Chapter 26
Machine Learning for Biometrics

Albert Ali Salah
Centre for Mathematics and Computer Science (CWI), The Netherlands

ABSTRACT

Biometrics aims at reliable and robust identification of humans from their personal traits, mainly for security and authentication purposes, but also for identifying and tracking the users of smarter applications. Frequently considered modalities are fingerprint, face, iris, palmprint and voice, but there are many other possible biometrics, including gait, ear image, retina, DNA, and even behaviours. This chapter presents a survey of machine learning methods used for biometrics applications, and identifies relevant research issues. The author focuses on three areas of interest: offline methods for biometric template construction and recognition, information fusion methods for integrating multiple biometrics to obtain robust results, and methods for dealing with temporal information. By introducing exemplary and influential machine learning approaches in the context of specific biometrics applications, the author hopes to provide the reader with the means to create novel machine learning solutions to challenging biometrics problems.

INTRODUCTION

Biometrics serves the identification of humans from their personal traits. As a rapidly growing field, it is initially pushed forward by a need for robust security and surveillance applications, but its potential as a natural and effortless means of identification also paved the way for a host of smart applications that automatically identify the user and provide customized services. With increasing awareness of its psychological, privacy-related and ethical aspects, there is no doubt that biometrics will continue to contribute to many technological solutions of our daily lives.

DOI: 10.4018/978-1-60566-766-9.ch026

The technology of biometrics relies on the input from a number of fields, starting with various kinds of sensors that are used to sample the biometric. Signal processing and pattern recognition methods are obviously relevant, as the acquired data need to be prepared for accurate and robust decisions. At its final stage, the system outputs a decision, which links the acquired and processed biometric trait to an identity. Algorithms and mathematical models developed by the machine learning community are frequently used in biometric systems to implement the decision function itself, but this is surely not the only contribution worth mentioning. We will show in this chapter that machine learning methods are useful in selecting appropriate feature representations that will facilitate the job of the decision function, in dealing with temporal information, and in fusing multi-modal information.

The goal of this chapter is to familiarize the machine learning researcher with the problems of biometrics, to show which techniques are employed to solve them, and what challenges are open in the field that may benefit from future machine learning applications. It is also intended to familiarize the biometrics researcher to the methods and ways of machine learning and its correct research methodology, and to provide the rudiments of a toolbox of machine learning. In the next section, we provide a general look at biometric systems, define some relevant terminology and broadly identify the research issues. The third section deals with learning and matching biometric templates. Since this is a very broad topic, a small number of demonstrative application examples are selected. The fourth section is on the use of dynamic information for biometric purposes, and it is followed by a section on the fusion of multiple biometrics. Before concluding, we give a machine learning perspective on how to evaluate a biometrics system.

A GENERAL LOOK AT BIOMETRIC SYSTEMS

The application area of biometrics with the grandest scale is in *border control*, typically an airport scenario. Within the national identity context, it is possible to conceive the storing and managing of the biometric information for the entire population of a country. A smaller scale application is *access control*, for instance securing the entrance of a building (*physical access control*) or securing a digital system (*logical access control*). In both applications, we have a *verification* (or *authentication*) problem, where the user has an identity claim, and a sampled biometric is checked against a stored biometric for similarity. In a sense, this is a one-class pattern classification problem.

The second important problem involves *identification*, where there is no identity claim, and the sampled biometric is matched against many stored *templates*. Checking passengers against a list of criminals, forensic applications, identification of individuals at a distance, or providing access in consumer products (e.g. fingerprint scanning on a laptop) would be typical applications. Depending on the application requirements, the problem may be sufficiently constrained to apply a discriminative approach.

The most important biometric modalities are fingerprint, face, iris, signature, palm print and voice. The biometric traits differ in their usability, convenience, security, and complexity. For providing access to a high-security facility, security is of primary importance, whereas a household appliance that identifies users via biometrics would strive to have maximum user convenience. Similarly, privacy can be a major determinant in the deployment of a particular biometric application. For this reason, a host of possible biometrics are considered for different applications, including DNA, gait, ear images, and even behaviours.

Basic Biometrics Terminology

In this section, we define frequently used terminology for research in biometrics. We have already defined identification and verification. The biometric *template* $\boldsymbol{t}_i$ (also called a *target*) is a recorded instance of the biometric trait during *enrollment* for subject *i*, and it is stored in a *gallery* of templates, denoted here with $T=\{\boldsymbol{t}_i \mid i=1...\ N\}$, where *N* is the number of *subjects* in the gallery. A *query* (or a *probe*) is a biometric recorded during the operation of the system, denoted with $\boldsymbol{B}$, and it is matched against the templates in the gallery.

The accuracy of a biometric authentication system is measured by its *false acceptance rate* (*FAR*) and its *false rejection rate* (*FRR)*, where the former is the probability of accepting an *impostor*, and the latter is the probability of rejecting a *genuine* claim. More formally, assume $s(\boldsymbol{B}_i,\boldsymbol{B}_j)$ denotes a similarity function for the recorded biometric, and τ is a threshold for accepting an identity claim. The identity claim for subject *d* with a recorded biometric $\boldsymbol{B}$ is accepted if and only if:

$$s(\boldsymbol{B}, \boldsymbol{t}_d) > \tau. \tag{1}$$

The threshold τ determines the convenience-security trade-off of the system, and a higher threshold means that more subjects will be rejected, thus a lower FAR will be obtained at the expense of a higher FRR. Low FAR is required for high-security systems, whereas low FRR means greater user convenience. Commonly the authentication performance of a system is reported with its FRR at 0.001 FAR.

As a side remark, we caution the reader to pay attention to the impostor model used for reporting the performance of a biometrics system. The impostor claims can be based on *zero-cost* attacks, where there is no particular effort on the part of the impostor, or on *informed* attacks. The former approach is exemplified by using all the test samples in a database for one genuine and *N-1* impostor claims, where *N* is the number of samples in the gallery. The latter approach is exemplified by creating expertly forged signatures for attacks on a signature-verification system, which is expected to produce a much higher FAR than the zero-cost approach.

The similarity function can be substituted by a distance function, in which case the threshold serves as an upper-bound for authentication. Especially in non-parametric models, the distance function plays an important role, and needs to be chosen carefully to suit the distribution of the data and the outliers. We briefly list frequently used distance functions here.

The most common distance function is the Euclidean distance:

$$D_{Euclidean}(\boldsymbol{x}, \boldsymbol{y})=\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2}$$

where *n* denotes the biometric vector dimensionality. For a known distribution of biometric samples with covariance denoted by *S*, the squared Mahalanobis distance gives less weight to deviations in directions along which the data are scattered:

$$D^2_{Mahalanobis}(\boldsymbol{x}, \boldsymbol{y})=(\boldsymbol{x} - \boldsymbol{y})^{\mathrm{t}} S^{-1}(\boldsymbol{x} - \boldsymbol{y}) \tag{3}$$

When feature mismatch has a fixed cost, one frequently uses the Hamming distance:

$$D_{Hamming}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} |x_i \oplus y_i| \tag{4}$$

where $(x \oplus y)$ is zero if $x = y$, and 1 otherwise. For binary vectors, this is equivalent to the XOR operation. For two finite point sets, the Hausdorff distance is proposed:

$$D_{Hausdorff}(X, Y) = \max(\sup_{\boldsymbol{x} \in X} \inf_{\boldsymbol{y} \in Y} D(\boldsymbol{x}, \boldsymbol{y}), \sup_{\boldsymbol{y} \in Y} \inf_{\boldsymbol{x} \in X} D(\boldsymbol{x}, \boldsymbol{y})) \tag{5}$$

where sup (*supremum*) denotes the least upper bound, and inf (*infimum*) denotes the greatest lower bound, respectively. This measure is used for instance in 3D face recognition, where each face may be represented as a point set.

