Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 1885–1906
Pattern Recognition Letters
www.elsevier.com/locate/patrec

2D and 3D face recognition: A survey

Andrea F. Abate, Michele Nappi *, Daniel Riccio, Gabriele Sabatino

Dipartimento di Matematica e Informatica, Università di Salerno, 84084 Fisciano (SA), Italy

Available online 26 January 2007

Abstract

Government agencies are investing a considerable amount of resources into improving security systems as result of recent terrorist events that dangerously exposed flaws and weaknesses in today's safety mechanisms. Badge or password-based authentication procedures are too easy to hack. Biometrics represents a valid alternative but they suffer of drawbacks as well. Iris scanning, for example, is very reliable but too intrusive; fingerprints are socially accepted, but not applicable to non-consentient people. On the other hand, face recognition represents a good compromise between what's socially acceptable and what's reliable, even when operating under controlled conditions. In last decade, many algorithms based on linear/nonlinear methods, neural networks, wavelets, etc. have been proposed. Nevertheless, Face Recognition Vendor Test 2002 shown that most of these approaches encountered problems in outdoor conditions. This lowered their reliability compared to state of the art biometrics. This paper provides an "ex cursus" of recent face recognition research trends in 2D imagery and 3D model based algorithms. To simplify comparisons across different approaches, tables containing different collection of parameters (such as input size, recognition rate, number of addressed problems) are provided. This paper concludes by proposing possible future directions.

Keywords: 2D/3D face recognition; Face databases

1. Face, the most attractive biometric

Many recent events, such as terrorist attacks, exposed serious weakness in most sophisticated security systems. Various government agencies are now more motivated to improve security data systems based on body or behavioural characteristics, often called biometrics (Perronnin and Dugelay, 2003). In general, biometric systems process raw data in order to extract a template which is easier to process and store, but carries most of the information needed. It is a very attractive technology, because it can be integrated into any application requiring security or access control, effectively eliminating risks associated with less advanced technologies that are based on what a person have or know rather than whom a person really is.

Perhaps the most common biometrics are fingerprints and iris, but many other human characteristics have been studied in last years: finger/palm geometry, voice, signature, face. Fig. 1 shows the spreading of the most popular biometrics in the last years from a commercial point of view. However, biometrics have drawbacks. Iris recognition is extremely accurate, but expensive to implement and not very accepted by people. Fingerprints are reliable and non-intrusive, but not suitable for non-collaborative individuals. On the contrary, face recognition seems to be a good compromise between reliability and social acceptance and balances security and privacy well. Its true that any identification system based on face-recognition technology poses several threats to civil rights (Johnson, 2004); first because impinges on the privacy of innocent people when false positives are investigated. Second, face-template data can be stolen and cannot be replaced (although anyone who loses a document can replace it easily). In spite of this, there are large numbers of commercial, security, and forensic applications requiring the use of face recognition technologies. Face recognition provides a lower security level in unconstrained acquisition conditions, but has the great advantage of being

* Corresponding author. Fax: +39 089 963 300.
E-mail addresses: abate@unisa.it (A.F. Abate), mnappi@unisa.it (M. Nappi), driccio@unisa.it (D. Riccio), gsabatino@unisa.it (G. Sabatino).

0167-8655/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2006.12.018

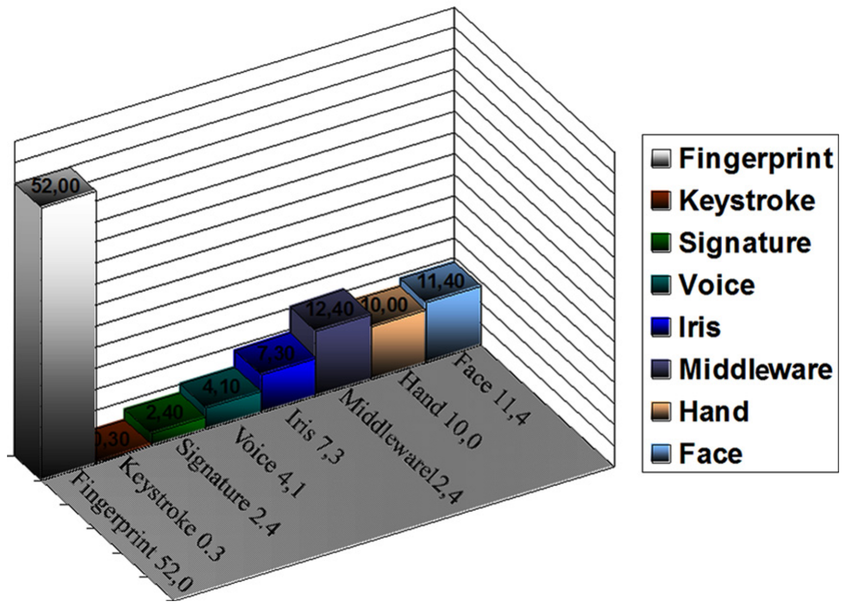

Fig. 1. The spreading of most popular biometrics.

able to work in places with large concourse of unaware visitors.

Face recognition systems fall into two categories: verification and identification. Face verification is a 1:1 match that compares a face image against a template face images, whose identity is being claimed. On the contrary, face identification is a 1:N problem that compares a query face image against all image templates in a face database to determine the identity of the query face. At last a third scenario, the watch list, has been proposed in Face Recognition Vendor Test (FRVT2002) (Phillips et al., 2002). The test individual may or may not be in the system database. The query face image is compared against all the face images in the database, computing a score for each one. All these scores are numerically ranked so that the highest score is first, and if a similarity score is higher than a given threshold, an alarm is raised.

In last decade, major advances occurred in face recognition, with many systems capable of achieving recognition rates greater than 90%. However real-world scenarios remain a challenge, because face acquisition process can undergo to a wide range of variations.

There are five key factors that can significantly affect system face recognition performances:

- *Illumination* variations due to skin reflectance properties and due to the internal camera control. Several 2D methods do well in recognition tasks only under moderate illumination variation, while performances noticeably drop when both illumination and pose changes occur.
- *Pose* changes affect the authentication process, because they introduce projective deformations and self-occlusion. Even if methods dealing with up to 32° head rotation exists, they do not solve the problem considering that security cameras can create viewing angles that are outside of this range when positioned. On the contrary, with exception of extreme expressions such as scream, the algorithms are relatively robust to *facial expression*.
- Another important factor is the *time delay*, because the face changes over time, in a nonlinear way over long periods. In general this problem is harder to solve with respect to the others and not much has been done especially for age variations.
- At last, *occlusions* can dramatically affect face recognition performances, in particular if they located on the upper-side of the face, as documented in literature.

