluates the features transmitted by the finger vein reader. After the smart card determines whether the presented finger matches with the pre-enrolled template, it transmits a response back to the host PC through the reader. One of the benefits of using the match-on-card system is its high security feature. Both the authentication algorithm and the template data are securely stored on a smart card that is inaccessible without taking validation procedures using Secure Application Module (SAM). Since these data is never transmitted outside the card, the risk of template duplication is extremely low. From a viewpoint of system administration, the risk management cost of the match-on-card system can be dramatically suppressed because the system does not need to provide protection for the template data (the card holders are responsible for their own templates, instead). Though the authentication processing time is slightly longer than other two kinds of readers (this is because the smart card CPUs are slower than the embedded CPUs or PCs), it does not make much difference especially for its primary usage, verification. For these reasons, match-on-card finger vein systems are currently the most popular biometric banking solution in Japan.

Security Features

Some finger vein readers have a security measure called ► **liveness detection**. It is very important for biometric systems in general to ensure that the enrolled biometric patterns are genuine. If a biometric device accepts any artifact mistakenly and enroll it as a genuine template, that can be used just like a normal key that can be used by anyone; if this happens, the security level of the biometric system becomes no higher than conventional keys and locks. In the actual applications, enrollment procedures typically require an administrator to be present (who will never allow users to enroll artifacts);

however, it is still beneficial to have this security measure because it is also used in the authentication procedure in order to ensure that the presented sample is from a live body. Liveness detection can be implemented by either hardware or software (or both) and there are many different methods to realize the functionality. The details of the method employed by finger vein readers are, however, not publicly available at the time of writing due to the secure nature of the functionality.

Another security feature that some finger vein readers have is the tamper-proof structure. This structure enables the system to identify that it has been tampered with, and in some cases, to disable itself when unauthorized person try to dismantle or reverse-engineer the system. This security measure is especially important when the biometric system is to be used by open public, for instance, ATMs. Just like liveness detection, the details of the tamper-proof structure are highly confidential and no finger vein manufacturer discloses the mechanism for security reasons.

Applications

- Banking transactions

Banking applications are currently the most popular application of finger vein biometrics. The first finger vein biometric ATM system was developed and introduced by Hitachi-Omron Terminal Solutions, Corp. in 2005. The biometric ATM was equipped with an open-scanning finger vein reader, as shown in Fig. 1, and adopted by one of the largest banks in Japan, Sumitomo Mitsui Banking Corporation (SMBC, Tokyo, Japan) [8] and later, was widely adopted by more than 60 financial institutions in Japan including Japan Post Bank Co., Ltd. [9, 10, 11]. According to a recent survey more than 80% of Japanese financial institutions that adopted biometric banking systems employ finger vein biometrics [12]. It is expected that more than 40,000 ATMs in Japan will be equipped with finger vein readers by the end of 2008, which will make up approximately 25% of ATMs of the country.

In typical finger vein banking systems, each account holder who wishes to have his or her biometric data enrolled visits a branch of the bank in person and enrolls two fingers at the teller counter after prescribed personal identification procedure. The templates are then stored in a smart card issued



Finger Vein Reader. **Figure 1** Finger vein reader implemented on an ATM.

by the bank, on which the matching process is executed during the authentication process (i.e., “*match-on-card*” technology). Since the matching process is executed against the two templates stored on the smart card, users can present either of the two enrolled fingers. Many of the finger vein ATM networks are connected to each other and the account holders can use their biometric bankcards at any ATM that belongs to the participating financial institutions.

- Door access control

Door access control is another popular application of the finger vein biometrics. The first commercial application of the finger vein biometrics was a door access control system called SecuaVeinAttestor® developed by Hitachi Engineering Co., Ltd. in 2002. (Please note that Hitachi Engineering Co., Ltd. reorganized its biometrics division into Hitachi Information and Control Solutions [13], Ltd. in 2006.) The door access control system is equipped with a ten-key pad, with which users type his or her ID number so that it can execute one-to-one matching (verification). It can also be used with proximity cards, which allow users to unlock the door without typing their ID numbers. In addition, a biometric door access control system has been developed that works with electric locks [14, 15]. A prototype automobile entry system using finger vein biometrics was demonstrated in the Tokyo Motor Show in 2005, which enables pre-registered users unlock the door just by holding the door handle (Fig. 2).



Finger Vein Reader. **Figure 2** Finger vein reader embedded on a door handle.

- Logical access control

Logical access control is also a popular and widely used application. Since host computers (PCs) are normally equipped with a CPU powerful enough to execute the matching process in real time, many finger vein readers for this application employ the match-on-PC architecture. Hitachi, Ltd. and Hitachi Software Engineering (Tokyo, Japan) [16] jointly developed a very compact finger vein reader for PC called PC-KCA100 in 2006. This match-on-PC finger vein reader has an application programming interface (API) based on the widely recognized international standard BioAPI 2.0, which enables it to easily communicate with many types of systems. The power consumption of PC-KCA100 is so small (less than 2.5 watts) that it can be driven by the power supplied by the USB interface only.

- Other applications

Amano Corporation (Kanagawa, Japan) [17] developed the first “time and attendance” terminal equipped with a finger vein reader called AGX250AV in 2007. This innovative terminal can store up to 1000 finger vein templates and authenticate the users without using an ID card. In addition to the convenience, AGX250AV eliminates inappropriate attendance records by impostors (this is known as “buddy punching”), which dramatically increases the reliability of the time information system. Alpha Locker System Co., Ltd. (Kanagawa, Japan) [18] developed the first finger vein biometric locker FB-BM in 2007. The biometric locker, which is aimed for public use, has some tens of doors that can be accessible by presenting a finger. FB-BM is capable of identifying a pre-enrolled finger by using one-to-many matching algorithm and does not require the users to specify which door to open before presenting their fingers.

Summary

Although finger vein biometrics is one of the latest biometric technologies, it has already established both technical and statistical feasibility. Finger vein readers have been successfully applied to a growing array of applications such as time and attendance or physical access control systems. Its banking applications remain one of the largest and the most successful set of applications for this state-of-the-art biometric modality; and it is anticipated that more than a quarter of ATMs in Japan will be equipped with finger vein readers by the end of 2008. Some financial institutions who adopted other hand vascular biometrics started to modify their systems to accept finger vein biometric data, or even replace their systems with finger vein readers. This trend is expected to continue as the number of finger vein readers increase, and it is very likely for the biometric technology to set a new standard in security applications in the very near future.

Related Entries

- ▶ [Finger Vein](#)
- ▶ [Finger Vein Feature Extraction](#)
- ▶ [Finger Vein Imaging](#)

References

1. Shimizu, K.: Optical trans-body imaging: feasibility of optical CT and functional imaging of living body. *Jpn. J. Medicina philosophica* **11**, 620–629 (1992) (in Japanese)
2. Kono, M., Ueki, H., Umemura, S.: A new method for the identification of individuals by using vein pattern matching of a finger. In: *Proceedings of the Fifth Symposium on Pattern Measurement*, pp. 9–12, Yamaguchi, Japan. (2000) (in Japanese)
3. Kono, M., Ueki, H., Umemura, S.: Near-infrared finger vein patterns for personal identification. *Appl. Opt.* **41**(35), 7429–7436 (2002)

4. Yanagawa, T., Aoki, S., Ohyama, T.: Human finger vein images are diverse and its patterns are useful for personal identification. MHF Preprint Series, MHF 2007–12, Kyushu University 21st Century COE Program, Development of Dynamic Mathematics with High Functionality (2007). <http://www2.math.kyushu-u.ac.jp/coe/report/pdf/2007-12.pdf>
5. UB Reader is developed by Hitachi-Omron Terminal Solutions, Corp. <http://www.hitachi-omron-ts.com>
6. International Biometric Group, LLC. <http://www.biometricgroup.com>
7. Theme, M., ed.: Comparative Biometric Testing Round 6 Public Report. The International Biometric Group (2006)
8. Sumitomo Mitsui Banking Corporation, <http://www.smbc.co.jp/global/index.html>
9. Japan Post Bank Co., Ltd. http://www.jp-bank.japanpost.jp/en_index.html
10. Mizuho Bank, Ltd. <http://www.mizuhobank.co.jp/english/>
11. Resona Bank, Limited. <http://www.resona-gr.co.jp/holdings/english/index.html>
12. Conducted by Hitachi-Omron, January 2008
13. Hitachi Information and Control Solutions, Ltd., http://www.hitachi-ics.co.jp/product/english/index_en.htm
14. Electric locks developed by Miwa Lock Co., Ltd. <http://www.miwalock.com/>
15. Hitachi, Ltd. <http://www.hitachi.com/>
16. Hitachi Software Engineering, Co., Ltd. <http://hitachisoft.jp/English/index.html>
17. Amano Corporation, <http://www.amano.co.jp/English/index.html>
18. Alpha Locker System Co., Ltd., <http://www.alpha-locker.com/index.html>

Finger Vein Scanner

- ▶ Finger Vein Reader

Fingermark Identification Procedure

- ▶ Fingerprint, Forensic Evidence of

Fingerprint

Fingerprint is an impression or image left on a surface by the friction skin of a finger.

- ▶ Anatomy of Friction Ridge Skin

Fingerprint Analysis

- ▶ Fingerprint Features

Fingerprint Authentication

- ▶ Fingerprint Indexing

Fingerprint Benchmark

- ▶ Fingerprint Databases and Evaluation

Fingerprint Binarization

Fingerprint binarization is the process of converting an 8-bit gray-scale fingerprint image into a 1-bit ridge image. This is virtually equivalent to thresholding. Post-processing for the binarized image, such as smoothing, is also important.

- ▶ Fingerprint Image Enhancement

Fingerprint Biometric

- ▶ Fingerprint Recognition, Overview

Fingerprint Capture

- ▶ Biometric Sample Acquisition

Fingerprint Characteristics

- Fingerprint Features

Fingerprint Classification

XUDONG JIANG

Nanyang Technological University, Nanyang Link,
Singapore

Synonyms

Fingerprint indexing; Fingerprint pre-matching; Fingerprint retrieval

Definition

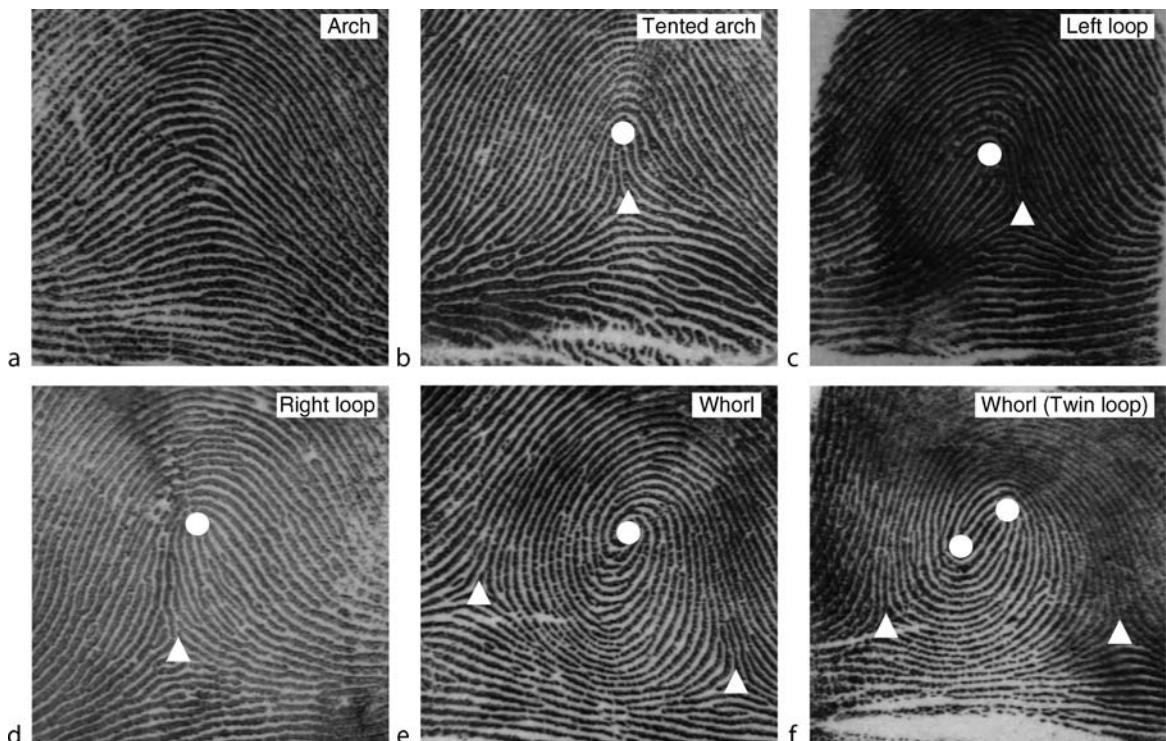
Fingerprint classification is a procedure in which fingerprints are grouped in a consistent and reliable way, such that different impressions of a same finger fall into a same group. It can be viewed as a coarse-level pre-matching procedure so that a query fingerprint needs to be further compared with only a smaller subset of fingerprints in the database belonging to the same group. It is often necessary to integrate a classification module into a fingerprint identification system to speed up the database search. A database can be partitioned into ► [human-interpretable fingerprint classes](#) based on Galton–Henry scheme or into ► [machine-generated fingerprint classes](#).

Introduction

A fingerprint recognition system captures a user's fingerprint and compares it with the information stored in a database to establish or to authenticate his/her identity. If an identity is claimed, the system compares the query fingerprint only with the template corresponding to this identity stored in the database. This one-to-one matching process is called fingerprint verification. If no identity is claimed, the system needs to compare the query fingerprint with all templates

stored in the database to establish the identity. This one-to-many matching process is called fingerprint identification. The extension of the one-to-one matching of a verification system to the one-to-many matching of an identification system increases the possibility of false positive matching. Comparing to the verification performance, both accuracy and speed may deteriorate significantly if a verification algorithm is naively extended to solve an identification problem. The performance deterioration could be very serious for large-scale identification systems as it is directly proportional to the number of fingerprints in the database [1]. This problem can be alleviated by reducing the search space of exact matching. Fingerprint classification, indexing, or retrieval techniques facilitate the reduction of the search space. They can be viewed as a coarse-level pre-matching process before further exact matching in an identification system. A query fingerprint is first compared to prototypes of the pre-specified classes, bins or clusters to find its class membership. Then, it is only necessary to compare the query fingerprint exactly with a subset of the database that has the same class membership. For example, if a database is partitioned into ten groups, and a query fingerprint is matched to two of the ten prototypes, then the identification system only needs to search two of the ten groups of the database for exact matching. This reduces the search space by fivefold if fingerprints are uniformly distributed in the ten groups.

The first rigorous scientific study on fingerprint classification was made by Sir Francis Galton in the late 1880s [2]. Classification was introduced as a means of indexing fingerprints to speed up the search in a database. Ten years later, Edward Henry refined Galton's work and introduced the concept of fingerprint "core" and "delta" points for fingerprint classification [3]. [Figure 1](#) shows the five most common classes of the Galton–Henry classification scheme where the core and delta points and the class names are shown. Henry's classification scheme constitutes the basis for most modern classification schemes. Most law enforcement agencies worldwide currently employ some variants of this Galton–Henry classification scheme. Although Galton–Henry scheme has some advantages, such as human-interpretable and rigid segmentation of a database, only a limited number of classes are applicable to the automated system. For example, most automated systems [4–8] can only classify fingerprints into five classes as shown in [Fig. 1](#). Moreover, fingerprints



Fingerprint Classification. **Figure 1** Six sample fingerprints from the five commonly used fingerprint classes (arch, tented arch, left loop, right loop, and whorl) under the Galton–Henry classification scheme where two whorl fingerprints are shown (a plain whorl and a twin loop whorl). Singular points of the fingerprints, called core and delta, are marked as *filled circles* and *triangles*, respectively. Note that fingerprints of an arch class have neither core nor delta.

are not evenly distributed in these classes and there are some ambiguous fingerprints that cannot be reliably classified even by human experts. Therefore, Galton–Henry scheme that partitions the database into human-interpretable fingerprint classes is not immune to errors and does not offer much selectivity for fingerprint searching in large databases.

In fact, it is not obligatory for an automated system to partition the database into human-interpretable fingerprint classes. In automatic fingerprint identification systems (AFIS), the objective of the classification is to reduce the search space. This objective can be accomplished by partitioning the database into machine-generated fingerprint classes in feature space as long as the classification is consistent and reliable. For example, some fingerprint index techniques [9, 10] can reduce the search space more efficiently than the Galton–Henry scheme. Continuous classification techniques [1, 11, 12] do not pre-classify the database, but

represent each fingerprint with a numerical feature vector. Given a query fingerprint, a class is formed by retrieving a portion of fingerprints from database whose feature vectors are close to that of the query fingerprint. Although these techniques can classify fingerprints into large number of classes, a query fingerprint needs to be compared with all fingerprints in the database, which could be time consuming for a large database. This problem can be circumvented by incorporating data clustering techniques in the ► [fingerprint retrieval](#) framework [12, 13].

Feature Extraction for Classification

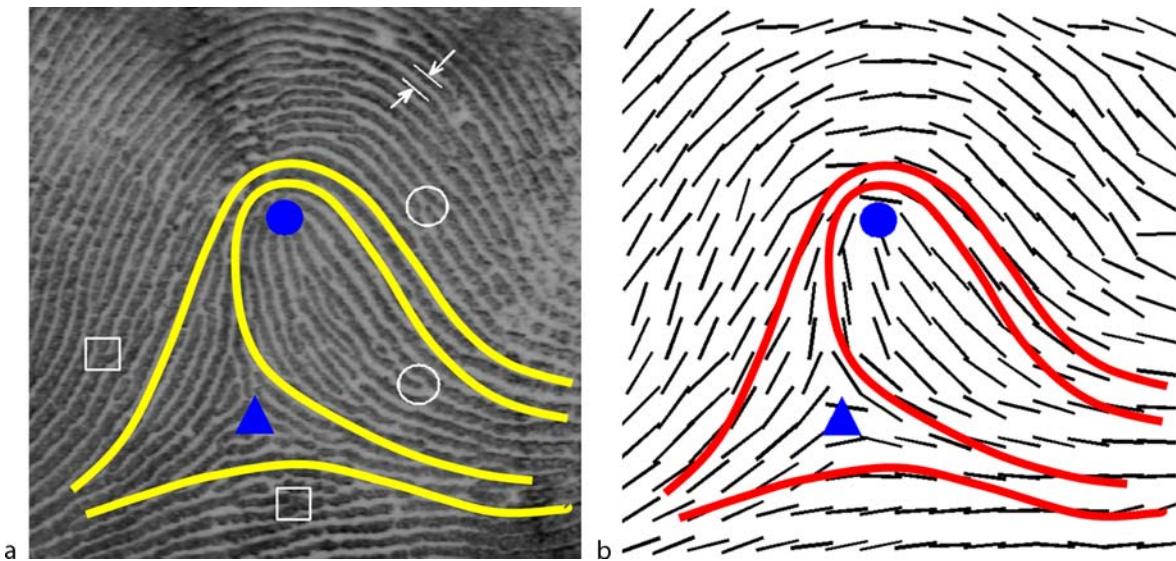
Not all measurements of a fingerprint image remain invariant for a given individual over the time of capture and can be used to discriminate between identities. The first step of fingerprint classification is to find

salient features that have low intra-class variation and high inter-class variation. Fingerprint image is an oriented texture pattern that contains ridges separated by valleys and exhibits two levels of feature as shown in Fig. 2. At the global level, the orientation field and the ridge frequency are two primitive and fundamental features. At the fine local level, the most prominent characteristics are the minutia points, where a ridge terminates or separates into two ridges.

An orientation field shown in Fig. 2b of a fingerprint shown in Fig. 2a contains information about the local dominant orientations of fingerprint ridges, from which some other features can be derived such as singular points and dominant ridge line flow as shown in Fig. 2. The dominant ridge flow is represented by a set of curves running parallel to the ridges lines but not necessarily coinciding with ridges and valleys. There are two types of singular points: core and delta points. A core point is the turning point of an inner-most ridge and a delta point is a place where two ridges running side by side diverge. Orientation field, dominant ridge flow, and singular points are useful features for classification. A local ridge frequency is the number of ridges per unit length along a

hypothetical segment orthogonal to the local ridge orientation. Its' inverse is the local ridge distance as shown in Fig. 2a. Although the local ridge distance varies across different fingers, it is difficult to serve as a reliable feature due to its high within-finger variation caused by the discontinuity of ridges and valleys and various unfavorable skin and imaging conditions. However, the average ridge distance over a fingerprint shows a stable and reliable feature and is employed in some approaches [12, 13].

Minutia points as shown in Fig. 2a are in general stable and robust to fingerprint impression conditions. They often serve as discriminative features for exact matching in most automatic fingerprint recognition systems. However, some fingerprint indexing approaches [9, 10] also use minutiae for coarse level fingerprint search. Another type of feature is the filter response of fingerprint image. Gabor filters are oriented band-pass filter with adjustable frequency, orientation, and bandwidth parameters. The responses of Gabor filters capture information of fingerprint local orientation, ridge frequency, and ridge discontinuity and hence can be used for both coarse level classification [5] and exact matching.



Fingerprint Classification. **Figure 2** A fingerprint image and its feature representation. The orientation field consisting of fingerprint local orientations is represented by short lines in (b). Core and delta points are marked in both (a) and (b) by filled circles and triangles, respectively. Two examples of ridge ending and ridge bifurcation, called minutia points, are enclosed by circles and squares in (a), respectively. An example of local ridge distance is shown by two arrows in (a). Three dominant ridge flow curves that can represent the Galton–Henry classes (here: right loop) are shown in both (a) and (b).

Classification Under Galton–Henry Scheme

Over the last four decades, many techniques have been developed for the automatic classification of fingerprints under Galton–Henry scheme, which can be coarsely assigned to one of these categories: rule-based, syntactic, structural, statistical and other approaches. While rule-based, syntactic and structural approaches are mainly used to partition the database into the human-interpretable fingerprint classes defined by Galton–Henry Scheme, statistical approaches are able to classify fingerprints into compact clusters in feature space.

The rule-based approaches codify the human expert knowledge of manual classification such as the singularity and the geometrical shape of ridge lines. It is not difficult to see from Figs. 1 and 2 that the five human-interpretable fingerprint classes can be determined by the number and location of the singular points plus some local ridge orientations. Fingerprints with neither core nor delta points are classified as arch. Whorls (plain whorl and twin loop whorl) have one or two cores and two deltas. Loops and tented arch contain only one core and one delta. Tented arch is discriminated from loops by examining the local orientations lying along the line connecting the core and delta points. The difference between these local orientations and the slope of the line is much smaller for a tented arch than loops. Left and right loops are distinguished by examining the local orientations around the core point with respect to the slope of the line [6]. Although a rule-based approach is simple and work well on rolled fingerprint with high image quality, robust and consistent detection of singular points in a poor quality fingerprint remains a difficult task. Thus, the rule-based approaches are in general sensitive to noise and cannot work on the partial fingerprint where the delta point is often missing.

A syntactic method represents a fingerprint by a sentence of a language extracted from the ridge flow or orientation field. For example, the three dominant ridge flow curves in Fig. 2 show the typical pattern of right loop. It is not difficult to see from Figs. 1 and 2 that, in general, the five human-interpretable fingerprint classes can be distinguished by such dominant ridge flow curves. In the syntactic approaches, a grammar is defined for each fingerprint class to build up sentences. Classification is performed by determining

which grammar most likely generates the sentence extracted from a query fingerprint. In general, syntactic methods tend to be robust in the presence of image noise but often require very complex grammars to struggle against the large intra-class and small inter-class variations. Complex grammars often result in unstable classification.

The structural approaches organize low-level features into higher-level structure. One approach partitions the orientation field into connected regions characterized by homogeneous local orientations [11]. For example, it is not difficult to identify some homogeneous orientation regions from the orientation field shown in Fig. 2b. A relational graph that shows the relations among these regions of a fingerprint contains discriminative information for classification. An inexact graph matching technique is exploited to compare the relational graphs with class-prototypes. As a robust and consistent partition of orientation field is not an easy task, a template-based matching is developed to guide the partitioning [11]. Another approach converts the two-dimensional fingerprint structure into one-dimensional sequence and exploits hidden Markov model for classification [8]. A set of horizontal lines across the fingerprint is used to extract a sequence of features. It captures information about the local orientations and ridge distances and thus has higher discrimination power than the orientation field alone. Since the structural approaches rely on global structural information, they can work on noisy images and are able to deal with partial fingerprints where some singular points are not available.

Statistical approaches extract a fix-size numerical feature vector from a fingerprint and exploit statistical classifiers, such as k -nearest neighbor classifiers, support vector machines and artificial neural networks. The feature vector can be constructed based on the orientation field [4, 11, 12] or the responses of Gabor filters [5]. As features extracted from different fingerprint regions show different discriminating power, some weighting schemes [4, 11, 12] or non-uniform spacing techniques [5, 13] are developed to put higher weights in more discriminative regions of fingerprint. Karhunen–Loëve (KL) transform and multi-space KL (MKL) transform [14] are also applied to reduce the dimensionality of feature vector. Statistical classifiers in general need to be trained with a fingerprint database. As Galton–Henry scheme defines the human-interpretable fingerprint classes rather than the natural

clusters of fingerprints in feature space, supervised training using fingerprint samples with known class labels is often applied. On the other hand, statistical approaches are able to classify fingerprints far beyond the Galton–Henry scheme into much more classes.

Classification with Machine-Generated Classes

The Galton–Henry Scheme does not offer much selectivity for fingerprint searching in large databases. Most automated systems [4–8] can only classify fingerprints into the five classes shown in Fig. 1 and the probabilities of the five classes are approximately 0.037, 0.029, 0.338, 0.317, and 0.279 for the arch, tented arch, left loop, right loop, and whorl, respectively [15]. The uneven distribution of these human-interpretable fingerprint classes further lowers the classification efficiency. In fact, for the application of the automated identification, it is often not obligatory to partition the database into human-interpretable fingerprint classes. Any classification scheme is in principle workable so long as different impressions of a same finger consistently fall into a same class. Instead of grouping fingerprints based on the visual appearance of fingerprint images, we can partition the database in the feature space into the machine-generated fingerprint classes, in the hope that more classes can be formed. However, there are always fingerprints located near the class boundaries regardless of how well the database is partitioned. These fingerprints are likely misclassified due to the large variations of different impressions of a same finger. To alleviate this problem, fingerprints are not pre-classified, but associated with numerical feature vectors. Given a query fingerprint, a fingerprint class is then formed by retrieving a portion of fingerprints from database whose feature vectors are similar or have small distance to that of the query fingerprint. Hence, this scheme is also called “continuous classification” [1, 11, 12].

Orientation field is often used to construct the numerical feature vector consisting of local orientations [4, 11–13]. Note that an orientation angle θ is a periodic variable with a period of 180° rather than 360° and has discontinuity at $\pm 90^\circ$ or 0° and 180° . The smallest and the largest angles in a period do not refer to two orientations far away but rather close to each other. The distance between two orientations

θ^P and θ^Q cannot be naively measured by $|\theta^P - \theta^Q|$, but rather by $\min(|\theta^P - \theta^Q|, 180^\circ - |\theta^P - \theta^Q|)$. Thus, the distance between two feature vectors cannot be computed by simple arithmetic such as Euclidean distance. To simplify the distance computation, an orientation angle θ is decomposed into two component, $\cos(2\theta)$ and $\sin(2\theta)$ [1, 11, 14] so that the similarity of two fingerprints can be measured by the convenient dot product of the two feature vectors. This also enables to put weights on different orientations, for example, $r[\cos(2\theta), \sin(2\theta)]$, where r is the weight of orientation θ . In fact, the similarity of two feature vectors can be measured by the consistency of the orientation differences. Thus, a similarity measure between two feature vectors $\mathcal{O}^P = (\theta_1^P, \theta_2^P, \dots, \theta_k^P, \dots)$ and $\mathcal{O}^Q = (\theta_1^Q, \theta_2^Q, \dots, \theta_k^Q, \dots)$ is defined by $|\sum_k r_k \exp[2j(\theta_k^P - \theta_k^Q)]| / \sum_k r_k$, where r_k are weights, $\exp[\cdot]$ is a complex exponential function and $|\cdot|$ is a magnitude operator [12, 13]. Besides the orientation field, the average ridge distance over the fingerprint is also used as an auxiliary feature in some approaches [12, 13].

Given a query fingerprint, a fingerprint class is formed by retrieving a number of fingerprints from the database whose feature vectors are nearest to that of the query fingerprint. Depending on application scenarios, different fingerprint retrieval strategies can be applied, such as a fixed distance threshold, or a fixed percentage of fingerprints in database to be retrieved, or some combination of the both [12]. In an identification system, fingerprint retrieval and exact matching can be integrated so that the retrieval threshold increases from a small value until the query fingerprint is matched with one of the retrieved templates by a matching algorithm. The threshold can increase by a fixed step or based on a fixed number of newly retrieved fingerprints. The incorporation of matching in the fingerprint retrieval may greatly improve the retrieval performance if a good matching algorithm is applied [1, 11, 12].

The continuous classification in general needs to compare the feature vector of a query fingerprint with those of all fingerprints in the database. The time consumption of fingerprint retrieval thus directly depends on the database size. For large database, the continuous classification could be time consuming. To circumvent the one-by-one exhausting comparisons of a query fingerprint with all templates, database is partitioned into clusters and hence the query fingerprint

only needs to be compared with the cluster prototypes [12, 13]. Since in general there are always some fingerprints near the cluster boundaries regardless of how well the clusters are formed, it is crucial to retrieve, instead of one, a few clusters. For the application of automated identification, this clustering based classification scheme is comparable to the Galton–Henry scheme in terms of the search speed that is independent to the database size. But the former has potential to achieve better classification accuracy and efficiency. Fingerprint database indexing [9, 10] is a closely related problem to this classification scheme. Different from the clustering based classification scheme, however, fingerprint indexing approaches [9, 10] utilize minutia points that most automated fingerprint matching algorithms rely on for the exact fingerprint comparison.

Classification Performance

The performance of a fingerprint classification system is usually measured in terms of accuracy or error rate, efficiency or penetration rate, and speed or computational complexity. The measurements of these performance indicators could be quite different on different fingerprint databases. Therefore, the performance comparison of different classification algorithms should be based on the same database. The NIST (National Institute for Standards and Technology) Special Database 4 is the most often used database for the classification performance evaluation. It contains 2,000 fingerprint pairs, uniformly distributed in the five Galton–Henry classes (see Fig. 1). Some approaches are tested on a reduced set (called Set 2), containing 1,204 fingerprints extracted from the database according to the real distribution of fingerprints.

The error rate is computed as the ratio of the number of misclassified fingerprints to the total number of samples in the test set. For a Galton–Henry classification system, a fingerprint is misclassified if it is placed in a class different from the human assigned one as the true class membership of a fingerprint is determined by human experts. For a system that is based on the machine-generated fingerprint classes, a query fingerprint is misclassified if the retrieved subset from database contains no fingerprint originating from the same finger as that of the query fingerprint. The error rate of a classification system in general

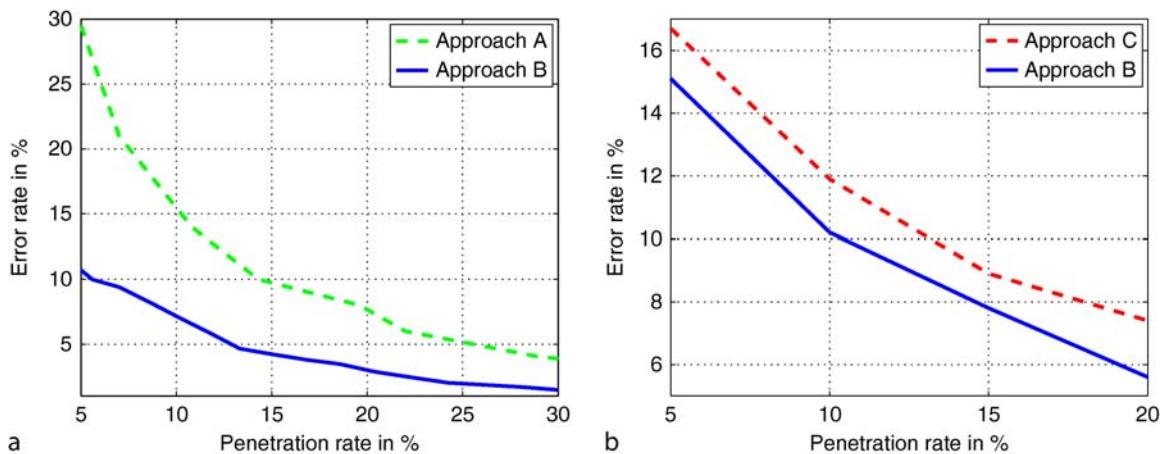
should be reported as a function of the penetration rate that is a performance indicator of the classification efficiency.

The classification efficiency is measured by the penetration rate defined as the average ratio of the number of fingerprints in a class to the total number of samples in the database [1, 11, 12]. If q_i represents the ratio of the number of fingerprints in class i to the total number of samples in database and p_i is the class occurrence probability, the penetration rate is calculated by $\sum_i p_i q_i$. For example, for the five Galton–Henry classes with the occurrence probabilities of 0.037, 0.029, 0.338, 0.317, and 0.279, respectively, the penetration rate of a error free classifier ($q_i = p_i$) is 0.2948, which lies between the penetration rates of 0.25 and 0.3333 for the four and three equal-sized classes, respectively.

Figure 3 illustrates the tradeoff between the classification error rate and the penetration rate of three techniques tested on two data sets. Obviously, lower classification error rate can be achieved at higher penetration rate. As higher classification accuracy and efficiency are measured by lower error rate and lower penetration rate, respectively, a lower curve indicates a better classification performance. Table 1 shows the classification results of some Galton–Henry scheme based approaches (the first seven rows) and the clustering based approach (the last two rows). All results are obtained from NIST Special Database 4. Some approaches are tested on the Set 2 and some approaches are tested on the second half of the database because they use the first half of the database to train their programs. Classification performance on the real distributed fingerprints is also resembled by the “weighted classes” shown in the third and the fifth columns. Note that Fig. 3 and Table 1 do not serve as a direct comparison between different algorithms due to different experimental settings and rate calculations. More information about the classification performances of these approaches can be found in the respective references [4–8, 10–12, 14].

Summary

The development of automatic fingerprint identification system for large database is a challenging task due to both accuracy and speed issues. Fingerprint classification as a tool to narrow down the searching space of



Fingerprint Classification. **Figure 3** Classification error rate against penetration rate: (a) approach A in [11] and B in [12] tested on the the NIST Special Database 4 Set 2 containing 1,204 fingerprint pairs; (b) approach B in [12] and C in [10] tested on the the second half of the NIST Special Database 4 containing 1,000 fingerprint pairs.

Fingerprint Classification. **Table 1** Classification error rates in % on NIST Special Database 4 of some Galton–Henry scheme based approaches (the first seven rows) and the clustering based approach (the last two rows)

Source	Five classes P.R. = 20%	Five weighted classes P.R. = 29.5%	Four classes P.R. = 28%	Four weighted classes P.R. = 29.7%	Test set
Candela et al. [4]	–	–	11.4	6.1	Second half
Karu and Jain [6]	14.6	11.9	8.6	9.4	Whole
Jain et al. [5]	10	7.0	5.2	–	Second half
Cappelli et al. [11]	–	12.9	–	–	Set 2
Cappelli et al. [14]	7.9	6.5	5.5	–	Second half
Senior [8]	–	–	–	5.1	Second half
Park and Park [7]	9.3	–	6.0	–	Whole
Jiang et al. [12]	5.3	3.3	3.5	3.2	Whole
Jiang et al. [12]	4.7	2.9	3.2	2.8	Set 2

The penetration rate is shown by the value of P.R. In the columns of “weighted classes”, error rates of different classes are weighted by the class occurrence probabilities in the calculation of the total error rate

exact matching can alleviate these difficulties. A lot of different techniques have been developed to automate the Galton–Henry classification scheme, thanks to its human-interpretability and rigid segmentation of a database. However, the Galton–Henry classification scheme that partitions the database into human-interpretable fingerprint classes does not reduce the search space significantly. The database partition based on the machine-generated fingerprint classes seems to be a more promising alternative for efficient reduction of the search space. For a classification

system that requires high accuracy, a fingerprint rejection engine can be applied to exclude poor quality fingerprints at a price of lower classification efficiency. Further research efforts are necessary to improve the classification performance.

Related Entries

- ▶ Fingerprint Features
- ▶ Fingerprint Indexing

- ▶ Fingerprint Recognition Overview
- ▶ Identification and Authentication

References

1. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: *Handbook of Fingerprint Recognition*. Springer-Verlag, New York (2003)
2. Galton, F.: *Finger Prints*. McMillan, London (1892)
3. Henry, E.: *Classification and Uses of Finger Prints*. Routledge, London (1900)
4. Candela, G.T., Grother, P.J., Watson, C.I., Wilkinson, R.A., Wilson, C.L.: PCASYS – A pattern-level classification automation system for fingerprints. Technique Report: NIST TR 5647 (1995)
5. Jain, A.K., Prabhakar, S., Hong, L.: A multichannel approach to fingerprint classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(4), 348–359 (1999)
6. Karu, K., Jain, A.K.: Fingerprint classification. *Pattern Recognit.* **29**(3), 389–404 (1996)
7. Park, C.H., Park, H.: Fingerprint classification using fast fourier transform and nonlinear discriminant analysis. *Pattern Recognit.* **38**(4), 495–503 (2005)
8. Senior, A.: A combination fingerprint classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(10), 1165–1174 (2001)
9. Bhanu, B., Tan, X.: Fingerprint indexing based on novel features of minutiae triplets. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(5), 616–622 (2003)
10. Tan, X., Bhanu, B., Lin, Y.: Fingerprint identification: Classification vs. indexing. In: Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance, pp. 151–156. Miami, Florida (2003)
11. Cappelli, R., Lumini, A., Maio, D., Maltoni, D.: Fingerprint classification by directional image partitioning. *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(5), 402–421 (1999)
12. Jiang, X.D., Liu, M., Kot, A.: Fingerprint retrieval for identification. *IEEE Trans. Inf. Forensics Secur.* **1**(4), 532–542 (2006)
13. Liu, M., Jiang, X.D., Kot, A.: Efficient fingerprint search based on database clustering. *Pattern Recognit.* **40**(6), 1793–1803 (2007)
14. Cappelli, R., Maio, D., Maltoni, D.: Fingerprint classification based on multi-space KL. In: Proceedings of Workshop on Automatic Identification Advanced Technologies, pp. 117–120 (1999)
15. Wilson, C.L., Candela, G.T., Watson, C.I.: Neural network fingerprin classification. *J. Artif. Neural Networks* **1**(2), 203–228 (1993)

Fingerprint Comparing

- ▶ Fingerprint Matching, Automatic

Fingerprint Compression

NIGEL M. ALLINSON

Department of Electronic and Electrical Engineering,
University of Sheffield, Sheffield, UK

Synonym

Fingerprint Image Compression

Definition

Image files can be reduced in size by exploiting either more optimized data representation and not compromising the faithful recovery of the source image - lossless compression, or permitting recovery to within some distortion criteria - lossy compression. Fingerprint images are relatively large detailed images, and their compression can alleviate operational problems of transmission and storage.

Introduction

Fingerprint images, whether prints obtained directly from live subjects or forensically recovered latents, are normally recorded at 500 dots or pixels per inch (ppi) resolution with an 8-bit grayscale, though there is an increasing tendency to use a higher resolution of 1,000 ppi that permits accurate rendering of individual sweat pores along the ridge lines. A single digit print has a minimum area of about 20 mm × 15 mm, which yields a raw image of about 120 kB; while a tenprint record card (full set of individual slaps, rolls, and palm prints) requires several 10 **MB of storage. As national Automatic Fingerprint Identification Systems (AFISs) can contain tens of millions of individual record cards, storage requirements can be easily in excess of 100 TB [1]. It was the rapid rise in storage requirements for developing AFIS installations that drove the need for effective compression of reference fingerprint images. With the availability of low-cost, very high-capacity mass memory, this requirement may not be so clear today as it was in the early 1990s. Perhaps of greater importance is the need to transmit, over restricted bandwidth channels, both reference prints and latents recovered from crime and other scenes to remote

locations. Clearly, the transfer of images to a remote AFIS or between AFISs requires agreed standards for image compression that do not adversely affect the usage of fingerprints as a reliable and robust biometric.

Mainstream image compression aims to meet both the requirements of reducing storage requirements and enabling faster transmission. Significant compression is possible due to the limited acuity of the human visual system especially our low sensitivity in detecting low-contrast spatially-fine detail, the limited spatial resolution of electronic displays and some printing processes, and that humans are often tolerant of some degree of visible distortion. The normal image compression standards have been developed (1) to provide satisfactory reductions in memory requirements for all types of scenes, whereas fingerprints are a very restricted type of image; and (2) to provide an acceptable viewing experience for a viewer, whereas fingerprint image may be studied in great detail by an expert examiner and may be submitted to extensive automatic processing on AFIS systems.

Fingerprint Image Standards

The typical ridge-valley period is approximately 500 μm with width of the ridge varying from about 100 - 300 μm . So a minimum image sampling frequency of 200–300 ppi would be sufficient to record unambiguously the friction skin details including local features or minutiae. FBI-compliant fingerprint scanners, and the resultant images, are specified at 500 ppi. The grey-scale resolution of some commercial fingerprint scanners is limited to only 2–3 bits though there is a general recognition that simple binary images, that is 1 bit deep, are unsatisfactory. For images to have the appearance of a continuous grey-scale approximately a minimum of 100 discrete levels are required, so current standards specify 8 bit or 1 byte deep grey-scale resolution (i.e., 256 possible levels). Sweat pores are smaller in diameter than a ridge width with an effective diameter that depends on whether the pore is open or closed. For the capture of pore structures, an image resolution of 1,000 ppi has been proposed [2] and is becoming to be accepted by the fingerprint community. These standards are embodied in current national and international standards for the data formats for the interchange of fingerprint information. These standards also recommend that the overall size of an

image should range from $406 \times 381 \text{ mm}$ ($800 \times 750 \text{ pixels}$) for a single digit to $1,397 \times 2,032 \text{ mm}$ ($2,750 \times 4,000 \text{ pixels}$) for a complete palm print. This translates to image sizes ranging from about 0.5 MB to 10.5 MB for 500 ppi images and, of course, four times larger for 1,000 ppi images.

Image Compression

Image compression can be delineated into lossless and lossy coding schemes. The former referring to an encoding process that permits the original image to be retrieved without any degradation. Methods include Run-Length-Encoding where a consecutive row of three or more pixels with identical grey-scale or color values are represented by a two-byte pair. This forms part of several well-known image file standards such as TIFF (Tagged Image Format File) and PCX (PC Paint-brush Exchange). Another approach utilizes entropy coding which assigns codes to grey-scale or color values so that code lengths match with the inverse probabilities of these values. Reductions in storage requirements are usually very modest – typically less than 2:1 compression – and will not be discussed further. However, some AFIS installations do employ lossless compression for their archived reference images.

Lossy compression means that it is possible only to recover the original image to within some distortion criteria. The normal criterion for the acceptability is based on the non-visibility of coding artifacts under normal viewing conditions or, at least, the acceptability of these artifacts. For fingerprint images, the criteria need to include the absence of any artifacts that could subsequently interfere with future processing and feature recognition – either by AFIS algorithms or by a human expert. With inappropriate compression, it is more likely that legitimate queries will not be matched that is the FNMR or FRR, depending on the application, will increase. Several approaches have been explored for lossy compression but the dominant technique is based upon transforming an image from the spatial domain to a second domain which is based on spatial frequencies. Such an approach exploits the non-random distribution of the spatial frequencies in the localized objects that make up an image and the non-uniform sensitivity to differing frequencies by our visual system. The basic structure of a transform-based image coder is illustrated in Fig. 1. The decoder, which



Fingerprint Compression. **Figure 1** Overall structure of generic transform-based image coder.

recovers the best approximation of the original image from the transformed one, is essentially the reverse process. The source encoder employs a (usually) linear transform such as the Discrete Fourier Transform, Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT) to convert the entire image or some region of it into the different representational domain. The quantizer reduces the number of bits employed to represent the coefficients of the previous transform. This loss in precision, a many-to-few mapping, is the major source of compression in the overall encoding process. Quantization may be performed on individual transform coefficients, termed Scalar Quantization; or on a group of coefficients, termed Vector Quantization. As there are usually some correlation between consecutive coefficients which can be usefully exploited, vector quantization is generally more efficient than the scalar. Small valued coefficients, below a predefined threshold value, are ignored. Quantizers may uniformly reduce the precision of coefficients regardless of their magnitude or, more likely, they implement a non-uniform approximation by giving greater weight to higher valued coefficients. The final stage is an entropy encoder which losslessly compresses the quantized coefficients to yield a smaller output code stream. Typical methods include Huffman and Arithmetic coding with both being variable-length coding schemes but the former applied to individual coefficients while the latter is applied to a group of coefficients. Image compression possesses an extensive literature and a useful introduction is provided in [3].

The most common image compression is the JPEG (Joint Photographic Expert Group) standard which is DCT-based; and only the baseline encoder will be discussed here that sequentially compresses a stream of 8×8 pixel blocks of the image. Each block progresses through each processing step to yield a compressed output data stream. As adjacent image pixels are highly correlated, the forward DCT is the basis for achieving data compression by focusing most of the energy into the lower spatial frequency bands. The DCT causes no loss to the source image but simply transforms it into a domain where they can be efficiently encoded. Each of the 64 DCT coefficients is uniformly quantized

according to a 64-element quantization table, which takes into account the falling sensitivity of the human eye to fine spatial details. After quantization, the quantized coefficients are ordered in a zig-zag sequence to assist the entropy encoding by placing low-frequency non-zero coefficients before high-frequency coefficients. The DC coefficient, which contains a significant fraction of the total image energy, is differently encoded. Decoding is essentially the reverse process. JPEG compression is efficient and simple to implement especially in dedicated hardware. Good compression rates can be achieved with little loss of perceived fidelity for naturalistic scenes up to 20:1 or 30:1 compression ratios. The use of 8×8 pixel blocks does at higher compression ratios create objectionable "blocking artifacts" especially in regions of low image contrast.

The basis function for JPEG is a discrete set of orthogonal cosine waves. Such sinusoidal waves are continuous in the spatial domain and are, in some sense, artificially truncated. There are numerous possible basis functions – some with limited support, that is they possess only a non-zero value for a limited interval. Wavelets are such a function which are defined over a limited distance and possess a zero average. From a single prototype wavelet function, the basis set is defined by a series of dilations and contractions of the prototype. Wavelets are a group of mathematical functions, of which the earliest example is the well-known Gabor function. These functions can be approximated as discrete filter structures. The variety of wavelet scales can be achieved efficiently using a cascade of high and low pass filters that decompose the image into several subbands, with each subband possessing optimal filter coefficients to match the image statistics for that band. Different numbers of subbands and their scope (bandwidth, orientation, etc), termed decomposition trees, are possible; and details of these, and other aspects of wavelet compression, are beyond the scope of this essay and the reader is referred to [4, 5]. As the wavelet transform is applied to the entire image and basis functions can overlap, there are no blocking artifacts. The type of artifact visible in highly compressed images is now low-level "ringing" around high-contrast edges. Wavelets, because of their local support, mimic the different

scales of receptive fields found in the human visual system and so produce visually more appealing images compared to JPEG images compressed to the same degree. JPEG, with its reliance on coding small blocks, is limited to moderate compression ratios; while DWT-based coding provides significant improvement in picture quality up to compression ratios of 70:1–100:1 for naturalistic scenes.

Fingerprint images are a very constrained class of image and as such it is reasonable to expect that more optimum forms of compression exist than provided by the standard methods. They are also exposed to more detailed scrutiny than most other images and they are subjected to extensive image processing and pattern recognition algorithms when submitted to an AFIS system. Though many quality metrics can be used to quantify the distortion introduced in lossy compression ranging from generic measures such as Peak-Signal-to-Noise (PSNR) to those developed specifically for fingerprints (such as the Image Quality Metric (IQM) [6]). PSNR, for 8-bit images, is defined as:

$$PSNR = 20 \log_{10} \left(\frac{255}{e_{mse}} \right) \quad (1)$$

where the mean square error (e_{mse}) is given by

$$e_{mse} = \frac{1}{MN} \sum_{m=1}^{M-1} \sum_{n=1}^{N-1} [u(m, n) - v(m, n)]^2 \quad (2)$$

where $u(\bullet)$ and $v(\bullet)$ are the original and compressed images respectively – each of size $M \times N$ pixels. Higher PSNR means less distortion with no distortion equating to a $PSNR = 48.13$ dB. It is a useful metric in comparing similar image types. However, the "ground truth" for any compression scheme is the effect it has on the ability of fingerprint experts to make the same decisions as for the corresponding uncompressed image and an AFIS system to recover the correct match (or, when searching for matches to a latent, to rank consistently the most likely tenprint candidates).

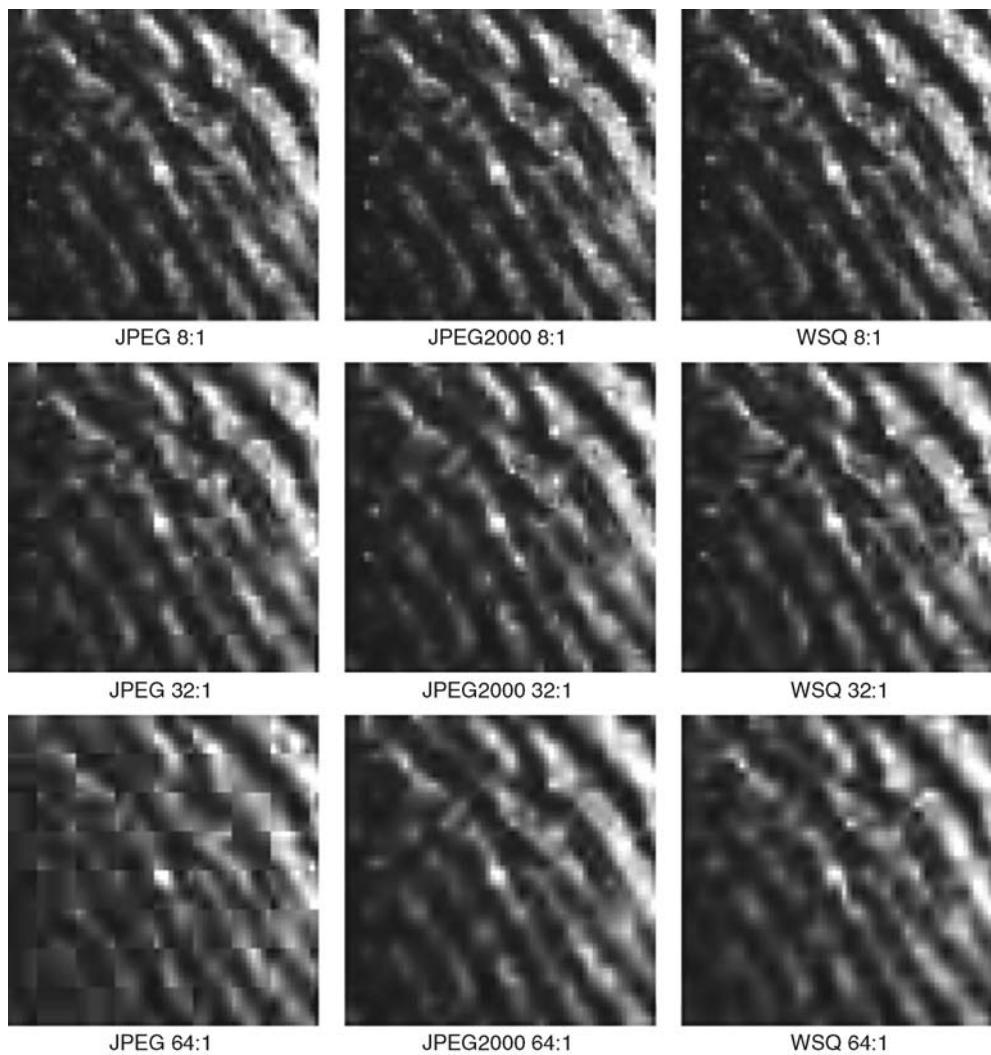
Common Fingerprint Compression Methods

Three main compression methods have been applied for the storage, transmission and, display of fingerprints namely, JPEG, WSQ, and JPEG2000. WSQ (Wavelet

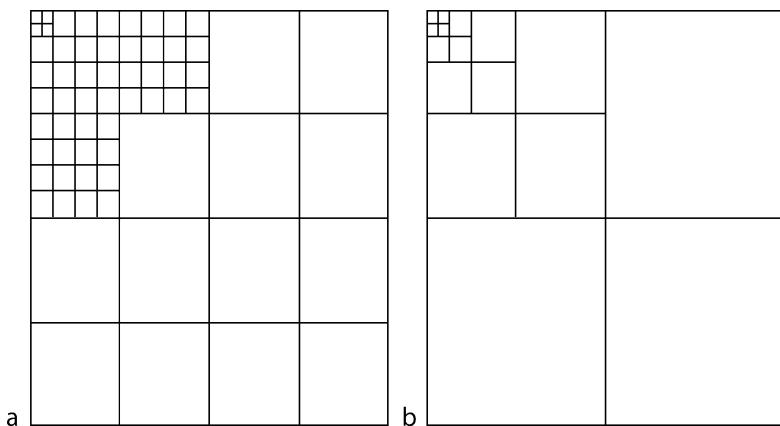
Scalar Quantization) was developed, by the FBI in association with Los Alamos National Laboratory and NIST, specifically to reduce the media storage requirements of the FBI's expanding AFIS facility by providing lossy compression over the range 10:1 to 20:1. It has become an accepted standard for 500 ppi image storage and transmission. A set of typical set of compressed image for a latent fingerprint for these three methods is shown in Fig. 2.

Though the unsuitability of JPEG compression has been known for some time [7] as it suffers from visible blocking artifacts and loss of fine details (e.g., ridge pores) even at relatively low compression rates, it is still employed in some systems. Wide-area AFIS systems with connections to remote terminals often employ such JPEG compression to return reference tenprint images of potential matches to bureaux. For example, the UK national AFIS (*Ident 1*) displays images of potential tenprint matches as 12:1 JPEGs. It is possible for examiners to discern deterioration in such images.

JPEG2000 [8] is a relatively new standard for general-purpose image compression which attempts to address the limitations of JPEG as well as incorporating many other functions. It offers both lossless and lossy compression, provides a tiled representation of images at multiple resolutions, nonuniform compression to preserve greater detail in some region of interest, and embedded metadata and security functions within the image file. Both WSQ and JPEG2000 schemes are DWT-based, but with major differences in the form of the decomposition tree, quantization, and entropy coding employed. The WSQ uses the Daubechies (9,7) filter [9] to perform the DWT and the same filter is the default for the lossy JPEG2000 transform. The JPEG2000 employs a dyadic decomposition tree, while WSQ employs a fixed structure with 64 subbands (Fig. 3). The greater decomposition structure of WSQ may enhance compression as it approximates to orthonormalization and be better suited to the high spatial frequency content of fingerprints over more general imagery. The decomposition structure influences the number and length of the zero coding runs and so enhances compression, while the bit-plane scanning order of JPEG2000 permits finer control to achieve an arbitrarily specified compression rate. Both schemes use scalar quantization with JPEG2000 having the quantization step varying in response to the dynamic range of the respective subband. While for WSQ, all quantizer steps are uniform except for a



Fingerprint Compression. **Figure 2** Comparative example of compressed image of 64×64 pixel region of latent.



Fingerprint Compression. **Figure 3** Schematics of decomposition trees, (a) WSQ and (b) JPEG2000 (default).

lengthened middle interval. For the last coding step, WSQ employs Huffman entropy coding while JPEG2000 uses scalar arithmetic or trellis-coded quantization. Much of the flexibility and power of the JPEG2000 comes from the Embedded Block Coding with Optimized Truncation (EBCOT) algorithm. Wavelet coefficients from small blocks of the image are processed using the EBCOT algorithm which adapts the quantizer based on the statistics of the source image at the bit-plane level. These blocks are tiled with integrated header information concerning with coding details, and further quantization can be performed on the final bit-stream. The flexibility of this final stage in JPEG2000 encoding greatly assists in providing more optimal compression. WSQ employs a much simpler approach and coding details is calculated for each image and the coefficients included in the file header.

Fingerprint Compression Performance

WSQ has proved to be satisfactory at compressing fingerprint images by factors up to about 20, though the original requirement was to compress to 0.75 bpp (i.e., 10.7:1 compression). Watson and Wilson [10] report that experiments using WSQ compressed images with three different matching systems under the conditions that a FAR is 0.001 is maintained and the maximum reduction in the TAR of less than 0.01 is permitted, then there is little effect on performance for compression ratios less than 20:1. A few studies have compared the relative merits of WSQ and JPEG2000. Figueroa-Villanueva et al. [11] showed a significant improvement for JPEG2000 over WSQ at 0.75 bpp compression in terms of PSNR and Receiver Operating Curves (ROCs) for different sources namely, capacitive sensor, optical sensor, and scanned inked prints. A study of JPEG2000 and WSQ interoperability [12] concluded that JPEG2000 produced a slightly lower quality reconstructed image compared to WSQ for the same file size. Most studies have focused on coding high-quality inked prints or live print capture from various sources, and not on poorer quality latents. One study that involved latents and performance on an operational AFIS system [13] concluded that for compression ratios less than 32:1, JPEG2000 consistently produced higher identification rates than WSQ. There was also strong indication that moderate degrees of compression facilitated improved identification rates

under normal operating procedures than uncom-pressed latent images.

Current national and international standards [14, 15] recommend that WSQ encoding is used for 500 ppi fingerprint images with compression limited to 15:1; but for images with resolutions greater than 500 ppi, 15:1 JPEG2000 should be employed. The UK national fingerprint system permits latents to be transmitted and submitted to *Ident 1* at 15:1 JPEG2000.

Conclusions

It may appear that satisfactory standards exist for compressing fingerprint images, certainly for the normal operational requirements associated with the effective transmission and storage of reference and livescan images. General enhancements in providing lower-cost, higher-capacity mass storage and increased bandwidth across both fixed and wireless data networks will reduce the pressure to develop new compression standards. However, fingerprints are an unusual and fairly well-defined class of image. Compression schemes such as JPEG2000 have been developed to cope well for a very wide variety of imagery and WSQ, though based on a principled consideration of the statistical properties of fingerprints, was developed prior to many more general advances in image compression. Recent proposals for improved fingerprint image compression are generally based on wavelet transformations but with more effective decomposition trees, optimized filter structures and coefficients through the use of genetic algorithm optimization, and vector quantization. The relationship between image enhancement and compression is not fully understood. There are suggestions that the filtering that occurs during compression may be advantageous in increasing identification, especially for latents. This is an avenue that needs to be explored further.

Related Entries

- ▶ [Fingerprint Image Enhancement](#)
- ▶ [Fingerprint Matching, Automatic](#)
- ▶ [Fingerprint Matching, Manual](#)
- ▶ [Fingerprint Recognition, Overview](#)

References

1. Komarinski, P.: Automated Fingerprint Identification Systems (AFIS). Elsevier Science Technology (2005)
2. Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST): Friction Ridge Digital Imaging Guidelines, 1 edn. (2001). URL <http://www.theiai.org/guidelines/swgfast/>
3. Ghanbari, M.: Image compression to advanced video coding. In: Standard Codecs, Telecommunications Series 49. Institution Electrical Engineers (2003)
4. Burrus, C., Guo, H., Gopinath, R.: Introduction to Wavelets and Wavelet Transforms: A Primer. Prentice Hall (1997)
5. Vetterli, M., Kovacevic, J.: Wavelets and subband coding. Prentice-Hall, Inc. (1995)
6. Nill, N.: Image quality evaluation - image quality of fingerprint (IQF). Tech. rep., The MITRE Corporation (2007)
7. Hopper, T., Preston, F.: Compression of grey-scale fingerprint images. In: Proceedings DCC '92. Data Compression Conference, pp. 309–318 (1992). DOI 10.1109/DCC.1992.227450
8. Skodras, A., Christopoulos, C., Ebrahimi, T.: The JPEG 2000 still image compression standard. IEEE Signal Process Mag **18**(5), 36–58 (2001)
9. Antonini, M., Barlaud, M., Mathieu, P., Daubechies, I.: Image coding using wavelet transform. IEEE Trans Image Process **1**(2), 205–220 (1992)
10. Watson, C., Wilson, C.: Effect of image size and compression on one-to-one fingerprint matching. Tech. rep., National Institute of Standards and Technology (2005)
11. Figueroa-Villanueva, M., Ratha, N., Bolle, R.: A comparative performance analysis of JPEG 2000 vs. WSQ for fingerprint image compression. In: Audio- and Video-Based Biometric Person Authentication, Lecture Notes in Computer Science, p. 1059. Springer Berlin/Heidelberg (2003)
12. Lepley, M.: JPEG 2000 and WSQ image compression interoperability. Tech. rep., The MITRE Corporation (2001)
13. Allinson, N.M., Sivarajah, J., Gledhill, I., Carling, M., Allinson, L.J.: Robust wireless transmission of compressed latent fingerprint images. **IEEE Transactions on Information Forensics and Security **2**(3), 331–340 (2007). DOI 10.1109/TIFS.2007.902684
14. American National Standards Institute: American national standard for information systems - data format for the interchange of fingerprint, facial, & other biometric information. URL <http://fingerprint.nist.gov/standard/>
15. International Organization for Standardization: Information technology - biometric data interchange formats - part 4: Finger image data. URL <http://www.iso.org/>

Fingerprint Contrast Enhancement

- Fingerprint Image Enhancement

Fingerprint Corpora

- Fingerprint Databases and Evaluation

Fingerprint Data Interchange Format

- Finger Data Interchange Format, Standardization

Fingerprint Databases and Evaluation

FERNANDO ALONSO-FERNANDEZ, JULIAN FIERREZ
 Biometric Recognition Group – ATVS,
 Escuela Politecnica Superior,
 Universidad Autonoma de Madrid,
 Campus de Cantoblanco, Madrid 28049, Spain

Synonyms

Fingerprint benchmark; Fingerprint corpora

Definition

Fingerprint databases are structured collections of fingerprint data mainly used for either evaluation or operational recognition purposes.

The fingerprints in databases for evaluation are usually detached from the identity of the corresponding individuals, are publicly available for research purposes, and usually consist of raw fingerprint images acquired with live-scan sensors or digitized from inked fingerprint impressions on paper. These databases are the basis for research in automatic fingerprint recognition, and together with specific experimental protocols, are the basis for a number of technology evaluations and benchmarks. This is the type of fingerprint databases further developed here.

On the other hand, fingerprint databases for operational recognition are typically proprietary, usually incorporate personal information about the enrolled people together with the fingerprint data, and can incorporate either raw fingerprint image data or some form of distinctive fingerprint descriptors such as minutiae templates. These fingerprint databases represent one of the modules in operational automated fingerprint recognition systems, and will not be addressed here.

Fingerprint Databases for Evaluation

Among all biometric techniques, fingerprint recognition is the most widespread in personal identification due to its permanence and uniqueness [1]. Fingerprints are being increasingly used not only in forensic investigations, but also in a large number of convenience applications, such as access control or online identification [2].

The growth that the field has experienced over the past two decades has led to the appearance of increasing numbers of biometric databases for research and evaluation purposes, either ► **monomodal** (one biometric trait sensed) or ► **multimodal** (two or more biometric traits sensed). Previous to the databases acquired within the framework of the International Fingerprint Verification Competition series, the only large, publicly available datasets were the NIST databases [3]. However, these databases were not well suited for the evaluation of algorithms operating with live-scan images [1] and will not be described here. In this section, the authors present some of the most popular publicly available biometric databases, either monomodal or multimodal, that include the fingerprint trait acquired with ► **live-scan sensors**.

FVC Databases

Four international Fingerprint Verification Competitions (FVC) have been organized in 2000, 2002, 2004 and 2006 [4, 5, 6, 7]. For each competition, four databases were acquired using three different sensors and the SFinGE synthetic generator [1]. Each database has 110 fingers (150 in FVC2006) with eight impressions per finger (12 in FVC2006), resulting in 880 impressions (1,800 in FVC2006). In the four competitions, the SFinGe synthetic generator was tuned to

simulate the main perturbations introduced in the acquisition of the three real databases.

1. In FVC2000 [4], the acquisition conditions were different for each database (e.g., interleaving/not interleaving the acquisition of different fingers, periodical cleaning/no cleaning of the sensor). For all the databases, no care was taken to assure a minimum quality of the fingerprints; in addition, a maximum rotation and a non-null overlapping area were assured for impressions from the same finger.
2. In FVC2002 [5], the acquisition conditions were the same for each database: interleaved acquisition of different fingers to maximize differences in finger placement, no care was taken in assuring a minimum quality of the fingerprints and the sensors were not periodically cleaned. During some sessions, individuals were asked to: (1) exaggerate displacement or rotation or, (2) have their fingers dried or moistened.
3. The FVC2004 databases [6] were collected with the aim of creating a more difficult benchmark because, in FVC2002, top algorithms achieved accuracies close to 100% [6]. Therefore, more intra-class variation was introduced. During the different sessions, individuals were asked to: (1) put the finger at slightly different vertical position, (2) apply low or high pressure against the sensor, (3) exaggerate skin distortion and rotation, and (4) have their fingers dried or moistened. No care was taken to assure a minimum quality of the fingerprints and the sensors were not periodically cleaned. Also, the acquisition of different fingers were interleaved to maximize differences in finger placement. Effects of quality degradation in fingerprint images can be observed in Fig. 1.
4. For the 2006 edition [7], no deliberate difficulties were introduced in the acquisition as it was done in the previous editions (such as exaggerated distortion, large amounts of rotation and displacement, wet/dry impressions, etc.), but the population was more heterogeneous, including manual workers and elderly people. Also, no constraints were enforced to guarantee a minimum quality in the acquired images and the final datasets were selected from a larger database (the BioSec multimodal database [8]) by choosing the most difficult fingers according to a quality index, to make the benchmark sufficiently difficult for an evaluation.



Fingerprint Databases and Evaluation. **Figure 1** Examples of quality degradation in fingerprint images due to factors like low/high pressure, dryness/moisture, dirt, etc.

BIOMET Multimodal Database

Five different biometric modalities are present in the BIOMET database [9]: audio, face image, hand image, fingerprint and signature. This database was designed with the additional goal of including unusual sensors (face images captured with an infrared camera and with a 3D acquisition system). The database consists of three different acquisition sessions. The number of individuals participating to the collection of the database was 130 for the first session, 106 for the second, and 91 for the last one, resulting in 91 individuals who completed the whole acquisition process. For fingerprint acquisition, an optical and a capacitive sensor were used. During the first acquisition campaign, only the optical sensor was used, whereas both the optical and capacitive sensors were employed for the second and third campaigns. The total number of available fingerprints per sensor in the BIOMET database is 6 for the middle and index fingers of each contributor.

MCYT Bimodal Database

A large biometric database acquisition process was launched in 2001 by four Spanish academic institutions within the MCYT project [10]. The MCYT database includes ten-print acquisition (MCYT Fingerprint subcorpus) and on-line signature (MCYT Signature subcorpus) samples of each individual enrolled in the database. A total of 330 individuals were acquired in the four institutions participating in the MCYT project. Regarding the MCYT Fingerprint subcorpus, for each individual, 12 samples of each finger were acquired using an optical and a capacitive sensor under different control conditions. The MCYT database has been extended

with the comprehensive BiosecurID multimodal database [11], which includes 8 different biometric traits from 400 donors collected in 4 sessions separated in time.

BioSec Multimodal Database

BioSec was an Integrated Project of the Sixth European Framework Programme which involved over 20 partners from nine European countries. The goal of BioSec was to leverage the integration of biometrics in a wide spectrum of everyday's applications. One of the activities within BioSec was the acquisition of a multimodal database. This database was acquired at four different European sites and includes face, speech, fingerprint and iris recordings. The baseline corpus [8] comprises 200 subjects with two acquisition sessions per subject. The extended version of the BioSec database comprises 250 subjects with four sessions per subject (about 1 month between sessions). Each subject provided in each session four samples of each of four fingers (left and right index and middle). Fingerprints were acquired using three different sensors. Some example images are shown in Fig. 2.

BioSecure Multimodal Database

The acquisition of the BioSecure Multimodal Database (BMDB) was jointly conducted by 11 European institutions participating in the BioSecure Network of Excellence. [11] The BMDB is comprised of three different datasets [12], namely:

1. *Data Set 1 (DS1)*, acquired over the Internet under unsupervised conditions (i.e., connecting to an



Fingerprint Databases and Evaluation. [Figure 2](#) Example fingerprint images of two fingers acquired with three different sensors (from the BioSec baseline corpus). Fingerprint images of the same finger are shown for a capacitive sensor (*left of each subplot*), an optical sensor (*center*) and a thermal sensor (*right*).

URL and following the instructions provided on the screen).

2. *Data Set 2 (DS2)*, acquired in a standard office room environment using a PC and a number of commercial sensors under the guidance of a human supervisor.
3. *Data Set 3 (DS3)*, acquired using two mobile ▶ [hand-held devices](#) under two acquisition conditions (controlled-indoor and uncontrolled-outdoor).

The three datasets of the BMDB include a common part of audio and video data. Additionally, DS2 includes signature, fingerprint, hand and iris data, and DS3 includes signature and fingerprint data. The three datasets were acquired in two different sessions (approximately 2 months between them). Pending yet to be distributed publicly, the BioSecure multimodal database has approximately 1,000 subjects in DS1, and 700 in DS2 and DS3. Fingerprint data in DS2 were acquired using an optical and a capacitive sensor. Fingerprint data in DS3 were acquired with a PDA.

The databases MCYT, BiosecurID, BioSec, and Bio-Secure have some commonalities that enable their integration for specific research studies, e.g., on time variability and sensor interoperability [12].

Fingerprint Evaluation Campaigns

The most important evaluation campaigns carried out in the fingerprint modality are the NIST Fingerprint Vendor Technology Evaluation (FpVTE2003) [13] and the four Fingerprint Verification Competitions (FVC), which took place in 2000 [4], 2002 [5], 2004 [6] and 2006 [7]. A comparative summary between FVC2004,

FVC2006 and FpVTE2003 is given [Table 1](#). An important evaluation is also the NIST Minutiae Interoperability Exchange Test (MINEX) [14].

Fingerprint Verification Competitions (FVC)

The Fingerprint Verification Competitions were organized with the aim of determining the state of the art in fingerprint verification. These competitions have received great attention both from academic and commercial organizations, and several research groups have used the FVC datasets for their own experiments later on. The number of participants and algorithms evaluated has increased in each new edition of the FVC. Also, to increase the number of participants, anonymous participation was allowed in 2002, 2004 and 2006. Additionally, the FVC2004 and FVC2006 were subdivided into: (1) *open category* and (2) *light category*. The light category aimed at evaluating algorithms under low computational resources, limited memory usage and small template size.

For each FVC competition, four databases were acquired using three different sensors and the SFinGE synthetic generator [1]. The size of each database was set at 110 fingers with eight impressions per finger (150 fingers with 12 impressions per finger in FVC2006). A subset of each database (all the impressions from ten fingers) was made available to the participants prior to the competition for algorithm tuning. The impressions from the remaining fingers were used for testing. Once tuned, participants submitted their algorithms as executable files to the evaluators. The executable files were then tested at the evaluator's site and the test data were not released until

Fingerprint Databases and Evaluation. **Table 1** Comparative summary between FVC2004, FVC2006 and FpVTE2003 (adapted from [6])

Evaluation	FVC 2004	FVC 2006	FpVTE 2003
Algorithms	Open category: 41 Light category: 26	Open category: 44 Light category: 26	Large scale test (LST): 13 Medium scale test (MST): 18 Small scale test (SST): 3
Population	Students	Heterogeneous (including manual workers and elderly people)	Operational data from a variety of U.S. Government sources
Fingerprint format	Flat impressions from low-cost scanners	Flat impressions from low-cost scanners	Mixed formats (flat, slap and rolled) from various sources (paper cards, scanners)
Perturbations	Deliberately exaggerated perturbations	Selection of the most difficult images according to a quality index	Intrinsic low quality fingers and/or non-cooperative users
Data collection	Acquired for this event	From the BioSec database	From existing U.S. Government sources
Database size	Four databases, each containing 880 fingerprints from 110 fingers	Four databases, each containing 1,800 fingerprints from 150 fingers	48,105 fingerprints from 25,309 subjects
Anonymous participation	Allowed	Allowed	Not allowed
Best average EER (over all the databases used)	2.07 % (Open category)	2.16 % (Open category)	0.2 % (MST, the closest to the FVC open category)

Fingerprint Databases and Evaluation. **Table 2** Results in terms of equal error rate (EER) of the best performing algorithm in each of the four databases of the FVC competitions

Database	2000	2002	2004	2006
DB1 (%)	0.67	0.10	1.97	5.56
DB2 (%)	0.61	0.14	1.58	0.02
DB3 (%)	3.64	0.37	1.18	1.53
DB4 (%)	1.99	0.10	0.61	0.27
Average	1.73	0.19	2.07	2.16

the evaluation concluded. In order to benchmark the algorithms, the evaluation was divided into: (1) ► **genuine attempts**: each fingerprint image is compared to the remaining images of the same finger, and (2) ► **impostor attempts**: the first impression of each finger is compared to the first image of the remaining fingers. In both cases, symmetric matches were avoided.

In Table 2, results of the best performing algorithm in each FVC competition are shown. Data in the 2000

and 2002 editions were acquired without special restrictions and, as observed in Table 2, error rates decrease significantly from 2000 to 2002, demonstrating in some sense the maturity of fingerprint verification systems. However, in the 2004 and 2006 editions, it is observed that error rates increase with respect to the 2002 edition due to the deliberate difficulties introduced in the data, thus revealing that degradation of quality has a severe impact on the recognition rates [15].

NIST Fingerprint Vendor Technology Evaluation (FpVTE2003)

The NIST Fingerprint Vendor Technology Evaluation (FpVTE2003) [13] aimed at: (1) comparing systems on a variety of fingerprint data and identifying the most accurate systems; (2) measuring the accuracy of fingerprint matching, identification, and verification on actual operational fingerprint data; and (3) determining the effect of a variety of variables on matcher accuracy. Eighteen different companies competed in the FpVTE, and 34 systems were evaluated.

Three separate subtests were performed in the FpVTE2003: (1) the large-scale test (LST), (2) the medium-scale test (MST), and (3) the small-scale test (SST). SST and MST tested matching accuracy using individual fingerprints, whereas LST used sets of fingerprint images. The size and structure of each test were designed to optimize competing analysis objectives, available data, available resources, computational characteristics of the algorithms and the desire to include all qualified participants. In particular, the sizes of MST and LST were only determined after a great deal of analysis of a variety of issues. Designing a well-balanced test to accommodate heterogeneous system architectures was a significant challenge.

Data in the FpVTE2003 came from a variety of U.S. Government sources, including low quality fingers of low quality sources. 48,105 sets of flat slap or rolled fingerprint sets from 25,309 individuals were used, with a total of 393,370 fingerprint images. The systems that resulted in the best accuracy performed consistently well over a variety of image types and data sources. Also, the accuracy of these systems was considerably better than the rest of the systems. Further important conclusions drawn from the FpVTE2003 included: (1) the number of fingers used and the fingerprint quality had the largest effect on system accuracy; (2) accuracy on controlled data was significantly higher than accuracy on operational data; (3) some systems were highly sensitive to the sources or types of fingerprints; and (4) accuracy dropped as subject age at time of capture increased.

NIST Minutiae Interoperability Exchange Test (MINEX)

The purpose of the NIST Minutiae Interoperability Exchange Test (MINEX) [14] was to determine the

feasibility of using minutiae data (rather than image data) as the interchange medium for fingerprint information between different fingerprint matching systems, and to quantify the verification accuracy changes when minutiae from dissimilar systems are used for matching fingerprints. ► **Interoperability** of templates is affected by the method used to encode minutiae and the matcher used to compare the templates. There are different schemes for defining the method of locating, extracting, formatting and matching the minutiae information from a fingerprint image [1]. In the MINEX evaluation, proprietary template formats were compared to the ANSI INCITS 378-2004 template standard.

The images used for this test came from a variety of sensors, and included both live-scanned and non live-scanned rolled and plain impression types. No latent fingerprint images were used. Participants submitting a system had to provide an algorithm capable of extracting and matching a minutiae template using both their proprietary minutiae format and the ANSI INCITS 378-2004 minutiae data format standard. The most relevant results of the MINEX evaluation are:

1. In general, proprietary templates lead to better recognition performance than the ANSI INCITS 378-2004 template.
2. Some template generators produce standard templates that are matched more accurately than others. Some matchers compare templates more accurately than others. The leading vendors in generation are not always the leaders in matching and vice-versa.
3. Authentication accuracy of some matchers can be improved by replacing the vendor's template generator with that from another vendor.
4. Performance is sensitive to the quality of the dataset. This applies to both proprietary and interoperable templates. Higher quality datasets provide reasonable interoperability, whereas lower quality datasets do not.

Related Entries

- Biometric Sample Acquisition
- Fingerprint Device
- Interoperability
- Performance
- Performance Evaluation

References

1. Maltoni, D., Maio, D., Jain, A., Prabhakar, S.: *Handbook of Fingerprint Recognition*. Springer, New York (2003)
2. Jain, A., Ross, A., Pankanti, S.: Biometrics: A tool for information security. *IEEE Trans. Inf. Forensics Secur.* **1**, 125–143 (2002)
3. NIST Special Databases and Software from the Image Group, <http://www.itl.nist.gov/iad/894.03/databases/defs/dbases.html>
4. Maio, D., Maltoni, D., Capelli, R., Wayman, J., Jain, A.: FVC2000: Fingerprint verification competition. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 402–412 (2002)
5. Maio, D., Maltoni, D., Capelli, R., Wayman, J., Jain, A.: FVC2002: Second fingerprint verification competition. *Proc. Intl. Conf. Pattern Recognit.* **3**, 811–814 (2002)
6. Capelli, R., Maio, D., Maltoni, D., Wayman, J.L., Jain, A.K.: Performance evaluation of fingerprint verification systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 3–18 (2006)
7. FVC2006. Fingerprint Verification Competition, <http://bias.csr.unibo.it/fvc2006/default.asp> (2006)
8. Fierrez, J., Ortega-Garcia, J., Torre-Toledano, D., Gonzalez-Rodriguez, J.: BioSec baseline corpus: a multimodal biometric database. *Pattern Recognit.* **40**, 1389–1392 (2007)
9. Garcia-Salicetti, S., Beumier, C., Chollet, G., Dorizzi, B., les Jardins, J., Lunter, J., Ni, Y., Petrovska-Delacretaz, D.: BIOMET: A multimodal person authentication database including face, voice, fingerprint, hand and signature modalities. *Lecture Notes Comput. Sci.* **2688**, 845–853 (2003)
10. Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., Gonzalez, J., Faundez-Zanuy, M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J., Vivaracho, C., Escudero, D., Moro, Q.: MCYT baseline corpus: a bimodal biometric database. *IEE Proc. Vision Image Signal Process* **150**, 395–401 (2003)
11. Fierrez, J., Galbally, J., Ortega-Garcia, J., et al: BiosecurID: A multimodal biometric database. *Pattern Anal. Appl.* **12** (2009)
12. Ortega-Garcia, J., Fierrez, J., Alonso-Fernandez, F., et al: The multi-scenario multi-environment BioSecure multimodal database (BMDB), *IEEE Trans. Pattern Anal. Mach. Intell.* **31** (2009)
13. Wilson, C., et al: Fingerprint Vendor Techonology Evaluation 2003: Summary of Results and Analysis Report. NISTIR 7123, <http://fpvte.nist.gov> (2004)
14. Grother, P., et al.: MINEX – Performance and interoperability of the INCITS 378 fingerprint template, NISTIR 7296, <http://fingerprint.nist.gov/minex> (2005)
15. ANSI-INCITS 378, Fingerprint Minutiae Format for Data Interchange, American National Standard (2004)

Fingerprint Device

- Fingerprint, Palmprint, Handprint and Soleprint Sensor

Fingerprint Encryption

► Fingerprints Hashing

Fingerprint Fake Detection

JEAN-FRANÇOIS MAINGUET
Grenoble, France

Synonyms

Liveness detection; Cut finger problem; Dead finger detection; Fake finger detection; Gummy bear finger; Latex finger; Liveness detection

Definition

Fingerprint fake detection is used to identify a fake finger, such as a cast made of latex. By extension, it also includes tests to detect a cut finger or dead finger, or a latent print remaining on a sensor after usage.

Introduction

In “Diamonds are forever” (1971) [1] James Bond took the identity of Peter Frank with a thin layer of latex glued on his fingertip to spoof Tiffany Case’s camera. James was using a simple fake finger, but the situation can be worse. With automated fingerprint recognition systems becoming more widely used, concerns over fingerprint fake detection have increased. In March 2005, a team of carjackers in Subang Jaya in Malaysia chopped off part of the owner’s left index finger, when they realized that his S-Class Mercedes Benz had a security feature which would immobilize the car without his fingerprint. Even with more reliable cut finger detectors in use, it is likely that this will happen again.

Security of a fingerprint-based system can be divided into two main areas:

1. The electronic security, which poses the question: “Is the electronic system, at the other end of the wires, a real trustful authorized fingerprint system?”

2. The liveness security, which asks a different question: "Is the object touching the sensor a real finger, alive and connected to a living person?"

Answers for electronic security deal with cryptography, using challenge-response schemes and cryptographic codes. Since the focus of this essay is to answer the second question, we will suppose that the electronic system is perfect and cannot be broken.

To begin, we know 100% security does not exist. However, what we would verify is that, "I'm Mr X, a living person not under threat and I agree to this action." Lacking the ability to read a person's mind, this is an impossible task. At the opposite end, a basic fingerprint system will identify a particular fingerprint image as likely the same one as registered in the template, which is only a small brick within a full security system.

To fill the gap, we need to acquire more information that will enable us to say "this is a real alive finger." If we can do that, then we have a good chance to know that a real person is making the transaction, rather a cast or cut finger being applied to the sensor. This will not answer the problem of detecting a person under threat, but it should be enough under normal usage, although some situations will never be detectable. For example, it will be impossible to detect a graft. In France, a man received two hands from a donor, a great medical achievement [2]. But at the same time, he received 10 brand new fingerprints! There is also the case of George who attempted to enter the US illegally on 24 September 2005 through the Nogales, Arizona Port of Entry during which time US Customs and Border Protection officers noted that his fingerprints had been surgically replaced with skin from his feet. George stated that this procedure had been done by a doctor in Phoenix to "clean" his identity [3]. But these should be extremely rare cases. What is primarily desired is to avoid anyone stealing a fingerprint to impersonate someone else. So, while it is impossible to create an absolute fake finger detection system, it is possible to make things extremely hard to be cracked.

Compromised Fingerprint

When someone creates a fake of one of the fingerprints and use it to spoof a fingerprint system, then we say that this fingerprint is compromised. With a smart

card (or a key), the smart card can be revoked. Further use of the card can be prevented and a new one can be created. But with fingerprints, this is limited to the 10 fingers. Biometric traits – the basis of biometrics cannot be revoked.

Liveness detection solves the compromised fingerprint problem. If the system can check that it is the real alive finger, then there is no possibility of using a fake.