Another important distance measure is based on correlation:

$$D_{Correlation}(\boldsymbol{x}, \boldsymbol{y}) = - \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{(n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)(n \sum_{i=1}^{n} \boldsymbol{y}_i^2 - (\sum_{i=1}^{n} y_i)^2)}} \tag{6}$$

The similarity function can also be replaced with a generative model, where the output takes on a probabilistic interpretation, and the threshold of acceptance is defined on the posterior probability for the biometric sample under the model associated with the claimed identity:

$$s(\boldsymbol{B}, \boldsymbol{t}_d) = P(M_d | \boldsymbol{B}) = \frac{p(\boldsymbol{B} | M_d) p(M_d)}{p(\boldsymbol{B} | M_d) p(M_d) + p(\boldsymbol{B} | I) p(I)} \tag{7}$$

Here, M_d denotes the generative model for person d and used in place of the template $\boldsymbol{t}_d$, I denotes a model for the impostor distribution, and $p(M_d)$ denotes the prior probability of having a genuine claim for person d. The probabilistic formulation is particularly suitable for biometric fusion approaches, where multiple biometric samples are treated as evidence, and the Bayes rule is again used for combining the modalities. We will deal with fusion of multiple biometric modalities in a dedicated section.

An alternative approach of authentication under a client and a generic impostor model is to use the likelihood ratio test:

$$s(\boldsymbol{B}, M_d, I) = \frac{p(\boldsymbol{B} | M_d)}{p(\boldsymbol{B} \mid I)} \tag{8}$$

where the impostor distribution under infinitely many classes can be replaced with $P(\boldsymbol{B})$. In (Bazen & Veldhuis, 2004), the authors demonstrate that this measure is more accurate than the posterior probability approach or the distance-based approach with a Euclidean metric for a fingerprint-based authentication application.

Often, the accuracy of the verification system is shown with a *receiver operator characteristic* (ROC)

or a *detection error trade-off* (DET) curve, where FRR is plotted as a function of FAR for a range of τ. Other accuracy indicators are the *equal error rate* (EER), which is the error rate at the point where FAR equals FRR in the DET curve, the *half-total error rate* (HTER), which is the average of FAR and FRR at a given threshold, the *genuine acceptance rate* (GAR), which is equal to 1 – FRR, and thus only takes into account the genuine claims, and the *weighted error rate* (WER) which is a weighted version of the EER. The weight *R* determines how much a false accept is harmful with respect to a false reject:

$$WER(R) = \frac{P_{FR} + RP_{FA}}{1 + R} \tag{9}$$

where P_{FR} and P_{FA} denote the probabilities of false reject and false accept, respectively.

For an identification system, the gallery templates can be ranked according to their similarity to the query. The accuracy of an identification system is often indicated by the *cumulative match characteristic* (CMC) curve, which plots the average rank of the correct gallery template in response to the query. The *rank-n recognition rate* is the correct identification percentage if we assume that a rank below (and including) *n* is acceptable for identification. The rank system makes sense if a human or a second (and possibly computationally more intensive) system will evaluate the top *n* candidates for a match with the probe. For the final identification system, the *rank-1 recognition rate* is the most important accuracy measure.

The Structure of a Biometrics System

The information flow in a biometrics system depends on the biometric modality, as well as the operation requirements of the system. For most biometric systems, the operation of the system has the following steps:

1. Acquisition: The acquisition of the biometric is the most important step in the system design. A clean signal that matches the gallery conditions well greatly facilitates recognition. For this reason, biometric evaluations usually test cases where the gallery is acquired under controlled environmental conditions with high quality sensors, and matching is performed under sub-optimal conditions, possibly with inferior sensors.
2. Pre-processing: The biometric data is prepared for the next stage of processing. Under pre-processing, we usually understand cleaning of noise artifacts, normalization, cropping, and signal enhancement operations. In voice biometrics, this step involves segmentation of the speech/non-speech signals. For iris biometrics, the pre-processing deals with illumination changes, contrast normalization, elimination of reflections, defocus and occlusion handling.
3. Registration and segmentation: The biometric template stored in the gallery and the recorded signal must be aligned for obtaining the best results in matching. For face authentication, this is usually performed by using facial landmarks (Tistarelli et al., 2008, Gökberk et al., in press). For iris biometrics, the iris area is segmented at this stage. For modalities with dynamic length, the registration and classification may be performed jointly. We inspect these methods in a separate section.

4. Feature selection and extraction: The raw biometric signal is rarely used for the purposes of classification. The feature extraction method depends on the modality, and influences the subsequent classification stage. For instance in fingerprint-based authentication, the structural features of the fingerprint (i.e. *minutiae*) are extracted at his stage. For iris biometrics, typical features are log-Gabor wavelet coefficients (Daugman, 1988), whereas for voice-based authentication Mel frequency cepstral coefficients are used (Huang et al., 2001).
5. Classification: The classifier decides whether the biometric signal $\boldsymbol{B}$ is generated by the model M or not. Depending on the problem being one of authentication or identification, the classification step assesses the biometric signal against a single model or multiple models. This step may also be followed by a score-based fusion step, if multiple biometric traits are recorded.

Challenges of Biometrics

There are several challenges faced by biometric systems. We mention some of these issues, as they are indicative of future research directions.

Processing and communication load: There is an obvious need for computationally efficient algorithms in biometrics, for several reasons. Some biometrics approaches require great computational resources. For instance in 3D face recognition, the data acquired from the sensor can be massive. The second reason is the real-time requirement imposed by application settings. It is now possible to match a query against millions of fingerprints in a second, but this eludes other modalities at the moment. There is also the issue for algorithms that assess the class-specific variability for a particular modality. For instance the newly acquired Multi-PIE face dataset (Gross et al., 2008) has more than 750.000 samples at 3072 x 2048 resolution (i.e. 6.3 million-dimensional samples), which makes the job of any learning algorithm very difficult.

Template protection: Security and privacy of biometric systems depend on the protection of stored biometric information. Once a biometric is compromised, it is irrevocable, as it is unique to a person. Studies are conducted on *biometric encryption*, where the biometric is combined with a random number to produce a secure and non-reversible hash for authentication.

Cancellable biometrics: For improved privacy, it is possible to store a transformed version of a biometric template. Each application will be associated with a different one-way transformation, and *function creep*, i.e. the use of stored template for other purposes, will be prevented. This kind of a requirement imposes extra constraints on the biometric matching algorithm.

Aging: The aging of the individual causes differences in the stored template and the acquired biometric signal. An aging face has more lines, and it can be difficult to recognize a person from a photograph taken 20 or 30 years ago. It is difficult to collect datasets that span long temporal acquisition periods. Therefore, artificial aging methods are applied to existing data collections to train systems that recognize an aging set of subjects.

Convenience vs. security: The parameterization of a biometric system may allow for a range of custom settings that trade off user convenience and security. For instance, biometrically enhanced entry system of an entertainment park offers high convenience by setting a very low threshold for user admittance, whereas the same system with a high threshold is usable in a high-security nuclear plant.