In order to assess how well proposed methods work when dealing with one or a combination of these variation, several face images databases have been built. The number and the typology of addressed problems (together with other parameters such as the number of tested databases, the size of the gallery and probe sets) are quite indicative of how robust face recognition methods are. This also motivated many researchers to generate several face databases which provides as many variations as possible on their images. FERET (Phillips et al., 2000), CMU-PIE (Sim et al., 2003), AR Faces (Martinez, 2002) represents one of the most popular 2D face image database collection. Each the database is designed to address specific challenges covering a wide range of scenarios. For example, FERET represents a good testing framework if one needs large gallery and probe sets, while CMU is more indicated when pose and illumination changes are the main problem. Finally, AR Faces is the only database providing natural occluded face images.

On the contrary, there are very few 3D face models databases and they contain very little amount of data. The most popular 2D face image and 3D face model databases are reported in Tables 1 and 2, together with their main characteristics: name, color/grayscale images, number of people and images per person, available distortions (illumination (i), pose (p), expression (e), occlusions (o), time delay (t), indoor/outdoor (i/o)), availability (for free/buyable) and web home page.

In contrast to the significant effort to build very large face databases, there is not a unique standard protocol to evaluate performances. Face recognition algorithm performance is typically characterized by correct identification rate, FAR (False Acceptance Rate) or FRR (False Rejection Rate) under closed-world assumptions. However Sherrah (2004) recently underlined the importance of mainly minimize the false alarm rate, because it is a more difficult criterion to minimize when designing a classifier. The FERET strategy (Phillips et al., 2000) and the FRVT (Facial Recognition Vendor Test) (Phillips et al., 2002) also give a great contribution for the standardization of the testing protocol.

Indeed, there is not a common benchmark database used to test existing algorithms, and FERET is an excellent attempt towards this direction. On the contrary, the main goal of the FRVT is the capabilities assessment for commercially available facial recognition systems with respect to changes in expression, illumination, pose and time delay. In the last edition of the FRVT (2002), FaceIt outperformed other vendors in the majority of the experiments, but performances are not yet comparable to other biometrics more largely used nowadays (fingerprints).

3D capturing process is becoming cheaper and faster, and for this reason recent works attempt to solve the problem directly on a 3D face model. The constant progress of the 3D capturing technologies influenced also the type of the recognition algorithms. In fact, the first algorithm was applied directly on clouds of points (after a suitable triangulation), while more recent ones directly work on a mesh, considering in some cases the information provided by both the 3D shape and texture. The 3D_RMA is an

Table 1
Most important face databases

Name	RGB/gray	Image size	Number of people	Pictures/person	Number of conditions	Available	Web address
AR Face Database*	RGB	576 × 768	126 70 Male 56 Female	26	i, e, o, t	Yes	http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html
Richard's MIT database	RGB	480 × 640	154 82 Male 74 Female	6	p, o	Yes	
CVL Database	RGB	640 × 480	114 108 Male 6 Female	7	p, e	Yes	http://www.lrv.fri.uni-lj.si/facedb.html
The Yale Face Database B*	Gray Scale	640 × 480	10	576	p, i	Yes	http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html
The Yale Face Database*	Gray Scale	320 × 243	15 14 Male 1 Female	11	i, e	Yes	http://cvc.yale.edu/projects/yalefaces/yalefaces.html
PIE Database*	RGB	640 × 486	68	~608	p, i, e	Yes	http://www.ri.cmu.edu/projects/project_418.html
The UMIST Face Database	Gray	220 × 220	20	19–36	p	Yes	http://images.ee.umist.ac.uk/danny/database.html
Olivetti Att – ORL*	Gray	92 × 112	40	10		Yes	http://www.uk.research.att.com/facedatabase.html
(JAFFE) Database	Gray	256 × 256	10	7	e	Yes	http://www.mis.atr.co.jp/~mlyons/jaffe.html
The Human Scan Database	Gray	384 × 286	23	~66		Yes	http://www.humanscan.de/support/downloads/facedb.php
The University of Oulu Physics-Based Face Database	Gray	428 × 569	125	16	i	Cost $50	http://www.ee.oulu.fi/research/imag/color/pbfd.html
XM2VTSDB	RGB	576 × 720	295		p	Frontal $153 Side $229.5	http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/
FERET*	Gray RGB	256 × 384	30.000		p, i, e, i/o, t	Yes	http://www.itl.nist.gov/iad/humanid/feret/

The '*' points out most used databases. Image variations are indicated by (i) illumination, (p) pose, (e) expression, (o) occlusion, (i/o) indoor/outdoor conditions and (t) time delay.

Table 2
Most important 3D face model databases

Name	Type	Data size	Number of people	3D Models/person	Number of conditions	Texture image	Available	Web address
3D RMA	Cloud of points	4000 points	120 106 Male 14 Female	3	p	No	Yes	http://www.sic.rma.ac.be/~beumier/DB/3d_rma.html
SAMPL	Range image	200×200	10	33 (for 2 sub) 1 (for 8 sub)	p, e	Yes	Yes	http://sampl.eng.ohio-state.edu/~sampl/
Univ. of York 1	Range image	–	97	10	p, e, o	No	Yes	http://www-users.cs.york.ac.uk/~tomh/3DfaceDatabase.html
Univ. of York 2	Range image	–	350	15	p, e	No	Yes	http://www-users.cs.york.ac.uk/~tomh/3DfaceDatabase.html
GavabDB	Tri-Mesh		61 45 Male 16 Female	9	p.e.	No	Yes	http://gavab.escet.urjc.es/recursos_en.html

Image variations are indicated by (p) pose, (e) expression, (o) occlusion.

example of a 3D face models database represented by clouds of points.

For long time it has been the only publicly available database, even if its quality is rather low. On the contrary 3D meshes are available today from newer technologies, but in most cases they are just proprietary databases. The rest of this paper is organized as follows: Section 2 describes recent 2D face recognition research trends, while emphasizing on results achieved; Section 3 analyzes what is currently preventing a wider adoption of face biometrics in commercial applications. In its closing paragraph, it also provides a more general way to evaluate performances for existing face recognition algorithms.

An opening discussion on 3D based face recognition models is presented in Section 4, while following subsections discuss both currently available acquisition procedures and most significant approaches present in literature until now. Finally, Section 5 closes the paper with some considerations on state of the art and possible future trends of the face biometric, suggesting several kinds of multimodality as a good compromise between reliability and social acceptance.