Attended/Unattended System

It is commonly admitted that an attended biometric system does not need any liveness detection because the supervisor "obviously" checks that a real alive person is present. In the case of fingerprints, this would be true if the supervisor was checking the finger: is the finger really connected to the body, and without any glued cast (Fig. 1)?

Fingerprint Fake Finger Detection Levels

There are three fake finger methods and detection levels described, starting from the easiest to the hardest to detect:

1. Latent print left on the sensor
2. Fake/copies:
 - a. Printed fingerprint image
 - b. Fake made of gelatin, latex, or other material



Fingerprint Fake Detection. Figure 1 Thin fake made of gelatine glued on a real finger.

- c. Thin layer of material glued to a real finger, including real skin cells grown in a laboratory
- 3. Original finger:
 - a. Cut out
 - b. Belonging to a dead person
 - c. Alive person under threat

Significant Developments in Fingerprint Spoofing

In the early 1990s, Ton van der Putte developed and improved a technique to fool the available biometrical fingerprint recognition systems. But when he contacted the manufacturers and showed them the security breach in their systems, it was ruled unimportant and nothing was done to solve it. In 2000, van der Putte and Jeroen Keuning decided to raise people's awareness and published an article [4] “*as a warning to those thinking of using new methods of identification without first examining the technical opportunities for compromising the identification mechanism.*” Using duplication with and without cooperation and material such as silicone rubber, van der Putte and Keuning

fooled four optical sensors and two silicon-based capacitance sensors.

In 2001, Kàkona [5] described how to spoof an optical fingerprint sensor using a printed fingerprint (Fig. 2) and reactivating latent fingerprints on the sensor's surface by breathing on it. In 2002, Thalheim et al. [6] tested five solid-state and two optical fingerprint sensors. Gummy bears were introduced by Matsumoto [7] in 2002. The experiments involved 11 commercially available fingerprint sensors, both optical and capacitive, using a new cheap material, gelatin.

Further studies from Kang [8] and Blommè [9] extended the previous work. Stén et al. [10] spoofed a capacitance sensor using hot glue for the negative mold and gelatin for the fake finger. Marie Sandström (2004) published her thesis [11], “Liveness Detection in Fingerprint Recognition Systems,” which gathered most of the available technologies at that time as well as experiment results on various sensors.

In 2006, Ton van der Putte updated his work [12] using additional material including silicon gel, acrylic paint, gelatin, gum arabic. Ongoing attempts to spoof fingerprint sensors continue to appear on the Internet; for instance, the Chaos Computer Club [13] used wood glue and published their results online (Fig. 3).



Fingerprint Fake Detection. **Figure 2** Printed fingerprint spoofing an optical sensor.



Fingerprint Fake Detection. **Figure 3** Wood glue using a printed fingerprint as negative.

Making a Fake Fingerprint

Making a fake fingerprint always requires a fingerprint image. The easiest way to get a good fingerprint image is to have the cooperation of the donor. This is rarely the case in the real world, except when the latent print is left on the sensor (Level 1). In that case, the donor completes a successful acquisition; later, the impostor “reactivates” the latent print by breathing on the sensor. This happened in the past with some optical systems and with some capacitance-based sensors. A simple algorithm rejecting an image previously acquired is generally enough to avoid this problem, while swipe-sensing just eliminates this possibility.

The required fingerprint image is not necessarily exactly the same as the original fingerprint of the donor. Minutia matching (which is the main matching

technology for fingerprints) only requires having the minutia locations and directions at the right place. It is possible, in theory, to create a fingerprint image with the right minutia locations that does not look like the original. This requires accessing the minutiae locations stored in the template, which should be ciphered. Work related to some form of automated reconstruction has been proposed, requiring only access to the matching score (hill-climbing) [14, 15]. This technique is far more difficult compared to obtaining the original fingerprint.

So in general, an impostor will take the easiest way to obtain the original fingerprint image. We will not deal here with the Level 3 which requires the original finger, cut, or belonging to a dead person. Obtaining the original image can be done with or without cooperation. With cooperation is the easiest way, and most

articles dealing with spoofing assume that the right finger is available to create a negative mould. Without cooperation will be the most common situation. Fortunately, stealing the fingerprint of someone else is not easy. Even for forensic professionals, it is hard to identify people from fingerprints left in a crime scene. Moreover, it is very difficult to select which fingerprint to use. It is likely that the forefinger is the most common finger used in a fingerprint system, but selecting the right fingerprint is not obvious.

Once the right image is obtained, image processing skills are generally required to enhance the fingerprint. Printed circuit technologies are often proposed to create a negative mould, but sometimes direct molding techniques, such as a rubber stamp (Fig. 4), can be used to get a positive.

With a negative mold, you need to create the positive cast that will be used to spoof the fingerprint sensor. Glue, latex, gelatin, and other materials have been proposed (Fig. 5), but the most difficult thing is to select the right material that properly fits the sensor. Latex may work for some sensors and not for others. Understanding the physics of the sensing techniques will help. So, at the end of the day, making a fake finger without cooperation is difficult, but far from being impossible.



Fingerprint Fake Detection. **Figure 4** Rubber stamp.

Liveness Measurement

To be able to detect a fake, we must first answer the question of what defines a live finger. Some activities related to liveness are:

1. Cellular metabolism with material transformation (protein)
2. Movement
3. Heat production (a sub-product)
4. Blood circulation for material delivery and heat transportation (regulation)

These activities have a number of signatures: physical, chemical, mechanical, nervous, geometrical, to name a few. Moreover, signification changes with the observation scale.

Detection methods can be active or passive. Active techniques involve a response to a stimulus, and can be voluntary or involuntary. It could be seen like a challenge-response as used in regular cryptographic techniques. Involuntary are reflexive challenge responses (removing your finger when you feel an electrical shock), while voluntary are behavioral challenge responses (how many vibrations did you feel?). Active detection is very interesting, because the nervous system up to the brain can be involved, which is a good marker of aliveness. But generally, active detection is not very practical from a user point of view, and nociceptive methods are not acceptable.

Passive techniques are linked to physiological activity of the finger. Here are some physiological data about fingers:

1. Cells, a bone, and a nail make a structure of about 1–10 cm³. Note that there is no muscle (and so electrical activity is coming from other areas)
2. Arterial blood brings all chemicals, oxygen, and heat and returns to the body through veins
3. Skin is composed of three layers:
 - a. Stratum corneum made of dead cells, more or less hydrated, 100 µm thick, variable electrical conductivity
 - b. Blood-free epidermis, 0.05–1 mm thick, made of proteins, lipids, melanin-forming cells
 - c. Dermis: dense connective tissues, capillaries arranged in vertical loops
4. Arteriovenous anastomoses, innervated by nerve fibers that regulate the blood flow of a factor of 30 in response to heat



Fingerprint Fake Detection. **Figure 5** Some moulded fakes: gelatine, plastic (negative), alginate, silicon.

5. Temperature range: 10°–40°C; not regulated
6. Skin emits some specific molecules (odor)
7. Skin presents some plasticity

Remark: The external layer of the skin is made of *dead* cells, which is not a favorable configuration for liveness detection!

Any liveness detection reader should read one or several data related to the previous list. Also, reading only one characteristic will not ensure that the read fingerprint is coming from a real finger: some material exhibiting the same plasticity than skin exists for instance.

Fingerprint Sensors with Liveness Detection

Few fingerprint sensor manufacturers claim to have some kind of liveness detection; and whenever claimed, little or evasive information is given. But, new techniques and ideas are being explored:

1. Maybe the most common liveness detection method is based on electrical measurements, using the conductivity and/or impedance of the skin. Some sensors can acquire fingerprints using electrical properties of the skin (RF-field, capacitance, electro-optical), and so require a conductive material to be spoofed. Non-conductive latex cannot work
2. Light transmission properties of the skin and/or the blood. Hospitals are using pulse oxymetry to measure the blood oxygenation, i.e., the percentage of oxyhemoglobin compared to deoxyhemoglobin. Two LEDs send infrared light through the finger to a photodiode, so it is some additional material aside the regular fingerprint sensor. Skin spectrum has also been proposed [16], using a wider range of colors
3. Perspiration induces detectable changes in time when looking at a series of images [17]
4. Distortion of the skin depends a lot on its plasticity [18]
5. Skin emits some specific molecules that can be detected (odor) [19]

Faking the Counter Measures

Any measurement can be faked:

1. Electrical method can be faked by the appropriate voltage applied on the sensing area (or even a simple connection to real skin while a fake is applied)
2. Optical methods can be faked by the appropriate plastic with the correct absorption characteristics
3. An optical sensor is made of photodiodes; it is always possible to send the appropriate light, synchronized with the light sent by the system
4. Cardiac pulse can be faked with the appropriate pump and pipes

But it is possible to make things very hard to spoof. For instance, the latest immigration control systems acquire the two forefingers at the same time, and so trying to spoof both sensors at the same time will be much harder.

Conclusion

Fake fingerprint detection will be an important feature of fingerprint sensors in the future, likely mandatory. We already know that a no fingerprint system will be 100% spoof-proofed, but several different sensors reading different information at the same time will be very hard to deceive. The “Swiss cheese” model applies here: each slice of cheese is not 100% secure, some holes exist. But more slices will stop most of threats... at the cost of each slice!

Related Entries

- ▶ Anatomy of Fingerprint
- ▶ Forensic Science
- ▶ Security

References

1. Danjaq, S.A.: DIAMONDS ARE FOREVER Copyright © (1971)
2. Transplantation des deux mains dans le service de chirurgie de transplantation de l'Hôpital Edouard Herriot/Hôpitaux

de Lyon. January 13, 2000 http://www.chu-lyon.fr/internet/relations_medias/2005/5ans_double_greffre/dossier_presse_5ans_double_greffre.pdf

3. US Department of Justice's US Attorney's office for Arizona press release, May 3 (2006)
4. van der Putte, T., Keuning, J.: Biometrical fingerprint recognition don't get your fingers burned. In: Proceedings of IFIP TC8/WG8.8 Fourth Working Conference on Smart Card Research and Advanced Applications, pp. 289–303. Kluwer, Dordrecht (2000) (<http://cryptome.org/fake-prints.htm>)
5. Kákona, M.: Biometrics: yes or no? (<http://home.i.cz/kakl/biometrics/Biometricsyesorno.htm>). Accessed Feb 11, (2009)
6. Thalheim, L., Krissler, J., Ziegler, P.M.: Body check: biometric access protection devices and their programs put to the test (<http://www.heise.de/ct/02/11/114/>)
7. Matsumoto, T., Matsumoto, H., Yamada, K., Hoshino, S.: Impact of artificial gummy fingers on fingerprint systems. In: Proceedings of SPIE on Optical Security and Counterfeit Deterrence Techniques IV, San Jose, CA, vol. #4677 (2002) (<http://cryptome.org/gummy.htm>)
8. Kang, H., Lee, B., Kim, H., Shin D., Kim, J.: A study on performance evaluation of the liveness detection for various fingerprint sensor modules. <http://www.springerlink.com/content/0df29gectgdkwrkl/>
9. Blommé, J.: Evaluation of biometric security systems against artificial fingers (<http://www.ep.liu.se/exjobb/isy/2003/3514/exjobb.pdf>). Accessed Feb 11, (2009)
10. Stén, A., Kaseva, A., Virtanen, T.: Fooling fingerprint scanners – biometric vulnerabilities of the precise biometrics 100 SC scanner. In: Proceedings of the Fourth Australian Information Warfare and IT Security Conference, Adelaide, Australia (2003) http://www.stdot.com/pub/ffs_article_asten_akaseva.pdf
11. Sandström, M.: Liveness detection in fingerprint recognition systems. Ph.D. thesis, Linkoping University, Sweden (2004)
12. van der Putte, T.: Workshop spoofing fingerprints, SAFE-NL, University of Twente <http://wwwes.cs.utwente.nl/safe-nl/meetings/08-06-2006/ton.pdf>
13. Chaos Computer Club, How to fake fingerprints? October 26, 2004 http://www.ccc.de/biometrie/fingerabdruck_kopieren.xml?language=en
14. Martinez-Diaz, M., Fierrez-Aguilar, J., Alonso-Fernandez, F., Ortega-Garcia, J., Siguenza, J.A.: Hill-climbing and brute-force attacks on biometric systems: a case study in match-on-card fingerprint verification. In: Proceedings of the 40th Annual IEEE International Carnahan conferences security technology, Lexington, Kentucky, pp. 151–159 (2006)
15. Ross, A., Shah, J., Jain, A.K.: Towards reconstructing fingerprints from minutiae points. In: Proceedings of SPIE Conference on Biometric Technology for Human Identification II, Orlando, FL, pp. 68–80 (2005)
16. Nixon, K.A., Rowe, R.K.: Multispectral fingerprint imaging for spoof detection. In: Jain, A., Ratha, N. (eds.) Proceedings of SPIE on Biometric Technology for Human Identification II, Bellingham, WA, vol. 5779, pp. 214–225 (2005)
17. Parthasaradhi, S., Derakhshani, R., Hornak, L., Schuckers, SAC.: Time-series detection of perspiration as a liveness test in fingerprint devices. IEEE Trans. Syst. Man Cybern. C Appl. Rev. **35**, 335–343 (2005)

18. Antonelli, A., Cappelli, R., Maio, D., Maltoni, D.: A new approach to fake finger detection based on skin distortion. International Conference on Biometric Authentication (ICBA'06), Hong Kong, China (2006)
19. Baldisserra, D., Franco, A., Maio, D., Maltoni, D.: Fake finger-print detection by odor analysis. International Conference on Biometric Authentication (ICBA'06), Hong Kong, China (2006)

Fingerprint Features

JOSEF BIGUN
 Embedded Intelligent Systems Center
 Halmstad University, IDE, Halmstad, Sweden

Synonyms

Fingerprint analysis; Fingerprint characteristics;
 Fingerprint signatures

Definition

Fingerprint features are parameters in epidermis images of a fingertip (the fingerprint) that can be utilized to extract information which is exclusively specific to a unique person. These parameters can be measured by computational techniques applied to a digital image obtained by a *fingerprint sensing* method, e.g., using live optical or solid-state scanners, and digitizing ink-rolled or latent fingerprint images. Such identity characterizing parameters include one or more specifics of ridge–valley direction and frequency, *minutiae*, and *singular points*. The fingerprint features should be reproducible and resilient to variation in the face of external factors such as aging, scars, wear, humidity, and method of collection.

Introduction

Fingerprints consist of ridges alternating with valleys that mostly run in parallel but also change direction smoothly or may terminate abruptly. Other patterns in nature that resemble fingerprints include Zebra skins, corals, and shallow sea-bottom. Such pattern variations can be parametrized and used to characterize the

fingerprints of individuals and to distinguish them from others. Identity establishment by fingerprint matching has been used by human experts long before the computer era, e.g., the nineteenth century contributors to the advancement of fingerprints, Jan. Purkyně, William Herschel, Alphonse Bertillon, Francis Galton, Edward Henry, Aziz-ul Haque, Chandra Bose, to name but few [1].

Caused by a foray of factors, low contrast and noisy images can compromise the reproducibility of fingerprint feature severely. Although the reason of poor image quality might be known, a better data acquisition is sometimes not a practicable option, e.g., latent fingerprints at a crime-scene, aging, scars and bruises, professional wear, etc. Accordingly, reproducibility is an important property of fingerprint features to be used. Another issue is their computational efficiency, if lacking it can hinder a practice of a fingerprint recognition method altogether, e.g., AFIS systems used in border-control, altogether.

Minutiae, to be discussed below in further details, represent the most widely used feature type by machine as well as human experts to determine if two fingerprints match. The geometric interrelationships of extracted minutiae, the spatial frequency between them or in their vicinity, and the local direction, contribute all to the strength of a minutiae based feature set so as to uniquely characterize a fingerprint. Another set of well-localized points is singular points. As will be detailed later, these are few, and one need larger neighborhoods to determine them in comparison to minutiae.

An important tool to characterize fingerprints is their *direction fields* since they are used in many operations of fingerprint processing. In the coming sections, we discuss direction field estimation, followed by minutiae, and singular points.

Direction Fields

The fingerprint direction fields are dense vector fields representing dominant local directions. A direction of an image-point (a pixel) is thus a property of its neighborhood; by itself no pixel can define a direction. Early direction fields were associated with local edges or lines and were approximated by the gradient of the image, $\nabla f = (\partial f / \partial x, \partial f / \partial y)^T$ where f is the local image, on digital lattices. Direction in this sense is the angle of the gradient and has already been used in

1960s, including in fingerprint applications. However, this concept hinders the use of effective signal processing tools, because a sinusoidal wave pattern (the local fingerprint) has a unique direction whereas half of its gradient directions in a fingerprint patch differ with 180° from the other half, resulting in a neither unique nor continuous representation if gradient angles would have defined feature spaces representing ridge directions. In turn this hinders efficient signal processing and inference which require rotation, scale-space, and interpolation operations.

Direction Fields by Structure Tensor

An effective cure to representation ambiguity of ridge and valley direction is to use the concept of iso-curve (points having the same gray-value), which suggests the use of 2×2 tensors naturally, in the quest of an optimal direction estimation in the total least squares sense. This is summarized next, where the notion of image refers to a local patch of a fingerprint.

If all iso-curves of an image has a common direction the image is said to be linearly symmetric, e.g., sinusoidal planar waves resembling most neighborhoods of fingerprints. Ideally, the unknown direction \mathbf{k} is optimal for an image $f(\mathbf{r})$ if the image is invariant to a translation in the amount of ϵ along the line \mathbf{k} where ϵ is small and can be positive as well as negative, and $\|\mathbf{k}\|=1$. Then the total translation error \mathcal{E}

$$\begin{aligned} \mathcal{E}(\mathbf{r}) &= f(\mathbf{r} + \epsilon\mathbf{k}) - f(\mathbf{r}) \\ &= \epsilon[\nabla f(\mathbf{r})]^T \mathbf{k} + \mathcal{O}(\epsilon^2) = \epsilon e(\mathbf{r}) = 0 \end{aligned} \quad (1)$$

will be zero for all \mathbf{r} if the gray-value patch f is translation invariant in the direction \mathbf{k} . Here $e(\mathbf{r})$ is the unit-error. Ignoring the quadratic term $\mathcal{O}(\epsilon^2)$, because ϵ represents small translations, if and only if the unit-error of translation in the (fixed) direction \mathbf{k}

$$e(\mathbf{r}) = [\nabla f(\mathbf{r})]^T \mathbf{k} = 0 \quad (2)$$

vanishes, (1) will vanish for *all* \mathbf{r} of the patch. Evidently, the unit-error will even vanish on a discrete sub-set of the points of the patch, as below

$$\begin{pmatrix} D_x f_1 & D_y f_1 \\ D_x f_2 & D_y f_2 \\ \vdots & \vdots \\ D_x f_M & D_y f_M \end{pmatrix} \begin{pmatrix} k_x \\ k_y \end{pmatrix} = \mathbf{D}\mathbf{k} = \mathbf{0} \quad (3)$$

where $D_x f_l = \partial f(\mathbf{r}_l)/\partial x$ and $D_y f_l = \partial f(\mathbf{r}_l)/\partial y$ with \mathbf{r}_l being a node of a grid having M nodes on the patch. The matrix \mathbf{D} is the set of gradients on the grid nodes, as indicated on the left in Eq. (3). using the continuous 2D Gaussian

$$g_{\sigma^2}(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (4)$$

the elements of \mathbf{D} , such as $D_x f_l$ and $D_y f_l$, can be preferably obtained by convolving the original discrete image with the discretized partial derivatives of the Gaussian. The parameter controlling the amount of smoothing the thus obtained *derivation filter* can apply is fixed by a certain $\sigma = \sigma_d$ in x and y directions as standard deviation, to avoid nonisotropic artificial bias. However, asking for nil (infinitesimal translation) error at every \mathbf{r}_l with a common \mathbf{k} may not be possible to fulfill in practice because f may not be perfectly linearly symmetric. The next best thing one can do is to solve the problem in the total least squares error sense such that $\|\mathbf{D}\mathbf{k}\|^2$ is minimized for a direction \mathbf{k} . The solution is given by the least significant eigenvector of the structure tensor, $\mathbf{S} = \mathbf{D}^T \mathbf{D}$, which is easy to obtain analytically as discussed in the following section. Alternatively, one can apply SVD numerically to \mathbf{D} yielding the same solution obtained by an eigenvalue analysis of \mathbf{S} . Before computing the direction, in practice one needs to incorporate a window function $\mu_l = \mu(\mathbf{r}_l)$ into the solution as well because the patch must be cut-out of a larger image. This can be conveniently done in the tensor-space (at the level of the outer-product of the gradients) and via a Gaussian, to obtain a mathematically tractable optimization [2, 3].

$$\begin{aligned} \mathbf{S} &= \mathbf{D}^T \mathbf{D} \\ &= \begin{pmatrix} \sum_l (D_x f_l)^2 \mu_l & \sum_l (D_x f_l)(D_y f_l) \mu_l \\ \sum_l (D_y f_l)(D_x f_l) \mu_l & \sum_l (D_y f_l)^2 \mu_l \end{pmatrix} \quad (5) \\ &= \sum_l (\nabla f_l \nabla f_l^T) \mu_l \end{aligned}$$

$$\begin{aligned} &= \lambda_{\max} \mathbf{k}_{\max} \mathbf{k}_{\max}^T + \lambda_{\min} \mathbf{k}_{\min} \mathbf{k}_{\min}^T \\ &= (\lambda_{\max} - \lambda_{\min}) \mathbf{k}_{\max} \mathbf{k}_{\max}^T + \lambda_{\min} \mathbf{I} \end{aligned} \quad (6)$$

Here μ_l is a discrete Gaussian with a certain σ_w that defines the extension of the local fingerprint patches, λ_{\max} , \mathbf{k}_{\max} are the most significant eigenvalue of \mathbf{S} and its corresponding eigenvector, delivering the largest error and the maximum variation direction, respectively. Similarly the λ_{\min} , \mathbf{k}_{\min} yield the

corresponding quantities for the least error and the direction of least variation respectively. Notice that \mathbf{k}_{\max} and \mathbf{k}_{\min} are always orthogonal (\mathbf{S} is symmetric), have unit lengths, and sum to identity tensorially, $\mathbf{k}_{\max}\mathbf{k}_{\max}^T + \mathbf{k}_{\min}\mathbf{k}_{\min}^T = \mathbf{I}$. Thus to represent the direction we could relate it to \mathbf{k}_{\max} , the normal of the ridges/valleys, as well as to \mathbf{k}_{\min} because knowing one determines the other. The representation of the direction is made by the tensor $\mathbf{k}_{\max}\mathbf{k}_{\max}^T$ rather than \mathbf{k}_{\max} because the tensor representation will map the two possible numerical representations of the normal \mathbf{k} and $-\mathbf{k}$ to the same (tensor) quantity avoiding the ambiguity inherent to vectors as representations of axes/directions.

Complex Representation of the Structure Tensor

There is a mathematically equivalent but a more convenient way of representing the structure tensor, by use of complex gradients [2, 4],

$$I_{20} = \sum_l (D_x f_l + i D_y f_l)^2 \mu_l = (\lambda_{\max} - \lambda_{\min}) e^{i 2 \varphi_{\max}} \quad (7)$$

$$I_{11} = \sum_l |D_x f_l + i D_y f_l|^2 \mu_l = \lambda_{\max} + \lambda_{\min} \quad (8)$$

with φ_{\max} being the direction angle of \mathbf{k}_{\max} and $i = \sqrt{-1}$.

The first benefit of complex representation is that the direction of the eigenvector is delivered by averaging (summation) squares of complex gradients, Eq. (6), in the argument of I_{20} , though in *double-angle representation* [5], and both eigenvalues are easily obtained by computing, $|I_{20}|$ and I_{11} . However easy to obtain, eigenvalues will not be necessary for many applications, as it is more useful to work with the sums and differences of them. This is because if λ_{\min} is very small, an acceptable way to conclude upon this fact is to compare it with λ_{\max} . Accordingly, when we obtain a large (magnitude) complex number I_{20} for a patch, it means that we have a good direction fit (linearly symmetric patch) and a reliable estimate of the common direction will be found right in the argument of I_{20} (in ► double angle representation), with the reservation that $|I_{20}|$ must be close to I_{11} . By contrast, if the error of the worst direction is not much worse than the best direction then the direction

fit is poor, making the corresponding argument angle meaningless automatically. Notice that $|I_{20}| \leq I_{11}$ and equality holds between the two quantities if and only if the iso-curve directions are aligned (linearly symmetric patch).

The next benefit is that the complex representation allows effective scale-space operations, including computation by subsampling, band-pass pyramids, extracting specific ridge frequencies (by changing σ_d and σ_w), and coarse-to-fine refinements, etc. by using the complex image $(D_x f_l + D_y f_l)^2$ and its (realvalue) magnitude image, $|D_x f_l + D_y f_l|^2$.

Direction Fields as Features

The fact that scalar products on complex number fields is well defined makes direction fields *descriptive features* which can be used as complements to other descriptive features. If two fingerprints are registered, meaning that the query image f^q and the reference image f^r are rotated and translated such that they are aligned, then the scalar product between the corresponding direction fields of the query, $I_{20}(f^q)$, and the reference, $I_{20}(f^r)$, fingerprints

$$b(f^r, f^q) = \frac{|\langle I_{20}(f^r), I_{20}(f^q) \rangle|}{\sqrt{\langle I_{20}(f^r), I_{20}(f^r) \rangle \langle I_{20}(f^q), I_{20}(f^q) \rangle}} \quad (9)$$

can be used as a belief in the match. Here the scalar product is

$$\langle I_{20}(f^r), I_{20}(f^q) \rangle = \sum_l I_{20}^*(f_l^r) I_{20}(f_l^q) \quad (10)$$

and the summation is applied either to a region, possibly weighted by some quality index [6–8], e.g. the common region of the fingerprint pair to be matched. The star as superscript denotes complex conjugation.

Direction Decomposition

A concept, i.e., closely related to direction fields is the decomposition of the original fingerprint in a set of images representing the (local) energy in quantized directions (typically 6–8 angles) and scales (typically 1–3 frequencies). Such decompositions can be obtained by a suitable Gabor filter bank independent of the direction field computations discussed earlier. Although the Gabor filter-bank filtered images can be interpolated to

generate accurate and dense direction fields [9], these have been mainly used to enhance fingerprints, and to estimate texture properties of fingerprints. The method suggested by [10] assumes that a landmark in each fingerprint of a pair to be matched is available or the pair is somehow registered with the corresponding landmarks. In regular concentric sectors of a circle (defined by a uniform polar grid of 5 radii and 16 angles) around the landmark, the average absolute deviations of Gabor-cosine filter responses (single frequency, eight directions) over the patch are computed. Called Finger-code, this set of texture measures constitutes a 640 dimensional ($5 \times 16 \times 8$) integer valued feature vector that can be used as a descriptive vector on its own or in conjunction with other features, Fig. 4.

Segmentation

In addition to their auxiliary or direct use to define descriptive features, the direction fields are also used in *segmenting fingerprints*. The latter refers to separating the image area that contains an acceptable quality of fingerprints, from the rest, typically the background. Because the fingerprint regions have a dominant orientation, meaning that there is a direction along which the gray-values change significantly faster than the orthogonal direction, the absolute and/or relative differences of the structure tensor

eigenvalues, λ_{\min} , λ_{\max} have been used to achieve segmentation [6, 11].

Minutiae

Minutiae are end-points of ridges or valleys of a fingerprint, occupying typically 0.1–0.5 mm on the skin, and are visible as 2–10 pixels in images captured at 500 dpi resolution. Minutiae are the most widely used features to match two fingerprints, for a variety of reasons, including that there is a great amount of human expertise in their use, and that it is difficult to reconstruct the original fingerprints only by the knowledge of minutiae, mitigating privacy concerns. A minutia can be of the type *termination* or *bifurcation*. A bifurcation of a ridge exists in conjunction with termination of a valley and vice-versa because the former engulfs the latter, by definition. This is known as *duality*. However, one must bear in mind that ridges appear as valleys and vice-versa depending on the sensing conditions, i.e., whether the dark pixels or the white pixels are ridges. Accordingly, the minutia-type, i.e., bifurcation or termination, as a descriptive feature is meaningful only if the interpretation ambiguity caused by sensing can be accounted for. Because from these two types of minutiae it is possible to derive other constellations, e.g., lake, spur, cross-over, Fig. 2, several national agencies relying on



Fingerprint Features. **Figure 1** Commonly used classes to categorize fingerprints [27]. (a) Arch, (b) Tented Arch, (c) Left Loop, (d) Right Loop, (e) Whorl, (f) Twin Loop.

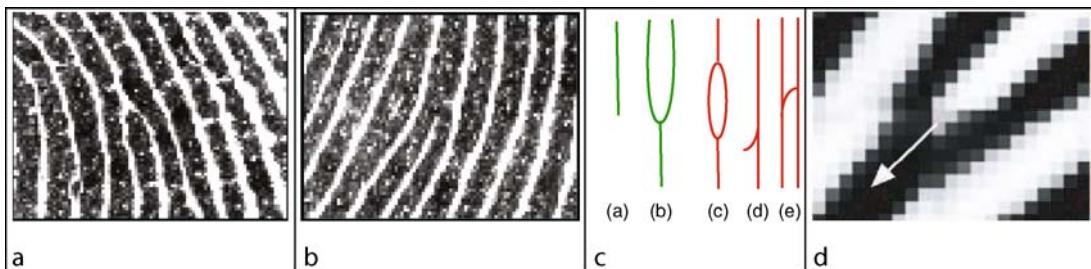
minutiae for their fingerprint processing base their taxonomy only on termination and bifurcation, e.g., FBI in USA [12]. Before minutiae extraction, *fingerprint enhancement* is applied if fingerprints are deemed noisy, usually according to an automatically extracted quality measure [7, 8, 13, 14].

Two main ways of minutiae extraction can be achieved by (1) by binary image processing operations, (2) by using gray-value image processing techniques.

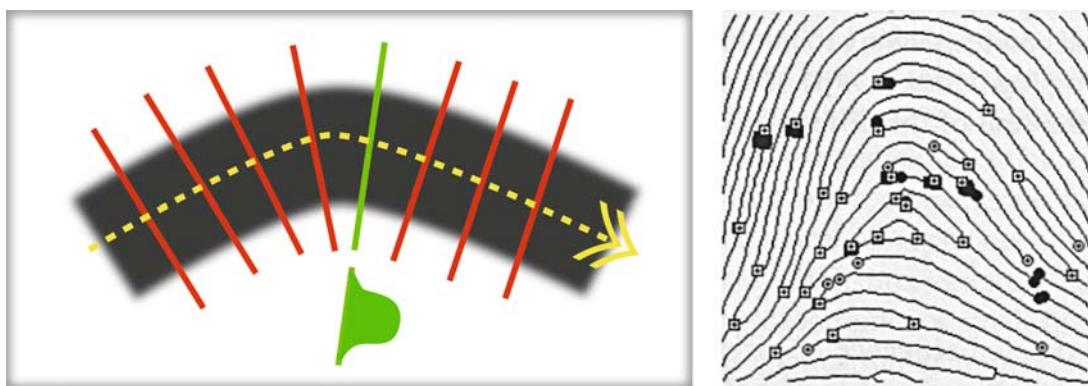
Assuming that the binary image of a fingerprint can be obtained and it has a reasonably high fidelity w.r.t. ridges, ► **fingerprint thinning** can be achieved by morphological operators (erosion and dilation) or by distance transforms [15–17]. A number of algorithms to extract minutiae from skeletonized binary images exist. It is common that at the beginning, there are several thousands of minutiae candidates of which only approximately 50 are real. Various criteria for

validating the endpoints, including the duality, a minimum length of the ridge or valley , are used to suppress spurious false minutiae [18].

However, minutiae detection based on binary images has a shortcoming, lack of robustness when used for low quality fingerprint images. Because ridge skeletons are obtained by applying a thinning method to the binarized fingerprint, the binary ridges should correspond to real ridges accurately if thinning procedure is to be successful. This puts high demands on the quality of the fingerprints, as well as the adaptiveness of the binarization since the resulting binary ridges might not represent the real ridges sufficiently well. Extracting minutiae from gray images, without passing through binarization, offers better opportunities in this respect. The ridges can be directly followed in the gray-value image by use of the direction field, and the gray-value ridge profiles [6, 11], Fig. 3.



Fingerprint Features. **Figure 2** Illustration of minutiae types and duality. (a) a ridge termination engulfed in a valley bifurcation; (b) vice-versa. (c) basic ridge types in green (termination, bifurcation) and derived types in red (lake, spur, crossover) (d) the direction of a minutia exemplified at a ridge-bifurcation.



Fingerprint Features. **Figure 3** Illustration of thinning and minutiae detection by ridge following in gray-images [11]. On the left, a segment of a ridge is represented. Gray-value profiles, like the one in green, are regularly sampled and tracked along the ridge, until a termination or a bifurcation is found. On the right, the result is shown where the white circles and squares represent terminations, and bifurcations respectively. The black circles and boxes are improvements of a postprocessing.

Alternatively, a large number of candidate minutiae can first be obtained, e.g., by detecting lack of linear symmetries during the direction field estimation, then a gray-value model of the minutiae, e.g., the parabolic appearance of terminations and bifurcations, can be enforced the candidates to retain the true minutiae [19], Fig. 4.

Minutia Direction When matching or registering two fingerprints the ▶ **minutia direction** is a valuable discriminative information. The minutiae directions can be either extracted from the direction field directly or from the direction of the binarized and thinned ridges, corresponding to minutiae locations, Fig. 2. The directions along with the type information (termination or bifurcation) are attached to minutiae coordinates.

Spatial Frequency Another descriptive feature which can be attached to minutiae positions is the spatial frequency information in the vicinity of minutiae. The spatial frequency is usually defined in terms of a direction in fingerprints and has different implementations [20]. One implementation is to use the average frequency of the ridge or the count of ridges in a fixed line segment orthogonal to the minutiae direction. Another implementation of the frequency measure is to count ridges or the average frequency along the line joining a pair of minutiae. Because pairs as well as triplet constellations of minutiae are commonly used in fingerprint matching, the frequency measures are

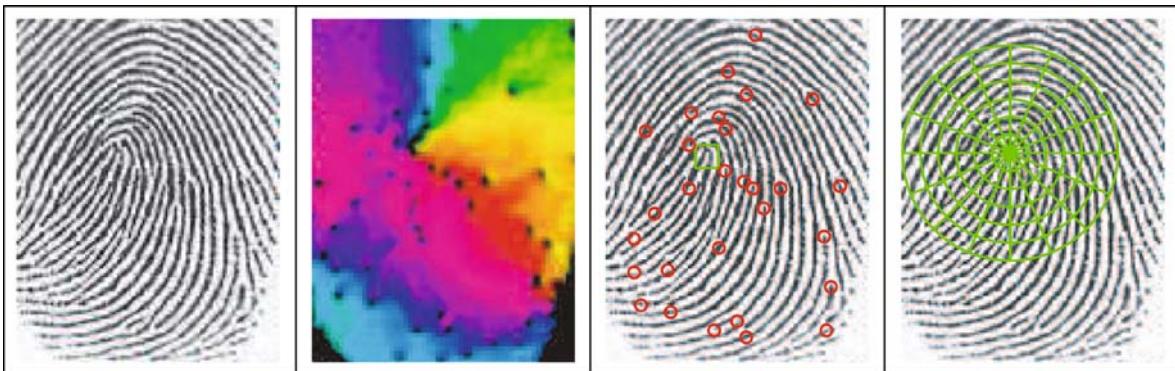
attached as a descriptive feature to the corresponding pairs or triplets.

Singular Points

Singular points are landmarks that are defined in large image patches (1–5 mm) compared to the size of minutiae. There are typically 1–2 singular points in a fingerprint though they may occasionally be missing or may be difficult to identify in a fingerprint. Three basic types can be discerned, loop (also known as core), whorl, and delta.

A major use of them is to classify a fingerprint typically into one of the six categories, (Left-loop, Right-Loop, Double-Loop, Arch, Tented-Arch, Whorl) which are different constellations of loops and deltas, Fig. 1. Such rough categorizations are employed to match, and to organize massive amounts of fingerprints data efficiently.

Loops can provide a unique intrinsic global orientation and position for a fingerprint, allowing an orientation and translation normalization of the fingerprint only on the basis of itself. Most whorls and deltas can provide a direction too, though these are in general not unique. Two singular points in the same image provide always a unique ▶ **intrinsic direction of fingerprint**. This normalization is a practical alternative to registration by minutiae or can be complementary. Every



Fingerprint Features. **Figure 4** Illustration of a use of direction field. From the *left*, the original image, the direction vector field color coded, the original superimposed with minutiae locations and the loop singularity [19, 24] are shown, respectively. On the *far right* the Fingercode grid placed, on the loop singularity, is shown [10]. The direction field image (*second*) represents the complex quantities I_{20} (7), where the argument of I_{20} is mapped to the Hue (color) of the HSV color model (same color indicates common direction) and the magnitude representing the quality of the direction fit is mapped to the Value (intensity).

fingerprint (the query, as well as every fingerprint in the database) is rotated and translated such that a reference point and a half-line that is well defined w.r.t. a singular point of the fingerprint become the origin and the positive x-axis. Two translation and rotation normalized fingerprints are then more efficiently matched – with minutiae or other features, because no rotation or translation compensation specific to the considered pair will be necessary.

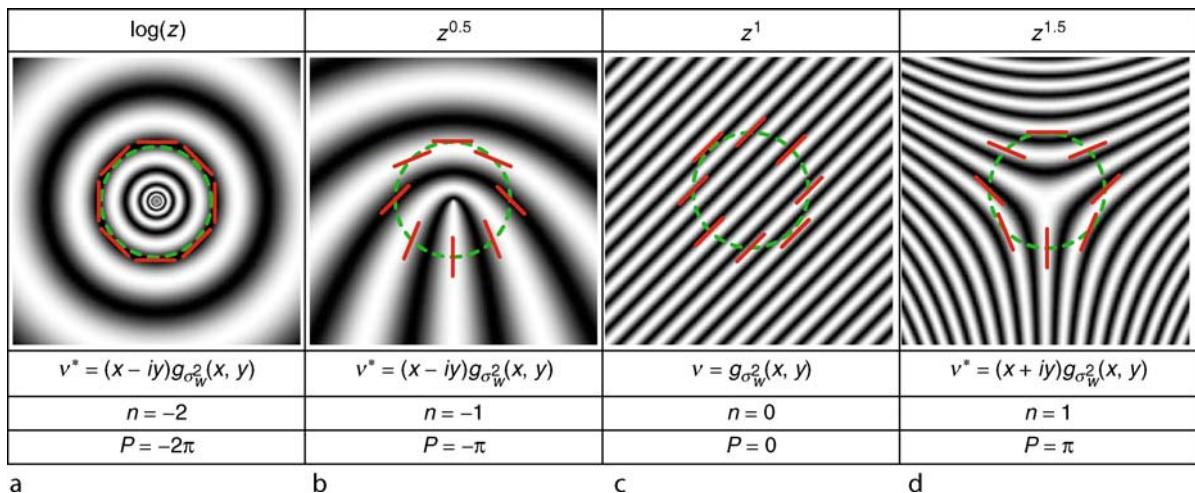
Finally, singular points can function as anchors to extract other descriptive features, e.g., the spatial frequency. One can count the ridges along a line joining two singular points, or along a line joining a minutia and a singular point, etc. The spatial frequency estimation issue is analogous to the one that has been discussed in conjunction with minutiae.

Singularities by Poincaré Index

One of the oldest singular point detection techniques used in fingerprint processing is the Poincaré index [21]. The index is defined for a path in a vector field and represents the total amount of angle change of the vectors along the curve. Assuming that the curve is closed and it is in the gradient field of a fingerprint then the Poincaré index, P , is given by

$$P = \oint \frac{\partial \theta}{\partial x} dx + \frac{\partial \theta}{\partial y} dy = \int \int \left(\frac{\partial^2 \theta}{\partial x \partial y} - \frac{\partial^2 \theta}{\partial y \partial x} \right) dx dy \quad (11)$$

where the function $\theta(x, y)$ represents the argument (angle) of the gradient vectors and the last expression is obtained by Green's Lemma. It is worth noting that even though the original fingerprint image is assumed differentiable (continuous) the gradient angle is not continuous, (π and $-\pi$) though its partial derivatives are. By laying the closed curve around a loop, a whorl, a regular (non-singular) point, and a delta, it can be concluded that P will assume $-2\pi, -\pi, 0$ and π radians, respectively. In Fig. 5 stylistic models of such fingerprint patches are shown along with segments of iso-curves (to which the gradients are orthogonal but are not shown for convenience). When one walks the dashed circle in full, the direction of iso-curves, and thereby the gradient angles change with the Poincaré index. This observation is used, typically along with the curve integral of Eq. (11), to determine if a candidate point is a whorl, loop, regular, or delta type. It is also possible to compute P according to the right hand side of the equation, by a double integral applied to the interior patch of the curve. By using the directions of linear-symmetry vector field, as opposed to those of the gradient field and the double integral [22]



Fingerprint Features. Figure 5 The top row shows the harmonic functions that generate the iso-curves of the patterns in the second row. The iso-curves (their linearized examples are shown as red line segments) are given by a weighted combination of the real and the imaginary parts of the respective harmonic functions, with a certain ratio between the weights, defining the direction parameter φ of each pattern [23]. The third row shows the filters, v where $g_{\sigma_w^2}$ is an ordinary Gaussian (4), to detect the singularity points and φ by a (complex) convolution applied to the direction field (12). The fourth row shows the symmetry order of the filters. The last row shows the Poincaré index of gradients.

suggested an alternative way of computing P . In this case the angles of the used vector field are continuous from the beginning so that no special care needs to be taken to achieve continuity at angles around π and $-\pi$. The resulting P must be divided by 2 to correspond to the gradient based Poincaré index.

Singularities by the Generalized Structure Tensor

A singular point can also be detected by use of the Generalized Structure Tensor (GST), which is an extension of the structure tensor to curvilinear (harmonic) coordinates [9, 23]. The fundamental idea is the same as that of the structure tensor – to find an (unknown) angle such that the patch remains invariant to a small translation along the found angle direction but in the curvilinear coordinates. It turns out that in this model, a singularity can be detected by complex filtering of the direction fields, already in the complex representation (7).

$$I'_{20} = \sum_l (D_x f_l + i D_y f_l)^2 v_l^* = (\lambda_{\max} - \lambda_{\min}) e^{i 2 \varphi_{\max}} \quad (12)$$

Here v_l is a filter specialized to detect a loop, a delta or a whorl, Fig. 5. The magnitude of a filter response, which is complex valued, encodes the likelihood that a location represents a singularity exactly in the same way as the ordinary structure tensor, but now the coordinates are harmonic, representing a pattern of a singularity, and the λ_{\max} and λ_{\min} are the error extrema due to translation in curvilinear trajectories having a certain direction. Likewise, its argument (angle) encodes the intrinsic orientation of the singularity (for loops and deltas their global inclination, for whorls the amount of chirality). The singularity filters can be implemented by derivatives of Gaussians which are separable, making them 1D filters. Because the complex feature space obtained from such filter responses are continuous both in their arguments and positions, scale-space filtering, e.g., coarse-to-fine refinement, is possible [24]. That the symmetry axes (intrinsic orientation) are available in the GST method is useful, because the obtained angle information can be used as a descriptive feature attached to the singular point coordinates, much like the use of minutiae orientations in fingerprint matching. Additionally, loop orientations alone allow a

normalization/registration of a fingerprint pair even if other singular points lack, and no minutiae are available.

Singularities by Other Methods

The methods discussed earlier can find singular points by modeling direction variations on closed curves (in practice a circle) or in regions containing a singularity. Methods which do not use closed paths are exemplified as follows. Such a method to obtain singularities is the early suggestion by [25] which models the direction variations along the horizontal scan lines. Information defining the location and the type of the singularity is contained in the direction information around the singular point and the horizontal lines contain only a part of this. This information is instead injected into the model in terms of orientation-change rules between scan lines. In [26], gradient vectors model half a circle, like “n.” Then generalized Hough transform is used to find a peak, suggesting the location of a loop. In contrast to GST and Poincaré index methods, the (loop) inclination is assumed to be approximately vertical, or a separate model is designed for alternative loop inclinations.

Summary

Descriptive features are used to match fingerprints. They include the locations of minutiae points, and the singular points. The location information can be enhanced with additional descriptive measurements including the local direction of the ridges and valleys at minutiae locations, the intrinsic orientation of singular points, the type of the singular points, ridge counts or average frequencies between minutiae as well as singular points. To extract such descriptive information direction maps are computed. Being texture measures, structure tensor representations of direction maps can also be used as descriptive features on their own if anchor points are available or in addition to minutiae based features. Similarly, Gabor filters can be used to obtain descriptive features if anchor points are available. Commonly used anchors for registration as well as descriptive features are the three basic singularity types, loops, whorls, and deltas. They can be detected and described independent of minutiae information.

Related Entries

- ▶ Fingerprint Enhancement
- ▶ Fingerprint Matching
- ▶ Fingerprint Quality
- ▶ Fingerprint Registration
- ▶ Fingerprint Sensing

References

1. Locard, A.: L'Identification des Récidivistes. A. Maloine, Paris (1909)
2. Bigun, J., Granlund, G.: Optimal orientation detection of linear symmetry. In: First International Conference on Computer Vision, ICCV, London, June 8–11, pp. 433–438. IEEE Computer Society, London (1987)
3. Kass, M., Witkin, A.: Analyzing oriented patterns. *Comput. Vision Graph. Image Process.* **37**, 362–385 (1987)
4. Bigun, J., Granlund, G., Wiklund, J.: Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE-PAMI* **13**(8), 775–790 (1991)
5. Granlund, G.: In search of a general picture processing operator. *Comput. Graph. Image Process* **8**(2), 155–173 (1978)
6. Ratha, N.K., Chen, S., Jain, A.K.: Adaptive flow orientation-based feature extraction in fingerprint images. *Pattern Recognit.* **28**(11), 1657–1672 (1995). URL [http://dx.doi.org/10.1016/0031-3203\(95\)00039-3](http://dx.doi.org/10.1016/0031-3203(95)00039-3)
7. Grother, P., Tabassi, E.: Performance of biometric quality measures. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(4), 531–543 (2007). URL <http://dx.doi.org/10.1109/TPAMI.2007.1019>
8. Fronthaler, H., Kolreider, K., Bigun, J., Fierrez, J., Alonso-Fernandez, F., Ortega-Garcia, J.: Fingerprint image quality estimation and its application to multi-algorithm verification. *IEEE Trans. Inform. Forens. Security* **3**(2): 331–338 (2008)
9. Bigun, J.: Vision with Direction. Springer, Heidelberg (2006)
10. Jain, A.K., Prabhakar, S., Hong, L., Pankanti, S.: Filterbank-based fingerprint matching. *IEEE Trans. Image Process.* **9**(5), 846–859 (2000). URL <http://dx.doi.org/10.1109/83.841531>
11. Maio, D., Maltoni, D.: Direct gray-scale minutiae detection in fingerprints. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(1), 27–40 (1997). URL <http://www.computer.org/tpami/tp1997/i0027abs.htm>
12. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of Fingerprint Recognition. Springer, Berlin (2003). URL <http://bias.csr.unibo.it/maltoni/handbook/>
13. Hong, L., Wand, Y., Jain, A.: Fingerprint image enhancement: Algorithm and performance evaluation. *IEEE-PAMI* **20**(8), 777–789 (1998)
14. Chen, Y., Dass, S.C., Jain, A.K.: Fingerprint quality indices for predicting authentication performance. In: Audio- and Video-Based Biometric Person Authentication, p. 160 (2005). URL http://dx.doi.org/10.1007/11527923_17
15. Xiao, Q., Raafat, H.: Fingerprint image postprocessing: A combined statistical and structural approach. *Pattern Recognit.* **24** (10), 985–992 (1991). URL [http://dx.doi.org/10.1016/0031-3203\(91\)90095-M](http://dx.doi.org/10.1016/0031-3203(91)90095-M)
16. Hung, D.C.D.: Enhancement and feature purification of fingerprint images. *Pattern Recognit.* **26**(11), 1661–1671 (1993). URL [http://dx.doi.org/10.1016/0031-3203\(93\)90021-N](http://dx.doi.org/10.1016/0031-3203(93)90021-N)
17. Shih, F.Y., Pu, C.C.: A skeletonization algorithm by maxima tracking on Euclidean distance transform. *Pattern Recognit.* **28** (3), 331–341 (1995)
18. Farina, A., Kovacs Vajna, Z.M., Leone, A.: Fingerprint minutiae extraction from skeletonized binary images. *Pattern Recognition* **32**(5), 877–889 (1999). URL <http://www.sciencedirect.com/science/article/B6V14-3WMK59F-D/2/bf21218ba618c9f63efb1663ea24a6f6>
19. Fronthaler, H., Kolreider, K., Bigun, J.: Local feature extraction in fingerprints by complex filtering. In: S.Z.Li et al. (ed.) International Workshop on Biometric Recognition Systems – IWBRS 2005, Beijing, Oct. 22–23, LNCS 3781, pp. 77–84. Springer, Heidelberg (2005)
20. Maio, D., Maltoni, D.: Ridge-line density estimation in digital images. In: International Conference on Pattern Recognition, vol I, pp. 534–538 (1998). URL <http://dx.doi.org/10.1109/ICPR.1998.711198>
21. Kawagoe, M., Tojo, A.: Fingerprint pattern classification. *Pattern Recognit* **17**, 295–303 (1984)
22. Bazen, A., Gerez, S.: Systematic methods for the computation of the directional fields and singular points of fingerprints. *IEEE-PAMI* **24** (7), 905–919 (2002)
23. Bigun, J., Bigun, T., Nilsson, K.: Recognition by symmetry derivatives and the generalized structure tensor. *IEEE-PAMI* **26**, 1590–1605 (2004)
24. Nilsson, K., Bigun, J.: Localization of corresponding points in fingerprints by complex filtering. *Pattern Recogn. Lett.* **24**, 2135–2144 (2003)
25. Wegstein, J.H.: An automated fingerprint identification system. Tech. Rep. Special Publication 500-89, National Bureau of Standards, NBS (1982). URL http://www.itl.nist.gov/iad/894.03/fing/Special_Publication_500-89.pdf
26. Novikov, S., Kot, V.: Singular feature detection and classification of fingerprints using Hough transform. In: E. Wenger, L. Dimtrov (eds.) Proc. of SPIE, vol. 3346, pp. 259–269 (1998)
27. Garcia, J.O., Aguilar, J.F., Simon, D., Gonzalez, J., Zanuy, M.F., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.J., Vivaracho, C., Escudero, D., Moro, Q.I.: MCYT baseline corpus: a bimodal biometric database. *IEE Proc. Vision Image Signal Process.* **150** (6), 395–401 (2003). URL http://ieeexplore.ieee.org:80/xpls/abs_all.jsp?isNumber=2825%28prod=JNL&arnumber=1263277&arSt=+395&ared=+401&arNumber=1263277

Fingerprint Identification

- ▶ Fingerprint Indexing
- ▶ Fingerprint Recognition, Overview

Fingerprint Image Compression

► Fingerprint Compression

The focus here is on the performance and limitations of current image enhancement techniques rather than on their algorithmic details. For this purpose, many samples including problematic images and their corresponding enhanced images are presented.

Fingerprint Image Digitalization and Density

Fingerprint images are digitized through either inked-print scanning or live scanning, most often with a resolution of 500 dpi and a depth of 8 bits (i.e., 256 gray levels) in compliance with the NIST standard [1]. The gray level is a value associated with each pixel representing its intensity or luminance. However, the term *density*, the degree of ink thickness on the paper surface, has been used throughout this section for the sake of illustration. Thus, the higher the density, the darker the ridges, and vice versa.

Ideally, the density of pores should be higher than that of valleys and the density of incipient ridges should be lower than that of true ridges. In fact, this is precisely what some feature extractors traditionally use to distinguish pores and incipient ridges from true ridges and valleys. Although most inked-print scanned images have continuous density, some live scanned images exhibit a sparse nature. It has been reported that the effective bit depth of some live scanners is only 2 or 3 bits [2, 3]. Obviously, such loss of information makes the subsequent processes virtually impossible to distinguish the key features.

Fingerprint Image Enhancement

MASANORI HARA
NEC Corporation, Tokyo, Japan

Synonyms

Fingerprint contrast enhancement; Ridge enhancement; Ridge extraction

Definition

Fingerprint image enhancement is the process of applying techniques to emphasize fingerprint images in order to facilitate the identification of ridge valley structures and hence their features.

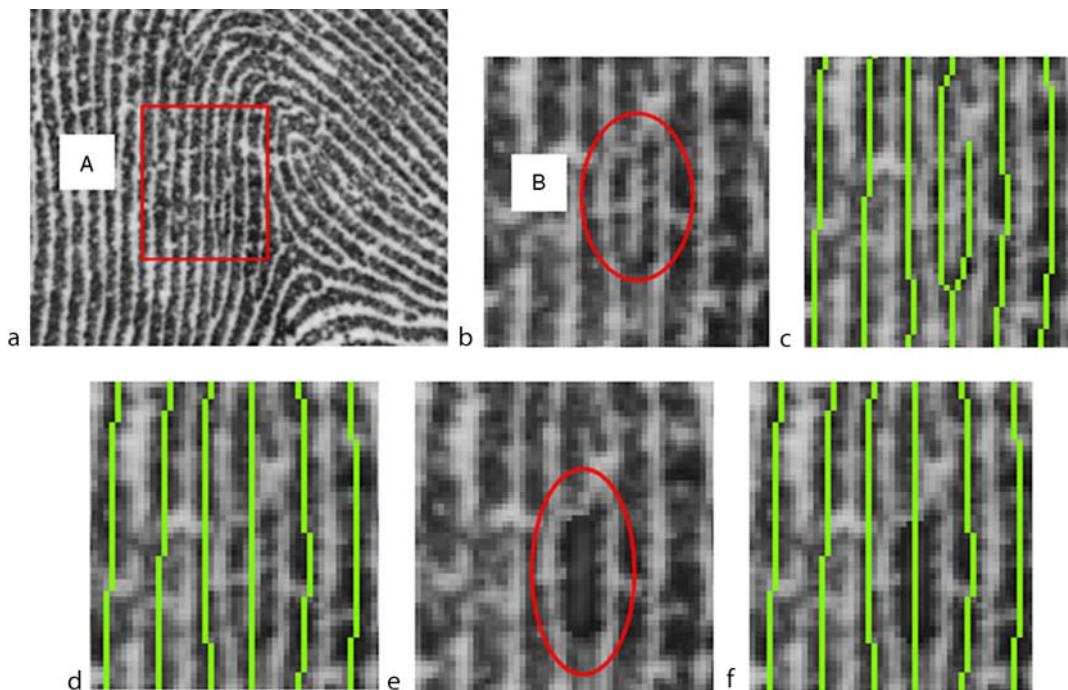
Introduction

Computerized fingerprint feature extractors more or less require some sort of image pre-processing or enhancement to improve perceptibility. In doing so, they need to contend with two major types of problems: one is associated with image contrast such as insufficient dynamic range, and the other is associated with adverse physical factors such as scars, blurs, creases, sweat pores, and incipient ridges. Fingerprint image enhancement aims to minimize the undesired effects caused by such elements in order to extract a sufficient number of reliable features, namely, minutiae and ► **fingerprint singularities** (cores and deltas). Broadly speaking, fingerprint image enhancement encompasses, but is not limited to, the intermediate steps such as contrast enhancement, pore and incipient ridge removal, ridge orientation and frequency estimation, foreground segmentation, and ridge enhancement filtering.

Recognition of Fingerprint Ridge

Since ridge orientation and frequency characterize the local ridge valley structure in the region of interest, the problem of fingerprint ridge recognition essentially simplifies to the task of estimating these two pieces of information. Therefore, local ridge orientation and frequency estimation play a key role in fingerprint ridge recognition.

Local orientation estimation, taking advantage of the fact that ridge orientation does not change suddenly when viewed locally, can “interpolate” ridge orientation even in obscured regions. However, this is not the case in frequency estimation; frequency can become rather unstable when there is a sudden change



Fingerprint Image Enhancement. **Figure 1** An Example of Unstable Frequency Estimation (NIST DB#27 002T).

Notes: (c) and (f) A skeleton image was extracted by one of the traditional algorithms [7] using the contextual ridge enhancement filter with a narrow spacing, which corresponds to higher frequencies. (d) A skeleton image was extracted by the same algorithm as in (c) and (f) with a wide spacing, which corresponds to lower frequencies.

in ridge spacing even in a clear, well-defined region, as illustrated in Fig. 1. The ridges in the region marked by the red square (region A) in Fig. 1(a) are nicely aligned in the vertical direction, whereas frequency significantly changes due to the presence of a spur, alternatively called a whisker, in the region marked by the red oval (region B) in Fig. 1(b). An average inter-ridge spacing is 8.7 pixels (0.435 mm) in region A whereas it is only 5.7 pixels (0.285 mm) in region B, which is narrower than the neighboring region by 35%. For this image, some feature extractions are able to correctly extract this narrow spur as shown in Fig. 1(c) but some others fail (Fig. 1(d)).

Even if the spur in region B is invisible as shown in Fig. 1(e), it is still easy to estimate ridge orientation in the region with a high degree of certainty. However, if this small region is contaminated with noise, most feature extractions incorrectly estimate the frequency in region B to be the same as that of its neighbor, which results in a failure to detect the spur (Fig. 1(f)). An ideal feature extractor should be able to mark this region as “indeterminate” because it is difficult even for human examiners to identify the spur confidently.

Intermediate Steps in Fingerprint Image Enhancement

A typical set of intermediate steps in fingerprint image enhancement includes:

1. Contrast enhancement or normalization.
2. Pore and incipient ridge removal.
3. Ridge orientation estimation.
4. Frequency estimation.
5. Foreground segmentation.
6. Ridge enhancement filtering.

Contrast Enhancement

Whatever features or structures there may be, either local or global, distinctiveness is important to appropriately separate one from the other. The conditions that are preferably satisfied may include uniform background density and a sufficiently wide dynamic range between ridges and valleys/background. If these conditions are fulfilled, a simple stretching and/or

thresholding should suffice. In reality, however, more elaborate and rigorous approaches are needed. Some real issues related to dynamic range are outlined in the following:

1. Uneven dynamic range.

A sample of uneven dynamic range is presented in Fig. 2(a). The ridges on the left (surrounded by the red oval) are substantially lighter than the ones on the right.

2. Uneven valley density.

A sample of uneven valley density is shown in Fig. 2(b). This is a latent image that is lifted from paper. The latent print is impressed on across the regions where letters (“O”, “A”, “N”) are printed. Here, the ridges and valleys cannot be recognized easily because their actual density considerably deviates from their original definition in which “the ridges are dark and the valleys are light”; the density of the valley on the letters is contrarily high and local dynamic range is extremely narrow, whereas the valley density in the plain region is low.

3. Noisy background.

The background containing leftover fingerprint images or stripe patterns resembling fingerprints makes it difficult to isolate true fingerprint patterns. Problematic live scanned images and an inked image are presented in Fig. 2(c, e) and 2(d), respectively.

Contrast enhancement is a technique to accommodate such problems by expanding the dynamic range of ridges and valleys. Adaptive histogram equalization is a popular contrast enhancement technique. Other popular techniques include a simple linear contrast stretching that uses the local minimum and maximum densities, and a density normalization that uses the local density mean and variance [2, 5]. Contrast-enhanced images of the sample problematic images are presented in Fig. 2(a') through (e').

Although contrast enhancement is a powerful tool, it has a drawback, i.e., it boosts background noise at the same time since it cannot selectively enhance only the targeted region unless some additional information is given. As shown in Fig. 2(a'), (c'), (d'), and (e'), the distinguishability of the foreground and background is lower than the original.

However, it should be stressed that it is still imperative to employ contrast enhancement when dealing with poor quality images, mainly latent images

(Fig. 2(b')), for example). Once the ridge valley structure becomes visible, it essentially boils down to the problem of identifying and analyzing ridge continuity.

Pore and Incipient Ridge Removal

Sweat pores are major obstacles in frequency estimation. There are a variety of methods for removing pores or at least for reducing their side effects. Some methods do not remove pores but they remove false minutiae that possibly originated from the pores. Other methods rely on the fact that pores are enclosed by darker pixels as shown in Fig. 3(a), and they can remove typical pores but not problematic pores such as continuous pores and swollen pores as shown in Fig. 3(b). Ridge structures such as lakes and spurs are easily confused with pores, leading to miscalculation of frequency if they are falsely filled in.

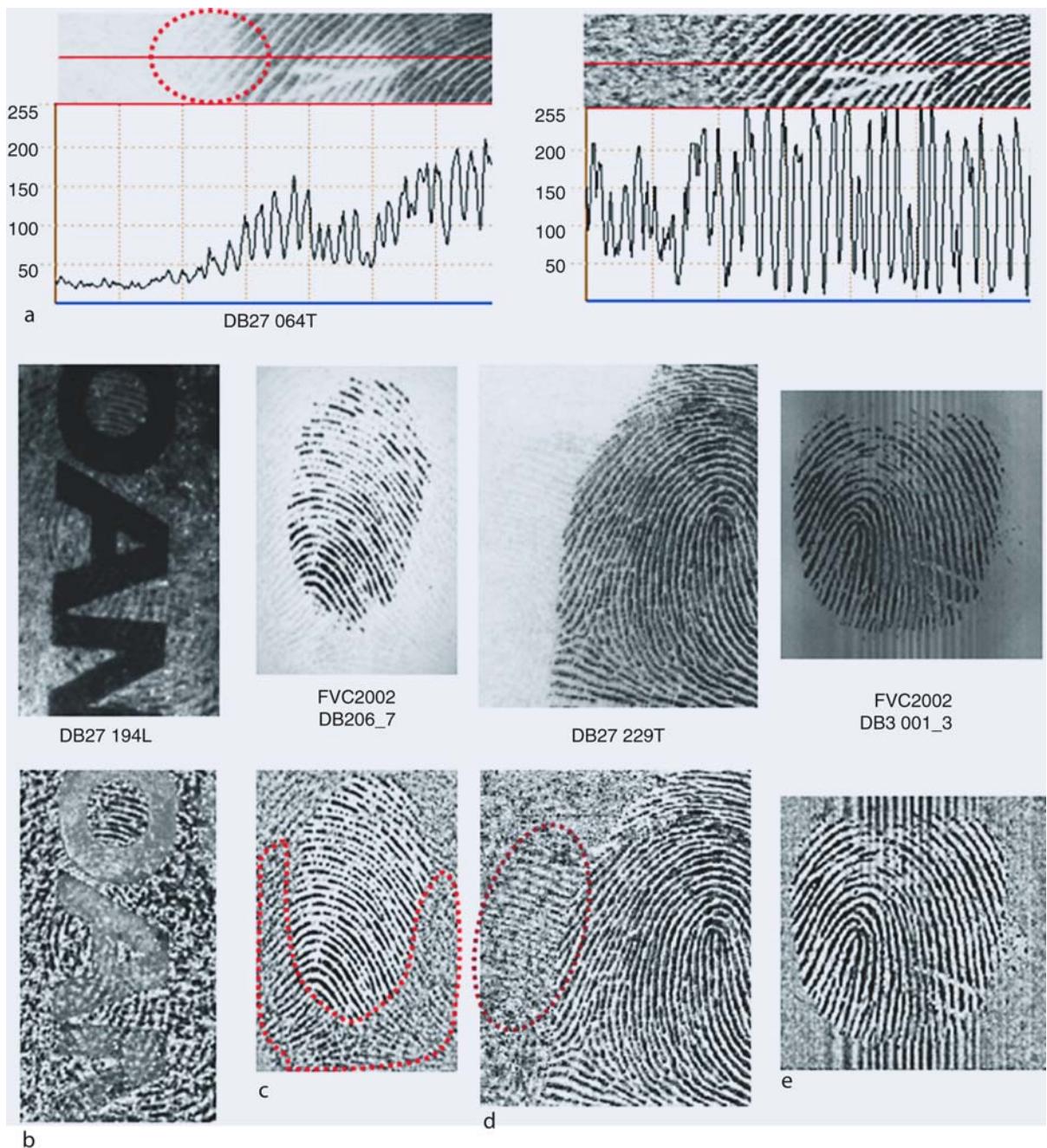
Incipient ridges are another obstacle in frequency estimation. The significant incipient ridges as shown in Fig. 3(c) can easily fool frequency estimation algorithms.

These two factors have not yet been fully explored, and their distinguishability plays an important role in improving fingerprint matching accuracy.

Ridge Orientation Estimation

Ridge orientation estimation is a fingerprint-specific image processing technique. A ridge orientation estimation algorithm was developed for a FBI system in the 1960s. In the 1960s and 1970s, many ridge orientation estimation algorithms set “slits” of predetermined orientations (8, 12, or 16 quantized orientations) and analyzed the density response [6–8]. The orientation slit having a higher amount of density change is indicative of the slit running perpendicular to the direction of the ridge flow. Similarly, the slit with a lower amount of density change is indicative of the slit running parallel to the direction of the ridge flow.

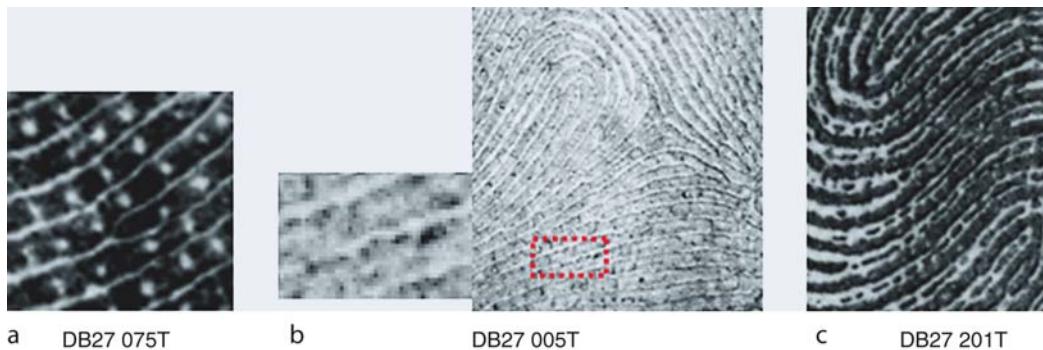
In the 1980s, more sophisticated methods were introduced to extract ridge orientation such as a method based on the gradient of two-dimensional vectors whose components are derivatives of densities at horizontal and vertical orientations [2, 5], and a method based on the two-dimensional Fourier transform [9, 10].



Fingerprint Image Enhancement. **Figure 2** Examples of Contrast-related Problematic Images. Notes: (a') through (e') The images were contrast-enhanced with one of the local adaptive stretching methods specialized for fingerprints [4].

In this process, the confidence level of ridge orientation is calculated. The difference in density fluctuation between the estimated orientation and its orthogonal orientation can be a base for confidence, and the power spectrum is another in the case of the Fourier analysis.

Since all these techniques estimate ridge orientation locally, the influence of adverse factors such as scars and smudges are not negligible and often lead to wrong estimation. In order to correct such anomaly, local orientation is examined for validity and re-estimated from its neighbor. This process is called



Fingerprint Image Enhancement. [Figure 3](#) Pores and Incipient Ridges.

ridge orientation smoothing, and several such techniques have been proposed [2, 5, 9, 11].

Ridge orientation smoothing also has a drawback. The orientation in the region where the ridge flow is not stable (e.g., in the proximity of the core and delta) cannot be properly corrected due to its interpolative nature. It also fails and propagates errors if the overall estimation quality is low because it is based on the assumption that the majority of orientations of neighboring regions are indeed correct.

Examples of problematic images in orientation estimation are presented in [Fig. 4](#). Orientations in the red oval in [Fig. 4\(a'\)](#) and [\(b'\)](#) are incorrectly estimated because of smudges and fragmented ridges.

One of suggested methods to improve estimation accuracy is to use global pattern types and prior knowledge of ridge flow. Once the core and delta have been extracted with high confidence, the global pattern shape can be estimated. This information can help estimate and adjust local ridge orientation more accurately.

Frequency Estimation

Frequency estimation is another fingerprint-specific image processing technique. Frequency is defined as the number of ridges per unit length and is often interchangeably referred to as the inverse of the inter-ridge distance. It is far more difficult to estimate than orientation, and that explains why most feature extractions in the 1960s and 1970s did not fully exploit this information.

In the 1980s, frequency analysis such as the two-dimensional Fourier transform was proposed

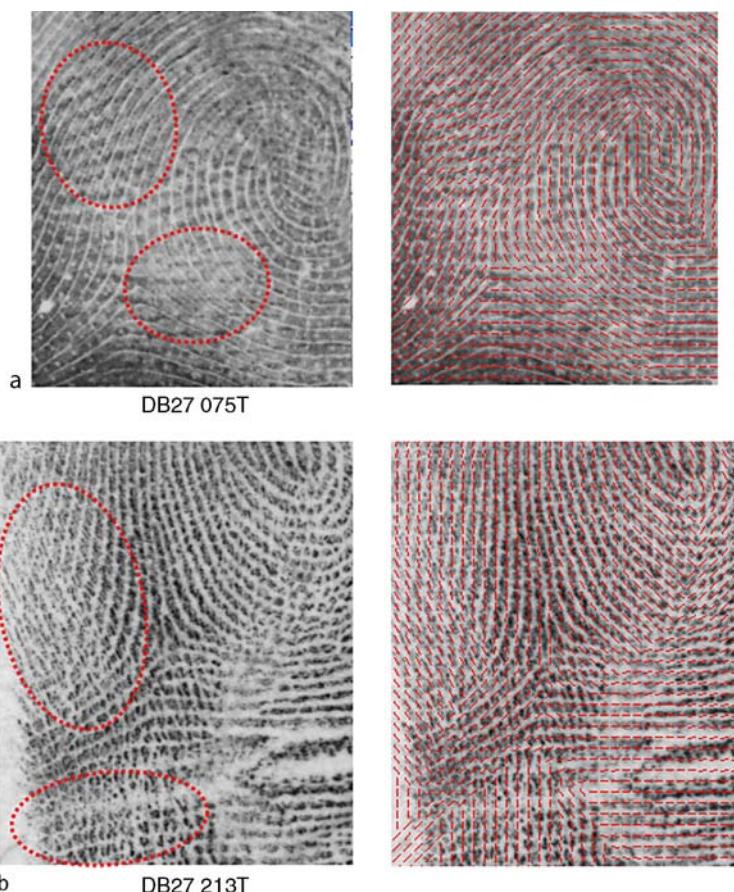
to estimate frequency [9, 10]. Another technique of frequency estimation analyzes peak intervals from gray-level profile orthogonal to the ridge orientation [2, 5].

In [Fig. 5\(a\)](#) an example of a problematic image with a sudden frequency change, denoted by the red oval is presented. In [Fig. 5\(b\)](#), the true frequency is reflected via some manual correction, and [Fig. 5\(c\)](#) shows an example of automatically estimated frequency image using one of the latest algorithms [12]. Dark density pixels correspond to the region where the inter-ridge spacing is narrow, that is, frequency is high. It can be observed that the frequency of the area with very narrow inter-ridge spacing is falsely estimated to be halved from its true value.

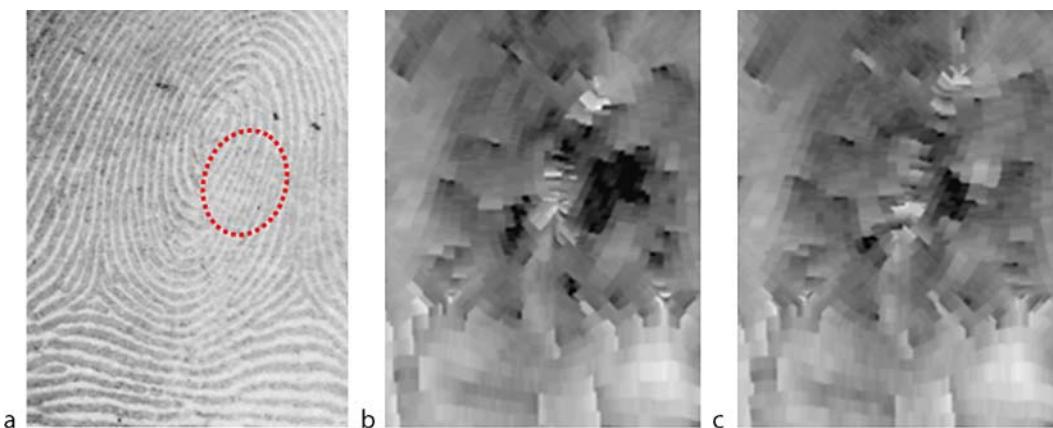
It is known that the presence of a minutia affects the structure of its surroundings and hence the corresponding local frequency. This often becomes a problem in frequency estimation where a strong frequency smoothing aimed to facilitate the estimation process can adversely eliminate true minutiae.

Foreground Segmentation

It is natural to conduct minutia extraction only in the foreground region to minimize the possibility of extracting false minutiae. Foreground segmentation aims to distinguish the fingerprint ridge region from the background. Some methods rely heavily on the confidence of the ridge orientation to define the foreground, whereas others rely on gray-level statistics as well. As already explained, gray-level analysis is not an ideal approach when dealing with very low quality images such as the ones shown in [Fig. 2](#).



Fingerprint Image Enhancement. [Figure 4](#) Problematic Images in Ridge Orientation Estimation. Notes: (a') and (b') A ridge orientation image was extracted by one of the traditional algorithms [7].



Fingerprint Image Enhancement. [Figure 5](#) Problematic Images in Frequency Estimation (NIST DB#27 073T). Notes: (b) True frequencies were calculated from an ideal skeleton image, which was manually generated so that skeleton curves correctly coincided with the original ridges. (c) Frequencies were calculated from an automatically extracted skeleton image using one of the recent algorithms.

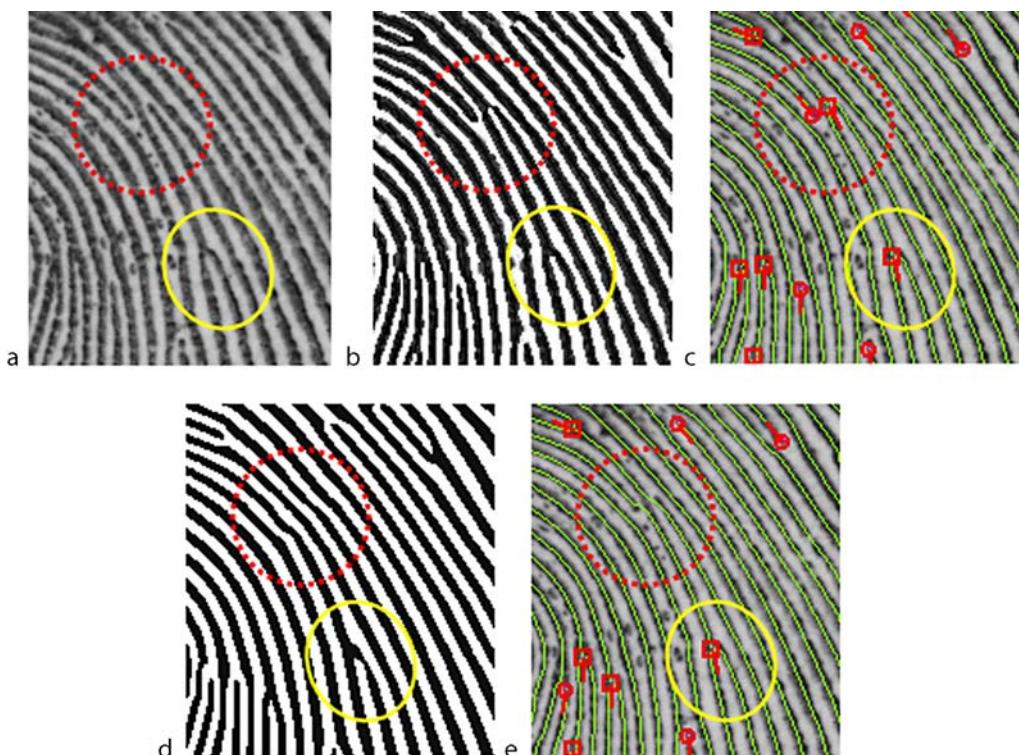
Ridge Enhancement Filtering

Ridge enhancement filtering is another fingerprint-specific image processing technique. In the 1960s and 1970s, most filtering techniques were labeled contextual. They used filtering masks similar to the ones used to estimate orientation, either fixed size or pre-determined variable frequencies. However, it was difficult for these techniques to flexibly adapt to very narrow or very wide ridges and spacing [6–8]. In the 1980s, a more sophisticated method based on the two-dimensional Fourier transform was proposed [9, 10]. In the 1990s and 2000s, Gabor filtering and wavelet filtering were introduced [2, 5, 13].

Conceptually, ridge enhancement filtering aims to “enhance” ridges by generating stripe patterns from scratch using the previously estimated orientation and frequency. Strong enhancement is effective for low quality images but at the risk of destroying the original ridge structure. The strength of filtering thus

needs to be controlled adaptively and depends on the field in which it is used: law enforcement and non-law enforcement. In the former case, the original ridge structure needs to be preserved as much as possible in order to improve compatibility with the examiners’ definition of minutiae since it still relies on manual processing such as latent minutia coding. This is important to improve latent-print matching accuracy, especially for fragmental latent prints with few minutiae. In order to match such latent prints, even unstable minutiae need to be incorporated to increase chances of hit. On the contrary, in the latter case, which is fully automatic, neither the original ridge structure has to be preserved nor is compatibility with the examiners’ definition critical.

With respect to minutia preserving ability, there are two types of minutiae to be considered: stable minutiae and unstable minutiae. The stable minutia is a minutia that is topologically isolated from other minutiae with no chance of interfering with other minutiae.



Fingerprint Image Enhancement. **Figure 6** Unstable and Stable Minutiae (NIST DB#27 076T). Notes: **(b)** The ridge image **(a)** was enhanced by one of the popular algorithms [9] with a relatively weak enhancement parameter. **(c)** A skeleton image and minutiae were automatically extracted from the image in **(b)**. **(d)** The image in **(b)** was enhanced by one of the popular algorithms [9] with a relatively strong enhancement parameter. **(e)** A skeleton image and minutiae were automatically extracted from the image in **(d)**.

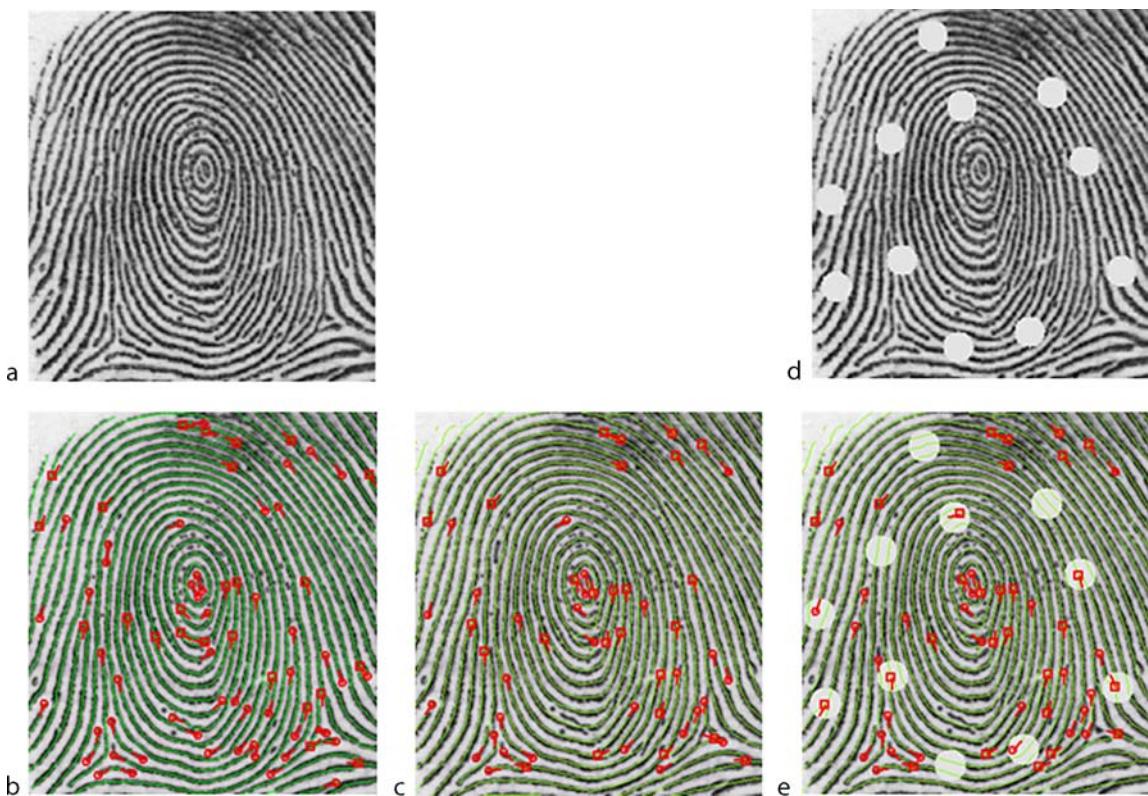
The unstable minutia is a minutia that may either remain unchanged or completely disappear depending on the physical conditions of its surroundings. Crossover minutiae are a typical example of unstable minutiae. In Fig. 6(a) a stable minutia is presented in the yellow circle and an unstable minutia in the red circle. Notice how the different levels of enhancement affect the extraction results. If the strength of the ridge enhancement filter is relatively mild, the crossover ridge structure and the corresponding minutiae are maintained (Fig. 6(c)). On the other hand, the crossover minutiae completely disappear when a strong filter is applied (Fig. 6(e)).

Despite this drawback, however, it is still beneficial to adopt strong filtering since it is capable of consistently extracting stable minutiae even from poor quality images as illustrated by the following example. The image in Fig. 7(b) represents an ideal, manually edited

minutiae of the image Fig. 7(a), containing a total of 76 minutiae, 55 of which are stable and 21 are unstable. The image in Fig. 7(d) is an artificially produced poor quality image by covering it with several circular “patches.” When a strong filter is applied to the images in Figs. 7(a) and (d), most of the 55 stable minutiae are correctly extracted as shown in Figs. 7(c) and (e), respectively. It should also be noted that this method is especially effective when the area of the overlapping region between the two images is large enough in which a sufficient number of stable minutiae exist.

Thus, filtering strength depends on the operational strategy, requirements, and target image characteristics.

Once fingerprint ridges are suitably enhanced, ► **fingerprint binarization** is then conducted to produce a black and white image and, finally, ► **fingerprint skeletonization** to generate a skeleton image.



Fingerprint Image Enhancement. Figure 7 Effects of Strong Ridge Enhancement Filter (NIST DB#27 076T).

Notes: (b) The ideal skeleton image was manually generated so that skeleton curves correctly coincided with the original ridges. Then, minutiae were extracted from the ideal skeleton image. (c) The ridge image in (a) was automatically enhanced by one of the popular algorithms [9] with a relatively strong enhancement parameter. Then, a skeleton image and minutiae were automatically extracted from that enhanced ridge image. (e) The ridge image in (d) was automatically enhanced by one of the popular algorithms [9] with a relatively strong enhancement parameter. Then, a skeleton image and minutiae were automatically extracted from that enhanced ridge image.