Score normalization and fusion: Multiple biometric modalities can be used in a single system, which requires efficient and robust score normalization and information fusion algorithms. The best fusion algorithms take into account the correlation structure of the related modalities.

LEARNING AND MATCHING THE BIOMETRIC TEMPLATE

The learning and matching of the biometric template is an integrated problem that encompasses feature selection/extraction and classification, as these steps are invariably intertwined. One may think that once an appropriate pre-processing and feature extraction method is selected, the choice of the classifier is a minor issue. This is not completely true, as literature is replete with studies that contrast various classification approaches for a fixed set of features with significantly different results. Here we give illustrative examples from face recognition domain, and refer the reader to (Kung et al., 2005) and (Petrovska-Delacrétaz et al., 2008) for many applications of learning algorithms for different biometric modalities and their fusion. Our exposition is restricted to two subspace methods for feature extraction to reduce dimensionality, the application of unsupervised clustering to facial landmarking, and combination of weak classifiers for face detection. A later section on dynamic information will deal with Bayesian approaches.

Classification with Subspace Methods

The subspace-based methods are best exemplified with applications to face recognition, which is a high-dimensional problem. The most frequently used baseline method for face recognition is the *Eigenface* method, introduced by Turk & Pentland (1991). This method is based on learning a subspace for faces, where the projected query and target images will be compared. Given a set of training images $\{T_1,... T_k\}$, the covariance matrix C that indicates the distribution of variance is computed:

$$C = \frac{1}{k}\sum_{i=1}^{k}(T_i - \mu)(T_i - \mu)^T \tag{10}$$

where μ denotes the average face. The eigenvectors of C with the greatest associated eigenvalues (i.e. the *principal components*) denote axes of maximal covariance, and a projection to these axes maximally spreads the projected points. In terms of reconstruction from the projected subspace, the principal component analysis (PCA) method is optimal. If the dimensionality of each image is assumed to be d, the covariance matrix has a dimensionality of (d x d), usually with $d >> k$. For this reason, a *singular value decomposition* based method is used to determine at most k eigenvectors for projection. Once the face images are projected to the subspace, several distance measures (e.g. L1, Euclidean) can be used for computing the similarity of the query and the target with nearest-neighbour classification. One measure that has received a lot of interest is the *Mahalanobis cosine distance*. If the p-dimensional subspace projected query is denoted by $\boldsymbol{u}=[u_1u_2...u_p]$, and the subspace-projected gallery template is denoted by $\boldsymbol{v}=[v_1v_2...v_p]$, denote their corresponding vectors in the Mahalanobis space with unit variance along each dimension as:

$$m_i = \frac{u_i}{\sigma_i} \qquad n_i = \frac{v_i}{\sigma_i} \tag{11}$$

where σ_i is the standard deviation for the i^{th} dimension of the p-dimensional eigenspace. Then the Mahalanobis cosine distance is given by:

$$d_{MC}(\boldsymbol{u}, \boldsymbol{v}) = \cos(\theta_{mn}) = \frac{\boldsymbol{mn}}{|\boldsymbol{m}||\boldsymbol{n}|} \tag{12}$$

with θ_{mn} denoting the angle between vectors $\boldsymbol{m}$ and $\boldsymbol{n}$ (Ramanathan et al., 2004).

The PCA projection is not optimized for discrimination of patterns. For this reason, *linear discriminant analysis* (*LDA*) is proposed as an alternative method that seeks the most discriminative subspace projection. Specifically, for c classes denote the grand mean of the training images with $\boldsymbol{\mu}$ and the class means with $\boldsymbol{\mu}_i$. Let the between-class scatter matrix be defined as:

$$S_B = \sum_{i=1}^{c} k_i (¼_i - ¼)(¼_i - ¼)^T \tag{13}$$

where k_i denotes the number of training samples for class i. Similarly, the within-class scatter matrix is defined as:

$$S_W = \sum_{i=1}^{c} \sum_{B_j \in T_i} (\boldsymbol{B}_j - \boldsymbol{\mu}_i)(\boldsymbol{B}_j - \boldsymbol{\mu}_i)^T \tag{14}$$

with T_i denoting training samples of class i. Assuming that S_W is non-singular, the optimal orthonormal projection W_{opt} is selected as the one that maximizes the ratio of the determinant of S_B to the determinant of S_W:

$$W_{opt} = \arg\max_W \frac{|W^T S_B W|}{|W^T S_W W|} = [W_1 \quad W_2 \quad \ldots \quad W_m] \tag{15}$$

where $\{\boldsymbol{w}_i | i=1,2,..., m\}$denotes the set of generalized eigenvectors of S_B and S_W corresponding to the m largest generalized eigenvalues λ_i:

$$S_B \boldsymbol{w}_i = \lambda_i S_W \boldsymbol{w}_i \tag{16}$$

The *Fisherface* method (Belhumeur et al., 1997) proposes an alternative criterion to equation (15) to overcome difficulties arising from singular S_W. In the proposed method, the images are projected to a PCA space first. This subspace has a sufficiently low dimensionality (equal to c, the number of classes) to make S_W non-singular. The classification, as before, uses a nearest neighbour approach.

Classifier Combination for Face Detection

Face detection is the essential pre-processing step in face biometrics. This application can be treated as a two-class learning problem, where the positive samples contain face images, and negative samples do not. Osuna et al. (1997) have successfully applied *support vector machines* (SVM) to this problem, and obtained good results. However, practical applications require very fast face detection, for which the SVM-based method was not adequate.

The combination of *weak* classifiers has given rise to one of the most frequently used algorithms in face detection, namely the Viola-Jones algorithm (Viola & Jones, 2001). The core of the method uses the AdaBoost algorithm proposed by Freund and Schapire (1997). In AdaBoost, a sequence of N labelled examples $<(\boldsymbol{x}_1, y_1), \ldots (\boldsymbol{x}_N, y_N)>$ is used (where $\boldsymbol{x}_i$ denote the samples, and y_i are binary labels) for finding a good combination of weak learners. The algorithm sets weights w_i for each sample, initialized by a prior distribution that can be uniform in the case no prior information exists. At every iteration of the algorithm, the normalized weights p_i are computed to form a distribution over samples, and the weak learner is called with this distribution to obtain a hypothesis $h_t: X \rightarrow [0,1]$. The error of this hypothesis is given by:

$$\varepsilon_t = \sum_{i=1}^{N} p_i^t |h_t(\boldsymbol{x}_i) - y_i| \tag{17}$$

The algorithm sets a weight update parameter $\beta_t = \varepsilon_t/(1-\varepsilon_t)$. The new weights are set as:

$$w_i^{t+1} = w_i^t \beta_t^{1-|h_t(\boldsymbol{x}_i)-y_i|} \tag{18}$$

After T iterations, the output of the algorithm is the hypothesis:

$$h_f(\boldsymbol{x}) = \begin{cases} 1 & if \quad \sum_{t=1}^{T} (\log 1/\beta_t) h_t(\boldsymbol{x}) \geq \frac{1}{2} \sum_{t=1}^{T} \log 1/\beta_t \\ 0 & otherwise \end{cases} \tag{19}$$

Viola and Jones use the AdaBoost algorithm to select features computed on sub-windows of the image that resemble Haar-wavelets computed in multiple scales (Viola & Jones, 2001). Once computed over the entire image, there are 180.000 possible features available for selection at each stage. In the Viola-Jones algorithm many of these features have final weights set to zero; only 6.000 features are selected. These features are learned in a cascade structure with 38 levels to reduce the processing load of the algorithm. Thus, an average of 10 feature evaluations per sub-window is reported in their experimental study. For the training, 4916 face images (plus their mirror images) and 9544 non-face images (for a total of 350 million sub-windows) were used. This method is currently used as the state-of-the-art benchmark in systems involving face detection[1].