2. Automatic face recognition: the old and the new

2.1. Linear/nonlinear projection methods

Automatic Face Recognition can be seen as a pattern recognition problem, which is very hard to solve due to its nonlinearity. Particularly, we can think of it as a template matching problem, where recognition has to be performed in a high-dimensional space. Since higher the dimension of the space is, more the computation we need to find a match, a dimensional reduction technique is used to project the problem in a lower-dimensionality space. Indeed, the Eigenfaces (Kirby and Sirovich, 1990) can be considered as one of the first approaches in this sense. An $N \times N$ image I is linearized in a N^2 vector, so that it represents a point in a N^2-dimensional space. However, comparisons are not performed in this space, but a low-dimensional space is found by means of a dimensionality reduction technique (Fig. 2). Kirby and Sirovich (1990) adopted the PCA (Principal Component Analysis). Thus, after the linearization the mean vector is calculated, among all images, and subtracted from all the vectors, correspond-

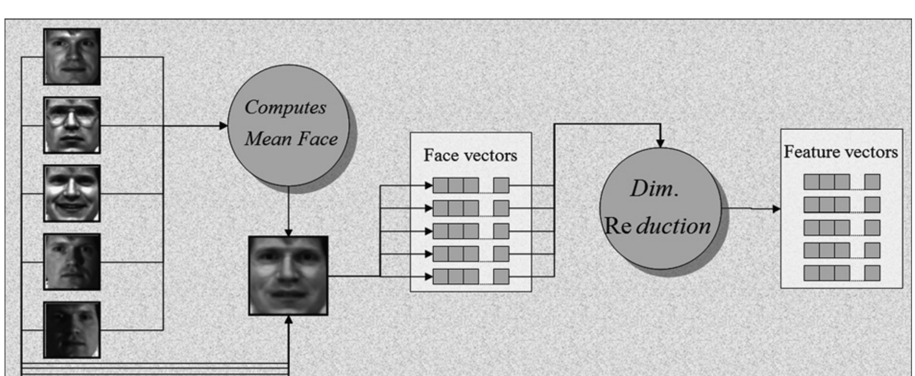

Fig. 2. The general scheme of the linear/nonlinear methods.

ing to the original faces. The covariance matrix is then computed, in order to extract a limited number of its eigenvectors, corresponding to the greatest eigenvalues. These few eigenvectors, also referred to as Eigenfaces, represent a base in a low-dimensionality space. When a new image has to be tested, the corresponding Eigenface expansion is computed and compared against the entire database, according to such a distance measure (usually the Euclidean distance). As the PCA is performed only for training the system, this method results to be very fast, when testing new face images. The PCA has been intensively exploited in face recognition applications, but many other linear projection methods have been studied too.

The LDA (Linear Discriminant Analysis) (Lu et al., 2003; Martinez and Kak, 2001) has been proposed as a better alternative to the PCA. It expressly provides discrimination among the classes, while the PCA deals with the input data in their entirety, without paying any attention for the underlying structure. Indeed the main aim of the LDA consists in finding a base of vectors providing the best discrimination among the classes, trying to maximize the between-class differences, minimizing the within-class ones.

The between- and within-class difference are represented by the corresponding scatter matrices S_b and S_w, while the ratio $\det|S_b|/\det|S_w|$ has to be maximized. Even if the LDA is often considered to outperform the PCA, an important qualification has to be done. Indeed the LDA provides better classification performances only when a wide training set is available, and some results discussed by Martinez and Kak (2001), confirm this thesis. Besides recent studies also strengthen this argument expressly tackling this problem referred to as the SSS (Small Sample Size) problem. In some approaches, such as the Fisherfaces (Belhumeur et al., 1997), the PCA is considered as a preliminary step in order to reduce the dimensionality of the input space, and then the LDA is applied to the resulting space, in order to perform the real classification. However it has been demonstrated in recent works (Chen et al., 2000; Yu and Yang, 2001) that, combining in this way PCA and LDA, discriminant information together with redundant one is discarded. Thus, in some cases the LDA is applied directly on the input space, as in (Chen et al., 2000; Yu and Yang, 2001). Lu et al. (2003) proposed an hybrid between the D-LDA (Direct LDA) and the F-LDA (Fractional LDA), a variant of the LDA, in which weighed functions are used to avoid that output classes, which are too close, can induce a misclassification of the input.

The DCV (Discriminant Common Vectors) (Cevikalp et al., 2005) represents a further development of this approach. The main idea of the DCV consists in collecting the similarities among the elements in the same class dropping their dissimilarities. In this way each class can be represented by a common vector computed from the within scatter matrix. When an unknown face has to be tested, the corresponding feature vector is computed and associated to the class with the nearest common vector. The main disadvantage of the PCA, LDA, Fisherfaces is their linearity. Particularly the PCA extracts a low-dimensional representation of the input data only exploiting the covariance matrix, so that no more than first- and second order statistics are used. In (Bartlett Marian et al., 2002) show that first- and second order statistics hold information only about the amplitude spectrum of an image, discarding the phase-spectrum, while some experiments bring out that the human capability in recognizing objects is mainly driven by the phase-spectrum. This is the main reason for which in (Bartlett Marian et al., 2002) the ICA are introduced as a more powerful classification tool for the face recognition problem. The ICA can be considered as a generalization of the PCA, but providing three main advantages: (1) It allows a better characterization of data in an n-dimensional space; (2) the vectors found by the ICA are not necessarily orthogonals, so that they also reduce the reconstruction error; (3) they capture discriminant features not only exploiting the covariance matrix, but also considering the high-order statistics.

2.2. The neural networks

A further nonlinear solution to the face recognition problem is given by the neural networks, largely used in many other pattern recognition problems, and readapted to cope the people authentication task. The advantage of neural classifiers over linear ones is that they can reduce misclassifications among the neighbourhood classes. The basic idea is to consider a net with a neuron for every pixel in the image. Nevertheless, because of the pattern dimensions (an image has a dimension of about 112×92 pixels) neural networks are not directly trained with the input images, but they are preceded by the application of such a dimensionality reduction technique.