Summary and Future Improvement

Fingerprint image enhancement is a very effective tool for improving ridge clarity. Undoubtedly, improvement in matching accuracy reported in the past two to three decades can be attributed to innovation in image enhancement techniques. Unfortunately, it is far from true if considered in terms of how close the automated fingerprint recognition got to the ability of human perception. This is because the current techniques that heavily rely on ridge orientation and frequency (and whatever information one can think of) are not capable of perceiving a fingerprint image as a fingerprint but just a collection of gray-scale pixels, and the circumstance has not changed in the course of over 40 years of research. This may change in the future if a leap forward in the computational neuroscience reveals the mechanism of human pattern recognition, but for the time being, a goal pro tempore is probably to find a way to extract information from unmodified gray images to avoid side effects of the image enhancement as far as possible.

Acknowledgments

The sample images are courtesy of the NIST and the University of Bologna [14, 15]. The author would like to thank Amane Yoshida for proofreading and rewriting in English. For more information on the technical details, refer to ‘Fingerprint Analysis and Representation’ in *Handbook of Fingerprint Recognition* by Maltoni Et al. [2].)

Related Entries

- ▶ Fingerprint Classification
- ▶ Fingerprint Features
- ▶ Fingerprint Image Quality

References

1. ANSI/NIST-ITL 1-2007 Fingerprint Standard – <http://fingerprint.nist.gov/standard/index.html>
2. Maltoni, D. et al.: *Handbook of fingerprint recognition*. Springer (2003)
3. Xia, X., O’Gorman, L.: Innovation in fingerprint capture devices. *Pattern Recognit.* **36**, 361–369 (2003)

4. Hara, M.: Image Density Conversion Method, Image Enhancement Processor, and Program Thereof (USP 20080050030A1 – Pending)
5. Hong, L. et al.: Fingerprint image enhancement: Algorithm and performance evaluation. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 777–789 (1998)
6. Stock, R.: Automatic fingerprint reading. In: The 1972 Carnahan Conference, on Electronic Crime Countermeasures, April 19–21, 1972
7. Asai, K. et al.: Automatic fingerprint identification. SPIE vol. 182, Imaging Application for Automated Industrial Inspection & Assembly (1979).
8. Capello, R. et al.: Method and apparatus for contextual data enhancement (USP 4,876,726)
9. Kamei, T. et al.: Image filter design for fingerprint enhancement. In: Proceedings International Symposium on Computer Vision, 109–114 (1995).
10. Chikkerur, S. et al.: Fingerprint enhancement using STFT analysis. *Pattern Recognit.* **40**, 198–211 (2007), 109–114 (1995)
11. Funada, et al.: System and method for processing fingerprint/palmprint image (USP 7,027,626)
12. Hara, M.: System for recognizing fingerprint image, method and program for the same (USP 20070036401A1 – Pending)
13. Paul, A. et al.: A study of image enhancement techniques for fingerprint identification. Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS’06) (2006)
14. NIST (National Institute of Standard and Technologies) Special Database #27 – <http://www.nist.gov/srd/nistsd27.htm>
15. FVC2002 Second Fingerprint Verification Competition Database – <http://bias.csr.unibo.it/fvc2002/>

Fingerprint Image Quality

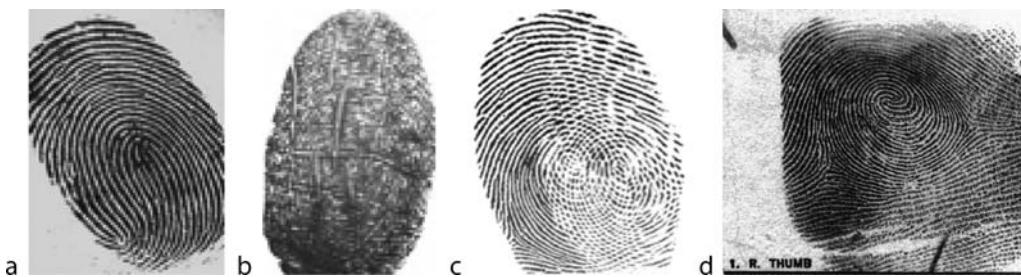
ELHAM TABASSI, PATRICK GROTHÉ
National Institute of Standards and Technology,
MD, USA

Synonym

Expected performance or utility of fingerprint image in an automated comparison environment

Definition

The intrinsic characteristic of a biometric signal may be used to determine its suitability for further processing by the biometric system or assess its conformance to preestablished standards. The quality of a biometric



Fingerprint Image Quality. Figure 1 Good quality fingerprint images (a) have clear pattern of ridge and valleys; however, poor quality fingerprint images (b) do not have easily distinguishable patterns. Poor quality images result in spurious and missed features, thus degrading the performance of the overall system. Poor quality samples can be due to distorted source like abraded skin (b), distortion in one or more steps of the process, e.g., capture (residual fingerprints on the platen in (c)) or compression, or low character source, the sample may subjectively be assessed as "good" quality, but a matcher may not be able to match it to its mate (d).

signal is a numerical value (or a vector) that measures this intrinsic attribute. Quality score is a quantitative expression of the utility, or predicted performance of a biometric sample in a comparison environment. This means that finger image quality scores should correlate to the observed false match and ► **false non-match rates** of the samples.

Introduction

With an increase in the need for reliable identity authentication, biometric recognition systems have been increasingly deployed in several different applications: government applications such as national ID card, border control; and commercial applications, such as physical access control, e-commerce, or mobile phone. Among all biometric modalities, fingerprint recognition is the most widespread due to its permanence and uniqueness [1].

A fingerprint is a pattern of friction ridges on the surface of a fingertip. A good quality fingerprint has distinguishable patterns and features that allow the extraction of features, which are useful for subsequent matching of fingerprint pairs. This viewpoint may be distinct from the human conception of quality. If, for example, an observer sees a fingerprint with clear ridges, low noise, and good contrast then he or she might reasonably say it is of good quality. However, if the image contains few minutiae points then a minutiae-based matcher would underperform. Thus, in the context of automated matching, the term quality should not be used to refer to the fidelity of the sample, but instead

to the utility of the sample to an automated system. Figure 1 shows examples of good and poor quality fingerprint images.

Automatically and consistently determining the quality of a given biometric sample for identification and/or verification is a problem with far-reaching ramifications. If one can identify low quality biometric samples, this information can be used to improve the acquisition of new data. This same quality measure can be used to selectively improve an archival biometric database by replacing poor quality biometric samples with better quality samples. Weights for multimodal biometric fusion can be selected to allow better quality biometric samples to dominate the fusion. All of these applications require that the quality of the biometric sample be determined prior to identification or verification. Most of these applications also require that quality of the biometric sample be computed in real-time during data acquisition.

Fingerprint Image Quality

Performance of an automated fingerprint recognition system is greatly affected by the degree of imperfection present in the finger image. Accuracy of current fingerprint recognition systems is high when high-quality samples are being compared [2] (Note that according to Minutia Interoperability Exchange Test 2004 (MINEX04) report, best single finger proprietary fingerprint recognition system performed at 0.0047 false non-match rate at 1% **false match rate**). However, performance degrades substantially as quality drops.

Although only a small fraction of input data are of poor-quality, the bulk of recognition errors can be attributed to poor-quality samples.

Degradation in fingerprint image quality reduces the amount of identifiable information in a fingerprint. Poor quality images cause spurious and missed features which decrease the likelihood of a correct verification and/or identification, while extremely poor quality samples might be impossible to verify and/or identify. The variation in performance for different quality levels is shown in Fig. 2. The five traces of Detection Error Tradeoff (DET) curves correspond to five different levels of quality as measured by NIST Fingerprint Image Quality (NFIQ) [3, 4]. NFIQ is an integer between 1 and 5 where 1 represents the highest quality and 5 the lowest (unusable) quality.

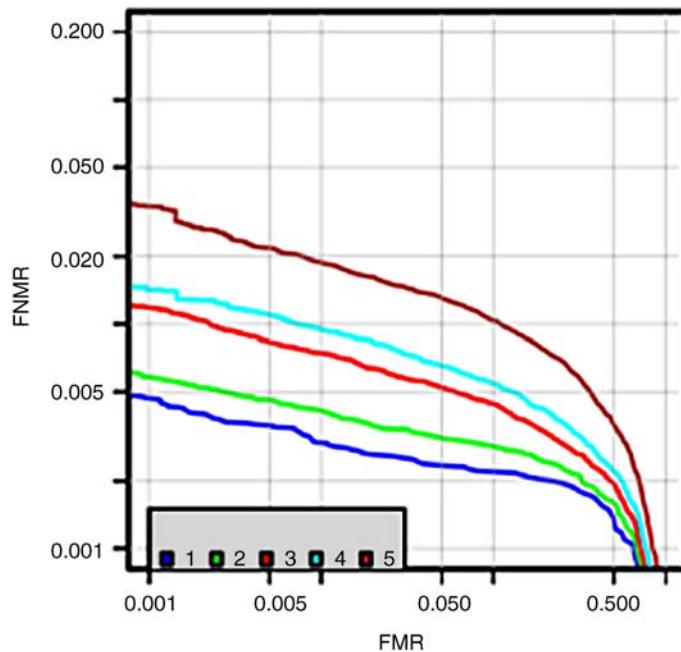
Several factors affect the quality of fingerprint images: user's skin condition, improper finger placement, scanner limitation or imperfection, impurities on the scanner surface and others. The cause of these imperfections can be classified in four groups: (1) *impairments in the source of biometric characteristics*: like scars, blisters, skin conditions such as wet or dry, age, occupation, etc.; (2) *user behavior*: such as improper finger placement, e.g., rotating finger or placing only tip of a finger

which cause capturing insufficient area of finger image; (3) *imaging*: e.g., low contrast, distortion, sampling error, insufficient dynamic range, etc.; and (4) *environment*: such as temperature, humidity, or unclean platen.

Fingerprint Image Quality Measures

It is widely accepted that a statement of a biometric sample's quality should be related to its recognition performance. That is, a quality measurement algorithm takes a signal or image, \mathbf{x} , and produces a scalar, $q = Q(\mathbf{x})$, which is predictive of error rates associated with the verification or identification of that sample. This predictive value of quality measures may be imperfect but valuable nevertheless. It should be noted that operationally the requirement for a scalar is not necessary: a vector could be stored and could be used. The fact that quality has historically been conceived of as scalar is a widely manifested restriction [5].

International Standards Organization (ISO) has recently established a biometric sample quality draft standard [6], in which quality score of a biometric sample is defined as predicted performance of the



Fingerprint Image Quality. Figure 2 Quality ranked detection error trade-off characteristics. Five traces correspond to five NFIQ levels. Fingerprint images with NFIQ=1 (highest quality) cause lower recognition error than images with NFIQ=5 (lowest quality).

sample in a comparison environment. It considers three components of quality: (1) *character*, which refers to quality of inherent physical features of the source, for example, a fingerprint with a scar has low character; (2) *fidelity*, which is the degree to which a sample is an accurate representation of its source, for example, distortion degrades fidelity; and (3) *utility*, which refers to contribution of a sample to the overall biometric recognition error rates and is related monotonically to the performance of biometric matchers. Character and fidelity of a sample positively or negatively impact the utility of the sample.

There are several fingerprint analysis approaches that gauge character and fidelity of fingerprint images. These measures are then summarized into a scalar (or a vector) quality score that is indicative of utility of the sample. Broadly fingerprint image analysis can be divided into local and global analysis methods [7]. Fingerprint local structure constitutes the main texture-like pattern of ridges and valleys within a local region while valid global structure puts the ridges and valleys into a smooth flow for the entire fingerprint. The quality of a fingerprint image is determined by both its local and global structures. Local feature analysis methods partition an image into nonoverlapping blocks and assign a quality score to each block which indicates the amount of useful information in that block for subsequent matching. Final image quality score can be computed by combining quality scores of the blocks. Global feature analysis examines continuity and uniformity of ridge-valley

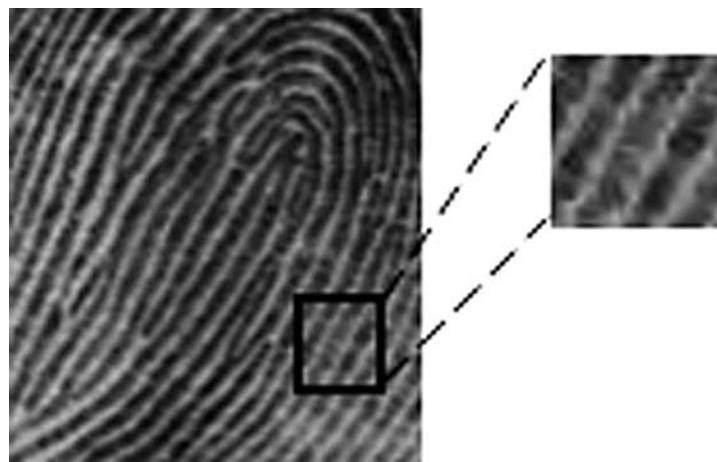
structure of a fingerprint image in a holistic manner and computes a global measure of fingerprint quality.

Global and local quality measures could be combined to obtain final quality score of a fingerprint image such that the overall quality score is a measure of matchability of the sample in an automated matching process, i.e., the derived quality score should be related to the biometric error rates that is likely to be realized when the sample is matched.

F 1. *Local Analysis*

To locally analyze a fingerprint image, it is divided into grids of blocks (Fig. 3). For each block, local features such as directional flow of ridges are computed which are then summarized into a quality score representing quality of the block. Each block should be large enough to contain sufficient ridge-valley information, at least two ridges per block. For example, for a fingerprint with a resolution of 500 ppi, each block could be 32×32 pixels. An overview of existing local analysis methods follows.

- Orientation certainty field:* A fingerprint image within a small block generally consists of ridges (dark pixels) separated with valley (light pixels) lines along the same orientation. High-quality blocks of a fingerprint image contain consistent ridge (or valley) orientation. Local angle information in each block can be used to compute local features. Lim et al. [8] computed energy concentration along the dominant direction of



Fingerprint Image Quality. Figure 3 Local analysis consists of partitioning a fingerprint image into small blocks. Local features such as orientation consistency or directional flow are extracted from each block. These features convey information useful for comparison of the image and therefore indicate quality of the block.

ridges by computing the ratio between two eigenvalues of the covariance matrix of a block's gradient vector. It gives an indication of how strong the energy is concentrated along the ridge-valley orientation. Chen et al. [9] measured orientation coherence in each block using gradient of the gray level image.

- b. *Ridge-valley structure:* Well-formed and clearly visible ridges are essential to the reliable detection of ridge endings and bifurcations, also known as minutia points. Ridges that are too close or too far apart, or ridges that are unreasonably thick or thin indicate that the finger image may not have captured properly, due to, e.g., pressing too hard or too soft (Fig. 4). Shen et al. [10] applied Gabor filter to image sub-blocks, to identify blocks with clear repetition of ridge and valley pattern as good quality blocks.
- c. *Pixel intensity or Directional contrast:* Region of good quality exhibits high directional contrast, which means that the ridges and the valleys are well separated with regard to gray values. High-quality blocks will exhibit large variance in gray levels while low-quality blocks will show small variance. [11–13] assess quality of each block based on its pixel intensity. Bolle et al. [14] used ratio of directional area to other nondirectional area as a quality measure.
- d. *Power Spectrum:* Ridge and valley structure in a high-quality block forms a periodic signal, which can be approximated either by a square wave or a sinusoidal wave with its frequency

lie in certain range. In frequency domain, a square wave exhibits a dominant frequency with sideband frequency components (sinc function), and a sinusoidal wave consists of one dominant frequency and minimum components at other nondominant frequencies. Therefore, existence of a dominant frequency component plus its frequency are indicative of high-quality blocks of fingerprint image. Poor quality blocks will not exhibit a dominant frequency or it will be out of the normal range of ridge frequency [12]. Hong et al. [15] modeled the ridge and valley pattern as sine wave, and computed the amplitude, frequency as well as the variance of the sine wave to decide the quality of the fingerprint. Nill and Bouzas [16] propose an objective image quality based on the digital image power of normally acquires scenes. Their system is designed to assess the quality of digital images and can be applied to fingerprint as well.

- 2. *Global Analysis* A good quality fingerprint exhibit smooth changes in ridge orientation across the entire fingerprint image except when a core or delta point occurs. Ratio of ridge to valley thickness should also be fairly constant throughout the whole image. [8] used local angle information in each block to assess continuity in orientation field between neighboring blocks and uniformity of ridge to valley thickness ratio. Chen et al. [13] computed a block's absolute difference in the orientation angle with its neighboring blocks as a measure of



Fingerprint Image Quality. Figure 4 Examples of (a) good, (b) thin, and (c) thick ridge structure. (b) and (c) pose challenge to automated matching system and hence are of lower quality than (a).

smoothness of the change in orientation angles among blocks. As mentioned earlier, the ridges of a finger image can be locally approximated by one sine wave with its frequency in a certain range. A region of interest (ROI) of the spectrum is defined as an annular region with radius ranging between the minimum and maximum typical ridge frequency values. For a more robust ridge structure (i.e., the better image quality) the energy will be more concentrated within the ROI. [9] measured the energy concentration in ring-shaped regions of the ROI by employing bandpass filters to extract the energy in each frequency band. Good quality images will have the energy concentrated in few bands while poor quality fingerprints will have a more diffused distribution.

3. *Overall Fingerprint Image Quality: prediction of performance:* It is desirable to combine local and global quality features into one scalar or a vector of quality such that the overall fingerprint image quality is related to the expected false match and false non-match of the image. The summarization can simply be the percentage of blocks classified as “good” or “bad” quality after a local analysis, or more elaborate combination methods such as weighted average of local qualities. For example, higher weights could be assigned to blocks closer to the centroid of a fingerprint since features extracted from blocks near the centroid have more useful and reliable information [9, 11]. Use of a classifier to nonlinearly combine local and global features was first proposed by Tabassi et al. [3, 4]. The method called NIST Fingerprint Image Quality NFIQ [3, 4] was developed to predict how far a genuine score would lie from its impostor distribution and is thus effective at improving false rejections while suppressing false acceptance errors. NFIQ extracts minutia, assigns a quality value to each minutia point, and measures orientation field, pixel intensity, and directional map to compute the following local and global features: number of foreground blocks, number of minutia, number of minutia that have quality value better than certain thresholds, percentage of foreground blocks of excellent, good, fair, and poor quality. A neural network was trained to classify the computed feature vectors into five levels 1–5 where NFIQ = 1 is the best quality and NFIQ = 5 is the

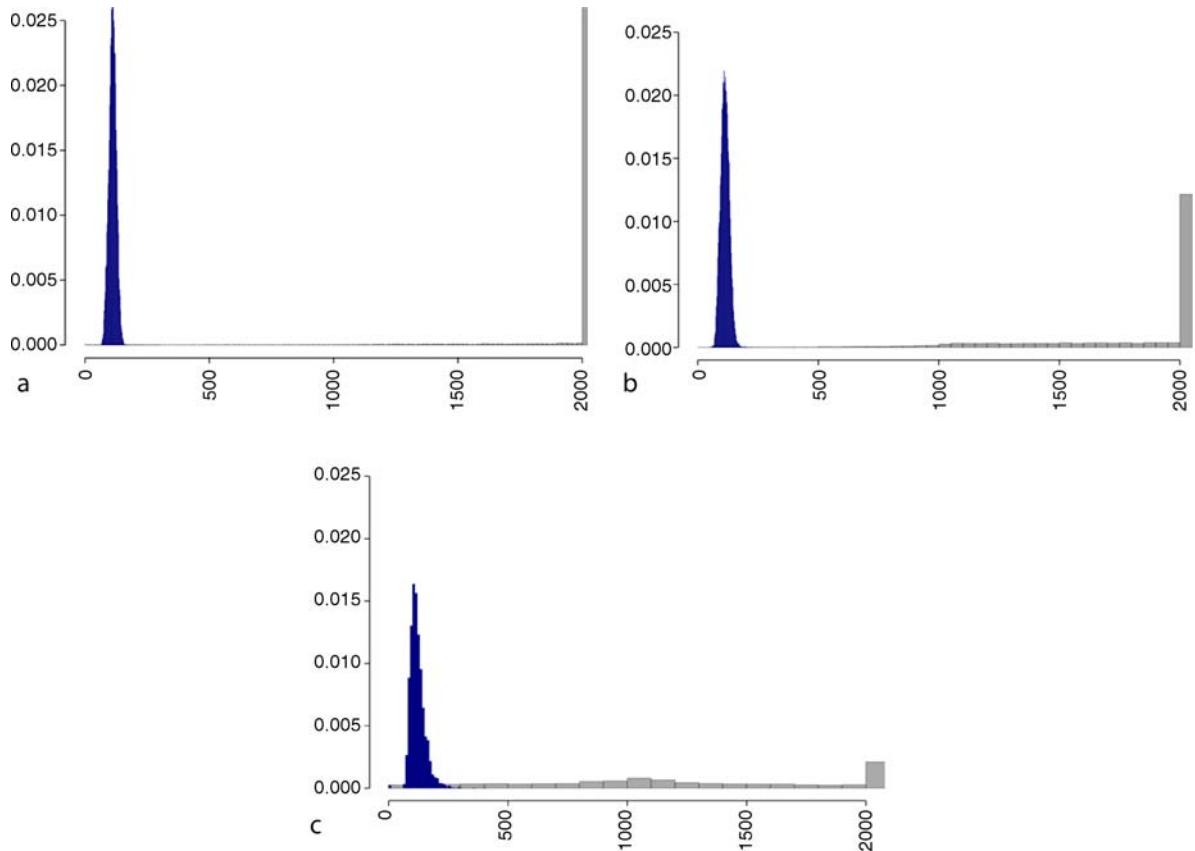
lowest quality. Figure 2 shows that the highest recognition performance is achieved for the best quality samples (NFIQ=1), and samples with lowest quality (NFIQ=5) have the lowest performance. The plots of Fig. 5 show, respectively, the genuine and impostor distributions for NFIQ values 1 (excellent quality), 3 (average quality), and 5 (poor quality). The overlapping of genuine and impostor for the poorest NFIQ (i.e., NFIQ = 5) means higher recognition errors for that NFIQ level while the almost complete separation of the two distributions for the best quality samples (i.e., NFIQ = 1) indicates lower recognition error. Source code for NFIQ algorithm can be found in [17].

Applications of Biometric Quality Values

This section describes the roles of a sample quality measure in the various contexts of biometric operations. The quality value here is simply a scalar summary of a sample that is taken to be some indicator of matchability. These uses of biometric sample quality are not fingerprint specific and can be generalized to other modalities like face or iris.

1. *Enrollment Phase Quality Assessment* Enrollment is usually a supervised process, and it is common to improve the quality of the final stored sample by acquiring as many samples as are needed to satisfy either an automatic quality measurement algorithm, a human inspector (a kind of quality algorithm), or a matching criterion (by comparison with a second sample acquired during the same session). Our focus on automated systems’ needs is warranted regardless of analyses of these other methods, but the authors do contend that naive human judgment will only be as predictive of a matcher’s performance as the human visual system is similar to the matching system’s internals, and it is not evident that human and computer matching are functionally comparable.

Specifically, human inspectors may underestimate performance on overtly marginal samples. Certainly human inspectors’ judgment may be improved if adequate training on the failure modes and sensitivities of the matcher is given to the



Fingerprint Image Quality. **Figure 5** Probability density of impostor scores is shown in blue and probability density of genuine scores is shown in gray. There is a higher degree of separation between the genuine and impostor distribution for better quality samples as measured by NFIQ. **(a)** Best. **(b)** Middle. **(c)** Worst.

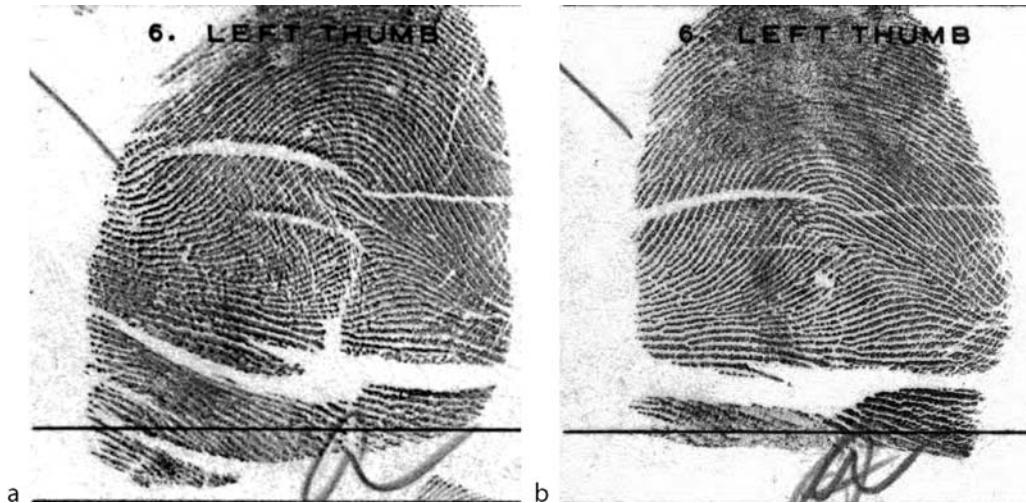
inspector, but this is often prohibitively expensive or time consuming and not scalable. Immediate matching also might not be predictive of performance over time because same-session samples usually produce unrealistically high match scores. For instance, Fig. 6 shows an example of two same-session fingerprint images that were matched successfully by three commercial vendors despite their obvious poor quality.

In any case, by viewing sample acquisition as a measurement and control problem in which the control loop is closed on the quality measure, a system gains a powerful means of improving overall sample quality.

2. *Quality Assurance* Finger image quality assessment algorithms may be used to monitor quality across multiple sites or over time. This is useful to signal possible performance problems ahead of some subsequent matching operation. Quality values

may be aggregated and compared with some historical or geographic baselines. Use of quality values in this role has been documented in [18]. The National Institute of Standards and Technology (NIST) has published a technical guidance toward quality summarization [19]. Quality summarization addresses the important issue of enterprise quality-assurance surveying by providing tools on how to combine quality scores of individual samples into one scalar representing quality of the whole database. Such a function would support identification of, e.g., defective sensors, underperforming sites, and seasonal or secular trends.

3. *Verification Quality Assessment* During a verification transaction, quality can be improved by closing an acquire–reacquire loop on either a match-score from comparison of new and enrollment samples or on a quality value generated without matching. Indeed it is common to implement an



Fingerprint Image Quality. **Figure 6** Example of same session captures of single finger that despite their poor quality (NFIQ=5) were matched correctly by three leading commercial matchers.

“up to three attempts” policy in which a positive match is a de facto statement that the sample was of good quality – even if the individual happens to be an impostor. Depending on the relative computational expenses of sample matching, reacquisition, and quality measurement, the immediate use of a matcher may not be the best solution. The key difference here (as compared with the enrollment-phase) is that quality values of both the enrollment and verification samples can be used to predict performance. This two-dimensional problem is distinct from the enrollment case where only one quality value is used.

4. *Identification Quality Assessment* Quality measurement in identification systems is important for at least three reasons. First, many users often do not have an associated enrollment sample. So a one-to-many match will be an inefficient and inconclusive method of stating whether the authentication sample had high quality. Second, in negative identification systems where users with an enrolled sample are motivated to evade detection, quality measurement can be used to detect and prevent submission of samples likely to perform poorly [20], which may help prevent attempts at spoofing or defeating detection. Third, identification is a difficult task: it is imperative to minimize both the false non-match rate (FNMR) and the false match rate (FMR). To the extent that consistently high-quality samples will produce high genuine scores, a high matching

threshold can be used and this will collaterally reduce FMR. But in large populations FMR becomes dominant, and this raises the question: can a quality apparatus be trained to be directly predictive of false match likelihood?

5. *Differential Processing* Quality measurement algorithms can be used to alter the subsequent processing of a sample. Such conditional activity are categorized as follows.
 - a. *Pre-processing Phase*
An identification system might apply image restoration algorithms or invoke different ► **feature extraction** algorithms for samples with some discernible quality problem.
 - b. *Matching Phase*
Certain systems may invoke a slower but more powerful matching algorithm when low-quality samples are compared.
 - c. *Decision Phase*
The logic that renders acceptance or rejection decisions may depend on the measured quality of the original samples. This might involve changing a verification system’s operating threshold for poor quality samples. For example, in multi-modal biometrics, the relative qualities of samples of the separate modes may be used to augment a fusion process [21, 22].
 - d. *Sample Replacement*
To negate the effects of template aging, a quality measurement may be used to determine whether

a newly acquired sample should replace the enrolled one. An alternative would be to retain both the old and new samples for use in a multi-instance fusion scheme.

e. *Template Update*

Again to address template aging, some systems instead combine old and new sample features. Quality could be used in this process.

Summary

Fingerprint quality measurement is an operationally important task. This paper enumerated ways in which it is useful to compute a quality value from a sample. In all cases the ultimate intention is to improve matching performance. The authors asserted therefore that quality algorithms should be developed to explicitly target matching error rates, and not human perceptions of sample quality. The term quality should not be equated to the acquisition settings of the sample, such as image resolution, dimensions in pixels, grayscale/color bit depth, or number of features. Though such factors may affect sample utility and could contribute to the overall quality score. We reviewed the existing practice of fingerprint local and global analysis. Local and global quality scores could be combined to form a vector of overall finger image quality. However, it is useful, even necessary for some applications, if local and global quality measures are summarized into a scalar which is predictive of error rates associated with the verification or identification of that sample.

Related Entries

- ▶ Biometric Sample Quality Standard
- ▶ Performance of Quality Measures

References

1. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: *Handbook of Fingerprint Recognition*. Springer, New York (2003)
2. Grother, P., et al.: MINEX: Performance and Interoperability of the INCITS 378 Fingerprint Template. National Institute of Standards and Technology, NISTIR 7296 edn. (2005). <http://fingerprint.nist.gov/minex04>
3. Tabassi, E., Wilson, C., Watson, C.: Fingerprint Image Quality, NFIQ. National Institute of Standards and Technology, NISTIR 7151 edn. (2004)
4. Tabassi, E., Wilson, C.L.: A novel approach to fingerprint image quality. In: ICIP (2), pp. 37–40 (2005)
5. Tilton, C., et al.: The BioAPI Specification. American National Standards Institute, Inc. (2002)
6. Benini, D., et al.: ISO/IEC 29794-1 Biometric Sample Quality Standard: Framework. JTC1 / SC37 / Working Group 3 (2008). <http://isotc.iso.org/isotcportal>
7. Alonso-Fernandez, F., Fierrez, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J., Frenthaler, H., Kollreider, K., Bigun, J.: A Comparative study of fingerprint image-quality estimation methods. *IEEE Trans. Inform. Forens. Secur.* **2**, 734–743 (2007)
8. Lim, E., Jiang, X., Yau, W.: Fingerprint image quality and validity analysis. In: IEEE proceedings of International Conference on Image Processing (ICIP), pp. 469–472. New York, USA (2002)
9. Chen, Y., Dass, S.C., Jain, A.K.: Fingerprint quality indices for predicting authentication performance. In: AVBPA, pp. 160–170 (2005)
10. Shen, L., Kot, A.C., Koo, W.M.: Quality measures of fingerprint images. In: AVBPA, pp. 266–271 (2001)
11. Ratha, N., Bolle, R.: Automatic Fingerprint Recognition Systems. Springer, New York (2004)
12. Lim, E., Toh, K.A., Saganthan, P.N., Jiang, X., Yau, W.Y.: Fingerprint image quality analysis. In: ICIP, pp. 1241–1244 (2004)
13. Chen, T.P., Jiang, X., Yau, W.Y.: Fingerprint image quality analysis. In: ICIP, pp. 1253–1256 (2004)
14. Bolle, R., et al.: System and methods for determining the quality of fingerprint images. US Patent 596356 (1999)
15. Hong, L., Wan, Y., Jain, A.K.: Fingerprint image enhancement: algorithm and performance evaluation. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(8), 777–789 (1998)
16. Nill, N., Bouzas, B.H.: Objective image quality measure derived from digital image power spectra. *Opt. Eng.* **31**(4), 813–825 (1992)
17. National Institute of Standards and Technology: NIST Biometric Image Software (NBIS) (2008). <http://www.itl.nist.gov/riad/894.03/nigos/nbis.html>
18. Ko, T., Krishnan, R.: Monitoring and reporting of fingerprint image quality and match accuracy for a large user application. In: Proceedings of the 33rd Applied Image Pattern Recognition Workshop, pp. 159–164. IEEE Computer Society (2004)
19. Tabassi, E., Grother, P.: Quality Summarization: Recommendations on Enterprise-wide Biometric Quality Summarization. National Institute of Standards and Technology, NISTIR 7244 edn. (2007)
20. Wein, L.M., Baveja, M.: Using fingerprint image quality to improve the identification performance of the u.s. visit program. In: Proceedings of the National Academy of sciences (2005). www.pnas.org/cgi/doi/10.1073/pnas.0407496102
21. Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J., Bigun, J.: Discriminative multimodal biometric authentication based on quality measures. *Pattern Recogn.* **38**(5), 777–779 (2005)
22. Tabassi, E., Quinn, G.W., Grother, P.: When to fuse two biometrics. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR-06. New York (2006). Biometric Workshop

Fingerprint Indexing

GEORGE BEBIS
University of Nevada, Reno, NV, USA

Synonyms

Continuous classification; Fingerprint retrieval; fingerprint authentication; fingerprint identification

Definition

When matching a query fingerprint to a large fingerprint database for identification purposes, a critical issue is how to narrow down the search space. Indexing provides a mechanism to quickly determine if a query fingerprint is in the database and to retrieve those fingerprints that are most similar with the query, without searching the whole database.

Introduction

Fingerprint matching is one of the most popular and reliable biometric techniques used in automatic personal identification. Typically, fingerprint matching is based on low-level features determined by singularities in the finger ridge pattern known as *minutiae*. To be practical, matching should be robust to translation, rotation, scale, shear, occlusion, and clutter. In this context, matching two fingerprints implies finding a subset of minutiae in the first fingerprint that best match to a subset of minutiae in the second fingerprint through a geometric transformation in an optimal sense.

There are two main applications involving fingerprint matching: *fingerprint authentication* and *fingerprint identification*. While the goal of fingerprint authentication is to verify the identity of a person, the goal of fingerprint identification is to establish the identity of a person. In this case, matching involves comparing a query fingerprint against a database of reference fingerprints to establish the identity of the query. An important issue in fingerprint identification is how to select the most similar fingerprint(s) to the query fingerprint from the fingerprint database. The easiest but least effective way to search a large database is to compare the query fingerprint with each

fingerprint in the database. Since usually there is no *a-priori* knowledge of possible correspondences between the query and the reference fingerprints, however, matching can be computationally too expensive, even for a moderate number of reference fingerprints.

A common approach to narrow down the search is by dividing the fingerprint database into smaller sets using *fingerprint classification*. The idea is to match the query fingerprint against fingerprints of the same type only. Although this approach can reduce the number of matches, it is not very effective since fingerprints are unevenly distributed (i.e., more than 90% of the fingerprints belong to only three classes [1]). Several *sub-classification* systems have been proposed to address this issue by further dividing some of the classes into more specific categories, however, these systems are much more complex and difficult to implement [1].

A more effective approach to narrow down the search space is to use *indexing*. In principle, indexing can quickly determine if a query fingerprint is in the database and to retrieve those reference fingerprints which are most similar to the query fingerprint, without searching the whole database. Therefore, methods based on indexing are less dependent on the size of the database. The main idea is to assign an index value to each fingerprint and match the query against those reference fingerprints having comparable indices only. Indexing methods have been very popular in computer vision for searching large databases of models in object recognition [2–4]. Therefore, many indexing schemes for finger identification have their roots in object recognition.

How Indexing Works

Indexing is a mechanism which, when provided with a key value, can rapidly access some associated data. Thus, instead of searching the space of all possible matches and explicitly rejecting invalid ones, indexing inverts the process so that only the most feasible matches are considered for matching. In essence, indexing serves as a “filtering” step which allows to verify a query fingerprint against the most similar fingerprints in the database only. To implement indexing, certain information about the reference fingerprints is prestored in an index structure. During identification, the index structure is accessed efficiently to narrow down the search.

Typically, a single index can be computed from the whole fingerprint or multiple indices can be computed from groups of local features. Using a single index, fingerprints are mapped to numerical vectors in a high-dimensional space through a similarity-preserving transformation. During identification, the query fingerprint is compared against those reference fingerprints which are close to the query in the multidimensional space. This approach, also known as *continuous classification* [5], is in essence a classification approach, however, the classes are not disjoint. Commonly, the orientation image is used in the mapping transformation, however, different transformations and distance measures have been proposed [5–8].

Using groups of local features, reference fingerprints are represented redundantly in the database by computing a separate index for each group of features and making an entry for each index [9–11]. This kind of redundancy provides robustness during identification by allowing the retrieval of reference fingerprints that match the query fingerprint only partially. Specifically, for each reference fingerprint, groups of features are extracted and an index is constructed from each group. The indexed locations are filled with entries containing information about the reference fingerprints. At a minimum, each entry contains information about the identity of the reference fingerprint and the group of features that generated the index.

During identification, the information stored in the index structure is used to quickly eliminate noncompatible matches between the query and the reference fingerprints. To reduce the number of false matches, geometric constraints can be used [11]. The reference fingerprints listed in the indexed locations are collected into a list of candidate fingerprints and the most often indexed fingerprints are selected for further verification. Verification works by computing the transformation between the candidate fingerprints and the query. Then, the candidate fingerprints are aligned with the query and their similarity to the query is estimated by finding the percentage of candidate features that have been aligned with query features.

An Example

Here is an example, based on [9, 10], to illustrate the use of indexing for fingerprint identification. In this

example, matching a pair of minutiae sets is performed by comparing minutiae triangles, formed by minutiae triplets, using geometric invariant features. In general, a pair of corresponding minutiae triangles provides enough information to compute a geometric transformation (e.g., similarity or affine) that potentially aligns the minutiae sets. To compute good alignments, voting can be applied in the transformation space to find transformations that are supported by many minutiae triangles [9]. A number of hypothetical transformations is obtained by considering transformations that have received a high number of votes. Each hypothetical transformation is then explicitly verified by counting the number of aligned minutiae.

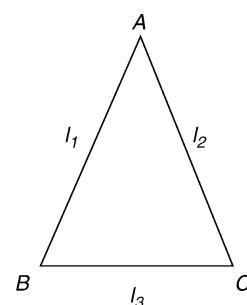
The indexing mechanism used in this example is based on geometric hashing [2]. Specifically, given a triplet of minutiae, three geometric invariants can be computed by considering the triangle formed by the minutiae triplet. The geometric invariants are based on the sides and angles of the minutiae triangle, as shown in Fig. 1, and remain unchanged under similarity transformations (i.e., translation, rotation, and scale). First, the sides of the triangle are sorted to avoid considering all possible orderings:

$$l_1 \leq l_2 \leq l_3$$

Then, we compute the following geometric invariants:

$$\begin{aligned} 0 &\leq \frac{l_1}{l_3} \leq 1 \\ 0 &\leq \frac{l_2}{l_3} \leq 1 \\ -1 &\leq \cos(A) \leq 1 \end{aligned}$$

where A is the angle between the smallest two sides. To compute an integer index, a simple hash function is applied on the geometric invariants which involves linear scaling followed by quantization. For each index,



Fingerprint Indexing. **Figure 1** A minutiae triangle defined by a minutiae triplet (A,B,C).

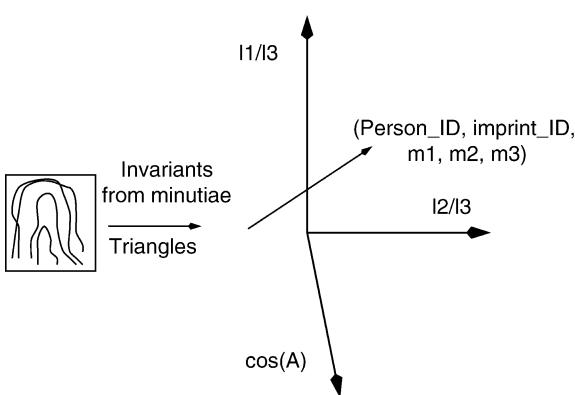
information is stored about the fingerprint and the minutiae triangle in a hash table. Each entry stored in the hash table has the following format:

$$(person_ID, print_ID, m_1, m_2, m_3)$$

where *person_ID* corresponds to the identity of the person whose fingerprint is considered, *print_ID* is an identification code for the particular fingerprint of that person, and m_i are the (x, y) coordinates of the m_i minutia in the triangle. [Figure 2](#) illustrates the indexing step.

During identification, each index generated by the query fingerprint is used to retrieve all reference fingerprints stored in the hash table under the same index. For each minutiae triangle, the lengths of the sides are computed, sorted in ascending order, and the geometric invariants are computed as before. Then, the invariants are scaled and quantized in the same manner. The resulting index is used to extract all entries from the database stored at the same index table location. To account for noise, entries stored in a small neighborhood around the indexed location could be also retrieved.

Several indexing-based approaches accumulate evidence about reference fingerprints by casting a vote for every entry stored in the indexed locations and by “histograming” the entries to pick the ones which have received a high number of votes. However, this approach takes into consideration only the number of votes received by a particular entry and not whether these votes are consistent among themselves. To introduce a measure of coherence, voting in the transformation space has been proposed [9]. The idea



Fingerprint Indexing. [Figure 2](#) Pre-storing information about the reference fingerprints using indexing.

is simply to consider transformations which form large clusters in the transformation space.

Each of the entries retrieved from the index table represents a hypothesized correspondence between minutiae triplets in the query and a reference fingerprint. Given this information, the transformation that best maps the query triplet to the reference triplet is computed. The computed transformation parameters are binned and, along with the *person_ID* and *imprint_ID*, form a key that indexes another data structure used for evidence accumulation. An eight-dimensional integer array is used to store the number of votes in the transformation space (i.e., six dimensions for the parameters of the transformation, one for the *person_ID* and one for the *imprint_ID*).

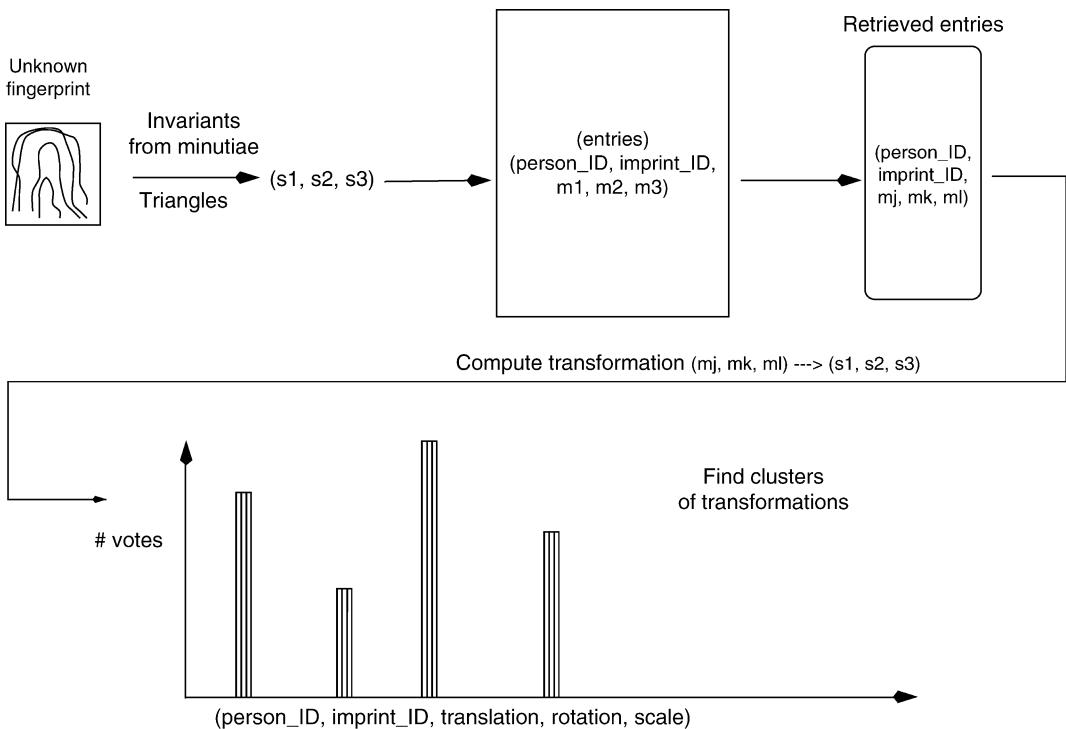
If a large number of minutiae points can be brought into correspondence by a transformation, then all the indices generated by the triangles formed by those minutiae points will yield close transformation parameters. Hence, a larger number of votes for a correct match will be accumulated. Although there might be a number of random correspondences between minutiae triplets in the query fingerprint and some arbitrary reference fingerprints, the likelihood of a number of consistent transformation parameters being generated by random correspondences is small, and the verification step will eliminate most of them. [Figure 3](#) illustrates these identification procedure based on indexing.

Practical Issues

Several important issues must be considered while employing indexing for fingerprint identification including: index construction, index selectivity, storage requirements, indexing mechanism, performance analysis, and error analysis. Each of them briefly discussed in the following section.

Index Construction

As illustrated in the earlier example, each index is typically constructed from groups of local features, such as minutiae triplets. In general, index construction should be based on features that are robust to fingerprint distortions, occlusion, and noise [11]. To reduce storage requirements, the computation of the index is based on geometric invariant features,



Fingerprint Indexing. **Figure 3** Illustration of the identification step using indexing.

that is, features that remain unchanged under certain geometric transformations. In the earlier example, we used length ratios and angles which are invariant to similarity transformations (i.e., translation, rotation, and scale). Other geometric invariant features include ridge count, triangle handedness, triangle type, triangle direction, and maximum side, minutiae density, and various ridge invariants [9, 11–13].

Index Selectivity

Although indexing is an attractive approach, very often it becomes less effective because of limited index selectivity. The issue of index selectivity relates to the discrimination power of the features considered for indexing. Features with low discrimination power give rise to very similar indices (i.e., low index selectivity). As a result, a large number of hypothetical matches can be generated during identification, making indexing ineffective. One way to deal with this problem is to increase the index dimensionality using larger groups of features, however, this would also increase memory requirements since the number

of groups increases exponentially with group size. Alternatively, additional information can be computed from each group and added to the index to increase its dimensionality. For example, the FLASH algorithm, introduced in [3] for object recognition and adopted in [9] for fingerprint identification, computes a nine-dimensional index from minutiae triangles. It should be mentioned that although this is an effective approach, it increases time requirements and raises the issue of computing the additional features fast and reliably. Recent studies using high-dimensional indices include [11] and [12].

Storage Requirements

Indexing methods have high storage requirements as they trade space for speed. For example, the number of entries to be indexed using minutiae triplets is of the order of $O(N^3)$ where N is the average number of fingerprint minutiae. If M is the number of fingerprints to be indexed, the total space requirements is of the order of $O(MN^3)$. To reduce storage requirements, geometric constraints can be used to limit the

number of minutiae triangles considered for indexing [9]. Alternatively, a unique topological structure can be associated with the fingerprint minutiae using the Delaunay triangulation [10, 13]. This approach considers only $O(N)$ minutiae triangles for indexing leading to significant memory savings and faster identification. A problem with this approach is that it is sensitive to noise and distortions (e.g., introduced by missing or spurious minutiae), however, both noise and distortion have only a local effect on the triangulation. Nevertheless, hierarchical matching schemes have been proposed to deal with these issues [14].

Indexing Mechanism

Hashing has been the most common indexing mechanism used both in fingerprint identification and object recognition. Hashing performs a range search, retrieving all points within a certain distance from the query point. However, the highest probability hypotheses can be discovered by observing just a few of the closest neighbors. Hashing is not efficient for nearest-neighbor search in high dimensions since it requires time exponential in the dimension of the space (i.e., the nearest neighbors might not lie in the same hash bin as the query point, but in one of the many adjacent bins). Moreover, “good” hash functions are required for distributing the data uniformly [15, 16]. In general, more effective indexing mechanisms can be employed, such as kd-trees [17], to retrieve only the k nearest points.

Kd-trees are data structures used to divide the data into hypercubes containing equal numbers of data. When a query point is presented, the boundaries between the hypercubes are used as decisions to discover the hypercube that contains the query point, and the data in this hypercube will be close matches. To guarantee that the matches in the hypercube containing the query point are in fact closer to the query point than data lying just over the boundary of the hypercube, it is necessary to examine neighboring hypercubes. This can make search quite slow. To deal with this issue, approximate nearest-neighbor schemes can be used which maintain good performance even in quite high dimensions (i.e., 10–20) and large number of data [18, 19]. These algorithms have been demonstrated to uncover the exact nearest neighbor a high percentage of the time and a very close neighbor in the remaining cases.

Performance Analysis

To analyze the performance of indexing schemes, it is typical to use identification rate versus ▶ penetration rate graphs. The ratio of fingerprints retrieved over the size of the database. These graphs show the identification rate achieved by varying the penetration rate. Typically, a low penetration rate with a high identification rate is desirable. Close to 99% identification accuracy with only 5% penetration rate is reported in [12] on DB1 from FVC2002. Alternative measures include the ▶ Correct Index Power (CIP) and the ▶ Correct Reject Power (CRP) [11]. CIP is defined as the number of correctly retrieved fingerprints over the size of the database while CRP is defined as the ratio of correctly rejected reference fingerprints over the number of query images not having a corresponding fingerprint in the database. Using the NIST-4 special database and extrapolating the results from 2,000 images to 30,000 images, Bhanu et al. [11] report a CIP rate of 50% using the top 100 candidate matches (i.e., 0.33% penetration rate). Using a smaller database (i.e., 400 image pairs) and assuming the top candidate match, they report a CIP rate of 96.2% for good quality images, 85.5% for fair quality images, and 83.3% for low quality images. Using the top five candidate matches, the CIP rate increases to 100, 99.2 and 98% correspondingly. The CRP rate reported using 200 query fingerprints not in the database was 100%.

Error Analysis

In the noiseless case, each indexed location will contain exactly the set of reference groups compatible with the query group used to access the index structure. In practice, however, several different sources of error must be taken into consideration to improve robustness. The most common source of errors is from the feature extraction step. Using minutiae triplets, for example, errors in the localization of the minutiae can lead to errors in the computation of the geometric invariants and, as a result, to errors in the computation of the indices. In this case, the correct entries will not be found in the indexed location but in a neighborhood around it. Several studies have considered the effect of localization errors on indexing performance

for object recognition [20]. Other studies model localization errors probabilistically in order to estimate the appropriate neighborhood size to retrieve the correct entries [15].

Summary

Indexing is an attractive method for reducing the number of matches when comparing a query fingerprint with a fingerprint database for identification purposes. This chapter reviewed the main concepts behind fingerprint indexing and discussed several critical issues to be addressed in practice.

Related Entries

- ▶ [Fingerprint Authentication](#)
- ▶ [Fingerprint Classification](#)
- ▶ [Fingerprint Identification](#)
- ▶ [Fingerprint Matching](#)

References

1. Maltoni, D., Maio, D., Jain, A., Prabhakar, S.: *Handbook on fingerprint recognition*. Springer, Berlin (2003)
2. Lamdan, Y., Schwartz, J., Wolfson, H.: Affine invariant model-based object recognition. *IEEE Trans. Robot. Automat.* **6**(5), 578–589 (1990)
3. Califano, A., Mohan, R.: Multidimensional indexing for recognizing visual shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(4), 373–392 (1994)
4. Bebis, G., Georgopoulos, M., Shah, M., da Vitoria Lobo, N.: Indexing based on algebraic functions of views. *Comput. Vision Image Understand.* **72**, 360–378 (1998)
5. Lumini, A., Maio, D., Maltoni, D.: Continuous versus exclusive classification for fingerprint retrieval. *Pattern Recogn. Lett.* **18**(10), 1027–1034 (1997)
6. Cappelli, R., Lumini, A., Maio, D., Maltoni, D.: Fingerprint classification by directional image partitioning. *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(5), 402–421 (1999)
7. Cappelli, R., Maio, D., Maltoni, D.: Multispace kl for pattern representation and classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(9), 977–996 (2001)
8. Li, J., Yau, W.Y., Wang, H.: Fingerprint indexing based on symmetrical measurement. *Int. Conf. Pattern Recogn.* **1**, 1038–1041 (2006)
9. Germain, R., Califano, A., Colville, S.: Fingerprint matching using transformation parameter clustering. *IEEE Computational Science and Engineering* **4**(4), 42–49 (1997)
10. Bebis, G., Deaconu, T., Georgopoulos, M.: Fingerprint identification using delaunay triangulation. In: *IEEE, International Symposium on Information, Intelligence, and Systems*, pp. 452–459 (1999)
11. Bhanu, B., Tan, X.: Fingerprint indexing based on novel features of minutiae triplets. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(5), 616–622 (2003)
12. Feng, J., Cai, A.: Fingerprint indexing using ridge invariants. *Int. Conf. Pattern Recogn.* **4**, 433–436 (2006)
13. Ross, A., Mukherjee, R.: Augmenting ridge curves with minutiae triplets for fingerprint indexing. In: Prabhakar S., Ross A. (eds.) *SPIE Defense and Security Symposium (Biometric Technology for Human Identification IV*, vol. 6539) (2006)
14. Uz, T., Bebis, G., Erol, A., Prabhakar, S.: Minutiae-based template synthesis and matching using hierarchical delaunay triangulation. In: *IEEE International Conference on Biometrics: Theory, Applications and Systems* (2007)
15. Wolfson, H., Rigoutsos, I.: Geometric hashing: An overview. *IEEE Comput. Sci. Eng.* **4**(4), 10–21 (1997)
16. Bebis, G., Georgopoulos, M., La Vitoria Lobo, N.: Using self-organizing maps to learn geometric hashing functions for model-based object recognition. *IEEE Trans. Neural Networks* **9**(3), 560–570 (1998)
17. Li, W., Bebis, G., Bourbakis, N.: Integrating algebraic functions of views with indexing and learning for 3D object recognition. *IEEE Workshop on Learning in Computer Vision and Pattern Recognition* (2004)
18. Nene, S., Nayar, S.: Closest point search in high dimensions. In: *Computer Vision and Pattern Recognition Conference*, pp. 859–865 (1998)
19. Beis, J., Lowe, D.: Shape indexing using approximate nearest-neighbor search in high-dimensional spaces. In: *Computer Vision and Pattern Recognition Conference*. pp. 1000–1006 (1997)
20. Grimson, W., Huttenlocher, D., Jacobs, D.: A study of affine matching with bounded sensor error. *Int. J. Comput. Vision* **13**(1), 7–32 (1994)

Fingerprint Individuality

Fingerprint Individuality is the study of the extent of which different fingerprints tend to match with each other. It is the most important measure to be ascertained when fingerprint evidence is presented in court as it reflects the uncertainty with the decision of the expert.

- ▶ [Individuality of Fingerprints](#)

Fingerprint Matching, Automatic

JIE TIAN, YANGYANG ZHANG, KAI CAO

Center for Biometrics and Security Research & The Key Laboratory of Complex System and Intelligence Science Chinese Academy of Sciences, Institute of Automation Zhongguancun Donglu, Beijing, China

Synonyms

Fingerprint comparing; Automatic

Definition

In contrast to manual fingerprint matching, automatic fingerprint matching can be efficiently operated on a computing machine following a series of preset procedures. Automatic matching compares two given fingerprint templates (raw images or extracted features) and returns their similarity score (in a continuous range) or a binary decision (matched/non-matched).

Introduction

With the increasing expansion of large-scale databases, manual fingerprint matching cannot satisfy the demand of efficiency in many applications. Automatic fingerprint matching simulates how human experts compare the fingerprints to measure the similarity between two given fingerprint templates or to determine whether they come from the same finger [1]. For most fingerprint matching procedures, experts calculate the similarity score of two templates and give the final judgment with a preset threshold. If the score exceeds the threshold, the compared templates are considered matched, otherwise they are non-matched. The templates are the representation of fingerprints, comprising extracted features or the raw images in case of no extraction. The features can be categorized into two kinds: local features (minutiae, pores) and global features (compressed raw fingerprint, ridge pattern, orientation and curvature map).

Fingerprint matching is one of the most important stages in ► **Automatic Fingerprint Identification System (AFIS)**. It is really difficult to match the different

impressions of the same finger and find the corresponding features reliably because of the following interferential factors. First, there are several kinds of transformation between two impressions, including linear transformation (translation, rotation, and scale) and non-linear distortion. The translation and rotation is caused by the differential finger placement with respect to the sensor surface during different acquisitions, which may result in a partially overlapped area. If the impressions are captured by different sensors with different resolutions, there exists scale variation in the transformation space. The non-linear distortion of fingerprints is inevitable because the capture is a process of mapping a three-dimensional finger to a two-dimensional impression. The pattern of distortion is firstly determined by finger pressure, finger condition, and the characteristics of sensors. Secondly, the quality of raw fingerprints are also influenced by the noise (fingerprint residues from the previous capture), skin condition (dryness, grease, skin disease), and the capture environment (humidity, temperature). Figure 1 displays three examples of these interferential factors in fingerprint matching. In addition, the algorithms of fingerprint enhancement and feature extraction are imperfect and often introduce some mistakes into the extracted features. Errors may be made and accumulated during each of the foregoing stages (orientation estimation, singular points detection and minutiae extraction). These objective factors are likely to generate spurious features or miss genuine features. All the above variations may make the templates from the same finger appear quite different, sometimes more severely than the similar templates from different fingers. Many fingerprint matching algorithms have been proposed in the scientific literature. Most of these algorithms are proved successful when dealing with good-quality fingerprints. However, fingerprint matching is still a challenging task due to the difficulty in matching low-quality, partial, or large-distorted fingerprints.

There have been a series of strategies to cope with the transformation between two fingerprints. In most of typical fingerprint matching processes, alignment is utilized to estimate the optimum linear transformation between two fingerprints. It rotates and translates one of the compared templates in order to make its features mostly overlap the corresponding features in another template. To achieve the optimum feature-pairing requires correctly calculating the parameters of translation and rotation. Note that scale has to be taken



Fingerprint Matching, Automatic. **Figure 1** Three examples of these interferential factors in fingerprint matching. **(a)** a pair of fingerprints with large translation and rotation; **(b)** a pair of poor-quality and partially overlapped fingerprints; **(c)** a pair of large-distorted fingerprints [22]. While the corresponding minutiae in blue rectangle are overlapped, the maximal distance of corresponding minutiae in red ellipse is above 100 pixels.

into account when the resolutions of fingerprints vary. Previous researches [2] prove that the performance of the matcher drastically decreased when the compared fingerprints originated from sensors with

different resolutions. Fingerprint alignment is certainly an important but time-consuming stage. Therefore, some algorithms [3] attempt to avoid this stage in fingerprint matching. For instance, experts construct

local feature structures invariant to the linear transformation for matching without prior global alignment. Such matching algorithms ignore the global relationship among local features and therefore may lose part of the discriminating information. On the other hand, non-linear distortion is universal during fingerprint acquisition, so it is needed to develop fingerprint matchers tolerant of the distortion. Some methods [4, 5] allow corresponding features to alter in the predetermined range (tolerance box). Others [6, 7] adopt local feature structures for matching because distortion affects to a lesser degree local areas. Few developers [8, 9] introduce an appropriate model to recover the distortion prior to matching. In general, tolerating more transformations may increase the successful percentage of not only ► **genuine matching** but ► **imposter matching**. When designing the matching algorithms, the degree of tolerance needs careful evaluation. Based on the calculated transformations, the correspondences between features can be established through the optimization methods.

Classification

Because fingerprint matching algorithms rely heavily on the stored features in the templates, they can be coarsely classified into three categories in terms of the selection of features:

- Local feature-based matching: The most popular local feature is minutia, which was earliest used in fingerprint matching technologies [5]. Minutiae features are extracted and stored in the templates as sets of points in the two-dimensional plane. They are usually described by the location, orientation, type, and other information in the neighborhood region. Most common minutiae matching are addressed as a point pattern matching problems and many approaches can be applied. Furthermore, several adjacent minutiae are constructed as local structures in various forms of minutiae, such as simplex [10] triangle [6] and so on.
- With the advent of high-resolution fingerprint sensors, more precise local features (pores and ridge contours [11]) are employed in fingerprint matching to satisfy the growing demand and requirements for accuracy. These algorithms usually align two different templates to establish the

correspondences between two sets of local features and calculate the similarity score combining all the matched features. Compared to other fingerprint features, local features have several advantages in terms of the template size and its discriminability, but they have inevitable drawbacks in practical usage. Sometimes it is difficult to exactly obtain local features due to its sensitivity to the fingerprint quality and capture area, which seriously degrades the performance.

- Global feature-based matching: The global features represent the fingerprint in a global perspective, many of which are more continuous and smooth everywhere except in some special regions. For poor-quality or partial fingerprints, global features can be extracted more reliably. It is too space-and time-consuming to directly store and compare the map/field of features pixel by pixel. To reduce the template size and simplify the matching, features can be approached with appropriate models and stored as a series of parameters. Global feature-based matching [12] overlaps two given templates with different transformation parameters and estimates the similarity score between the corresponding cells. Compared to local features, the global features have less distinctness, so they are often exploited together with other features or in the preprocessing stage of fingerprint matching.
- Combined feature-based matching: Since the local and global features are somewhat independent and capture contemporary information, it is reasonable to improve the discriminating ability of matching by fusing features. The approaches in this category [13, 14] combine the local and global features in the matching stage with available feature-level fusion strategies. The combination can reinforce the individuality of fingerprints and improve the performance for fingerprint systems on large-scale databases.
- How to select features is pivotal for the effect of feature combination. It is proved that combining the irrelative features will bring the most obvious improvement of accuracy or efficiency. On the other hand, fusing local and global features may result in additional time or memory cost, so the appropriate hierarchical strategy can be utilized to reduce resource consumption. For instance, due to the complexity of alignment, two fingerprints can

be pre-aligned by the modeled orientation field (global feature). Then the similarity score is calculated based on the minutiae (local feature). Pre-alignment is more efficient while matching on the large-scale database.

Performance Evaluation

Performance evaluation is necessary for understanding the limitations and advantages of a fingerprint matching algorithm and addressing its appropriate applications. The performance can be evaluated from different aspects: accuracy, resource consumption, scalability and sensor interoperability.

- The accuracy of matching is evaluated based on the distribution of similarity score in genuine and impostor matching. Genuine matching compares two fingerprint templates from the same finger, whereas impostor matching is for two fingerprint templates from different fingers. The overall accuracy can be illustrated by Receiver Operation Characteristics (ROC) curve, which shows the dependence of False Non-match Rate (FNMR) on False Match Rate (FMR) at all thresholds. A series of indicators are adopted to quantify the accuracy containing Equal Error Rate (EER – the point where FNMR and FMR yield the same value), FMR100 (the lowest FNMR for FMR $\leq 1\%$), FMR1000 (the lowest FNMR for FMR $\leq 0.1\%$) ZEROFMR (the lowest FNMR for FMR $\leq 0\%$), and ZEROFNMR (the lowest FMR for FNMR $\leq 0\%$) [15, 16]. Accuracy usually attracts most of the attention in common applications, but the algorithms cannot just be characterized by these indicators.
- Resource consumption can be measured through three aspects: the amount of storage, time, and memory required by the algorithms. The storage cost is measured by the average/maximum size of template for each database. The efficiency is indicated by the average/maximum time in genuine/impostor matching and the memory requirement is measured by the average/maximum size of allocated memory in genuine/impostor matching. The variation of the indicators through the whole database reflects the stability of the tested algorithms.
- The scalability reflects the degradation of the accuracy with the growing scale of database, which is

available in one-to-many matching. It should be evaluated on different-scale databases through observing the relationship between the aforesaid indicators and the scale of database. The international competition FpVTE2003 [17] adopted three different-scale databases to evaluate the scalability of the tested algorithm.

- Sensor interoperability denotes the ability to handle the templates obtained from different sensors [18]. The features in templates are sensitive to different characteristics of multiple sensors in a fingerprint system. The comparison of measures (accuracy, efficiency) between *intra-sensor* matching (comparing templates from the same sensor) and *inter-sensor* matching (comparing templates from different sensors) somewhat reflects the interoperability of the matching approach. Research on sensor interoperability is at its fledgling stage and so far there have been no authorized databases or indicators for quantified evaluation.

Performance during evaluation is relative to many objective conditions. Accuracy is influenced by both the characteristics of database (size, average quality, distortion) and the testing ► **protocol**, while resource consumption relies on the hardware capability, so it is meaningless to evaluate the matching approach without considering these conditions. An authentic evaluation should be conducted on the databases that have independent training/testing parts and sufficient fingerprints, and calculate the statistical indicators with a reasonable protocol. Because the above indicators are statistical results, it should be reported how believable the evaluation of these statistics really are. The problem can be addressed by computing the confidence intervals on the distribution of these values [19]. The accuracy of these confidence interval estimates is ascertained by both correct estimation strategies and correct dataset sampling.

The comparison of performance among various matching algorithms is always a controversy. Different algorithms have different advantages and disadvantages; therefore it is unfair to directly conclude one better than the other. Some research displays experiment results conducted on the proprietary databases using different protocols. This makes it difficult to compare the performance fairly. Evaluating and comparing these indicators among different algorithms is required to operate on the

same public databases with the same authoritative protocol and testing environment, such as FVC and FpVTE.

Application

In AFIS, the fingerprint matching can be applied in two distinct models: verification and identification. The verification model is a one-to-one matching (1:1) in which a user states his/her identity by means of an ID and proves it with a fingerprint. A new fingerprint sample taken from the user is compared with the user's previously registered or stored fingerprints. The comparison only occurs once between the input fingerprint image and the selected sample from the database following the claim of the user. If the fingerprints are successfully matched, the user is verified as who he/she is claiming to be, and granted all the privileges and access of the stated user. On the contrary, the identification process is a one-to-many matching (1:N), in which a user need not claim his or her identity. A new impression is taken from the user and compared to the existing fingerprints of registered or stored users in the databases. The identification can be implemented with a sequence of verification between the input template and the query templates in the database. Fingerprint identification requires searching the database for a matched template or several candidates, which is a process more complex than verification. Although satisfactory performances have been reported for fingerprint verification, both the efficiency and accuracy of identification deteriorate seriously by simple extension of a 1:1 verification procedure to a 1:N identification system. It is still necessary to improve the performance of fingerprint matching in the large-scale fingerprint database. Fingerprint classification and indexing techniques are proved effective to narrow down the searching space of verification, which will speed up the identifying process.

Different kinds of applications focus on different requirements. For the same algorithm, the matching threshold can be modulated or other parameters configured to realize trade-off among these performance indicators. There exists a strict relationship between accuracy and resource consumption, FMR and FNMR of each algorithm. For instance, both FMR and FNMR are actually the functions of matching threshold. The decrease in value makes the

algorithm more tolerant to the transformations (lower FNMR), but increases the possibility of incorrectly matching two templates from different fingers (higher FMR). Contrarily, if the value increases, the algorithm performs with higher FNMR and lower FMR. According to the given application, the threshold is carefully chosen as suitable for the special requirement. It is difficult to develop a matching approach omnipotent in every scenario, therefore different applications may at times need different algorithms. The embedded applications (mobile phone, identity card) emphasize limited resources and put significant strain on the recognition reliability, because high performance fingerprint matching approaches tend to be computationally intensive. In this case, we tend to adopt these algorithms with lower resources consumption. In contrast, the resource-unlimited applications equipped with adequate resources attach more importance to accuracy rather than the computation and storage expenditure. For instance, fingerprint matching in network security operates on the distributed computer system with a "Trustworthy Authority + Remote Client" mode, where extreme accuracy is the most crucial target. In these situations, we choose the algorithms that have more accuracy despite of the possible computational complexity.

Summary

Recently, there have been great advances in the research on automatic fingerprint matching. However, the various applications of AFIS in personal identification desire further improvement of the performance of matching algorithms. Recent research demonstrates that fusion in different levels (feature, score, decision) is effective in improving the performance in many aspects, attracting increasing interest. Besides the features, the fusion of multiple independent matchers [20, 21] is likely to ameliorate the accuracy of fingerprint matching.

Related Entries

- ▶ [Fingerprint Classification](#)
- ▶ [Fingerprint Clustering](#)
- ▶ [Fingerprint Matching, Manual](#)
- ▶ [Fingerprint Recognition](#)
- ▶ [Fingerprint Templates](#)

References

1. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of Fingerprint Recognition. Springer, Berlin (2003)
2. Ross, A., Jain, A.K.: Biometric sensor interoperability: a case study in fingerprints, BioAW 2004. LNCS **3087**, 134–145 (2004)
3. Bazen, A.M., Gerez, S.H.: An intrinsic coordinate system for fingerprint matching. In: Third International Conference on Audio- and Video-Based Biometric Person Authentication, Halmstad, Sweden (2001)
4. Luo, X.P., Tian, J., Wu, Y.: A Minutia matching algorithm in fingerprint verification. Fifteenth ICPR **4**, 833–836 (2000)
5. Jain, A.K., Lin, H., Bolle, R.: On-line fingerprint verification. IEEE Trans. Pattern Recogn. Machine Intell. **19**(4), 302–314 (1997)
6. Kovacs-Vajna, Z.M.: A fingerprint verification system based on triangular matching and dynamic time warping. IEEE Trans. Pattern Recogn. Machine Intell. **22**(11), 1266–1276 (2000)
7. Chen, X.J., Tian, J., Yang, X.: A new algorithm for distorted fingerprints matching based on normalized fuzzy similarity measure. IEEE Trans. Image Process. **15**(3), 767–776 (2006)
8. Bazen, A.M., Gerez, S.H.: Fingerprint matching by thin-plate spline modelling of elastic deformations. Pattern Recogn. **36**(8), 1859–1867 (2003)
9. Ross, A., Dass, S., Jain, A.K.: A deformable model for fingerprint matching. Pattern Recogn. **38**(1), 95–103 (2005)
10. He, Y.L., Tian, J., Li, L., Yang, X.: Fingerprint matching based on global comprehensive similarity. IEEE Trans. Pattern Analy. Machine Intell. **28**(6), 850–862 (2006)
11. Jain, A.K., Chen, Y., Demirkus, M.: Pores and ridges: high-resolution fingerprint matching using level 3 features. IEEE Trans. Pattern Recogn. Machine Intell. **29**(1), 15–27 (2007)
12. Jain, A.K., Prabhakar, S., Lin, H., Pankanti, S.: Filterbank-based fingerprint matching. IEEE Trans. Image Process. **9**(5), 846–859 (2000)
13. Gu, J., Zhou, J., Yang, C.: Fingerprint recognition by combining global structure and local cues. IEEE Trans. Image Process **15**(7), 1952–1964 (2006)
14. Wang, X.C., Li, J.W., Niu, Y.M.: Fingerprint matching using OrientationCodes and PolyLines. Pattern Recogn. **40**(11), 3164–3177 (2007)
15. <http://bias.csr.unibo.it/fvc2002/perfeval.asp>
16. Cappelli, R., Maio, D., Maltoni, D., Wayman, J.L., Jain, A.K.: Performance evaluation of fingerprint verification systems. IEEE Trans. Pattern Recogn. Machine Intell. **28**(1), 3–18 (2006)
17. <http://fpvte.nist.gov>
18. Ross, A., Jain, A.K.: Biometric sensor interoperability: a case study in fingerprints. Proc. Biometric Authentication: ECCV 2004 International Workshop **3087**, 134–145 (2004)
19. Bolle, R.M., Ratha, N.K., Pankanti, S.: An evaluation of error confidence interval estimation methods. Fifteenth ICPR **3**, 103–106 (2004)
20. Ross, A., Jain, A.K., Reisman, J.: A hybrid fingerprint matcher. Pattern Recogn. **36**(7), 1661–1673 (2003)
21. Bhakar, S., Jain, A.K.: Decision-level fusion in fingerprint verification. Pattern Recogn. **35**(4), 861–874 (2002)
22. Chen, X.J., Tian, J., Yang, X.: An algorithm for distorted fingerprint matching based on local triangle features set. IEEE Trans. Inform. Forensics Security **1**(2), 169–177 (2006)

Fingerprint Matching, Manual

HERMAN BERGMAN¹, ARIE ZEELENBERG²

¹Certified Fingerprint Expert, San Francisco, USA

²Senior Advisor Fingerprints, National Police Force, The Netherlands

Synonyms

Identification; Individualization; Minutial

Definition

Identification has been defined as the determination by a fingerprint examiner that two examined images of friction ridge skin are deposited by the same source (finger, palm or foot), with the goal of determining the identity of a donor. If this can be established it is generally accepted within the discipline that given the uniqueness or ► individuality of friction ridge skin, this fingerprint can be attributed to this donor at the same time excluding all others. (In this contribution an expert for practical reasons is referred to as “he”. Female experts should not feel excluded but may comfort themselves with the idea that with respect to erroneous identifications also the male form is used)

Fingerprint Matching: Manual

The matching process described here applies to marks or latent prints found at a crime scene or on pieces of evidence associated with a crime. Those marks tend to be incomplete and of lesser quality than ► comparison prints. The process where known prints are compared, one to one or one to many, to verify an identity has become an increasingly automated process. Because of the amount of quality and quantity of data available and the accuracy of current Automated Fingerprint Identification Systems (AFIS) this process can be applied in a “lights out” mode or monitored by examiners.

This automated process to determine individuality is generally referred to as “matching” and is executed by matching algorithms. For the process where latent prints or marks are analyzed and compared by an examiner the more generic term identification or individualization is used rather than matching.

The identification process is a one to one comparison and starts after a similar print is found which cannot be excluded as being the same at face value. Three possible scenarios can lead to this:

A candidate can be the result of an AFIS search in which the similarity of the extracted features is calculated against known exemplars in a digital repository.

If one of the best resembling candidates cannot be excluded it might be eligible for input in an identification process.

Second, a candidate can be selected after manual comparison of one or more named suspects.

Third, a candidate may be found through a manual search of a physical fingerprint repository. This last occasion becomes increasingly rare because physical fingerprint repositories and manual searching become distinct by the broad use of AFIS.

The process by which the expert examines possible candidates focuses more on elimination based on differences than on weighing of similarities. At this stage the examiner searches for differences in the overall pattern formed by the ridges which is considered the first of three levels of information that are generally distinguished [1]. They are addressed to as ► **the first**, ► **second** and ► **third level** detail.

When an expert manually compares a mark against known, or comparison prints he visually assesses the main aspects of the ridge flow and/or a discernible pattern and a chosen target group of ► **minutiae** which he can relate to a recognizable area or location in the mark such as a delta, core or along the type lines.

This information is used to eliminate compared prints, this exclusion may be a very fast process. At one glance an expert may see that a compared donor shows 10 whorl patterns in the fingertips while he is looking for a loop. Even so a donor with a number of loops to the right with high ridge counts between the delta and core can be excluded definitively if the mark has a low ridge count. If no exclusion on ridge flow is possible because it is similar the remaining print will be compared keeping the target group in mind and looking for differences in the known print at the given positions relative to known locations. If he initially finds small clusters in a similar sequence he will then expand the assessed area both in the mark and the known exemplar.

If the print does not originate from the same source he will quickly find discrepancies and the comparison print will be excluded. If exclusion fails, the candidate will be included in the identification process.

The identification process

The generally accepted methodology for the identification process of friction ridge impressions is known as ACE-V [1] or a variation of this [2]. ACE-V is the acronym for Analysis, Comparison and Evaluation followed by Verification by another expert. ACE-V was first introduced by R.A. Huber [3] and later by D. Ashbaugh [1] for the examination of friction ridge skin. This methodology is generally accepted in forensics as a universal protocol to promote reproducibility and objectivity and should allow for the validation of the stated conclusions by reference to the process through which they are constructed.

It has been argued that ACE-V may not fully provide the requirements [4] necessary for an identification technique which should be explicit and defined in more detail [5]. Professors van Koppen and Crombag [6, 7] proposed the use of a descriptive model and a decision making model in forensic identification of ear-, lip- and fingerprints.

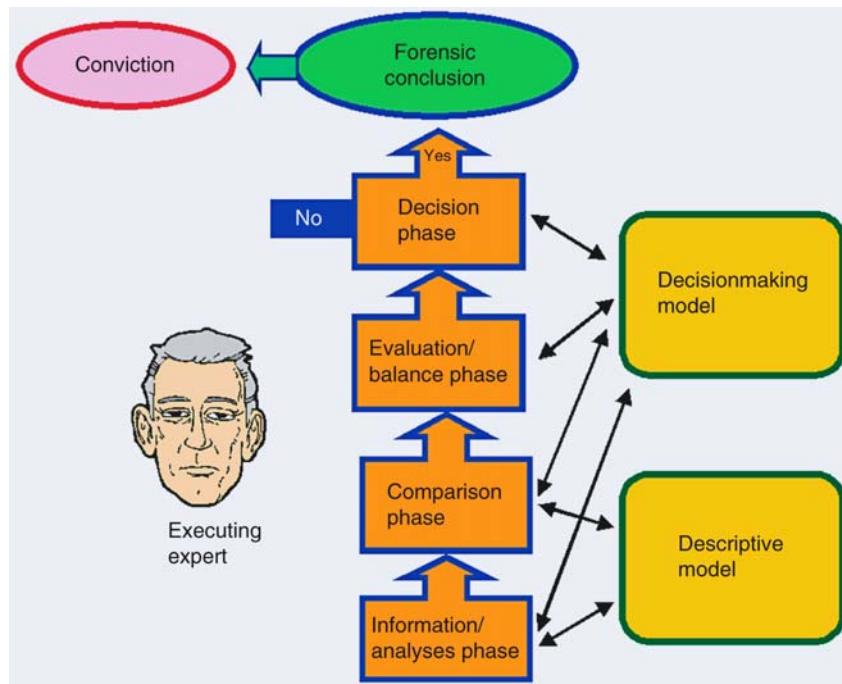
The Interpol European Expert Group on Fingerprint Identification (IEEGFI) report [2] not only describes a method similar to ACE-V (Fig. 1), but also provides both a descriptive model and a decision making model [6].

These models present a common terminology, grounds for establishing the value of features, rules of thumb, describe the pitfalls and provide good guidance for decision-making with respect to details and the overall decision of identification. It is essential to reproduce the whole process rather than to confine reproducibility to the conclusion.

The IEEGFI uses the word information phase as a synonym for the analyses phase and addresses the evaluation as the balance phase.

Analyses

A thorough and objective analysis of the latent print is the basis of a sound process, an unbiased establishment of the quantity and quality of available data is the aim. The analysis is the establishment of features and their properties and values recorded in a combination of mental and explicit written notes of all observed data. A copy of the image of the latent can be marked up in order to document observations. All three levels of information that are regarded as properties of friction ridge skin are assessed to determine their reliability



Fingerprint Matching, Manual. Figure 1 Diagram ACE.

and value, taking into account the influence of development technique(s) used, the exhibit, distortion, surface, deposition pressure, matrix, and anatomical aspects. Ambiguous Galton features of which the exact location cannot be seen at face value can still be established by ► tracing. In these instances the ridge detail and the exact appearance of the detected feature are unknown and may add little weight to the value of the latent and, subsequently, to the comparison. Nevertheless, it can be helpful to check whether certain Galton features in the comparison print are at least not in conflict with the latent.

Although a good practise in all cases, it is acknowledged that not all latent prints require such an in-depth analysis. In instances of high quality latent prints with unambiguous and/or an abundance of data, the analysis can be very quick.

However, it should be stressed that with low quality and quantity latent prints a full in-depth analysis is essential. The importance and depth of the analysis is inversely proportional to the quality of the latent print.

The IEEGFI II proposed a special procedure, “► The need for a questionable ID procedure” for complex examinations [6]. The examiner has to form an opinion about the quality, quantity, and reliability of the observed data in the latent print and on the

basis of this he has to decide whether the latent print has sufficient potential to relate it to its unique source. If that is the case he moves on to the comparison phase.

Comparison

The latent and the comparison prints are placed side by side enabling accurate comparison and the preservation of observations.

The data obtained in the analysis phase form the basis and guide for the comparison process and should be leading. During comparison not only data in the latent are checked against the comparison print, but also data found in the comparison print are cross-checked whether or not they are present in the latent.

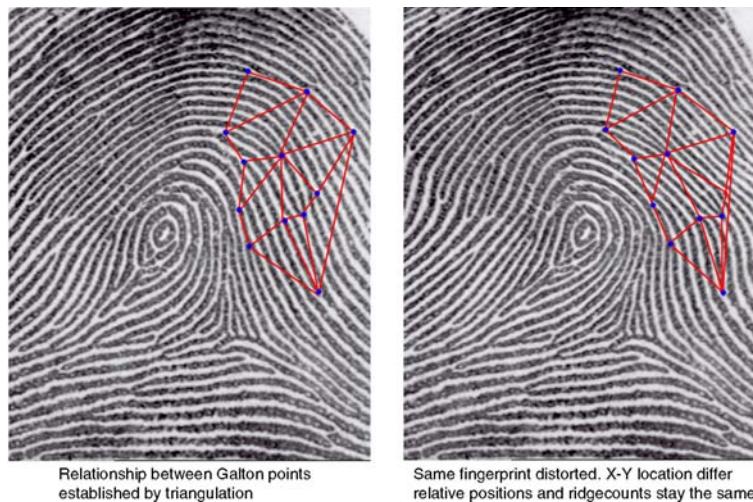
The relations of all features within the configuration are checked through triangulation [8]. This is done by following the ridges or furrows and counting the number of intervening ridges between features along a virtual connecting line. The relative location aspects and relations to other features in the latent have to be within tolerance compared to the features in the corresponding locations of the comparison print. (The direction in which the neighboring feature is found is

checked towards the general ridge flow and relative to the connecting line with other minutiae in the same area.) Due to the flexibility of the skin the interrelationship of features can be disturbed, but as in the case of a stretched spider web the relative positions remain the same (Fig. 2).

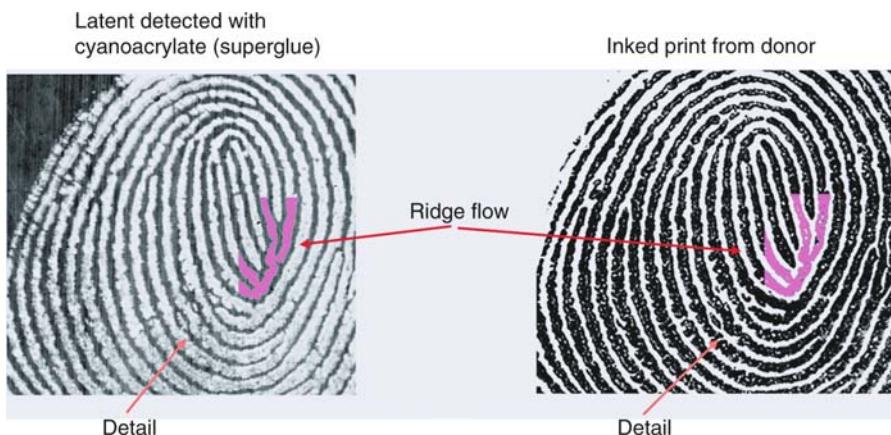
At this stage the expert also looks for similar third level detail which he relates to the location of second level detail.

It can be a very powerful contribution to each individual minutia and to the whole of a print but its accurate representation is dependent on a large number of variables such as pressure, moisture, the surface,

and the detection technique. Reliable third level detail in latent prints is a gift rather than a given fact. It is often difficult to draw a distinction between third level detail and anomalies. Matching third level detail is not very common and often calls for rationalization. Instead, the relationship of the minute events amongst them and with minutiae is more often studied. A large pore on the edge of a line followed by a small one in the centre of a ridge, the flow of an individual ridge like a recognizable river bed, a small dot lying in front of a tapered ridge ending are examples of ridge detail that, if similar, can be very significant contributions to the weight of the comparison (Fig. 3).