Unsupervised Learning for Feature Localization

Exact localization of facial features in faces ensures better *registration*, subsequently allowing better classification of face images. These prominent features that guide registration are called *anchor points* or *landmarks*. The learning problem for landmarks has two aspects:

1) Learning the appearance of each landmark.
2) Learning the structural relationship between the landmarks.

It is possible to learn the appearance of each landmark via unsupervised models. For this purpose, features are extracted from around the landmark location from all images in the training set, for each landmark j. Then, a model G_j is fit to the feature distribution. In (Hamouz et al., 2005), complex Gabor features are extracted, and a Gaussian mixture model (GMM) is used for learning the feature distribution. The structural information is taken into account at a later stage. For all the candidates of one landmark (i.e. locations with likelihood exceeding a threshold), two other landmark locations are estimated from the location ground-truth given in the training set, and candidates selected for evaluation. This produces a certain number of landmark triplets for assessment. An SVM classifier is used to produce the final decision.

The GMM (including *k-means clustering* as a special case) is the most frequently used method of clustering in biometrics. The fitting of the GMM involves two decisions: the number of components, and the covariance restrictions imposed on the components. Assuming a particular covariance shape will simplify the observation model, and reduce the number of parameters. For an observation feature of size d, the Gaussian distribution has d parameters for the mean vector. The number of free parameters for the covariance can be as low as one, if a spherical covariance is adapted with the shape of $\sigma^2 I$. Setting σ to unity will result in the *k-means* model. A diagonal covariance will have d parameters, and a full covariance $d(d-1)/2$ parameters. For a mixture model, the number of parameters is multiplied by K, the number of mixture components, plus $(K-1)$ for the component priors. Each simplification corresponds to a different assumption in the model space, and may have adverse effects on the learned model. A diagonal covariance matrix will correspond to assuming independence of observed features, whereas a spherical covariance will also imply that the noise model is the same for each feature dimension.

It should be noted that in modelling a very complex distribution, a large number of simpler Gaussian distributions may be a better solution than fewer full-covariance distributions, particularly due to computational reasons. For instance in the MIT Lincoln Laboratory's successful speaker verification system, a *universal background model* with 2048 diagonal-covariance Gaussian components was employed (Reynolds et al., 2000). An alternative approach for reducing dimensionality is given in (Salah & Alpaydın, 2004), where the mixture components are expressed as *factor analysis* models, which assume that the data are generated in p_i-dimensional manifolds. In this case, each component indexed with i has $d(p_i+1)$ parameters for expressing the covariance, which means that the complexity scale between diagonal- and full-covariance models can be fully explored.

The number of components in a Gaussian mixture can be determined by adapting a criterion of model complexity. In the method proposed by Figueiredo & Jain (2002) a large number of components are fit to the given data set, using the *expectation-maximization* (EM) algorithm. Then, small components are eliminated one by one, each time running EM to convergence. When a single component remains, all converged models are evaluated with a *minimum description length* (MDL) criterion, and one model is

selected. The shape of the covariance must be manually selected at the onset of the algorithm. This is the approach used in (Hamouz et al., 2005) for modelling landmark features.

In (Salah & Alpaydın, 2004), the *incremental mixtures of factor analysers* (IMoFA) algorithm was proposed, which automatically determines the number of components, as well as the covariance complexity for each component separately. In this approach, a single-component factor analyser with a single factor is used to initialize the algorithm. At each iteration, the mixture is grown by adding a component or by adding a factor to an existing component. The algorithm terminates when the likelihood is no longer increasing for a separately monitored validation set.

The IMoFA method was used on Gabor features extracted from sub-windows of face images for landmark localization in (Salah et al., 2007). However, only the best candidate for each landmark was evaluated in the subsequent structural analysis step. The proposed algorithm for this stage (the GOLLUM algorithm) selects one landmark triplet for affine normalization, and models the rest of the landmarks after transformation with Gaussian distributions. The landmark candidates are checked against their expected locations by a thresholded Mahalanobis distance. The landmarks failing the check can be automatically corrected by the back-projection of their expected locations.

DYNAMIC INFORMATION

Biometric information can be recovered from traits with a temporal dimension. A typical example would be the gait of a person, which is unobtrusive and has high user-acceptability, thus capable of serving in a user-convenience scenario. While gait is inherently dynamic, it may be possible to recover temporal information from traits that are usually processed in a static way. For instance the signature or the handwriting of a person can produce dynamic information with appropriate equipment. Similarly, the face is ordinarily a static biometric modality, but sequences of faces in a video stream or the facial expression changes of a person can serve as a dynamic biometrics (Chen et al., 2001).

What makes dynamic biometrics particularly challenging is the variability in the size of the feature vectors. To access the toolbox of methods that work with fixed-dimensionality vectors (e.g. neural networks, PCA, nearest neighbour methods, etc.) *dynamic time warping* may be applied (Kruskal and Liberman, 1983). This method aligns two sequences S_i and S_j by associating cost functions with insertions, deletions, and substitutions, and by locally minimizing the Levenshtein distance, i.e. the minimum number of operations needed to transform S_i into S_j. For feature vectors that have straightforward interpolation functions, resizing the feature vector with linear and quadratic functions to a predetermined size is a computationally cheap alternative. For face recognition, this is the most frequently employed method. For signatures recognition, a re-sampling procedure that produces a fixed-length feature can be employed.

The most frequently used approach in dealing with variable-sized feature vectors is to associate probabilities with feature dynamics. *Dynamic Bayesian networks* (DBNs) are a broad class of graphical models that are capable of incorporating temporal dynamics in this manner. Most frequently used DBNs are *Kalman filters*, and *hidden Markov models* (HMMs). A good survey of learning with DBNs can be found in (Ghahramani, 1998).

It is also possible to extract dynamic information from an otherwise static source by introducing an interest operator, and simulating a dynamic information extraction process. In (Salah et al., 2002), several interest operators are used jointly to derive a saliency map of a facial image, followed by the simulation

of saccadic movements of the fovea in the human visual system to visit the most interesting locations of the image in the order of decreasing saliency. The content of the foveal window is processed via local neural network experts, whereas the location information is mapped to the states of an observable Markov model. After each saccade, Markov models of each person in the gallery are consulted to produce a conditional probability, and if the probability of a particular class exceeds a certain threshold, the identification is effected. With this method, it is possible to inspect only a small part of the facial image before taking a decision, and the whole image is analysed only for difficult cases. Thus, the length of the final feature vector is determined on the fly, and its variability is a blessing, rather than a curse, as it serves reducing the temporal complexity of the identification process. In (Bicego et al., 2003) overlapping windows are extracted from face images, scanned in a regular fashion, and the wavelet coefficients computed from each window are then classified using HMMs.