A first solution to this problem has been given by Cottrell and Fleming (1990), which introduced a second neural net, that operates in auto-association mode (Fig. 3). At first, the face image, represented by a vector x, is approximated by a new vector h with smaller dimensions by the first network (auto-association), and then h is finally used as input for the classification net. Cottrell and Fleming also shown that this kind of neural network does not behave better than the Eigenfaces even if in optimal circumstances. Other kind of neural networks also have been tested in face recognition, in order to exploit their particular properties. For examples Self Organizing Map (SOM) are invariant with respect to minor changes in the image sample, while convolutional networks provide a partial invariance with respect to rotations, translations and scaling. In general, the structure of the network is strongly dependent on its application field, so that different contexts result in quite different networks. In a recent work, Lin et al. (1997) presented the Probabilistic Decision Based Neural Network, which they modelled for three different applications (a face detector, an eyes localizer and a face recognizer). The flexibility of these networks is due to their hierarchical structure with nonlinear basis functions and a competitive

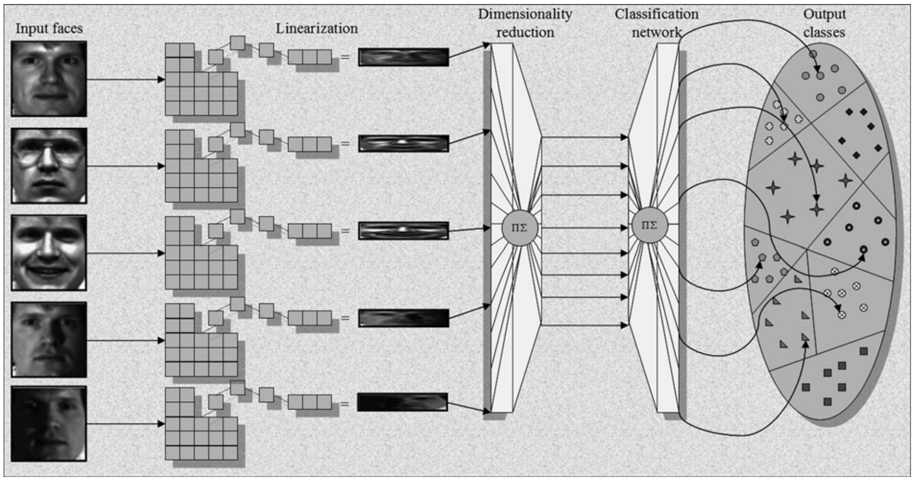

Fig. 3. The structure of a neural network based approach.

credit assignment scheme, which shown the capability of recognizing up to 200 people.

At last, Meng et al. (2002) introduced a hybrid approach, in which, through the PCA, the most discriminating features are extracted and used as the input of a RBF neural network. The RBFs perform well for face recognition problems, as they have a compact topology and learning speed is fast. In their work the authors also face the problem of the overfitting: the dimension of the network input is comparable to the size of the training set; of the overtraining: high dimension of the input results in slow convergence, small sample size: the sample size has to exponentially grow for having a real estimate of the multivariate densities when the dimension increases; the singular problem: if the number of training patterns is less than the number of features, the covariance matrix is singular. In general, neural networks based approaches encounter problems when the number of classes increases. Moreover, they are not suitable for a single model image recognition task, because multiple model images per person are necessary in order for training the system to "optimal" parameter setting.

2.3. Gabor filters and wavelets

The Gabor filters represent a powerful tool both in image processing and image coding, thanks to their capability to capture important visual features, such as spatial localization, spatial frequency and orientation selectivity. In the most cases the Gabor filters are then used to extract the main features from the face images. Indeed, in (Lades et al., 1993) they have been applied to specific areas of the face region, corresponding to nodes of a rigid grid. In each node of the grid the Gabor coefficients are extracted and combined in jets. The nodes are linked to form such a Dynamic Link Architecture, so that the comparisons among different subjects can be made by means of a graph matching strategy. Wiskott et al. (1997) further expanded on DLA and developed a Gabor wavelet based elastic bunch graph matching method (EBGM) to label and recognize human faces. Furthermore, comparisons are made in two consecutive steps: a rigid alignment of the grid only accounts for global transformations, such as translations and scale, then the local misplacement of the grid nodes is evaluated by means of a Graph Similarity Function. Generally, dynamic link architecture is superior to other face recognition techniques, in terms of rotation invariant; however, the matching process is computationally expensive. Perronnin and Dugelay (2003) proposed a further deformable model, whose philosophy is similar to the EBGM. They introduced a novel probabilistic deformable model of face mapping, based on a bi-dimensional extension of the 1D-HMM (Hidden Markov Model). Given a template face F_T, a query face F_Q and a deformable model M, the proposed method try to maximize the likelihood $P(F_T|F_Q, M)$. There are two main differences between this method and the original EGM. First of all the HMM is extended to the 2D case to estimate $P(F_T|F_Q, M)$, automatically training all the parameters of M, so taking into account for the elastic properties of the different parts of the face. Secondly, the model M is shared among all faces, so the approach works well also when little enrolment data is available. On the contrary, a quite different approach has been proposed by Liu (2004) (Fig. 4). A mother wavelet is defined and forty Gabor filters are derived, considering five scales and eight orientations. Each of these filters is convolute with the input image, resulting in forty filtered copies of the face image. To encompass all the features produced by the different Gabor kernels, the resulting Gabor wavelet features are concatenated to derive an augmented Gabor feature vector. Then, in order to reduce the dimensionality of the feature vector, both the PCA and the Enhanced Fisher Linear Discriminant Model (EFM) are investigated. The use of Gabor filters renders this method very robust to changes in expression and illumination; however they dramatically increase the computational cost of the method,

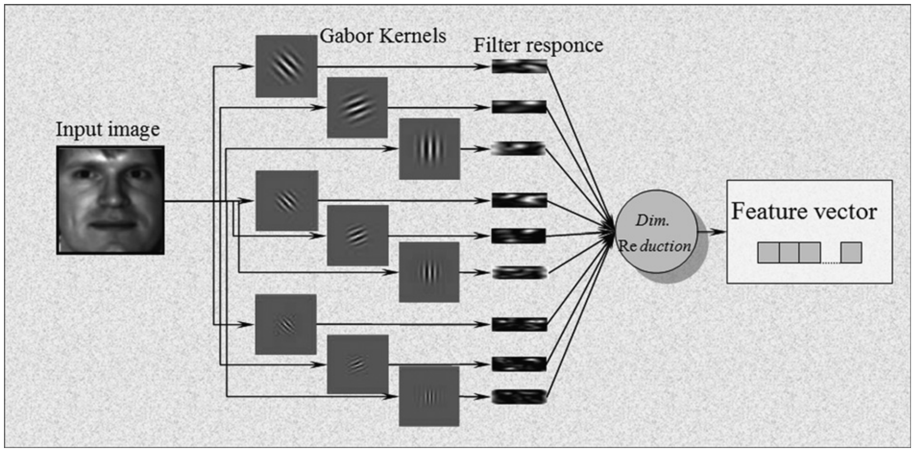

Fig. 4. The convolution of Gabor kernels and dimensionality reduction of the filter responses.