Fingerprint Matching, Manual. Figure 2 Triangulation/Distortion.



Fingerprint Matching, Manual. Figure 3 Third level detail of exceptional quality, organic shapes that are found in a similar fashion on exact corresponding locations.

Overall similarities should be apparent and demonstrable and be primarily based upon findings obtained in the analysis phase. Thus, avoiding the implementation of features found in the supposed original into the latent. When marking corresponding features it is important to establish the existence, the relations and their significance. For each individual point of similarity the quality may differ. If a point is clear and shows corresponding ridge detail its value is significantly higher than points that do not have these properties.

Dissimilarities and/or discrepancies should be detected, assessed, noted and accounted for. Any explanation of dissimilarities should preferably reflect the observations made during the analysis phase. An opinion has to be formed whether the differences in appearance are considered distortions or discrepancies. In the case of discrepancies, the conclusion should be an exclusion or/and inconclusive.

There is a distinct difference between the comparison of minutiae and ridge detail. Minutiae must be the same and ridge detail can be the same. Whereas the basic properties of the Galton points are firmly established during the analysis phase true third level detail is often only acknowledged and confirmed during comparison taking the supposed original as the blueprint. This carries the risk of a picking attitude of the expert who may select everything that appears to be similar and ignore all that is not.

Further, this promotes the risk of circular reasoning [6] or “gestalt analyses” [4], instead of proving origin by the similarities one “proves” similarities by the assumed origin.

It has been discussed that the ACE-V protocol is a recurring and reversible process [9, 10]. Opinions vary however whether or not the process should be totally recurring, and reversible (or up to a certain level) or that attempts should be made to confine it to a more linear process wherever possible. With a recurring and reversible process the risk of inserting information of the “known” exemplar into the unknown is higher than in a strict linear process in which ACE-V is executed once in the exact order.

The risk of making a (subconscious) decision early in the comparative process and the potential influence of it must be recognized [11]. The comparison must be an unbiased “step by step” building process ensuring that the data in the latent and comparison print match, with nothing in disagreement which cannot be

logically explained and accounted for. The decision must be made at the end of the process only.

An expert who has executed the process of searching and elimination has performed an initial and incomplete analysis directed towards elimination and/or the search process. Since he has singled out a comparison print for the identification process he has arrived at a preliminary conclusion about possible identification. With an eye to the “half baked” analysis and the preliminary conclusion it is advisable that the expert renounces himself from the identification process.

Evaluation and Preliminary Conclusion

Requirements for the conclusion of identification as provided by SWGFAST [12] are; agreement of sufficient friction ridge detail; determined by a competent examiner; applied to a common area in both impressions; based on quantity and quality of friction ridge detail; without any discrepancy and a reproducible conclusion. The total volume in agreement is a composition of coherent qualitative and quantitative information.

In the USA, after a 3-year study by a Standardization Committee, the use of a numerical standard was discouraged by the adoption of a resolution at a conference of the International Association for Identification which stated: “no scientific basis exists for requiring that a predetermined minimum of friction ridge features must be present in two impressions in order to establish positive identification”.

Sufficiency is since left to the discretion of the expert and measured against his training, knowledge and experience, and his personal standard. SWGFAST [12] relates reproducibility primarily to the format of ACE-V and to the conclusion. This position is known as the Expert Opinion System or the holistic approach [2, 13].

In many other countries a numerical standard is used as an aid to measure sufficiency which is called the Empirical Standard Approach [2, 13]. This standard expresses a minimum number of minutiae in agreement that is used as a common, empirical reference and a tool to guide the process, to facilitate verification and to obtain and guarantee quality.

In either system if an expert decides that in his opinion identification is justifiable because the equation is both sufficient and cogent he will put it up for verification [14].

Verification

The postulated conclusion should be reproducible by another examiner applying the same methodology. This is accomplished by the verification phase of the ACE-V methodology.

The reliability of a conclusion can be checked and demonstrated by an independent verification. Verification can be limited to another expert independently arriving at the same conclusion or by repeating and checking the whole examination of the initial examiner. The verification process should have the characteristics of scrutiny rather than confirmation of the conclusion. (Also see mistakes.)

If the verifier is satisfied that the process and the conclusion meet the requirements, then the conclusion is confirmed and the identification is established.

Conclusion

The conclusion of identification is a verified opinion that the investigated latent and the comparison print come from the same source. It also implicates the expectation of reproducibility, i.e., any other examiner using the same methodology should arrive at the same conclusion. Given the empirical, biological and statistical support for friction ridge skin uniqueness or individuality, an identified fingerprint is attributed to a single donor [15].

Charting

The use of a computer screen during the analysis and side by side comparison of friction ridge images can be of tremendous help in the examination process. The data and the relations of the configuration can be cross-checked, in particular with ambiguous information. Details can be better observed and compared by enlarging and/or enhancement of the images to optimize the perceptibility of the characteristics in print. This not only increases the quantity and accuracy of the data observed [16], but also makes it easier to value and appreciate the similarities and dissimilarities. At the same time similarities can be marked up, printed and saved for documentation purposes.

In order to meet the requirement of demonstrability of all the phenomena upon which the expert bases

his findings and conclusion this tool is indispensable. It also facilitates consultation and discussion amongst experts.

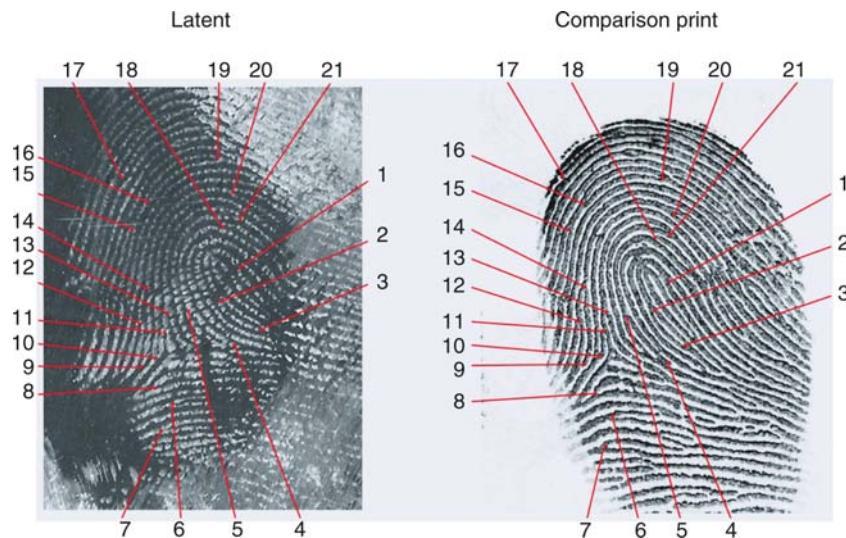
Historically, court charts have been produced in which coinciding minutiae have been marked up and numbered. Court charts can be a useful tool to demonstrate some of the findings but are just meager illustrations of a very complex process and should not be taken as ultimate proof. (The simple argument for that is the fact that in the past with erroneous identifications court charts were produced with even extensive numbers of marked similarities (Fig. 4).)

Mistakes

Error rates, an endless source of scientific debates, philosophies and semantics, will not be covered nor decided here but known errors will be discussed. It is obvious that, in relation to the immense numbers of identifications effected over the more than hundred years of fingerprinting, the number of erroneous identifications that surfaced is extremely low. In a study by Simon Cole [17] 22 erroneous identifications were investigated for the period from 1920 to 2004. Even if the number is tripled or multiplied by 10 incorporating a number of dark figures the positive ratio against the millions of identifications performed remains. Some support for this positive ratio is also found in data collected during comparison training exercises [18].

This does not implicate that mistakes are regarded as part of the system and inevitable, on the contrary. Every mistake is one too many and can do irreparable harm to innocent people. The profession should take all possible measures to prevent them.

Another ongoing debate is whether the mistakes can be attributed to flaws in the technique or to human error or whether the two can be separated at all. However, it is clear that erroneous identifications are discovered and exposed by experts. This is a strong indication that the human factor is dominant. Second, it is important to note that the examination of questioned identifications can be repeated and checked endlessly. When opinions differ upon sufficiency the comparsion should be regarded as inconclusive. In general, experts view mistakes very seriously. They believe that making a mistake is the worst thing that can happen to them and may discredit them in the eyes of their contemporaries.



Fingerprint Matching, Manual. **Figure 4** Example of a court chart with 21 coinciding Galton characteristics marked.

The paradox is that the acceptance of the susceptibility for human error by experts should be the basis for a quality system, whilst very often the initial response of experts to such criticism is defensive rather than open minded. This blocks the feedback essential for a quality cycle.

A preliminary analysis of mistaken identifications revealed the following factors:

- It concerned a border line latent with respect to quality, quantity or both.
- There was no apparent relationship between the organization or level of experience of the expert (s) involved.
- Verification was degenerated to confirmation rather than scrutiny.
- Experts were biased by domain irrelevant information.
- Discrepancies were ignored or erroneously attributed to distortion.
- Applied tolerances were too wide given the quality of the latent. This is another paradox; “the worse the print the larger the tolerances applied”; experts may attribute differences to the lack of quality and distortion and “explain them away” something they may not do with an image of good quality. Thus, bad quality may not only conceal real discrepancies, but also provide an excuse for it at the same time.

In general, there is a growing opinion that a number of psychological factors may potentially contribute to cognitive and decision-making errors [19].

Examples are; the primacy effect, when information is judged in the light of an early opinion; and confirmation biases like myside bias and truth bias [11] are found in all types of fields as well as in ordinary life. One major concern is that sufficiency may be established after the comparison process and as such after a conscious or subconscious decision is made about identity. This makes the expert more vulnerable to bias [19].

Studies have been done to enhance insight into the potential influences of bias during the examination of fingerprints [20].

Infallible or Reliable?

Some have criticized the profession for the explicit or implicit claim of infallibility [17, 21].

The apparent reliability of fingerprint identification for decades may have created this image as reflected by the proverbial expression “as reliable as a fingerprint”. This meant an image so strong that all other forensic techniques were compared against it, much like the introduction of DNA that was erroneously labeled the “genetic fingerprint”.

Responsible experts never claim infallibility because this is an unsustainable and unscientific position.

In retrospect, however, fingerprints in general can claim a record of great reliability, but as in any human endeavor mistakes occur so safeguards have to be in place.

The main ground for quality is the acceptance of fallibility by individuals and communities. With that in mind, instruments to achieve a solid conclusion, the rigorous application of the methodology, Quality Assurance protocols, training, testing and transparency, will be applied and maintained with conviction and can be further improved.

Per individual case reliability of a conclusion can be reached and demonstrated by verification, peer review and counterchecks by independent experts. This process can be repeated over and over again without affecting the material.

Related Entries

- ▶ Classification
- ▶ Feature Extraction
- ▶ Fingerprint Classification
- ▶ Fingerprint Matching Automatic
- ▶ Individuality

References

1. Ashbaugh, D.: Quantitative-Qualitative friction ridge analysis, p. 105. CRC Press, Boca Raton, FL (1999)
2. Method for Fingerprint Identification, Interpol European Expert Group for Fingerprint Identification Report I <http://www.interpol.int/Public/Forensic/fingerprints/WorkingParties/default.asp>
3. Huber, R.A.: Expert witness. *Criminal Law Quarterly*, 2, 276–295 (1959)
4. Rudin, N., Inman, K.: The proceedings of lunch The CAC News of the California Association of Criminalists, 2nd Quarter (2005) <http://www.cacnews.org/>
5. A Review of the FBI's Handling of the Brandon Mayfield Case, Office of the Inspector General, http://www.usdoj.gov/oig/special/s0601/PDF_list.htm pp. 7, 198–199. (2006)
6. Method for Fingerprint Identification, Interpol European Expert Group for Fingerprint Identification Report II. <http://www.interpol.int/Public/Forensic/fingerprints/WorkingParties/default.asp>
7. van Koppen, P.J., Crombag, H.H.: Over Oren, Lippen en Vingers. *Nederlands Juristenblad* (2000)
8. Hare, K.: Proportional analysis: The science of comparison. *J. Forensic Ident.* **53**, 700 (2003)
9. Vanderkolk, J.R.: ACE-V: A Model. *J. Forensic Ident.* **54** (2004)
10. Mc Kasson, S.C., Richards, C.A.: Speaking as an expert: A guide for the identification sciences from the laboratory to the courtroom, 131–138 (1998)
11. Nickerson, Raymond S.: Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* **2**, 175–220 (1998)
12. Scientific Working Group on Friction Ridge Analyses, Study and Technology http://www.swgfast.org/Standards_for_Conclusions_ver_1_0.pdf
13. C. Champod, et al., Fingerprints and other ridge skin impressions, 27–31 (2004)
14. Thornton, J.: “Setting Standards in the Comparison and Identification” In: 84th Annual Training Conference of the California State Division of IAI Laughlin, Nevada, May 9 (2000) <http://www.latent-prints.com/Thornton.htm>
15. Moenssens, A.A.: Is fingerprint identification a science? http://forensic-evidence.com/site/ID/ID00004_2.html
16. Langenburg, G.M.: A statistical analysis of the ACE-V methodology: Analysis stage. *J. Forensic Ident.* **54** (2004)
17. Cole, S.A.: More than zero: Accounting for error in latent fingerprint identification. *J. Crim. Law Criminol.* **95** (2005)
18. Wertheim, K., Langenburg, G. Moenssens, A.: “A report of latent print examiner accuracy during comparison training exercises”. *J. Forensic Ident.* **56**, 55–93 (2006)
19. Itiel, D., Charlton, D., Péron, A.E.: “Why are experts prone to error?” *Forensic Sci. Int.* **156**, 74–78 (2006)
20. Schiffer B., Champod, C.: The potential (Negative) influence of observational biases at the analysis stage of fingermark individualization. *Forensic Sci. Int.* **167**, 116–120 (2007)
21. Saks, M., Koehler, J.: The coming paradigm shift in forensic identification science. *Science*, **309**, 892 (2005)

Fingerprint Pre-Matching

- ▶ Fingerprint Classification

Fingerprint Quality

The intrinsic characteristic of a fingerprint image that may be used to determine its suitability for further processing by the biometric system or assess its conformance to pre-established standards is fingerprint quality. The quality of a biometric signal is a numerical value (or a vector) that measures this intrinsic attribute.

- ▶ Individuality of Fingerprints

Fingerprint Reading

► Biometric Sample Acquisition

Fingerprint Recognition, Overview

DAVIDE MALTONI
DEIS, University of Bologna, Italy

Synonym

Fingerprint Biometric

Definition

Fingerprint recognition allows a person to be verified or identified through the analysis and comparison of his or her finger dermal ridges. Fingerprint recognition was one of the first techniques used for automatically identifying people and today is still one of the most popular and effective biometric techniques.

Introduction

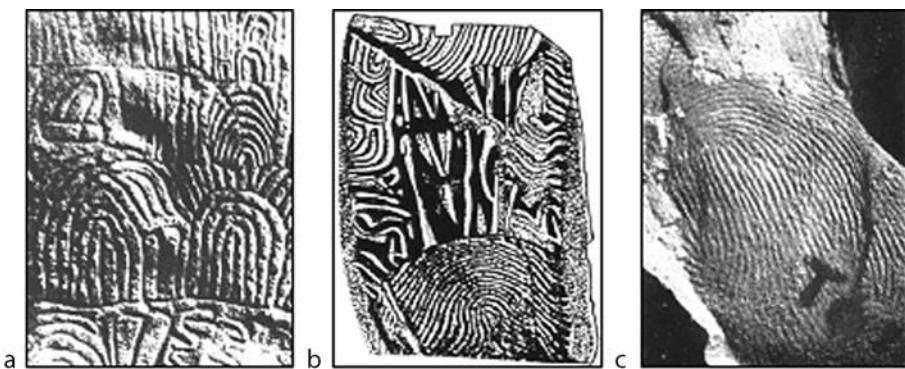
A fingerprint is the representation of the dermal ridges of a finger [1]. Dermal ridges form through a combination of genetic and environmental factors; the genetic code in DNA gives general instructions on the way skin should form in a developing fetus, but the specific way it forms is a result of random events such as the exact position of the fetus in the womb at a particular moment. This is the reason why even the fingerprints of identical twins are different [2]. Fingerprints are fully formed at about 7 months of fetus development and finger ridge configurations do not change throughout the life of an individual, except in case of accidents such as severe cuts on the fingertips. This stability makes fingerprints a very attractive biometric identifier. Several mathematical models based on the ► [anatomy of friction ridge skin](#) were developed over the years to quantify ► [fingerprint individuality](#) [3] and to prove that

finding two persons with identical fingerprints is extremely unlikely. This does not imply that fingerprint recognition is a perfect technique: in fact, various kinds of errors can affect fingerprint acquisition and processing thus requiring to introduce thresholds to decide if two fingerprint impressions are similar enough to be considered belonging to the same person. As for any biometric technique, a sound performance evaluation (see ► [fingerprint databases and evaluation](#)) is extremely important to estimate the accuracy of a fingerprint-based biometric system and to understand if it is well-suited for a particular application. Recent independent evaluation campaigns such as FVC2006 [4] proved that state-of-the art fingerprint recognition algorithms are nowadays very accurate (i.e., EER less than 0.1% for a database collected with a large area optical scanner).

History

Human fingerprints have been discovered on archaeological artefacts and historical items (Fig. 1). Although these findings prove that ancient people used fingerprints for a number of purposes, it was not until the late sixteenth century that the modern scientific fingerprint studies were initiated [5]. In 1686, Marcello Malpighi, a professor of anatomy at the University of Bologna, noted the presence of ridges, spirals and loops in fingerprints. Henry Fauld, in 1880, was the first to scientifically suggest the individuality of fingerprints based on an empirical observation. At the same time, Herschel asserted that he had practiced fingerprint recognition for about 20 years. In the late nineteenth century, Sir Francis Galton conducted an extensive study on fingerprints; in 1888 he introduced the ► [minutiae](#) features for fingerprint matching. Another important advance was made in 1899 by Edward Henry, who established the well-known “Henry system” of ► [fingerprint classification](#).

In the early twentieth century, fingerprint recognition was formally accepted as a valid identification method and became a standard routine in forensics [5]. Fingerprint identification agencies were set up worldwide and criminal fingerprint databases were established; for instance, the FBI fingerprint identification division was set up, in 1924, with a database of 810,000 fingerprint cards. With the rapid expansion of fingerprint recognition in forensics, operational



Fingerprint Recognition, Overview. **Figure 1** Examples of archaeological fingerprint carvings and historic fingerprint impressions: (a) Neolithic carvings (Gavrinis Island); (b) standing stone (Goat Island, 2000 B.C.); (c) an impression on a Palestinian lamp (400 A.D.). Figures courtesy of A. Moenssens and R. Gaensslen.

fingerprint databases grew so large that manual fingerprint identification (see ► [fingerprint matching, manual](#)) became infeasible; for example, the total number of fingerprint cards in the FBI fingerprint database stands well over 200 million and is continuously growing. With thousands of requests being received daily, even a team of more than 1300 fingerprint experts were not able to provide timely responses to these requests. Starting in the early 1960s, the FBI, Home Office in the UK, and Paris Police Department began to invest a large amount of effort in developing Automatic Fingerprint Identification Systems (► [AFIS](#)). Based on the observations of how human fingerprint experts perform fingerprint recognition, three major problems in designing AFIS were identified and investigated: digital fingerprint acquisition, local ridge feature extraction, and ridge characteristic pattern matching. Their efforts were so successful that today almost every law enforcement agency worldwide uses an AFIS. These systems have greatly improved the operational productivity of law enforcement agencies and reduced the cost of hiring and training human fingerprint experts.

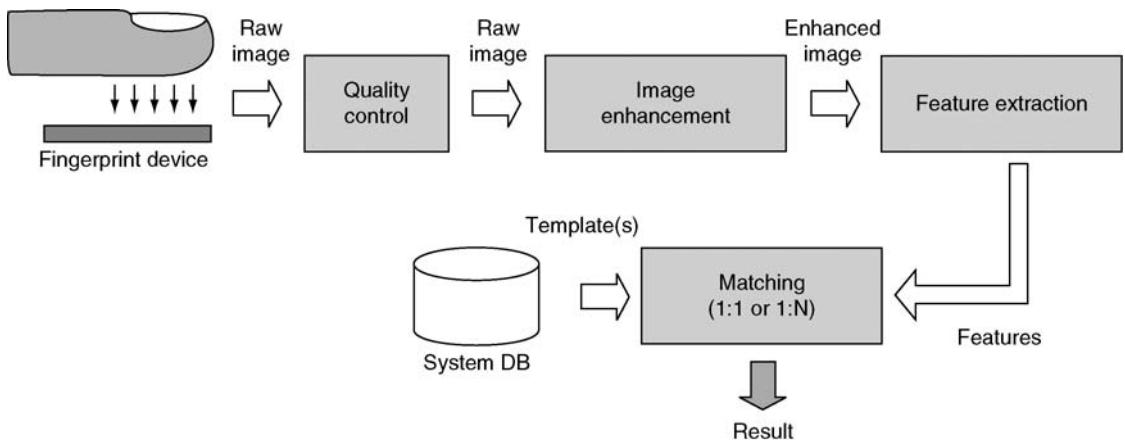
Automatic fingerprint recognition technology has now rapidly grown beyond forensic applications. On the one side, together with face, fingerprint is the main biometric modality for electronic documents (e-passport, visas, ID cards, etc) used to enforce border crossing and citizen security. On the other side, thanks to a very good performance/cost tradeoff, fingerprint-based biometric systems are becoming very popular and are being deployed in a wide range of commercial applications such as logon to computers and networks, physical access control, ATMs.

Components of a Fingerprint Recognition System

The block diagram of a fingerprint-based recognition system is depicted in [Fig. 2](#).

A fingerprint is acquired through a live-scan ► [fingerprint device](#) that allows to simply and quickly capture a digital fingerprint image: most of the fingerprint devices sample the pattern at 500 DPI (Dots per Inch) and produce an 8-bit gray-scale raw image (see [Fig. 3](#)). Some devices also include fake detection mechanisms (see ► [fingerprint fake detection](#)) that allow to reveal spoofing attacks carried out with fake fingers.

The acquired raw image is then passed to a quality control module that evaluates if the fingerprint sample quality is good enough to correctly process it and to extract reliable features. In case of insufficient quality, the system rejects the sample and invites the user to repeat the acquisition; otherwise, the raw image is passed to an ► [image enhancement](#) module whose goal is improving the clarity of the ridge pattern, especially in noisy region, to simplify the subsequent feature extraction. Special digital filtering techniques, known as contextual filtering [1], are usually adopted at this stage; the output enhanced image can still be a gray-scale image or become a black-and-white image. The ► [feature extraction](#) module further processes the enhanced image and extracts a set of features from it. This feature set often includes minutiae but, depending on the matching algorithm, other features (e.g., local orientation, local frequency, singularities, ridge shapes, ridge counts, parts of the enhanced image, etc.) can be extracted in conjunction with (or instead of) minutiae.



Fingerprint Recognition, Overview. [Figure 2](#) Block diagram of a fingerprint-based recognition system.



Fingerprint Recognition, Overview. [Figure 3](#) Example of fingerprint images from FVC2006 databases [4].

Finally, the fingerprint matching module (see ▶ [fingerprint matching, automatic](#)) retrieves from a system database one or more templates (see ▶ [fingerprint templates](#)) and matches it/them with the features extracted from the current sample. Most of the matching algorithms, following the well established manual method (see ▶ [fingerprint matching, manual](#)), compare two fingerprints by searching the spatial correspondence of a minimum number of minutiae; this is not a simple task because of the large variations (e.g., displacement, rotation, skin condition, distortion, noise, etc.) that can characterize two fingerprint images acquired from the same finger at different times. If the system is operating in verification mode, the user has been required to claim his identity and therefore just one template is retrieved from the database and matched with the current sample; if the system is

operating in identification mode the current sample is matched against all the database templates to check if one of them is sufficiently similar.

Protecting fingerprint templates is very important to avoid attacks to fingerprint-based biometric systems [6] and to preserve user privacy: cryptography techniques can be used to this purpose (see ▶ [Fingerprints Hashing](#)).

Large-Scale Automatic Fingerprint Identification Systems

Large-scale automatic fingerprint identification systems (AFIS) are used in forensic and civil government applications. The basic functioning of these systems is the same as described in the previous section, but a number of ad-hoc optimizations are employed to effectively and efficiently store, retrieve and match millions of fingerprints in a few seconds. In the past, special dedicated hardware and storage devices were used to guarantee the required throughput; nowadays, most of the AFIS cores run on conventional hardware (e.g., cluster of personal computers) and the software is the main responsible of the system efficiency. Fingerprint classification and ▶ [fingerprint indexing](#) are the two main techniques used to speed-up a fingerprint search in a large database [1]. The former allows to split the database in a number of partitions and to limit the search to the partition to which the searched sample belongs to. The latter enables sorting the database templates according to the similarity with the searched

sample, so that the probability to find a mate in the first attempts increases significantly. Even if the capacity of mass storage devices is continuously growing, storing fingerprints as uncompressed raw images would require too much space (nowadays AFIS must store billions of fingerprint images) and would increase the time necessary to transmit a fingerprint record over a network; to alleviate this problem, without compromising recognition accuracy, specific ► **fingerprint compression** techniques such as WSQ (Wavelet Scalar Quantization) have been developed by researchers.

Related Entries

- Biometrics, Overview
- Biometric Recognition
- Fingerprint Anatomy
- Fingerprint Classification
- Fingerprint Compression
- Fingerprint Databases and Evaluation
- Fingerprint Fake detection
- Fingerprint Features
- Fingerprint Image Enhancement
- Fingerprint Indexing
- Fingerprint Individuality
- Fingerprint Matching, Automatic
- Fingerprint Matching, Manual
- Fingerprint Image Quality
- Fingerprint Templates
- Fingerprints Hashing

References

1. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: *Handbook of fingerprint recognition*. Springer, New York (2003)
2. Jain, A.K., Prabhakar, S., Pankanti, S.: On the similarity of identical twin fingerprints. *Pattern Recognit.* **35**(11), 2653–2663 (2002)
3. Pankanti, S., Prabhakar, S., Jain, A.K.: On the individuality of fingerprints. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(8), 1010–1025 (2002)
4. The Fourth International Fingerprint Verification Competition (FVC2006) <http://bias.csr.unibo.it/fvc2006>.
5. Lee, H.C., Gaenslen, R.E.: *Advances in fingerprint technology*. 2nd Edn. Elsevier, New York (2001).
6. Cappelli, R., Lumini, A., Maio, D., Maltoni, D.: Fingerprint image reconstruction from standard templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(9), 1489–1503 (2007)

Fingerprint Representation

- Fingerprint Templates

Fingerprint Retrieval

Fingerprint retrieval is a procedure that draws a subset of fingerprints from a database stored on a computer system based on some similarity measure between a query fingerprint and the fingerprints in the database. The ultimate goal of fingerprint retrieval is not to find a group of fingerprints similar to the query fingerprint, but to get back the fingerprint originating from the same finger as that of the query fingerprint. Hence, success or failure of the fingerprint retrieval is determined by whether the retrieved subset contains the fingerprint originating from the same finger as that of the query fingerprint.

- Fingerprint Classification
- Fingerprint Indexing

Fingerprint Sample Synthesis

RAFFAELE CAPPELLI

Biometric System Laboratory, DEIS, University of Bologna, Cesena, Italy

Synonyms

Synthetic fingerprint generation; Synthetic fingerprints; Artificial fingerprints

Definition

Fingerprint sample synthesis is the generation of images similar to human fingerprints, through parametric models that simulate the main characteristics of such biometric data and their modes of variation. The image synthesis is typically performed by a computer program that, starting from some input

parameters, executes a sequence of algorithmic steps that finally produce a synthetic fingerprint image.

Introduction

With the increasingly adoption of fingerprint recognition systems, driven by their very appealing accuracy/cost tradeoff, methodical and accurate performance evaluations of fingerprint recognition algorithms are needed. Unfortunately, this requires large databases of fingerprints, due to the very small rates of error necessary for the procedure. For instance, according to [1], in order to support a claim of FMR less than 1/10,000 (the requirement for verification applications in [2]), 30,000 impostor matches from at least 250 individuals should be performed without observing any false match error. On the other hand, collecting large databases of fingerprint images is expensive both in terms of money and time, boring for both the people involved and for the volunteers, and problematic due to the privacy legislation that protects such personal data. FVC competitions [3] are examples of technology evaluations, where real fingerprint databases have been collected to test different algorithms, but do not constitute lasting solutions for evaluating and comparing different algorithms; in fact, since FVC databases are made available to the participants after the competition to let them improve the technology, they expire once “used,” and new databases have to be collected for future evaluations.

Fingerprint synthesis is a feasible way to address the issues just cited, since it allows large databases of images to be easily generated and used for testing fingerprint recognition systems without infringing on privacy.

A fingerprint synthesis method typically consists of two main steps: first, a ridge pattern, which represents the unique and immutable characteristics of a “synthetic finger,” is generated according to a given model; then, one or more “fingerprints” of the synthetic finger are generated by simulating the main factors that make the fingerprints of a given human finger different each other.

Physical Ridge Pattern Models

Physical ridge pattern models are based on some hypothesized physical mechanisms of fingerprint formation during embryogenesis.

The crucial period of fingerprint development in humans starts at the 10th week of pregnancy [4], when the epidermis consists of three layers (outside layer, intermediate layer and basal layer). It is then observed that the basal layer of the epidermis becomes undulated toward the surface, forming the so-called “primary ridges,” whose development ends at about the 17th week of pregnancy: at this stage the geometry of the epidermal ridge pattern is determined for life and becomes visible on the skin surface in subsequent weeks.

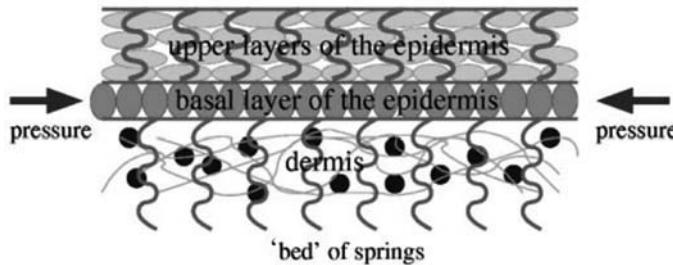
Several theories for fingerprint pattern formation have been proposed in the scientific literature [4], including cell proliferation phenomena, mechanical interaction between the extracellular matrix and fibroblasts in the dermis, reaction-diffusion models.

In a study by Sherstinsky and Picard [5], a complex method which employs a dynamic non-linear system called “M-lattice,” is introduced. The method is based on the reaction-diffusion model first proposed by Turing in 1952 to explain the formation of animal patterns such as zebra stripes. Although this work is aimed at optimally binarizing a fingerprint image, the underlying ridge-line model could be used as a basis for synthetic generation.

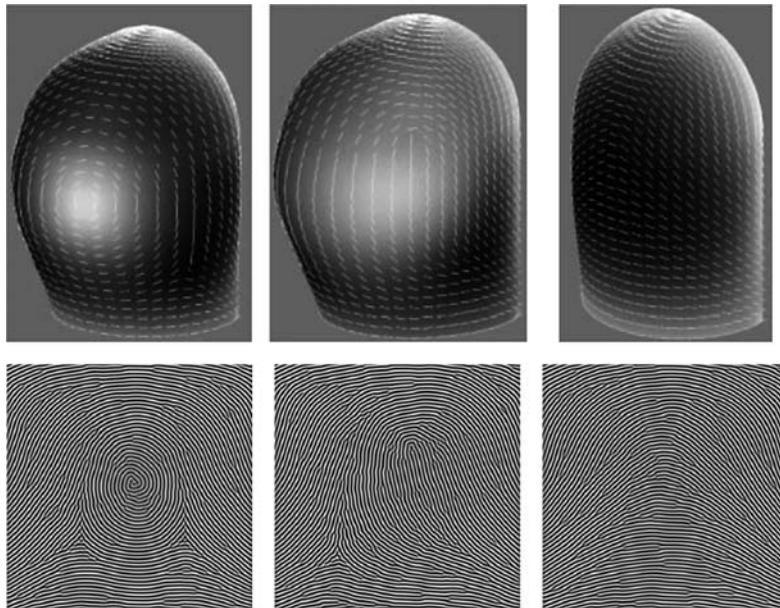
An interesting model was proposed by Kücken [4, 6], based on the following hypotheses:

1. Fingerprint patterns are created by forces that are induced by differential growth of the epidermis’ basal layer (as argued by Cummins [7] from the observed dependency of the pattern class on the fingertip geometry)
2. Non-uniform growth of the epidermis’ basal layer results in compressive stress that leads to buckling, creating the primary-ridges [8]

Kücken considers the basal layer as an elastic sheet trapped between the neighboring tissues of the intermediate epidermis layer and the dermis (Fig. 1) and studied the buckling process by means of the von Karman equations, which describe the behavior of a thin curved sheet of elastic material. The analysis of those equations confirmed that the direction of the ridges is roughly perpendicular to the direction of greatest stress; Kücken postulated that two factors mainly contribute to generate the compressive stress in the basal layer: (1) resistance at the nail furrow and at the major flexion creases of the finger (boundary effects); (2) the regression of the “volar pads” at the time of fingerprint development. Volar pads are



Fingerprint Sample Synthesis. **Figure 1** The basal layer of epidermis: Kücken and Newell [6] assumes that due to differential growth, a compressive stress act on this layer.



Fingerprint Sample Synthesis. **Figure 2** Simulation of three common fingerprint patters (from left to right: whorl, loop, and arch) using the model proposed in [6].

temporary eminences of the skin surface that form during the 7th week of pregnancy and start to digress at about the 10th week. From studies of embryos, monkeys and malformed hands, it has consistently been observed that highly rounded pads at the fingertips exhibit whorls; less well-developed pads show loops, where the direction of the loop opening is determined by the asymmetry of the pad; small indistinct pads give rise to arches.

Computer simulations have shown results consistent with the above observations and hypothesis; Fig. 2 shows how an almost periodic pattern very similar to human fingerprints can be generated by applying Kücken's model: the three main fingerprint classes can be simulated and ▶ minutiae are present in regions

where ridge patches with different directions and/or wavelength meet.

Statistical Ridge Pattern Models

Statistical ridge pattern models aims to reproduce realistic-looking fingerprints without starting from embryological hypothesis. Such models are based on the empirical analysis of real fingerprints, from which statistical data about the main characteristics of the patterns are derived and parameterized into appropriate equations or synthesis algorithms.

In 1999, Kosz published some interesting results concerning fingerprint synthesis based on a

mathematical model of ridge patterns and minutiae [9]; further details on this technique have been provided online by Bicz [10] in 2003. According to this model, a fingerprint can be described by a wave pattern:

$$f(x, y) = \cos(\varphi(x, y)) \quad (1)$$

where:

$$\varphi(x, y) = \varphi_0(x, y) + \varphi_M(x, y) \quad (2)$$

is a function that defines the phase of the wave structure as the sum of two parts: φ_0 , which describes the global “shape” of the ridge lines, and φ_M , which describes the minutiae. According to the model introduced by bicz [10], φ_M can simply generate n minutiae by adding n spatially-shifted arctangent functions:

$$\varphi_M(x, y) = \sum_{i=1}^n \arctan\left(\frac{y - y_i}{x - x_i}\right) \quad (3)$$

where (x_i, y_i) is the location of minutia i . Figure 3 shows a synthetic pattern generated by using the above equations.

In 1993, Sherlock and Monro [11] proposed an orientation model that allows a consistent ▶ orientation field to be computed from the sole knowledge



Fingerprint Sample Synthesis. Figure 3 A simple synthetic pattern generated by equations (1)–(3), with $\varphi_0(x, y) = 20 \cdot 2\pi \cdot \sqrt{x^2 + y^2}$ and $\{(x_i, y_i)\} = \{(0.2, -0.25), (-0.2, -0.37), (0.0, 0.2), (-0.25, 0.3), (0.2, 0.43)\}$.

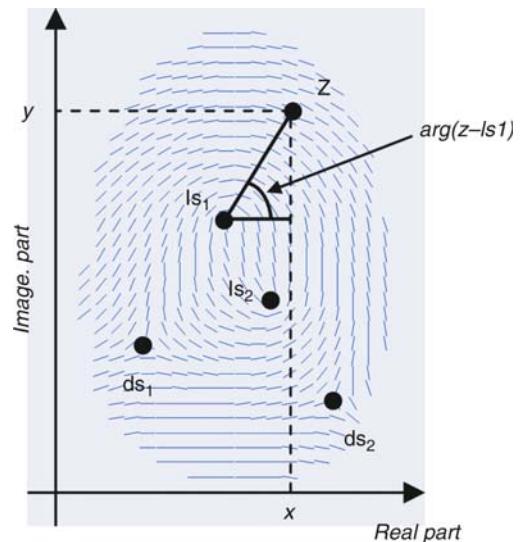
of the position of fingerprint ▶ singularities (loops and deltas). In this model, the image is located in the complex plane and the local ridge orientation is the phase of the square root of a complex rational function whose singularities (poles and zeros) are located at the same place as the fingerprint singularities (loops and deltas). Let \mathbf{ls}_i , $i = 1..n_c$ and \mathbf{ds}_i , $i = 1..n_d$ be the coordinates of the loops and deltas respectively. The orientation θ at each point $\mathbf{z} = [x, y]$ is calculated as:

$$\theta = \frac{1}{2} \left[\sum_{i=1}^{n_d} \arg(\mathbf{z} - \mathbf{ds}_i) - \sum_{i=1}^{n_c} \arg(\mathbf{z} - \mathbf{ls}_i) \right] \quad (4)$$

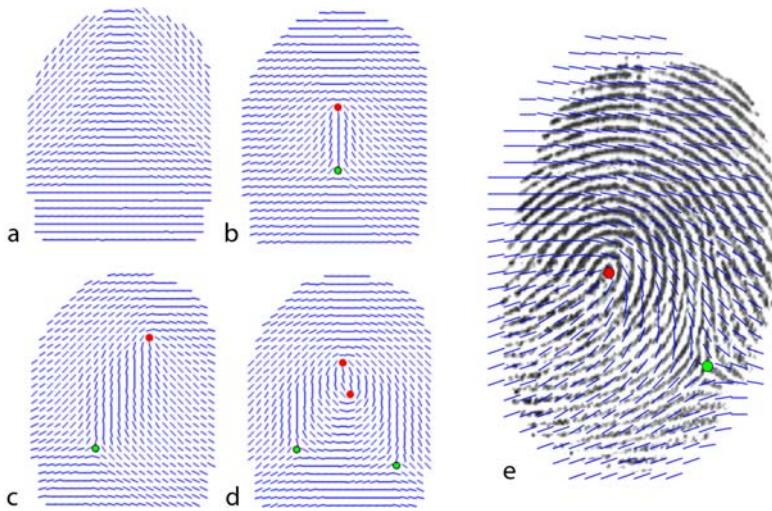
where the function $\arg(\mathbf{c})$ returns the phase angle of the complex number \mathbf{c} (see Fig. 4).

The Sherlock and Monro model may be exploited for generating synthetic orientation fields by first randomly choosing a fingerprint class and then randomly selecting the positions of the singularities, according to the class-specific constraints (for instance, in a left loop, the delta must be on the right side of the loop). Figure 5 shows some examples of orientation fields generated by this model.

However, in nature the ridge-line flow cannot be completely determined by the singularity type and position. In 1996, Vizcaya and Gerhardt proposed a variant of the Sherlock and Monroe model that



Fingerprint Sample Synthesis. Figure 4 Sherlock and Monro model: each element of the orientation field is considered as a complex number.



Fingerprint Sample Synthesis. **Figure 5** An example of Arch (a), Tented Arch (b), Right Loop (c) and Whorl (d) orientation field as generated by the Sherlock and Monroe model. In (e), an example of left-loop orientation field superimposed to a real left-loop fingerprint with coincident singularity positions.

introduces more degrees of freedom to cope with the orientation variability that may characterize orientation fields with coincident singularities. The orientation θ at each point \mathbf{z} is calculated as:

$$\theta = \frac{1}{2} \left[\sum_{i=1}^{n_d} g_{ds_i}(\arg(\mathbf{z} - \mathbf{ds}_i)) - \sum_{i=1}^{n_c} g_{ls_i}(\arg(\mathbf{z} - \mathbf{ls}_i)) \right] \quad (5)$$

where $g_k(\alpha)$, for $k \in \{ls_1, \dots, ls_{n_c}, ds_1, \dots, ds_{n_d}\}$, are piecewise linear functions capable of locally correcting the orientation field with respect to the value given by Sherlock and Monroe model:

$$g_k(\alpha) = \bar{g}_k(\alpha_i) + \frac{\alpha - \alpha_i}{2\pi/L} (\bar{g}_k(\alpha_{i+1}) - \bar{g}_k(\alpha_i)) \quad (6)$$

for $\alpha_i \leq \alpha \leq \alpha_{i+1}$, $\alpha_i = -\pi + \frac{2\pi i}{L}$.

Each function $g_k(\alpha)$ is defined by the set of values $\{\bar{g}_k(\alpha_i) | i = 0..L-1\}$, where each value is the amount of correction of the orientation field at a given angle (in a set of L angles uniformly distributed between $-\pi$ and π). If $\bar{g}_k(\alpha_i) = \alpha_i \forall i \in \{0..L-1\}$ (i.e. $g_k(\alpha)$ is the identity function), the model coincides with that of Sherlock and Monroe.

Figure 6a and **b** show two examples of orientation fields generated according to the Vizcaya and Gerhardt model; these images are definitely more realistic than those in **Fig. 5**. The superiority of the Vizcaya and Gerhardt model in approximating existing ridge patterns is also evident from the comparison between **Fig. 6c** and **d**.

In 2000, Cappelli et al. introduced a ridge pattern generation approach based on the following steps [12]:

1. Orientation field generation
2. Frequency map generation
3. Ridge pattern generation

Step 1 adopts the Vizcaya and Gerhardt model for generating the orientation field starting from the positions of loops and deltas; for generating arch type patterns (which do not contain any singularity), a simple sinusoidal function, whose frequency and amplitude are tuned to control the arch curvature and aspect, is used.

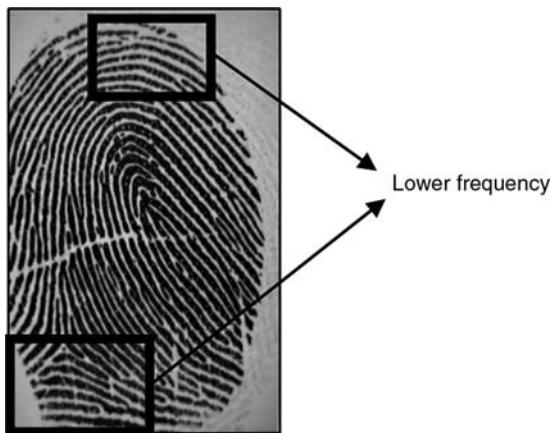
Step 2 creates a frequency map (see Fingerprint Feature Extraction) on the basis of some heuristic criteria inferred by the visual inspection of a large number of real fingerprints (for instance, in the regions above the northernmost loop and below the southernmost delta, the ridge-line frequency is often lower than in the rest of the fingerprint, see **Fig. 7**).

Finally step 3, given an orientation field and a frequency map as input, generates a ridge line pattern by iteratively enhancing an initial image (containing one or more isolated points) through ► **Gabor filters**. The filters are applied at each pixel (x, y) and adjusted according to the local ridge orientation ϕ_{xy} and frequency v_{xy} :

$$\begin{aligned} & gabor(r, s : \phi_{xy}, v_{xy}) \\ &= e^{-\frac{(r+s)^2}{2\sigma^2}} \cdot \cos[2\pi v_{xy}(r \sin \phi_{xy} + s \cos \phi_{xy})] \end{aligned} \quad (7)$$



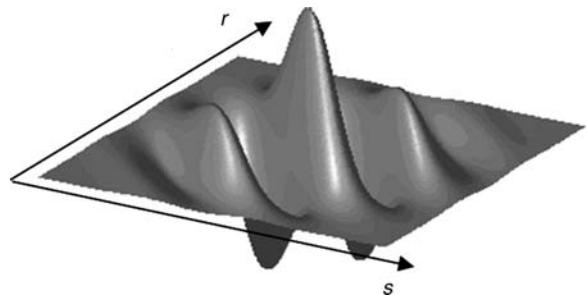
Fingerprint Sample Synthesis. **Figure 6** An example of Right Loop (**a**) and Whorl (**b**) orientation fields, as generated by the Vizcaya and Gerhardt model. In (**c**) and (**d**) the orientation fields produced by the two models, for a given fingerprint, are compared.



Fingerprint Sample Synthesis. **Figure 7** An example of a right-loop fingerprint where the ridge-line frequency is lower in the regions above the loop and below the delta.

Parameter σ , which determines the bandwidth of the filter, is set according to the frequency, so that the filter does not contain more than three effective peaks (see Fig. 8).

While one could reasonably expect that iteratively applying “striped” filters to random images would simply produce striped images, very realistic minutiae are generated at random positions. Based on their experiments, in [12] the authors argue that minutiae primarily originate from the ridge-line disparity produced by local convergence/divergence of the

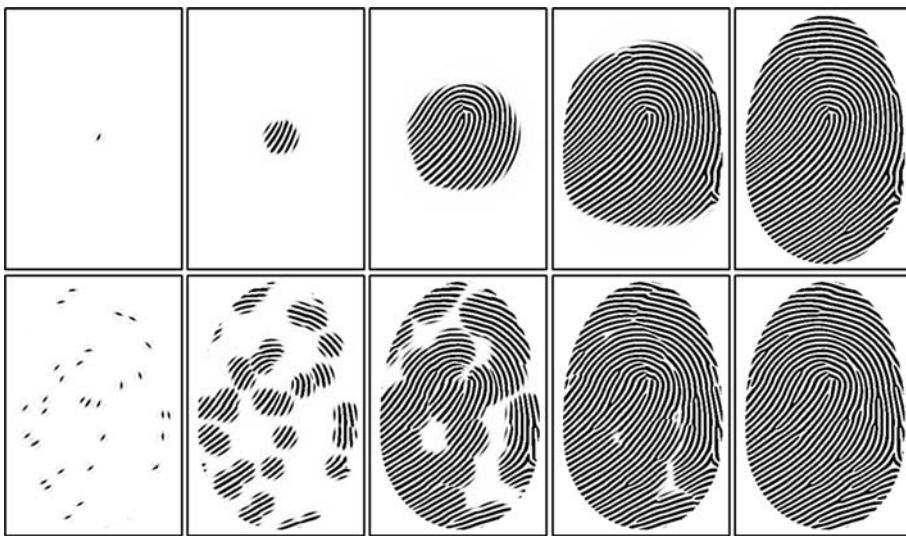


Fingerprint Sample Synthesis. **Figure 8** An example of Gabor filter used in step 3: note that the bandwidth is adjusted so that the filter does not contain more than three peaks.

orientation field and by frequency changes. In Fig. 9, examples of the iterative ridge-line generation process are shown; the authors experimentally found that increasing the number of initial points determines a more irregular ridge pattern richer of minutiae: this is not surprising, since expanding distinct image regions causes interference where regions merge, thus favoring the creation of minutiae (see Fig. 10).

Generation of Synthetic Fingerprint Impressions

Several factors contribute in making the impressions of a real finger substantially different when captured



Fingerprint Sample Synthesis. **Figure 9** Some intermediate steps of a fingerprint-generation process starting from a single central point (top) and from a number of randomly located points (bottom). Usually, increasing the number of initial points determines a more irregular ridge pattern richer of minutiae.



Fingerprint Sample Synthesis. **Figure 10** Genesis of a minutia point during the merging of the two regions originated by two different initial points.

by an on-line acquisition sensor (see ▶ [Fingerprint Device](#)):

1. Displacement in x and y direction and rotation
2. Different touching areas
3. Non-linear distortions produced by non-orthogonal pressure of the finger against the sensor
4. Variations in the ridge-line thickness given by pressure intensity or by skin dampness
5. Small cuts or abrasions on the fingertip
6. Background noise and other random noise

In 2002, Cappelli et al. proposed an evolution of the approach introduced in [13], which is able to simulate most of the above factors, thus generating very realistic fingerprint impressions. Starting from a synthetic ridge-line pattern, the main steps involved in the simulation of a fingerprint impression are: (1) Variation of the ridge thickness; (2) Skin distortion; (3) Noising and global translation/rotation; (4) Background generation. The subsections that follow briefly

describe the various steps, as they were proposed by Cappelli [14].

Variation of the Ridge Thickness

Skin dampness and finger pressure against the sensor platen have similar effects on the acquired images: when the skin is dry or the pressure is low, ridges appear thinner, whereas, when the skin is wet or the pressure is high, ridges appear thicker (see Fig. 11).

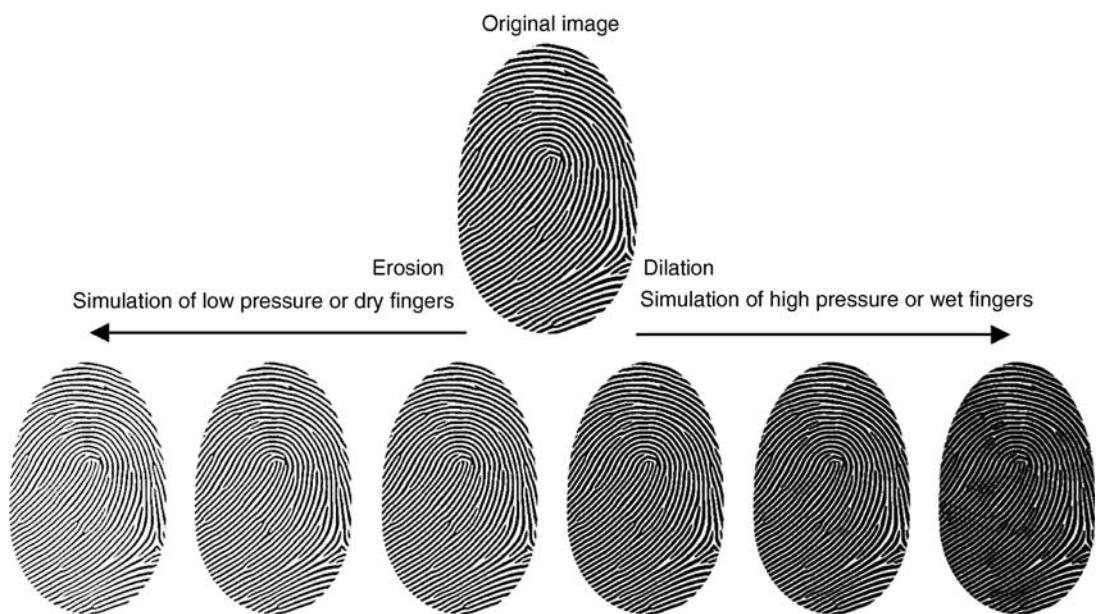
Morphological operators (see Image Preprocessing) are applied to the ridge line pattern, to simulate different degrees of dampness/pressure. In particular, the erosion operator is applied to simulate low pressure or dry skin, while the dilation operator is adopted to simulate high pressure or wet skin (see Fig. 12).

Skin Distortion

One of the main aspects that distinguish the different impressions of the same finger is the presence of non-linear distortions, mainly due to skin deformations



Fingerprint Sample Synthesis. **Figure 11** Three impressions of the same real finger as captured when the finger is dry, normal and wet, respectively.



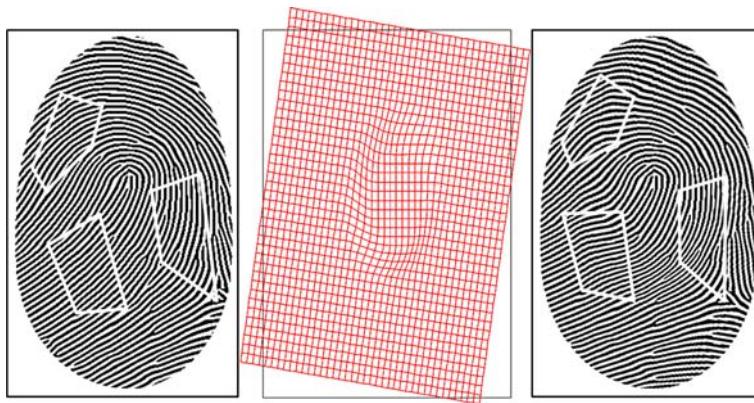
Fingerprint Sample Synthesis. **Figure 12** The application of different levels of erosion/dilation to the same ridge line pattern.

according to different finger placements over the sensing element (see Fig. 13). In fact, due to the skin plasticity, the application of forces, some of whose components are not orthogonal to the sensor surface, produces non-linear distortions (compression or stretching) in the acquired fingerprints (see ▶ [Fingerprint Matching](#)).

In “Synthetic Fingerprint Generation” [14], the skin-distortion model introduced by Cappelli, Maio, and Maltoni [15] is exploited. While in the latter, the distortion model was applied to re-map minutiae points, in order to improve fingerprint matching,



Fingerprint Sample Synthesis. **Figure 13** Two impressions of the same real finger where a few corresponding minutiae are marked to highlight distortion.



Fingerprint Sample Synthesis. [Figure 14](#) A synthetic ridge line pattern (on the left) and a distorted impression (on the right); the equivalent distortion of a square mesh is shown in the middle. To better highlight the non-linear deformations, some corresponding minutiae are connected by white segments in both the fingerprint images.

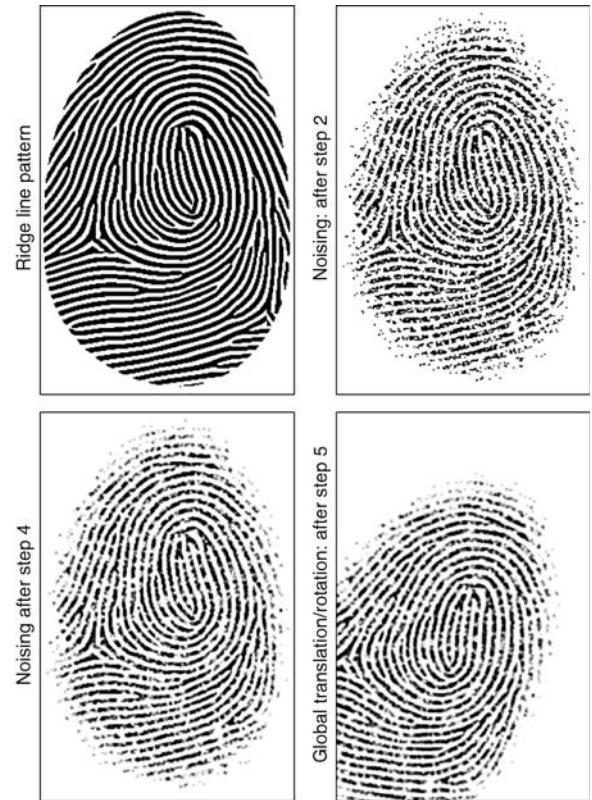
here the mapping has to be applied to the whole image, in order to simulate realistic distorted impressions. In [Fig. 14](#), a ridge line pattern and its distorted impression are shown.

Noising and Global Translation/Rotation

During fingerprint acquisition, several issues contribute to deteriorate the original signal, thus producing a gray-scale noisy image: irregularity of the ridges and their different contact with the sensor surface, presence of small pores within the ridges, presence of very-small-prominence ridges, gaps and cluttering noise due to non-uniform pressure of the finger against the sensor. Furthermore, the fingerprint is usually not perfectly centered in the image and can present a certain amount of rotation. The noising phase sequentially performs the following steps:

1. Isolate the valley white pixels into a separate layer. This is simply performed by copying the pixels brighter than a fixed threshold to a temporary image
2. Add noise in the form of small white blobs of variable size and shape. The amount of noise increases with the inverse of the fingerprint border distance
3. Smooth the resulting image with a 3×3 averaging box filter
4. Superimpose the valley layer to the image obtained
5. Rotate and translate the image

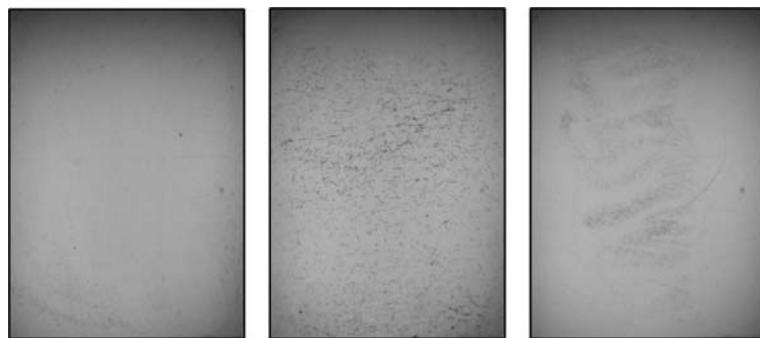
Steps 1 and 4 are necessary to avoid an excessive overall image smoothing. [Figure 15](#) shows an example where the intermediate images produced after steps 2, 4 and 5 are reported.



Fingerprint Sample Synthesis. [Figure 15](#) An example of noising and global translation/rotation, where the intermediate images produced after steps 2, 4 and 5 are reported.

Background Generation

The output of the previous step is a fingerprint that appears realistic, but the image background is completely white. In order to generate backgrounds



Fingerprint Sample Synthesis. [Figure 16](#) Examples of background-only images (acquired from an optical sensor) used for training the background generator.



Fingerprint Sample Synthesis. [Figure 17](#) Three synthetic images with backgrounds generated according to the model in [14].

similar to those of fingerprints images acquired with a given sensor, a statistical model based on the KL transform (see ▶ Dimensionality Reduction) is adopted. The model requires a set of background-only images as a training set (see Fig. 16): a linear subspace that represents the main variations in the training backgrounds is calculated and then used to randomly generate new backgrounds.

Figure 16 shows some examples of the background images (obtained from an optical acquisition sensor) used as a training set for the background generation step; Fig. 17 reports three synthetic fingerprints with backgrounds generated according to the above-described model.

Related Entries

- ▶ Anatomy of Fingerprint
- ▶ Biometric Sample Synthesis
- ▶ Fingerprint Classification

- ▶ Fingerprint Databases and Evaluation
- ▶ Fingerprint Features
- ▶ Fingerprint Singularities, Minutiae, Pores

References

1. UK Biometrics Working Group: Biometric Evaluation Methodology (2002)
2. Biometric Information Management and Security, American National Standards Institute, X9.84, 2001
3. Cappelli, R., Maio, D., Maltoni, D., Wayman, J.L., Jain, A.K.: Performance evaluation of fingerprint verification systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(1), 3–18 (2006)
4. Kücken, M.: Models for fingerprint pattern formation. *Forensic Sci. Int.* **171**(2–3), 85–96 (2007)
5. Sherstinsky, A., Picard, R.W.: Restoration and enhancement of fingerprint images using M-lattice – a novel non-linear dynamical system. Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, pp. 195–200 (1994)
6. Kücken, M., Newell, A.C.: A model for fingerprint formation. *Europhys. Lett.* **68**(1), 141 (2004)

7. Cummins, H.: Epidermal-ridge configurations in developmental defects, with particular reference to the ontogenetic factors which condition ridge direction. *Am. J. Anat.* **38**, 89–151 (1926)
8. Bonnevie, K.: Studies on papillary patterns in human fingers. *J. Genet.* **15**, 1–111 (1924)
9. Kosz, D.: New numerical methods of fingerprint recognition based on mathematical description of arrangement of dermatoglyphics and creation of minutiae. In: Mintie, D. (ed.) *Biometrics in Human Service User Group Newsletter*, (1999)
10. Bicz, W.: "The idea of description (reconstruction) of fingerprints with mathematical algorithms and history of the development of this idea at Optel," (Optel, 2003). <http://www.optel.pl/article/english/idea.htm>. Accessed 18 Dec 2007
11. Sherlock, B.G., Monro, D.M.: A model for interpreting finger-print topology. *Pattern Recognit.* **26**(7), 1047–1055 (1993)
12. Cappelli, R., Erol, A., Maio, D., Maltoni, D.: Synthetic finger-print-image generation. Proceedings of the 15th International Conference on Pattern Recognition, vol. 3, pp. 475–478. Barcelona (2000)
13. Cappelli, R., Maio, D., Maltoni, D.: "Synthetic fingerprint-database generation." Proceedings of the 16th International Conference on Pattern Recognition, vol.3, pp. 744–747. Québec City (2002)
14. Cappelli, R.: Synthetic fingerprint generation. In: Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S. (eds.) *Handbook of Fingerprint Recognition*. Springer, New York (2003)
15. Cappelli, R., Maio, D., and Maltoni, D.: "Modelling plastic distortion in fingerprint images". Proceedings of the Second International Conference on Advances in Pattern Recognition (ICAPR2001), Rio de Janeiro, pp. 369–376 (2001)

Fingerprint Scan

- Biometric Sample Acquisition

Fingerprint Sensor

- Fingerprint, Palmprint, Handprint and Soleprint Sensor

Fingerprint Signatures

- Fingerprint Features

Fingerprint Singularity

Fingerprint singularity is defined as a core or delta of a fingerprint. Minutiae have dual correspondence between the normal image and density-inverted image, that is, terminations appear as bifurcations and vice versa. However, singularity does not have such trait.

- Fingerprint Image Enhancement

Fingerprint Skeletonization

Fingerprint skeletonization, also referred to as thinning, is the process of reducing the width of binarized ridgelines to 1 pixel. Standard thinning algorithms are applicable. Modified methods based on local ridge orientation have been proposed to improve skeletonization accuracy. Post-processing for skeleton image, such as skeleton adjustment, is also important.

- Fingerprint Image Enhancement

Fingerprint Templates

WEI-YUN YAU

Institute for Infocomm Research, Agency for Science, Technology & Research, Singapore

Synonym

Fingerprint representation

Definition

A fingerprint template is a set of stored fingerprint features extracted from the fingerprint of a user. It is stored during the enrollment process to represent the actual owner of the fingerprint. It is subsequently compared directly to the fingerprint features of the query fingerprint in order to establish whether the

query fingerprint is obtained from the same person as the actual owner. It should be noted that the original fingerprint or its enhanced or compressed form is not a fingerprint template.

Introduction

As discussed in the section on general biometrics, the operation of a fingerprint recognition system, just like any other biometric system, follows a common process flow as shown in Fig. 1.

A fingerprint sensor is required to capture the fingerprint image which is then processed by a feature extractor to obtain the unique features of the fingerprint. If the user is new, the features extracted are stored in a database, typically along with other personal details of the new user such as name, and identification number. This set which stores the fingerprint feature is commonly known as a *fingerprint template*. The process by which this is done is called the *enrollment* process. Subsequently, when the user wants to use the fingerprint recognition system, the fingerprint features extracted from the fingerprint image acquired live or as a query image provided into the system are compared against the stored *fingerprint template(s)*. If the comparison involves only one *fingerprint template* from the database, such as when the user key in the name to retrieve the enrolled fingerprint template, the comparison process is called *verification*. Alternatively, the comparison can be done against all the fingerprint templates stored in the database and such a process is referred to as *identification*.

Composition of Fingerprint Template

In general, a fingerprint template contains the unique features extracted from the fingerprint image.

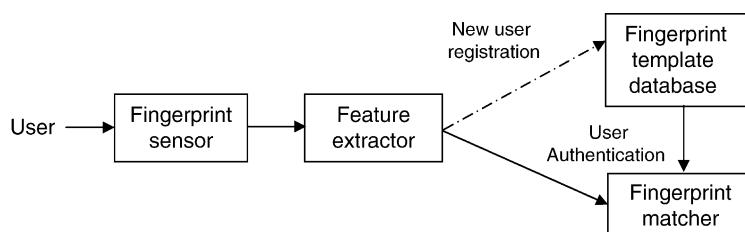
However, the exact content varies according to the type of algorithm used to extract and match the fingerprint. Nevertheless, if the stored file is merely the enhanced or compressed fingerprint image or the original fingerprint image itself, it is not considered a fingerprint template. There are two general types of algorithm used in fingerprint feature extraction and matching, namely, minutia-based and pattern-based or ridge feature-based [1, 2]. Minutia arises when a fingerprint ridge comes to an end (called ridge ending) or when it forks out into two ridges (called bifurcation). A sample fingerprint image with the detected minutiae is presented in Fig. 2. Minutia detection is a complex process and is thus beyond the purview of this contribution.

Each minutia, F_i , can be represented by a parameter vector $F = (x, y, \varphi, t)^T$ where (x, y) is the coordinate in the image, φ the local ridge direction and t the type of a minutia (i.e., bifurcation or ridge ending). The basic composition of the minutia template, S , of a fingerprint image is then the set of all n valid minutia parameter vectors found in the fingerprint image given by:

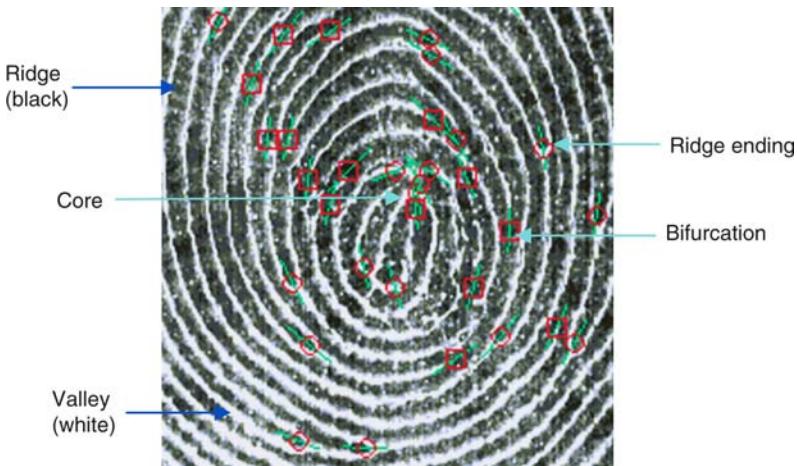
$$S = F_k = (x_k, y_k, \varphi_k, t_k), k = 1, 2, \dots, n \quad (1)$$

Apart from the minutia information, the ridge count [2] between two minutiae, which is the number of ridges intersecting a straight line joining two minutiae, is commonly used and included in the template. The non-minutia data commonly extracted and included in the template are the location, direction and the number of core and ▶ delta points. There are many other details that can be extracted and included in the template such as a short ridge line information associated to the minutia, and the number and type of minutia encountered by the straight line used in the ridge count with the aim to improve the performance of fingerprint matching.

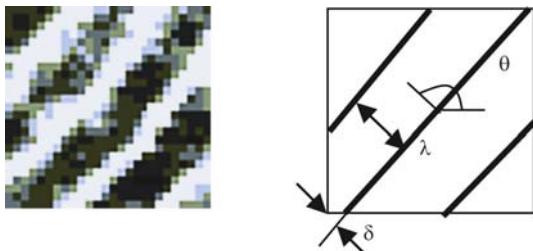
A popular pattern-based approach is the FingerCode [3] approach. The fingerprint image is tessellated



Fingerprint Templates. Figure 1 Process Flow of a Fingerprint Recognition System.



Fingerprint Templates. **Figure 2** A Fingerprint Image Showing Detected Ridge Endings and Bifurcations.



Fingerprint Templates. **Figure 3** A Sample Set of Spectral Triplets Representation (right) of a Fingerprint Image Block (left).

into sectors and bands with respect to a reference point, such as the ► core point. A bank of Gabor filters is then applied to each cell in the tessellation, bounded by the boundary delineated by the sector and band of the tessellation. For every cell, the average absolute deviation of each filter response over all the pixels in the cell is computed and used as an element of the feature vector, called the FingerCode. To approximate rotation invariant, rotate the FingerCode cyclically. Thus, the combined FingerCode becomes the template for this approach. Another proposed approach includes dividing the fingerprint image into small blocks. For each block, the spectral information that describes the fingerprint pattern in that block as closely as possible is obtained. This can be done using Discrete Fourier Transform, Gabor Filterbanks or by selecting the spectral component from a predefined set of spectral triplets [4] (see Fig. 3). The parameters describing the spectral component for each block, such as $(\theta, \lambda, \delta)$ for the spectral triplets, is quantized to limited discrete

values and then stored as a feature vector. The set of feature vectors for all the blocks in the region of interest of the fingerprint image is then stored as the fingerprint template.

Storage of Fingerprint Template

Usually fingerprint templates are stored in a central database residing in a central database server. To perform a match, the extracted features from the query fingerprint image are sent to the server. Such a model requires connection from the point where the query image is acquired to the central server. Since the fingerprint template is stored in a central server, it carries a notion of “big brother” which is a cause of concern for the privacy advocates as such a system is capable of tracking an individual. Another model for the storage of fingerprint template is the distributed database concept. The fingerprint templates are stored at each unit where the use of the fingerprint system is the most common. However, if a user wishes to be recognized in the other system, the fingerprint template has to be sent to the unit in advance, usually via a central server which acts as a backup and synchronization unit. Such a model is usually preferred to the central database model if the usage spans over a large geographical area. Instead of depending on a database, the fingerprint template can also be stored in a token such as a smartcard, memory stick or thumb drive in a fully decentralized model. A smartcard is usually preferred since it is generally regarded as more secure. The user carries the token containing his or her fingerprint

template with him/her. To perform fingerprint matching, the user has to present the token and the fingerprint template stored in the token is retrieved for matching with the query features. Alternatively, the matching is performed inside the token itself with the query features sent into the token. As such, the use of the fingerprint system is not dependent on any connectivity. Also, the number of users can easily be scaled when managing the template database. Unfortunately, once the token is lost, the genuine user is unable to use the system.

Template Synthesis

Various types of fingerprint sensors are available that capture live fingerprint images. Traditionally, those used for law enforcement purposes require a sensor with a sensing area of at least $2.54\text{cm} \times 2.54\text{cm}$. The consumer version is usually smaller, about a quarter of the size or even smaller. However, such a small sensor is not able to image the complete portion of the finger that touches the sensor. Consequently, if the user does not position his/her finger such that the contact portion of the skin is largely similar to the portion used during enrollment, then the matching will fail. This will result in false non-match, causing inconvenience to the user. To solve this, an image **mosaicking** technique has been proposed [5] for constructing a composite fingerprint from an image sequence of partial fingerprints. This is done by applying a low pass filter to smooth the images and then compressing the intensity to the range of [10, 20]. The images are then aligned using the Iterative Closest Point algorithm [6] before superimposing the aligned images to form the ► **mosaic**. Another **mosaicking** technique for rolled fingerprints has been proposed in [7]. Alternatively, a minutia-based template synthesis approach to combine the various fingerprint templates obtained from the small fingerprint images into a composite template which resembles the template obtained using a larger fingerprint image has been developed in [8].

Minutia-based template synthesis is performed by finding all the correspondent or matched minutiae between two fingerprint images, I_R and I_1 , to be synthesized. Based on the matched minutiae, an affine transformation that maps the minutiae from I_1 to I_R is then determined. This is repeated until all the other fingerprint images are synthesized. The experimental

comparison [9] revealed that the template synthesis approach is faster, less affected by elastic deformation, and is more suitable for larger partial images while the image **mosaicking** approach is more appropriate when accurate performance for small partial images is required.

Template Improvement

Fingerprint images are often corrupted by noise, imaging artifacts and affected by the skin condition (wet, dry), amount of pressure exerted when touching the sensor, etc. This causes the occurrence of missing minutiae (valid minutiae are not detected) or spurious minutiae (false minutiae detected). Thus, accurate detection of minutia is a very challenging task. If many dropped or spurious minutiae are present in the fingerprint template, the usability of the fingerprint recognition system is affected. To expect the user to re-enroll regularly may cause a lot of convenience. The purpose of template improvement is to improve the fingerprint template using multiple fingerprint images captured over a period of time [10]. For each minutia in the template, it is initialized with a default certainty level. When a query fingerprint submitted after a time interval is matched above a predefined threshold, the template and the certainty level associated with each minutia will be updated. This is done by finding those unmatched minutiae in the template within the region which overlaps with the query fingerprint and then reducing their certainty level by a predefined weight, α . Next, all the unmatched minutiae found in the query fingerprint outside of the overlapping region will be included in the template using the template synthesis technique but with a reduced certainty level of $(1 - \alpha)$. Then all minutiae in the template with a certainty level lower than a predefined threshold will be removed. In this way, spurious minutiae can be eliminated after many unmatched iterations while the missing minutiae can be incorporated.

Template Interchange

There are many ways in which a fingerprint system can generate a fingerprint template. In order to facilitate the interchanging of fingerprint templates among the

various fingerprint systems from different vendors, fingerprint templates have to be coded in a consistent manner. This is defined by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). The ISO/IEC documents specify the standards for the minutiae-based template [11], the pattern-based template [4], and the combined minutia and pattern template, called the pattern skeletal [12]. The standards specifying consistent formats for the construction of fingerprint templates comprise 3 sections:

1. A header which describes the generic information about the template and its source.
2. A normative section which describes all the mandatory features to be included and the manner in which they have to be coded.
3. A non-normative section which allows for other non-mandatory information to be included in the fingerprint template.

Image Reconstruction from Template

It is often assumed that the minutiae-based template cannot be used to construct back the corresponding original fingerprint, partly because the way the data is stored in the template is proprietary. However, it has been shown that this is possible with a standard template [13], such as those defined by the ISO/IEC [9]. The general idea is to reconstruct the orientation pattern using the orientation modeling approach [14] and the fingerprint area based on the template information. Subsequently, the ridge pattern is developed by applying a high gain Gabor filter adjusted to the local frequency and orientation and then rendering it to make the image look realistic [2].

Summary

A fingerprint template contains the unique features of a fingerprint image and can be used for fingerprint matching. The exact composition of the template is dependent on the algorithm used to extract the unique features. Nevertheless, international standards exist to facilitate the interchanging of the template. It can be stored in a database which can either be centrally

managed, distributed or stored in portable tokens instead of a database. Template synthesis and template improvement techniques can be used to improve the performance of the system when dealing with small fingerprint sensor and poorly enrolled fingerprint templates respectively.

Related Entries

- Enrolment
- Fingerprint matching
- Minutia

References

1. Jain, A.K., Hong, L., Pankanti, S., Bolle, R.: An Identity-authentication System Using Fingerprints. *Proceedings of the IEEE* **85**(9), 1365–1388 (1997)
2. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: *Handbook of Fingerprint Recognition*. Springer-Verlag, New York (2003)
3. Jain, K., Prabhakar, S., Hong, L., Pankanti, S.: Filterbank-based Fingerprint Matching. *IEEE Trans. Image Process.* **9**(5), 846–859 (2000)
4. Standards document ISO/IEC 19794–3: Biometric Data Interchange Formats – Part 3: Finger Pattern Spectral Data
5. Jain, A.K., Ross, A.: Fingerprint Mosaicking, In: *Proceedings of the International Conference on Acoustic, Speech Signal Processing*, vol. 4, pp. 4064–4067. Washington, DC (2002)
6. Besl, P.J., McKay, N.D.: A Method for Registration of 3D Shapes, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(2), 239–256 (1992)
7. Bolle, R.M., Ratha, N.K., Connell, J.H.: Image Mosaicking for Rolled Fingerprint Construction, In: *Proceedings of International Conference on Pattern Recognition*, vol. 2, pp. 1651–1653 (1998)
8. Yau, W.Y., Toh, K.A., Jiang, X.D., Chen, T.P., Lu, J.W.: On Fingerprint Template Synthesis, In: *Proceedings of the 6th International Conference on Control, Automation, Robotics and Vision 5–8* (2000)
9. Moon, Y.S., Yeung, H.W., Chan, K.C., Chan, S.O.: Template Synthesis and Image Mosaicking for Fingerprint Registration: An Experimental Study, In: *Proceedings of the IEEE Conference on Acoustics, Speech, Signal*, vol. 5, pp. V-409–V-412 (2004)
10. Jiang, X., Ser, W.: Online Fingerprint Template Improvement, *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(8), 1121–1126 (2002)
11. Standards document ISO/IEC 19794–2: Biometric Data Interchange Formats – Part 2: Finger Minutiae Data
12. Standards document ISO/IEC 19794–8: Biometric Data Interchange Formats – Part 8: Finger Pattern Skeletal Data

13. Cappelli, R., Lumini, A., Maio, D., Maltoni, D.: Fingerprint Image Reconstruction from Standard Templates, *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(9), 1489–1503 (2007)
14. Vizcaya, P., Gerhardt, L.: A Nonlinear Orientation Model for Global Description of Fingerprints, *Pattern Recognit.* **29**(7), 1221–1231 (1996)

Fingerprint Thinning

Obtaining a 1 pixel wide digital skeleton of ridges.

► Fingerprint Features

the use of Automatic Fingerprint Identification Systems (AFIS). The fourth process, forensic evaluation, is an expert-based process, built on procedure, training, and experience. The procedure and practice vary a lot between countries, principally regarding the threshold used for forensic identification. Most of the European and South American countries favor a quantitative approach based on a numerical standard when the USA, UK, and most of the Scandinavian countries have adopted a qualitative approach based on the experience and knowledge of the dactyloscopist. For both approaches, the decision is an expert opinion that is deterministic: exclusion, inconclusive or identification. As the current practice is not error-free and partly based on the subjective probabilities of the dactyloscopists, efforts are made to develop a new approach based on a logical inference model and statistical probabilities, in order to assist the dactyloscopists in producing a logical, testable, and quantitative evaluation of the fingerprint evidence.

Fingerprint, Forensic Evidence of

DIDIER MEUWLY

Netherlands Forensic Institute, The Hague,
The Netherlands

Synonyms

Fingermark identification procedure; Automatic finger-print identification system; Forensic evaluation of fingerprints and fingermarks.

Definition

Forensic evidence of fingerprint is the field of forensic expertise related to the inference of the identity of source from the examination of all the friction ridge skin, namely the fingers, the palms, the toes, the soles, and their marks. But for the sake of simplicity, the text is mainly focused on fingerprints and fingermarks. The extreme variability of the fingerprints derives firstly from the knowledge of the morphogenesis of the papillary ridges pertaining to embryology and, secondly, from statistical researches pertaining to dactyloscopy. This variability is mainly used in four different processes within forensic science: identity verification, forensic intelligence, forensic investigation, and forensic evaluation. The first three processes are based on

Nomenclature

At the end of the 19th century William Herschel and Henry Faulds expressed the principles of the forensic use of fingerprints and fingermarks: the use of fingerprints and fingerprint databases for the identification of serial offenders and the use of fingermarks to establish a link between a crime scene or an object and an individual. In literature, confusion exists between the term fingerprint and fingermark. This article uses a uniform terminology: the finger dermatoglyphics and their standard rolled inked impressions are named fingerprints, whereas recovered traces left by unprotected fingers are named fingermarks. In criminal records, reference prints are collected using forms named ten-print cards.

Individuality of the Fingerprint

Confusion surrounds the terms *identity*, *identify*, and *identification* in forensic science. This is clearly demonstrated in popular practice, when the perpetrator of an infringement is said to be “identified from her/his fingerprints”. The perpetrator is not identified, but individualized. What is proved by the fingerprints is individuality. To individualize a human being on the basis of fingermarks in forensic science ultimately consists in determining if an individual is the source of

the fingermark linked to the criminal activity [1]. The individuality of fingerprints derives firstly from the knowledge of the morphogenesis of the papillary ridges pertaining to embryology and, secondly, from statistical researches pertaining to dactyloscopy.

Morphogenesis

The friction ridge skin morphogenesis offers a biological basis to explain the variability in friction ridge patterns. The morphogenesis of the human hands and feet starts during the 6th week of the estimated gestational age (EGA). The pattern of ridge skin is established from the 10th week to the 14th week of EGA when the basal layer of the volar epidermis becomes folded and forms the primary ridges. This process is influenced by the volar pads, local eminences of subcutaneous tissue in well-defined locations of the volar surfaces. It is conjectured that the inversion of the volar pads creates tensions in the epidermis that align the ridge pattern [2]. From this moment on up to the 16th week of EGA, the tissues growing under the dermis, named volar pads, induce physical stress in the cell layers constituting this dermis. This physical stress forms a two-dimensional structure of ridges on the palms, the soles, the fingers tips, and the toes. From the 16th to the 24th week of EGA, the dermis matures; secondary dermal ridges start to develop between the primary dermal ridges and bridges, named dermis papillae, appear between the apex of the primary and secondary ridges. After 24 weeks of EGA, the development of the dermis is finalized and the epidermis is gradually formed by cell development from the dermis, named papillary ridges. In its final stage, the papillary ridges grow as a three-dimensional structure based on the two-dimensional pattern. The anchorage of this

epidermal structure in the dermis ensures the stability and the permanence of the dermatoglyphics. Therefore a permanent modification or destruction of the dermatoglyphics can only occur in case of destruction of the dermis [3].

Variability of the Fingerprint

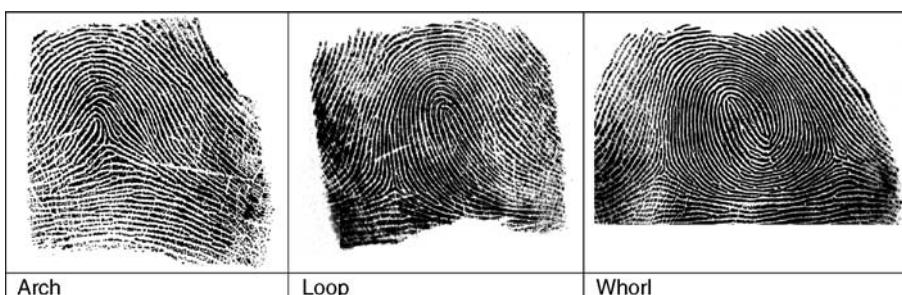
The fingerprint is expressed through the interaction of genotype, development, and environment; therefore this biometric modality is qualified as epigenetic, similar to the iris of the eye but contrarily to a DNA sequence, from which by instance a DNA profile is extracted, that is genetically determined. The information content in the fingerprint ridges is structured in three levels named the general pattern, the minutiae, and the third level details.

General Pattern

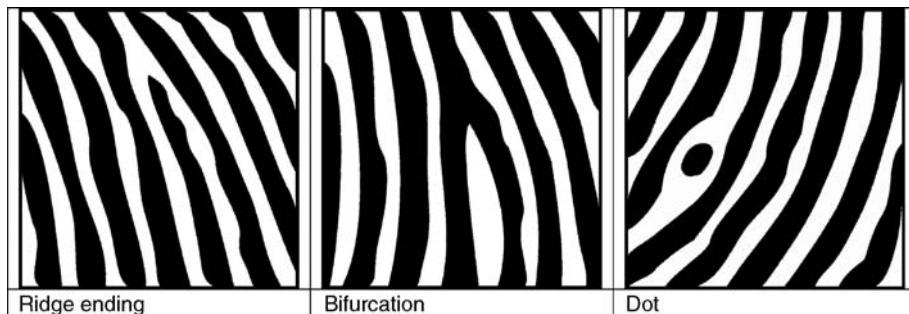
The general shape of the ridge flow, named general pattern, is to some degree indirectly genetically inherited and is classified in three generic types: arches (simple or tented), loops (left or right), and whorls (including various composite forms). The approximate center of the general pattern is named the core, and the small area where 3 flows of ridges meet to form a triangular pattern is called a delta. Arches have no delta, loops have 1 delta, and whorls, 2 deltas (Fig. 1).