Storing multiple templates for a single user can make a biometrics system more robust, at the cost of increased computational complexity. For dynamic biometrics, it may be a necessity to store multiple templates. For instance in signature recognition, a person may have an elaborate, slow-dynamics signature, and a fast-dynamics, quick and dirty signature for rapid signing. For this purpose, it will be necessary to quantify the similarity of biometric signals under different dynamic models. Given two different biometric signals $\boldsymbol{B}_i$ and $\boldsymbol{B}_j$, and two DBNs M_i and M_j trained for responding maximally to these signals (or to a set of signals containing them), a similarity (or affinity) score can be computed as:

$$s(\boldsymbol{B}_i, \boldsymbol{B}_j) = \frac{1}{2}\left\{\frac{1}{\tau_j}\log P(\boldsymbol{B}_j|M_i) + \frac{1}{\tau_i}\log P(\boldsymbol{B}_i|M_j)\right\} \tag{20}$$

where τ_i is the authentication threshold for model M_i. Panuccio et al. proposed another measure of similarity (2002), which also takes into account the model quality, i.e. how well M_i models the signal B_i:

$$s(\boldsymbol{B}_i, \boldsymbol{B}_j) = \frac{1}{2}\left\{\frac{P(\boldsymbol{B}_i|M_j) - P(\boldsymbol{B}_i|M_i)}{P(\boldsymbol{B}_i|M_i)} + \frac{P(\boldsymbol{B}_j|M_i) - P(\boldsymbol{B}_j|M_j)}{P(\boldsymbol{B}_j|M_j)}\right\} \tag{21}$$

In both approaches, $P(\boldsymbol{B}_j \mid M_i)$ denotes the probability of observing $\boldsymbol{B}_j$ under the model M_i. Once the similarity matrix is computed, it is possible to employ clustering methods to group the signals into several clusters of differing dynamics, and train one model per cluster for the final system.

For a typical biometrics application, an obvious problem with these approaches is that limited enrollment time requirement leaves the system with very few examples, whereas the dimensionality can be large. One way of dealing with the dimensionality is to constrain the models appropriately, effectively reducing the number of parameters in the process. If the dynamics have a well-defined order, as in a signature, where the order of letters do not change, some transitions may be pruned from the model, leaving less parameters to estimate. An example is using left-to-right HMMs instead of fully connected HMMs, where each state represents a dynamic that is executed once, before continuing to the next dynamic. Depending on the particular case, the time spent in one state may be short or long. Another method is to constrain the observation distribution at every state of the model. Typically Gaussian models or Gaussian mixtures are used for observation probability distribution, where the complexity can be tuned by imposing restrictions on the covariance.

The temporal dynamics can also be taken into account by constant online modification of static models, instead of a single dynamic model that traces the complete evolution of the dynamic. This is particularly useful for dynamics with indeterminate duration, for instance in camera-based surveillance applications. An illustrative example is video-based biometrics, where the motion of bodies or faces is analysed. This necessitates reliable foreground-background segmentation, which is frequently tackled by learning statistical models for foreground objects and/or the background image. For static views, Stauffer and Grimson have proposed an influential method, where one adaptive Gaussian mixture model per background pixel is trained (1999). In their approach, the recent history of each background pixel X_t is modelled by a mixture of K Gaussian distributions, and the probability of a particular pixel belonging to the background at any time is given by:

$$P(X_t) = \sum_{i=1}^{K} \pi_i N(X_t, \mu_i, \sigma_i^2 I) \tag{22}$$

where π_i is the component prior, μ_i is the mean, and σ_i is the standard deviation that specifies the spherical covariance shape of the component. Pixels not falling within 2.5 standard deviations from the mean of any component are assumed to be evidence for new components, and replace the component with the smallest prior. At any time, the first few distributions that account for a fixed portion of the data are considered to belong to the background.

MULTIPLE BIOMETRICS AND INFORMATION FUSION

Information fusion in biometrics serves a twofold purpose. First and foremost, the system requirements may dictate an operational description beyond the technological provisions of a single biometric modality, either in terms of security, or user convenience. Multiple biometrics can be used to design systems to fit more demanding requirements. The second purpose of using multiple modalities relates to user-convenience. It is known that some biometric modalities (like fingerprints) are not usable for a small but non-negligible percentage of the population (Newham, 1995), and consequently, a non-discriminating biometric authentication scheme needs to accommodate these users.

Assume two biometric signals $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are recorded from a single person b in two different modalities with models $M_1 = \{s_1, \boldsymbol{t}_d^{\,1}, \tau_1\}$ and $M_2 = \{s_2, \boldsymbol{t}_d^{\,2}, \tau_2\}$, respectively, where s_i denotes the similarity function, $\boldsymbol{t}_d^{\,i}$ is the biometric template and τ_i is the threshold of authentication. The combined biometric system can be made more secure than either of the individual biometric systems by requiring:

$$s_1(\boldsymbol{B}_1, \boldsymbol{t}_b^1) > \tau_1 \wedge s_2(\boldsymbol{B}_2, \boldsymbol{t}_b^2) > \tau_2 \tag{23}$$

or it can provide for more robust (e.g. for cases where the data acquisition conditions are impaired for one modality), or more convenient (e.g. letting the user decide which modality to use) systems by allowing alternative authentication venues, at the cost of reduced system security:

$$s_1(\boldsymbol{B}_1, \boldsymbol{t}_b^1) > \tau_1 \vee s_2(\boldsymbol{B}_2, \boldsymbol{t}_b^2) > \tau_2 \tag{24}$$

This is the simplest biometric fusion scenario at the decision level of individual systems. It is possible to effect fusion of biometric information at the raw data level, at the feature level, at the matching score level, or at the decision level. Another dimension of fusion is the architecture, which can be serial or parallel. In (Gökberk et al., 2005), the authors contrast a serial (hierarchical) fusion scheme for 3D face recognition in which the first classifier ranks the most plausible classes, after which the second classifier operates on a reduced set of classes, with parallel fusion schemes in which all classifier outputs are jointly assessed. The parallel approach has increased real-time operation cost, but its accuracy is superior to that of the serial, and both fusion approaches excel in comparison to individual classifiers.

When using multiple systems, we usually deal with scores that have different ranges and meanings. The system proposed in (Gökberk et al., 2005) sidesteps this issue by operating on rank scores, but the appropriate normalization of scores is nonetheless an important issue. The normalization can be performed for each score domain separately, bringing the scores to a common scale for combination (*transformation-based score fusion*). Weighted combination of scores is a simplified version of this approach, and most of the fusion approaches reported in the literature aim at learning an appropriate set of weights for individual biometric modalities. For the authentication task, this can also be achieved by treating the genuine and impostor scores as two separate classes, and applying supervised classification methods. For instance in (Ross & Jain, 2003), the scores from different modalities (face, fingerprint, hand geometry) are combined with:

1. The SUM rule, which corresponds to a weighted average of the scores, where learning involves finding the best set of weights on the training set,
2. A decision tree, which was used to learn a number of rules to solve the two-class classification problem of distinguishing impostor and genuine claims,
3. A linear discriminant classifier, which is obtained by projecting the scores to a subspace that minimizes the within-class variance and maximizes the between-class variance, for the same two-class problem.

The SUM rule has been found to perform better than the alternatives, and has the additional benefit of having a threshold to calibrate the system for increased security or convenience. In this kind of a *classifier-based fusion scheme*, the distribution of genuine and impostor scores is usually not balanced. A zero-cost impostor model implies that the samples available for impostor access are *N-1* times greater than the number of genuine access samples, *N* being the number of subjects, and assuming equal number of samples per subject.