requiring that each kernel is convolved with the input image. A faster wavelet based approach has been proposed by Garcia et al. (2000), which presented a novel method for recognition of frontal views of faces under roughly constant illumination. It is based on the analysis of a wavelet packet decomposition of the face images, because very fast implementations of this procedure are available in hardware. Each face image is first located and then described by a subset of band filtered images containing wavelet coefficients. From these wavelet coefficients, which characterize the face texture, they build compact and meaningful feature vectors, using simple statistical measures. Then, they show how an efficient and reliable probabilistic metric derived from the Bhattacharrya distance can be used in order to classify the face feature vectors into person classes, so that even very simple statistical features can provide a good basis for face classification.

2.4. Fractals and Iterated Function Systems (IFSs)

The IFS theory (Iterated Function Systems) (Riccio and Nappi, 2003) has mainly been developed in the framework of the still image coding and subsequently, it has been extended to the image indexing in view of its capability to describe the image content in a very compact way (Fig. 5). Furthermore the fractal code of an image is invariant with respect to a wide set of global transformations, such as rotations (multiples of $\pi/2$), contrast scaling and channel shifting, just to cite some of them. In (Kouzani et al., 1997) the fractal code of a face image is used for training a neural network, which works as classifier on the face database. The authors claim a recognition rate of 100%, with a false rejection rate of 0. However, experiments have been conducted on a non-standard database, with very few images (150 frontal view face images).

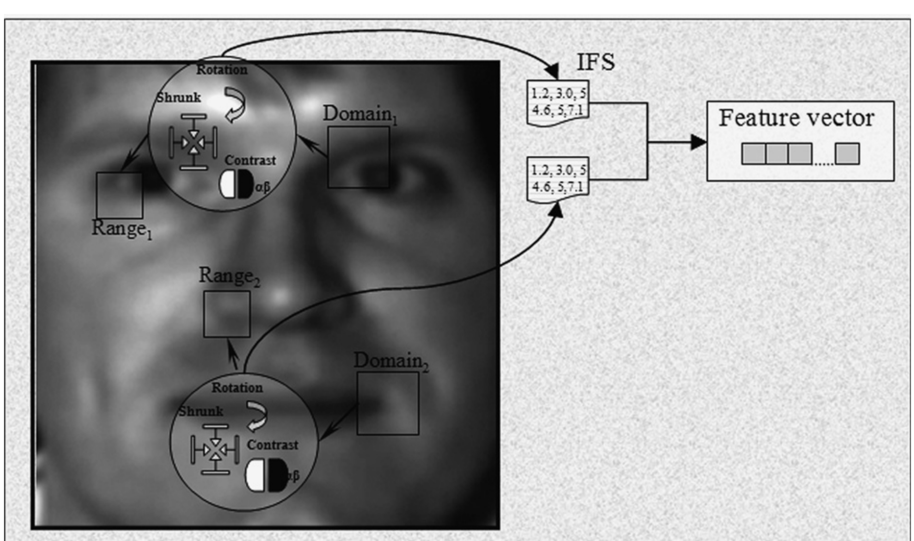

Fig. 5. The feature extraction process of fractal based techniques.

The real advantage of such a fractal based recognition system is that it is less computationally expensive than linear, nonlinear and NN-based techniques and no retraining of the system is necessary when new images are added. Subsequently Tan and Yan proposed a different IFS based strategy (Tan and Tan, 1999), in which the IFS of the face image is associated as a feature vector and stored in a database. When a new face has to be identified, only one iteration of every IFS code in the database is applied to the test image and the one producing the greater PSNR (Picked Signal to Noise Ratio) is indicated as the correct identity. The linear cost of the testing operation represents the real limit of this technique. Indeed when the face database becomes considerable, the time spent comparing all the faces could be not negligible. The better way to cope this drawback is to store the information extracted during the coding process, in such a feature vector. In fact, Komleh et al. investigated on the discriminating power of the IFS parameters (Ebrahimpour-Komleh et al., 2001). At first they considered contrast, brightness, range/domain match and rotation separately and then combined all together, obtaining a single feature vector for each face image. Their results point out that to combine the IFS parameters gives better recognition rate, than testing each one singly. A further crucial limit is that fractal coding is asymmetrical, as the coding process is much more expensive than decoding, so that IFS based face recognition algorithm are often slow.

2.5. Thermal and hyperspectral

Performances of the classical image based techniques are satisfactory when the face image are acquired under controlled conditions, but most of these methods cannot deal with distortions, such as change in illumination, mainly when the images are taken under unconstrained conditions. Someone tried to overcome this problem investigating algorithms, which use some kind of input different from intensity images; the infrared imagery could be an example. The infrared imagery works on the subsurface features of the face and some recent works demonstrated that it can be really considered as a biometric feature. The most interesting aspects is the grasp of the main differences between intensity image based and infrared image based face recognition technique, in order to highlight the advantages of this new biometric.

An interesting work in this sense has been proposed by Chen et al. (2003). Indeed, they presented a wide set of experiments, in which they compare performances of the infrared and intensity image based techniques, showing that the former outperform the latter, when the testing image are taken in unconstrained condition, mainly with respect to the illumination changes. On the contrary, Socolinsky and Selinger (2004a,b) investigated the matter of recognition accuracy when a noticeable lapse of time occurs between the acquisitions of the training and testing data. The data they used in their experiments have been taken in different sessions and in outdoor conditions, but results confirm that to combine the information coming from visible and infrared images improve recognition performances. A further contribution is given by Buddharaju et al. (2004). In their approach a fuzzy based segmentation process is applied in order to extract the region of interest. The Gabor filters are used to extract the main features, which are then used as the input of the Bessel forms. At last a Bayesian classifier is used in order to perform the recognition task. Buddharaj shows in the experiments that this technique can achieve results comparable with respect to the ones obtained by the Eigenfaces. However, an important drawback, working with thermal images is represented by their dependence on the temperature of the skin during the acquisition process. Nevertheless, it has been observed that unlike thermal imagery, the hyperspectral signature of the face is less dependent on the temperature than the thermal radiance. Furthermore, the spectroscopy has been also widely investigated in the biomedicine and remote sensing applications, assessing that different people show a high variability of the hyperspectral properties of the facial tissue, but these features are constant for the same person across the time and under different illumination condition. These observations suggest investigating them as a possible biometric. Indeed, in (Pan et al., 2003) such a technique based on the hyperspectral images has been applied on a database of 200 subjects, which has been acquired under different pose, expression and illumination conditions. Each face is represented using spectral reflectance vectors that are extracted from small facial regions, while the Mahalanobis distance is adopted as similarity measure.