Minutiae

In addition to the general ridge flow of the ridges, deviations appear along the papillary ridges. They are named minutiae, and can be classified in three basic types: ridge ending, bifurcation, and dot (Fig. 2). All other denominations, employed at the convenience of



Fingerprint, Forensic Evidence of. **Figure 1** Examples of fingerprint with different general patterns: arch, loop, and whorl.



Fingerprint, Forensic Evidence of. [Figure 2](#) The three basic structures of minutiae: the ridge ending, the bifurcation, and the dot.

the users or for statistical purposes, are a combination of two or three minutiae of basic type.

The minutiae contribute the most to the selectivity of the fingerprint, due to the combination of their spatial arrangement along the ridges and their intrinsic characteristics: type, location, and orientation. The selectivity offered by a minutiae configuration present on a fingerprint or on a fingermark is a function of their number, type, and topology (relative position and orientation on the ridges).

The process underlying the development of the minutiae is not known yet, but models offered by mathematical biology and empirical studies suggest that it is epigenetic [2]. For ridge endings, bifurcations, and dots, more correlations are observed on fingerprints of monozygotic twins as opposed to dizygotic twins [4]. Correlations are also observed between the number of minutiae and the finger number, which can be explained by the fact that the surface of the fingertip of the thumb is bigger than the surface of the fingertip of the little finger. The relative frequencies of the minutiae type are correlated with gender, but no difference has been observed between the fingerprint characteristics of the left and right hands [5].

Third Level Details

The study of the friction ridge details may be further subdivided into the description of ridge contours or edges, and the position and the shape of the pores [6]. However, the degree of agreement between dactyloscopists on the value of these latter characteristics is limited so far, and no systematic study supports the different opinions.

Statistical Research

The first statistical investigations were conducted at the end of the nineteenth and at the beginning of the twentieth century, but the initial models were developed on the basis of unrealistic premises: it was presumed that each minutiae type appeared with the same probability and independently of each other on the ridge skin surface. More sophisticated models were developed later during the twentieth century, first including the unbalance between the minutiae type (e.g., the bifurcations are more rare than the ridge endings) and then including the uneven density of the minutiae (e.g., the density of minutiae increases in the centre and delta zones) [7].

Statistical studies mainly focus on the second level features and especially the spatial arrangement of minutiae, while studies of other fingerprint features remain too seldom. These studies on minutiae provide extremely valuable fundamental knowledge about the degree of randomness of minutiae configurations, but they cannot be used yet for the deployment of large-scale, case-specific statistical evaluation of the fingermark evidence. Current statistical models simplify reality, emphasizing the statistical behavior of minutiae, and adopting a restricted view of the overall factors like the general pattern, the main ridge flows, the ridge edges, or the pores. Nevertheless, this new approach aims to offer a uniform framework and a transparent methodology to the dactyloscopists. Coupled to a logical inference model originating in the Bayes theorem, these models aim to assist them in producing a logical, testable, and quantitative evaluation of the fingerprint evidence based on statistical probabilities [8].

Classification of Fingerprints and Fingermarks

Manual Classification

For about a century the classification of fingerprints based on general patterns allowed the dactyloscopists to limit the search for the source of an unidentified fingermark to a specific section of their databases of fingerprint reference files. Francis Galton proposed the first system of fingerprint classification in 1891, and the development and practical application of dactyloscopy for forensic use were materialized in 1892 with the publication of his manual of dactyloscopy. This led to the acceptance of fingerprints in Great-Britain and the British Empire. In 1900, Henry modified the classification system of Galton, which remained the most widely used system in the world under the name of Galton-Henry. In 1891, Vucetich began to collect the first ten print cards databases based on the ideas of Francis Galton and developed another classification system, which was adopted by some South-American countries. The size of the ten print cards databases increased progressively during the twentieth century, and the workability was maintained sophisticating the indexation system, but to the cost of a trade-off between selectivity and reliability. The coexistence of several classification systems around the world limited the interoperability of the manual classification between different systems. In the second part of the twentieth century, manual classification was slowly abandoned and replaced by computerized classification systems named Automatic Fingerprint Identification Systems (AFIS) [9].

Automatic Classification

Development

From the mid-1960s, research on automation of fingerprint identification started. USA and Japan concentrated on automation of the high-volume ten-print workload, while France and the UK focused more on automation of fingermark identification. After a decade of effort, digitization of the ten-print card and automatic designation of minutiae were effective enough for the USA and the UK to produce automatic fingerprint reader systems. This advancement opened the possibility to digitize the ten print card

records and to store the standard impressions and the demographic data of individuals (e.g., name, citizenship, and date of birth) in a computerized database.

Forensic Uses of AFIS Technology

AFIS technology was initially developed to assist the dactyloscopists with computers in the identity verification process of individuals through their fingerprints. This process consists in searching the ten fingerprints of an individual in the database of standard impressions to verify if he or she is already present in the database and, if present, to check his or her demographic data. The AFIS technology has achieved enough maturity to ensure an identity verification process that is virtually error-free from the technological point of view, even if clerical mistakes in the database or in the running of the process can never be excluded.

In the 1990s the improvement of both AFIS and computer technologies allowed for the processing of fingermarks, exploited in two forensic processes. Fingermarks can be used for forensic investigation, in order to establish a link between a crime scene or an object and an individual. They can also be used for forensic intelligence to establish links between several crimes, even if the potential for links using marks depends on their limited quality.

In the 2000s the improvement of the computer mass-storage, in terms size and affordability, favored the constitution of large-scale palmprints databases. This development allowed for an extension of forensic investigation and forensic intelligence based on palmmarks. In most countries, the constitution of large scale palmprints databases is an ongoing process.

The challenge of standardization has only been solved recently, through the use of a common format, developed by the American National Institute for Standards and Technology (NIST), facilitating the computerized exchange of fingerprint and fingermark data between countries and agencies [10].

Individualization of Fingerprints and Fingermarks

History

The criminalist Edmond Locard enounced the first rule establishing a minimum number of minutiae

necessary for fingermark identification. During 1911–1912 he initiated the discussion of a numerical standard for the forensic identification of fingermarks, suggesting the following rule:

1. If more than 12 minutiae (“concurring points”) are present, and the fingermark is sharp, then the identity is certain. The imperative requirement for the absence of significant differences is implicit.
2. With 8–12 concurring points, the case is borderline and the certainty of the identity depends on
 - a. The sharpness of the fingermark.
 - b. The rarity of the type.
 - c. The presence of the core of the general pattern and the delta in the usable part of the mark.
 - d. The presence of pores.
 - e. The perfect and obvious similarity of the print and the mark regarding the width of the papillary ridges and valleys, the direction of the lines, and the angular value of the bifurcations.

In these instances, the certainty of the identification can only be established following a discussion of the case by at least two competent and experienced specialists.

3. With less than 8 minutiae, the fingermark cannot provide certainty for the identification, but only a presumption proportional to the number of minutiae available and their clarity.

Principally the first two parts of this rule were largely adopted by the community of the dactyloscopists but, unfortunately, the third part of the rule remained largely ignored [5].

Current Practice

The current dactyloscopic practice has evolved from the body of knowledge developed about the fingerprint individuality and the forensic use of fingermarks. It is formalized in a 4-step procedure named ACEV (Analysis-Comparison-Evaluation-Verification).

This procedure consists in the analysis of the fingermark followed by the analysis of the fingerprint, the comparison of the fingermark and the fingerprint, the evaluation and the decision based on the observed similarities and discrepancies between the fingermark and the fingerprint, and the verification of the findings by a second dactyloscopist.

Despite the formalization of the identification procedure, the practice varies between continents and countries, and even within some countries. The evaluation step, in particular, is based either on a quantitative threshold or on a qualitative threshold.

Quantitative Threshold: Presence of a Numerical Standard

A majority of European and South American countries favor a purely quantitative approach for forensic individualization, by fixing a numerical standard and considering qualitative aspects such as the third level details as secondary. A formal identification is established only if a minimal number of corresponding minutiae between the observed mark and the fingerprint – and an absence of significant differences – is put in evidence.

The numerical standard differs between countries and sometimes also between agencies in the same country: Italy (16-17); UK (before 2000) (16); Belgium, France, Israel, Greece, Poland, Portugal, Romania, Slovenia, Spain, Turkey, South American Countries (12); Netherlands (10-12); Germany (8-12); Switzerland (before 2008) (8-12); and Russia (7) [5].

Qualitative Threshold: Absence of Numerical Standard

Until 1970, the fingerprint identification procedure in the USA was also based on a numerical standard of 12 points, and below this threshold, qualitative factors in the comparison were taken into consideration. In 1970, a commission of experts from the International Association for Identification (IAI) was established to study the question of the relevancy of a fixed numerical standard for dactyloscopy. The following resolution was adopted by the IAI in 1973: “The International Association for Identification, based upon a 3-year study by its Standardization Committee, hereby states that no valid basis exists for requiring a predetermined minimum of friction ridge characteristics that must be present in two impressions in order to establish positive identification.”

It was accepted that the concept of identification could not be reduced to counting fingerprint minutiae, because each identification process represents a unique set of features available for comparison purposes; the identification value of concurring points between a fingerprint and a fingermark depends on a variety of conditions that automatically excludes any minimum standard.

In 1995, during a conference meeting on fingermark detection techniques and identification hosted in Ne'urim, Israel, 28 scientists active in the field of dactyloscopy, representing 11 countries, unanimously approved a resolution that is a slight variation of the IAI 1973 resolution. The Ne'urim declaration states that "no scientific basis exists for requiring that a predetermined minimum number of friction ridge features must be present in two impressions in order to establish a positive identification."

Decision Process

A formal identification is established when the dactyloscopists reach a decision threshold. They evaluate the contributions to individuality on a quantitative level (numerical standard), or on a qualitative level (absence of numerical standard), and the size of the relevant population of potential sources of the fingermark is set to its maximum, independently of the circumstances of the case [5].

On the basis of their evaluation, most dactyloscopists report three types of qualitative opinion: identification, exclusion, and inconclusive. As their evaluation is deterministic, they also make an implicit use of their own subjective probabilities of the rarity of the characteristics used to substantiate their opinion. They refine these subjective probabilities through training and experience, but they rarely consider results from research, particularly in the fields of embryology and statistics.

Admissibility of the Fingerprint in the USA

Like for other forensic disciplines, the scientific status of fingerprint identification has been questioned since 1993, when the Supreme Court of the USA handed down its ruling in *Daubert v. Merrell Dow Pharmaceuticals* (1993, Inc., 509 US, 579). Previously the main criterion for the admissibility of expert testimony in the federal courts of the USA was the Frye standard, which requires the general acceptance of the methods by the relevant scientific community. Daubert gave federal judges much greater discretion in deciding admissibility. It suggested that they consider (1) whether a theory or technique can be tested, (2) whether it has been subject to peer review, (3) whether standards exist for applying the technique, and (4) the technique's error rate. Although it is possible to test and validate methods for the forensic individualization of fingermarks, the research on this topic is still very limited.

The admissibility of fingerprint evidence, as being scientific in nature, has been subject to a Daubert hearing in the case U.S. v. Mitchell (1999, U.S. District Court for the Eastern District of Pennsylvania, Criminal), followed by Daubert hearings in more than 20 other fingermark cases. In the same case, U.S. v. Mitchell, the FBI provided calculations based on experiments carried out on an AFIS system. Random match probabilities of 10^{-97} and 10^{-27} were claimed respectively for complete fingerprints and partial fingermarks. These extraordinary numbers have been obtained by an extreme extrapolation of the probability density of the score using a postulated model, but they are so far from reality that it is surprising that they were admitted as evidence. Until January 2002, all Daubert hearings on fingermark cases led to the full admissibility of fingermark evidence in the courtroom. Judicial notice was given to the fact that fingerprints are permanent and unique [5].

January 2002 coincides with the first decision that proposes to limit expert testimony on fingerprint identification. Indeed in *U.S. v. Llera Plaza* (188F. Supp. 2d 549, 572–73 (E.D. Pa. 2002)), the defense "Motion to Preclude the United States from Introducing Latent Fingerprint Identification Evidence" has been partly successful. Judge Pollak held that a dactyloscopist could not give an opinion of identification, and required that the expert limits his testimony to outline the correspondences observed between the mark and the print, leaving to the court the assessment of the significance of these findings. That led the Government experts to ask for reconsideration bringing to the debate background documents in relation to the move of the UK toward the abandonment of the 16 point standard. Judge Pollak later reversed his opinion, and admitted the evidence.

Two cases of wrongful fingermark identification following the case of the Scottish police officer Shirley McKie perpetuated this controversy. In the first case the American Stephan Cowans was convicted by fingerprint identification, but later exonerated by DNA analysis. In the second case, the American Brandon Mayfield was wrongly associated with the 11 March 2003 Madrid bombing, by means of fingerprint to a latent mark revealed by the Spanish National Police on a plastic bag containing detonators recovered from a stolen van associated with these bombings. Three FBI experts and an independent court-appointed expert all identified Mayfield as the donor of the mark. Mayfield, a lawyer based in the US State of Oregon, came to the

FBI's attention when one of the latent marks sent by the Spanish authorities through Interpol gave a hit against his name on the FBI integrated AFIS (IAFIS), containing about 440 millions of fingerprints from 44 millions of persons. Brandon Mayfield was arrested, and remained in custody for a few weeks until the Spanish dactyloscopists, who immediately had raised issues with this identification, finally identified the mark with the finger of an Algerian suspect.

The FBI offered an apology and published a research report in the beginning of 2004 in which the existing FBI procedures were investigated extensively. This report showed that the mistake in this case was not owed to the methods the FBI used, but was the consequence of "human error" which cannot be excluded. The problem with this frequently used explanation is that the method and the human cannot be separated in case of an activity at which the human acts as a measuring instrument as is the case in traditional dactyloscopy [11].

An extensive research by the General Inspector of the US department of Justice appeared in January 2006 in which a clear analysis was given of the facts and circumstances causing the incorrect identification [12]. According to this report, an important factor in the Mayfield case was that when a search is performed using a very large database, there will always be a reference print which strongly looks like the unknown mark. A positive consequence of these cases is that they initiated a move towards a much more open discussion about the misidentifications in the forensic fingerprint field.

Analysis of the Current Practice

Research in embryology and statistics clearly do not legitimate the reduction of fingerprint individuality to counting minutiae. Indeed the scope of features is much broader than minutiae alone, and the nature of the papillary individuality prevents the adoption of any predefined number of ridge characteristics necessary for identification, without significant differences [13]. It is axiomatic that no two fingerprints are identical, as no two entities of any kind can be identical to each other. A common misconception lies in the fact that the features of individuality of the fingerprint is often attributed to the fingermark. As already described by Locard, in criminalistics, the transfer of material is logically never perfect. In dactyloscopy, the transfer of the pattern from the fingerprint ridges to the fingermark is accompanied by two types of loss

of information: quantitative, due to the limited size of the trace, and qualitative, due to distortion, blurring, bad resolution, and loss of pore and edge details.

The challenge for dactyloscopy is about the ability to quantify the information available for the individualization process in a partial distorted fingermark, and not to prove the individuality of the friction ridge skin. The first step in the quantification of the evidential value of fingermark evidence consists in estimating the similarity between the features of this fingermark and those of the fingerprint considered as potential source of this mark. The second step consists in estimating the typicality or the rarity of these features, and the third step, in reporting the similarity-typicality ratio as evidential value. This concept encapsulates a continuum of values for individualization of the fingermarks ranging from very high to very low, depending on the feature analyzed. Therefore, the forensic individualization process of fingermarks cannot be considered as a binary decision process, but has to be envisaged as a purely probabilistic assessment of the value of evidence, as it is for any type of evidence [14].

Probabilistic models, which are applicable to fingermark individualization [15], have been proposed and accepted by forensic scientists in other forensic areas – i.e., DNA, microtraces and speaker recognition [16]. The absence of extensive statistical analysis on fingerprint variability can be viewed as the main reason to prevent giving qualified opinions. Statistical data only support and comfort identification statements used by dactyloscopists but, according to Stoney, "we must realize that to reach absolute identification, or its probabilistic equivalent, through an objective process is not possible. Probabilities are objective when they can be tested and reproduced" [17].

Future Perspectives

The statistical studies applied to fingerprints and fingermark individualization provide valuable knowledge about the statistical behavior of various types of features, mainly the minutiae, and to a more limited extent, the pores, but they do not provide a robust tool to assess the probability associated with a given configuration of features for several reasons: none of the proposed models has been subjected to an extended empirical validation, and the assumptions about

the features used in these models have not been fully explored.

The research possibilities are huge, mainly in three different directions. The first is a refinement and an empirical validation of the model-based approaches developed in earlier studies [8]. The second is the development of data-driven approaches taking advantage of the capabilities of the current AFIS systems, embedding large fingerprint and fingermark databases, high computation capabilities, and sophisticated pattern recognition techniques. The third direction is to explore the morphogenesis process from the point of view of mathematical biology, with the aim to determine the contribution of the genetic, environmental, and the other factors, which influence the features defined in the three levels of information present in the fingerprint. These studies require the availability of large samples of fingermarks and fingerprints and a clear definition of the features used by the examiners to compare fingermarks with fingerprints.

Related Entries

- Automatic Fingerprint Matching
- Fingerprint Classification
- Fingerprint Databases and Evaluation
- Fingerprint Features
- Fingerprint Individuality
- Fingerprint Matching, Manual
- Individuality of Fingerprints
- Latent Fingerprint Experts

References

1. Meuwly, D.: Forensic individualization from biometric data. *Sci. Justice* **46**(4), 205–213 (2006)
2. Kuecken, M., Newell, A.C.: Fingerprint formation. *J. Theor. Biol.* **235**, 71–83 (2005)
3. Wertheim, K., Maceo, A.: The critical stage of friction ridge and pattern formation. *J. Forensic Ident.* **52**(1), 35–85 (2002)
4. Jain, A.K., Prabhakar, S., Pankanti, S.: On the similarity of Identical Twin Fingerprints. *Pattern Recognit.* **35**(12), 2653–2663 (2002)
5. Champod, C., et al.: Fingerprints and other Ridge Skin impressions. CRC press, London (2004)
6. Ashbaugh, D.R.: Qualitative-quantitative friction ridge analysis – An introduction to basic and advanced ridgeology. In: Geberth, V.J. (ed.) Practical Aspects in Criminal and Forensic Investigations. CRC Press, Boca Raton, FL (1999)

7. Stoney, D.A.: Measurement of fingerprint individuality. In: Lee, H.C., Gaenslen, R.E. (eds.) Advances in Fingerprint Technology, pp. 327–388. CRC Press, Boca Raton, FL (2001)
8. Neumann, C., et al.: Computation of likelihood ratios in finger-print identification for configurations of any number of minutiae. *J. Forensic Sci.* **52**(1), 54–64 (2007)
9. Berry, J., Stoney, D.A.: History and development of fingerprinting. In: Lee, H.C., Gaenslen, R.E. (eds.) Advances in Fingerprint Technology, pp. 1–40. CRC Press, Boca Raton, FL (2001)
10. McCabe, R. M. (ed.): Data format for the interchange of finger-print, facial, scar mark & tattoo (SMT) Information, American National Standard ANSI/NIST-ILT 1-2000, July (2000)
11. Fine, G.E.: A Review of the FBI's handling of the Brandon Mayfield case. 2006, Office of the Inspector General, U.S. Department of Justice
12. Office of the Inspector General, United States Department of Justice. A Review of the FBIs Handling of the Brandon Mayfield Case: Unclassified Executive Summary, Washington, DC (2006)
13. Champod, C.: Dactyloscopy: Standards of proof, In: Siegel, J. (ed.) Encyclopedia of Forensic Science. Academic, London. (2000)
14. Taroni, F., Champod, C., Margot, P.: Forerunners of Bayesianism in early forensic science. *Jurimetrics* **38**, 183–200 (1998)
15. Good, I.J.: Weight of evidence and the Bayesian likelihood ratio, In: Aitken, C.G.G. (ed.) Statistics and the Evaluation of Evidence for Forensic Scientists. Wiley, Chichester, UK (1995)
16. Aitken, C.G.G., Taroni, F.: Statistics and the evaluation of evidence for forensic scientists. Wiley, Chichester, UK (2004)
17. Stoney, D.A.: What made us ever think we could individualize using statistics. *J. Forensic Sci. Soc.* **31**(2), 197–199 (1991)

Fingerprint, Palmprint, Handprint and Soleprint Sensor

GEPPY PARZIALE
Cogent Systems, Inc., South Pasadena, CA, USA

Synonyms

Fingerprint device; Fingerprint sensor; Handprint sensor; Palmprint device; Palmprint sensor; Soleprint device; Soleprint sensor

Definition

A fingerprint or palmprint or handprint or soleprint sensor is a transducer that converts the ridge-valley structure of a person's hand or foot sole to an electrical signal. Generally, the sensor *reads* the difference of

pressure, temperature, light, electrical capacity or other kinds of energies are measured between the ridges and the valleys. Then, this difference is converted into an electrical digital signal that is encoded as an image representing the ridge–valley pattern. Different technologies can be applied to achieve this conversion and each of them brings advantages and disadvantages.

It is important to highlight that the output signal is a representation of the real-world ridge–valley pattern. Hence, if F is a ridge–valley pattern of a real-world finger tip and s is the transfer function of a device, the output signal is $F' = s(F)$ and $F' \neq F$.

Introduction

The similarity of the ridge–valley pattern of the epidermis present on finger tips, palms, and soles [1] allows to use the same physical principles for capturing fingerprints, palmprints and soleprints. The devices using these technologies can be grouped into a single family, known as ► **livescan furrow devices** or shortly, *livescan devices*.

The technological advancement of livescan devices has been mainly driven by the research done in the fingerprint recognition field more than the palmprint and soleprint modalities. The reason has to be found in a more convenient use of fingerprint devices, instead of the larger, heavier, and more power-consumer palmprint and soleprint ones. Moreover, fingerprint being the oldest biometric means used to identify people, large collection of data have always been available. This facilitated the development of algorithms for fingerprint recognition, pushing experts and scientists to focus mainly on this modality more than palmprints and soleprints.

Ink-on-Paper Method

The oldest approach to capture the furrow pattern is represented by the ink-on-paper method. Even if we cannot consider this as a real sensing technology, it is important to mention it here, since ink-on-paper is still widely used to collect palmprints, fingerprints, handprints, and soleprints. Moreover, it represents a strong obstacle for the advancement and the introduction of new capture technologies. The reason has to be found in the existence of very large databases collected using this method during the last ten decades. When a

new technology is introduced on the market, it must have a high degree of interoperability with the ink-on-paper method to ensure the continuity of the use of these databases, because a fingerprint or palmprint representation different than the legacy one would make the comparison very difficult. Thus, the representation of the ridge–valley pattern provided by the ink-on-paper method still represents the *model* that the modern technology tries to imitate.

The ink-on-paper capture approach consists in covering the ridge–valley pattern with black ink. Then, the print is obtained impressing the inked skin onto a white paper applying a small pressure. The resulting print is represented by a black mark for each ridge, while nothing is left in correspondence of each valley. The quantity of the ink applied on the skin and the pressure applied onto the paper during the impression are very important factors influencing the quality of the final result. In spite of other approaches, this technique does not suffer the skin condition problems (dry skin, wet skin, etc.), which are instead very difficult to overcome in the case of the other capture methods.

In some applications, the capture of fingerprints is performed rolling the finger onto the paper. This is done to acquire as much information as possible of the finger tip that can be used during an identification. The impression obtained with this approach is called ► **rolled-equivalent fingerprint**. Using dedicated image processing algorithms, rolled-equivalent fingerprints can also be obtained rolling the finger on the sensing surface of a sensor.

Collecting fingerprints, palmprints, handprints, and soleprints with the ink-on-paper method is still widely used, because it still represents the cheapest way to collect these biometric data. In Spain, the registration of all new born children is done applying ink on the baby soles of the feet and impressing them on a paper. In some Asian countries, inked fingerprints are used to register civilians during elections to avoid double voting. In Switzerland, Spain, Germany, and many other Countries, criminals are still registered by inking the tips of their fingers and their hands to collect the ten flat and rolled fingerprints and palms.

Once the fingerprints are collected on the paper, they can be digitalized using a flatbed scanner and then stored in digital format. This approach is still the most used by an Automated Fingerprint Identification System (AFIS) or an Automated Palmprint and Fingerprint Identification System (APFIS).

To improve the user convenience, especially in civilian applications, a special transparent oily substance is used in place of the black ink. In this way, the fingerprinted person does not need to wash her/his hands many times to remove the inconvenient residues of the black ink.

Nowadays, palmprints are becoming more and more popular in crime investigation, especially for latent comparison, since recent studies have demonstrated that more than 30% of the latent prints found on a crime scene belongs more likely to palms than fingers. Generally, the ink-on-paper palm capture consisted of inking the lower palm and the impress it on a paper. Nowadays, the reduction of the cost of the digital storage space allows to store larger quantity of data. Thus, the most modern approach consists of capturing the full handprint consisting of the three ► **palm segments** (lower, upper and writer palms).

Sensor Characteristics

Before describing the modern fingerprint sensor technologies, their main characteristics are highlighted in this article. These features define the application range in which the sensor can be used. For some applications, livescan devices have to pass very strict tests. The most famous and required certification is the *FBI fingerprint scanner certification*, covered in the Appendix F of the Criminal Justice Information Service (CJIS) Electronic Fingerprint Transmission Specification [2]. A list of FBI certified livescan devices is available at <http://www.fbi.gov/hq/cjis/iafis/cert.htm>.

The first important feature for a livescan device is the *Image Resolution*, which describes the ability of a sensor to distinguish, detect, and/or record physical details of the ridge–valley pattern. It represents the number of pixels in a unitary length and is expressed in *pixel-per-inch* or shortly, ppi. Typical image resolution values are 500 and 1,000 ppi. The first value is the most common and it is used in majority of the applications and products present in the market. The 1,000 ppi is mainly used for criminal investigation, especially for palmprints. The interest in analyzing the so-called third level details of a ridge–valley pattern is now pushing the manufacturers to introduce new devices with very high resolution (1,500–5,000 ppi).

Radiometric Resolution or *Image Depth* or *Dynamic Range* determines how finely a sensor can represent

or distinguish differences of intensity. It is usually expressed as a number of gray levels or bits, for example, 8 bits or 256 gray levels which is typical of finger-print image.

The *Modulation Transfer Function* (MTF) or *Spatial Frequency Response* is another important parameter. Spatial frequency is typically measured in cycles or line pairs per millimeter (*lp/mm*). The more extended the response, the finer the detail and the sharper the image. MTF is the contrast at a given spatial frequency *f* relative to contrast at low frequencies and it can be computed with the following Eq. (1):

$$MTF = 100\% \frac{C(f)}{C(0)}, \quad (1)$$

where $C(f) = (V_{max} - V_{min})/(V_{max} + V_{min})$ is the contrast at frequency *f*, and $C(0) = (V_W - V_B)/(V_W + V_B)$ is the low frequency contrast. V_B , V_W , V_{min} and V_{max} represent the luminance for black areas, the luminance for white areas, the minimum luminance for a pattern near spatial frequency *f* and the maximum luminance for a pattern near spatial frequency *f* respectively.

All the optical features of a sensor can be measured using special targets. To test the quality of a device, a manufacturer must purchase these targets and test the accuracy of all its optical features.

The *Geometric Image Accuracy* represents the absolute value of the difference $D = X - Y$, between the distance *X* measured between any two points on the target and the distance *Y* measured between those same two points on the output image. This is a very important parameter especially for devices having a very large capture area.

The *Grayscale Linearity* (GL) represents the capacity of a device to reproduce the gray level values correctly. A target with gradually varying grayscale levels is used for this scope. The grayscale levels on the output image are compared with the grayscale levels on the input target to measure the accuracy of the representation.

The *Signal-to-Noise Ratio* (SNR) is measured using another special target representing a grayscale level reference. This reference can be a white-colored and a black-colored target. An image is generated from these two targets and compared point by point with the reference.

The *Framerate* represents the number of frames a sensor can generate per time unit. It is measured in *frames/s* and it is a very important parameter when the

object (finger, palm or hand) movements are implied during a capture (sweep devices). To improve the final image quality, many sensors acquire more images of the same finger or palm during an acquisition. The captured images are then combined to produce the final image.

The *Shutter-speed* is the time that a detector needs to capture a single image.

Other important sensor characteristics that can change the application range of a device are the communication interface type (USB, Firewire, Ethernet, etc.), the sensor dimensions and weight, the Mean-Time-Before-Failure (MTBF), the self-powering capacity (if the power of the device comes from the communication interface) and obviously the price.

Optical Sensors

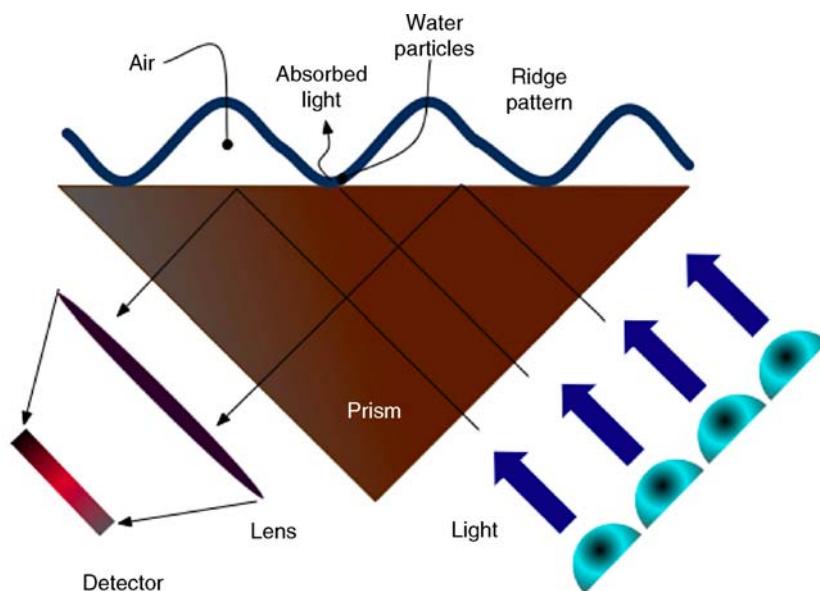
Sensors using light to discriminate between ridges and valleys represent the oldest technology to capture fingerprints, palmprints, and soleprints with no need of inking hands and feet.

The most widely used optical capture principle is known as Frustrated Total Internal Reflection (FTIR) and highlighted in Fig. 1. The sensor contains an optical prism and one of its faces is used as the ► **platen** that must be touched by the finger to produce an image. A monochromatic light source enters the prism and is reflected in accordance with each valley

of the skin. Then, it is collected by a detector (CMOS or CCD array). The light is absorbed in accordance with each ridge touching the platen. The lack of reflection allows the ridges (appearing dark in the image) to be discriminated by the valleys (appearing bright in the image). The 3D ridge-valley structure plays here an important role: presenting to the platen a picture or a drawing of a fingerprint does not produce any image. On the other hand, molding the shape of the ridge-valley pattern with special materials (latex, silicon, etc.) and touching the platen with it produces an image that cannot be distinguished by the image obtained by the real ridge-valley pattern (spoofing).

This capture technology is strongly influenced by the skin conditions. When the skin is too dry, the ridges do not completely adhere to the glass platen and thus, an image with very low contrast is obtained. On the other hand, very wet fingers produce an uniform black spot image, because during the finger pressure, the sweat accumulates in accordance with each valley and the light of the LEDs is fully absorbed. The result is a uniform dark spot with very low contrast between ridges and valleys. To overcome these problems, before each capture the user is asked to clean her/his hands and wet them with special non-toxic substance providing the right quantity of wet on the skin.

Another problem related to optical devices is represented by the so called ► **halo effect**. When the wet skin touches the colder platen, the moisture starts to



Fingerprint, Palmprint, Handprint and Soleprint Sensor. **Figure 1** Total frustrated internal reflection (► **TFIR**) principle.

condense. This results in a halo on the final image reducing the ridge–valley contrast. To avoid the halo effect, the platen has to be warmed up. This process is expensive in terms of current consumption especially for palmprint and soleprint devices with very large platen. The warming process requires a certain period of time and thus, these devices cannot be used immediately after they are switched on. This is sometimes impractical for some applications.

The physical size is another limitation of the optical sensors. The length of the optical path (the path traversed by the light from its source to the detector) cannot be significantly reduced without introducing severe optical distortions on the final image. The use of small mirrors and special lenses can help in keeping the same path length in a small space, but the manufacturing costs drastically increase and the robustness of the device decreases.

Nowadays, optical sensors represent the maturest technology in the market for capturing fingerprints, palms and soleprints. The large production of this kind of devices is reducing their price more and more. Since they are rugged and less sensitive to environmental factors than other technologies, optical sensors are spreading very fast and blocking the penetration of other capture technologies into the market.

Palmprint devices, ► [slap or four-four-two devices](#) and ► [Rolls Capture Devices](#) are only available based on this technology for the quality it can provide also in the case of devices with large platens. This is why palmprint sensors are only available based on optical technology.

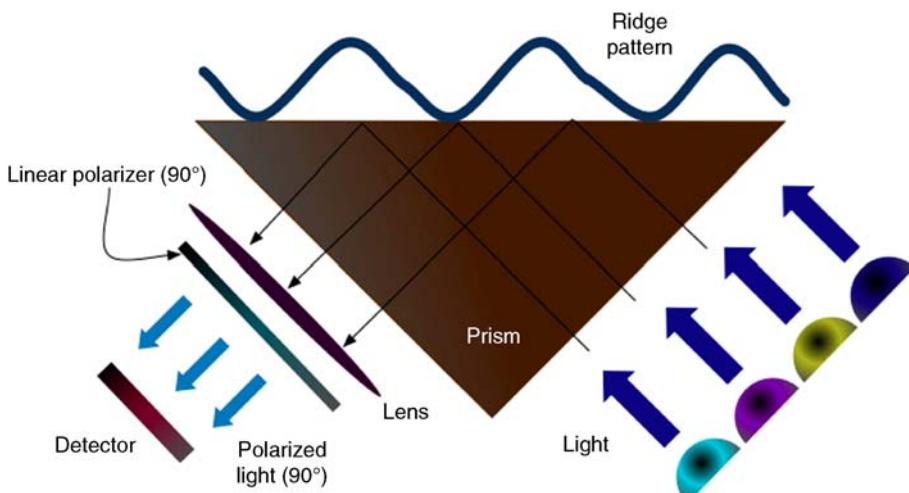
Optical Multispectral Sensors

To improve the ridge detail representation provided by the FTIR method, a novel approach to capture fingerprint has been recently proposed [3–5]. This approach is called Multispectral Imaging (Fig. 2) and uses multiple illumination wavelengths rather than a single monochromatic illumination commonly used in the FTIR approach. The orthogonal configuration of linear polarizers emphasizes this multispectral light, which penetrates the surface of the skin. The light then undergoes multiple scattering events before emerging from the skin toward the image array. In avoiding the optical phenomenon of the FTIR, the multispectral imaging sensor is capable of collecting more identifying data from the finger than the FTIR sensor. Currently, only fingerprint devices are available based on this technology.

Optical Multispectral Imaging is also claimed to be capable of detecting fake fingers obtained with organic or synthetic materials. The difference between the spectral characteristics of the skin and these materials is known and can be used to detect fake fingerprint.

Optical Contactless or Touchless Sensors

When a finger touches or rolls onto a surface, the elastic skin deforms. The quantity and direction of the pressure applied by the user, the skin conditions, and the projection of an irregular 3D object (the finger) onto a



Fingerprint, Palmprint, Handprint and Soleprint Sensor. **Figure 2** Multispectral imaging principle.

2D flat plane introduce distortions, noise and inconsistencies on the captured fingerprint image. To overcome these problems, a new approach to capture fingerprints has been proposed [6, 7], called touchless or ► **contactless fingerprinting**. Because of a lack of contact between the finger and any rigid surface, the skin does not deform during the capture and the repeatability of the measure is improved.

The approaches used to capture a fingerprint based on touchless technology can be grouped in two main families: ► **Reflection-based Touchless Finger Imaging (RTFI)** and ► **Transmission-based Touchless Finger Imaging (TTFI)**. Figure 3 highlights the two approaches. In the RTFI approach, the light generated by monochromatic light sources and reflected on the finger skin is collected by the detector. In the TTFI approach, the light penetrating the finger is collected by the detector positioned in front of the ridge-valley pattern.

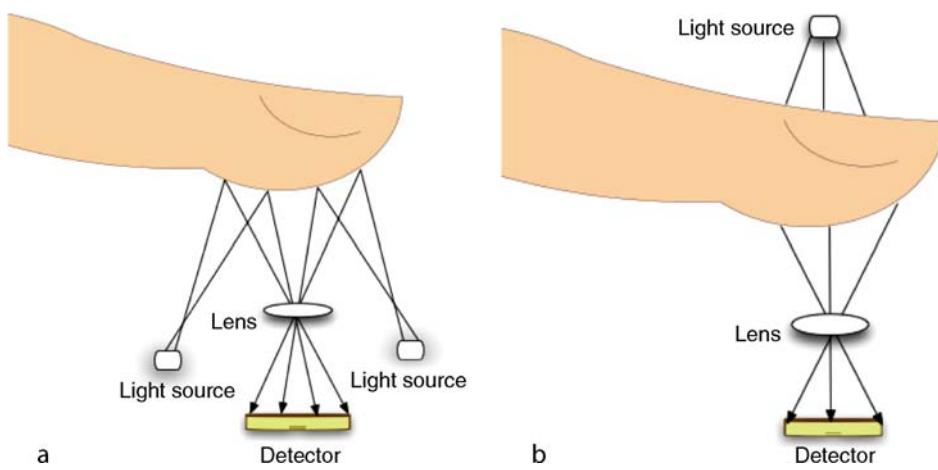
Since both the light reflecting on or penetrating the valleys and the light reflecting on or penetrating the ridges are collected by the detector, the final image has a contrast lower than that in the traditional FTIR technology. This has a huge impact on the minutiae extraction algorithm and thus, the advantage of a lack of skin deformation is negatively compensated by this low contrast. Moreover, the illumination not being perfectly perpendicular to the skin surface, shadowing effects of the ridges on the valley provide a wrong representation of small details (minutiae, pores, island, branches, etc.). Sophisticated

illumination techniques are required to avoid this representation problem and increase the final image contrast. The consequence is an increase of the size and final costs of these devices.

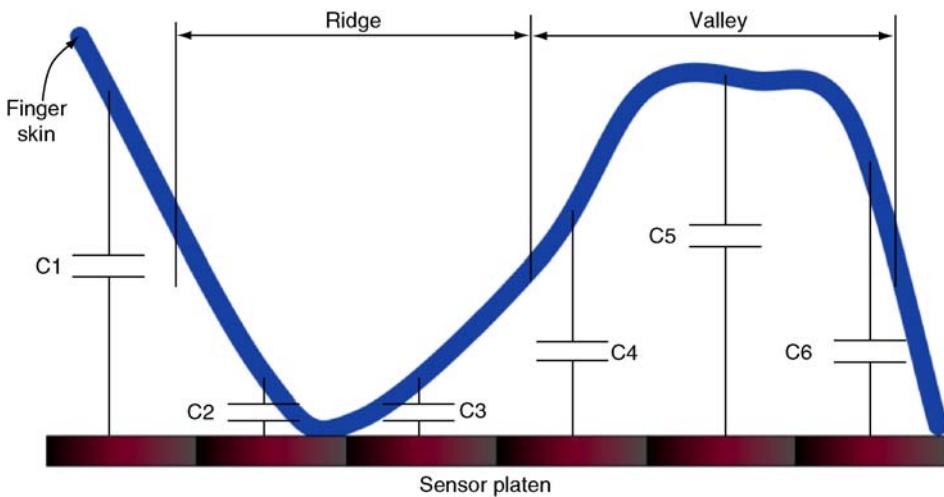
Another disadvantage of this technology is represented by the easy methods that can be used to attack these devices, which cannot be definitively used for high-security applications. In contrast to the FTIR case, where the ridge-valley 3D structure is important to generate an image, the touchless approach cannot discriminate between a 2D and a 3D pattern. Hence, presenting a photograph or a simple drawing of a fingerprint to the sensor, a new fingerprint image similar to the synthetic one is generated and the access is granted. Finger positioning, sensor usability, and user convenience must be still addressed.

Solid-State Sensors

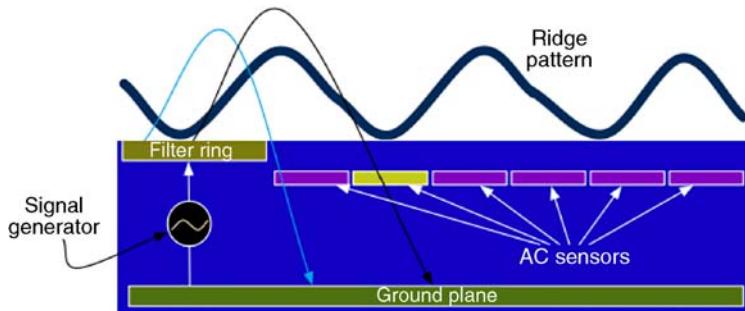
The first solid-state fingerprint capture device appeared on the market only in the middle of 1990s. It was a CMOS sensor capable of measuring the electrical capacity between the finger skin and the sensing surface (Fig. 4), which is composed by many squared pixels. Each pixel and the corresponding skin portion can be considered as an electrical capacitor with capacity $C = \epsilon A/d$, where A represents the pixel area and d the distance between the skin and the pixel and ϵ is the permittivity (a constant depending on the material) of the dielectric contained between the two capacitor



Fingerprint, Palmprint, Handprint and Soleprint Sensor. Figure 3 ► Touchless or contactless capture approach: (a) reflection-based touchless finger imaging; (b) transmission-based touchless finger imaging.



Fingerprint, Palmprint, Handprint and Soleprint Sensor. **Figure 4** Capacitive principle for the capture of the ridge–valley structure.



Fingerprint, Palmprint, Handprint and Soleprint Sensor. **Figure 5** Radio Frequency Field principle used to capture the ridge–valley structure.

plates. Each pixel produces a graylevel value proportional to its distance from the skin.

Another approach used to capture the ridge–valley pattern is based on the Radio Frequency (RF) electrical field (Fig. 5). A signal generator produces a low-level RF field traveling through the finger. The signal is then collected by AC sensors after being attenuated by the finger skin. The attenuation level of the signal is a function of the ridges and the valleys; the sensor array calculates the attenuation to synthesize the fingerprint structure. RF signal can be dynamically optimized in frequency and level to obtain the best possible image.

Using pyroelectric materials, it is possible to measure the difference of temperature between ridges and valleys. This approach is used while the finger is swiped on the small sensor surface (Fig. 6). This type of devices are called ▶ sweep sensors [8, 9]. The thermal

sensing elements detect temperature difference between valleys and ridges during the finger movement. This technology is claimed to overcome the skin condition issues of optical sensors. However, the resulting images are not rich in gray level values, i.e., dynamic range. Sweep sensors are very attractive because of their small size and low cost. This makes easier their integration in handheld and mobile devices.

The big advantage of the solid-state technology is represented by their smaller dimensions and lower costs with respect to the optical technology. Since they can be manufactured very thin and their power consumption needs are very low, solid-state sensors can be mounted on cards, handheld devices or laptops and used as logon means. This has an implication on the range of applications in which solid-state fingerprint sensors can be involved with respect to the



Fingerprint, Palmprint, Handprint and Soleprint Sensor. **Figure 6** An example of sweep sensor.

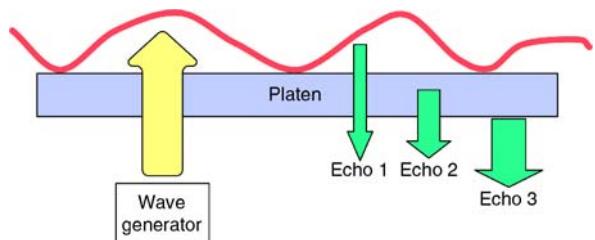
optical devices. However, external environmental factors (temperature, humidity, dust, etc.) are the major drawbacks of this technology. The sensing area is a chip completely open to the external world. Thus, special electrostatic protection methods must be used to avoid that external electrostatic charges destroy the chip surface. The same human skin can be the cause of the surface destruction, since the human body is usually electrically charged. Dust is another common vehicle of electrostatic charges that can quickly and easily degrade the sensing surface characteristics.

The use of solid-state sensor is mainly limited to fingerprint for their small sensing area. Palmprint and soleprint would require very large silicon areas that would make these sensors completely unaffordable in term of costs.

Even if their introduction on the market has been revolutionary and the expert envision new kinds of applications for fingerprint recognition (domotic, health-care, id-card and credit-card protection, etc.) their reduced lifetime and their high sensitivity to the external environmental factors limit the wide-spreading of these devices.

Ultrasonic Sensors

The ability to obtain images using ultrasound is based upon the reflection and transmission coefficients of



Fingerprint, Palmprint, Handprint and Soleprint Sensor. **Figure 7** Capture principle of an ultrasonic capture device.

ultrasound as it propagates through media of varying acoustic impedance. What makes sound waves valuable for the imaging of the ridge–valley pattern is that they can both reflect and pass through objects. The characteristics of sound waves make it possible for high-frequencies to pass through substances and accurately measure the ridges and valleys of a fingerprint even if in presence of dirt, grease, ink, moisture, dye, or other substances routinely found on fingers.

The capture principle of a ultrasonic device is highlighted in Fig. 7. An ultrasonic wave generator produces high-frequency sound impulses. These impulses reflect on each material found on their path producing echos. The strength of each echo depends on the material and the shape of the object on which they were generated. Special receptors are used to translate the echos in an electrical signal.

Livescan imaging the fingerprints of children 5 years and younger is a technically challenging task, since the ridge structure is usually very fine and contains high “spatial frequencies,” meaning that the ridges very close together. The spatial frequency of the fingerprint directly determines the resolution that the imaging device needs to accurately image the finger. Most live-scan fingerprint scanners have been designed to image adult fingers where a high-resolution scan is unnecessary. High-resolution ultrasonics is the only technology that can reliably and repeatedly capture clear and useful images of a young child’s fingerprint.

Next Generation

Although most of the technologies mentioned earlier are quite new (some of them are still in the prototyping phase), the research and the development continues to

bring new ideas to this field. The study of the physiology and the formation of the furrow pattern allowed to propose new fingerprint and palmprint capture approaches. It is important to mention here the Optical Coherence Tomography (OCT) which is an interferometric, noninvasive, optical tomographic imaging technique offering millimeter penetration (approximately 2–3 mm in tissue) with micrometer-scale axial and lateral resolution. OCT is like an optical version of ultrasound imaging. The technique is already routinely used in medicine, but has not had a forensic application until now. The technique provides a transparent 3D structural picture by sending light through the pattern of natural secretions left on a surface by a finger and combining the reflected beam with a “reference beam” produced by bouncing light from a laser off a mirror. This produces an interference pattern at a photodetector the same as those found in a digital camera which can then be used to reconstruct an image of the original fingerprint.

This technology together with multispectral and touchless imaging must be still further developed to demonstrate their superiority with respect to the FTIR approach that still remains the most used method to capture fingerprints and palmprints.

Summary

Livescan furrow sensors represent a family of devices used to capture fingerprints, palmprints, handprints, and soleprints. The same anatomical characteristics of the skin present on finger tips, palms, and soles allow the use of the same technology for the capture of these biometric traits.

The ink-on-paper method is first method used to capture fingerprints and palmprints. Optical devices try to overcome the inconvenience of the ink on the skin and provide a good alternative method to the legacy ink-on-paper. Solid-state sensors are very attractive for their very small size and reduced costs, but they can only be used to capture fingerprints. Moreover, environmental factors limit the life time of these devices.

Multispectral and touchless imaging technologies try to overcome the limitation of the optical devices, but their relatively higher costs and very low interoperability with legacy technology limit their wide-spreading.

Related Entries

- Biometric Recognition
- Biometric Sensor and Device, Overview
- Fingerprint Recognition, Overview
- Fingerprint Verification

References

1. Ashbaugh, D.R.: Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology. CRC, Boca Raton (1999)
2. Criminal Justice Information Services Division: Electronic Fingerprint Transmission Specification. Department of Justice (1999)
3. Rowe, R.K., Nixon, K.A.: Fingerprint enhancement using a multispectral sensor. In: Proceedings of SPIE Conference on Biometric Technology for Human Identification, vol. 5779. pp. 81–93. Orlando, USA (2005)
4. Rowe, R.K., Corcoran, S.P., Nixon, K.A., Ostrom, R.E.: Multispectral Imaging for Biometrics. In: Proceedings of SPIE Conference on Spectral Imaging: Instrumentation, Applications, and Analysis, vol. 5694, pp. 90–99. Orlando, USA (2005)
5. Rowe R.K., Nixon, K.A., Butler, P.W.: Advances in Biometrics. Sensors, Algorithms and Systems, chap. Multispectral Fingerprint Image Acquisition. Springer, Berlin (2008)
6. Parziale, G.: Advances in Biometrics. Sensors, Algorithms and Systems, chap. Touchless Fingerprint Technology. Springer, Berlin, New York, USA (2008)
7. Parziale, G., Diaz-Santana, E.: The surround imager: a multi-camera touchless devide to acquire 3D rolled-equivalent fingerprints. In: Proceedings of IAPR International Conference on Biometrics (ICB), pp. 244–250. Hong Kong, China (2006)
8. Parziale, G., Bishof, H.: Image reconstruction and on-the-fly minutiae extraction of fingerprints acquired with sweep sensors. In: Proc. of 28th Workshop of the Austrian Association of Pattern Recognition, pp. 241–248. Austria (2004)
9. Clausen, S.: Advances in Biometrics Sensors, Algorithms and Systems, chap. A single-line AC capacitive fingerprint swipe sensor. Springer, New York, USA (2008)

Fingerprints Hashing

JEAN-FRANÇOIS MAINGUET
Grenoble, France

Synonyms

Biometric encryption; Cancelable biometrics; Fingerprint encryption; Fuzzy extractor; Fuzzy vault; Intricated biometrics

Definition

Fingerprint hashing is merging fingerprint recognition and cryptographic methods. The aim is to perform a recognition using fingerprint while, at the same time, hiding the private information related to the fingerprint, thus enabling public fingerprint templates.

Introduction

Keeping a database in a safe place is not easy. Even with good encryption methods and special care, databases containing sensitive information, such as bank account numbers, are vulnerable to being compromised. Nobody wants something like that to happen when dealing with fingerprint identification.

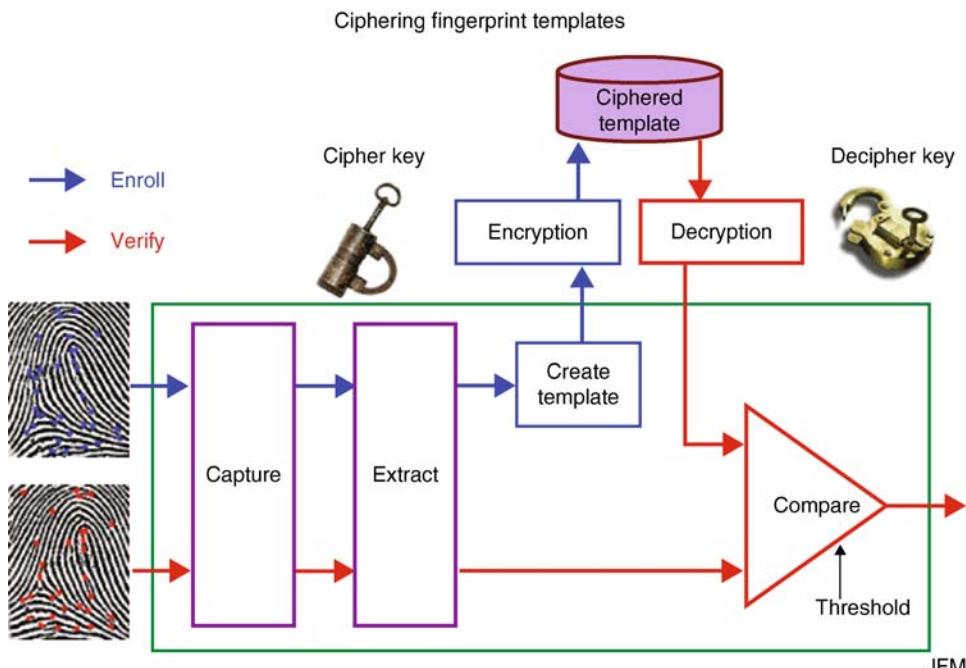
Security of a fingerprint-based system can be divided into two main areas:

1. The electronic security, which poses the question: “Is the electronic system, at the other end of the wires, a real trustful authorized fingerprint system?”
2. The liveness security, which asks a different question: “Is the object touching the sensor a real finger, alive and connected to a living person?”

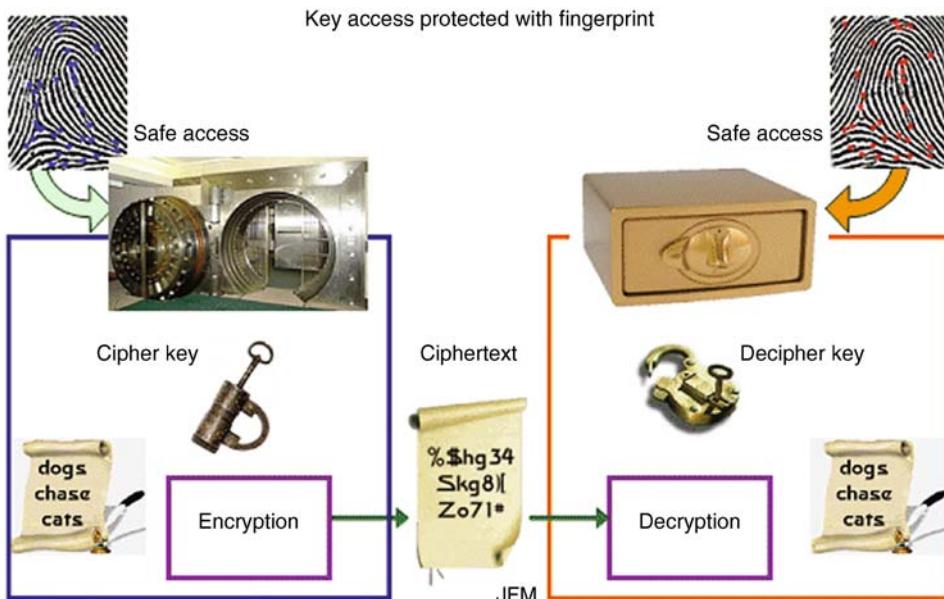
Liveness security is not addressed in this essay. Fingerprint hashing is part of the electronic security solution and deals with encryption.

Any biometric system requires the storage of a template (or reference). For fingerprint systems, the most common method consists in storing the minutiae. This information is considered as private information and should be protected and ciphered, not only for privacy reasons, but also against template replacement. This prevents a hacker from replacing the owner's minutiae, or using reverse engineering to get the minutiae locations and create a fake fingerprint. Even if biometric data cannot be considered as secret (they are public information in its cryptographic sense, always hiding ones face or voice is impossible), it is important to protect biometric data and the additional data that goes with them (name, bank account or whatever).

The template storage problem is generally solved using encryption. A template database can be created and protected using a single key pair (Fig. 1). Everything is fine up to the date when the key is compromised or badly protected, allowing hackers to access the data. However, in the case of fingerprints, this is ones very private information that is stored, and so it seems desirable to have an even better security scheme.



Fingerprints Hashing. Figure 1 Protecting fingerprint templates: Key is not protected.



Fingerprints Hashing. **Figure 2** Protecting access to cryptographic secret keys using a fingerprint system: Fingerprint template is not protected.

Another problem occurs when one tries to create a system enabling encryption/decryption features (like

► [Pretty Good Privacy \(PGP\)](#)). One needs to protect the access to the encryption/decryption module, which is done using a password. If one tries to replace the password with a fingerprint (Fig. 2), then one faces the problem of protecting the template, and cannot use the encryption/decryption scheme, because it is not yet enabled! It is the same problem as “you cannot put the key of the safe inside the safe itself.” One needs another safe.

There is also an additional problem from a security point of view. The result of matching is only one bit of information that is easy to find and hack (too low entropy). It would be better to eliminate this weakness.

Desirable Features, Definitions

A better fingerprint system includes:

1. The storage of the template (minutiae) in a non-reversible way. It is still possible to perform a match, but it is impossible to recover the original minutiae and impossible to derive the secret key.
2. It is possible to revoke (cancel) a template. If a template is not to be used anymore, it is possible to forbid its use and create a new one.

3. There is no step with a single yes/no bit corresponding to the match/no match result.

Properties #1 and #2 are generally linked, because it would be very impractical and dangerous to use a transform that is unique. Each individual would have a unique number ID for his or her whole life, impossible to change.

Fingerprint hashing is the use of a non-reversible transform (similar to a hash function) over a fingerprint. It is also called “cancellable biometrics,” because it is possible to cancel or to revoke the template. Fingerprint hashing involves using some kind of cryptographic scheme, similar to a hash function, but it is not a hash function.

Property #3 requires a stronger merge between biometrics and cryptography. Having all the properties at the same time is pretty hard to achieve and to prove, but has been originally proposed under the name of “Biometric Encryption” [1, 2]. Unfortunately, “biometric encryption” can be a simple combination of a biometric template and a simple encryption scheme. But it is much more; it is a real merge. In quantum cryptography, the word “► **intricated**” is used to designate the non-separable nature of some properties in quantum mechanics, and so “Intricated Biometrics” seems a better designation.

Cancelable Biometrics

Fingerprint hashing seems pretty close to password protection. A password is protected using a hash function, which is basically a method to transform some data into a relatively small number, the hash value, sometimes called fingerprint (which causes confusion), because of its uniqueness property (no collision should occur). A hash function is not reversible, and in most cases, some original data is lost as the resulting hash value is much shorter. This works well for password storage. You just need to apply the same hash function to the proposed password and perform a bit-to-bit comparison for checking. It is not useful to regain access to the original password.

Unfortunately, this scheme cannot apply to a fingerprint, because you never enter exactly the same fingerprint image. Every acquisition is different, and usual hash functions will return a different value, forbidding a further comparison. The problem is much more complex as it needs to be accepted that there will be some variability of the data. Fingerprint hashing must use a non-reversible transform like a hash function, but the comparison stops here. The other properties of a hash function, such as fixed length and uniqueness, are not required, but there needs to be a comparison, a match at the end, as depicted in Fig. 3.

General concepts related to cancelable biometrics have been discussed by Ratha et al. [3]. Davida et al. [4] added data to create a non-reversible template. Linnartz and Tuyls [5] proposed the use of specific shield functions before a hash function.

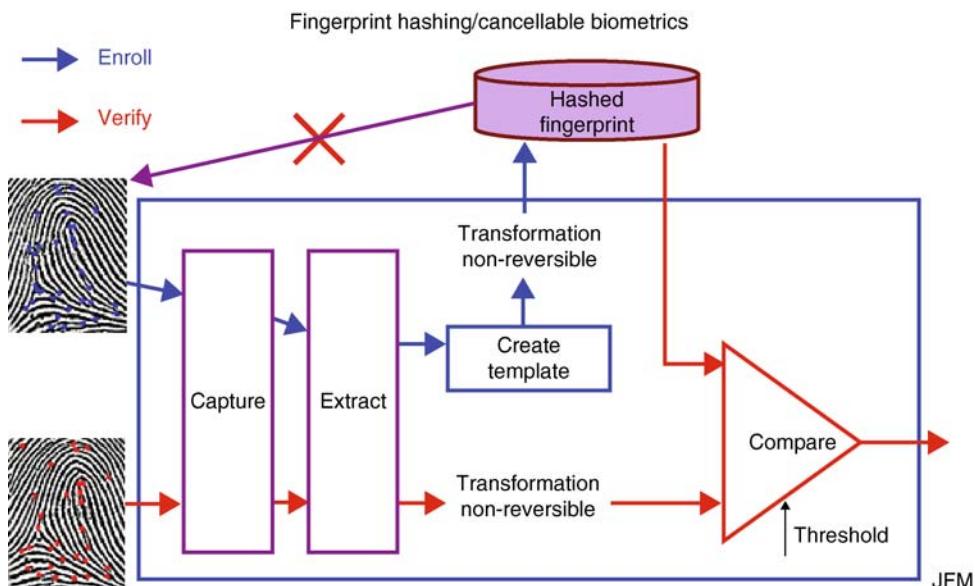
Lumini and Nanni [6] proposed the BioHash, a combination of a hash code, a Gram–Schmidt normalization and using a Hamming distance for comparison. This has been tested using the FVC-2002 fingerprint database.

Boult et al. [7] proposed another scheme called BioToken, and also tested on the FVC-2002 and 2004 fingerprint databases, which showed some enhancements of the accuracy of the system.

As usual in cryptography, proving that the transform is non-reversible or reversible with an extremely long computation time is very hard to achieve. Although there are some good reasons to believe that some solutions exhibit the right properties, nothing is mathematically proven yet. It took a long time for the security of regular cryptographic schemes to be accepted, and biometrics is in a similar situation, still in its infancy.

Intricated Biometrics

Cancelable biometrics shows interesting features, but still shows the potential weakness of the

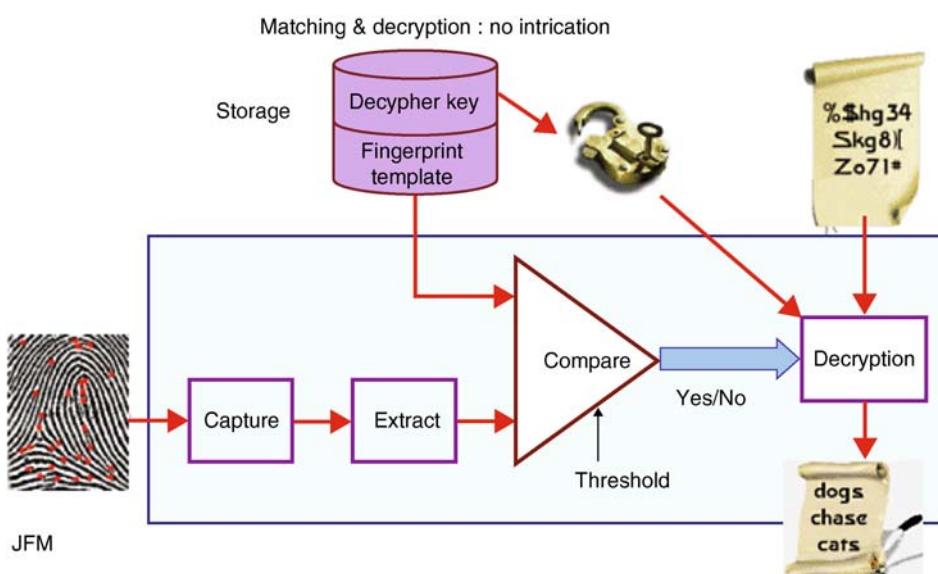


Fingerprints Hashing. **Figure 3** Fingerprints hashing/cancellable biometrics: It is not possible to extract the original biometric data from the template as the transform is non-reversible.

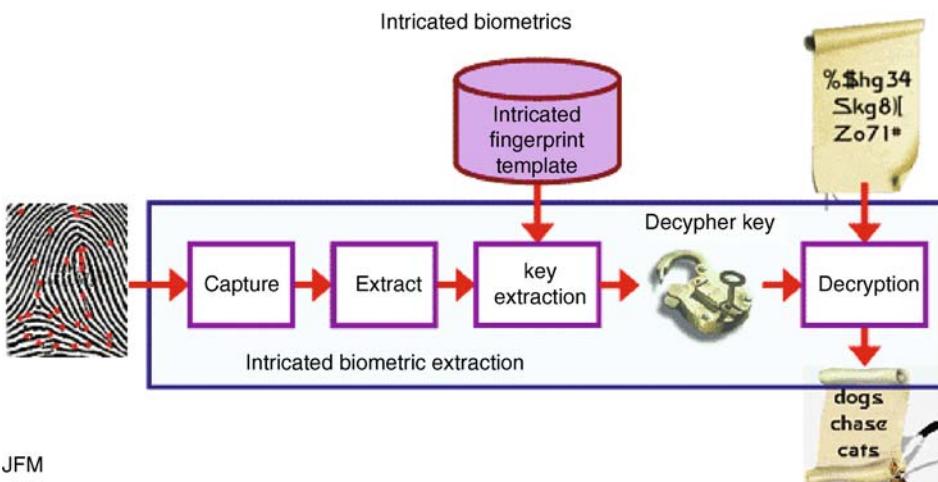
match/no-match bit. This type of weakness may not be as critical when speaking of a physical or logical access, because of the need of a go/no go answer. But in most cases, the aim of the biometrically enabled system is to provide a service, and secure systems always use somewhere a cryptographic key when a transmission is involved in a non-secured environment.

A simple scheme uses the result of the match to enable the decipher key, as shown in Fig. 4. The

template and the key are not protected. This requires external means; another secret key and method. There is still the one bit match/no match result. Intricated biometrics proposes to merge the decipher key with the template, so that neither the biometric template (the minutiae) nor the decipher key can be obtained from the stored template alone; they are intricated (Fig. 5). The intricated biometric template can be stored anywhere, even in a non-secured area.



Fingerprints Hashing. Figure 4 A simple use of biometrics to use a secret key: fingerprint and secret key are not yet protected, and the 1-bit match/no match still exists.



Fingerprints Hashing. Figure 5 Intricated biometrics: It is not possible to get the secret key and the original fingerprint data from the intricated fingerprint template, and the extracted key appears only for a short while for deciphering. The 1-bit match/no match step is eliminated.

When the decipher key is to be used, the live fingerprint can be scanned. If the extracted minutiae corresponds to the stored template, then the right decipher key will be regenerated, immediately used to decipher the message (this is the service) and destroyed (the key never leaves the secure area). If the extracted minutiae are not the genuine minutiae, a key is still generated but not the correct key. The message is then incorrectly deciphered, giving a meaningless result. At the end, there is no information revealed, which is a very good property of a secure system, and it is not possible to apply a scheme such as hill-climbing, based on access of the matching score.

It is possible to reach these objectives, but it is hard to achieve and to prove, especially for fingerprints.

Cryptography is Accurate; Biometrics is Fuzzy

In cancelable biometrics, a function similar to a hash function had to be applied, but the data variability was a problem. Intricated biometric involves re-generating a cryptographic key and the same problem. Every bit must be correct; no error is allowed. With biometrics, there is always some uncertainty. Each time a fingerprint is scanned or applied to a sensor, it may not be exactly the same area. The person may have a new cut or scar; the finger could be dirty, wet, or dry.

A partial solution would be to extract a stable sequence from a fingerprint image, always the same, and then combine it with a cryptographic key. This is like extracting a stable signal from a noisy, fuzzy environment. Some research proposed the use of error-correcting code, with the so-called “fuzzy extractor” [8, 9] that can be applied to different biometric modalities, and then specifically over fingerprint databases [10, 11].

The “► **fuzzy vault**” was proposed in 2002 by Juels et al. [12]. The proposal involved secret being merged with biometric data such as minutiae that does not need to be in a specific order. ► **Chaff points** [13] are added to hide the genuine minutiae. The experiment was later enhanced using lattice [14], tested on the FVC-2002 database and enhanced with helper data by Uludag et al. [15, 16].

Soutar et al. [2] proposed in 1999 using filters to extract stable characteristics of the fingerprint and then merged them with a secret.

One example scheme is:

- Enroll
 - A set of M minutiae is extracted from a fingerprint.
 - A secret key is divided into M pieces of data; each piece is linked to one minutiae.
 - Random chaff points are added, corresponding to non-existing minutiae and wrong pieces of secret key.
- Recognition
 - A live set of N minutiae are extracted from a live fingerprint.
 - The matching minutiae enable extraction of the correct piece of the secret key.

As the live minutiae may not be exactly the same, it is important to introduce some kind of redundancy for the secret key. A subset of the M enrolled minutiae is needed to perform a match. Lagrange interpolation has been proposed to recover the full secret key, with the advantage of not depending on the order of the points or minutiae.

But some problems arise:

- *Brute force attack*: It is important to add enough chaff points to hide the genuine points and to create too many possible combinations for a brute force attack to succeed.
 - Generating chaff points is not a simple operation, because care must be taken to avoid flaws. The chaff points must be indistinguishable from genuine points. It is a similar problem to random number generators, where it is difficult to prove that they are really random. Always using the same chaff points would make it too easy to find them.
 - Matching minutiae for key extraction will likely require more computation.
 - Chaff points may lead to wrong alignments, especially with poor fingerprints, making minutiae matching less robust.
 - Intricated fingerprint template requires more memory space than a simple template (but another key and program would be needed for protection).

Conclusion

Fingerprint hashing (intricated biometrics) seems to be the ultimate protection scheme. This is not a proven technology yet, but achieving the objectives would lead

to a better protection of privacy without worrying about databases.

Related Entries

- [Encryption, Biometric](#)
- [Fake Finger Detection](#)
- [Fingerprint Features](#)
- [Fingerprint Matching, Automatic](#)
- [Fingerprint Templates](#)

References

1. Cavoukian, A., Stoianov, A.: Biometric encryption: A positive-sum technology that achieves strong authentication, security and privacy. White paper, Information and privacy commissioner of Ontario, March (2007)
2. Soutar, C., Roberge, D., Stoianov, A., Gilroy, R., Vijaya Kumar, B.V.K.: Biometric Encryption, chap. 22, McGraw-Hill (1999)
3. Ratha, N.K., Connell, J.H., Bolle, R.M.: Enhancing security and privacy in biometrics-based authentication systems. *IBM Syst. J.* **40**(3), 614–634 (2001)
4. Davida, G.I., Frankel, Y., Matt, B.J., Peralta, R.: On the relation of error correction and cryptography to an off-line biometric based identification scheme. In: Proceedings of the Workshop on Coding and Cryptography, Paris, France, pp. 129–138 (1999)
5. Linnartz, J.P., Tuyls, P.: New shielding functions to enhance privacy and prevent misuse of biometric templates. In: Proceedings of the Fourth International Conference on Audio and Video based Biometric Person Authentication, Guildford, UK, pp. 393–402 (2003)
6. Lumini, A., Nanni, L.: An improved biohashing for human authentication. *Pattern Recognit.* **40**, 1057–1065 (2007)
7. Boult, T.E., Scheirer, W.J., Woodworth, R.: Revocable Fingerprint Bitokens: Accuracy and Security Analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07), Minneapolis, USA, pp. 1–8, 17–22 June (2007)
8. Dodis, Y., Reyzin, L., Smith, A.: Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In: Proceedings of the Eurocrypt 2004, pp. 523–540 (2004)
9. Burnett, A., Byrne, F., Dowling, T., Dury, A.: A biometric identity based signature scheme. In: Proceedings of the Applied Cryptography and Network Security Conference, New York, USA (2005)
10. Costanzo, C.R.: Biometric cryptography: Key generation using feature and parametric aggregation. Online techreport, School of Engineering and Applied Sciences, Department of Computer Science, The George Washington University, October (2004)
11. Al-Tarawneh, M.S., Khor, L.C., Woo, W.L., Dlay, S.S.: Crypto key generation using contour graph algorithm. In: Proceedings of the 24th IASTED International Multi-Conference Signal Processing, Pattern Recognition and Applications, Innsbruck, Austria, February (2005)
12. Juels, A., Sudan, M.: A fuzzy vault scheme. In: Lapidot, A., Teletar, E. (eds.) *Proceedings of the IEEE International Symposium on Information Theory*, p. 408. IEEE Press (2002)
13. Chang, E.-C., Li, Q.: Hiding secret points amidst Chaff. In: *Proceedings of the Eurocrypt*, Saint Petersburg, Russia (2006)
14. Zheng, G., Li, W., Zhan, C.: Cryptographic key generation from biometric data using lattice mapping. In: *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, Washington, DC, USA, pp. 513–516. IEEE Computer Society (2006)
15. Uludag, U., Jain, A.K.: Fuzzy fingerprint vault. In: *Proceedings on Workshop: Biometrics: Challenges Arising from Theory to Practice*, August 2004, pp. 13–16 (2004)
16. Uludag, U., Jain, A.: Securing fingerprint template: Fuzzy vault with helper data. In: *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, June 2006, pp. 163–170 (2006)

First Level Detail

This reflects the general flow of the papillary ridges which may form certain patterns such as arches, loops, whorls, and deltas.

- [Fingerprint Matching, Manual](#)

Fisher Criterion

Fisher criterion is a discriminant criterion function that was first presented by Fisher in 1936. It is defined by the ratio of the between-class scatter to the within-class scatter. By maximizing this criterion, one can obtain an optimal discriminant projection axis. After the sample being projected on to this projection axis, the within-class scatter is minimized and the between-class scatter is maximized.

- [Non-linear Techniques for Dimension Reduction](#)

Fixed Pattern Noise

It is characterized by the same pattern of “hot” pixels occurring with images taken under the same conditions of temperature and exposure.

- [Face Device](#)

Focal Distance

The distance that is required between the iris acquisition device and the iris, for the system to be able to acquire and accurately recognize.

- ▶ Iris Acquisition Device

Focal Length

With respect to a lens or mirror, the distance from the lens or mirror at which a parallel beam of light rays will be focused to the smallest size possible for the lens or mirror. The focal length of a simple converging (convex) lens can be measured by focusing the rays from the sun to the smallest point possible and measuring the distance from the image to the lens.

- ▶ Face Device
- ▶ Iris Device

Footprint Comparison

- ▶ Forensic Barefoot Comparisons

Footstep Identification

- ▶ Footstep Recognition

Footstep Recognition

RUBEN VERA RODRIGUEZ¹, NICHOLAS W. D. EVANS^{1,2}, JOHN S. D. MASON¹

¹Swansea University, Singleton Park, Swansea, SA2 8PP, UK

²Institut Eurécom, 2229 route des Crêtes, 06560 Sophia-Antipolis, France

Synonyms

Footstep identification; Footstep verification

Definition

Footstep recognition is a relatively new biometric and is based on the study of footstep signals captured from persons walking over an instrumented sensing area. Since the biometric information is embedded in a time varying signal, thereby implying some form of action (in this case those of walking or running for example), footsteps can be included in the group of behavioral biometrics.

Introduction

Footstep recognition was first suggested as a biometric in 1977 by Pedotti [1], but it was not until 1997 when Addlesee et al. [2] reported the first experiments. Since then the subject has received relatively little attention in the literature and so it is perhaps of little surprise that reported performances fall short of those achievable with other, more popular, and researched biometrics. However, recent work has demonstrated the real potential of the footstep biometric which is certainly not without its appeal.

One significant benefit of footsteps over other, better known biometrics is that footstep signals can be collected covertly with minimal client cooperation. Other benefits lie in the robustness to environmental noise (a limiting aspect of speaker recognition) or lighting variability (as in the case of face recognition). There is, however, a number of new challenges to be addressed. Footsteps can exhibit a high degree of intra-class variability, i.e., different footwear, persons carrying heavy baggage and different walking speeds, all

extraneous factors which make footstep recognition an extremely challenging task.

In addressing these difficulties among others, researchers have investigated footstep signals using different sensor approaches. Systems reported in the literature include the extraction of footstep positions using video cameras, acoustic-based approaches which capture the sound of footsteps [3] and, by far the most common, under-floor contact or tactile-based sensors. These approaches range from simple ON/OFF sensors that indicate the position of the footstep [4–7] to more sophisticated sensors that capture transient pressure [1, 2, 8–13]. Pressure sensors generally measure the ground reaction force (► GRF). An example GRF profile for a single footstep signal captured from the sensor approach reported in [13] is shown in Fig. 1. Generally there are two peaks to the GRF profile, the first peak is attributable to the heel strike and the second to the toe push-off as the body is propelled forward. Figure 1 also illustrates some of the most common geometric features (maximum, minimum and mean values) as used in the works of [9, 12, 14] for subsequent classification.

Reported performances vary widely. The most statistically meaningful results obtained for footstep recognition with an identification protocol relate to a database comprised of 1,680 footsteps from 15 persons [9]. Here

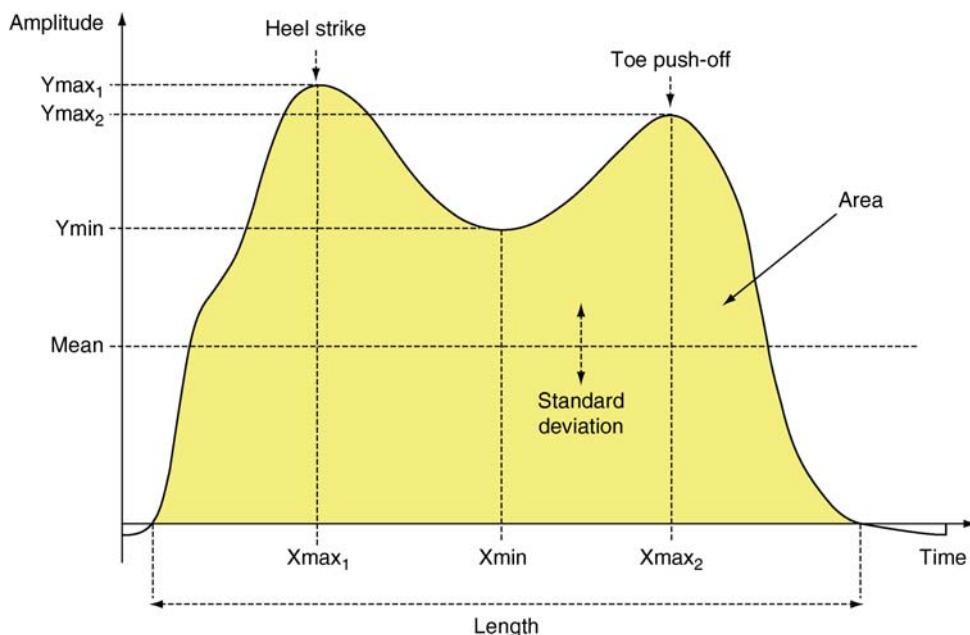
an accuracy of 93% was reported. For the case of verification as a protocol, best results relate to a database comprised of 3,147 footsteps from 41 persons [15]. Equal error rates (EERs) of 9.5 and 13.5% are reported for development and evaluation sets respectively. Results to date are promising and show that the use of footsteps as a biometric warrants further investigation.

The following sections present an overview of different applications of footstep signals and a review of published literature which has investigated the use of footsteps more specifically for biometrics.

F

Applications

It is possible to classify different biometric techniques according to the original application of the biometric signal. In the case of the fingerprint and hand geometry biometrics, signals are captured with the sole application of biometrics; whereas for speech, for example, the main application is communication, and biometrics can be considered a secondary application. Other biometrics such as the footsteps are in the middle of this range. A footstep is an action that can be captured for several applications. Potential uses of footstep signals in the



Footstep Recognition. Figure 1 Example of a GRF profile against time for a single footstep. The first peak corresponds to the heel strike and the second corresponds to the toe push-off.

literature include medicine, surveillance, smart homes, multimedia, and biometrics, none of them dominating and therefore this overview presents the entire spectrum.

In the field of medicine, footstep signals have been used to analyse different gait deficiencies by comparing normal and pathological patterns of footstep pressure signals. Following early work on biomechanics, in 1977 Pedotti [1] studied the three orthogonal components of the GRF signal using a square force plate with four piezoelectric transducers placed in the corners, similar to other systems used later for biometrics [2, 9, 10, 12]. He studied visually around 4,500 footsteps from 65 normal and 165 pathological subjects and observed stride symmetry between the left and the right feet for normal subjects but not for pathological subjects; furthermore, Pedotti noted low intra-person variability, leading to one of the first suggestions to the use of footsteps as a biometric. Commercial products today provide high resolution pressure image sequences from thin sensor mats created by printing processes. These systems are used in medicine to study for example the plantar pressure profiles, identify asymmetries between left and right feet, review dynamic weight transfer and local pressure concentrations, or identify areas of potential ulceration amongst others.

More focused on the detection of footsteps for surveillance applications, footstep signals have been used to detect human presence in a determined area. The work described in [3] reports some experiments carried out with a database comprised of five people walking ten times toward a microphone. The aim of the research was not only on footstep detection but person identification using mel-cepstrum analysis. Other work reported in [16] used piezoelectric accelerometers to detect impulses induced by walking. Footsteps were identified from three or more impulses where the sensor was excited at its resonant frequency, having satisfactory results in most occasions.

One particularly appealing application of footstep signals is found in the field of smart homes. In 2000 Mori et al. [17] developed a robotic room where multiple sensors were distributed in several locations. Footstep signals were collected from a distribution of force sensing resistors (FSRs) to specify human position in the room. A total number of 252 FSRs were installed in a $200\text{ mm} \times 200\text{ mm}$ lattice shape. More recent work on the same floor [4] (2002) increased the spatial resolution of the sensors to a 64×64 switch sensor array in a 500 mm^2 space. With this higher resolution,

experiments determined the positions of a human and a four-wheeled cart and distinguished between them. In 2004 Murakita et al. [5] reported a system for tracking individuals over an area of 37 m^2 employing basic block sensors of 18 cm^2 . The system was capable of tracking two different people when separated by more than 1.4 m but failed to track people in a crowded area due to the low spatial resolution and a low capture rate of 5 Hz. Making use of the hardware developed for the Active Floor [2], in 2001 Headon and Curwen [18] used the vertical component of the GRF and a hidden Markov model (HMM) classifier to recognise different movements including stepping, jumping, or sitting down. Applications of such a system exist in safety (i.e., fall detection for the elderly) and entertainment (i.e., video games). More recently, in 2008 Liau et al. [19] developed a system which used load cells over an area of $4\text{ m} \times 4\text{ m}$ to track people and addressed the cross-walking problem where the paths of two or more people intersect.

Footstep signals have also been used for multimedia applications. In 1997 Paradiso et al. [20] developed a system which he called The magic carpet to be used in an audio installation where users created and modified complex musical sounds and sequences as they wandered about the carpet. The sensor floor comprised a 16×32 grid of piezoelectric wires in an area of $1.8\text{ m} \times 3\text{ m}$ carpet. Later in the same year, the same laboratory developed a system installing PVDF (polyvinylidene fluoride) and FSR sensors into a dancing shoe [21]. The goal was to capture many degrees of expression and use them to drive music synthesizers and computer graphics in a real-time performance. More recently, in 2005 Srinivasan et al. [8] developed a portable pressure sensing floor constructed of modular high resolution pressure sensing mats. A sensor mat comprised 2,016 sensors made from a pressure sensitive polymer and covered an area of $62\text{ cm} \times 53\text{ cm}$, sampling each sensor at a frequency of 30 Hz. Initial applications of the system were to study interactive dance movement and video game controlling.

Review of Footsteps as a Biometric

Review of footsteps as a biometric is now an addressed work in the open literature which considers the use of footsteps specifically as a biometric. One of the first investigations into footstep recognition was reported

by UK researchers in 1997 [2]. They reported experiments on a database of 300 footsteps signals that were captured from 15 walkers in one session. The system was comprised of four load cells measuring the vertical component of the GRF and placed on the corners of a tile working at a sampling frequency of 250 Hz. They divided the database into train and test and an identification accuracy of 91% was achieved with an HMM classifier and samples from the GRF of a single footprint signal as features.

In 2000, and using a similar sensor approach, a group in the USA reported results on a database of 1,680 footprint signals collected from 15 persons using a frequency sampling of 150 Hz [9]. Signals were collected from both left and right feet and different footwear having 20 footsteps per condition using half of them for training and half for testing. Ten geometric features were extracted from the GRF of a single footprint signal including the mean value, the standard deviation, maxima, and minima, etc. They considered each combination of user, foot, and shoe type as a cluster. Then a nearest neighbour classifier was used to measure the Euclidean distance of a footprint from the test set to each cluster. An identification accuracy of 93% was reported regardless of whether the correct shoe or foot was given. In 88% of the cases, a user's footprint was more similar to other footprints for that same user than for another user, concluding from these results that footwear does not greatly affect the ability of their approach to identify the user by his footprints.