An additional difficulty is that most classification approaches would not allow the tuning of the system for a particular FAR or FRR requirement. Suppose the system is required to operate at 0.001 FAR, and a neural network classifier is selected for solving the genuine-impostor classification problem. The training regime needs to be adjusted appropriately, and most likely the final classifier output needs to be post-processed (e.g. by applying a threshold on the posterior) to bring the final system into the desired operation range.

In (Nandakumar et al., 2008) the genuine and impostor match scores from different modalities are modelled with Gaussian mixtures, and the decision follows a likelihood ratio test. This is a *density-based score fusion* scheme, as opposed to transformation-based or classifier-based fusion schemes. The difficulty in density-based fusion is the fitting of the model; it is not uncommon that the limited number of samples available for statistical learning rules out all but the simplest models. However, if the underly-

ing joint score density is known, it is possible to formulate an optimal fusion scheme. The application of the likelihood ratio test is similar to the case in equation (8). We summarize this method here, along with further justification of its use.

Assume $\boldsymbol{B}$= [B_1, B_2, ..., B_k] denotes the biometric match scores from *k* different modalities. Assume further that we know the joint densities of the genuine claims and impostor claims, denoted with $p(\boldsymbol{B}|M_d)$ and $p(\boldsymbol{B}|I)$. Let Ψ denote a statistical test for the null-hypothesis H_0:$\boldsymbol{B}$ is associated with a claim from an impostor versus H_1: $\boldsymbol{B}$ is associated with a genuine claim. $\Psi(\boldsymbol{B})$ is the binary valued function for accepting or rejecting the null hypothesis. The justification proposed by (Nandakumar et al., 2008) for using the likelihood-ratio test rests on the Neyman-Pearson theorem, which states that for testing H_0 against H_1 there exists a test Ψ and a constant η such that:

$$P(\Psi(\boldsymbol{B}) = 1|H_0) = \alpha \tag{25}$$

and

$$\Psi(\boldsymbol{B}) = \begin{cases} 1, & when \quad \dfrac{p(\boldsymbol{B}|M_d)}{p(\boldsymbol{B}|I)} \geq \eta, \\ 0, & when \quad \dfrac{p(\boldsymbol{B}|M_d)}{p(\boldsymbol{B}|I)} < \eta. \end{cases} \tag{26}$$

The more important part of the theorem states that if a test Ψ satisfies these equations for some constant η, then it is the *most powerful test* for testing H_0 against H_1 at level α. Consequently, the authors propose to employ the ratio of $p(\boldsymbol{B}|M_d)$ to $p(\boldsymbol{B}|I)$ for giving the authentication decision, and η is selected in a way that the likelihood ratio test maximizes the GAR at the specified FAR. In practice, the true genuine and impostor score densities are unknown, and they are approximated with the help of the training set, typically with mixture models. One obvious difficulty is that due to user convenience issues, there are very few (typically less than five) genuine scores per subject.

The likelihood-ratio test can further be refined by learning subject-specific parameters for the threshold η or for the class-conditional densities (Poh & Kittler, 2007). The impostor model can be a single model (an example of which is the universal background model frequently used in speaker verification) or it can be a set of alternative models, in which case it is called a *cohort*. The latter approach is less popular, as it implies increased computational complexity.

A discriminative biometric modality will result in a good separation of the genuine and impostor score distributions, and produce higher ratios (i.e. a more confident decision). The fusion should take into account the discriminative power of the contributing biometrics, but it can also consider an assessment of the acquisition conditions and the *quality* of the recorded biometric signal, which can be different for each instance, depending on the acquisition conditions and the particular sensor used. Once a set of quality measures are selected, a *quality-based fusion* approach can be adopted. In (Nandakumar et al., 2008), the quality measures are taken into account by replacing the genuine and impostor score distributions with joint distributions of scores and corresponding quality measures. In (Chatzis et al., 1999) a vector quantization approach was taken to define score categories based on quality, and the biometric

scores are fuzzified to give a robust decision. In (Maurer & Baker, 2007) Bayesian belief networks were employed to incorporate the quality as an effect on the prior.

A recent evaluation campaign under the European FP6 Network of Excellence BIOSECURE assessed biometric fusion methods under *cost* and *quality* considerations (Poh et al., in press). Here, cost refers to the computational cost of verification, which is important for real-time operation. Scores obtained from fingerprint and face scanners were associated with a cost of use, and with several automatically derived quality measures. 22 score-level fusion systems were evaluated in the campaign, including dynamic fusion approaches where biometric modalities were consulted in a sequential fashion until a desired level of authentication confidence was reached or all the scores were exhausted.

Biometric information can be fused with ancillary information that has little discriminative value, but nonetheless can increase the robustness of the system in question. By making the assumption that these so-called *soft biometric traits* (e.g. height, gender) are independent, they can be fused with a strong biometric:

$$s(\boldsymbol{B}, \mathbf{t}_b) = a_0 \log P(\boldsymbol{t}_b^0 | \boldsymbol{B}_0) + \sum_{i=1}^{m} a_i \log P(\boldsymbol{B}_i | \boldsymbol{t}_b^i) \tag{27}$$

where $\boldsymbol{B}_0$ is the primary biometric with a large weight a_0, and $\boldsymbol{B}_i$ are the soft biometrics with relative weights a_i (Jain et al., 2004).

The usefulness of fusion increases if the individual modalities that are fused are not highly correlated. For this reason, it is important to understand the role of the multimodal databases when analysing fusion results. Collecting biometric information can be a meticulous process, and the anonymity of the subjects is imperative. Since collecting and associating multiple biometrics from individuals is perceived to be a threat to anonymity, some of the studies in the literature employ *chimeric databases*, which are created by randomly combining biometric information from different modalities under virtual identities. Since the correlation structure is not retained, the added value of fusion will be overestimated under a chimeric database.

To illustrate this point, consider a system with face and palm-print modalities. In a natural database, male faces will be associated with larger palm-prints and the face and palm skin colours will be highly correlated, whereas a chimeric dataset will lose these correlations. One potential solution is to employ a small set of natural multimodal samples and use unsupervised clustering to model the covariance structure. Then recorded biometrics can independently be assigned to the nearest cluster, and used to construct realistic chimeric datasets by probabilistic selection of samples from within clusters. Some biometric combinations like faces and fingerprint have low correlation, and thus more suitable for chimeric databases.

EVALUATING A BIOMETRICS SYSTEM

The practice of biometrics often requires the evaluation of a system under scrutiny to determine the operating region, as well as the comparison of several algorithms in terms of their computational cost, accuracy, and robustness. Several important considerations that are common practice in the machine learning community should guide the biometrics researcher in the evaluation of the developed systems.

This section provides a short list of good practices. For a detailed exposition, see (Mansfield & Wayman, 2002).

Appropriate experimental setup: The model development and reporting of the test results should be performed on different datasets. When reporting results, accuracies on the training set and the test set should be reported in conjunction, as the amount of difference in accuracy is indicative of the generalization capability of a model. While developing the models, the test set should not be consulted at all, as doing so would imply that the test set is contributing to the learning process, and the accuracies must be appropriately adjusted. A separate validation set can be set aside from the training set during model selection and learning.

Statistical tests for comparison: The claim of superiority of one algorithm over another cannot be based on a higher accuracy alone; it should be supported with statistical tests. Especially with benchmark databases, the accuracies reported in the literature tend to become very high over time, and it is important to demonstrate the success of new algorithms over alternative approaches with tests administered under exactly the same protocol. Producing mean and standard deviation of accuracy results with randomized runs allows comparisons in terms of distance in standard deviations, which serves as a heuristic to decide the significance of obtained accuracy differences. The reader is referred to (Yildiz & Alpaydin, 2006) for statistical testing methodology to determine the algorithm with the smallest expected error given a data set and a number of learning algorithms.