3. Open questions in face recognition

The Automatic Face Recognition (AFR) can be thought as a very complex object recognition problem, where the object to be recognized is the face. This problem is even more difficult to solve, since the search is done among objects belonging to the same class. Besides, in most cases, no more than one visible image is available to train the system and different problems rise when images are acquired under uncontrolled conditions. The sensibility of the classifiers to illumination and pose variations are the main problems researchers have been facing until now, while a smaller effort has been made to cope with occlusions and age variation problems. Therefore, recent works can be classified depending on their main contribution in order to address some of these problems.

3.1. The changes in illumination

Ambient lighting changes greatly within and between days and among indoor and outdoor environments. Due to the 3D structure of the face, a direct lighting source can cast strong shadows that accentuate or diminish certain facial features. It has been shown experimentally and theoretically for systems based on Principal Component Analysis that differences in appearance induced by illumination are larger than differences between individuals. Since dealing with illumination variation is a central topic in computer vision numerous approaches for illumination invariant face recognition have been proposed. In (Adini et al., 1997), investigate the way, in which changes in illumination can affect performances of some face recognition methods. They define three different classes in order to grade the methods: the shape from shading approaches, which extract the shape information of the face, from one or more of its views, the representation based methods, which try to get a characterization of the face invariant to illumination changes and the generative methods, which produce a wide set of synthetic images containing as variations as possible. The authors deduced that none of the experimented technique (edge map, 2D Gabor Filters, first and second derivatives of the gray level images) is able to solve the problem by itself and the results they report seems to confirm this hypothesis.

Notwithstanding this, several efforts have been made in order to achieve better performances in uncontrolled conditions. Indeed, Gao and Leung (2002) extended the edge map technique defining a new approach, namely the Line Edge Map, in which the face contours are extracted and combined in segments, which are then organized in lines. The Hausdorff distance has also been modified in order to manage these new feature vectors. Besides, they also describe a new prefiltering criterion for screening the whole set of individuals before to perform the real testing operation. The method has been tested on several conditions for pose and illumination and the results show that this approach outperforms other methods, such as Linear Subspaces or Eigenfaces, presented in (Belhumeur et al., 1997). However, the Fisherfaces remain superior thanks to their capability to maximize the between-person variability, minimizing the within-person differences. This suggests that combining several linear methods, performances can be further improved. Indeed, an in-depth study on the performances of the linear methods when changes in illumination occur has been conducted by Li et al. (2004). The examined techniques have been compared with respect to both recognition rate and time/memory complexity. The authors observe that the LDA combined with a generalization of the SVD (Singular Value Decomposition), outperforms all the other methods. Nevertheless this hybrid is less adaptable to general face recognition problems, owning to its computational cost. Therefore, the authors suggest that to combine the LDA with the QR decomposition could represent the optimal choice in most cases, since it provides almost comparable performances with the LDA/SVD approach with a lower cost. On the contrary, the PCA and the PCA + LDA (Fisherfaces) perform worse of all the other methods. To overcome limits introduced by the linearity of the abovementioned strategies, nonlinear methods, such as the ICA, have been studied. One of the most recent work has been proposed by Kim et al. (2003). The face is split in different regions that overlap on the boundaries. For each class containing all the elements belonging to the same face region the residual space (the space spanned by the PCA after removing a few leading Eigenfaces) is computed and the ICA is applied to. The results underline that the PCA components in the residual space are the same that in the normal space, while ICA components are different, so that performances improve. Moreover, to split the face in several regions simplifies the statistical model of the illumination variations, making the recognition task more robust with respect to changes.

On the contrary, not much has been made yet on generative methods. One of the few generative methods has been proposed by Georghiades et al. (2001). The face shape and the albedo are extracted by few subject images, by means of a shape from shading algorithm. The 3D model is then used to synthesize a wide set of face views in different pose and illuminations. This method is based on the main hypothesis that for a fixed pose of the object/face, all its views under different illuminations form a convex cone in the image space. For every subject and pose the convex cone is then computed and approximated by means of a low-dimensional linear subspace. In the testing phase the pose of the subject is esteemed and using the Euclidean distance the identity of the subject with the nearest convex cone is assigned. This method is superior to many others in terms of the recognition rate (e.g., the Eigenfaces). Nevertheless the computational cost of the training phase is non-trivial, since all the synthetic views extracted from the 3D model must be processed. There are 3D-based methods, such as the 3D-Morphable Models discussed in Section 4.2.1, which overcome this problem.

3.2. The changes in pose

In many face recognition scenarios the pose of the probe and gallery images is different. For example, the gallery image might be a frontal "mug-shot" and the probe image might be a 3/4 view captured from a camera in the corner of a room. Approaches addressing pose variation can be classified into two main categories depending on the type of gallery images they use.