While focused toward the study of gait, a group from Switzerland [10] developed in 2002 a system fusing data acquired from 3 tiles of 4 piezo force sensors each and video cameras. A database of 480 footprints was collected from 16 persons walking barefoot using a sampling frequency of 300 Hz. The database was further divided into train and test. They studied different feature extraction techniques as geometric features from GRF [9] and phase plane (as area within the curve, position of the loop, maxima, minima, etc.). The best verification performance was achieved using the power spectral density (PSD) of the derivative GRF of footprints signals in the band of 0–20 Hz with generalized principal component analysis (GPCA), obtaining a verification EER of 9.5% with an Euclidean distance classifier.

A Korean group reported a system in 2003 [6] that used 144 simple ON/OFF switch sensors in a total area of 1 m × 3 m. Stride data (connected footprints) was

collected from ten persons who each one walked 50 times across the ubiFloor resulting in a database of 500 walking samples. Then the database was divided into training, validation, and testing data randomly. The position of several connected footprints was used as users walking features instead of the pressure of one footprint, as proposed in [2, 9]. An accuracy of 92% was reported with a multi-layer perceptron (MLP) neural network used as an experimental identification method.

In 2004 a group from Finland investigated footprint recognition using electro mechanical film (EMFi). Long strips of the sensor material were laid over an area covering 100 m². A database of around 440 footprint signals (of both feet) was collected from 11 persons at a frequency rate of 100 Hz. In their publication [11] they reported experiments with a two level learning vector quantisation (LVQ) based classifier and considered three consecutive footprints of a person to carry out a single test. On the first level each of the three single footprint signals was classified independently, and on the second level the decisions of the three consecutive footprints were taken into account having a final acceptance if a majority of the footprints were classified to the same class. The recognition rate reported was 89% of accuracy with an 18% of rejection rate. In the same year they reported different experiments [14] based on the same database. Geometric features were extracted from the GRF profiles as in [9] and first FFT coefficients. Using a distinction-sensitive LVQ (DSLVQ) classifier for a single footprint, an identification accuracy of 70% was achieved. Later in 2005, they presented experiments in [22] combining different feature sets using a two level classifier. On the first level three different feature sets were extracted from a single footprint as geometric features from the GRF as in [14], FFT of GRF with PCA, and FFT of the derivative GRF with PCA. Then, a product rule was used to combine the three results obtained. On the second level different footprints from the same person were combined using an average strategy. These experiments were done for two classifiers: LVQ and a MLP neural network. Results were better for MLP classifier in all cases, having a recognition rate of 79% for the case of a single footprint and a 92% for three consecutive footprints.

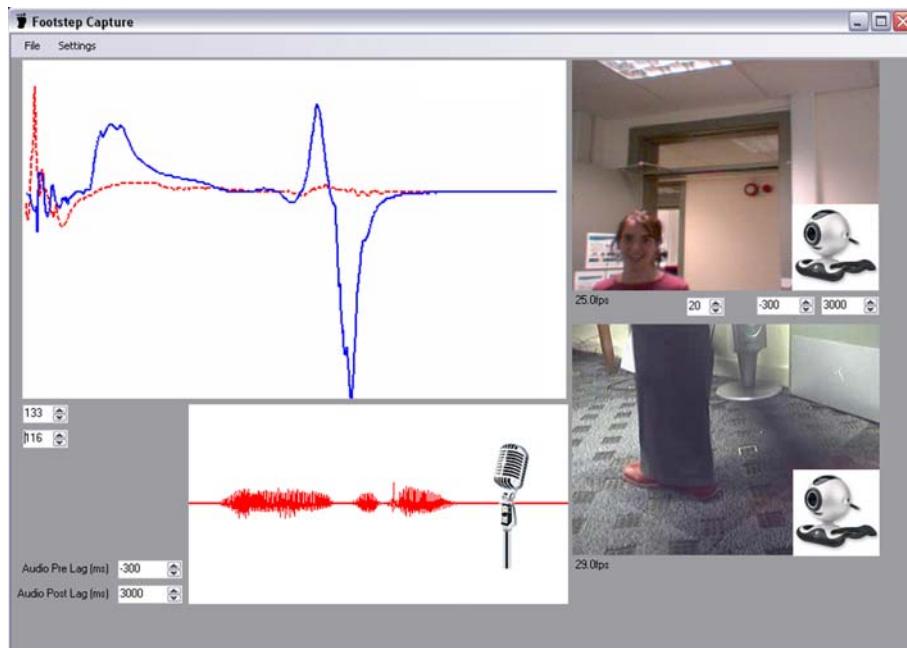
In 2005 a group from Southampton (UK) [7] reported trials with a system comprising 1,536 sensors arranged in a 3 m × 0.5 m rectangular strip with an individual sensor area of 3 cm². A database of 180

signals was collected from 15 people without wearing footwear at a frequency of 22 Hz. Each person walked over the mat 12 times and in each case two complete gait cycles (4 foot falls) were captured. Three features were extracted: stride length, stride cadence, and heel-to-toe ratio. An identification accuracy of 80% was reported using a nearest neighbor classifier to measure the Euclidean distance between each feature vector and the mean feature vector of the experimental population, i.e., the whole database. This work along with the early work of [6], differs from other published material in using binary signals rather than sampled waveforms and capture stride information from a short series of footfalls. Stride characteristics are also considered by [11, 22] as stated above.

In 2006 another group from Southampton [12] investigated a system similar to the work in [2, 9]. A database of 400 signals was collected from 11 people. Using geometric features extracted from GRF profiles as in [9] an identification accuracy of 94% was achieved using a nearest neighbor classifier in the same way as in [7].

More recently, in 2007, a research group from Swansea (UK) presented in [13, 15] experiments obtained with a database comprised of 3,174 footsteps from 41 different persons in different sessions and shoes

from two piezoelectric transducers sampled at a frequency of 1,024 Hz. The database was further divided into independent development and evaluation datasets adopting a standard best practice evaluation strategy, and therefore, presenting more statistically meaningful results and potentially more reliable predictions of performance. The database is freely available to the research community [23]. Due to the amount of data collected, a semi-automatic footstep capture system was developed to facilitate automatic labeling and rapid manual validation. Figure 2 shows a screenshot of the footstep capture system user interface. A microphone captured a spoken ID used for automatic speaker recognition to label the data (bottom part of Fig. 2); and two video cameras, one recording the face and the other the foot (top and bottom right part of Fig. 2 respectively), were used for manual data validation; the sensor responses are illustrated in the top left part of Fig. 2 as a function of time (horizontal axis). For feature extraction, two approaches were followed, namely geometric and holistic. The geometric approach was based on the extraction of main characteristic points of the footstep profile: the area, mean, length, maxima/minima, etc. The holistic approach was based on both sensor outputs and the GRF profile after PCA to reduce dimensionality of the data. In [13] two different



Footstep Recognition. Figure 2 Screenshot of the footstep capture system software developed in [13, 15].

classifiers, a nearest neighbor and SVM were also compared and findings were as expected that SVM outperforms the NN, and surprisingly holistic features outperforms the geometric features. Results of 9.5 and 11.5% EER were obtained for development and evaluation sets respectively for holistic features with an SVM classifier. Following best-practice, a formal assessment protocol was defined for the footstep recognition evaluation presented in [15]. The protocol reflects that utilized by the international NIST speaker recognition evaluations. Also, an optimization of the two feature approaches was carried out obtaining results of 9.5% EER for the development set and 13.5% EER for the evaluation set using optimized holistic features with an SVM classifier. EER given of 13.5% corresponds to 1,697 errors of each class (false acceptance and false rejection) from a total number of 25,143 tests. Such simple analysis allowing comparison across systems comes from adopting the task with verification. Work is ongoing with a multi-sensor stride capture system with the primary goal of improving confidence in the assessment of footsteps as a biometric.

Table 1 presents a comparison of this related work. The second column shows that relatively small database sizes is a common characteristic of the earlier work certainly judged in relation to other biometric evaluations where persons are normally counted in hundreds or thousands and the number of tests perhaps in many thousands. A maximum number of 16 persons and 1,680 footstep examples were gathered in all cases except in [13, 15] which reports results on 3,147 footsteps and 41 persons. In each case, except for [7, 12], the databases are divided into training and testing sets, but none use independent development and evaluation sets, with exception of [13, 15], a limitation which makes performance predictions both difficult and unreliable. Identification, rather than verification, was the task considered in all but three of the cases, the exceptions being [10, 13, 15]. Identification has the benefit of utilizing the available data to a maximum but suffers from well known scalability problems in terms of the number of classes in the set. Also, it is interesting to point out that some systems present classification results for stride data (consecutive

Footstep Recognition. **Table 1** A comparison of different approaches to footstep recognition 1997–2007

Group, year	Database (total steps/ persons)	Technology	Features	Classifier	Results
The ORL Active Floor (UK) 1997 [2]	300 steps, 15 persons	Load cells	Sub sampled GRF	HMM	ID rate: 91%
The Smart Floor (USA) 2000 [9]	1,680 steps, 15 persons	Load cells	Geometric from GRF	NN	ID rate: 93%
ETH Zurich (Switzerland) 2002 [10]	480 steps, 16 persons	Piezo force sensors	Power spectral density	Euclidean density	Verif EER: 9.5%
Ubifloor (Korea) 2003 [6]	500 steps, 10 persons	Switch sensors	Position of several steps	MLP neural net.	ID rate: 92%
EMFi Floor (Finland) 2004 [11, 14, 22]	440 steps, 11 persons	Electro mechanical film	Geometric from GRF, and FFT	MLP neural net. and LVQ	Best ID rate [22] of 92% using three footsteps as test
Southampton University (UK) 2005 [7]	180 steps, 15 persons	Resistive (switch) sensors	Stride length, cadence and heel-to-toe ratio	Euclidean distance	ID rate: 80%
Southampton University (UK) 2006 [12]	400 steps, 11 persons	Load cells	Geometric from GRF	NN	ID rate: 94%
Swansea University (UK) 2007 [13, 15]	3,174 steps, 41 persons	Piezoelectric sensors	Geometric and holistic	SVM	[15] Verif EER: 9.5% for Devel, 13.5% for Eval

footsteps) [6, 7, 11, 14, 22] while the rest only for a single footprint [2, 9, 10, 12, 13, 15]. In [22] an identification accuracy of 79% using a single footprint as a test was improved to 92% when three consecutive footsteps were used. This equates to a relative improvement of 16%.

Summary

Footstep recognition is a relatively new biometric relative to other biometrics in terms of the research reported in the literature. As reviewed, footprint signals have been used for different applications, thus different capture systems have been developed. In the field of biometrics the same trend is observed; researchers have developed systems with different sensors, extracting different features, and with different assessment protocols. Recently, in 2007, the world's first freely available footprint database was released to the research community [23]. Of particular importance to this development is, not only the size of the database both in terms of the number of footprints and clients, but the standard, best practice evaluation protocols that accompany the database. For the first time researchers will be able to develop and assess new approaches on a common and meaningfully sized database. As has happened for many other biometric modalities, it is hoped that this will stimulate new interest in the footprint biometric, lower the cost of entry and provide a solid foundation for future research.

Given its current state of development the future of footprint recognition research is difficult to predict. Some obvious avenues include new features and novel normalization approaches to reduce the effects of extraneous factors. Other possibilities include further investigation into connected footprints, i.e., stride information, information that isn't captured by single footprint systems. This research would explore the middle ground between footprints and gait. Gait is another biometric that finds applications in different areas such as in medicine, the sports industry, and biometrics. In the biometrics context, gait aims to recognise persons from a distance using walking characteristics extracted from video recordings. In contrast, footprints are a more controlled biometric due to the fixed, constrained sensing area. It would thus seem natural for future research to investigate the fusion of the two biometrics.

Related Entries

► Gait Recognition

References

- Pedotti, A.: Simple equipment used in clinical practice for evaluation of locomotion. *IEEE Trans. Biomed. Eng.* **BME-24**(5), 456–461 (1977)
- Addlesee, M.D., Jones, A., Livesey, F., Samaria, F.: The ORL active floor. *IEEE Pers. Commun.* **4**(5), 35–41 (1997)
- Shoji, Y., Takasuka, T., Yasukawa, H.: Personal identification using footprint detection. In: Proceedings of 2004 International Symposium on Intelligent Signal Processing and Communication Systems, pp. 43–47 (2004)
- Morishita, H., Fukui, R., Sato, T.: High resolution pressure sensor distributed floor for future human–robot symbiosis environments. In: Proceedings of 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 2, pp. 1246–1251 (2002)
- Murakita, T., Ikeda, T., Ishiguro, H.: Human tracking using floor sensors based on the Markov chain Monte Carlo method. In: Proceedings of the 17th International Conference on Pattern Recognition (ICPR), vol. 4, pp. 917–920 (2004)
- Yun, J.S., Lee, S.H., Woo, W.T., Ryu, J.H.: The user identification system using walking pattern over the ubiFloor. In: Proceedings of International Conference on Control, Automation, and Systems, pp. 1046–1050 (2003)
- Middleton, L., Buss, A.A., Bazin, A.I., Nixon, M.S.: A floor sensor system for gait recognition. In: Proceedings of Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05), pp. 171–176 (2005)
- Srinivasan, P., Birchefield, D., Qian, G., Kidane, A.: A pressure sensing floor for interactive media applications. In: Proceedings of the 2005 ACM SIGCHI International Conference, vol. 265, pp. 278–281 (2005)
- Orr, R.J., Abowd, G.D.: The smart floor: a mechanism for natural user identification and tracking. In: Proceedings of Conference on Human Factors in Computing Systems, pp. 275–276 (2000)
- Cattin, C.: Biometric authentication system using human Gait. Swiss Federal Institute of Technology, Zurich. PhD Thesis (2002)
- Suutala, J., Pirttikangas, S., Riekki, J., Roning, J.: Reject-optimal LVQ-based two-level classifier to improve reliability in footprint identification. *Lecture Notes Comput. Sci.* Springer, Berlin **3001**, 182–187 (2004)
- Gao, Y., Brennan, M.J., Mace, B.R., Muggleton, J.M.: Person recognition by measuring the ground reaction force due to a footprint. In: Proceedings of Ninth International Conference on Recent Advances in Structural Dynamics (2006)
- Vera-Rodriguez, R., Evans, N.W.D., Lewis, R.P., Fauve, B., Mason, J.S.D.: An experimental study on the feasibility of footprints as a biometric. In: Proceedings of 15th European Signal Processing Conference (EUSIPCO'07), pp. 748–752. Poznan, Poland (2007)

14. Suutala, J., Roning, J.: Towards the adaptive identification of walkers: automated feature selection of footsteps using distinction-sensitive LVQ. In: Proceedings of International Workshop on Processing Sensory Information for Proactive Systems, pp. 61–67 (2004)
15. Vera-Rodriguez, R., Lewis, R.P., Evans, N.W.D., Mason, J.S.D.: Optimisation of geometric and holistic feature extraction approaches for a footprint biometric verification system. In: Proceedings International Summer School for Advanced Studies on Biometrics for Secure Authentication. Alghero, Italy (2007)
16. Mazarakis, G.P., Avaritsiotis, J.N.: A prototype sensor node for footprint detection. In: Proceedings of the Second European Workshop on Wireless Sensor Networks, pp. 415–418 (2005)
17. Mori, T., Sato, T., Asaki, K., Yoshimoto, Y., Kishimoto, Y.: One-room-type sensing system for recognition and accumulation of human behavior. In: Proceedings of 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 1, pp. 344–350 (2000)
18. Headon, R., Curwen, R.: Recognizing movements from the ground reaction force. In: Proceedings of the 2001 Workshop on Perceptive User Interfaces, vol. 15, pp. 1–8. Orlando, USA (2001)
19. Liau, W.H., Wu, C.L., Fu, L.C.: Inhabitants tracking system in a cluttered home environment via floor load sensors. IEEE Trans. Autom. Sci. Eng. 5(1), 10–20 (2008)
20. Paradiso, J., Abler, C., Hsiao, K., Reynolds, M.: The magic carpet: physical sensing for immersive environments. In: Proceedings of CHI'97, pp. 277–278. Atlanta, USA (1997)
21. Paradiso, J., Hu, E.: Expressive footwear for computer-augmented dance performance. In: Proceedings of the First international Symposium on Wearable Computers. IEEE Computers Society Press, pp. 165–166. Cambridge, USA (1997)
22. Suutala, J., Roning, J.: Combining classifiers with different footprint feature sets and multiple samples for person identification. In: Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, pp. 357–360 (2005)
23. S.U.: Footstep recognition at Swansea University. Available at <http://eeswan.swan.ac.uk>

produce through its use. The terms outsole print, imprint and impression, or footwear print, imprint and impression are collectively called footwear marks.

► Footwear Recognition

Footwear Recognition

MARIA PAVLOU, NIGEL M. ALLINSON

University of Sheffield, Mappin Street, Sheffield, UK

Synonyms

Outsole pattern matching; Shoepoint matching

Definition

Footwear recognition is the process of acquiring, identifying, and verifying the marks of the outsole (underside) patterns of a shoe. These marks arise as a result of the normal use of footwear in many conditions and environments. Footwear recognition can be used by the police and other law enforcement agencies in the identification of crime suspects.

Introduction

Although footwear recognition in a strict sense is not a biometric, it does provide a very useful source of intelligence and potential evidence in the application of forensics for policing and security. As shoes are fairly personal items of apparel with usually an extended period of ownership by their wearer, they could be termed a “near-biometric.” Similar to latent fingerprints, ►footwear marks are very frequently left behind on surfaces at crime scenes [1]; and they can be more commonly recovered than fingerprints for some crime categories. A number of methods are then used to develop and collect these ►scene marks to provide useful evidential clues by linking patterns of movement of suspect individuals (at crime scenes), and can even provide strong courtroom evidence by matching a mark to an individual shoe. This useful resource has gained recent interest internationally, even resulting in

Footstep Verification

► Footstep Recognition

Footwear Marks

Footwear marks is an umbrella term describing the various types of marks that an item of footwear can

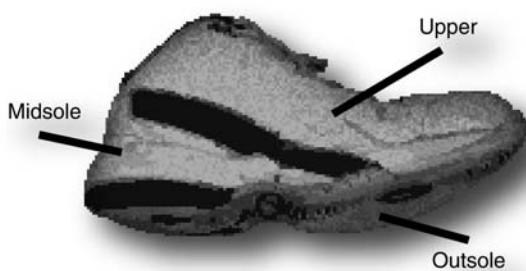
legislative changes in the UK [2] where collected footwear evidence is treated in the same way as fingerprint and DNA evidence. Namely, they have to be provided at time of arrest, and can be held and searched on local/national computer systems.

Etiology, Detection and Recovery

The typical shoe comprises of three parts (see Fig. 1) – upper, midsole, and outsole. The footwear upper is generally constructed from a variety of hard wearing fabrics or leather and is fashioned and colored in a multitude of ways. The upper holds the foot firmly in place and provides suitable support. The midsole holds the inner sole and is also used to fasten the uppers to the outsole. The outsole is the underside of the footwear, made of a durable leather, rubber or polyplastic, which provides traction and cushioning for the wearer. Manufacturers have made great efforts in the design of the outsole for the benefit of the wearer in varying activities by incorporating functional and decorative ► **tread patterns**. More commonly worn leisure footwear or sneakers typically have intricate tread designs based on the shoe model theme or manufacturer logos. What is important here is that the tread pattern is usually very distinctive to any design of shoe model just like the friction skin ridge patterns of fingers are unique to an individual.

Similar to latent fingerprints, it is the contact of the outsole with various surfaces that results in the formation of a footwear mark in a number of ways. This can be from the deposition of dry material such as dust or dirt, or wet materials such as water, blood or mud,

onto a surface. The removal of material from a surface may also form a mark, leaving a negative impression for example when stepping into and out of a shallow pool of blood, while an indented impression can be formed in a soft substrate such as snow or clay. Accordingly for each type of mark there are numerous methods by which these are detected, recorded, and preserved. Details can be found in [1, 3]. Briefly, these range from using specialized lighting methods, such as oblique and multispectral lighting, and chemical developers to enhance hard to see traces which can then be photographed. Several lifting techniques are also used to capture deposited particle materials onto a fixing substrate such as a ► **gelatin pad**. When a footwear mark is left in a soft material, such as snow, specialized plaster or molten sulphur can be used to produce a cast of the impression. Finally a print of the outsole can be made directly if available. This is done using dusting techniques, such as fine aluminum powder and then pressing onto a transparent gel sheet. More commonly the outsole can be impregnated with a dye, and printed onto paper, or with an oil-based liquid and printed on special sensitized paper – a method commonly called Printscan (see Fig. 2). This last method is the technique most employed in Police ► **custody suites** to produce impressions of a suspect's shoes. The resulting impression can then be used for one-to-one comparisons to provide forensic evidence, or can be scanned for computer-based processing. The overriding aim of all these development and recovery techniques is to obtain as true and unaltered a representation of the mark as possible for later processing and examination.



Footwear Recognition. Figure 1 Components of a typical athletic shoe, comprising the upper, midsole, and outsole.

Uniqueness and Application

Intuitively the uppers of an item of footwear are immediately more useful in identifying the make or model of a shoe. This is because of their styling, coloring, and the presence of manufacturer logos, with detailed information being readily obtainable from outlets and manufactures. However the uppers and any associated markings are rarely encountered as forensic evidence. The impressions and marks produced by the outsole are more readily found and can contain sufficient characteristic information to ascertain the manufacturer, model, and potentially the wearer. These characteristics originate in a number of



Footwear Recognition. **Figure 2** Making an outsole pattern print on sensitized paper with the Printscan method.

ways starting from the manufacture process. Footwear manufactures use a number of processes for the production of outsoles [1] resulting in a varying degree of ► **process artifacts** and defects remaining in the final product. These are one of the three useful characteristics of an outsole, which also comprises the outsole tread pattern and the accumulated wear-and-tear artifacts. Of these characteristics only the wear-and-tear artifacts are unique provided they have occurred due to a random process where something is added or taken away from the outsole that either causes or contributes to making the outsole unique. Such artifacts include nicks, cuts, scratches and ► **feathering** of the rubber material due to the normal usage of the footwear. These can be called “individual characteristics” while the outsole tread pattern and other manufacture defects are termed “class characteristics” which are distinct to a particular model of footwear and the process of its production, such as its outsole mold.

The identification and use of these characteristics have different meaning and implications. In a forensic setting the class characteristics are important as they provide information on the manufacture and model of the footwear worn and also its size. This is useful when restricting a suspect list based on physical build/size, accessibility of rare/expensive items and even geographical distribution of crimes. Once candidates of the same footwear class are available only then can comparisons be made on individual characteristics.

Forensic examiners will look for common individual characteristics between items of recovered footwear, their reproduced marks and marks found as evidence which provide conclusive links and can be used as court room evidence. Provided there are sufficient individual characteristics between an outsole and a recovered mark it may be possible to state that the outsole created the outsole mark to court room standards of evidence. Outside the forensic setting, footwear class characteristics can be very useful for screening and intelligence gathering. The footwear of a suspect can be collected and from which its class characteristics are ascertained. Usually suspects will be offenders detained or held on an unrelated offense and may have their footwear proactively compared with evidence collected at an earlier time in relation to other offenses in a local area, such as burglaries. If a link is made between an arrestee and marks found at crime scenes then it can provide law enforcement officers with some important information with which to interrogate while the arrestee is still in custody.

Comparison and Identification Methods

The comparison of class and individual characteristics of an outsole have been largely carried out by experienced forensic professionals, as an intimate

knowledge of footwear marks and their etiology is required. However, some efforts have been made to automate the identification of outsole patterns and their associated shoe model based on class characteristics. While this task is feasible, the task of automatically verifying an outsole-to-mark or a mark-to-mark based on class and individual characteristics is much more difficult. Current nation-wide footwear databases in the UK comprise over 15,000 shoe models and are constantly expanding as manufacturers introduce new styles. Crime scene marks are recovered by diverse techniques and the resulting impressions are often of poor quality, confounded by details of the underlying surface and may only represent a partial impression of the entire outsole. Such factors will make the fully automatic identification or verification of outsoles marks a very difficult, if not impossible, option.

Automatic matching of footwear patterns has not been reported much in the literature. Early works [4] have employed semi-automatic methods of ► [manually annotated](#) footwear print descriptions using a codebook of shape and pattern primitives, for example, wavy lines, geometric shapes, and logos. Searching for an example impression then requires its encoding in a similar manner to that used for the reference database. This process is laborious and can be the source of poor performance as similar patterns may be inconsistently encoded by different users. It is still, however, predominantly used by Police Forces across the UK and elsewhere. Two major factors are attributed to this; firstly, the ease of understanding the coding methodology and its primitives (i.e., their visual intuitiveness), and secondly, a lack of proven and accepted automated systems based on rigorous standards and robust performance.

One early work [5] employed an intuitive coding scheme based on shapes automatically generated from footwear images using various ► [image morphology](#) operators. The spatial positioning and frequencies of these shapes were used for classification with a neural network. Unfortunately, the authors did not report any performance statistics for their system. Later works do not follow this approach but instead use template matching approaches based on the pattern representation in a suitable transform space. Fractal representations were used in [6, 7] with a mean square noise error method for classification. They

report good results; however the dataset used was small and contained no spatial or rotational variations. In [8], Fourier Transforms (FT) are used for the classification of full and partial marks of varying quality. The FT provides a degree of invariance to translations and encodes spatial frequency information. By incorporating duplicate rotated templates a degree of rotation invariance was also possible. Their approach was weak on first rank precision, and this may have been due to the large variation in print quality. Also, the footwear prints were processed globally and hence noise in the images could have hindered the quality of useful encoded local information evident in the print. Despite these failings this approach is promising and shows the importance of encoding local information.

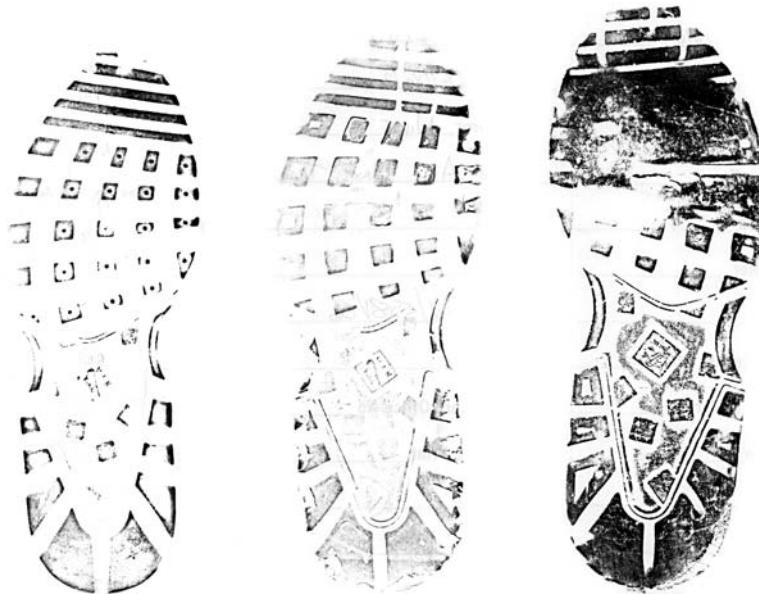
In [9] local image features (LIF) were used to make one-to-one comparisons of patterns in a database of footwear marks which had been subjected to added noise and transformed in various ways. This work compared a number of the aforementioned approaches, for which LIF performed very well, indicating that local image information is crucial to good matching performance especially when considering partial marks. However as test samples were generated from the training set it is not clear how this template matching approach would perform on untrained test samples or in a one-to-many and pattern search scenario.

Moving away from a template-based approach, work in [10] employed a histogram approach by quantizing edge directions into 5° intervals. In order to cope with rotational variations of the image, and hence translation of the histogram bins, an FT was also applied. Their approach is useful as now the pattern information is described in a compact way based on the content of edges, however, due to histogram normalization and the use of the FT, this approach is not effective with partial marks.

The flexibility of histogram-based encoding is effective in many image retrieval and indexing tasks as has been demonstrated in [11]. It is interesting that the manual coding schemes still used for footwear impressions bear a strong resemblance to basic histogram coding methods. The work proposed by [12] proceeds by encoding footwear patterns into a compact vector space model using a feature-rich codebook of local feature descriptions. The codebook is a quantized set of features derived by applying such feature detectors

as maximally stable extremal regions (MSERs), followed by the use of robust feature descriptors. A query shoe impression can be coded using this code-book and compared against others. This approach

is fully automated and yet still preserves some similarity to manual coding approaches in that the coded features resemble the use of annotated features in the semi-automated systems described previously.



Footwear Recognition. **Figure 3** Outsole marks captured using the Printscan method showing different stages of wear and print quality. From left to right the pattern shown has increasing wear. Note the change in tread pattern appearance.



Footwear Recognition. **Figure 4** Various footwear marks collected from scenes of crime. From left to right, a print in blood on soft tiles, outsole imprint in mud, an electrostatic lift of dirt from carpet, a casting, and a gel lift of a dusty mark.

This approach is able to cope with rotational changes and queries in the form of partial impressions. Though good performance is reported in this study (e.g., for a reference set of 374 different footwear models, a precision at first rank of 87% was obtained), it is difficult to compare the performance of differing approaches as no common datasets have been employed nor is there agreement on testing methodologies.

A number of issues are not addressed by these methods. One is the problem of dealing with large appearance changes in footwear marks as the outsole becomes worn over time and where the mark is strongly degraded or partially missing (see Fig. 3). Some work by Su et al. in [13, 14] looked at methods for assessing and improving image quality of footwear marks collected using the Printscan process. Even if good quality marks are obtainable there is still work to be done on how to compare marks obtained by different methods, for example between a casting and a gel lift (see Fig. 4). Additionally, a lack of standards for the digital capture of marks and the absence of an agreed and openly available dataset are issues that still need to be addressed.

Summary

The use of footwear impressions both recovered from crime scenes and acquired from suspect's shoes have a significant role to play as a "near-biometric" in forensic investigations. It has, despite its long history as an important tool of forensics, remained until recently largely forgotten. This will undoubtedly change in the near future with a much greater development and application of mainstream biometric tools and methodologies.

Related Entries

- ▶ [Forensic Applications, Overview](#)
- ▶ [Forensic Barefoot Comparison](#)

References

1. Bodziak, W.J., ed.: *Footwear Impression Evidence*. CRC Press, Boca Raton (2000)
2. Parliament: *Serious Organised Crime and Police Act 2005*, Elizabeth II. The Stationery Office (2005)
3. Hilderbrand, D.S., ed.: *Footwear, The Missed Evidence*. Staggs Publishing (1999)

4. Mikkonen, S., Astikainen, T.: Databased classification system for shoe sole patterns-identification of partial footwear impression found at a scene of crime. *J. Forensic Sci.* **39**, 1227–1236 (1994)
5. Geraarts, Z., Keijzer, J.: The image-database rebezo for shoeprints with developments on automatic classification of shoe outsole designs. *Forensic Sci. Int.* **82**, 21–31 (1996)
6. Bouridane, A., Alexander, A., Nibouche, M., Crookes, D.: Application of fractals to the detection and classification of shoeprints. In: *Proceedings International Conference on Image Processing*. Volume 1. pp. 474–477 (2000)
7. Alexander, A., Bouridane, A., Crookes, D.: Automatic classification and recognition of shoeprints. In: *Proceedings Seventh International Conference on (Conf Image Processing and Its Applications* Publ. No. 465). Volume 2. 638–641 (1999)
8. de Chazal, P., de Chazal, P., Flynn, J., Reilly, R.: Automated processing of shoeprint images based on the fourier transform for use in forensic science. *Trans. Pattern Anal. Mach. Intell.* **27**, pp. 341–350 (2005)
9. Su, H., Crookes, D., Bouridane, A., Gueham, M.: Local image features for shoeprint image retrieval. In: *British Machine Vision Conference 2007*. (2007)
10. Zhang, L., Allinson, N.: Automatic shoeprint retrieval system for use in forensic investigations. In: *5th Annual UK Workshop on Computational Intelligence* (2005)
11. Everingham, M., Zisserman, A., Williams, C.K.I., Van Gool, L.: The PASCAL Visual Object Classes Challenge 2006 (VOC2006) Results (2006)
12. Pavlou, M., Allinson, N.: Automatic extraction and classification of footwear patterns. In: *Intelligent Data Engineering and Automated Learning*, IDEAL 2006. pp. 721–728 (2006)
13. Su, H., Crookes, D., Bouridane, A.: Thresholding of noisy shoeprint images based on pixel context. *Pattern Recognit. Lett.* **28**, 301–307 (2007)
14. Su, H., Bouridane, A., Crookes, D.: Image quality measures for hierarchical decomposition of a shoeprint image. *Forensic Sci. Int.* **163**, 125–131 (2006)

Force Field Feature Extraction

The overall objective in defining feature space is to reduce the dimensionality of the original pattern space, while maintaining discriminatory power for classification. To meet this objective in the context of ear biometrics a novel force field transformation which treats the image as an array of mutually attracting particles that act as the source of a Gaussian force field has been developed. Underlying the force field there is a scalar potential energy field, which in the case of an ear takes the form of a smooth surface that

resembles a small mountain with a number of peaks joined by ridges. The peaks correspond to potential energy wells and to extend the analogy the ridges correspond to potential energy channels. Since the transform also happens to be invertible, and since the surface is otherwise smooth, information theory suggests that much of the information is transferred to these features, thus confirming their efficacy. Force field feature extraction, using an algorithm similar to gradient descent, exploits the directional properties of the force field to automatically locate these channels and wells, which then forms the basis of the characteristic ear features.

► Physical Analogies for Ear Recognition

Force Field Transform

An invertible linear transform which transforms an image into a force field by pretending that pixels have a mutual attraction proportional to their intensities and inversely to the square of the distance between them rather like Newton's Law of Universal Gravitation. Each pixel is assumed to generate a spherically symmetrical force field so that the total force $\mathbf{F}(\mathbf{r}_j)$ exerted on a pixel of unit intensity at the pixel location with position vector \mathbf{r}_j by a remote pixel with position vector \mathbf{r}_i and pixel intensities $P(\mathbf{r}_i)$ is given by the vector summation,

$$\mathbf{F}(\mathbf{r}_j) = \sum_i \left\{ P(\mathbf{r}_i) \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} \forall i \neq j \right\}. \quad (1)$$

To calculate the force field for the entire image, this equation should be applied at every pixel position in the image. In practice this computation would be done in the frequency domain using Eq. 2 where \mathfrak{I} stands for FFT and \mathfrak{I}^{-1} stands for inverse FFT.

$$\text{forcefield} = \sqrt{M \times N} \mathfrak{I}^{-1} [\mathfrak{I}(\text{unitforcefield}) \times \mathfrak{I}(\text{image})]. \quad (2)$$

► Physical Analogies for Ear Recognition

Forensic

Forensic is the use of science or technology in the investigation and establishment of facts or evidence in the court of law.

► Skull, Forensic Evidence of

Forensic Anthropology

Forensic anthropology is the application of physical anthropology in special cases with forensic importance, such as to identify the skeletonized human remains.

► Skull, Forensic Evidence of

Forensic Applications, Overview

CHRISTOPHE CHAMPOD

Institut de Police Scientifique, Ecole des Sciences Criminelles, Université de Lausanne, Switzerland

Introduction

The use of biometric data is a decisive process in **► forensic science** that helps to establish a person's identity or associate two unknown persons. Forensic scientists realized that physiological or behavioral data could help to inform about, sort, and potentially individualize the persons involved in criminal offences. It is the case when (1) an unknown individual (living or his/her remains) has to be identified, (2) when biometric traces left by unknown individuals during activities of interest have to be traced back to their sources, or (3) when biometric traces have to be linked together in a series. Situations (1) and (2) require comparison between biometric information gathered from unknown sources and material of known (or declared as

such) origin, either on a one-to-one or on a one-to-many basis. In the latter case, data of known origin are organized in a database, allowing one-to-many searches. The third situation (3) compares biometric material from unknown sources and groups them according to potential (yet unidentified) sources. This last activity may involve the use of a database or may be carried out on a case-by-case basis.

Both situations (2) and (3) take advantage of what is known in forensic science as “Locard’s exchange principle”. Locard suggested in 1920 that forensic scientists can take advantage of the traces (such as fibers, paint, firearms discharge residues, dust, blood-stains, etc.) and marks (e.g., finger marks, footwear marks, toolmarks, etc.) exchanged between actors (e.g., a victim and an offender), and associated objects or scenes involved in criminal activities. The systematic search (potentially helped with detection techniques), preservation and analysis of these marks and traces will help to establish the relation between the actors, objects or scenes, and reconstruct the course of activities. The biometric features that can be helpful are many (and tend to increase in our modern society) and may consist of the following types of traces or marks:

1. Handwritten notes, including disputed signatures.
2. Finger marks and, by extension, any marks of friction ridge skin.
3. Barefoot impressions.
4. Bloodstain, semen stain, saliva stain, and any other biological fluids.
5. Earmarks.
6. Impression left by a face on airbags in car crashes [1].
7. Images of individuals (including images showing face, ears, or other identifying features) in stills or videos taken either from CCTV systems or any image-recording device.
8. Recording of a voice utterance of an individual (either in analog or digital form).

Biometric features have been used in forensic science for many centuries for some attributes (such as handwriting) and more recently for others (e.g., DNA profiling or recognition of faces from CCTV surveillance camera). However, in many respects, the application of biometry in forensic science is in contrast to the deployment of biometric systems in other areas (such as access control) this is mainly due to the

unpredictable nature in terms of quality and quantity of the biometric features available to the forensic scientist, especially when they are collected as traces. These contrasts have been detailed elsewhere [2].

This overview is intended to cover the standard use of biometric features by forensic experts, either manually or with semi-automated or automated systems. While distinguishing the types of biometry helps to structure the text, it also helps to distinguish between the investigative and evaluative use of the techniques.

In the investigative mode, marks and traces are questioned to provide leads that may help to focus the inquiry, without making any reference to a potential source at the outset. Typical questions are as follows:

1. Can we link these criminal incidents the biometric evidence recovered?
2. On the basis of the recovered material, can we make any inference regarding the sex, age, ethnicity, or other physical attributes of the donor?
3. By comparing this biometric entry with a forensic database, is it possible to prepare a short list of potential donors?

In the evaluative mode, the questions are directed towards a source that is available, to provide control material to be compared against the unknown material. The case is then focused on one or more identified persons, with a view to help assess whether they are the source of the recovered material.

The references used hereinafter are limited to major textbooks or papers for each evidence type considered. Some evidence types are largely covered in specific chapters of this encyclopedia and address specifically the issues associated with automatic recognition.

Admissibility of Biometric Evidence in Court

The question of admissibility is crucial when the scientific element is used in the evaluative mode, especially when it is expected to find and present identification evidence in a courtroom. In the United States of America, the criterion for admissibility was traditionally based on the *Frye rule* (1923), which invites the judge – acting as a gatekeeper – to assess if the

technique has gained general acceptance within the relevant scientific community. It was under *Frye* that numerous human identification evidence types such as fingerprints, handwriting evidence, and DNA gained acceptance. Earprint, however, failed to pass the *Frye* test. The *Frye* standard has been revised following a ruling of the US Supreme court in *Daubert v. Merrell Dow Pharmaceuticals* (1993) and its progeny [3]. *Daubert* gave an interpretation of the Federal Rule of Evidence (FRE) and required the judge, still acting as gatekeeper, to assess more than only the general acceptance and include five criteria:

1. Whether the expert's technique or theory can be or has been tested, that is, whether the expert's theory can be challenged in some objective sense.
2. Whether the technique or theory has been subject to peer review publication.
3. The known or potential rate of error of the technique or theory when applied.
4. The existence and maintenance of standards and controls.
5. Whether the technique or theory has been generally accepted in the scientific community.

Nowadays in the United States of America, at the federal level, *Daubert* is in force. States may apply either *Frye* (or similar state decisions) or *Daubert*.

Daubert led to an increased number of challenges in court and forced the forensic community to articulate in detail the foundations of their disciplines, even in areas that had gained acceptance in US courtrooms for numerous years (see <http://www.daubertontheweb.com/>).

In Europe, there is no specific admissibility rule regarding scientific evidence. The principle of the judges' free evaluation of the evidence prevails. Hence, it is not surprising to see currently, a limited debate in the European jurisprudence regarding the admissibility of identification evidence.

From Anthropometry to Fingerprinting and AFIS

In the face of the absence of any reliable means of identifying recidivists, Alphonse Bertillon proposed in 1881 a classification and retrieval method based on anthropological measures. He took advantage of the fact that bone lengths remain constant in adulthood.

It varies from individual to individual and can be measured with reasonable precision. Eleven precise measurements (height of the individual, length of outstretched arms, height of trunk, length and width of the head, length of the left middle finger, the left foot, left forearm, and right ear) combined with a mention of the color of the iris, were proposed to establish an anthropometric form for each arrested individual. This anthropometric record was completed with a photograph of the face and a standardized description of particular marks that can be filled and retrieved. The system was essentially used as an investigative measure to help identify individuals arrested multiple times (in time and place). The combination of anthropometric measurement, forensic photograph, and the standard description of the face was coined "Bertillonage". A rapid spread of Bertillonage has been observed at the turn of the 20th century across the world police departments and penal institutions [4]. The limitations of this technique were quickly noticed: (1) uneven distributions of the measurements; (2) correlation between features; (3) inter-operator variations, and (4) the imperative need of the body of the individual because no anthropometric traces are left on crime scenes (at the time).

In 1880, Herschel, a colonial administrator in India, published his proposal to use finger prints to identify individuals. At the same time (in 1880), Faulds, a medical missionary in Japan, proposed using finger prints for investigative identification purposes as well, as finger marks could be detected on crime scenes. Fingerprinting became a credible alternative to anthropometry for identification of habitual offenders, when Galton presented in 1892 the basic axioms of fingerprinting, including permanence (based on Herschel's work and data), discriminative power (Galton published the first statistical model on the fingerprint variability), and the possibility of reliably classifying fingerprints into three basic patterns. The classification method was then greatly improved by Henry and gained a large acceptance in English-speaking countries. Almost simultaneously, Vucetich, proposed a simpler system, based on Galton's initial proposal, that proved very successful for small to medium size databases. From the early 1900s, fingerprinting became the sole means of identification of habitual offenders throughout the world [5]. The use here is investigative (search for a potential candidate in a repository of fingerprint forms) and evaluative (verification on a

one-to-one basis), based on ten print records (taken from living or dead individuals).

An additional benefit of papillary ridge skin is that marks are left on the crime scene and are either readily visible or detectable using adequate detection techniques [6]. The first cases of identification of criminals through the finger marks left by them are attributed to Vucetich (1892) and Bertillon (1902). These marks can be searched against a fingerprint file or, of late, in an AFIS (Automatic Fingerprint Identification System). With the increase of tenprint cards and the difficulty of searches based on papillary marks, research in automatic retrieval processing systems took off in parallel with the technological advances since the 1960s. All forensic AFIS are nowadays largely based on minutiae matching both for finger and palm impressions [7]. The main advantage of an AFIS is the ability to compare a single print or mark, as well as a tenprint card, to the whole database. Note that AFIS provides a list of best candidates (according to a scoring/ranking metric). The identification process is not carried out by the system, but processed manually by an expert (through a dedicated user interface) in exactly the same way as if the potential candidate prints were suggested as a result of the usual police inquiry.

Other Biometric Characteristics used for Human Identification Purposes

The use of *deoxyribonucleic acid (DNA)*, a chain of nucleotides contained in the nucleus of our cells, has been a major breakthrough in forensic science to help in the identification of unknown individuals or biological samples left by them. Nuclear DNA can be extracted from all biological tissues (blood, saliva, urine or semen, from hair (with roots) and skin cells left by contact with the skin). For identification purposes (of the living or dead), samples are obtained from blood, saliva, or bones. The most common analysis of nuclear DNA is focused on STRs (short tandem repeats) [8]. STRs are repetitive sequences at a given location of the DNA molecule, of non-coding nature, which show a large and well-documented polymorphism. At a given locus, one individual will show two specific numbers of repetitions of the given sequence of nucleotides. These two numbers called *alleles* then give the biometric template for that locus. Note that one allele results from the genetic transmission from

the biological father, and the other from the biological mother. The constitution of DNA databases to assist investigation is simple and has been developed in most countries.

Forensic applications of DNA profiling for human identification are numerous for STR analysis and cover (1) the comparison of an unknown profile (for example, from human remains) of an individual to a database of known profiles or profiles of potential relatives; (2) filiation testing when putative genitors are available or alternatively with ancestors or descendants. The first introduction of DNA profiling in forensic science dates back to 1986 (the Pitchfork case in the UK), but the large development of practices started in the 1990s. Before that time, biological fluids were analyzed using blood grouping determination or analysis of various proteins or enzymes [9]. Most of these forensic analyses have been abandoned for identification purposes in favor of DNA profiling because of the limited sensitivity and discriminating power of these systems.

When nuclear DNA cannot be analyzed (typically because of the degradation of DNA), mitochondrial DNA (contained in the mitochondria and inherited through maternal lines) can be used. Its discriminating power, however, is much lower than STR nuclear DNA analysis.

Dental features are mainly used in the identification of human remains in cases of missing persons or mass disasters [10]. The features used range from the standard dental record (indication of missing teeth, restorations, crowns, etc.) to dental radiographs (tooth contours, relative positions of neighboring teeth, and shapes of the dental work). These anatomical features have shown very good stability and variability and the teeth serve as a suitable repository of manmade operations that will leave various marks and shapes. Alphanumeric data can easily be organized in databases and such systems are used operationally in cases of mass disasters (<http://www.interpol.int/Public/DisasterVictim/Default.asp>).

Forensic analysis of soft tissues can help in the identification of remains. Analysis of scars, incised wounds, burn marks, trauma, and medical/surgical intervention are typical either in the investigative or evaluative mode [11]. When no tissue is left, *forensic anthropology* becomes an essential part of forensic and archeological investigations [11, 12]. Following the recovery of unidentified skeletal remains, the forensic anthropologist

can assist in guiding the investigation to identify the sex, ethnic origin, stature, and age (and if it is a woman's remains, whether she had gave birth) of the deceased. This information is investigative in nature. The same applies to cranio-facial reconstruction from the skull to help in the search for a deceased person [13]. When reference material is made available (X-ray images from ante mortem medical documentation), its comparison with post mortem data can help to establish identity through the analysis of morphology, fractures, medical interventions on the bones, and frontal sinus shapes [14]. Most of these areas have not been subjected to extensive automation research [15].

Other Biometric Marks Left Following Activities of Forensic Interest

DNA profiles can be obtained from the marks left behind by activities of forensic interest, typically from stains of blood, saliva, urine or semen, and from hair (with roots) and skin cells left by mere contact. Extracts are amplified using a sensitive and selective DNA replication method known as *Polymerase Chain Reaction* (PCR). In practice sensitivity to levels below 100 pg of DNA (a few cells) can be achieved. Such sensitivity widens the investigative possibilities, allowing the analysis of biological stains of very limited quantity. These profiles can be used to compare a DNA profile obtained from biological material against profiles from known individuals. If a correspondence is obtained, then this information can be used as evidence in court. It is important to stress that a match between two DNA profiles does not establish conclusively an identification of sources. Indeed, although the selectivity of DNA profiling is very high, there exists a probability of random association. In addition, DNA analysis offers some investigative capabilities gathered through the systematic comparison of DNA profiles coming from various scenes or familial searches against the DNA database. Another investigative aspect is the use of specific DNA analysis (SNP for single nucleotide polymorphism) to infer (within defined uncertainty boundaries) iris or hair color, skin pigmentation, and ancestry background [16].

The morphology of the *ear* was considered by Bertillon as the most identifying part of an individual. This modality was thus quickly used for identification purposes in forensic cases, either on photographs

(or still images from video recording [17]) or on *ear marks* left on crime scenes, for instance, on doors. Forensic ear or ear print comparison is traditionally completed by skilled examiners according to published principles and protocols [18]. It is important to note that there is a big gap in terms of quality between a well-taken photograph of a ear and its impression on a door; hence, the strength of the evidence may vary from case to case as a function of the quality and the extent of the available material. Ear print examination can be used in the investigative phase to constitute a series based on the collected marks or to estimate the height of the donor, or in an evaluative manner, to associate a recovered mark with the ears of a designated individual.

Barefoot impressions can be left either on crime scenes or inside the shoes [19]. In the first instance, their investigation will help to assess the sequence of events and associate or exclude a given individual from being at the source of these marks. In the second, the analysis can help to assess whether a given individual is the habitual wearer of the shoe. Their use in criminal investigations predates the use of finger marks [20]. Barefoot impressions have shown a very high discrimination power and allow, when the quality of the mark is adequate, to bring powerful evidence of the identity of the sources in court.

Bite marks can be left on various substrates (the skin of a victim, some food, etc.) and can be compared against the control material from potential donors. A full account of their detection and analysis can be found in [21]. If bite mark analysis is to continue to play a role in the judicial process, there is an urgent need for high quality studies that meet the levels of forensic and scientific scrutiny applied to the other disciplines within the criminal justice system [22, 23].

From time to time, *lip marks* can be recovered from objects that came into contact with lips. Forensic lip print analysis is a very anecdotal area.

Handwriting and signature are biometric attributes with a long history in forensic science. The principles and procedures used by forensic experts to assign questionable handwritten documents to known individuals are described in [24]. The forensic expert tries to assess existing similarities and dissimilarities between control and recovered samples through a subjective estimation of the individuality and variability of the material at hand. At the moment, the automatic techniques used for handwriting and signature recognition are in their

infancy, especially in forensic science. Following the *Daubert* challenges, the field has been the focus of an increased scrutiny as to its scientific underpinning. It has led to a new body of research that shows the fertile avenues of collaboration between biometric computer science and forensic science [25–27].

Analog/Digital Biometric Information Recorded in Investigations

Forensic speaker recognition can be defined as any process using speech signals to determine if a specific individual was the speaker of a specific declaration. Experts may reach opinions from a variety of techniques used alone or in combination: auditory comparison, visual comparison of spectrograms, and semi-automatic methods for extraction of specific parameters (e.g., formant frequencies). Auditory comparisons are more likely to be conducted by phoneticians. They assess voice characteristics (voice, speech, language, and linguistic) either subjectively or objectively (using signal processing tools). The visual spectrographic approach was first proposed in 1962. In 1976, the US National Academy of Sciences recommended the use of this approach in forensic cases cautiously [28]. Automatic speaker recognition is also used in forensic science (see related entry in this Encyclopedia). Several characterization and modeling tools have been developed for automatic speaker recognition. All are sensitive to voice modification in recording and transmission conditions and their performance worsens when the conditions deteriorate. In forensic cases, the recording conditions of the trace and the reference materials are rarely similar or ideal, but rather record in different and unconstrained conditions, i.e., through mobile communications (GSM) transmission and with background noise. Due to these factors, the comparison is often undertaken under adverse conditions.

Facial images are more and more available for forensic investigations. Forensic face recognition is generally carried out by dedicated experts using approaches based either on morphological analysis of facial structures, anthropometric measurements, or image superimposition [29]. The morphological approach is based on a nomenclature for the description of the physiological aspects of the nose, the forehead, and the ear. Additional information, such as facial wrinkles and scars, can also be used. As the description

is rather subjective, variations between operators are observed. In addition, the features of the same individual change due to expression changes, photographic angles or aging, and the demonstration of their statistical independence is often weak. The anthropometric approach can be described as the quantification of physiological proportions between specific facial landmarks. This method is only used for the comparison of faces with the same orientation. In order to avoid any scale and absolute size differences between photographs, ratios are calculated from these landmarks. Lighting conditions, camera distortions, camera positioning, facial orientation, facial expressions, and aging may impact the measures. The superimposition-based approach is the juxtaposition or the superimposition of facial images, taken under the same acquisition conditions (the orientation, the pose, and the size).

These three main comparison approaches do not yet consider automatic face recognition (a subject covered in several chapters of this Encyclopedia). Automatic face recognition systems have a large role to play in the future, not only in dealing with the face as such but also taking advantage of lips [30] or other features. But before introducing any automatic face recognition in court, a full and systematic assessment of the system should be conducted under realistic conditions, using fit-for-purpose forensic efficiency measures.

The prevalence of images or videos in modern society opens the route to the development of new types of forensic biometry (some are already covered in this Encyclopedia, e.g. gait analysis). Is it not rare to observe anatomical features on images that can help towards the identification of the individual captured on these images? These features can be *skin details, scars, veins and tattoos*. Biometric developments are still in their early stages [31].

References

1. Yamazaki, K., Imaizumi, K., Kubota, S., Atsuchi, M., Noguchi, K., Yosino, M.: Experimental study on personal identification from faceprint on vehicle's airbag. Japanese Journal of Science and Technology of Identification. **9**(1), 19–27 (2004)
2. Dessimoz, D., Champod C.: Linkages between biometrics and forensic science. In: Flynn, P.J., Jain, A.K., Ross, A. (eds.) Handbook of Biometrics, pp. 425–459. Springer, New York (2007)
3. Saks, M.J., Faigman, D.L.: Expert evidence after Daubert. Annu. Rev. Law Soc. Sci. **1**(1), 105–130 (2005)

4. Cole, S.: Suspect identities: A history of fingerprinting and criminal identification. Harvard University Press, Cambridge, MA (2001)
5. Berry, J., Stoney, D.A.: The history and development of fingerprinting. In: Lee, H.C., Gaenslen, R.E. (eds.) Advances in Fingerprint Technology, 2nd edn. pp. 1–40. CRC Press, Boca Raton, FL (2001)
6. Champod, C., Lennard, C.J., Margot, P.A., Stoilovic, M.: Fingerprints and other Ridge Skin Impressions. CRC Press, Boca Raton, FL (2004)
7. Komarinski, P.: Automated fingerprint identification systems (AFIS). Elsevier, New York (2005)
8. Butler, J.M.: Forensic DNA typing, 2nd edn. Elsevier, Burlington, MA (2005)
9. Gaenslen, R.E.: Sourcebook in Forensic Serology, Immunology, and Biochemistry. US Department of Justice, National Institute of Justice, US Printing Office, Washington, DC (1983)
10. Sweet, D., Pretty, I.A.: A look at forensic dentistry – Part 1: The role of teeth in the determination of human identity. Br. Dent. J. **190**(7), 359–366 (2001)
11. Black, S.M. (ed.): Forensic Human Identification: An Introduction. CRC Press, Boca Raton, FL (2006)
12. Pickering, R.B., Bachman, D.C.: The Use of Forensic Anthropology. CRC Press, Boca Raton, FL (2000)
13. Iscan, M.Y., Helmer, R.P. (ed.): Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification. Wiley, New York (1993)
14. Christensen, A.M.: Assessing the variation in individual frontal sinus outlines. Am. J. Phys. Anthropol. **127**(3), 291–295 (2005)
15. Falguera, J.R., Falguera, F.P.S., Marana, A.N.: Frontal sinus recognition for human identification. In: Vijaya Kumar, B.V.K., Prabhakar, S., Ross, A.A. (eds.) Biometric Technology for Human Identification V. In: Proceedings of the SPIE; 2008 March 18, 2008; Orlando, FL. SPIE; 2008. p. 69440S–9
16. Frudakis, T.: Molecular photofitting: Predicting Ancestry and Phenotype using DNA. Academic Press, Burlington, MA (2008)
17. Hoogstrate, A.J., van den Heuvel, C., Huyben, E.: Ear identification based on surveillance camera images. Sci. Justice. **41**(3), 167–172 (2001)
18. van der Lugt, C.: Earprint Identification. Elsevier Bedrijfsinformatie, Gravenhage (2001)
19. Kennedy, R.B., Yamashita, A.B.: Barefoot morphology comparison: A summary. J. Forensic Ident. **57**(3), 383–413 (2007)
20. Caussé, S.: Des empreintes sanguines des pieds, et de leur mode de mensuration. Annales d'hygiène publique et de médecine légale. 1854;1 (2ème série):175–89
21. Dorion, B.J. (ed.): Bitemark Evidence. Marcel Dekker, New York (2005)
22. Pretty, I.A.: The barriers to achieving an evidence base for bitemark analysis. Forensic Sci. Int. **159**(Suppl 1), S110–S20 (2006)
23. Bowers, C.M.: Problem-based analysis of bitemark misidentifications: The role of DNA. Forensic Sci. Int. **159**(Suppl 1), S104–S9 (2006)
24. Huber, R.A., Headrick, A.M.: Handwriting Identification: Facts and Fundamentals. CRC Press, Boca Raton, FL (1999)
25. Marquis, R., Schmittbuhl, M., Bozza, S., Taroni, F.: Quantitative characterization of morphological polymorphism of handwritten characters loops. Forensic Sci. Int. **164**, 211–220 (2006)
26. Schomaker, L.: Advances in writer identification and verification. In: Ninth International Conference on Document Analysis and Recognition – ICDAR 2007, pp. 1268–1273 (2007)
27. Srihari, S., Huang, C., Srinivasan, H.: On the discriminability of the handwriting of twins. J. Forensic Sci. **53**(2), 430–446 (2008)
28. Bolt, R.H., Cooper, F.S., Green, D.M., Hamlet, S.L., McKnight, J.G., Pickett, J.M. et al.: On the Theory and Practice of Voice Identification. National Research Council, National Academy of Sciences, Washington, DC (1979)
29. Iscan, M.Y.: Introduction of techniques for photographic comparison: Potential and problems. In: Iscan, M.Y., Helmer, R.P. (eds.) Forensic Analysis of the Skull, pp. 57–70. Wiley-Liss, Inc., New York (1993)
30. Choraś, M.: Human lips as emerging biometrics modality. In: Image Analysis and Recognition: 5th International Conference, ICIAR 2008, Póvoa de Varzim, Portugal, June 25–27, 2008 Proceedings, p. 993–1002. Springer, Berlin (2008)
31. Jain, A., Lee, J.-E., Jin, R.: Tattoo-ID: Automatic tattoo image retrieval for suspect and victim identification. In: Advances in Multimedia Information Processing – PCM 2007, pp. 256–265 (2007)

Forensic Barefoot Comparisons

BRIAN A. YAMASHITA¹, ROBERT B. KENNEDY²

¹Forensic Identification Operations Support Services, National Services and Research, Royal Canadian Mounted Police, Ottawa, ON, Canada

²Royal Canadian Mounted Police (retired), Ottawa, ON, Canada

Synonyms

Barefoot morphology comparison; Footprint comparison

Definitions

Forensic barefoot comparison, or barefoot morphology comparison, describes the comparison of impressions of the weight-bearing areas of feet in an attempt to include or exclude a suspect as someone linked to a crime scene. A bare or socked foot impression found at the crime scene can be compared to inked barefoot

impressions and footprint casts taken from a suspect. Similarly, a link to footwear matched to a crime scene can be determined by comparing the insoles of the crime scene footwear to footwear seized from a suspect, or to inked impressions and casts taken from a suspect.

Introduction

Barefoot morphology comparison refers to the examination of the weight-bearing areas on the bottom of a human foot, when ridge detail is not present, to establish a link between the bare foot of an individual and a footprint impression found at a crime scene [1–3]. In the case of footwear linked to a crime scene, comparison can be made to shoes seized from a suspect, or to inked impressions or casts taken from a suspect. Research has indicated that the shapes of footprints are sufficiently variable to make it possible to include (as having possibly made the impression) or exclude (as definitely not having made the impression) a suspect as being the person who created a particular footprint at a crime scene [3]. As an example, Fig. 1 shows barefoot impressions taken from identical twins,

illustrating that even twins can be differentiated based on their footprints.

When a crime scene is being examined, it is common to find footprints that might be those of the perpetrator of the crime. If ridge detail is developed in a barefoot impression, the comparison to a suspect's foot can be carried out in exactly the same fashion as a fingerprint comparison [4, 5]. If enough ridge detail, with sufficient clarity, is available for comparison, a positive identification may be forthcoming. However, if the barefoot impression is smudged or unclear for any other reason, or if it is a socked impression, then recourse can be made to barefoot morphology comparison.

Similarly, when a shoe has been positively identified back to a crime scene, and no suspect has been found in possession of the footwear in question, recourse can again be made to barefoot morphology comparison. This can be accomplished by comparing the impressions on the insole inside the crime scene shoe to the impressions inside a similar shoe worn by the suspect, or to inked impressions and casts seized from the suspect.

The comparison itself is similar to a toolmark or tire track comparison, where the foot has acted like a tool or a tire in creating an impression at the crime scene. The shapes of various parts of the foot are compared to see if there is correspondence between the crime scene impression and the suspect exemplars. In a footwear example, the impression to be compared has been made on the insole of the identified footwear.



Forensic Barefoot Comparisons. Figure 1 Barefoot impressions taken from identical twins, illustrating the variability of footprints, even for twins.

Background

Although footprints in general look quite similar, it has long been assumed that careful examination of barefoot impressions could be used to differentiate between people. Historically, in various societies, trackers have been trained to be able to pick up someone's trail and to follow the person based on their footprints [3].

Footprint evidence was presented in court as early as the late nineteenth century, when a criminal was convicted in 1888 based on his footprint. Other cases have since been documented in the forensic literature, mainly from Europe and North America [6]. Much of this early casework was based on the assumption that

footprints were unique to the individual, without many studies to support this hypothesis.

Whenever a new means of including or excluding suspects is being introduced in court, the basis for the comparison must be justified. Some early work on footprint variability was carried out in India [7], while in North America, Dr. Louise Robbins, an anthropologist, carried out studies on the individuality of footprints in the 1970s [8]. In the 1980s, the Federal Bureau of Investigation (FBI) collected and compared footprint impressions from hundreds of volunteers to show how variable barefoot impressions might be [9, 10].

The Royal Canadian Mounted Police (RCMP) began research in this area in the 1990s [11]. Inked barefoot impressions were collected from thousands of volunteers for entry into a computerized database. As each footprint was measured and entered, it was compared with previous impressions in the database to ensure that another foot did not share the same measurements. A statistical analysis of the impressions was carried out to illustrate how variable barefoot impressions are [12]. Even with a limited number of samples and measurements, probabilities on the order of 1 in a billion were achieved.

Collecting Evidence [3]

When a bare or socked foot impression is found at a crime scene, the investigator must document the evidence correctly. Photographs, with a scale included, should be taken. If required, barefoot impressions can sometimes be enhanced *in situ* using fingerprint powder or chemical techniques, especially for impressions in blood. Impressions can then be lifted, or, depending on the surface, the entire impression can be removed from the scene. All of this evidence will have to be sent to the expert who is doing the final barefoot morphology comparison.

Similarly, when footwear impressions are found, they should be thoroughly documented. Again, enhancement techniques can be used to make the impressions more visible. If accidental characteristics are noted, there is the possibility of positively linking a shoe to the impression found at the crime scene.

If a suspect is arrested, his feet and his foot impressions must be well-documented. Several photographs, including a scale, should be taken of the feet to ensure that the tops, sides, and bottoms are all recorded. Foam

impressions should be taken for later casting. Inked standing and walking impressions should also be obtained. Standing and walking impressions should also be obtained with the suspect wearing a pair of socks.

If the case involves footwear impressions, attempts should be made to seize similar footwear from the suspect. The best comparison would be of a shoe insole with another shoe insole. However, the feet of the suspect should also always be recorded in the same manner as described above.

The Comparison Process [3]

Barefoot morphology comparison should only be undertaken by an adequately-trained specialist. The RCMP has conducted barefoot comparison courses in North America and Europe, training forensic specialists from several countries. A group of doctors has recently formed a forensic podiatry sub-committee within the International Association for Identification (IAI), currently establishing its own criteria for training and standards.

As in any other physical comparison, class characteristics are compared first. The overall size of the foot and the number of toes making contact with the ground would be considered class characteristics, and can be used to quickly eliminate a suspect foot.

The shape and placement of the toes, the shape of the ► metatarsal ridge, the length and width of the arch, and the contour of the heel can be examined and compared (see Fig. 2). Any unexplained feature can be used to eliminate a suspect, while correspondence of features means that the suspect foot remains included as a possible source of the crime scene impression. Some examiners will make positive identifications based on foot morphology [13], while others will only go as far as a strong likelihood that the same foot made both the crime scene print and the exemplar [3]. Because the uniqueness of barefoot impressions has not been proven, and crime scene and inked impressions are not exactly reproducible, current RCMP policy advises examiners against making positive identifications. Positive identifications may be possible when flexion creases, marks, and scars are visible [14].

When the impression is three-dimensional, like a footprint in mud, it is important to try to obtain a



Forensic Barefoot Comparisons. **Figure 2** Comparing the toes and the contour of the metatarsal ridge.

three-dimensional replica of the suspect's foot for comparison. The use of foam impression material and dental casting material should be considered.

In the case of footwear that has been positively identified back to the crime scene, and where footwear has been seized from the suspect, the outer areas of the shoe can be examined to confirm that wear on the shoes is similar. The barefoot impressions made on the insoles can then be compared in much the same fashion as the barefoot comparisons described above. The inside uppers of the footwear can also be examined to look for agreement or disagreement of wear and damage. If footwear cannot be seized from the suspect, then the examiner must make-do with inked impressions and casts of the suspect's bare feet, bearing in mind the changes in morphology caused by the foot being constricted in the shoe.

In Court

Barefoot morphology comparison is not yet a routinely-accepted forensic technique. As such, it is still vigorously challenged when it is presented in court. In Canada, the testimony has been successfully defended in Mohan and Voir Dire hearings [15], (*R v Dimitrov* was overturned because the jury put too much emphasis on "could have been" testimony, and not because of the technique itself) while in the US it has withstood the scrutiny of "Rule 702," Frye, and Daubert challenges. An early case that was sent back on appeal when the judge felt that not enough background research had been done has been successfully re-tried in light of more recent published research. In essence, the courts have started to

recognize the scientific foundation upon which the evidence is based.

Besides Canada and the United States, barefoot comparison testimony has been tendered in several countries around the world. In Israel, barefoot morphology comparison testimony was appealed all the way up to the highest court, and upheld. As the technique becomes more accepted, prosecutors and defense lawyers will soon start to look for opportunities where this type of evidence might be useful.

Conclusion

Barefoot morphology comparison refers to the comparison of the weight-bearing areas of feet in an effort to include or exclude a suspect as being linked to a crime scene. Background research has established the variability of barefoot impressions, justifying their use in a forensic context. Crime scene investigators should be aware of this technique and should always be looking for suitable evidence of this kind.

Related Entries

- ▶ Earprints
- ▶ Fingerprints

References

1. Bodziak, W.J.: *Footwear Impression Evidence*, pp. 381–411. CRC Press, Boca Raton, FL (2000)

2. Kennedy, R.B.: Bare footprint marks. In: Siegel, J.A., Saukko, P.J., Knupfer, G.C. (eds.) *Encyclopedia of Forensic Sciences*, pp. 1189–1195. Academic Press, London (2000)
3. Kennedy, R.B., Yamashita, A.B.: Barefoot morphology comparisons: a summary. *J. Forensic Ident.* **57**(3), 383–413 (2007)
4. Lemieux, M.: Histoire de Pied/a foot story. *Identif. Can.* **25**(4), 16–17 (2002)
5. Watkins, D., Brown, K.C.: The case of the toe print. *J. Forensic Ident.* **57**(6), 870–873 (2007)
6. McCafferty, J.D.: The shoe fits. *The Police J.* **28**(2), 135–139 (1955)
7. Puri, D.K.S.: Footprints. *Int. Police Rev.* **187**, 106–111 (1965)
8. Robbins, L.M.: The individuality of human footprints. *J. Forensic Sci.* **23**(4), 778–785 (1978)
9. Lovejoy, O.C.: Methods of Footprint Analysis. Seminar in Footprint and Shoeprint Identification. Federal Bureau of Investigation, 29 April 1984
10. Bodziak, W.J., Monson, K.L.: Discrimination of individuals by their footprints. Paper presented at 11th meeting of International Association of Forensic Sciences, Vancouver, BC (1987)
11. Kennedy, R.B.: Uniqueness of bare feet and its use as a possible means of identification. *Forensic Sci. Int.* **82**(1), 81–87 (1996)
12. Kennedy, R.B., Chen, S., Pressman, I.S., Yamashita, A.B., Pressman, A.E.: A large-scale statistical analysis of barefoot impressions. *J. Forensic Sci.* **50**(5), 1071–1080 (2005)
13. DiMaggio, J.A.: The Foot as a Forensic Tool. Paper presented at the 55th annual meeting of the American Academy of Forensic Sciences, Chicago, IL (2003)
14. Massey, S.L.: Persistence of creases of the foot and their value for forensic identification purposes. *J. Forensic Ident.* **54**(3), 296–315 (2004)
15. Richard, C.: Case law: [2002] Ontario superior court R. V. Arcuri. *Identif. Can.* **25**(4), 18–20 (2002)

Forensic DNA Evidence

T. HICKS, R. COQUOZ
Institut de police scientifique, Ecole des sciences criminelles, Lausanne, Switzerland

Synonyms

DNA analysis; DNA profiling; DNA typing

Definition

Deoxyribonucleic acid (DNA) is a large molecule present in all living cells (e.g., animals, plants, viruses). As a tape allows the storage of a recording, DNA allows the storage of genetic information. It consists of two long chains of

nucleotides twisted in a double helix. There are four types of nucleotides designed by the name of their bases: Adenine (A), Guanine (G), Cytosine (C) and Thymine (T). The genetic information is encoded in the sequence of nucleotides of the DNA molecule. Part of its name originates from its localization in the nuclei of the cell. However, the acronym is also used for ► **mitochondrial DNA** (mtDNA), which is the DNA present in the mitochondria of the cells. It is transmitted only by the mother.

Introduction

Although DNA profiling has high discriminating power and can help establishing the biological identity of a person, it is not used as a biometric yet. Indeed, the results of the analysis are not immediate (nor yet amenable to full automation), the cost of analysis is high, and there are contamination and transfer issues. Moreover, as parents transmit biological material to their children, DNA can be intrusive and reveal unknown family relationships.

Most of the text that follows is based on [1]. Another standard text in English is [2].

DNA: Basic Concepts

Each human cell contains biological information; each of the cells thus contains the same DNA. As DNA is a very long molecule (three billion base pairs), it is organised into 23 small bundles: the chromosomes. Each child receives two DNA: one from the mother and one from the father. Each person (and each cell from this person) has thus two copies of each chromosome, one from the mother and one from the father, giving a total of 46 chromosomes. They are numbered in pairs from 1 to 22, the 23rd pair being the sex chromosomes X and Y. On each of the chromosomes, there are genes (i.e., a zone where the DNA codes how to make a protein). The number of genes is estimated to be around 20,000–25,000. When reading DNA from one extreme to the other, one will encounter a code “START” to indicate that one can read from here how to make a given protein, and a code “STOP” to indicate that this is the end of the genetic information necessary for that protein. The code for the following protein does not begin immediately after the “STOP”

message and there are usually several thousands nucleotides in between that do not code genetic information. These are called noncoding DNA or junk DNA and represent 98% of the genetic material. The geographical distribution of the genes and the noncoding DNA is in principle identical across all individuals in a given species.

A geographical nomenclature has been derived to designate a given localization on the DNA strand of human chromosomes: the locus (pl. loci). For each locus, one copy is received from the father and one copy from the mother. These two copies are called alleles: therefore for each locus, a person will have two alleles. The set of alleles owned by a person for a locus is called his/her genotype. If the alleles transmitted by the father and the mother are the same, the individual is homozygote for this locus. If the two alleles transmitted are different, then the individual will be heterozygote at this locus. The loci are symbolized by a four-digit code, that is particularly useful for noncoding DNA. The first letter of the code is D (for DNA), the second is the chromosome number (1, ..., 22, X, Y), the third element indicates the sequence type (S: a unique sequence; Z a sequence that has several copies, at different localizations, on the same chromosome; F a sequence that is part of a family with similar sequences that are encountered on several chromosomes); the last digits are a unique number that generally correspond to the order in which the sequences were discovered. As an example, the locus D18S51 is a DNA region on chromosome 18, with the serial number 51. This locus is commonly used in forensics because of its polymorphism and is present in different commercial kits (e.g., SGMplus ABI; Powerplex 16 Promega).

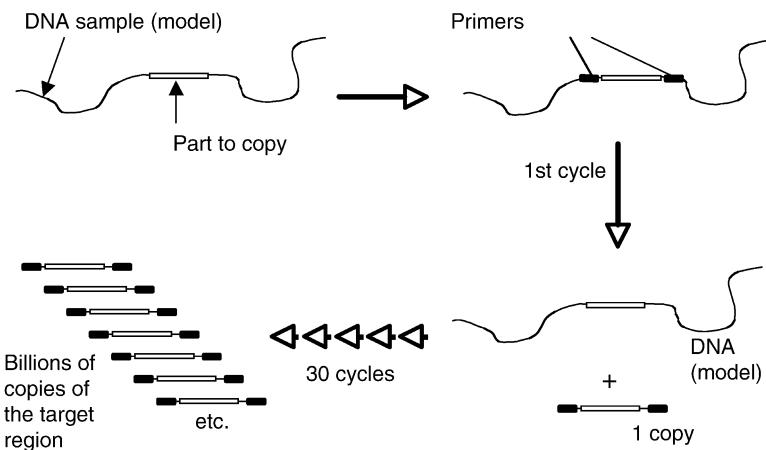
Since DNA is stable over the lifetime of an individual, and is reasonably resistant to chemical degradation, it is a good candidate for use in forensic science. One must further find a sensitive and not too expensive technique that enables the analysis of parts of the DNA that vary between individuals: the genetic markers. These markers should show as much as polymorphism as possible. There are two types of polymorphisms: sequence polymorphism (the nature of the nucleotides themselves differ from individual to individual, e.g., mtDNA, SNPs) and length polymorphism (the difference between individuals is based on the length of a given repetitive sequence of nucleotides, e.g., STR).

In 1985, thanks to the discovery of repetitive sequences, Sir Alec Jeffreys first applied the analysis of DNA to forensic science. As discussed previously, there are coding and non coding DNA. The repetitive sequences are zones in the DNA, where noncoding DNA seems to stutter. The human DNA has a very large number of different repetitive sequences (30% of the genome) and the length of the repetitive sequences varies between individuals. If the length of the repetitive sequence is larger than six nucleotides, one speaks of VNTR (Variable Number Tandem Repeats, Nakamura). If it is equal or smaller to six, one will speak of STR (Short Tandem Repeat). STRs have also been named microsatellites and VNTRs minisatellites. Nowadays, STRs are the standard targets of the routine analysis of forensic samples. Companies offer different type of kits allowing the analysis of several STRs at the same time (multiplex STR analysis). An example of some kits available is given below:

Forensic DNA Analysis

Before analysis, one has to sample the DNA on the crime scene and to sample the individual. As genetic information is theoretically the same in every cell, saliva is generally used for the later. On crime scenes, chemical tests to detect saliva, sperm, or blood can be used to help in finding the invisible stains. If a stain is detected, the object is either cut, or swabbed. As, it is possible to detect very low levels of DNA, it is highly recommended to wear gloves and a face mask when collecting DNA.

Once DNA has been collected, it will be extracted, purified, and quantified. Because there is often very little material in forensic samples, the specific zones of DNA that are the target of the analyses will be first amplified using a technique called Polymerase Chain Reaction (PCR). PCR is often compared to a DNA photocopier, where a given DNA segment is copied a given number of times. First, the zone to be amplified is delimited using two primers: one is placed at the beginning of the sequence and one at the end. During the copying cycles of the PCR process, copies of the original DNA zone will be produced. After one cycle, there will be the original DNA and one copy; after two cycles, there will be 2 more copies and after 30 cycles there will be about a billion copies (Fig. 1).



Forensic DNA Evidence. **Figure 1** From [1] representing the PCR Process.

Multiplex STR Analysis

PCR can be used to copy simultaneously several DNA fragments: one uses different primers, each corresponding to the DNA zone that has to be copied. This simultaneous amplification of different DNA zones is called Multiplex PCR. The PCR products are analyzed using Capillary Electrophoresis. This technique allows the separation of DNA fragments according to their sizes while they travel through a thin capillary. The large fragments move more slowly than small fragments. A peak is displayed on the results' graph, when the detector at the end of the capillary detects DNA molecules. The detector is able to differentiate between up to five different dyes. The detector can recognize STR alleles that have the same length but are labeled with different dyes.

Theoretically, one can amplify dozens of different fragments in one operation, which saves considerable time. However, the design of multiplex STR analysis kits is not straightforward: the numerous primers involved must not interfere with each other; the set of alleles of one STR must be recognized from the alleles of the other STRs, either through the use of different labeling dyes or because they are in different size ranges. The different STR will be chosen if possible on different chromosomes, so that the transmission of the alleles from generation to generation can be considered independent.

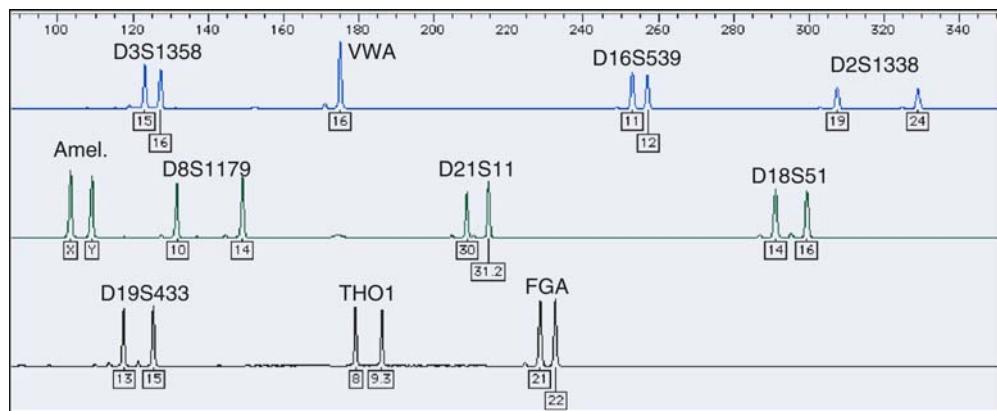
PCR is unable to amplify efficiently large fragments. It is thus not possible to increase indefinitely the number of STR analyzed simultaneously with a multiplex kit. The practical limit is around 15–20

STRs. An example of a profile obtained with Multiplex STR analysis is shown below. The vast majority of forensic DNA analyses done today is multiplex STR analysis. It is the golden standard and fulfills most of the needs of forensic DNA analysis.

Figure 2 shows the result of an analysis performed on an individual with the SGMplus kit.

Nonautosomal DNA

An autosome is a nonsex chromosome. Nonautosomal DNA is a DNA that originates either from a sex chromosome (i.e., X, Y), or from mitochondria (mtDNA). Sex chromosome DNA is used to study the paternal heritage: if there is no mutation, a son will have the same Y chromosome than his brothers, his father, and his paternal lineage; a daughter will have one of her X chromosomes identical to her father and her paternal lineage. Mitochondrial DNA is used to study the maternal heritage: a mother will transmit her mitochondrial DNA to her children, grandchildren, etc. These markers are useful for parentage testing. Since there are hundreds of copies of mtDNA per cell, mtDNA can also be used when the trace is degraded or/and there is very little nuclear DNA (e.g., hair without roots). STRY may also be useful in rape cases, when the male DNA of the aggressor is not detected in the DNA profile because of the presence of vast amounts of female DNA in the mixture. When speaking of alleles observed for Y chromosome STR or mtDNA and STRX, one speaks of haplotype and not genotype, as a person owns only one single allele.



Forensic DNA Evidence. Figure 2 Figure from [1] showing the result of an analysis performed on an individual with the SGM plus hit. The raw data are decomposed by color: the labels below the peaks show the designation of the detected alleles. She name above is the STR name. For most STRs, our individual has two peaks and is heterozygote for these makers. For VWA, this person presents only one peak and is therefore homozygote for that particular marker.

Low Template DNA Analysis

In essence, PCR allows a sensitivity at the single molecule level. However, for the standard DNA profiling, the potential of PCR is curbed to avoid the stochastic effects and contamination problems associated with an extreme sensitivity. The sensitivity of standard DNA profiling is thus adjusted to provide a DNA profile only when there is DNA from at least 50 cells in the evidence material. The concept of low template level analysis, or as previously known LCN (Low Copy Number), covers strategies designed to go beyond this limit. The usual strategy to reach such an extreme sensitivity is to increase the number of copying cycles of the PCR (e.g., 34 cycles instead of 28) [3]. Increasing sensitivity does not go without caveats: artefacts arise with alleles dropping in and dropping out. This added sensitivity also increases the issue of contamination. The use LCN profiles should induce very strict protocols to avoid contamination on the scene and at the laboratory. It requires cautious interpretation. In the Omagh Bombing, the use of low template level profiles was challenged, but was shown to be scientifically robust in the review conducted by Professor Caddy, Dr Linacre, and Dr Taylor [4].

SNPs and DNA Chips

SNPs (Single Nucleotide Polymorphisms) occur on average every 1,200 nucleotides: that means that

depending on the individual the nucleotide in that position will differ. In theory there might be four variants (A, T, G, C) for the nucleotide; however, because of the reality of the evolution process only two variants are usually observed. The main caveat of SNPs is their limited sequence polymorphism. Multiplex analysis of a large number of SNPs can however overcome this limitation. SNPs can be analyzed using miniaturized devices called DNA chips. These are miniaturized systems allowing the analysis of hundreds of SNPs. The disadvantages of SNPs are their low polymorphism, and their very limited capacity to handle DNA mixture cases. SNP analysis certainly has good prospects for specific applications. ► **Mitochondrial DNA** polymorphisms are in essence SNPs. SNP analysis has better chances of success for highly degraded samples than STR analysis. Some specific SNPs have the capacity to provide morphological information (hair, eye colour, ...) and Y chromosome SNPs should be able to provide useful ethnic information. Some companies (e.g., 23andme; Decode genetics; ...) have started to offer to the public wide ranging SNP analyses providing information on their ancestry, and predispositions to possible diseases (see, for example, <https://www.23andme.com/> or <http://www.decode.com/>). Although their approach is controversial and not forensically oriented, it is a good example of the power of the SNP analysis technologies. Applications of SNPs in context of forensic intelligence will be briefly mentioned later.

DNA Sequencing

This method consists in reading the sequence of the nucleotides of small pieces of DNA. It is the standard tool for mtDNA analysis. More recent “whole genome DNA sequencing” technologies are being developed. They are massively parallel sequencing approaches with the potential to provide the complete sequence of an individual in a single process. Their potential for forensic DNA analysis is not yet clear.

DNA used as Evidence and Interpretation

DNA (human, animal or vegetal) can be used in diverse areas: nature and species preservation, food control, missing persons, mass disaster identification, serious, and volume crime. As parents transmit part of their DNA to their children, parentage testing using DNA (autosomal and nonautosomal) is also common, whether for civil cases, historical cases (e.g., Thomas Jefferson, Nicolas II), immigration, or genealogy. DNA techniques are very sensitive and sources of DNA are diverse (e.g., blood, saliva, sperm, dandruff, skin, hair (nuclear DNA if root, mtDNA on shaft, bones, teeth...)), which explains the great potential of forensic DNA analysis.