Cross-database evaluation: The problem of overlearning plagues many learning algorithms. Considering that most biometrics databases are constituted by samples acquired under similar conditions (i.e. same environment, same sensor, similar pre-processing and compression, etc.) it is important to assess the generalization ability of proposed algorithms by cross-session recordings, where training and testing are performed on data recorded in different sessions with temporal separation. Better still are the cross-database evaluations, where the training and testing are performed with different databases. Sometimes a *world model* is used to learn the conditions specific to the particular data acquisition environment. This is a small and representative data sample acquired under the test conditions that is used in addition to the training set.

Challenges and evaluation campaigns: The independent testing of the biometrics systems has been a vital practice. The efforts of the U.S. National Institute of Standards and Technology (NIST) have been instrumental and exemplary in this respect, and a number of campaigns have been conducted by NIST and other institutions to evaluate biometric systems independently (see for example, Philips et al. 2007). Among these, we should mention NIST's Face Recognition Vendor Test, speaker recognition challenges, Iris Challenge Evaluation, Face Recognition Grand Challenge and the recent Multiple Biometrics Grand Challenge. Apart from NIST, important biometric assessment campaigns are NVU iris challenges, Univ. of Bologna Fingerprint Verification Competition, M2VTS and BANCA challenges, and the BioSecure Multimodal Evaluation Campaign. These are valuable venues to try baseline algorithms with clearly defined protocols, to see the shortcomings of present approaches, and to objectively evaluate the performance of a new system. The reader is referred to (Petrovska-Delacrétaz et al., 2008) and to references therein for more information on biometric evaluation campaigns.

Links to databases and codes: Table 1 includes some links for the most important databases for different biometric modalities. For OpenSource code of baseline systems the reader is referred to the web resources of Association BioSecure[2], which includes OpenSource reference systems,

Table 1. Links to some important biometrics databases.

Name	Modality	Link
BANCA	face video	www.ee.surrey.ac.uk/CVSSP/banca
CASIA	iris & gait palmprint	www.cbsr.ia.ac.cn
FERET	2D face	www.itl.nist.gov/iad/humanid/feret/feret_master.html
FRGC	2D-3D face	face.nist.gov/frgc
FVC	fingerprint	atvs.ii.uam.es/databases.jsp
MCYT	signature & fingerprint	atvs.ii.uam.es/databases.jsp
NIST	speech	www.ldc.upenn.edu (multiple corpora available from Linguistic Data Consortium)
PolyU	palm-print	www.comp.polyu.edu.hk/biometrics
USF	gait	figment.csee.usf.edu/GaitBaseline (includes code)
XM2VTS	face & speech	www.ee.surrey.ac.uk/CVSSP/xm2vtsdb

publicly available databases, assessment protocols and benchmarking results for several modalities (2D and 3D face, speech, signature, fingerprint, hand, iris, and talking-face).

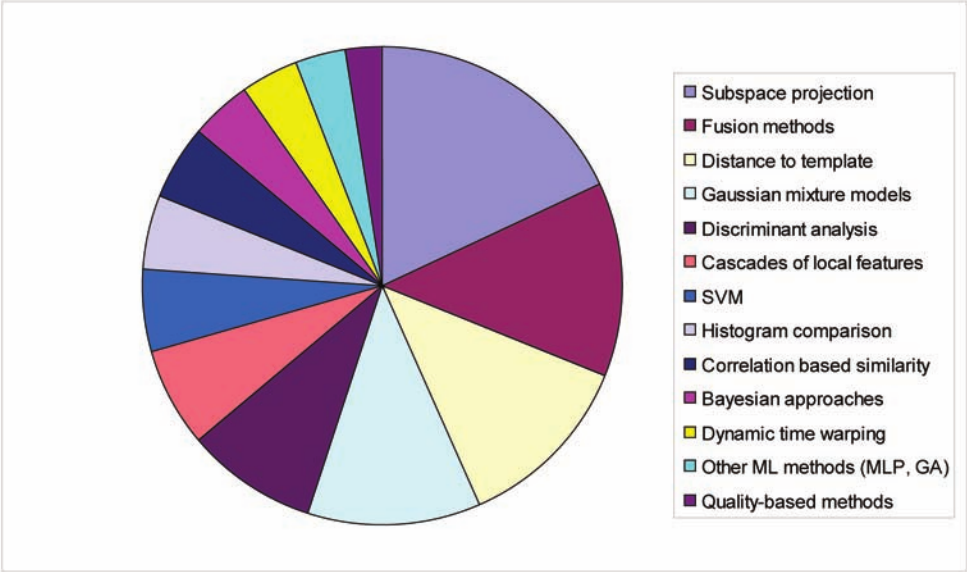
Applications: Figure 1 shows the distribution of various ML approaches in biometrics, as gleaned from the 125 papers presented at the International Conference on Biometrics (Lee & Li, 2007). The most frequently used methods are based on learning a suitable subspace projection to reduce the dimensionality. Non-parametric methods that store templates for each enrolled subject and look at the distance to the query from the template are employed in more than 10 per cent of the papers, with different distance measures. Subspace projections favour Mahalanobis-based distance functions, whereas Hamming distance and histogram comparison techniques are frequently applied in iris recognition. Gaussian mixture models are particularly prominent in speech applications; HMM and other dynamic Bayesian approaches are used in signature and speech applications. About 15 per cent of all the papers favoured a fusion approach.

CONCLUSION

The rapid advance in technology, particularly the production of cheaper and better sensors enables computer systems to automatically identify people. This capability satisfies an increasing need for security and smarter applications, and gains wide acceptance thanks to strict control for privacy and ethical concerns.

In this chapter, we have given a glimpse of machine learning methods that are relevant in making biometric technology a reality. These methods are actively used in deployed systems, but there is ever the need for faster and more accurate algorithms. We hope that the application examples and references we have provided will serve the reader in creating novel machine learning solutions to challenging biometrics problems.

Figure 1. Distribution of different ML approaches to biometric problems.



ACKNOWLEDGMENT

This research is supported by the Dutch BRICKS/BSIK project. The author thanks Dr. Eric Pauwels and an anonymous reviewer for valuable comments.

REFERENCES

Bazen, A. M., & Veldhuis, R. N. J. (2004). Likelihood-ratio-based biometric verification. *IEEE Trans. Circuits and Systems for Video Technology*, *14*(1), 86–94. doi:10.1109/TCSVT.2003.818356

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*, 711–720. doi:10.1109/34.598228

Bicego, M., Castellani, U., & Murino, V. (2003). Using hidden Markov models and wavelets for face recognition. In *Proc. of the Int. Conf. Image Analysis and Processing* (p. 52).

Bolle, R. M., Connell, J. H., Pankanti, S., Ratha, N. K., & Senior, A. W. (2004). *Guide to biometrics*. New York: Springer-Verlag.

Chatzis, V., Bors, A. G., & Pitas, I. (1999). Multimodal decision-level fusion for person authentication. *IEEE Trans. System, Man, and Cybernetics, part A, 29*(6), 674-680.

Chen, L. F., Liao, H. Y. M., & Lin, J. C. (2001). Person identification using facial motion. In *Proc. of the Int. Conf. Image Processing* (pp. 677-680).