Multi-view face recognition is a direct extension of frontal face recognition in which the algorithms require gallery images of every subject at every pose. In face recognition across pose we are concerned with the problem of building algorithms to recognize a face from a novel viewpoint, i.e., a viewpoint from which it has not previously been seen. Linear subspaces have been extended in order to deal also with the problem of pose changes. Indeed Okada and von der Malsburg (2002) present a framework for recognizing faces with large 3D pose variations, by means of parametric linear subspace model for representing each known person in the gallery. The authors investigate two different linear models: (1) the LPCMAP model, that is a parametric linear subspace model, combining the linear subspaces spanned by principal components (PCs) of training samples and the linear transfer matrices, which associate projection coefficients of training samples onto the subspaces and their corresponding 3D head angles; (2) the PPLS model, that extends the LPCMAP by using the piecewise linear approach, that is a set of local linear models, each one providing continuous analysis and synthesis mappings, enabling to generalize to unknown poses by interpolation. The experimental results show that the recognition system is robust against large 3D head pose variations covering 50° rotation along each axis. While significantly compressing the data size, the PPLS system performed better than the LPCMAP system. However, the number of known people is relatively small and the samples included some artificialities which might accidentally increase the performance. Another drawback is that the recognition systems uses pixel-wise landmark locations for representing facial shape and deriving head pose information, but finding landmark locations in static facial images with arbitrary head pose is an ill-posed problem. Then, Gross et al. (2002) proposed to use the light-field to achieve a greater robustness and stability solving the problem of pose variation in face recognition. The light-field is a 5D function of position (3D) and orientation (2D), which specifies the radiance of light in free space. In particular, the authors apply the PCA to a collection of light-fields of faces of different subjects, obtaining a set of eigen light-fields, while the mean light-field could also be estimated and subtracted from all of the light-fields. Since, any image of the object corresponds to a curve in the light-field. One way to look at this curve is as a highly occluded light-field, from which the eigen coefficients can be calculated yet, especially for objects with simple reflectance properties such as Lambertian. Then input face images are vectorized in light-field vectors, next used for training and testing the system. They test the eigen light-field method on the CMU (PIE) database and the FERET database, showing that it outperforms both the standard Eigenfaces algorithm and the commercial FaceIt system. Overall, it is observed that the performance improvement of eigen light-fields over the other two algorithms is more significant on the PIE database than on the FERET database, because the former contains more variation in pose than the latter.

3.3. The occlusion

One of the main drawbacks of the appearance-based paradigm (e.g., PCA), is its failure to robustly recognize partially occluded objects. One way to deal with partially occluded objects (such as faces) is by using local approaches. In general, these techniques divide the face into different parts and then use a voting space to find the best match. However, a voting technique can easily misclassify a test image because it does not take into account how good a local match is. In (Martinez, 2002) in order to cope with this problem, each face image is divided into k different local parts. Each of these k local parts is modelled by using a Gaussian distribution (or, equivalently, with a mixture of Gaussians), which accounts for the localization error problem. Given that the mean feature vector and the covariance matrix for every local subspace are drawn out and the probability of a given match can be directly associated with the sum of all k Mahalanobis distances. This approach differs from previous local PCA methods in that it uses a probabilistic approach rather than a voting space. In his work the author investigates on the amount of the occlusion that can be handled by the proposed approach, and the minimum number of local areas needed to successfully identify a partially occluded face. Martinez demonstrated experimentally that the suppression of 1/6 of the face does not decrease accuracy, while even for those cases where 1/3 of the face is occluded, the identification results are very close to those obtained without occlusions. He also has shown that worse results are obtained when the eye area is occluded rather than the mouth area. The probabilistic approach proposed by Martinez is only able to identify a partially occluded face, while Kurita et al. (2003) proposed a method that also reconstructs the occluded part of the face and detects the occluded regions in the input image, by means of an auto-associative neural network. At first the network is trained on the non-occluded images in normal conditions, while during the testing the original face can be reconstructed by replacing occluded regions with the recalled pixels. The training data set consisted of ninety three 18×25 8-bits images, while the trained network has been tested using three types of test data: pixel-wise, rectangular, and sunglass. In the results the authors claim that the classification performance is not decreased even if 20–30% of the face images is occluded. On the other hand, this method suffers from two of the main problems of the NN-based approaches: the system retraining in case of new enrolments and the little availability of training samples. Moreover, a method, which is able to deal with both occlusions and illumination changes, has been proposed by Sahbi and Boujemaa (2002). They presented a complete scheme for face recognition based on salient feature extraction in challenging conditions. These features are used in a matching process that overcomes occlusion effects and facial expressions using the dynamic space warping which aligns each feature in the query image, if possible, with its corresponding feature in the gallery set. Once features have been extracted, they construct a binary image which is subdivided into regions describing shape variation between different faces. They model the statistical deviation of each feature in the face model with respect to its corresponding matched features in each candidate face of the gallery set, and they introduce a matching class for each extracted and matched feature from the face model. This matching class expresses the possible deviation of this feature (modelled using a Gaussian distribution) with respect to the gallery images. Tests have been performed using the Olivetti and ARF public databases, noting that for little occlusion and rotation, the matching process succeeds, so the precision of recognition is guaranteed to be unchangeable with respect to small occlusions and rotation effects.

3.4. The age

Many of the considered techniques drop in performances, when the time lapse between the training and testing images is not negligible. This makes clear that all the introduced methods do not take into account for problems due to the age variations. Some strategies overcome this problem periodically upgrading the gallery or retraining the system. Nevertheless this not very suitable solution only applies to those systems granting services, which perform the authentication, task frequently, while it is impractical in other situations, such as low enforcement. Alternatively the age of the subject could be simulated trying to make the system more robust with respect to this kind of variation. Several techniques for the age simulation are given in literature: Coordinate Transformations, Facial Composites, Exaggeration of 3D Distinctive Characteristics, but none of these methods has been investigated in the face recognition framework. In a recent work Lanitis and Taylor (2000) and Lanitis et al. (2002) proposed a new method based on age functions. Every image in the face database is described by a set of parameters b, and for each subject the best age function is drawn depending on his/her b. The greatest advantage of this approach is that different subject-based age functions allow taking into account for external factors which contribute towards the age variations. The authors tested this approach on a database of 12 people, with 80 images in the gallery and 85 in the probe. They reported an improvement of about 4–8% and 12–15% swapping the probe and gallery set. In both the experiments the mean age of the subjects has been simulated, before performing the recognition task. Notice that the number of the subject in the database is very small, emphasizing the absence of a standard FERET-like database, which systematically models the age variations. However to improve the robustness of the face recognition systems with respect to changes in age is an interesting and still unexplored aspect in low enforcement applications, mainly for the prediction of facial appearance of wanted/missing persons.

3.5. Is there a more general way to state a technique better than others?

Methods presented in previous sections have both advantages and drawbacks. State which one is the best is very difficult and strongly depends on what is required the system to do. Moreover, most of these approaches have been tested on different datasets.

One way to make a more general evaluation is to pick a set of significant parameters, rather than considering the recognition rate only. As shown in Table 3 the parameter set includes several aspects that need to be taken into account when testing. Examples are number and database characteristics, probe dimension and gallery sets, input size and so on.

It is quite interesting to analyze the way in which these parameters can drive a more accurate comparative study of face recognition algorithms. Obviously, the greater the number of used databases is, the thorough the assessment of the performances can be. On the contrary, the connection between the dimension of the input and the effectiveness of the method is less self-evident. In general, to speed up training/testing tasks, the higher the computational complexity is, the smaller the dimension of the input images can be.