DNA evidence can help the court in assessing three types of propositions [5]: offence level propositions (e.g., the suspect has raped the victim vs. the suspect has not raped the victim), activity level propositions (e.g., the suspect has had sexual intercourse with the victim vs. the suspect has not had sexual intercourse with the victim) and source level propositions (e.g., the recovered DNA comes from the suspect vs. the recovered DNA originates from a person unrelated to the suspect). The higher the hierarchy (from source to offence level), the more information will be required to inform an opinion. Through DNA profiling and other means, the forensic scientist is usually only able to provide useful evidence for source and activity levels. If addressing source level proposition, the scientist will take into account the rarity of the DNA profile, estimating the match probability (Weir and Evett), [6]. When addressing activity level propositions, the forensic scientist will in addition take into account transfer and persistence of DNA (or/and blood pattern in the presence of blood, see [7]), as well as the relevance of the trace to the alleged activities.

DNA interpretation is a very large subject area [8, 6, 9–11] and (NRC reports I and II). With the large number of polymorphisms available, DNA of an individual can certainly be considered as unique. Thus, DNA profiles are frequently viewed as unique in the general public, but they are not. Only a limited number of polymorphisms are examined in DNA profiling, providing DNA profiles that are indeed very rare, with match probabilities smaller than 1 in a billion. The value of DNA evidence is usually assessed using the likelihood ratio approach (also called the Bayesian approach). The evidence is evaluated considering two propositions (e.g., the prosecutor’s and the defence’s): in the given example, the forensic scientist would, for example, assess the probability of the evidence given that the suspect has had sexual intercourse with the victim and the probability of the evidence given that the suspect denies knowing the victim. This approach insures an unbiased interpretation of the evidence.

DNA Used as an Intelligence Tool

With the launching of DNA databases in the 1990s, DNA has become a very useful intelligence tool. Most countries have national DNA databases or are in the process of doing so. The largest are the USA database and the England and Wales DNA database. Each country has its own legislation and own set of STR loci (in Europe there is a set a “core” loci that are used by all EU countries). The most common program used for storing and searching the DNA profiles is CODIS (Combined DNA Index System). The program was developed by the FBI and the private firm SAIC. The profiles are stored in different indexes (e.g., forensic profiles index, offender index, victim index, staff index, and missing person’s relatives index). This allows comparing only profiles that should be (e.g., the profiles from relatives of missing persons will not be compared to crime scene profiles). Criteria for entering a profile in a database vary according to the country and the size of the database. DNA databases rely on the fact that the vast majority of crimes is committed by a small proportion of the population, that tends to re offend and on the fact that DNA profiles are extremely rare. This last prerequisite is not fulfilled with so-called partial DNA profiles (i.e., does not present results for all loci in the given kit because of DNA degradation), or with mixture DNA

Forensic DNA Evidence. Table 1 From [1] Table showing an example of multiplex kits available. Highlighted in dark gray are the markers chosen by the European Network of Forensic Science Institutes to warrant compatibility across countries; in light and dark gray the markers chosen for the USA combined DNA Index System (CODIS) database

Marker	Supplier	ABI	ABI	ABI	ABI	ABI	Promega	Promega	Promega	Serac	Serac	Biotype
Kit Name	Cofiler	Profiler	Profiler Plus	SGM Plus	Identifier	Sefiler	Powerplex 1	Powerplex 2	Powerplex 16	Powerplex ES	MPX2	MPX3
TH01	x	x		x	x	x	x	x	x	x	x	x
VWA		x	x	x	x	x	x	x	x	x	x	x
D21S11		x	x	x	x	x	x	x	x	x	x	x
FGA	x	x	x	x	x	x	x	x	x	x	x	x
D8S1179		x	x	x	x	x	x	x	x	x	x	x
D18S51		x	x	x	x	x	x	x	x	x	x	x
D3S1358	x	x	x	x	x	x	x	x	x	x	x	x
TPOX	x	x		x	x	x	x	x	x	x	x	x
CSF1PO	x	x		x	x	x	x	x	x	x	x	x
D13S317	x	x	x	x	x	x	x	x	x	x	x	x
D7S820	x	x	x		x	x	x	x	x	x	x	x
D5S818	x	x	x		x	x	x	x	x	x	x	x
D16S539	x		x	x	x	x	x	x	x	x	x	x
D2S1338		x	x	x	x					x	x	x
D19S433		x	x	x	x					x	x	x
Penta-D								x	x	x	x	x
Penta-E								x	x	x	x	x
SE33								x	x	x	x	x

profiles. That type of DNA profiles has to be used with more caution in conjunction of the database, especially, if the suspect population is large (i.e., the searched database is large). When there is no other forensic evidence than DNA to limit the suspect population, then the discriminating power of the technique must be higher, than if there are other means (e.g., partial fingerprints, modus operandi, micro-traces, and traditional police investigation information).

New approaches of the DNA database involve the use of partial profiles and familial searching for intelligence purpose. In general, to limit adventitious matches, partial profiles with less than six loci, for example, are generally not entered in the database, but they could be used to generate intelligence. Familial searching aims at helping the investigation when no match is found in the database. The technique consists in looking for profiles that share alleles with the crime scene profile. As it is more common for relatives to share part of their DNA than unrelated persons, there are examples where it was possible to find in the database a close relative of the offender. The Forensic Science Service (UK) has been able to solve a couple of famous cases using this method. There are ethical issues to consider when using this technique [12].

The analysis of some SNPs can predict physical characters (such as red hair, eye colour, ...) based on the analysis of the crime scene sample (see www.dnrintprint.com).

Conclusion

The advent of DNA analysis and DNA databases has revolutionised forensic science, police investigation and the whole criminal justice system. It is anticipated that automation will play tomorrow an even more important role than today. With the advent of ultra-sensitive methods, the relevance of the recovered material, the questions of transfer and persistence of DNA will become the core of interpretation. Today, the research in this area is still scarce and needs to be developed. Regarding the techniques, analysis of STR is here to stay for years. Other fascinating techniques for SNP analysis or even whole genome sequencing are coming. But they will not improve much the performance of DNA profiling and their use will require very difficult validations before they

can diffuse widely into the routine practice. As it was portrayed in the film "Gattaca" by Mike Nichols, 1997, it seems clear that DNA on a chip and DNA as a biometric system at the finger tip will be available one day. But that day is not yet at the horizon.

Related Entries

- LCN DNA/Low Template Level

References

1. Coquoz, R., Taroni, F.: Preuve par l'ADN –la génétique au service de la justice. Presses Polytechniques et Universitaires Romandes, Lausanne (2006)
2. Butler, J.M.: Forensic DNA Typing. Biology, Technology, and Genetics of STR Markers. 2nd edn. Elsevier Academic, Burlington, MA (2005)
3. Gill, P., Whitaker, J., Flaxman, C., Brown, N., Buckleton, J.: In investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. Forensic Science International **112**(1), 17–40 (2000)
4. Caddy, B., Linacre, G.R., Taylor, A.M.T.: A Review of the Science of Low Template DNA Analysis. (2008) doi: http://police.homeoffice.gov.uk/publications/operational-policing/Review_of_Low_Template_DNA_1.pdf?view=Binary
5. Cook, R., Evett, I.W., Jackson, G., Jones, P.J., Lambert, J.A.: A hierarchy of propositions: deciding which level to address in casework. Science Justice **38**, 231–240 (1998)
6. Buckleton, J., Triggs, C., Walsh, S.J.: Forensic DNA Evidence Interpretation. CRC, Boca Raton (2005)
7. Bevel, T., Gardner, R.: Blood Pattern Analysis with an Introduction to Crime Scene Reconstruction, 3rd edn. CRC, New York (2008)
8. Evett, I.W., Weir, B.S.: Interpreting DNA Evidence – Statistical Genetics for Forensic Scientists. Sinauer, Sunderland (1998)
9. Aitken, C.G.G., Taroni, F.: Statistics and the Evaluation of Evidence for Forensic Scientists, 2nd edn. Wiley, Chichester (2004)
10. National Research Council, Committee on DNA Technology in Forensic Science, Board on Biology, Commission on Life Sciences. DNA Technology in Forensic Science, National Academy Press, Washington, D.C. (1992)
11. National Research Council, Committee on DNA Forensic Science. The Evaluation of Forensic DNA Evidence. National Academy, Washington, D.C. (1996)
12. Williams, R., Johnson, P.: Forensic DNA Databasing: A European Perspective. Interim Report. University of Durham, Durham (2005)
13. Robertson, B., Vignaux, G.A.: Interpreting Evidence – Evaluating Forensic Science in the Courtroom. Wiley, Chichester (1995)

Forensic Evaluation of Fingerprints and Fingermarks

- Fingerprint, Forensic Evidence of

Forensic Identification Based on Dental Radiographs

- Dental Biometrics

Forensic Science

Forensic science refers to the applications of scientific principles and technical methods to the investigation of criminal activities, in order to establish the existence of a crime, to determine the identity of its author(s) and their modus operandi.

- Forensic Applications, Overview

Forensic Speaker Recognition

- Voice, Forensic Evidence of

Forgery Attempt

Active forgery attempt is an impostor attempt in which an individual tries to match the stored template of a

different individual by presenting a simulated or reproduced biometric sample, or by intentionally modifying his or her own biometric characteristics.

- Influential Factors to Performance

Forgery Sign

Synonyms

Forgery signature; Impostor sign; Mimicked sign

Definition

Forgery sign is an illegal sign by simulating or tracing a genuine signature. There are two kinds of forgery signs such as “substitution or random” and “freehand or skilled.” The former is called as “zero effort” forgery, because the forger uses his or her own signature instead of the signature to be tested. The later includes signatures imitated as closely as possible by simulating or tracing a genuine signature.

- Signature Matching

Forward-Backward Algorithm

The Forward–Backward algorithm is the conventional, recursive, efficient way to evaluate a Hidden Markov Model, that is, to compute the probability of an observation sequence given the model. This probability can be used to classify observation sequences in recognition applications.

- Hidden Markov Models

Fourier Transform

Mathematically, the continuous Fourier transform is one of the specific forms of the Fourier analysis.

It transforms the original function in the time-domain into another function in the frequency domain. The term “Fourier transform” can refer to either the frequency domain representation of a function or to the process/formula that transforms one function to another.

- ▶ Face Recognition, Component-Based
- ▶ Image Pattern Recognition
- ▶ Iris Encoding and Recognition Using Gabor Wavelets
- ▶ Iris Recognition Using Correlation Filters

Fovea

The fovea is a small depressed region at the center of the macula, the central area of the retina. There, the inner retinal layers are shifted aside, allowing light to pass unimpeded to the photoreceptors. Only tightly packed cones, and no rods, are present at the foveola, the center of the fovea. The elongated axons of these cone cell bodies are called Henle fibers. The fovea is the region of maximum visual acuity.

- ▶ Anatomy of Eyes

Fragile Bits

- ▶ Iris Template Extraction Via Bit Inconsistency and GRIT

Fraud Deterrence

- ▶ Fraud Reduction, Applications
- ▶ Fraud Reduction, Overview

Fraud Mitigation

- ▶ Fraud Reduction, Applications
- ▶ Fraud Reduction, Overview

Fraud Reduction, Applications

VICTOR MINCHIH LEE
International Biometric Group, New York, NY, USA

Synonyms

Biometric fraud reduction; Duplicate detection; Fraud deterrence; Fraud mitigation

Definition

Fraud is conventionally defined as the deliberate perversion or withholding of veracity in order to induce another to surrender something of value. In the context of biometrics, the item of value is typically an identity or a privilege associated with an identity.

For the purposes of this entry, fraud reduction in a biometric applications context refers to the use of biometric technology’s duplicate detection capabilities to deter, inhibit, and mitigate fraud. Duplicate detection refers to the discovery of multiple identities claimed by a single, given individual.

Introduction

For numerous decades, individuals have sought to misrepresent their identities for the sake of obtaining benefits and privileges to which they are not properly entitled. Such fraud can be costly – both financially and politically. For fiscal year 2007/2008, the United Kingdom’s Department for Work and Pensions estimated that it overpaid about £2.7 billion in housing-related benefits, alone, due to fraud and error [1]. From October 2002 to September 2005, the US Justice Department indicted 40 voters (21 noncitizens) for illegal voting or voter registration fraud [2].

In 2000, following the United States' presidential election, a study in Georgia, USA, discovered over 15,000 deceased individuals on the state's active voting rolls. The US Federal Election Commission also discovered 502,968 names on Alaska's 1998 voter rolls – yet only 437,000 eligible voters were estimated by the statewide census conducted that year [3]. In both cases, several thousand invalid, but influential, votes could have been cast in close elections by individuals assuming others' identities or fake identities.

In the past, such fraudulent actions were enabled by the tendency to ascertain identity based upon documentation with little – if any – connection to the distinctive characteristics of the legitimate document holder, aside from often replaceable or forgeable photographs. Authenticity of transactions was assured more by anti-forgery, document-oriented techniques (such as watermarks, holographs, security strips, microlines, intaglio printing, etc.) rather than by examination of the document bearer.

The advent of biometrics, however, has enabled a shift of focus from predominantly documents to a mix of documents and individuals. In 1858, the United Kingdom's William Herschel of the Civil Service of India was precocious in his decision to capture employee palmprints to help distinguish amongst his native Indian staff on paydays [4]. Today, automated biometric capture and processing systems allow for quick determinations and verifications of identity. They enable the detection of individuals who may assume multiple nominal identities through various documents, but who are really the same, single entity.

This duplicate detection capability deters and inhibits fraud in applications including:

1. Benefits issuance and disbursement
2. Voter registration
3. Visa shopping
4. Border control
5. Consumer recognition and
6. Time and attendance monitoring

This entry introduces and provides examples of the first four aforementioned applications. Consumer recognition (including check cashing) and time and attendance monitoring are both addressed in other entries.

Benefits Issuance and Disbursement

Governments are often responsible for the proper and equitable distribution of benefits to their qualified citizenry. With large populations of potential benefits recipients, however, it can be a logistics challenge to keep track of who is a qualified recipient and whether or not they have previously claimed a given benefit and are attempting illegitimately to reclaim the same benefit (a phenomenon sometimes referred to as “► double dipping”).

Biometrics can help mitigate the problems associated with such challenges. Initially, when biometrics are first captured and associated with a given identity in an enrollment process, biometrics are particularly vulnerable and dependent on the legitimacy and robustness of ► breeder documents. But once a ► nominal identity has been paired with a biometric, a government can be relatively certain that whenever that biometric is presented, it is presented by the individual with that same nominal identity and not an imposter. This is because of the fundamental assumption and belief that biometrics based on individual physiological and behavioral characteristics are more difficult to steal and forge than are documents.

One example of the benefits that can ensue is the deployment of fingerprint recognition technology by the United States of America's Texas Health and Human Services Commission (HHSC). HHSC sought to ensure that Texas' limited Medicaid benefits be distributed only to the truly needy. HHSC wanted to make sure that it was paying for services actually rendered and delivered only to those authorized to receive Medicaid benefits. By leveraging fingerprint biometrics, HHSC sought to confirm that authorized Medicaid recipients were indeed physically at treatment facilities when Medicaid benefits were disbursed [5].

Another example of biometrics applied to fraud reduction in benefits issuance is the Andhra Pradesh Ration Card Entitlements program in India. In this deployment, first announced on 16 June 2005, iris recognition systems were employed in the issuance of food ration cards by the state of Andhra Pradesh [6]. Through the incorporation of biometrics in the program, Andhra Pradesh officials seek to deter its citizenry from selling or sharing their food ration cards, as well as returning to claim multiple cards under different nominal identities.

Voter Registration

The validity of elections critically depends on ensuring that only legitimate voters vote and that the general democratic principle of “one voter, one vote” is followed. As with benefits issuance, biometric systems can help in determining who is voting, that they are authorized to vote, and that they are not voting, or registering to vote, multiple times. Biometrics can also act as a fraud deterrent by, for example, facilitating the forensic identification of a person who attempts to vote using a deceased individual’s credentials.

One example of biometrics in a voter registration application is the Bangladeshi Voter Registration Project. This project, run by the Bangladesh Army and Bangladesh Elections Commission, utilized fingerprint biometric technology to register voters for Bangladesh’s 2008 general elections and to issue national identity cards. Four fingerprints were captured from each registrant and checked to see if they matched those captured from a prior registrant [7].

The Bangladeshi Voter Registration Project follows similar voter registration deployments conducted in countries like Mexico, Mozambique, and Nigeria. In Mexico, the Instituto Federal Electoral implemented a multi-biometric system that uses fingerprint and facial recognition to analyze historical voter rolls for duplicates, as well as to vet new voters against existing voter rolls [8].

Visa Shopping

In 2007, the European Parliament recognized a challenging problem facing several European Union member states: visa applicants rejected by a Schengen country were applying to other Schengen countries in the hopes of finding one that would issue them a visa. They were “visa shopping.” In some cases, applicants would present forged documents as part of the visa application process.

To counter such attempts, the European Parliament established the Visa Information System (VIS). VIS is a database that contains fingerprint and face images and associates the collected biometrics with visa applicants’ biographical data, as well as the dates and locations of application attempts [9]. Authorized officials responsible for border security can now better detect if a visa applicant has previously been rejected

by another Schengen nation and is presenting falsified documents.

One result of the implementation of VIS has been that some visa applicants who have been rejected previously, or who have reason to believe they may be rejected, have mutilated their own fingers to avoid being processed against VIS. Others have attempted to perpetuate fraud by trying to alter their fingerprints using often painful processes with low chances of success, given that fingerprints extend beyond the epidermis.

Border Control

As with visa shopping deployments, border control deployments seek to protect national borders by determining with greater certainty who is entering (and, sometimes, exiting) a nation. These biometric applications help deter and expose entrance document fraud and identity fraud. They differ from visa shopping deployments insofar as they are oriented more towards identity and credential verification at the arrival and departure stages, rather than at the registration or application stages.

One example of a border control fraud reduction effort using biometrics is the United Arab Emirates’ Iris Expeltee Tracking System, deployed in 2003 and run by the Abu Dhabi Police. United Arab Emirates (UAE) officials were concerned about foreigners expelled from the UAE subsequently attempting to reenter the country after changing their name and/or nationality and then obtaining a new passport. The Iris Expeltee Tracking System involves collecting iris images from all expelled foreigners. Arriving Passengers have their irises scanned at the UAE borders to verify that they were not formerly expelled. In 2005, the UAE reported catching approximately 32,850 previously expelled individuals [10].

Another program, the Canadian Passenger Accelerated Service System (CANPASS), uses iris recognition to allow preapproved, low risk travelers to clear Canadian customs and immigration without having to present documentation to border officials. Applicants who are approved to participate in CANPASS have their irises enrolled into the system. They then present their irises at designated kiosks at participating border environments (e.g., airports) for quick passage into Canada. This system not only helps reduce the

chance of forged documents passing inspection due to human error, it also allows border control officers to focus more on persons of greater interest who are more likely to be potential fraudsters.

Related Entries

- ▶ Asset Protection
- ▶ Binding of Biometric and User Data
- ▶ Biometric Encryption
- ▶ Consumer Recognition
- ▶ Forgery Sign
- ▶ Fraud Reduction, Applications
- ▶ Liveness and Anti-Spoofing
- ▶ Spoofing
- ▶ Time and Attendance

References

1. UK Department for Work and Pensions, Information Directorate: Fraud and Error in the Benefit System: October 2006 to September 2007 (May 2008)
2. Urbina, I.: Voter ID Battle Shifts to Proof of Citizenship, New York Times. <http://www.nytimes.com/2008/05/12/us/politics/12vote.html?ref=opinion> (12 May 2008). Accessed 2 Sept 2008
3. Samples, J.: The Motor Voter Act and Voter Fraud, CATO Institute. <http://www.cato.org/testimony/ct-js031401.html> (14 May 2001). Accessed 2 Sept 2008
4. Palm Print Recognition, National Science and Technology Council. <http://www.biometrics.gov/Documents/PalmPrintRec.pdf>. Accessed 2 Sept 2008
5. Front End Medicaid Fraud Reduction Pilot Program Based on Biometric Front-End System, Texas Health and Human Services Commission. http://www.hhsc.state.tx.us/OIE/RFP/FrontEnd/FingerImaging_RFI.pdf (30 June 2003). Accessed 2 Sept 2008
6. LGE Iris Tech Win in India Redefines Biometric Scalability. Findbiometrics.com. <http://www.findbiometrics.com/article/115>. Accessed 2 Sept 2008
7. BIO-key and Tiger IT Bangladesh Voter Registration Project Nearing Completion, BIO-key International, Inc. <http://www.reuters.com/article/pressRelease/idUS121192+17-Jun-2008+PRN20080617> (17 June 2008). Accessed 3 Sept 2008
8. The Mexican Instituto Federal Electoral (IFE): Mexico deploys multi-biometric voting system. Biometric Technol. Today. 14(5), 3–4 (2006)
9. KableNet: EU aims to stop ‘visa shopping’, The Register. http://www.theregister.co.uk/2007/06/08/schengen_visa_data/ (8 June 2007). Accessed 3 Sept 2008

10. Lieutenant Mohammad Al-Mualla: The UAE Iris Expellees Tracking and Border Control System. http://www.biometrics.org/bc2005/Presentations/Conference/2%20Tuesday%20September%2020/Tue_Ballroom%20B/Lt.%20Mohammad%20UAE2005.pdf. Accessed 4 Sept 2008
11. Study Report on Biometrics and E-Authentication. INCITS/M1 Ad Hoc Group on Biometrics in E-Authentication (2006)
12. Anderson, K.: Consumer Fraud in the United States: Second Survey. Federal Trade Commission, Washington, DC (2007)
13. Lee, V.: Transcript of Authentication Technologies FTC Proof Positive Workshop Session. Federal Trade Commission, Washington, DC (2007)
14. Progress in Tackling Benefit Fraud: Thirty-first Report of Session 2007–2008, UK House of Commons, Committee of Public Accounts. Accessed 2 June 2008

Fraud Reduction, Overview

VICTOR MINCHIH LEE

International Biometric Group, New York, NY, USA

Synonyms

Biometric fraud reduction; Fraud deterrence; Fraud mitigation; Identity theft reduction

Definition

Fraud is conventionally defined as the deliberate perversion or withholding of veracity to induce another to surrender something of value. In the context of biometrics, the item of value is typically an identity or a privilege associated with an identity.

Fraud can assume a variety of forms ranging from phishing to scams to hacking. In the specific case of biometrics, fraud can also consist of spoofing, or the presentation of an artifact designed to imitate a legitimate biometric.

Fraud reduction in a biometric context entails both the use of biometric technology to deter, inhibit, and mitigate fraud, as well as efforts to counter the exploitation of biometric system vulnerabilities through illegitimate submissions.

Introduction

With the increased reliance of modern society on technology, fraudsters have developed new exploitative techniques to prey upon the unsuspecting and the vulnerable. Internet merchants offer wares at hard-to-resist discounts, but never deliver their patrons any of the purchased goods. Sophisticated counterfeiters create fake currency and checks that are frequently difficult to distinguish from their genuine counterparts. Identity thieves send realistic, yet illegitimate, emails designed to harvest passwords and identity information from the careless and the inexperienced.

The cost of fraudulent activity, which includes tangible sums lost and expenditures to recover stolen goods, identities, or privileges, can be enormous. In 2005, British insurer Norwich Union estimated the cost of fraud in UK to be around £16 billion, or roughly 1.4% of UK's gross economic output (<http://news.bbc.co.uk/1/hi/business/4463132.stm>). In 2006, this number soared to £40 billion according to conservative UK government estimates (<http://www.timesonline.co.uk/tol/news/uk/article633540.ece>). The same year, identity theft alone victimized over 8 million people in the US to the tune of more than US \$49 billion, according to the California Office of Privacy Protection. Unfortunately, loss due to fraudulent activities is likely to increase as fraudsters expand their reach thanks to increased globalization and rapid development of technology.

Technology, however, can be used both to abet, as well as combat, fraud. Biometric technologies and systems, especially, enable the deterrence, inhibition, and mitigation of fraud. For the purpose of this discussion, biometrics are defined as technologies that perform automated measurement of human physiological or behavioral characteristics to determine or authenticate identity.

Biometrics, thus, revolve around the concept of identity. Identity, in turn, is often seen as a proxy for trustworthiness, whether through linkage of identity to a historical and/or transactional record, or through connection of identity to a privilege or right.

Trustworthiness is the first victim of any fraudulent act. Security, robustness, and confidence of identity are therefore critical. Biometrics, if employed judiciously and with appreciation for the limitations and vulnerabilities of the technology, can satisfy such crucial needs.

Biometrics Vis-à-Vis Alternative Authentication/Identification Technologies

Biometric Advantages in Fraud Reduction

Currently several non-biometric methods and technologies exist to help combat fraud by serving an ► authentication or ► identification function. Examples of such technologies include smart cards, tokens, fobs, passwords, and personal identification numbers (PINs). Generally, these technologies can be divided into two categories: those based on something one has and those based on something one knows.

Biometrics introduces a third, complementary aspect: authentication and identification technologies based on something one *is*. By focusing on the elements that are inherent to an individual, biometrics offer additional protections that are unavailable or weaker through more traditional authentication/identification technologies. These include:

- Convenience
- Accountability
- Security

Convenience

Biometrics can obviate the carrying of tokens or cards that can be lost, misplaced, or – more saliently – stolen, leading to fraudulent access or unauthorized transactions. Biometrics can also eliminate the need to remember passwords or PINs. Often, people select simplistic passwords that can be easily guessed or hacked because they fear that they will forget more complex passwords.

Losing or forgetting a biometric, however, is considerably more difficult, especially for biometrics primarily dependent on physiological, rather than behavioral characteristics. Whereas one can suffer brain damage and forget how to sign one's name, rendering signature recognition useless, misplacing or forgetting a finger that is integrally tied to the rest of one's body is harder to accomplish. While violent criminals can go to extremes to remove a finger from a target so as to gain access to a fingerprint-secured facility, it is markedly more challenging to trick a target into unwittingly giving up such a personal feature.

Accountability

Biometrics are excellent technologies when transferability is of concern. Instead of relying on force or compulsion, fraudsters achieve success through convincing, coercion, and deception designed to encourage victims to surrender, or provide access to, a privilege, right, or item of value. Sophisticated scams, for example, can lead victims consciously and willingly to hand over precious access cards or passwords to perpetrators of fraud.

Biometric characteristics, however, are distinct and very personal. Their transference from one individual to another can be, as mentioned earlier, extremely challenging. This can contribute markedly to accountability, in addition to deterring and inhibiting fraud. If it is difficult for a fraudster to trick an individual into giving up their biometric, then any action taken that can be linked to that biometric is likely to have been undertaken by the legitimate possessor of the biometric in question. This makes it difficult to believe excuses in which a misdeed was allegedly committed by another who fraudulently obtained one's biometric characteristics.

With more traditional authentication/identification technologies, however, transferability can translate into reduced accountability. One fraudster could borrow access cards or passwords that allow him to take advantage of services or privileges intended for another. There could even be complicity in this effort – something that would take a high degree of personal sacrifice if biometrics were involved.

Biometrics can also add an element of accountability by deterring and inhibiting fraudulent attempts at establishing or relying on multiple identities. In the past, for instance, certain fraudsters with notorious histories of cashing bad checks would assume several identities so as to avoid the stigma and troubles accompanying their negative transactional histories. The introduction of biometric technologies and systems, however, has helped identify and address problems of multiple registrations by linking personal, biometric characteristics, rather than just nominal identities, to transactional histories and other historical records. This has also aided in the combating of fraudulent acts including multiple civil ID registrations and visa shopping.

Additionally, in some cases, biometrics deter and mitigate acts of fraud by encouraging or necessitating

the leaving behind of distinct, personal characteristics. For example, some prospective culprits may think twice before acting if they are aware that their criminal and fraudulent activity could potentially result in the leaving behind of biometric markers, such as their latent fingerprints. Those who proceed anyway and ignore the concern of enrolling in a biometric system and leaving behind an image or template of a distinct, personal characteristic could possibly be identified later and tracked by the biometrics they previously presented, a potential advantage for law enforcement and means of mitigating the severity and impact of a fraudulent act (e.g., by catching fraudsters before they are able to take advantage of the captured item of value or privilege).

Security

As described above, biometrics offers security and anti-fraud advantages over more traditional authentication/identification technologies with respect to identity transference, establishment of multiple fake identities, and loss or forgetting of credentials. They can render certain fraudulent activities – like phishing – almost irrelevant.

Biometric characteristics also provide additional anti-fraud security benefits thanks to their inherent nature; compared with passwords/PINs and cards/fobs/tokens, biometric characteristic is generally more difficult to capture, steal, replicate, and fake. Cards, for instance, are often designed to be robust, yet flexible enough that, in case they are lost, a replacement can be relatively easily created. PINs can be sniffed out through tracking or hidden monitoring technologies. They can also be readily discovered, in several cases, through brute force and trial-and-error techniques. Replication of a compromised PIN is then no more complicated than re-entering the newly revealed PIN.

It can be challenging, however, to create a replica of a biometric characteristic that has sufficient enough fidelity to work with a targeted biometric system. Creating a plausible fake iris, for example, often requires more effort than just copying electronic data onto a new smart card or retying a password (in which cases the artifact will be identical to the genuine sample). This is due in part to liveness detection, a security function that is built into several biometric systems. Liveness detection, a fraud countermeasure, deters or

inhibits the presentation of artifacts, called spoofs, as legitimate biometric characteristics. Examples of live-ness detection include: measurement of finger perspiration over time, 2D Fourier spectrum analyses, and behavioral reactions to cues (e.g., blinking upon command).

In addition, several biometric systems rely on templates, rather than full images of biometric characteristics, for reasons that range from privacy to cost to efficiency of data management and processing. Attempting to regenerate or reverse-engineer a complete biometric image from a select template is a very challenging, if not, at times, outright impossible, task. Also, trying alternative, brute-force techniques to recreate a biometric characteristic could take extremely lengthy and impractical periods of time, given the vast number and variability of components that make up many biometric characteristics.

Furthermore, biometric systems can often be costly, expensive, and technologically complex. Spending large sum of money to obtain a biometric device for study and identification of vulnerabilities and penetration points may not be cost effective. Likewise, even those who have the resources and know-how to create fake biometric characteristics, however, may find the effort of doing so to be cost-inefficient, especially when the value of the item or privilege being protected is outweighed by the cost of fraudulently obtaining it.

In some of the Panasonic's US offices, for example, hand geometry biometric readers are employed for time and attendance functions vis-à-vis custodial staff. Though several special effects and novelty item firms have the ability to create fake hand models, the cost and effort entailed in obtaining a suitable spoof would probably exceed the financial return of an extra hour of pay.

Authentication/Identification Trifecta

While biometrics offer significant advantages over more traditional and conventional authentication/identification technologies, it is important to note that this does not mean that biometrics should be employed *in lieu of* these other technologies. When issues of fraud, as well as security and protection of identity, are at stake, it may be optimal to leverage all proven options, especially given the potentially high cost of fraud and the ease with which fraud can often be committed.

Also, as will be discussed in the following section, biometrics have their own inherent vulnerabilities. These potential weaknesses can sometimes be mitigated by adopting complementary technologies which can provide an extra – if not necessarily equally effective – layer of defense. Where fraud is involved, the need for security may outweigh convenience and cost; such scenarios encourage reliance on an authentication trifecta that consists of:

- Something you *have*
- Something you *know*
- Something you *are*

Biometric Vulnerabilities

While biometric technologies can prove to be relatively robust and effective tools in fraud reduction through deterrence, inhibition, and mitigation, biometric systems themselves are not immune to fraudulent and exploitative attacks. These attacks can be classified according to three overarching categories:

- Input level attacks
- Processing and transmission level attacks
- Backend and storage level attacks

Input Level Attacks

Input Level Attacks generally fall into one of three categories:

- Spoofing attacks
- Bypassing attacks
- Overloading attacks

Spoofing attacks consist of attempts to deceive biometric system sensors into accepting an artifact as a legitimate biometric sample, typically for false enrollment, verification, or identification purposes. Spoofing attacks are usually considered to be attempts at breaking into biometric systems that are predominantly physiological in their focus. Biometric systems that are predominantly behaviorally based revolve less around the creation of spoof items and more around careful observation and practiced imitation of legitimate behavior.

Bypassing attacks consist of attempts to circumvent biometric system processes by creating artificial

Examples of spoof types for five established biometric modalities include:

Fingerprint	Face	Iris	Hand geometry	Voice recognition
Prostheses	Prostheses	Prostheses	Prostheses	Audio playback recordings
Props/Models/Gag items	Masks/Disguises	Video playback recordings	Props/Models/Gag items	Audio composite recordings
Photograph imitations	Photograph imitations	Photograph imitations		
Residual prints		Imprinted contact lenses		
Latent prints				

failures during enrollment or recognition so as to skip the biometric system altogether. One example of a bypassing attack would be to alter the quality of a biometric characteristic in such a way that a biometric system has difficulty in acquiring that characteristic. This could, for example, entail artificially filing down fingerprints so that there is a failure to enroll. The risk, as a result, is that an individual could then possibly be excused from biometric system recognition requirements and permitted to use a less robust authentication/identification system.

Closely related to bypassing attacks are variants called overloading attacks. In an overloading attack, a fraudster attempts to defeat or circumvent a biometric system by damaging or overwhelming the biometric sensor(s). Overloading attacks can range from flashing strobe lights against an optical sensor to presenting artificial heat sources to near-infrared-based sensors to short circuiting of sensitive sensors using liquids. As with bypassing attacks, the goal of an overloading attack is to either reduce the robustness, precision, and accuracy of the targeted biometric system and/or to encourage the substitution of the biometric recognition method with a less robust authentication/identification process and mechanism.

Processing and Transmission Level Attacks

Processing and transmission level attacks generally fall into one of three categories:

- Hacking
- Skimming/Sniffing
- Hill-Climbing

Processing and transmission level attacks are, strictly speaking, lesser acts of deception and fraud and more direct, technically based invasions. However, the result of success in any such attack on a biometric system could enable future acts of fraud, so it is important to be aware of these potential vulnerabilities.

Hacking, as herein defined, consists of electronically based attempts to penetrate a biometric system by altering the operation and functionality of the system through non-physical modifications and subterfuge (often at the code or system communications levels). A hacker could change the enrollment or recognition algorithms of a biometric system, lowering thresholds to accommodate less robust performance and security checks. They could program the system to forward them the copies of legitimate samples or instruct the system to allow them special, otherwise unauthorized, access.

Skimming and sniffing refers to techniques by which data is captured – often surreptitiously – during communication or processing of the information. Skimming devices, for example, could be designed to read and copy biometric data being submitted on a smart card to a biometric system for comparison against a live sample. This data could then be illegally replicated. Sniffing could occur if monitoring programs are put in place to capture data packets being sent from the capture sensor to the backend for verification.

Hill-climbing attacks first consist of the presentation of a test biometric sample to a biometric algorithm for comparison against an enrolled sample. A match score is then obtained and studied so that a new test sample can be presented for re-comparison and the achievement of a higher match score. This process is re-iterated until the biometric system's threshold has been discovered and is penetrable.

Backend and Storage Level Attacks

Backend and storage level attacks generally fall into two categories:

- Infiltration
- Implantation

As with the process and transmission level, the backend and storage levels are susceptible to malicious hacking. Skilled hacker-fraudsters could alter the permission levels tied to specific images or templates stored in databases. They could infiltrate the backend and alter the way biometric data that is classified and stored. More of concern, they could perhaps steal biometric characteristics data and try to generate spoofs using the information captured.

In addition, acts of fraud can be facilitated if fraudsters are able to gain unauthorized or complicit access to backend and storage databases of biometric information to perform acts of implantation. In this attack, fraudsters might implant their own biometric characteristics into a targeted biometric system's database. By doing so, fraudsters would be able to appear as legitimately authorized individuals with free access to the rights or privileges otherwise secured by the biometric system.

Countermeasures

In order to counter – or at least inhibit – the three aforementioned types of attacks, certain countermeasures can be enacted. These countermeasures can be classified according to the level of attack they are best suited to address.

At the input level, spoofing is typically counteracted by the attempt to determine whether a live, real human sample is being presented to the capture device. This is, as mentioned earlier, called liveness detection and is based on the assumption that, with the exception of some cadaver recognition applications, a legitimate biometric will always be presented by the live possessor of that biometric characteristic.

As for bypassing and overloading attacks, countermeasures include increased ruggedization of capture devices and sensor equipment, conscientious form factor design (e.g., creating shielding from external light sources that could be potentially malevolent), supervision of enrollment and recognition submission

processes, as well as rigorous fallback procedures and processes. After all, those who seek to accomplish fraud will often target the weakest link. If this means taxing a biometric system out so that, for example, access to a secure facility can be obtained through a potentially more fallible human guard inexperienced at identifying fake identity documents, which will often be the strategy of choice for motivated fraudsters.

At the processing and transmission levels, countermeasures may entail proven information systems and information technology security techniques, such as the use of firewalls and encryption. After all, at a certain level, biometric data is often converted into streams of digital data that should be accorded no less than the rudimentary security protections already commonplace for digital information that is less personally sensitive. In addition, best practices should be implemented, such as requiring the use of data transmission shields (that limit the range at which data can be sniffed from contactless smart chips), as well as strict limitation of access to matching score data.

At the backend and storage levels, highly advisable countermeasures include firewalls, as well as extensive auditing functions and logs of modifications executed (whether they are additions, subtractions, or alterations of biometric data). A best practice countermeasure would also be the frequent, though not necessarily habitual or scheduled, review of random images and templates for evidence of tampering, alteration, missing presence, or unexpected presence.

In order to further deter or inhibit the abuse of biometric systems by fraudsters, biometric systems may also be designed with the following four countermeasures:

- ► **Multifactor** or ► **multimodal** authentication requirements
- Randomization of modality
- System challenges
- Emphasis on internal/subcutaneous characteristics

By adopting multifactor or multimodal systems, deployers increase the challenge for fraudsters by requiring them to defeat several disparate systems for which the optimum exploitation and penetration techniques may be very different. Though there is a convenience tradeoff, the security that accrues can be significant, particularly when security of identity is at stake. The main caveat, however, is that potential fraudsters are not encouraged to pretend to be unable to use one of the modalities so as to simplify their, say, spoofing task.

Examples of liveness detection methods for five established biometric modalities include:

Fingerprint	Face	Iris	Hand geometry	Voice recognition
Spectroscopic analysis	Reactivity to Cues, Commands, and stimuli (e.g., – blinking)	Photonic and spectrographic analysis	Required contact with specifically- placed prongs	Recitation of randomly generated passphrases
Temporal variation in perspiration		Reactivity to stimuli (e.g., – pupil dilation)		
		Ink/Dye detection		
		Timestamping and byte scrambling		

To provide a little more balance between convenience and security, a multimodal system could still be employed, but only with one or two randomly selected biometric modalities required for authentication/identification. Variations could also be introduced within a single modality (e.g., requiring submission of a right index finger, one day, and submission of a left thumb on the next day).

The authentication/identification systems could also be designed to issue randomized as well as cued challenges – even if the original submission would otherwise have been acceptable. At the very least, this implementation would provide the opportunity to obtain two biometric samples. Where the samples are unusually similar, extra caution might be merited in case a spoof is involved, as the likelihood that a person will be as infallible as to place their biometric so consistently is slim.

Finally, biometric systems can be deployed and designed so as to focus on internal or subcutaneous characteristics that are generally much more difficult to capture surreptitiously, as well as to forge or modify.

Biometrics and Fraud: Looking to the Future

While a lot of focus has been placed on the design and utilization of biometrics to deter, inhibit, and – to a lesser degree – mitigate non-biometric fraud, comparatively little attention has been paid to the consequences and implications when fraudsters are successful in compromising biometric data. Because biometric characteristics are so intrinsic and relatively immutable, this is an issue of particular concern that can impact the successful deployment of the technologies.

Cancelable/Changeable Biometrics

To address concern over the immutability of biometrics, research has been conducted by entities like IBM and the Korean Biometrics Engineering Research Center into cancellable biometrics, also known as changeable biometrics. The high-level concept of such research has been to look into altering biometric data that is captured before it is actually fully processed and stored in template form. In this way, a compromised biometric can theoretically be revoked and a new algorithm can generate a novel distortion of the affected individual's biometric characteristic – essentially giving them a new biometric.

However, one should keep in mind that if a fraudster is able to get hold of the original source biometric characteristic (or an equivalent spoof), this approach would not suffice, as the fraudster would then still be able to regenerate a new cancelable/changeable biometric characteristic just as easily as the legitimate bearer of the original source biometric.

Nominal Identities Versus Biometric Identities

As biometric systems increasingly protect sensitive data and items or access privileges of high value, the incentive fraudulent activity to exploit them will also increase. And at some point, as the case has been with virtually every major security technology in the past, biometric data will be compromised.

One of the important ways in which the impact of such compromise can be mitigated is to sever, whenever feasible and reasonable, the connection between

an individual's nominal identity and their biometric identity. If, for instance, a deployment merely requires a determination as to whether a given individual, represented by their biometric characteristics, should be granted access to a given secure location, then there is no need to link permanently the individual's name and background information to their biometric data after an initial background check has been conducted.

Whereas names have often served as proxies for trustworthiness or transactional histories, biometrics can now serve this purpose going forward. With biometrics there is also the possibility of selecting different biometric aspects for accreditation or validation given each distinct application or deployment. A single biometric characteristic, thus, has the flexibility to serve in a variety of functions that process that biometric differently – without making that biometric into a universal identifier rife with the problems of overuse that have plagued the US social security number.

In the scenario described above, if a fraudster compromises one biometric system, the damage is mitigated insofar as other systems and deployments may still be protected, in addition to sensitive and private information tied to one's nominal identity.

Valuing Biometric Data

One of the remaining challenges with respect to biometrics and fraud is the determination of how to value biometric data. This is especially important as fraudsters increasingly target not just data protected by biometrics, but biometric data, itself.

Traditionally, items have been valued based on three factors:

- Scarcity
- Uniqueness
- Demand

With biometrics, however, such a framework for assessing value is of little use: virtually each and every given biometric characteristic is inherently distinct (if not unique), scarce, and of high demand for both the possessor and potential imposters/fraudsters. It would seem, therefore, that all biometric characteristics should be deemed priceless or at least assigned extremely lofty values.

However, this would be impractical in an age of risk calculations and need by insurance companies, governments, and other entities realistically to quantify the impact and cost of fraud. Therefore, valuation of a biometric is best conducted according to a different set of three factors:

- Value of the Biometrically-Protected Item or Privilege
- Range of Utility
- Spoofability

In addition, whenever a biometric system is designed, careful consideration needs to be taken as to whether templates or images should be used. Generally, images will be more valuable from perspectives concerned with forensics, interoperability, and scalability. Templates, however, will be more desirable from an identity-protecting perspective. Thus, from a fraud reduction perspective, the guiding principle should be that templates, which are more limited than images, normally, should be employed whenever possible in lieu of images. To achieve this balance, a negative incentive should be implemented such that there will be stiffer legal penalties for compromised biometric image data versus biometric template data.

Related Entries

- ▶ Asset Protection
- ▶ Binding of Biometric and User Data
- ▶ Biometric Encryption
- ▶ Consumer Recognition
- ▶ Forgery Sign
- ▶ Fraud Reduction, Applications
- ▶ Liveness and Anti-Spoofing
- ▶ Spoofing
- ▶ Time and Attendance

References

1. Study Report on Biometrics and E-Authentication. INCITS/M1 Ad Hoc Group on Biometrics in E-Authentication (2006)
2. Anderson, K.: Consumer Fraud in the United States: Second Survey. Federal Trade Commission, Washington, DC (2007)
3. Lee, V. et al.: Transcript of Authentication Technologies FTC Proof Positive Workshop Session. Federal Trade Commission, Washington, DC (2007)

4. Fraud and Error in the Benefit System: October 2006 to September 2007. UK Department for Work and Pensions, Information Directorate (May 2008)
5. Progress in Tackling Benefit Fraud: Thirty-first Report of Session 2007–2008. UK House of Commons, Committee of Public Accounts (2 June 2008)
6. Counting the cost of UK fraud, BBC News, 24 November 2005. <http://news.bbc.co.uk/1/hi/business/4463132.stm> (29 September 2008)
7. Woolcock, N.: Cost of fraud spirals to £40bn, TimesOnline, 9 September 2006. <http://www.timesonline.co.uk/tol/news/uk/article633540.ece> (29 September 2008)
8. Ratha, N. et al.: Cancelable Biometrics: A Case Study in Fingerprints (2006)
9. Unisys. Consumers Worldwide Overwhelmingly Support Biometrics for Identity Verification, 26 April 2006. Press release (30 September 2008)
10. Use of Biometric Identification Technology to Reduce Fraud in the Food Stamp Program, United States Department of Agriculture, Food and Nutrition Service, December 1999. <http://www.fns.usda.gov/oane/MENU/Published/fsp/FILES/ProgramIntegrity/biomeval.htm> (30 September 2008)
11. Identity Fraud: Prevalence and Links to Alien Illegal Activities, United State General Accounting Office, 25 June 2002. <http://www.gao.gov/new.items/d02830t.pdf> (30 September 2008)
12. Martinez-Diaz, M. et al.: Hill-Climbing and Brute-Force Attacks on Biometric Systems: A Case Study in Match-on-Card Fingerprint Verification, http://fierrez.ii.uam.es/docs/2006_ICCST_HillClimbingAttackMoC_Martinez.pdf (30 September 2008)
13. Schuckers, S. et al.: Issues for Liveness Detection in Biometrics. http://www.biometrics.org/html/bc2002_sept_program/2_bc0130_DerakhshabiBrief.pdf (30 September 2008)

Freeman Chain Code (FCC)

Freeman Chain Code (FCC) is a compact method for representing the contours of an object, first made popular by Herbert Freeman.

► Hand Data Interchange Format, Standardization

Function Creep

This refers to the use of data beyond the purposes originally intended at the time of data collection. For biometrics, this usually means using the data for

purposes other than identification, as when a face image is used to determine gender or ethnicity.

► Privacy Issues

Fundamental Frequency, Pitch, F0

The fundamental frequency or F0 is the frequency at which vocal chords vibrate in voiced sounds. This frequency can be identified in the sound produced, which presents quasi-periodicity, the pitch period being the fundamental period of the signal (the inverse of the fundamental frequency). Pitch is more often used to refer to how the fundamental frequency is perceived.

► Speech Analysis

Fusion Network Topology

The network architecture including sensors, communication channels, and fusion processing. The fusion processing may be distributed due to physical constraints in the system. If a communication channel between two sensors is long, it may be beneficial to fuse all the sensors at one end of the channel so that only the single fused decision is sent through the channel. The topology is directly impacted by the physical layout of the sensor network.

► Fusion, Decision Level

Fusion, Biometric

See Multi-biometrics.

► Biometric Algorithms
 ► Multibiometrics and Data Fusion Standardization
 ► Multiple Experts

Fusion, Confidence Level

- Fusion, Score-Level

Fusion, Data Level

- Fusion, Sensor-Level

Fusion, Decision-Level

LISA OSADCIW, KALYAN VEERAMACHANENI
Syracuse University, Syracuse, NY, USA

Synonyms

Distributed detection; Distributed inference making;
Multiple classifier fusion; Statistical signal processing

Definition

Decision level fusion falls under a broader area known as distributed detection systems and is the process of selecting one hypothesis from multiple M hypotheses given the decisions of multiple N sensors in the presence of noise and interference. In biometrics, decision level fusion creates a single decision from typically two hypotheses, imposter or genuine user, from multiple biometric sensor decisions, which may or may not be identical sensors. Often, decision level fusion is implemented to save communication bandwidth as well as improve decision accuracy. A statistical performance model for each biometric sensor is needed a priori to support the system wide optimization in terms of two error rates: false acceptance rate, admitting an imposter, and false rejection rate, rejecting the genuine user. A weighted sum of these two errors is a useful objective function. This provides the designer with the flexibility to weigh one error more than the other error. Decision level fusion may be done at one processor, centrally, or at multiple processors, distributed.

Introduction

In biometric decision level fusion, the biometric sensors send their final decisions through a communication network that finally fuses these decisions at a fusion center. Optimal decision fusion theory can be applied to these problems. In distributed detection systems, the number of decisions a sensor can make varies as well as the ► **fusion network topology**. It may be more advantageous to fuse a few sensors at a local node before transmitting the information over a long distance to the final fusion processor [1]. This complicates the fusion problem by introducing different fusion network topologies. Decision level fusion remains at the foundation of the problem, however.

The decision level fusion problem in the biometric area is typically one in identifying the user as a genuine user or an imposter with the final decision made by a central fusion processor [2–5]. This is referred to as a ► **parallel fusion network**. The advantages of fusion are twofold. The first advantage is a more accurate final decision by using multimodal, multiple and diverse, biometric sensors, which provide significantly more information to base a decision. Secondly, communication bandwidth needs, which are great as more sensors are networked, remain relatively constant if the decisions instead of the full observation or measurement are communicated.

The fusion accuracy of the sensor decisions relies on the accuracy of the statistical models for the sensors and an optimally designed fusion rule. The biometric verification problem may be posed as a ► **binary hypothesis** testing problem with the match score(s) serving as observations. The two hypotheses are

$$\begin{aligned} H_0: & \text{Imposter Identified} \\ H_1: & \text{Genuine User Identified} \end{aligned}$$

Probability of false alarm,

$$P_{\text{FA}} = P(u = 1|H_0) \quad (1)$$

and probability of false rejection,

$$P_{\text{FR}} = P(u = 0|H_1) \quad (2)$$

In the Bayesian formulation, these two errors are weighted by costs and summed into a single cost function called the Bayesian risk function. The Bayesian risk function is

$$R = P(H_0) \times C_{\text{FA}} \times P_{\text{FA}} + P(H_1) \times C_{\text{FR}} \times P_{\text{FR}}, \quad (3)$$

where $P(H_0)$, a priori probability of an imposter, $P(H_1)$ a priori probability of a genuine user C_{FA} , cost of false acceptance, and C_{FR} , cost of false rejection. In the worst-case scenario, one assumes equal a priori probabilities. Thus, we get

$$R = C_{\text{FA}} \times P_{\text{FA}} + C_{\text{FR}} \times P_{\text{FR}}. \quad (4)$$

We can rewrite Eq. 4, using a single cost factor by replacing the cost of false acceptance by

$$C_{\text{FA}} = 2 - C_{\text{FR}}. \quad (5)$$

This simplifies the problem to one design parameter to optimize if the a priori probabilities of genuine users and imposters are assumed to be fixed.

Decision Level Fusion with Single Bit Information

Often Gaussian distribution functions are used as the statistical sensor models for a binary hypothesis problem. Each sensor has a different Gaussian distribution function as shown in Fig. 1 for each hypothesis: genuine user and imposter. Higher observation values are typically associated with a positive user identification or the genuine user hypothesis. This leads to the distribution on the right side of the plot in Fig. 1. The imposter has a lower mean. Sensor 1 must measure a score that exceeds a threshold for comparison purpose to decide if it has the

genuine user. The error rates are simply the areas under the distribution corresponding to the opposite hypothesis or wrong side of the threshold. False acceptance probability is the area to the right of the threshold under the imposter distribution. False rejection probability is the area to the left of the threshold under the genuine user distribution. A single bit of 1 denotes that the user is detected while the 0 is for the imposter [7].

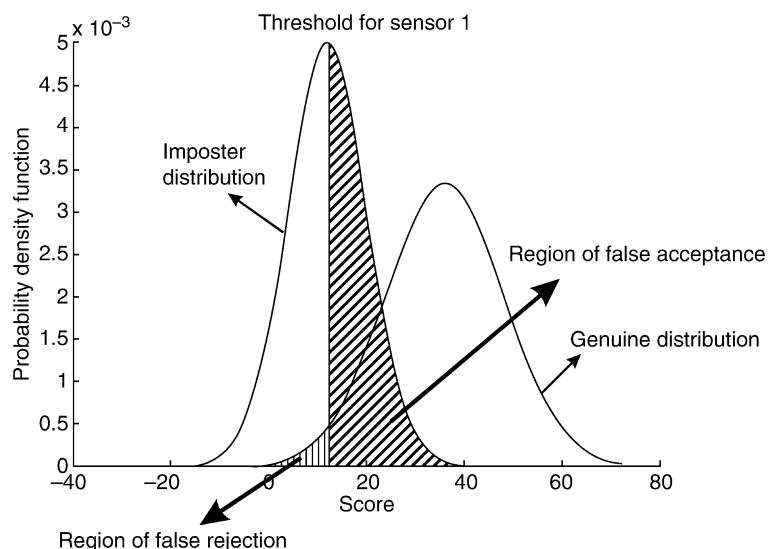
The threshold divides the entire decision region into region of acceptance and region of rejection. If a user's matching score happens to fall above the thresholds, he/she is considered as genuine. If the user's matching score falls below threshold he/she is considered as imposter [1, 2, 5, 6]. This process using the threshold, λ_i , for sensor i is given by

$$u_i = \begin{cases} 1, & y_i \geq \lambda_i \quad \forall i \\ 0, & y_i < \lambda_i \end{cases} \quad (6)$$

Let $[u] = [u_1, u_2, \dots, u_n]$, be the combined vector of decisions represented by 1s and 0s for all the sensors. These decisions are combined using a fusion rule of

$$u_f = f([u]). \quad (7)$$

The complete set of fusion rules for the 2-sensor case is given in Table 1 [1]. There are 16 possible rules for 2 sensors or $(2^2)^N$ with N sensors as 2. The fusion rule can be written as a 4-bit vector, where each bit represents the final fused decision given the



Fusion, Decision-Level. Figure 1 Illustration of thresholding process, Decision Regions, and Error Regions, for given Gaussian conditional density functions.

Fusion, Decision-Level. **Table 1** All Possible Fusion Rules for 2 Sensors

$u_1 u_2$	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}
0 0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0 1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1 0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1 1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

sensor decisions, $[u_1 \ u_2]$. Since, there are two hypotheses, the fusion rules are based on Boolean logic. For example, in **Table 1**, f_2 is a rule based on AND logic. The final decision is “1” only when both the sensors say “1” and is “0” otherwise. Similarly, f_9 is “NAND”, and f_8 is “OR”.

For 3 sensors, there are 8 possible vectors requiring a fused decision. Hence, the fusion rule is 8 bits long, and the number of possible fusion rules for this problem is $(2^2)^3$ or 64 rules for 3 sensors.

The error probabilities as in **Eqs. 1** and **2** for the entire system after fusion is estimated using

$$GP_{FA} = \sum_{[u]} P(u_f = 1 | [u], H_0) P([u] | H_0). \quad (8)$$

Assuming independence, **Eq. 8** can be calculated using the statistical models and

$$\begin{aligned} GP_{FA} &= \sum_{[u]} P(u_f = 1 | [u], H_0) \prod_{i=1}^n P(u_i | H_0) \\ &= \sum_{[u]} P(u_f = 1 | [u], H_0) \prod_{i=1}^n \int_{y_i} P(y_i | H_0) dy_i \end{aligned} \quad (9)$$

In case of correlation [8], however, the product disappears resulting in a multivariate integral or

$$\begin{aligned} &= \sum_{[u]} P(u_f = 1 | [u], H_0) \left(\int_{y_1} \int_{y_2} \dots \int_{y_n} f_{Y_1, Y_2, \dots, Y_n} \right. \\ &\quad \left. (y_1, y_2, \dots, y_n | H_0) dy_1 dy_2, \dots, dy_n \right) \end{aligned} \quad (10)$$

Similarly,

$$GP_{FR} = \sum_{[u]} P(u_f = 0 | [u], H_1) P([u] | H_1). \quad (11)$$

Assuming independence, (11) can be calculated using

$$GP_{FR} = \sum_{[u]} P(u_f = 0 | [u], H_1) \prod_{i=1}^n P(u_i | H_1). \quad (12)$$

In case of correlation, the multivariate integral arises as before giving

$$\begin{aligned} &= \sum_{[u]} P(u_f = 0 | [u], H_1) \left(\int_{y_1} \int_{y_2} \dots \int_{y_n} f_{Y_1, Y_2, \dots, Y_n} \right. \\ &\quad \left. (y_1, y_2, \dots, y_n | H_1) dy_1 dy_2, \dots, dy_n \right). \end{aligned} \quad (13)$$

This multivariate integral can only be calculated using numerical methods. Since there are 2^N combinations of local decisions for N sensors, this integral must be evaluated $2^N - 1$ times to estimate each error. This operation can be very expensive computationally as the number of sensors increases. An alternative is using the Bahadur–Lazarfeld expansion, which enables the estimation of the error probabilities using “ $n-1$ ” evaluations of integrals [8].

The Bayesian risk function is now given by,

$$R = P_0 C_{10} P(u_0 = 1 | [u], H_0) + P_1 C_{01} P(u_0 = 0 | [u], H_1). \quad (15)$$

Optimal Fusion Rule

For independent sensors, however, the optimal fusion rule is the ► likelihood ratio test [5]. For fixed thresholds, the optimal fusion rule can be obtained by using the likelihood ratio as in

$$\begin{aligned} \frac{P(u_1, u_2, u_3, \dots, u_n | H_1)}{P(u_1, u_2, u_3, \dots, u_n | H_0)} &> \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}. \\ u_0 = 1 & \\ u_0 = 0 & \end{aligned} \quad (16)$$

Fusion, Decision Level. **Table 2** Means and Standard Deviations of the Gaussian Distributions Under both the Hypothesis for Different Sensors

Hypothesis/Parameter	H_0/μ_0	H_0/σ_0	H_1/μ_1	H_1/σ_1
Sensor 1	47.375	43.864	144.514	12.843
Sensor 2	67.755	52.633	251.209	23.008

Fusion, Decision-Level. **Table 3** Thresholds for the 3 Sensors for Single Bit Information

Sensor	Threshold	FAR	FRR
1	95.945029260756	0.13409060569213	0.00007790554000
2	185.799498919171	0.01245661704014	0.00223574300430

$P(u_1, u_2, u_3, \dots, u_n | H_h)$ in (16) can be replaced by $\prod_{i=1}^n \int_{y_i} P(y_i | H_h) dy_i$ in case of independence or $\int_{y_1} \int_{y_2} \dots \int_{y_n} f_{Y_1, Y_2, \dots, Y_n}(y_1, y_2, \dots, y_n | H_h) dy_1 dy_2, \dots, dy_n$ in case of correlation. Optimal fusion rule can be employed when the thresholds are fixed. The optimal fusion rule as in Eq. 16 minimizes the Bayesian risk function.

In the case of independence [5], the optimum rule simplifies to

$$\begin{aligned} & \sum_{i=1}^N \left[u_i \log \left(\frac{1 - F_{FR}}{F_{AR}} \right) + (1 - u_i) \log \left(\frac{F_{FR}}{1 - F_{AR}} \right) \right] \\ & u_f = 1 \\ & > \log \left(\frac{C_{FA}}{2 - C_{FA}} \right). \quad (17) \\ & u_f = 0 \end{aligned}$$

Independent Pair of Biometric Sensors

Consider two sensors with conditional distributions under both the hypotheses given by the familiar Gaussian distribution. A Gaussian distribution is characterized by $N(\mu, \sigma)$ with a different mean, μ , and standard deviation, σ , for each hypothesis as mentioned earlier. Table 2 gives the parameters of the Gaussian distributions used for the 2 sensors in this example. In Table 3, the threshold that achieves the false alarm rate and false rejection rate given is specified for both sensors. Using these thresholds as well

Fusion, Decision-Level. **Table 4** Optimal Fusion Rule Under Assumption of Independence

CFA	Optimal Fusion Rule
0.2	Majority Voting Rule
0.6	Majority Voting Rule
1	(1 OR 2) AND 3
1.2	(1 OR 2) AND 3
1.5	(1 OR 2) AND 3
1.9	(1 OR 2) AND 3

as the error rates in the optimal fusion rule of Eq. 17, we give the rules in the right column for the specified costs in the left. Thus, different rules become optimum as the error rates are weighted differently. If the sensor is replaced with a more accurate biometric sensor, the rule selection will change. Finally, if the sensors are correlated, the original rule in Eq. 16 must be applied and performance computed accordingly.

Related Entries

► Multiple Experts

References

- Varshney, P.K.: Distributed Detection and Data Fusion, Springer. Springer-Verlag, New York, Inc (1997)

2. Prabhakar, S., Jain, A.: "Decision-level Fusion in Fingerprint Verification", *Pattern Recognit.* **35**, 861–874 (2002)
3. Osadciw, L., Varshney, P., Veeramachaneni, K.: "Optimum Fusion Rules for Multimodal Biometric Systems", Foresti, G.L., Regazzoni, C.S., Varshney, P.K. Chap. 15, *Multisensor Surveillance Systems: The Fusion Perspective*, Kluwer (2003)
4. Veeramachaneni, K.: "An Evolutionary Algorithm Based Dynamic Thresholding for Multimodal Biometrics" Masters thesis, School of Electrical and Computer Engineering, Syracuse University (2003)
5. Chair, Z., Varshney, P.K.: "Optimal Data Fusion in Multiple Sensor Detection Systems", *IEEE Trans. Aerosp. Electron. Syst.* **22**(1), 98–101 (1986)
6. Tang, Z.-B., Pattipati, K.R., Kleinman, D.L.: "An Algorithm for Determining the Decision Thresholds in a Distributed Detection Problem", *IEEE Trans. Syst. Man Cybern.* **21**, 231–237 (1991)
7. Veeramachaneni, K., Osadciw, L., Varshney, P.: «Adaptive Multimodal Biometric Management Algorithm », *IEEE Trans. Syst. Man Cybern.* **35** (2005)
8. Kam, M., Zhu, Q., Gray, W.S.: "Optimal Data Fusion of Correlated Local Decisions in Multiple Sensor Detection Systems," *IEEE Trans. Aerosp. Electron. Syst.* **28**, 916–920 (1992)

Introduction

Feature level fusion is an example of an early fusion strategy, i.e., the biometric evidence from multiple sources are consolidated *before* invoking the matcher. In this scheme, multiple feature sets are integrated in order to generate a single template that is expected to be more robust than the individual feature sets. When the feature sets to be integrated are homogeneous (e.g., multiple measurements of a person's hand geometry), a single feature vector can be computed as a weighted average of the individual feature sets. When the feature sets are nonhomogeneous (e.g., features of different biometric modalities like face and hand geometry), they can be concatenated to form a single feature set. Feature selection schemes are employed to reduce the dimensionality of the ensuing feature set [1]. Concatenation is not possible when the feature sets are incompatible (e.g., fingerprint minutiae and eigen-face coefficients).

If the feature sets to be combined originate from the same feature extraction algorithm (thus, a single modality is assumed) then feature level fusion can be used for template update or template improvement as discussed in the following section.

1. *Template update:* The template in the database can be updated based on the evidence presented by the current feature set in order to reflect (possibly) permanent changes in a person's biometric. Hand geometry systems use this process to update the geometric measurements stored in the database in order to account for changes in an individual's hand over a period of time. A simple scheme would be to take the average of the two feature vectors corresponding to the two instances of the biometric signal and use the average feature vector as the new template (Fig. 1).
2. *Template improvement:* In the case of fingerprints, the minutiae information available in two impressions can be combined by appropriately aligning the two prints and removing duplicate minutia thereby generating a larger minutia set. This process, known as template improvement, can also be used to remove spurious minutiae points that may be present in a feature set. While template update is used to accommodate temporal changes in a person's biometric, the purpose of template improvement is to increase the number of features (*and* decrease the number of spurious features) in the template.

Fusion, Feature-Level

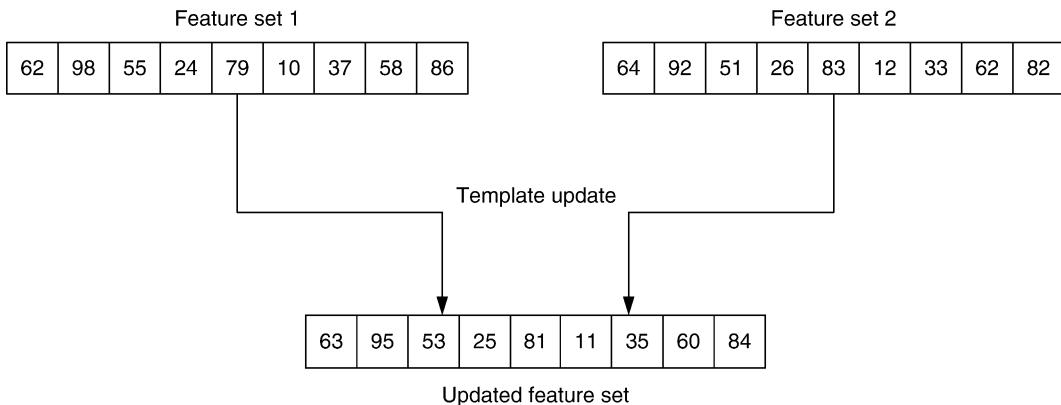
ARUN ROSS
West Virginia University, Morgantown, WV, USA

Synonym

Feature Fusion

Definition

In feature-level fusion, the feature sets originating from multiple biometric sources are consolidated into a single feature set by the application of appropriate feature normalization, transformation, and reduction schemes. The primary benefit of feature-level fusion is the detection of correlated feature values generated by different biometric algorithms thereby identifying a compact set of salient features that can improve recognition accuracy. Eliciting this feature set typically requires the use of ▶ **dimensionality reduction** methods and, therefore, feature-level fusion assumes the availability of a large number of training data. Feature-level fusion algorithms can also be used for template update or template improvement.



Fusion, Feature-Level. [Figure 1](#) A template update procedure may be viewed as a feature fusion scheme. In this example, the nine-dimensional feature set of a user (“Feature Set 1”) is updated based on the evidence presented by the current feature set (“Feature Set 2”), via the averaging scheme.

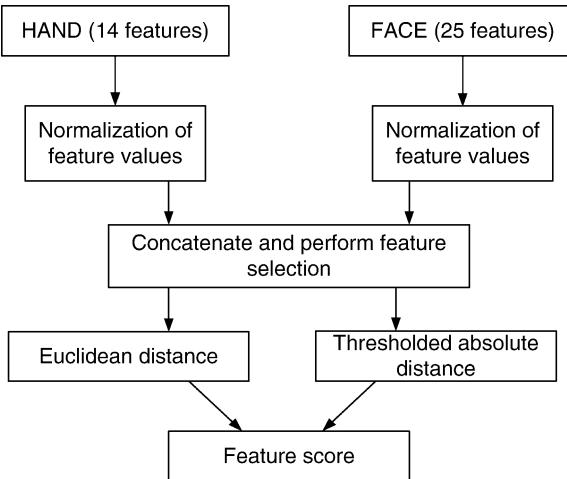
Several template improvement algorithms have been discussed in the literature for fingerprints. Jiang and Ser [2] propose a template improvement scheme where a reliability measure is associated with each extracted minutia point. This reliability measure is updated as minutiae evidence from newly acquired impressions is made available. The parameters of a minutia point (i.e., its x-y location and orientation) are updated via a weighted average scheme; even the “type” of the minutiae (i.e., ridge-ending or ridge-bifurcation) is altered if necessary. Template improvement is applicable only when the new fingerprint impression is accurately aligned with the stored one. The authors use the match score to determine if two impressions are accurately aligned. During the verification stage, only those minutia points whose reliability measure is above a certain threshold are used in the matching process. The authors show that their scheme results in (1) the elimination of spurious minutiae points, (2) the addition of missed minutiae points, (3) the relabeling of incorrect minutiae types and, consequently, (4) a general improvement in matching performance. Other algorithms for minutiae template improvement have been discussed in [3, 4].

Feature Fusion Scheme

How does one consolidate feature sets originating from different algorithms and modalities? Feature level

fusion is difficult to achieve in such cases because of the following reasons:

1. The relationship between the feature spaces of different biometric systems may not be known.
 2. The feature sets of multiple modalities may be incompatible. For example, the minutiae set of fingerprints and the eigen-coefficients of face are irreconcilable. One is a variable length feature set (i.e., it varies across images) whose individual values parameterize a minutia point; the other is a fixed length feature set (i.e., all images are represented by a fixed number of eigen-coefficients) whose individual values are scalar entities.
 3. If the two feature sets are fixed length feature vectors, then one could consider augmenting them to generate a new feature set. However, concatenating two feature vectors might lead to the ► curse-of-dimensionality problem [5] where increasing the number of features might actually degrade the system performance especially in the presence of small number of training samples. Although the curse-of-dimensionality is a well known problem in pattern recognition, it is particularly pronounced in biometric applications because of the time, effort and cost required to collect large amounts of biometric (training) data.
 4. Most commercial biometric systems do not provide access to the feature sets used in their products. Hence, very few biometric researchers have



Fusion, Feature-Level. [Figure 2](#) The procedure adopted by Ross and Govindarajan [1] to perform feature level fusion.

focused on integration at the feature level and most of them generally prefer fusion schemes that use match scores or decision labels.

If the length of each of the two feature vectors to be consolidated is fixed across all users, then a feature concatenation scheme followed by a dimensionality reduction procedure may be adopted. Let $X = \{x_1, x_2, \dots, x_m\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ denote two feature vectors ($X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$) representing the information extracted from two different biometric sources. The objective is to fuse these two feature sets in order to yield a new feature vector, Z , that would better represent an individual. The vector Z of dimensionality k , $k < (m + n)$, can be generated by first augmenting vectors X and Y , and then performing feature selection or feature transformation on the resultant feature vector in order to reduce its dimensionality. The key stages of such an approach are described as follows (also see [Fig. 2](#)).

Feature Normalization

The individual feature values of vectors $X = \{x_1, x_2, \dots, x_m\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ may exhibit significant differences in their range as well as form (i.e., distribution). Augmenting such diverse feature values will not be appropriate in many cases. For example, if the

x_i 's are in the range [0,100] while the y_i 's are in the range [0,1], then the distance between two augmented feature vectors will be more sensitive to the x_i 's than the y_i 's. The goal of feature normalization is to modify the location (mean) and scale (variance) of the features values via a transformation function in order to map them into a common domain. Adopting an appropriate normalization scheme also helps address the problem of outliers in feature values. While a variety of normalization schemes can be used, two simple schemes are discussed here: the min–max and median normalization schemes.

Let x and x' denote a feature value before and after normalization, respectively. The min–max technique computes x' as

$$x' = \frac{x - \min(F_x)}{\max(F_x) - \min(F_x)}, \quad (1)$$

where F_x is the function which generates x , and $\min(F_x)$ and $\max(F_x)$ represent the minimum and maximum of all possible x values that will be observed, respectively. The min–max technique is effective when the minimum and the maximum values of the component feature values are known beforehand. In cases where such information is not available, an estimate of these parameters has to be obtained from the available set of training data. The estimate may be affected by the presence of outliers in the training data and this makes min–max normalization sensitive to outliers. The median normalization scheme, on the other hand, is relatively robust to the presence of noise in the training data. In this case, x' is computed as

$$x' = \frac{x - \text{median}(F_x)}{\text{median}(|(x - \text{median}(F_x))|)}. \quad (2)$$

The denominator is known as the Median Absolute Deviation (MAD) and is an estimate of the scale parameter of the feature value. Although, this normalization scheme is relatively insensitive to outliers, it has a low efficiency compared to the mean and standard deviation estimators. Normalizing the feature values via any of these techniques results in modified feature vectors $X' = \{x'_1, x'_2, \dots, x'_m\}$ and $Y' = \{y'_1, y'_2, \dots, y'_n\}$. Feature normalization may not be necessary in cases where the feature values pertaining to multiple sources are already comparable.

Feature Selection or Transformation

Augmenting the two feature vectors, X' and Y' , results in a new feature vector, $Z' = \{x'_1, x'_2, \dots, x'_m, y'_1, y'_2, \dots, y'_n\}$, $Z' \in \mathbb{R}^{m+n}$. The curse-of-dimensionality dictates that the augmented vector of dimensionality ($m + n$) need not necessarily result in an improved matching performance compared to that obtained by X' and Y' alone. The feature selection process is a dimensionality reduction scheme that entails choosing a minimal feature set of size k , $k < (m + n)$, such that a criterion (objective) function applied to the training set of feature vectors is optimized. There are several feature selection algorithms in the literature, and any one of these could be used to reduce the dimensionality of the feature set Z' . Examples include sequential forward selection (SFS), sequential backward selection (SBS), sequential forward floating search (SFFS), sequential backward floating search (SBFS), “plus l take away r ” and ▶ branch-and-bound search (see [6, 7] for details). Feature selection techniques rely on an appropriately formulated criterion function to elicit the optimal subset of features from a larger feature set. In the case of a biometric system, this criterion function could be the Equal Error Rate (EER); the d-prime measure; the area of overlap between genuine and impostor training scores; or the average GAR at predetermined FAR values in the ROC/DET curves corresponding to the training set (see [1]).

Dimensionality reduction may also be accomplished using feature *transformation* methods where the vector Z' is subjected to a linear or a nonlinear mapping that projects it to a lower dimensional subspace. Examples of such transformations include the use of principal component analysis (PCA), independent component analysis (ICA), multidimensional scaling (MDS), Kohonen Maps, and neural networks [8]. The application of a feature selection or feature transformation procedure results in a new feature vector $Z = \{z_1, z_2, \dots, z_k\}$ which can now be used to represent the identity of an individual.

Examples of Feature Level Fusion

Ross and Govindarajan [1] discuss feature level fusion as applied to three different scenarios: (1) multialgorithm, where two different face recognition algorithms based on Principal Component Analysis (PCA) and

Linear Discriminant Analysis (LDA) are combined; (2) multisensor, where the three different color channels of a face image are independently subjected to LDA and then combined; and (3) multimodal, where the face and hand geometry feature vectors are combined. The general procedure adopted in [1] is summarized as follows.

1. Let $\{X_i, Y_i\}$ and $\{X_j, Y_j\}$ be the feature vectors obtained at two different time instances i and j . Here, X and Y represent the feature vectors derived from two different information sources. The corresponding fused feature vectors may be denoted as Z_i and Z_j , respectively.
2. Let s_X and s_Y be the normalized match scores generated by comparing X_i with X_j and Y_i with Y_j , respectively, and let $s_{match} = (s_X + s_Y)/2$ be the fused match score obtained using the simple sum rule.
3. A pair of fused feature vectors, Z_i and Z_j , are then compared using two different distance measures: the Euclidean distance (s_{euc}) and the Thresholded Absolute Distance or TAD (s_{tad}). Thus,

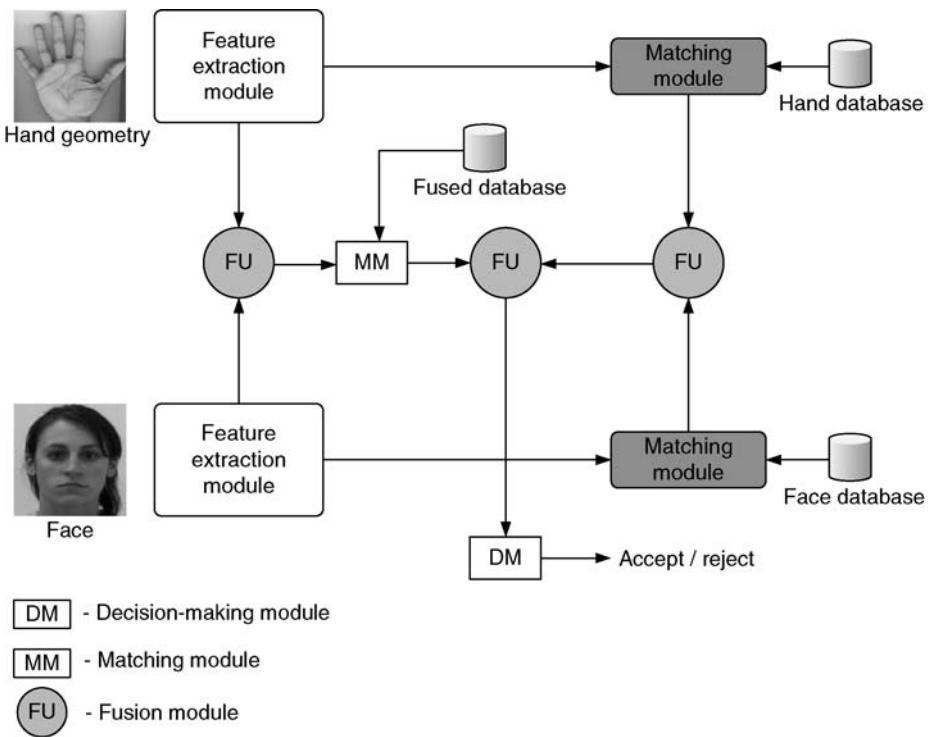
$$s_{euc} \sum_{r=1}^k (z_{i,r} - z_{j,r})^2 \quad (3)$$

$$s_{tad} \sum_{r=1}^k I(|z_{i,r} - z_{j,r}|, t). \quad (4)$$

Here, $I(u, t) = 1$, if $u > t$ (and 0, otherwise), t is a prespecified threshold, and k is the dimensionality of the fused feature vector. The thresholded absolute distance measure determines the *number* of normalized feature values that differ by a magnitude greater than t . The s_{euc} and s_{tad} values are consolidated into one feature level score, s_{feat} via the simple sum rule (Fig. 2). This retains information at the match score level (s_{match}) as well as the feature level (s_{feat}).

4. Finally, the simple sum rule is used to combine s_{match} and s_{feat} in order to obtain the final score s_{tot} (Fig. 3).

The authors compare the matching performances obtained using s_{match} and s_{tot} in all three scenarios. Results indicate that feature level fusion is advantageous in some cases. The feature selection scheme ensures that redundant or correlated feature values are detected and removed before invoking the matcher. This is probably one of the key benefits of performing fusion at the feature level [9].



Fusion, Feature-Level. **Figure 3** The flow of information when data from the feature level and match score level are combined in a multibiometric system [1].

Chibelushi et al. [10] discuss a scheme to combine the features associated with the voice (audio) and lip shape (video) of an individual in an identification system. Fourteen mel-frequency cepstral coefficients (MFCC) and 12 geometric features are extracted from the audio and video streams to represent the voice and shape of the lips, respectively. The PCA and LDA transformations are used to reduce the dimensionality of the concatenated feature set. The authors demonstrate that the use of feature level fusion in their system is equivalent to increasing the signal-to-noise ratio (SNR) of the audio signal thereby justifying the use of lip shape in the fusion module. Other examples of feature level fusion can be found in [11] (face and iris) and [12] (hand geometry and palmprint).

Summary

Feature-level fusion represents an early fusion strategy in which multiple feature sets are consolidated in order to generate a more robust template. These feature sets can emerge (1) from a single biometric algorithm operating on different biometric samples (e.g.,

two images of the right hand of a single subject), or (2) from multiple biometric algorithms. If the feature sets to be combined originate from the same biometric algorithm (thus, a single modality is assumed), then feature level fusion can be used for template update or template improvement. If the feature sets originate from multiple biometric algorithms, then a concatenation procedure can be used to integrate them. The concatenation procedure has a feature normalization and a feature selection (or transformation) stage resulting in a compact set of salient features that can be used by the matcher. The primary advantage of such an approach is the elimination of redundant features thereby improving matching accuracy. In some cases, it may be advantageous to design a hybrid system that combines the outputs of score-level fusion and feature-level fusion. The disadvantages of feature-level fusion include the need to design a new matcher and to acquire a large number of training samples.

Related Entries

- ▶ Multibiometrics

References

1. Ross, A., Govindarajan, R.: Feature Level fusion using hand and face biometrics. In: Proceedings of SPIE Conference on Biometric Technology for Human Identification II. vol. 5779, pp. 196–204. Orlando, USA (2005)
2. Jiang, X., Ser, W.: Online fingerprint template improvement. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 1121–1126 (2002)
3. Moon, Y.S., Yeung, H.W., Chan, K.C., Chan, S.O.: Template synthesis and image mosaicking for fingerprint registration: an experimental study. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). vol. 5, pp. 409–412. Montreal, Canada (2004)
4. Yau, W.Y., Toh, K.A., Jiang, X., Chen, T.P., Lu, J.: On fingerprint template synthesis. In: CD-ROM Proceedings of Sixth International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore (2000)
5. Jain, A.K., Chandrasekaran, B.: Dimensionality and sample size considerations in pattern recognition practice. In: Krishnaiah, P., Kanal L.N. (eds.) *Handbook of Statistics*, Vol. 2, Vol. 2, pp. 835–855. North-Holland, Amsterdam (1982)
6. Pudil, P., Novovicova, J., Kittler, J.: Floating search methods in feature selection. *Pattern Recognit. Lett.* **15**, 1119–1124 (1994)
7. Jain, A.K., Zongker, D.: Feature selection: evaluation, application, and small sample performance. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 153–158 (1997)
8. Jain, A.K., Duin, R.P.W., Mao, J.: Statistical pattern recognition: a review. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 4–37 (2000)
9. Kumar, A., Zhang, D.: Biometric recognition using feature selection and combination. In: Proceedings of Fifth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), pp. 813–822. Rye Brook, USA (2005)
10. Chibelushi, C.C., Mason, J.S.D., Deravi, F.: Feature-level data fusion for bimodal person recognition. In: Proceedings of the Sixth International Conference on Image Processing and Its Applications, vol. 1, pp. 399–403. Dublin, Ireland (1997)
11. Son, B., Lee, Y.: Biometric authentication system using reduced joint feature vector of iris and face. In: Proceedings of Fifth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), pp. 513–522. Rye Brook, USA (2005)
12. Kumar, A., Wong, D.C.M., Shen, H.C., Jain, A.K.: Personal Verification using palmprint and hand geometry biometric. In: Fourth International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA), pp. 668–678. Guildford, UK (2003)

Fusion, Image Level

- Fusion, Sensor-Level

Fusion, Measurement Level

- Fusion, Score-Level

Fusion, Physics-Based

Physics-based fusion makes use of the physical characteristics of the multispectral image acquisition process. In this fusion scheme, information on the spectral response of the sensor, the transmittance of the liquid crystal tunable filter (when used), the spectral reflectance of the object being imaged, and the spectral power distribution of the illuminant, are used, separately or in combination, as weights for the different sub-spectral images for their fusion.

- Multispectral and Hyperspectral Biometrics

Fusion, Quality-Based

NORMAN POH

CVSSP, FEPS, University of Surrey Guildford, Surrey GU2 7XH, UK

Synonym

Quality-dependent fusion

Definition

Quality-based fusion refers to the use of *quality measures* in combining several biometric system outputs. Quality measures are an array of measurements quantifying the degree of excellence or conformance of biometric samples to some predefined criteria known to influence the system performance. Examples of quality measures for face biometrics are focus, contrast, and face detection reliability; and for iris biometrics are iris texture richness, the area of iris used for matching, and iris detection reliability. In quality-based fusion, the

match scores of biometric samples of higher quality are considered more important, i.e., given higher weights, in order to compute the final combined score.

Introduction

Quality-based fusion in the context of multibiometric systems is more challenging than multi-algorithmic systems because quality measures of the different biometric modalities are not comparable. This implies that quality-based fusion techniques have to necessarily consider the joint space of scores and quality measures, hence taking into account not only the dependency among scores themselves but also the dependency between scores and quality measures.

Prior studies in this direction include but are not limited to [1–4]. Nandakumar et al. proposed a likelihood ratio-based approach to achieve quality dependent score fusion [1]. This is a generative approach to model the relationship between scores and quality measures of the same modality. The likelihood of scores and quality measures of different biometric modalities are combined using the product rule, hence, realizing a naive Bayes classifier. The result is that the less informative modalities will produce likelihood ratios close to one and will therefore not influence the final combined score.