Daugman, J. (1988). Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *36*, 1169–1179. doi:10.1109/29.1644

Figueiredo, M. A. T., & Jain, A. K. (2002). Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 381–396. doi:10.1109/34.990138

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. doi:10.1006/jcss.1997.1504

Ghahramani, Z. (1998). Learning dynamic Bayesian networks. In C. L. Giles & M. Gori (Eds.), *Adaptive processing of sequences and data structures* (pp. 168-197). Berlin, Germany: Springer-Verlag.

Gökberk, B., Salah, A. A., & Akarun, L. (2005). Rank-based decision fusion for 3D shape-based face recognition. In *Proc. of the Int. Conf. Audio- and Video-based Biometric Person Authentication* (LNCS 3546, pp. 1019-1028).

Gökberk, B., Salah, A. A., Alyüz, N., & Akarun, L. (in press). 3D face recognition: Technology and applications. In M. Tistarelli, S. Z. Li, & R. Chellappa (Eds.), *Biometrics for surveillance and security*. London: Springer-Verlag.

Gross, R., Matthews, I., Cohn, J., Kanade, T., & Baker, S. (2008). Multi-PIE. In *Proc. of the 8th IEEE Conf. on Automatic Face and Gesture Recognition*, Amsterdam.

Hamouz, M., Kittler, J., Kamarainen, J. K., Paalanen, P., Kalviainen, H., & Matas, J. (2005). Feature-based affine-invariant localization of faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(9), 1490–1495. doi:10.1109/TPAMI.2005.179

Huang, X., Acero, A., & Hon, H. (2001). *Spoken language processing*. Upper Saddle River, NJ: Prentice Hall.

Jain, A. K., Dass, S. C., & Nandakumar, K. (2004). Soft biometric traits for personal recognition systems. In *Proc. of the Int. Conf. Biometric Authentication*, Hong-Kong.

Kruskal, J., & Liberman, M. (1983). *The symmetric time-warping problem: From continuous to discrete*. Reading, MA: Addison-Wesley.

Kung, S. Y., Mak, M. W., & Lin, S. H. (2005). *Biometric authentication: A machine learning approach*. Upper Saddle River, NJ: Prentice Hall Professional Technical Reference.

Lee, S.-W., & Li, S. Z. (2007). *Advances in biometrics* (LNCS 4642). Berlin, Germany: Springer Verlag.

Maurer, D. E., & Baker, J. P. (2007). Fusing multimodal biometrics with quality estimates via a Bayesian belief network. *Pattern Recognition*, *41*(3), 821–832. doi:10.1016/j.patcog.2007.08.008

Nandakumar, K., Chen, Y., Dass, S. C., & Jain, A. K. (2008). Likelihood ratio-based biometric score fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 342–347. doi:10.1109/TPAMI.2007.70796

Newham, E. (1995). *The biometrics report*. SJB Services.

Osuna, E., Freund, R., & Girosi, F. (1997). Training support vector machines: An application to face detection. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 130-136).

Panuccio, A., Bicego, M., & Murino, V. (2002). A hidden Markov model-based approach to sequential data clustering. In *Structural, syntactic and statistical pattern recognition* (LNCS 2396, pp. 734-742). Berlin, Germany: Springer-Verlag.

Petrovska-Delacrétaz, D., Chollet, G., & Dorizzi, B. (Eds.). (2008). *Guide to biometric reference systems and performance evaluation.* London: Springer-Verlag.

Phillips, P.J ., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K. W., Schott, C. L., & Sharpe, M. (2007). *FRVT 2006 and ICE 2006 large-scale results*. NISTIR 7408.

Poh, N., Bourlai, T., Kittler, J., Allano, L., Alonso, F., Ambekar, O., et al. (in press). Benchmarking quality-dependent and cost-sensitive multimodal biometric fusion algorithms. *IEEE Trans. Information Forensics and Security*.

Poh, N., & Kittler, J. (2007). On the use of log-likelihood ratio based model-specific score normalisation in biometric authentication. In S.-W. Lee & S. Z. Li (Eds.), *Advances in biometrics: Proc. of the ICB* (LNCS 4642, pp. 614-624). Berlin, Germany: Springer-Verlag.

Ramanathan, N., Chellappa, R., & Chowdhury, A. K. R. (2004). Facial similarity across age, disguise, illumination, and pose. In *Proc. of the Int. Conf. Image Processing* (pp. 1999-2002).

Reynolds, D., Quatieri, T., & Dunn, R. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, *10*, 19–41. doi:10.1006/dspr.1999.0361

Ross, A., & Jain, A. (2003). Information fusion in biometrics. *Pattern Recognition Letters*, *24*, 2115–2125. doi:10.1016/S0167-8655(03)00079-5

Salah, A. A., & Alpaydın, E. (2004). Incremental mixtures of factor analysers. In *Proc. of the Int. Conf. on Pattern Recognition* (pp. 276-279).

Salah, A. A., Alpaydın, E., & Akarun, L. (2002). A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(3), 420–425. doi:10.1109/34.990146

Salah, A. A., Çınar, H., Akarun, L., & Sankur, B. (2007). Robust facial landmarking for registration. *Annales des Télécommunications*, *62*(1-2), 1608–1633.

Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 246-252).

Sun, Z., Wang, Y., Tan, T., & Cui, J. (2005). Improving iris recognition accuracy via cascaded classifiers. *IEEE Trans. Systems, Man, and Cybernetics. Part C: Applications and Reviews*, *35*(3), 435–441. doi:10.1109/TSMCC.2005.848169

Tistarelli, M., Bicego, M., Alba-Castro, J. L., Gonzàlez-Jimènez, D., Mellakh, A., Salah, A. A., et al. (2008). 2D face recognition. In D. Petrovska-Delacrétaz, G. Chollet, & B. Dorizzi (Eds.), *Guide to biometric reference systems and performance evaluation* (pp. 217-266). London: Springer-Verlag.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86. doi:10.1162/jocn.1991.3.1.71

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Xudong, J., & Ser, W. (2002). Online fingerprint template improvement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(8), 1121–1126. doi:10.1109/TPAMI.2002.1023807

Yildiz, O., & Alpaydin, E. (2006). Ordering and finding the best of $K>2$ supervised learning algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(3), 392–402. doi:10.1109/TPAMI.2006.61

Zhang, D., Wai-Kin, K., You, J., & Wong, M. (2003). Online palmprint identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(9), 1041–1050. doi:10.1109/TPAMI.2003.1227981

KEY TERMS AND DEFINITIONS

biometric: A biometric is a personal or a behavioural trait that can be used to identify a person.

Biometric Template: The biometric template of a person is a pre-recorded biometric sample stored in a database for later authentication.

Authentication: Authentication is the decision process where a biometric is sampled from a person with an identity claim, and the sampled biometric is compared to a biometric template stored previously for this person to validate the identity claim.

Biometric System: A biometric system involves a set of sensors to record a biometric from the users of the system, a database of stored biometric templates, and an authentication algorithm by which the recorded biometric is compared to the template.

Biometric Fusion: The use of multiple biometric samples in a biometrics system. These samples can be collected through different modalities, resulting in multimodal fusion, or multiple samples from a single modality or even a single sensor can be employed for fusion.

ENDNOTES

1 The code is available in the OpenCV library, at http://sourceforge.net/projects/opencvlibrary

2 http://biosecure.it-sudparis.eu/AB/