Fierrez-Aguilar et al. proposed a quality-based fusion realized using a support vector machine (SVM) [2]. In their context, quality measures were manually annotated and were used in two ways. First, they were used to control the penalty function of the SVM learning criterion. Second, during inference, quality measures were also used to weigh the relative influence of the respective modalities and the joint decision making process. Intuitively, the approach enables the multimodal system to focus on the single modality of dominant quality or for comparable qualities on the joint decision making system. Unfortunately, as a result of the SVM training strategy the joint decision making system is optimized for good quality data only.

Bigun et al. proposed the Bayesian Conciliation method [3]. This method relies on two components known as a client and an impostor supervisor. The client supervisor estimates the expected true authenticity score of a claim based on its expertise in recognizing client data (likewise for the impostor supervisor).

The final decision is made by taking into account the different expertise of the two supervisors and choosing the one which comes closest to its goal, which is defined as zero for impostor supervisor and one for client supervisor. Effectively, the supervisor adapts to each identity claim as a function of the quality of the input data.

Kryszczuk et al. proposed a *derived* quality measure [4] instead of raw quality measures as done in [1–4]. The derived quality measure, or the confidence is defined as the posterior probability of making the correct decision given some observed evidences, which include both the system output and raw quality measures. In the context of bimodal fusion, this means that if the decision of two systems are in conflict (different), one takes the decision of the system which is more likely to be correct.

Kittler et al. proposed a framework to incorporate the quality information in fusion from a pattern recognition perspective [5]. In this framework, various levels of system output dependency, i.e., whether they belong to the same modality or to different modalities, are considered.

Last but not least, Poh et al. proposed a generative approach to estimate the joint density of scores and quality measures by first clustering the quality measures into discrete hidden states [6]. This approach assumes that the scores and quality measures are independent given the discrete quality state/cluster. This approach is sensible because similar quality measures in a cluster will share similar statistical properties and thus they can be combined by the same fusion classifier, and vice versa for dissimilar quality measures.

Quality-Based Fusion from the Pattern Recognition Perspective

Let $x \in \mathbb{R}^R$ be a vector of output scores of R experts, $q \in \mathbb{R}^P$ be a vector of P quality measures and $k \in \{C, I\}$ be one of the two possible classes of users, i.e., genuine users or clients and impostors. From the Bayesian point of view, the *generative* and *discriminative* approaches which incorporate the quality information directly can be written as follows:

$$y_{com}^{llr} \equiv f^{llr}(x, q) = \log \frac{p(x, q|C)}{p(x, q|I)} \quad (1)$$

$$y_{com}^{prob} \equiv f^{prob}(x, q) = P(C|x, q) \quad (2)$$

In practice (2) is approximated by:

$$P(C|x, q) \approx \text{sigmoid}(f^{disc}(x, q)) = \frac{1}{1 + \exp(f^{disc}(x, q))} \quad (3)$$

where the output $f^{disc}(x, q) \in [-\infty, \infty]$ does not have to be associated with probability. $f^{disc}(x, q)$ is known as a discriminative function and very often, based on the sign of its output, one classifies x as either belonging to a client or an impostor. One can implement $f^{llr}(x, q)$ using any density estimator, e.g., Gaussian Mixture Model and Parzen windows [7]; $f^{prob}(x, q)$ using logistic regression [8] or any neural network [6] with the sigmoid activation function; and $f^{disc}(x, q)$ using a support vector machine [9], linear or quadratic discriminant functions and their variant [8], and neural networks.

The conventional fusion approaches without using the quality information can also be divided into either generative or discriminative. They can be written in similar ways as in (1) and (2) except that q is not used as part of the observations.

Classifier Design and System Output Dependency

Considered here is the case where the system outputs, x , can be obtained from the same biometric modality or from different modalities. For this reason, $x_{m,i}$ is introduced to denote the i -th classifier of the m -th biometric modality. There are I_m systems for the m -th modality and M biometric modalities are available. As a result, the number of systems available for fusion is $\sum_m I_m$.

In general, higher dependence is expected among the system outputs sharing the same biometric modality and, in contrast, independence when the system fuses different biometric modalities. By assuming different types of system output dependency, the following three types of fusion architecture are identified, in increasing levels of complexity:

1. *Multi-stage single processing (MSSP)*. This architecture is a result of assuming independence among all the system outputs despite the fact that systems sharing the same biometric modality may be

dependent. It can be written as:

$$y_{com}^{MSSP} = \prod_m \prod_i P(C|x_{m,i}, q) = \prod_m \prod_i f^{prob}(x_{m,i}, q) \quad (4)$$

Note that since $f^{prob}(x_{m,i}, q)$ operates on a single system at a time, it can be considered as a quality-dependent score normalization procedure. It is therefore not a deterministic one-to-one mapping function as studied in [10] but rather a function of $x_{m,i}$ and q jointly. Note that discriminative functions $f^{disc}(x, q)$, e.g., a Support Vector Machine (SVM), do not output scores which satisfy the axiomatic properties of probabilities and cannot therefore be used in conjunction with a product fusion rule. Instead, the sum rule may be more appropriate, i.e.,

$$y_{com}^{MSSP} = \sum_m \sum_i f^{disc}(x_{m,i}, q) \quad (5)$$

By doing so, one implicitly assumes that the class-conditional distributions of the outputs $f^{disc}(x_{m,i}, q)$ across all m and i are comparable. This is, in general, not the case, thus implying the need for normalizing the outputs. Fortunately, this can be avoided by normalizing the input to the function $f^{disc} : \mathbb{R}^{R+P} \rightarrow \mathbb{R}$ instead of its output. Suppose that each of the \mathbb{R}^{R+P} input elements is normalized to having zero mean and unit variance (across all the training examples), and the same complexity of $f^{disc}(x_{m,i}, q)$ is used for all m and i , then the outputs $f^{disc}(x_{m,i}, q)$ will be comparable. For the generative approach, using the sum rule, i.e.,

$$y_{com}^{MSSP} = \sum_m \sum_i f^{llr}(x_{m,i}, q), \quad (6)$$

is a direct implication of assuming independence among the output of systems $x_{m,i}$ for all m and i .

2. *Multi-stage joint processing (MSJP)*. This architecture takes into consideration the dependency among system outputs derived from the same biometric modality yet ignores the dependency of the system outputs coming from different biometric modalities. It can be written as:

$$y_{com}^{MSJP} = \prod_m P(C|x_m, q) = \prod_m f^{prob}(x_m, q), \quad (7)$$

where x_m denotes a vector the components of which are the system outputs sharing the m -th biometric modality, i.e., $x_m \equiv [x_{m,1}, \dots, x_{m,I_m}]$.

The practical implication of this architecture is that one designs a fusion classifier per biometric modality and then combines all M resulting fusion classifiers using a fixed rule, e.g., the product rule for $f^{prob}(x_m, q)$ and the sum rule for $f^{disc}(x_m, q)$ and $f^{llr}(x_m, q)$.

3. *Single-stage joint processing (SSJP)*. This architecture does not assume system output independence. It can be written as:

$$y_{com}^{SSJP} = P(C|x, q) = f^{prob}(x, q), \quad (8)$$

where x is a vector containing all the system outputs, i.e., $x = \{x_{m,i} | \forall_{i,m}\}$. The function $f^{prob}(x, q)$ is simply replaced by $f^{llr}(x, q)$ when using a ▶ **generative classifier** and by $f^{disc}(x, q)$ when using a ▶ **discriminative classifier**.

In the discussion that follows, the focus is on training the discriminative function $f^{disc}(x, q)$. However, the discussion generalizes to the functions $f^{llr}(x, q)$ and $f^{prob}(x, q)$. For this reason, the generic term $f(x, q)$ is used and refer to one of the three particular fusion algorithms, i.e., $f^{llr}(x, q)$, $f^{prob}(x, q)$, or $f^{disc}(x, q)$, only when necessary.

The Complexity of Modeling Scores and Quality Measures: A Generative Approach

In the generative approach, modeling the joint space of x and q is difficult since q is not directly relevant to the classification task. For example, if one uses a mixture of Gaussian components to estimate the joint density, one would use many more components than one does if one models just x . This problem is particularly acute when the dimension of q is large. One way to reduce the complexity (the number of components and their associated parameters) is to first cluster the quality measures and then learn the density of x for each cluster. This strategy was reported in [6]. Instead of modeling $p(x, q)$ directly, Poh et al. proposed to factorize it into $p(x|q)p(q)$ where,

$$p(x|q) = \sum_Q p(x|Q)P(Q|q) \quad (9)$$

where Q is a cluster state and $P(Q|q)$ is the posterior probability of Q given the observation q . Since Q is not observed (hidden), it has to be integrated out, hence,

explaining the sum over Q in (9). In [6], it turns out that one does not need to model $p(q)$ to implement a quality-based fusion classifier.

The solution of (9) is more elegant than the one that directly estimates $p(x, q)$. This is because the density $p(x|Q)$ has only R dimensions, i.e., the dimension in x , whereas $p(x, q)$ has $R + P$ dimensions. As a result, one can potentially face the curse of dimensionality when modeling $p(x, q)$, especially in the situation where x is small and q is large in dimension. In brief, this curse means that modeling the increased number of dimensions may be less effective since this is not necessarily supported by an exponential increase in the number of training samples. In fact, there is only a fixed number of training samples to design one fusion classifier. Note that when q is one dimensional, the classifier should be more appropriately called a quality-dependent score normalization procedure.

The realized quality-based fusion via (9), when written in the form of (1), is:

$$f^{llr}(x, q) = \log \frac{\sum_Q p(x|C, Q)p(Q|q)}{\sum_Q p(x|I, Q)p(Q|q)} \quad (10)$$

The Complexity of Modeling Scores and Quality Measures: A Discriminative Approach

Similar to the generative approach, jointly estimating x and q is also a challenging problem for the discriminative approach. Suppose that, one uses a linear function in $f(x, q)$ to distinguish the client class from the impostor one. In this case a weight will be associated with each element in x and q . The result after training is that magnitude of the weight associated with q will be comparatively small because q has no discriminative information. This suggests that using nonlinear function of $f(x, q)$ may be more useful.

One way to introduce non-linearity is by using some kind of expansion between x and q , i.e., $x \otimes q$, where \otimes is called a *tensor product*. Note that x and q are not vectors of the same length. If x has R elements and q has P elements, then $x \otimes q$ will result in $P \times R$ elements and each element is a product between a pair of the elements in x and q . Therefore, when training $f(x, q)$, the fusion classifier must be fed with inputs $[x, q, x \otimes q]$ instead of $[x, q]$.

When one uses $[x, q, x \otimes q]$, the linear function can be written as:

$$\begin{aligned} f(x, q) &= \sum_i \sum_j w_{i,j} x_i q_j + \sum_i w_i x_i + \sum_j v_j q_j \\ &= \sum_i x_i \left(\underbrace{\sum_j q_j w_{i,j} + w_i}_{\text{under-braced term}} \right) + \underbrace{\sum_j v_j q_j}_{\text{under-braced term}}, \end{aligned} \quad (11)$$

where the weight $w_{i,j}$ is associated with $x_i q_j$, the weight w_i is associated with x_i , and v_j is associated with q_j . In this notation, x_i is an element of vector x and q_j is an element of vector q . (11) clearly shows that the resulting classifier is *linear* except that the weight is modified *dynamically* by the quality measures via the first under-braced term. The second under-braced term shows that q *dynamically* adjusts the decision threshold.

Several possible “arrangements” are outlined in **Table 1**, presented in the order of increasing complexity, i.e., the number of parameters. $f([x, q])$ is written to explicitly refer to the second arrangement, $f([x, x \otimes q])$ to refer to the third arrangement, etc. The second column shows the four possible arrangements, i.e., the way the features are used as input to a fusion algorithm. The third column shows the resulting discriminative linear function $f^{disc}(x, q)$. While similar analyzes cannot be done for the linear discriminative function $f^{prob}(x, q)$ (due to the sigmoid function) and for the generative function $f^{lhr}(x, q)$, our purpose in showing the elements in the expanded input vector along with their associated weight parameters is to illustrate the complexity of each arrangement. For instance, the first arrangement, i.e., $f([x])$, does not use any quality information. The second arrangement, i.e., $f([x, q])$ does not contain any interaction between x and q . However, it considers the case where the

Fusion, Quality-Based. **Table 1** The complexity of the function $f(x, q)$ when implemented using a linear classifier, in increasing level of complexity due to different input arrangements

No.	Arrangement	The resulting function $f^{disc}(x, q)$	No. of parameters
1	$[x]$	$\sum_i x_i w_i$	R
2	$[x, q]$	$\sum_i x_i w_i + \sum_j q_j v_j$	$R+P$
3	$[x, x \otimes q]$	$\sum_i x_i (\sum_j q_j w_{i,j} + w_i)$	$R \times (P+1)$
4	$[x, q, x \otimes q]$	$\sum_i x_i (\sum_j q_j w_{i,j} + w_i) + \sum_j v_j q_j$	$R+P+R \times P$

decision threshold may be modified by q . In the third arrangement, one creates a linear classifier whose weights can change dynamically as a function of q . The last arrangement, i.e., $f([x, q, x \otimes q])$ or (11), is the most general one since it contains all possible interactions between x and q of the first three arrangements. In [5], it was shown that the last three arrangements achieve superior results compared to the first one (without considering the quality information) across many intramodal and multimodal fusion tasks.

The quality-enhanced discriminative fusion classifier with the input $[x, x \otimes q]$ (the third arrangement) is structurally very similar to the one proposed in [11] where a reduced polynomial discriminative function was used. In our case, one can use *any* discriminative classifier to implement it. This is an elegant solution because one does not need to design a dedicated fusion algorithm such as those proposed in [2, 3, 11] to achieve the same goal any longer.

Related Entries

- ▶ Biometric Sample Quality
- ▶ Face Sample Quality
- ▶ Feature-level Fusion
- ▶ Fingerprint Image Quality
- ▶ Iris Image Quality
- ▶ Multiple Classifier Systems
- ▶ Multibiometrics
- ▶ Score-level Fusion

References

1. Nandakumar, K., Chen, Y., Dass, S., Jain, A.: Quality-based score level fusion in multibiometric systems. In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR), pp. 473–476. Hong Kong (2006)
2. Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J., Bigun, J.: Kernel-based multimodal biometric verification using quality signals. In: Defense and Security Symposium, Workshop on Biometric Technology for Human Identification, Proceedings of SPIE, vol. 5404, pp. 544–554 (2004)
3. Bigun, J., Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J.: Multimodal biometric authentication using quality signals in mobile communications. In: 12th International Conference on Image Analysis and Processing, pp. 2–13. Mantova (2003)
4. Kryszczuk, K., Richiardi, J., Prodanov, P., Drygajlo, A.: Error handling in multimodal biometric systems using reliability

- measures. In: Proceedings of the 12th European Conference on Signal Processing. Antalya, Turkey (2005)
5. Kittler, J., Poh, N., Fatukasi, O., Messer, K., Kryszczuk, K., Richiardi, J., Drygajlo, A.: Quality dependent fusion of intra-modal and multimodal biometric experts. In: Proceedings of SPIE Defense and Security Symposium, Workshop on Biometric Technology for Human Identification, vol. 6539 (2007)
 6. Poh, N., Heusch, G., Kittler, J.: On Combination of Face Authentication Experts by a Mixture of Quality Dependent Fusion Classifiers. In: LNCS 4472, Multiple Classifiers System (MCS), pp. 344–356. Prague (2007)
 7. Bishop, C.: Neural Networks for Pattern Recognition. Oxford University Press (1999)
 8. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer (2001)
 9. Vapnik, V.N.: Statistical Learning Theory. Springer (1998)
 10. Jain, A., Nandakumar, K., Ross, A.: Score normalisation in multimodal biometric systems. *Pattern Recognit.* **38**(12), 2270–2285 (2005)
 11. Toh, K.A., Yau, W.Y., Lim, E., Chen, L., Ng., C.H.: Fusion of Auxiliary Information for Multimodal Biometric Authentication. In: LNCS 3072, International Conference on Biometric Authentication (ICBA), pp. 678–685. Hong Kong (2004)

Fusion, Rank-Level

AJAY KUMAR

Department of Computing, The Hong Kong Polytechnic University

Synonym

Biometric Fusion, Rank-Level

Definition

Rank level fusion is the method of consolidating more than two identification results to enhance the reliability in personal identification. In multimodal biometric system, rank level fusion can be used to combine the biometrics matching scores from the different biometric modalities (for example face, fingerprint, palmprint, and iris). It can also be used for performance improvement in unimodal biometric system by combining multiple classifier output that use different classifiers (K nearest neighbor, neural network, support vector machine, decision tree, etc.), different training set, different architectures (different number of layers or transfer function in neural network), or different

parameter values (different kernels in support vector machine or different K in K nearest neighbor).

Introduction

The majority of biometric system deployed using feature extraction from a single biometric modality and a particular classification procedure to determine the identity on an individual. The perfect solutions for user identification are often difficult to achieve, mainly due to the large number of user classes and the imperfection in the feature extraction process. Therefore, the improvement in the user identification results using the simultaneous extraction of features and classifiers of different types has been investigated. The combination of potentially conflicting decisions in multimodal or unimodal biometric system employing different classifiers can be achieved in several ways: at feature, score, and decision level. In general, the improvement in identification accuracy is achieved by selecting combination mechanism that can take advantage of strengths of individual classifiers while suppressing their weakness.

Any biometric recognition system is capable of generating matching scores for the input user with those of the enrolled possible identities. The set of all the possible user identities can be ranked by sorting the matching scores in the descending order. Thus a biometric system can identify an unknown user by generating ranks, i.e., integer numbers for each of the possible user identity. The rank level fusion refers to the mechanism of combining such output ranks from the various biometrics ► **matchers** (subsystems), to consolidate the combined output ranks to establish the identity of an individual with higher confidence. The matching score contains more information than ranks and therefore matching score level fusion schemes are believed to be more flexible. However, the rank level fusion schemes do not require ► **transformation** of ranks from various biometrics matchers into a common domain and are simpler to implement. Several decision level fusion schemes only use ► **top choice** (rank) from each of the biometric classifiers, which is likely to be sufficient for biometric systems with small number of users. However, with the increase in number of enrolled identities or users, the correct rate for top choices drops, the ► **secondary choices** often contain near misses that should not be overlooked and are made use of in the rank level fusion.

Methods for Combining Ranks

The voting techniques proposed by different researchers [1–3] for consolidating rank output from the different biometric matchers will now be introduced. Given the ranked list of user identities returned by M different biometric matchers, let $r_i(k)$ be the rank assigned to the user k by the i th matcher. The user identity for k th user is assigned by computing the fused rank score m_k from all the M matchers.

1. *The Highest Rank Method.* In this method, the user identity is ascertained from the highest ranks returned by the individual matchers. Each of the possible user identity receives M ranks, each from the M matchers. The fused rank score m_k for every possible user identity k is computed from the minimum (highest) of these M ranks. The user identities are then sorted in the order of fused rank scores to obtain the combined or new ranking from all M matchers. Any ties in the fused rank scores (m_k) are randomly broken to obtain linearly ordered combined ranking. These ties are due to a number of user identities sharing the same combined ranks and depend on the number of employed matchers. The chances of the occurrences of such ties will be smaller, if the number of enrolled user identities are large and the number of matchers employed in the fusion are small. The advantage of this method lies in the utilization of strength of each of the biometric matchers. However, large number of matchers can result in more ties in the combined ranking, which is the major problem in this method. Therefore this method is considered useful in biometric systems combining small number of matchers with large number of enrolled users.

2. *Borda* (Named for the French scientist *Jean-Charles de Borda* (1733–1799) who formulated this preferential voting system.) *Count Method.* The Borda count is the generalization of majority vote and the most commonly used method for ▶ unsupervised rank level fusion. It is the voting method in which each matcher gives priority to all possible user identities. Each matcher ranks the fixed set of possible user identities in the order of its preference. For every matcher, the top ranked user identity is given N votes, the second ranked candidate identity is given $N-1$ votes and so on. Then for every possible user identity, the votes from all the matchers are added. The

user identity that receives the highest number of votes is assigned as the winner or the true user identity.

$$m_k = \sum_{i=1}^M r_i(k) \forall k, \quad k = \{1, 2, \dots, N\}. \quad (1)$$

The Borda count score m_k represents strength of agreement among different biometric matchers. The Borda count method assumes statistical independence, i.e., ranks assigned to a given user by different matchers are independent. This assumption is often made in practice but it may not be true. The Borda count method is particularly considered suitable for combining the biometrics matchers with large number of user identities that often generate the correct user identities *near* the top of list (ranks) but not *at* the top. This method is efficient, simple, and does not require any training. However, it assumes that all matchers are equally correct. This may not be the case when some matchers are more likely to be correct than others. Therefore, weighted Borda count method has been suggested to utilize the strength of individual matchers.

3. *Weighted Borda Count Method.* The performance of different biometric matchers is not uniform, for example, a biometric matcher using iris images is expected to perform better than those matchers using hand geometry or face images. Therefore, modification of Borda count method by assigning corresponding weights to the ranks produced by individual matchers has been suggested. The fused rank scores in weighted Borda count method are computed as follows:

$$m_k = \sum_{i=1}^M w_i r_i(k), \quad (2)$$

where the w_i represents the weights assigned to the i th matcher. The weight w_i are assigned to reflect the significant of each matcher and can be computed from the overall assessment of the performance. The weights are computed during the training phase using logistic regression (as detailed in [3]) or using more sophisticated machine learning techniques.

4. *Bayes Fuse.* The Bayes fuse is the ▶ supervised rank level fusion method based on Bayesian inference. Each of the possible user identity is ranked

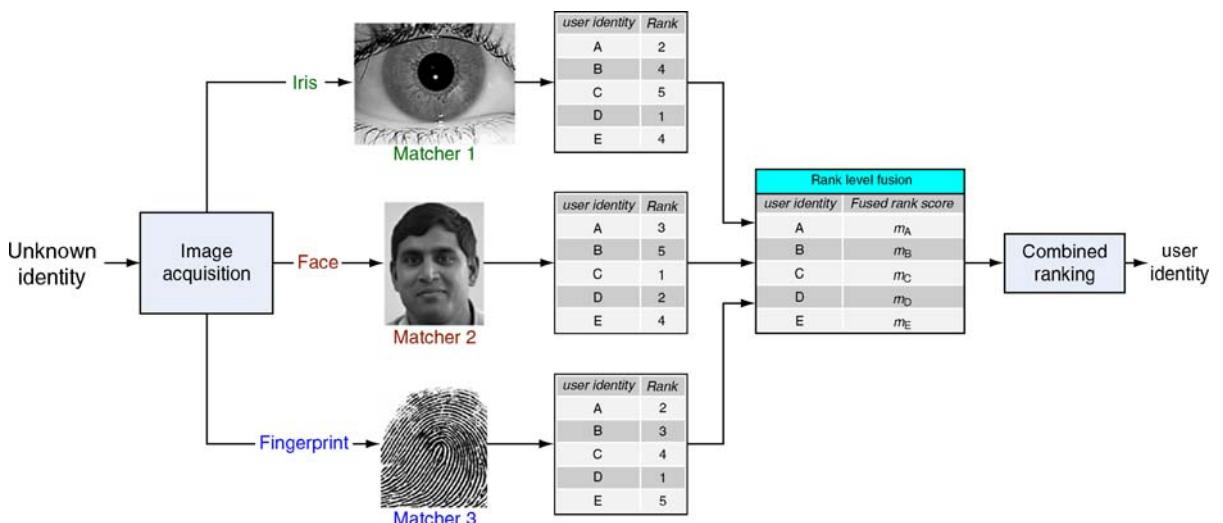
according to the fused rank scores computed as follows:

$$m_k = \sum_{i=1}^M \log \frac{\Pr[m_k(i)|genuine]}{\Pr[m_k(i)|impostor]}, \quad (3)$$

where $\Pr[m_k(i)|impostor]$ is the probability that an imposter user would be ranked to $m_k(i)$ by the i th matcher and $\Pr[m_k(i)|genuine]$ is the probability that a genuine user would be ranked to $m_k(i)$ by the i th matcher. These two likelihood probabilities are computed from the training data during training phase. The above equation is easily derived [2] from the estimation of two posterior probabilities, each for the genuine and imposter class, using Bayes rule. The combined ranks generated using Eq. (3) makes a common naive Bayes assumption, i.e., individual ranks assigned to the user identities by M matchers are independent. The training phase in Bayes fuse method required the collection of simple statistics about the distribution of ranks among various user identities. The rank level fusion using Bayes fuse was originally introduced for information retrieval but is equally useful in biometrics fusion.

Example

The four different rank level fusion methods discussed above can be better clarified with a simple example in multimodal biometric fusion. This example illustrates the combination of three different biometric matchers (Fig. 1), using iris, fingerprint, and face image, to



Fusion, Rank-Level. **Figure 1** An example of multimodal biometric system employing rank level fusion.

generate matching scores. These matching scores are internally sorted to produce different ranking among the possible user identities. There are only five different users (user A, user B, user C, user D, and user E) and 1, 2, ..., 5 represents the ranks for the possible input user identity with 1 being the highest rank/possibility. Let the weights of different matchers computed from the training data using linear regression be 0.5, 0.15, 0.35 for the matcher 1, matcher 2, and matcher 3 respectively. Let the probability that a genuine user be ranked at ranks (1, 2, 3, 4, 5) be (0.8, 0.1, 0.06, 0.02, 0.02), (0.5, 0.42, 0.06, 0.01, 0.01), and (0.6, 0.2, 0.08, 0.07, 0.05) for matcher 1, matcher 2, and matcher 3 respectively. Similarly the prior probabilities for an imposter user be ranked at ranks (1, 2, 3, 4, 5) have been obtained from the training data and are listed as (0.2, 0.9, 0.94, 0.98, 0.98), (0.5, 0.58, 0.94, 0.99, 0.99), and (0.4, 0.8, 0.92, 0.93, 0.95) respectively for matcher 1, matcher 2, and matcher 3.

Let us now compute the fused rank scores (m_A , m_B , m_C , m_D , m_E) and the new rankings for each of the four methods discussed in previous section.

1. *Highest Rank.* The fused rank scores using highest rank level method are shown in Table 2. The fused rank score for user A (m_A) will be 2 (highest rank or minimum of 2, 3, 2). The ties for m_C and m_D are randomly broken and the combined ranking is also shown in Table 1. The highest rank method achieves highest ranking for C and therefore the unknown input identity is user C.

Fusion, Rank-Level. **Table 1** Example for consolidating ranks using unsupervised rank level fusion methods

User identity	Highest rank method			Borda count method		
	Fused rank score	Combined ranking		Fused rank score	Combined ranking	
A	m_A	2	3	m_A	7	2
B	m_B	3	4	m_B	12	4
C	m_C	1	1	m_C	10	3
D	m_D	1	2	m_D	4	1
E	m_E	4	5	m_E	13	5

Fusion, Rank-Level. **Table 2** Example for consolidating ranks using supervised rank level fusion methods

User identity	Weighted borda count method			Bayes fuse method		
	Fused rank score	Combined ranking		Fused rank score	Combined ranking	
A	m_A	2.15	2	m_A	-6.34	2
B	m_B	3.8	3	m_B	-10.93	4
C	m_C	4.05	4	m_C	-6.48	3
D	m_D	1.15	1	m_D	1.47	1
E	m_E	4.35	5	m_E	-11.43	5

2. *Borda Count.* The fused rank scores using Borda count are computed as follows: $m_A = (2 + 3 + 2) = 7$, $m_B = (4 + 5 + 3) = 12$, $m_C = (5 + 1 + 4) = 10$, $m_D = (1 + 2 + 1) = 4$, $m_E = (4 + 4 + 5) = 13$. Thus, m_D is lowest and user D achieves highest combined ranking ([Table 1](#)).
3. *Weighted Borda Count.* The fused rank scores $m_A = (2 \times 0.5 + 3 \times 0.15 + 2 \times 0.35) = 2.15$. Similarly rank fused scores for rest of the users can be computed and are shown in [Table 2](#).
4. *Bayes Fuse.* The prior probabilities that each of the ranks are true (untrue), i.e., belongs to the genuine (impostor) class, can be obtained from the training data and are provided in the problem. The fused rank score for user A can be computed using (3) as follows: $m_A = \log(0.1/0.9) + \log(0.06/0.94) + \log(0.2/0.8) = -6.34$. The rest of the fused rank scores and the combined rankings are displayed in [Table 2](#).

Summary

In the biometrics literatures, one can find several examples [1, 3, 4, 6] of above rank level fusion methods to consolidate the outputs from different matchers. Bhatnagar et al. [4] employs a variation of Borda count method that uses partitioning of templates

to consolidate the combined ranks. Highest rank method employed by Rautiainen and Seppanen [6], is referred as lowest rank method since it chooses the minimum rank from the list of dissimilarity score instead of conventional maximum rank methods that employ highest ranks from the list of similarity scores. Several other variations of Borda count method have also been developed in the literature [7]; *Nenson's method* that uses successive elimination from Borda count that are below average Borda count or *Quota Borda method* that includes the quota element in counting ranks. However, they have not yet been investigated for their utility in the biometrics literature.

A survey of biometrics on various fusion techniques [5] suggests that the rank level fusion method is less preferred method of fusion while score level fusion continues to be the most popular method. The rank level fusion can be more useful in combining decisions from a large number of biometric matchers and such large systems has not yet been evaluated in the biometrics literature.

References

1. Lee, Y., Lee, K., Jee, H., Gil, Y., Choi, W., Ahn, D., Pan, S.: “Fusion for multimodal biometric identification,” Proc. ABVPA 2005, LNCS 3546, 1071–1079 (2005)

2. Aslam, J.A., Montague, M.: “Models for metasearch,” In: Proceedings of the 24th ACM SIGIR Conference on Research and Development in Information Retrieval, Sep. 2001, pp. 379–381 (2001)
3. Ho, T.K., Hull, J.J., Srihari, N.: “Decision combination in multiple classifier systems.” IEEE Trans. Pattern Anal. Mach. Intell. **16**, 66–75 (1994)
4. Bhatnagar, J., Kumar, A., Saggar, N.: “A novel approach to improve biometric recognition using rank level fusion,” Proc. CVPR 2007, Minneapolis, MN, pp. 1–6, (2007)
5. Ross, A., Nandakumar, K., Jain, A.K.: Handbook of Multibiometrics, Springer Verlag (2006)
6. Rautainen, M. and Seppanen, T.: “Comparison of visual features and fusion techniques in automatic detection of concepts from news video,” In: Proceedings of the IEEE International Conference on Multimedia and Expo., ICME 2005, Amsterdam, pp. 932–935 (2005)
7. Dummett, M.A.E.: Principles of Electoral Reform, Oxford University, New York (1997)

Fusion, Score-Level

ARUN ROSS¹, KARTHIK NANDAKUMAR²

¹West Virginia University, Morgantown, WV, USA

²Institute for Infocomm Research A* STAR,
Fusionopolis, Singapore

Synonyms

Fusion at the confidence level; Fusion at the measurement level; Match score fusion

Definition

In score-level fusion the match scores output by multiple biometric matchers are consolidated in order to render a decision about the identity of an individual. Typically, this consolidation procedure results in the generation of a single scalar score which is subsequently used by the biometric system. Fusion at this level is the most commonly discussed approach in the biometric literature primarily due to the ease of accessing and processing match scores (compared with the raw biometric data or the feature set extracted from the data). Fusion methods at this level can be broadly classified into three categories: density-based schemes, transformation-based schemes and classifier-based schemes.

Introduction

A match score is the result of comparing two feature sets extracted using the same feature extractor. A *similarity* score denotes how “similar” the two feature sets are, while a *distance* score denotes how “different” they are. Consequently, a high similarity score between a pair of feature sets indicates a good match whereas a high distance score indicates a poor match.

In score-level fusion the match scores output by multiple biometric matchers are combined to generate a new match score (a scalar) that can be subsequently used by the verification or identification modules for rendering an identity decision (alternatively, the fusion process may directly result in a decision). Fusion at this level is the most commonly discussed approach in the biometric literature primarily due to the ease of accessing and processing match scores (compared with the raw biometric data or the feature set extracted from the data). Fusion methods at this level can be broadly classified into three categories [1]: density-based schemes, transformation-based schemes and classifier-based schemes.

Density-Based Fusion schemes

Let $\mathbf{s} = [s_1, s_2, \dots, s_R]$ denote the scores emitted by multiple matchers, with s_j representing the match score of the j th matcher, $j = 1, \dots, R$. Further, let the labels ω_0 and ω_1 denote the genuine and impostor classes, respectively. Then, by ▶ Bayes decision theory [2], the probability of error can be minimized by adopting the following decision rule. (This is known as the Bayes decision rule or the minimum-error-rate classification rule under the 0-1 loss function [2]).

Assign $\mathbf{s} \rightarrow \omega_i$ if

$$P(\omega_i|\mathbf{s}) > P(\omega_j|\mathbf{s}), i \neq j, \quad \text{and} \quad i, j = 0, 1. \quad (1)$$

Here, the *a posteriori* probability $P(\omega_i|\mathbf{s})$, $i = 0, 1$, can be derived from the class-conditional density function $p(\mathbf{s}|\omega_i)$ using the Bayes formula, i.e.,

$$P(\omega_i|\mathbf{s}) = \frac{p(\mathbf{s}|\omega_i)P(\omega_i)}{p(\mathbf{s})}, \quad (2)$$

where $P(\omega_i)$ is the *a priori* probability of observing class ω_i and $p(\mathbf{s})$ denotes the probability of encountering \mathbf{s} . Thus, Eq. (1) can be re-written as

Assign $\mathbf{s} \rightarrow \omega_i$ if

$$\frac{p(\mathbf{s}|\omega_i)}{p(\mathbf{s}|\omega_j)} > \tau, i \neq j, \text{ and } i, j = 0, 1 \quad (3)$$

where $\frac{p(\mathbf{s}|\omega_i)}{p(\mathbf{s}|\omega_j)}$ is known as the *likelihood ratio* and $\tau = \frac{P(\omega_1)}{P(\omega_0)}$ is a predetermined threshold. The density $p(\mathbf{s}|\omega_i)$ is typically estimated from a training set of match score vectors, using parametric or nonparametric techniques. However, a large number of training samples are necessary to reliably estimate the joint-density function $p(\mathbf{s}|\omega_i)$ especially if the dimensionality of the feature vector \mathbf{s} is large. In the absence of sufficient number of training samples (which is typically the case when the multibiometric system is first deployed or if its parameters are subsequently adjusted), it is commonly assumed that the scalar scores s_1, s_2, \dots, s_R are generated by R independent random processes. This assumption permits the density function to be expressed as

$$p(\mathbf{s}|\omega_i) = \prod_{j=1}^R p(s_j|\omega_i), \quad (4)$$

where the joint-density function is now replaced by the product of its marginals. The marginal densities, $p(s_j|\omega_i), j = 1, 2, \dots, R, i = 0, 1$, are estimated from a training set of genuine and impostor scores corresponding to each of the R biometric matchers. Equation (4) results in the *product rule* which combines the scores generated by the R matchers as,

$$s_{prod} = \prod_{j=1}^R \frac{p(s_j|\omega_0)}{p(s_j|\omega_1)}. \quad (5)$$

Kittler et al. [3] modify the product rule by further assuming that the *a posteriori* probability $P(\omega_i|\mathbf{s})$ of class ω_i does not deviate much from its *a priori* probability $P(\omega_i)$ resulting in the *sum rule*:

$$s_{sum} = \frac{\sum_{j=1}^R p(s_j|\omega_0)}{\sum_{j=1}^R p(s_j|\omega_1)}. \quad (6)$$

Similar expressions can be derived for combining the match scores using the max, min, and median rules [1, 3]. All the aforementioned rules implicitly assume that the match scores are *continuous* random variables. Dass et al. [4] relax this assumption and represent the univariate density functions (i.e., the marginals in Eq. (4)) as a mixture of discrete as well as continuous

components. The resulting density functions are referred to as generalized densities. The authors demonstrate that the use of generalized density estimates (as opposed to continuous density estimates) significantly enhances the matching performance of the fusion algorithm. Furthermore, they use ▶ copula functions to model the correlation structure between the match scores s_1, s_2, \dots, s_R and, subsequently, define a novel fusion rule known as the copula fusion rule.

Transformation-Based Fusion schemes

Density-based schemes, as stated earlier, require a large number of training samples (i.e., genuine and impostor match scores) in order to accurately estimate the density functions. This may not be possible in most multibiometric systems due to the time, effort, and cost involved in acquiring labeled multibiometric data in an operational environment. In such situations, it may be necessary to *directly* combine the match scores generated by multiple matchers using simple fusion operators (such as the simple sum of scores or order statistics) without first interpreting them in a probabilistic framework. However, such an approach is meaningful only when the scores output by the matchers are comparable. To facilitate this, a score normalization process is essential to transform the multiple match scores into a common domain (it must be noted, however, that some score normalization schemes do require a large number of training samples as seen in the following section). The process of score normalization entails changing the location and the scale parameters of the underlying match score distributions in order to ensure compatibility between multiple score variables. A few of the commonly discussed score normalization methods are described in this article.

The simplest normalization technique is the *min–max* normalization. Min–max normalization is best suited for the case where the bounds (maximum and minimum values) of the scores produced by a matcher are known. In this case, the minimum and maximum scores can be easily transformed into 0 and 1, respectively. However, even if the match scores are not bounded, the minimum and maximum values for the given set of training match scores can be estimated prior to applying min–max normalization. Let s_j^i denote the i th match score output by the j th matcher,

$i = 1, 2, \dots, N; j = 1, 2, \dots, R$ (R is the number of matchers and N is the number of match scores available in the training set). The min–max normalized score, ns_j^t , for the test score s_j^t is given by

$$ns_j^t = \frac{s_j^t - \min_{i=1}^N s_j^i}{\max_{i=1}^N s_j^i - \min_{i=1}^N s_j^i}. \quad (7)$$

When the minimum and maximum values are estimated from the given set of match scores, this method is not robust (i.e., the method is sensitive to outliers in the data used for estimation). Min–max normalization retains the original distribution of scores except for a scaling factor and transforms all the scores into a common range [0, 1]. Distance scores can be transformed into similarity scores by subtracting the normalized score from 1.

Decimal scaling can be applied when the scores of different matchers are on a logarithmic scale. For example, if one matcher has scores in the range [0, 10] and the other has scores in the range [0, 1000], the following normalization could be applied to transform the scores of both the matchers to the common [0, 1] range.

$$ns_j^t = \frac{s_j^t}{10^{n_j}}, \quad (8)$$

where $n_j = \log_{10} \max_{i=1}^N s_j^i$. In the example with two matchers where the score ranges are [0, 10] and [0, 1000], the values of n would be 1 and 3, respectively. The problems with this approach are the lack of robustness and the implicit assumption that the scores of different matchers vary by a logarithmic factor.

The most commonly used score normalization technique is the *z-score* normalization that uses the arithmetic mean and standard deviation of the training data. This scheme can be expected to perform well if the average and the variance of the score distributions of the matchers are available. If the values of these two parameters are not known, then they can be estimated based on the given training set. The z-score normalized score is given by

$$ns_j^t = \frac{s_j^t - \mu_j}{\sigma_j}, \quad (9)$$

where μ_j is the arithmetic mean and σ_j is the standard deviation for the j th matcher. However, both mean and standard deviation are sensitive to outliers and hence, this method is not robust. Z-score

normalization does not guarantee a common numerical range for the normalized scores of the different matchers. If the distribution of the scores is not Gaussian, z-score normalization does not preserve the distribution of the given set of scores. This is due to the fact that mean and standard deviation are the optimal location and scale parameters only for a Gaussian distribution. While mean and standard deviation are reasonable estimates of location and scale, respectively, they are not optimal for an arbitrary match score distribution.

The *median* and *median absolute deviation* (MAD) statistics are less sensitive to outliers as well as points in the extreme tails of the distribution. Hence, a normalization scheme using median and MAD would be relatively robust and is given by

$$ns_j^t = \frac{s_j^t - med_j}{MAD_j}, \quad (10)$$

where $med_j = \text{median}_{i=1}^N s_j^i$ and $MAD_j = \text{median}_{i=1}^N |s_j^i - med_j|$. However, the median and the MAD estimators have a low efficiency compared to the mean and the standard deviation estimators, i.e., when the score distribution is not Gaussian, median and MAD are poor estimates of the location and scale parameters. Therefore, this normalization technique does not preserve the input score distribution and does not transform the scores into a common numerical range.

Cappelli et al. [5] use a *double sigmoid function* for score normalization in a multibiometric system that combines different fingerprint matchers. The normalized score is given by

$$ns_j^t = \begin{cases} \frac{1}{1+\exp\left(-2\left(\frac{s_j^t-\tau}{\alpha_1}\right)\right)} & \text{if } s_j^t < \tau, \\ \frac{1}{1+\exp\left(-2\left(\frac{s_j^t-\tau}{\alpha_2}\right)\right)} & \text{otherwise,} \end{cases} \quad (11)$$

where τ is the reference operating point and α_1 and α_2 denote the left and right edges of the region in which the function is linear. The double sigmoid function exhibits linear characteristics in the interval $(\tau - \alpha_1, \tau - \alpha_2)$. While the double sigmoid normalization scheme transforms the scores into the [0, 1] interval, it requires careful tuning of the parameters τ, α_1 and α_2 to obtain good efficiency. Generally, τ is chosen to be some value falling in the region of overlap between the genuine and impostor score distributions, and α_1 and α_2 are set so that they correspond to the

extent of overlap between the two distributions toward the left and right of τ , respectively. This normalization scheme provides a linear transformation of the scores in the region of overlap, while the scores outside this region are transformed nonlinearly. The double sigmoid normalization is very similar to the min–max normalization followed by the application of a two-quadrics (QQ) or a logistic (LG) function as suggested by [6]. When the values of α_1 and α_2 are large, the double sigmoid normalization closely resembles the QQ-min–max normalization. On the other hand, the double sigmoid normalization can be made to approach the LG-min–max normalization by assigning small values to α_1 and α_2 .

The *tanh-estimators* introduced by Hampel [7] are robust and highly efficient. The tanh normalization is given by

$$ns_j^t = \frac{1}{2} \left\{ \tanh \left(0.01 \left(\frac{s_j^t - \mu_{GH}}{\sigma_{GH}} \right) \right) + 1 \right\}, \quad (12)$$

where μ_{GH} and σ_{GH} are the mean and standard deviation estimates, respectively, of the genuine score distribution as given by Hampel estimators. Hampel estimators are based on the following influence (ψ)-function:

$$\psi(u) = \begin{cases} u & 0 \leq |u| < a, \\ a * sign(u) & a \leq |u| < b, \\ a * sign(u) * \left(\frac{c-|u|}{c-b} \right) & b \leq |u| < c, \\ 0 & |u| \geq c, \end{cases} \quad (13)$$

where

$$sign\{u\} = \begin{cases} +1, & \text{if } u \geq 0, \\ -1, & \text{otherwise.} \end{cases} \quad (14)$$

The Hampel influence function reduces the influence of the scores at the tails of the distribution (identified by a , b , and c) during the estimation of the location and scale parameters. Hence, this method is not sensitive to outliers. If many of the points that constitute the tail of the distributions are discarded, the estimate is robust but not efficient (optimal). On the other hand, if all the points that constitute the tail of the distributions are considered, the estimate is not robust but its efficiency increases. Therefore, the parameters a , b , and c must be carefully chosen depending on the amount of robustness required which in turn depends on the amount of noise in the available training data.

Fusion, Score-Level. **Table 1** Summary of score normalization techniques.

Normalization technique	Robustness	Efficiency
Min-max	No	High
Decimal scaling	No	High
Z-score	No	High
Median and MAD	Yes	Moderate
Double sigmoid	Yes	High
Tanh-estimators	Yes	High

Mosteller and Tukey [8] introduce the biweight location and scale estimators that are robust and efficient. But, the *biweight estimators* are iterative in nature (initial estimates of the biweight location and scale parameters are chosen, and these estimates are updated based on the training scores), and are applicable only for Gaussian data. A summary of the characteristics of the different normalization techniques discussed in this article is shown in **Table 1**. The min–max, decimal scaling and z-score normalization schemes are efficient, but are not robust to outliers. On the other hand, the median normalization scheme is robust but inefficient. Only the double sigmoid and tanh-estimators have both the desired characteristics, namely, robustness and efficiency.

Once the match scores output by multiple matchers are transformed into a common domain they can be combined using simple fusion operators such as the sum of scores, product of scores or order statistics (e.g., maximum/minimum of scores or median score).

Classifier-Based Fusion schemes

In the verification mode of operation, the match scores generated by the multiple matchers may be input to a trained pattern classifier, such as a neural network, in order to determine the class label (genuine or impostor). In this approach, the goal is to directly estimate the class rather than to compute an intermediate scalar value. Classifier-based fusion schemes assume the availability of a large representative number of genuine and impostor scores during the training phase of the classifier when its parameters are computed. The component scores do not have to be transformed into a common domain prior to invoking the classifier.

In the biometric literature several classifiers have been used to consolidate the match scores of multiple matchers. Brunelli and Falavigna [9] use a HyperBF network to combine matchers based on voice and face features. Verlinde and Cholet [10] compare the relative performance of three different classifiers, namely, the k-Nearest Neighbor classifier using vector quantization, the decision tree classifier, and a classifier based on the logistic regression model while fusing the match scores originating from three biometric matchers. Experiments on the M2VTS database show that the total error rate (sum of the false accept and false reject rates) of the multimodal system is an order of magnitude less than that of the individual matchers. Chatzis et al. [11] use classical k-means clustering, fuzzy clustering and median radial basis function (MRBF) algorithms for fusion at the match score level. The proposed system combines the output of five different face and voice matchers. Each matcher provides a match score and a quality metric indicating the reliability of the match score. These values are concatenated to form a ten-dimensional vector that is input to the classifiers. Ben-Yacoub et al. [12] evaluate a number of classification schemes for fusion including support vector machine (SVM) with polynomial kernels, SVM with Gaussian kernels, C4.5 decision trees, multilayer perceptron, Fisher linear discriminant, and Bayesian classifier. Experimental evaluations on the XM2VTS database consisting of 295 subjects suggest the benefit of score level fusion. Bigun et al. [13] propose a novel algorithm based on the Bayesian classifier that takes into account the estimated accuracy of the individual classifiers (i.e., matchers) during the fusion process. Sanderson and Paliwal [14] use a support vector machine (SVM) to combine the scores of face and speech experts. In order to address noisy input, they design structurally noise-resistant classifiers based on a piece-wise linear classifier and a modified Bayesian classifier.

Summary

In a multibiometric system, fusion at the score level offers the best tradeoff between amount of information that is available and ease of fusion. Hence, score level fusion is typically adopted by most multibiometric systems. Although a wide variety of score level fusion techniques have been proposed in the literature,

these can be grouped into three main categories, viz., density-based, transformation-based and classifier-based schemes. The performance of each scheme depends on the amount and quality of the available training data. If a large number of match scores is available for training the fusion module, then density-based approaches such as the likelihood ratio test can be used. Estimating the genuine and impostor distributions may not always be feasible due to the limited number of training samples that are available. In such cases, transformation-based schemes are a viable alternative. The nonhomogeneity of the match scores presented by the different matchers raises a number of challenges. Suitable score normalization schemes are essential in order to transform these match scores into a comparable domain. The sum of scores fusion method with simple score normalization (such as min-max) represents a commonly used transformation-based scheme. Classification-based fusion schemes consolidate the outputs of different matchers into a single vector of scores which is input to a trained classifier. The classifier then determines if this vector belongs to the “genuine” or “impostor” class.

Related Entries

- ▶ Fusion, Quality-Based
- ▶ Fusion, User-Specific
- ▶ Multibiometrics

References

1. Ross, A., Nandakumar, K., Jain, A.K.: *Handbook of Multibiometrics*. 1st edn. Springer, New York, USA (2006)
2. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*. Wiley, New York (2001)
3. Kittler, J., Hatef, M., Duin, R.P., Matas, J.G.: On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 226–239 (1998)
4. Dass, S.C., Nandakumar, K., Jain, A.K.: A principled approach to score level fusion in multimodal biometric systems. In: *Proceedings of Fifth International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pp. 1049–1058. Rye Brook, USA (2005)
5. Cappelli, R., Maio, D., Maltoni, D.: Combining fingerprint classifiers. In: *Proceedings of First International Workshop on Multiple Classifier Systems*, pp. 351–361. Cagliari, Italy (2000)

6. Snelick, R., Uludag, U., Mink, A., Indovina, M., Jain, A.K.: Large Scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 450–455 (2005)
7. Hampel, F.R., Rousseeuw, P.J., Ronchetti, E.M., Stahel, W.A.: Robust Statistics: The Approach Based on Influence Functions. Wiley, New York (1986)
8. Mosteller, F., Tukey, J.W.: Data Analysis and Regression: A Second Course in Statistics. Addison-Wesley, Reading, MA, USA (1977)
9. Brunelli, R., Falavigna, D.: Person Identification using multiple cues. *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 955–966 (1995)
10. Verlinde, P., Cholet, G.: Comparing decision fusion paradigms using k-NN based classifiers, decision trees and logistic regression in a multi-modal identity verification application. In: Proceedings of Second International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), pp. 188–193. Washington D.C., USA (1999)
11. Chatzis, V., Bors, A.G., Pitas, I.: Multimodal decision-level fusion for person authentication. *IEEE Trans. Syst. Man Cybernet. Part A: Syst. Humans* **29**, 674–681 (1999)
12. Ben-Yacoub, S., Abdeljaoued, Y., Mayoraz, E.: Fusion of face and speech data for person identity verification. *IEEE Trans. Neural Networks* **10**, 1065–1075 (1999)
13. Bigun, E.S., Bigun, J., Duc, B., Fischer, S.: Expert Conciliation for multimodal person authentication systems using bayesian statistics. In: First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA), pp. 291–300. Crans-Montana, Switzerland (1997)
14. Sanderson, C., Paliwal, K.K.: Information Fusion and Person Verification Using Speech and Face Information. Research Paper IDIAP-RR 02-33, IDIAP (2002)

Fusion, Sensor-Level

AFZEL NOORE, RICHA SINGH, MAYANK VASTA
West Virginia University, Morgantown, WV, USA

Synonyms

Fusion, data level; Fusion, image level

Definition

Sensor level fusion combines raw biometric information that can account for inter-class and intra-class variability and facilitate decision making based on the fused raw information. A typical sensor level fusion algorithm first integrates raw biometric data either

obtained from different viewpoints (for example, mosaicing several fingerprint impressions) or obtained from different sensors (for example, multimodal biometric images). The integrated data is then processed and discriminatory biometric features are extracted for matching. This level of fusion can be operated in both verification and ► **identification** modes. Few examples of sensor level fusion are: fingerprint mosaicing, multispectral face image fusion, and multimodal biometric image fusion.

Introduction

The concept of biometric information fusion is motivated from classical multi-classifier systems that combine information from different sources and represent using a single entity. Performance driven systems that use multiple biometric characteristics are known in multibiometric system [1]. These systems have several advantages over unimodal biometric systems such as tolerance to noise and malfunction, universality, and improved accuracy. Multibiometric systems are broadly classified into five levels of fusion.

1. *Sensor level fusion.* Raw data obtained directly from the sensors are fused without any feature extraction and represented as a single unit. This level of fusion is also known as data level fusion or image level fusion (for image based biometrics).
2. *Feature level fusion.* Data obtained from different sensors are first subjected to feature extraction algorithms and the feature sets are combined to generate a new feature vector which is subsequently used for recognition.
3. *Match score level fusion.* Features extracted from individual biometric modalities are first matched to compute the corresponding match scores. Match scores obtained from different biometric systems are then combined to generate a fused match score.
4. *Decision level fusion.* Decisions of individual biometric classifiers are fused to compute a combined decision. This level of fusion is also known as abstract level fusion because it is used when there is access to only decisions from individual classifier's.
5. *Rank level fusion.* With identification systems, rank level fusion involves combining identification ranks obtained from multiple unimodal biometrics. The output of rank level fusion is a consolidated rank that is used for final decision.

This article focuses on sensor level fusion and provides a comprehensive overview of the methodologies involved. In this level of fusion, first the raw data obtained from the sensors are combined to generate a fused data. An application oriented feature extraction algorithm is then used to compute the features from the fused data and matching is performed. [Figure 1](#) illustrates the basic concept of sensor level fusion.

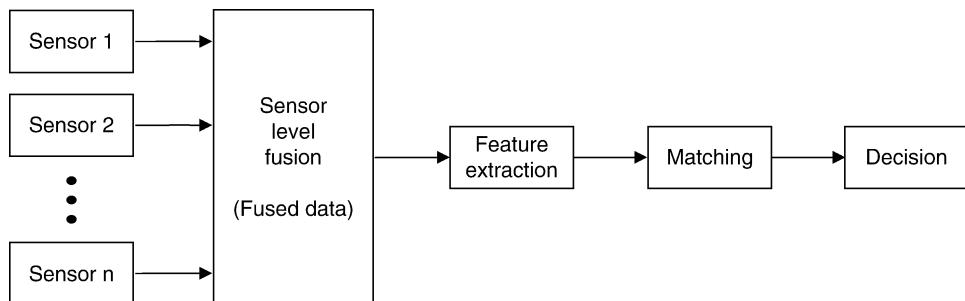
Sensor level fusion can be broadly classified into three categories: (1) single sensor multi-samples, (2) multi-sensor, and (3) multimodal. This article is organized to accentuate various algorithms proposed in each fusion category.

Sensor Level Fusion: Single Sensor Multi-Samples

In these systems, multiple samples of a single biometric modality are acquired using a single sensor and the information is combined to account for variations that can occur in a biometric modality. For example, as shown in [Fig. 2](#), different profiles of a face image can be combined to obtain a fused representation of face image that can address the challenges due to pose variations [\[2\]](#). In this category of sensor fusion,

image ▶ [mosaicing](#) techniques are used for integrating information obtained from several impressions or view points, to augment the biometric content and to enhance the verification/identification performance. Singh et al. [\[2\]](#) describe the concept of mosaicing in biometrics as an exercise in information fusion when multiple images of a subject's biometric information are fused into a single entity in the image domain itself. Therefore, this could be viewed as fusion at the sensor level.

Mosaicing was first introduced in biometrics by Ratha et al. [\[3\]](#). A rolled fingerprint image is generated from several partial fingerprint images using segmentation and blending algorithms assuming that the partial fingerprints are spatially registered. The performance of this fingerprint mosaicing algorithm is evaluated using different blending algorithms. The mosaicing algorithm generates rolled fingerprint image that is very close to the ground truth and improves the minutiae count that is useful for recognition. Further, Jain and Ross [\[4\]](#) proposed the use of iterative closest point algorithm to seamlessly register the ridges of two fingerprint images and to generate a composite fingerprint image. Recently, Ross et al. [\[5\]](#) employed thin-plate splines (TPS) to model the non-linear deformation in fingerprint images and integrate it



Fusion, Sensor-Level. [Figure 1](#) Basic concept of sensor level fusion.



Fusion, Sensor-Level. [Figure 2](#) Combining profile and frontal face images using mosaicing technique. (a) Profile and frontal face images and (b) Mosaiced face image.

in the mosaicing process. This algorithm first aligns two fingerprint images using coarse alignment (affine model) followed by TPS based fine alignment. Once the fingerprint images are registered, a mosaiced fingerprint image is obtained by applying simple pixel averaging based blending method.

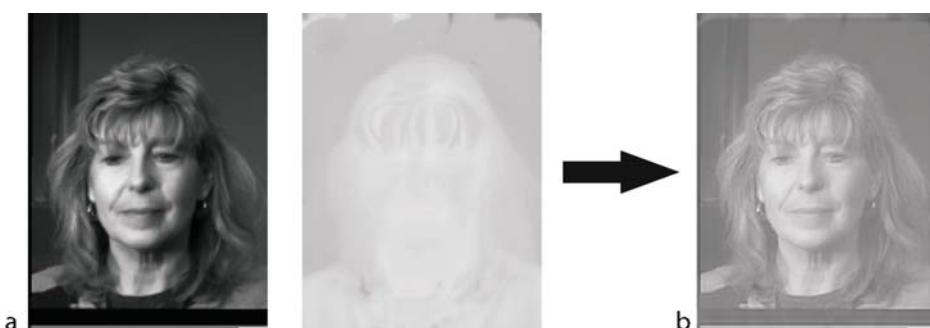
Mosaicing has also been applied to face biometrics. Yang et al. [6] describe an algorithm to create panoramic face mosaics. The acquisition system consists of five cameras that simultaneously obtains five different views of a subject's face. Based on the manually marked control points, the algorithm uses a series of linear transformations and smoothing operations on component images to generate a face mosaic. Unlike fingerprint mosaicing, face mosaicing requires specific feature extraction algorithm. Two different schemes to represent the panoramic image were proposed: one in the spatial domain and another in the frequency domain. Experimental evaluation on a database of 12 individuals shows that the face mosaicing algorithm improves identification accuracy in both the spatial and frequency domains. In [7], Liu and Chen describe a face mosaicing algorithm in which the human head is approximated with a 3D ellipsoidal model. The face, at a certain pose, is viewed as a 2D projection of this 3D ellipsoid. All 2D face images of a subject are projected onto this ellipsoid via geometrical mapping to form a texture map which is represented by an array of local patches. Matching is accomplished by adopting a probabilistic model to compute the distance of patches from an input face image. An identification accuracy of 90% on different databases has been reported. In [2], Singh et al. proposed a face mosaicing algorithm that can perform mosaicing in visible spectrum domain as well as in short wave infrared domain.

The algorithm first registers the component face images using two stage registration algorithm and then a face mosaic is generated using multi-resolution splines based blending algorithm. Facial features are encoded using a generic feedforward hierarchical model-based feature extraction algorithm that extracts local facial features using the fundamentals of a biological visual system. Experiments conducted on three different face databases indicate that the proposed face mosaicing algorithm offers significant benefits by accounting for pose variations that are commonly observed in face images. Moreover, the mosaicing algorithm requires less time for matching compared to the score level fusion and also reduces the memory requirements.

Sensor Level Fusion: Multi-Sensors

In this category of sensor level fusion, multiple samples of a single biometric modality are obtained using multiple sensors and the information is combined such that the fused multi-sensor information improves the recognition performance. In general, the information obtained from multiple sensors are complementary to each other and can account for the intra-class variability. For example, as shown in Fig. 3, multi-spectral face images obtained using visible spectrum and infrared sensors can be fused to minimize the intra-class variations due to illumination and expression.

Multi-spectral face image fusion is the classical model for this level of fusion. Face recognition algorithms generally use visible spectrum images for recognition because the reflectance property yields a



Fusion, Sensor-Level. **Figure 3** Multi-spectral face image fusion. **(a)** Visible and infrared spectrum images and **(b)** Fused image.

clear representation of facial features to differentiate between two individuals. However, visible spectrum images also possess several other properties which affect the performance of recognition algorithms. For example, changes in lighting affect the representation of visible spectrum images and can influence feature extraction. Other variations in facial appearance such as hairs, wrinkles, and expression are also evident in visible spectrum images and these variations increase the false rejection rate of face recognition algorithms. To address the challenges posed by visible spectrum images, researchers have used infrared images for face recognition [8]. Among all infrared spectrum images, long wave infrared (LWIR) images possess several properties that are complementary to visible images. Visible spectrum captures the electromagnetic energy in the range 0.4–0.7 μm, whereas long wave infrared or thermal images are captured in the range of 8–12 μm. Thermal images represent the heat pattern of the object and are invariant to illumination and expression. Face images captured in long wave infrared spectrum have less intra-class variation and help to reduce the false rejection rate of recognition algorithms. These properties of long wave infrared and visible images can be combined to improve the performance of face recognition algorithms.

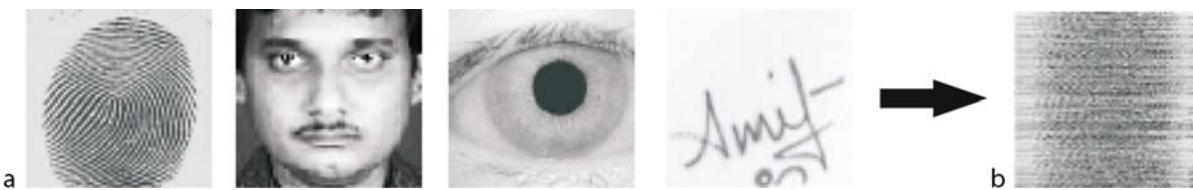
In literature, researchers have proposed several multi-spectral face image fusion algorithms [8]. Bebis et al. [9] proposed an image fusion algorithm in wavelet domain using genetic algorithm. In this algorithm, multi-spectral face images are first transformed into wavelet domain and a multiresolution representation is obtained. Then, a genetic algorithm is used to select the most appropriate wavelet coefficients at pixel level. Finally, inverse ▶ **wavelet transform** is applied to generate a fused face image and Eigenface based algorithm is used for feature extraction and matching. The genetic fusion algorithm suffers from making a good choice of fitness function. Kong et al. [10] proposed a wavelet based multi-spectral face image fusion algorithm in which the visible and infrared spectrum images are first registered using affine transformation. An empirical weighting scheme is then applied on the registered multi-spectral face images in wavelet domain to obtain the composite face image. Although the algorithm is straightforward, the generic empirical weighting scheme is not sufficient to address the inter-class and intra-class variability in face images. Recently, Singh et al. [11] proposed a 2v-granular support vector

machine (2v-GSVM) based multi-spectral face image fusion algorithm. This algorithm first registers multi-spectral face images using mutual information based registration algorithm. Then, a 2v-GSVM learning scheme is invoked in wavelet domain to learn the properties of the multi-spectral face images at different resolution and granularity levels, determine optimal information and combine them to generate a fused image. Finally, texture features are extracted from the fused image for recognition. Experimental results show that 2v-GSVM based fusion algorithm can address the challenges due to illumination, expression, and occlusion variations. This algorithm provides improved verification accuracy (>94%) compared to other image fusion schemes.

Another example of sensor level fusion with multiple sensors is fusing 2D and 3D facial information. Lu et al. [12] describe a semi-automatic sensor level fusion algorithm that integrates range and texture features for improved face recognition performance. Combining 3D shape information with registered 2D texture information using iterative closest point algorithm improves the face identification performance. The authors report that the algorithm is robust to arbitrary view, lighting, and facial appearance. However, the algorithm is computationally expensive and suffers due to non-rigid variations.

Sensor Level Fusion: Multimodal

In most of the multimodal biometric systems, such as bimodal system with face and fingerprint, fusion is performed at match score level or decision level. Very limited research is undertaken to perform sensor level fusion in a multimodal system. In this category of sensor level fusion, multimodal biometric images are fused to address issues such as universality, memory storage, small sample size recognition, and recognition performance. Further, an efficient sensor level fusion algorithm has advantages due to the availability of fused raw information from where the representative composite biometric features can be extracted and used for matching. The main challenge lies in developing fusion algorithm that can account for inter-class and intra-class variability in multimodal biometric images. An example of multimodal biometric image fusion is shown in Fig. 4.



Fusion, Sensor-Level. **Figure 4** Multimodal image fusion. **(a)** Image pertaining to different biometric modalities and **(b)** Fused and scrambled image.

Jing et al. [13] propose a sensor level fusion algorithm that generates a composite image from face and palmprint biometrics. Circular Gabor filters are first applied on face and palm print images to generate 32 filtered responses of each biometric data. These filtered responses are concatenated to generate a fused image. A pixel normalization scheme is then used to minimize variations due to imaging conditions. Finally, kernel discriminative common vectors are extracted from the fused image and radial basis function based neural network is used for classification. The fusion algorithm improves the recognition performance and is an effective solution for the small sample size recognition problem. Noore et al. [14] proposed discrete wavelet transformation based image fusion algorithm that generates a composite image by combining multimodal biometric images. The algorithm starts with transforming biometric images into wavelet domain and generating composite image by amalgamating the wavelet coefficients. The composite image is then scrambled using a secret encoding key generated with Fibonacci transforms. The algorithm not only improves the recognition performance but also reduces the memory requirements and provides resilience to common image processing attacks such as smoothing, cropping, JPEG 2000 compression, and filtering.

Future Research Directions

As discussed in previous sections, sensor level fusion has several advantages. However, compared to other levels of fusion, this level of fusion is less explored and requires further research to address the limitations of current research. First and foremost is to further improve the recognition accuracy. Researchers have shown that for certain applications, sensor level fusion algorithms do not provide better results compared to match score level fusion algorithms [5, 11]. This is mainly because existing algorithms do not effectively

reconcile the information that is useful for recognition. We believe that existing sensor level fusion algorithms fail in some cases because during information fusion it is possible that redundant and less discriminatory features become predominant. Furthermore, there is a lack of generalized sensor level fusion algorithms that can be used for different biometric scenarios or applications. For instance, genetic algorithm based multi-spectral image fusion algorithm can not be directly used for multimodal image fusion. Additional research is required to design an effective and generalized sensor level fusion algorithm which can be applied to different biometric modalities. Every sensor level fusion algorithm requires specific feature extraction algorithm that can effectively extract discriminatory biometric information from the composite image or data. This requirement is not mandatory with a generalized sensor level fusion algorithm. Therefore, a generalized algorithm can be easily incorporated in commercial systems and can conform to data fusion standards.

Another important research issue is to unify the sensor level fusion in a ► [unification framework](#) that reconciles multiple fusion algorithms. Originally proposed by Vatsa et al. [15], a biometric unification framework combines multiple fusion algorithms by dynamically selecting the most appropriate fusion algorithm depending on the input evidences such as quality and other priors. Currently, the unification framework includes only the match score fusion algorithms. However, with proper modifications, the unification framework can be expanded to include multi-level fusion algorithms that can address the operational needs of biometric systems and provide better recognition performance.

Related Entries

- [Data Fusion](#)
- [Face Recognition](#)

- ▶ Identification
- ▶ Verification

References

1. Ross, A., Nandakumar, K., Jain, A.K.: *Handbook of Multibiometrics*. 1st edn. Springer, New York (2006)
2. Singh, R., Vatsa, M., Ross, A., Noore, A.: A mosaicing scheme for pose-invariant face recognition. *IEEE Trans. Syst. Man Cybern. Part B* **37**(5), 1212–1225 (2007)
3. Ratha, N.K., Connell, J.H., Bolle, R.M.: Image mosaicing for rolled fingerprint construction. In: *Proceedings of International Conference on Pattern Recognition*, pp. 1651–1653 (1998)
4. Jain, A., Ross, A.: Fingerprint mosaicking. In: *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, pp. 4064–4067 (2002)
5. Ross, A., Shah, S., Shah, J.: Image versus feature mosaicing: a case study in fingerprints. In: *Proceedings of SPIE Conference on Biometric Technology for Human Identification III*, pp. 620,208-1–620,208-12 (2006)
6. Yang, F., Paindavoine, M., Abdi, H., Monopoly, A.: Development of a fast panoramic face mosaicing and recognition system. *Opt. Eng.* **44**(8), 087 005/1–087 005/10 (2005)
7. Lu, X., Jain, A.K.: Pose-robust face recognition using geometry assisted probabilistic modeling. In: *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 502–509 (2005)
8. Kong, S., Heo, J., Abidi, B., Paik, J., M., A.: Recent advances in visual and infrared face recognition - a review. *Comput. Vision Image Understand.* **97**(1), 103–135 (2005)
9. Bebis, G., Gyaourova, A., Singh, S., Pavlidis, I.: Face recognition by fusing thermal infrared and visible imagery. *Image Vision Comput.* **24**(7), 727–742 (2006)
10. Kong, S., Heo, J., Bougħorbel, F., Zheng, Y., Abidi, B., Koschan, A., Yi, M., M., A.: Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition. *Int. J. Comput. Vision* **71**(2), 215–233 (2007)
11. Singh, R., Vatsa, M., Noore, A.: Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recognit.* **41**(3), 880–893 (2008)
12. Lu, X., Jain, A.K.: Integrating range and texture information for 3D face recognition. In: *Proceedings of Workshop on Applications of Computer Vision*, pp. 156–163 (2005)
13. Jing, X.Y., Yao, Y.F., Zhang, D., Yang, J.Y., Li, M.: Face and palmprint pixel level fusion and Kernel DCV-RBF classifier for small sample biometric recognition. *Pattern Recognit.* **40**(11), 3209–3224 (2007)
14. Noore, A., Singh, R., Vatsa, M.: Robust memory efficient data level information fusion of multi-modal biometric images. *Inf. Fusion* **8**(4), 337–346 (2007)
15. Vatsa, M., Singh, R., Noore, A.: Unification of evidence theoretic fusion algorithms: A case study in level-2 and level-3 fingerprint features. In: *Proceedings of IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pp. 1–6 (2007)

Fusion, User-Specific

JULIAN FIERREZ, JAVIER ORTEGA-GARCIA
 Biometric Recognition Group – ATVS, Escuela Politecnica Superior, Universidad Autonoma de Madrid, Campus de Cantoblanco, Madrid, Spain

Synonyms

Adapted fusion; Local fusion; Target-dependent fusion; User-dependent fusion

Definition

User-specific fusion in the framework of biometrics, initially devised for score fusion in the verification mode, refers to techniques used for information fusion in which there is a specific fusion function for each user enrolled in the system. These fusion functions are retrieved and used for information integration in the same way the enrolled templates corresponding to the claimed identities are retrieved and used for matching.

User-specific fusion techniques find application in several biometric fusion scenarios, e.g., multi-modal fusion, where some subjects may be not adequate for recognition based on specific modalities (these evidences can be ignored or given less importance in the information fusion step), or multi-algorithm fusion, where some subjects may be better recognized based on particular algorithms (their fusion functions can be adapted to give more importance to those algorithms).

The biggest challenge for effective user-specific fusion is the need for user-specific training data, which is usually very scarce. Recent user-specific fusion techniques exploit the usually scarce training data by considering also for training the information provided by background users. These new techniques are known as adapted user-specific fusion.

System Model

The following nomenclature is used throughout the essay. Given a multi-biometric verification system consisting of a number of uni-modal systems, each one computes a similarity score between an input biometric pattern and the enrolled pattern or model of the

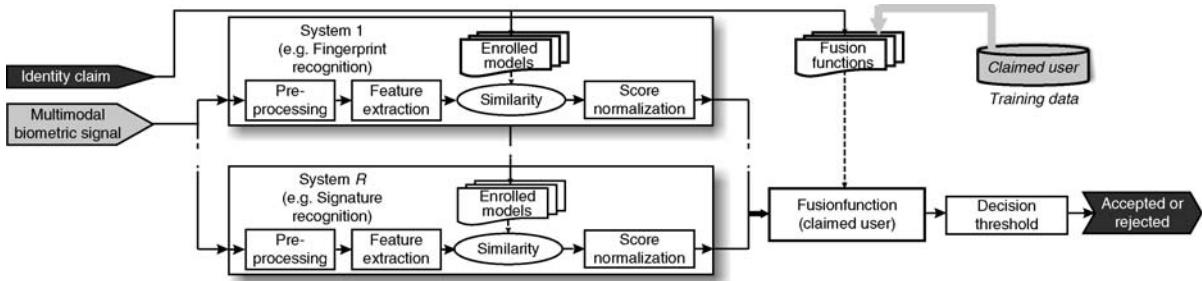
given claimant. The similarity scores are then normalized to a given score range. Let the normalized similarity scores provided by the different uni-modal systems be combined into a multi-modal score. The design of a fusion scheme consists in the definition of a function which maps a multi-modal score to a fused real value, so as to maximize the separability of client and impostor fused score distributions. This function may be fixed or trained (see the entry in this encyclopedia on Multi-biometrics) by using a set of training scores (scores known to be genuine or impostor).

The aim in user-specific fusion is to obtain the best score fusion function for a particular user, resulting in the system model shown in Fig. 1.

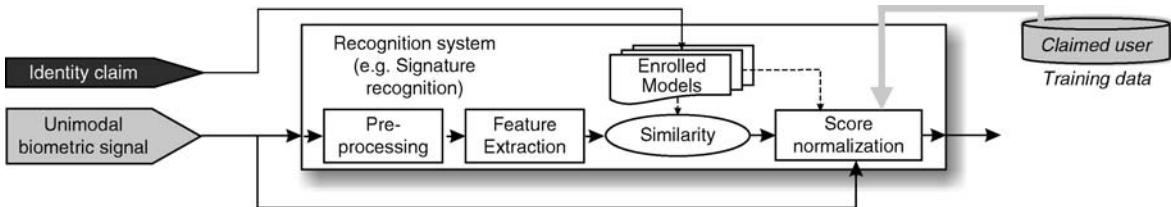
User-Specific Multi-Biometrics

User-specific multi-biometric verification can be achieved not only by making the fusion functions user-specific as shown in Fig. 1, but also other processing modules, such as the score normalization and the decision processing blocks. In the first case, each individual system will be used as indicated in Fig. 2, in the latter case the overall system diagram will be as indicated in Fig. 3.

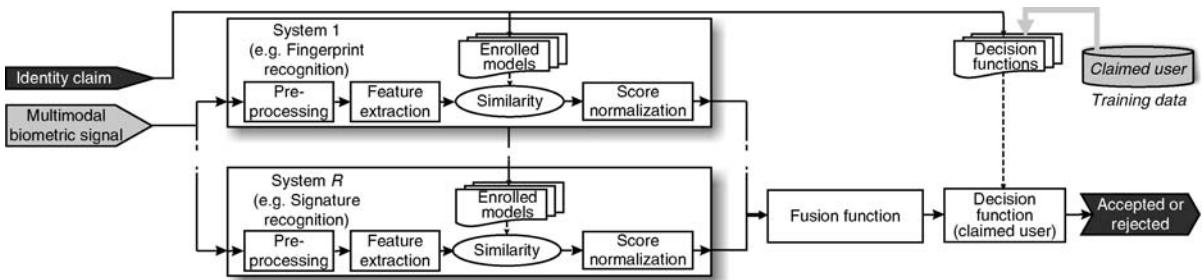
On one hand, user-specific score normalization has been traditionally studied for individual behavioral biometric modalities in which there are large variations between users (such as speech [1] or signature [2]),



Fusion, User-Specific. **Figure 1** System model of multi-biometric verification with user-specific score fusion.



Fusion, User-Specific. **Figure 2** System model of biometric verification with user-specific score normalization.



Fusion, User-Specific. **Figure 3** System model of multi-biometric verification with user-specific decision functions.

where their application is very effective to compensate the problems related to the heterogeneity between users. When user-specific score normalization is used in one of the systems being combined in a multi-biometric setup, the resulting approach can be seen as integrating the multi-biometric data in a user-specific way [3]. Despite the success of user-specific score normalization in individual modalities, and the success of fusion techniques, few efforts have been reported in the literature studying the combined use of both techniques to make the most out of the usually scarce user-specific training data.

On the other hand, the use of user-specific decisions in multi-biometrics has been typically studied in combination with user-specific score fusion. In this case, it has been demonstrated that it is better to use the available training data for computing user-specific fusion functions instead of user-specific decision schemes [4].

User-Specific Fusion

The idea of exploiting user-specific parameters at the score level in multi-modal biometrics was introduced, to the best of our knowledge, by [5]. In that work, user-independent weighted linear combination of similarity scores was demonstrated to be improved by using either user-specific weights or user-specific decision thresholds, both computed by exhaustive search on the testing data. The idea of user-specific fusion parameters was also explored by [6]. Other attempts to personalize multi-modal biometrics include the use of the claimed identity index as a feature for a global trained fusion scheme based on neural networks [7], computing user-specific weights using lambness metrics [8], and using personalized Fisher ratios [9].

The existing score fusion approaches can be classified as global or local depending first on the fusion function (i.e., user-independent or user-specific fusion strategies) and secondly on the decision making process (i.e., user-independent or user-specific decision thresholds), resulting in [10]: global-learning-global-decision (GG), local-learning-global-decision (LG), and similarly GL and LL. Some example works on user-specific multi-biometrics using this classification are: LG [4, 5, 6, 7, 8, 10, 11], GL [4, 5, 10], and LL [4, 10].

User-specific score fusion is confronted with a great challenge: the scarcity of user-specific training scores.

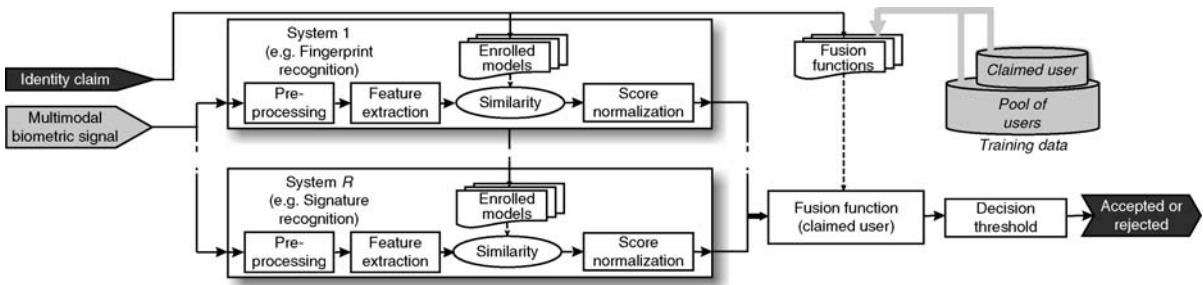
For overcoming this challenge, the simultaneous use of user-specific and background information has been proposed for training the user-specific fusion functions, in what has been called *adapted user-specific fusion*. This approach can be seen as a particular case of a more general type of approaches, referred to as ► **adapted fusion** [11]. In these approaches, a baseline fusion function is first constructed based on some general knowledge of the problem at hand, and then adjusted during the operation of the system. The adaptation can be based on ancillary information such as: the user being claimed (adapted user-specific fusion), quality measures of the input biometrics (quality-based fusion [12], see related entry in this encyclopedia), or other kind of environmental information affecting the various information channels being fused.

Adapted User-Specific Fusion

Adapted methods in the context of user-specific fusion refer to the use of both global and local information for learning the fusion functions.

The idea of adapted learning is based on the fact that the amount of available training data in localized learning is usually not sufficient and representative enough to guarantee good parameter estimation and generalization capabilities. To cope with this lack of robustness derived from partial knowledge of the problem, one can exploit the information provided by background global data. In general, the relative balance between the background information (pool of users) and the local data (specific user) is performed as a tradeoff between both kinds of information.

The system model of adapted user-specific score fusion is shown in Fig. 4, where we can see that the fusion function of a given user is trained with two sets of training data, both including both genuine and impostor matching scores. The first training set consists of scores corresponding to the user being claimed. The second set consists of scores corresponding to a pool of background users different to the user being claimed. By considering these two sets simultaneously, the resulting adapted user-specific fusion schemes outperform the traditional user-independent fusion (also known as ► **global fusion**, in which only the pool of users is used for training), and the traditional user-specific fusion depicted in Fig. 1 (also known as ► **local fusion**, in which only data from the claimed



Fusion, User-specific. **Figure 4** System model of multi-modal biometric authentication with adapted user-specific score fusion.

user is used for training). This affirmation has been demonstrated experimentally in various scenarios, such as multi-algorithm speaker verification [3, 13], and multi-modal verification combining on-line signature and fingerprint traits [4, 11].

Related Entries

- ▶ [Fusion, Quality-Based](#)
- ▶ [Multi-Algorithm Systems](#)
- ▶ [MultiBiometrics](#)
- ▶ [Multi-Modal Systems](#)

References

1. Doddington, G., Liggett, W., Martin, A., Przybocki, M., Reynolds, D.: Sheeps, goats, lambs and wolves: a statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. In: Proceedings of International Conference on Speech and Language Processing, ICSLP (1998)
2. Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J.: Target dependent score normalization techniques and their application to signature verification. IEEE Trans. Syst. Man Cybernet., Part C **35**, 418–425 (2005)
3. Poh, N., Kittler, J.: Incorporating model-specific score distribution in speaker verification systems. IEEE Trans. Audio Speech Lang. Process. **16**, 594–606 (2008)
4. Fierrez-Aguilar, J., Garcia-Romero, D., Ortega-Garcia, J., Gonzalez-Rodriguez, J.: Adapted user-dependent multimodal biometric authentication exploiting general information. Pattern Recognit. Lett. **26**, 2628–2639 (2005)
5. Jain, A.K., Ross, A.: Learning user-specific parameters in a multi-biometric system. In: Proceedings of IEEE International Conference on Image Processing, ICIP, vol. 1, pp. 57–60 (2002)
6. Wang, Y., Wang, Y., Tan, T.: Combining fingerprint and voice biometrics for identity verification: An experimental comparison. In: Zhang, D., Jain, A.K. (eds.) Proceedings of International Conference on Biometric Authentication, ICBA, Springer LNCS-3072, pp. 663–670 (2004)
7. Kumar, A., Zhang, D.: Integrating palmprint with face for user authentication. In: Proceedings of Workshop on Multimodal User Authentication, MMUA, pp. 107–112 (2003)
8. Snelick, R., Uludag, U., Mink, A., Indovina, M., Jain, A.K.: Large scale evaluation of multimodal biometric authentication using state-of-the-art systems. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 450–455 (2005)
9. Poh, N., Bengio, S.: An investigation of f-ratio client-dependent normalisation on biometric authentication tasks. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, vol. 1, pp. 721–724 (2005)
10. Toh, K.A., Jiang, X., Yau, W.Y.: Exploiting local and global decisions for multimodal biometrics verification. IEEE Trans. Signal Process. **52**, 3059–3072 (2004)
11. Fierrez-Aguilar, J., Garcia-Romero, D., Ortega-Garcia, J., Gonzalez-Rodriguez, J.: Bayesian adaptation for user-dependent multimodal biometric authentication. Pattern Recognit. **38**, 1317–1319 (2005)
12. Fierrez-Aguilar, J., Ortega-Garcia, J., Gonzalez-Rodriguez, J., Bigun, J.: Discriminative multimodal biometric authentication based on quality measures. Pattern Recognit. **38**, 777–779 (2005)
13. Fierrez-Aguilar, J., Garcia-Romero, D., Ortega-Garcia, J., Gonzalez-Rodriguez, J.: Speaker verification using adapted user-dependent multilevel fusion. In: Oza, N.C., Polikar, R., Kittler, J., Roli, F. (eds.) Proceedings of International Workshop on Multiple Classifier Systems, MCS, Springer LNCS-3541, pp. 356–365 (2005)

Fusion, Wavelet-Based

Wavelet-based fusion has been widely used in literature. The wavelet transform is a data analysis tool that provides a multi-resolution decomposition of an image. Wavelet-based pixel-level data fusion is used on two or more sets of probe images. Given two

registered images I_1 and I_2 of the same object from two sets of probe images (two different spectral bands in this case), a two-dimensional discrete wavelet decomposition is performed on I_1 and I_2 to obtain the wavelet approximation coefficients (a_1, a_2) and detail coefficients (d_1, d_2). The wavelet approximation and detail coefficients of the fused image, a_f and d_f , are then calculated as follows:

$$a_f = W_{a_1} \times a_1 + W_{a_2} \times a_2 \quad \text{and}$$
$$d_f = W_{d_1} \times d_1 + W_{d_2} \times d_2,$$

where $W_{a_1}, W_{a_2}, W_{d_1}$, and W_{d_2} are weights determined either empirically or according to some selected rule. The two-dimensional discrete wavelet inverse transform is then performed to obtain the fused image.

► [Multispectral and Hyperspectral Biometrics](#)

Fuzzy Extractor

- [Encryption, Biometric](#)
- [Fingerprints Hashing](#)

Fuzzy Vault

Fuzzy vault is where a secret key is hidden behind some biometric data which are fuzzy and noisy by nature.

- [Fingerprints Hashing](#)