While it is clear that more information is carried by larger input, some studies shows that recognition is still possible on 18×25 greyscale images (Kurita et al., 2003). Most cameras used in video surveillance applications still provide low resolution images, making methods working on smaller images more suitable than others.

However, high resolution images and videos made possible by recent technologies and presented in the upcoming FRVT2005 confirm that the higher the resolution is, the better performances are. The probe and gallery set size also has to be taken into account mainly with respect to the SSS problem. It is well known that only one image is available for training in most real situations, while the identification is performed many times. This suggests that the smaller the gallery set is, the higher the capability of extracting discriminant features is. This can be further improved by a large probe set. It makes sense then to minimize the ratio (gallery size)/(probe size).

Many research results show that several approaches are more sensitive to changes in high frequencies than to low ones. This is not a desirable property, because low frequencies carry most of the invariant information about the identity of the subject, while high frequencies are often affected by changes in environmental conditions. Therefore, the usefulness of a time lapse between sessions providing the images of the gallery and probe set becomes apparent.

As stated in Section 3 five important open questions still need to addressed. Thus, the larger the number of the addressed problems is, the higher the adaptability to real-world applications can be esteemed. Finally, all the methods exposed so far require some kind of input preprocessing; and this could significantly reduce the usefulness of a face recognition algorithm suggesting that the system flexibility increases when normalization on input data is reduced.

Based on these considerations is then possible to investigate which techniques provide a better approximation of pinpointed parameters. The PDBNN based algorithm seems to provide the best experimentation. It addresses most of the problems, while experiments conducted on three different databases with a large number of images reported a high recognition rate. As further example, the LEM approach can be considered. The recognition rate is lower than other methods such as Th-Infrared (Buddharaju et al., 2004) or LDA (Lu et al., 2003), but it has been tested on more databases and it addresses three different problems rather than one. This highlights the robustness of the method.

Table 3
The main information about the experimental results of most of the discussed methods

Method		Databases	Image size	Max|*G*|–Max|*P*|	Time lapse	Recog. rate (%)	Expr.	Ill.	Pose	Occl.	Age
Authors	Name										
Martinez and Kak (2001)	PCA	AR-Faces	85 × 60	100–250	No	70		No	No	No	No
Martinez and Kak (2001)	LDA	AR-Faces	85 × 60	100–250	No	88		No	No	No	No
Belhumeur et al. (1997)	Fisherfaces	YALE		144–16	No	99.6	Yes	Yes	No	No	No
Yu and Yang (2001)	Direct LDA	ORL	112 × 92	200–200	No	90.8	Yes	Yes	Yes	No	No
Lu et al. (2003)	DF-LDA	ORL	112 × 92	200–200	Yes	96		Yes	No	No	No
		UMIST	112 × 92	160–415	No	98		No	No	No	No
Cevikalp et al. (2005)	DCV	Yale	126 × 152	15–150	No	97.33		Yes	No	No	No
		AR-Faces	229 × 299	350–350	Yes	99.35					
Bartlett Marian et al. (2002)	ICA	FERET	60 × 50	425–421	Yes	89	Yes	No	No	No	No
Lin et al. (1997)	PDBNN	SCR	80 × 20	320–1280	No	100	Yes	Yes	Yes	No	No
		FERET		200–200	No	99	Yes	Yes	No	No	No
		ORL			No	96		Yes	Yes	No	No
Joo Er et al. Meng et al. (2002)	RBF	? ORL	160 × 120	300–300		98.1	Yes		Yes	No	No
		PropertyDB				100					
Perronnin and Dugelay (2003)	HMM	FERET	128 × 128	500–500	No	97	Yes	No	No	No	No
Lades et al. (1993)	DLA	PropertyDB	128 × 128	88–88	No	90.3	Yes		Yes	No	No
Liu (2004)	Gabor EFM	FERET	128 × 128	200–100	No	99	Yes	No	No	No	No
		ORL	128 × 128	200–200	No	100	Yes	No	Yes	No	No
Wiskott et al. (1997)	EGM	FERET	256 × 384	250–250	No	80	Yes		Yes	No	No
		PropertyDB		108–//		90	Yes		Yes	No	No
Garcia et al. (2000)	WPA	MIT	480 × 640	155–155		80.5	Yes	Yes		No	no
		FERET	256 × 384	200–400		89					
Kouzani et al. (1997)	IFS	PropertyDB	64 × 64	100–100		100		No	No	No	No
Tan and Tan (1999)	IFS	ORL	92 × 112	200–//	No	95				No	No
Ebrahimpour-Komleh et al. (2001)	IFS	MIT	480 × 640	90–90		90			Yes	No	No
Chen et al. (2003)	Th-Infrared	PropertyDB		166–166	No	98	Yes	Yes	No	No	No
Socolinsky and Selinger (2004b)	Thermal	PropertyDB	99 × 132	770–2310	Yes	93	Yes	Yes	No	No	No
Buddharaju et al. (2004)	Th-Spectrum	Equinox		225–2500		86.8	Yes		Yes	No	No
Pan et al. (2003)	Hyperspectral	PropertyDB		200–1200	Yes	92	No	Yes	No	No	No
Open question methods											
Gao and Leung (2002)	LEM	Bern		40–160	No	72.09	Yes		Yes	No	No
		AR-Faces		112–336		86.03		Yes	No	No	No
		Yale		15–150		85.45		Yes	No	No	No
Kim et al. (2003)	ICA	Subset of AR Faces, Yale, ORL, Bern and FERET	46 × 56	1685–1490		98		Yes	Yes	No	No
Li et al. (2004)	LDA/GSVD LDA/QR	CMU_PIE/Pose27		68–1360	No	100 99.53	No	Yes	No	No	No
		YaleB/Pose00		80–432	No	99 98.03		Yes	No	No	No
Georghiades et al. (2001)	Cones Gen.	Yale B	36 × 42	450–4050	No	97	No	Yes	Yes	No	No
Okada and von der Malsburg (2002)	Linear Subspaces	ATR-Database		2821–804	No	98.7	n	No	Yes	No	No
Gross et al. (2002)	Eigen Lights	CMU-PIE		5304–5304	No	36	No	Yes	Yes	No	No

Table 3 (*continued*)

Method		Databases	Image size	Max|G|–Max|P|	Time lapse	Recog. rate (%)	Expr.	Ill.	Pose	Occl.	Age
Authors	Name										
Martinez (2002)	Martinez	AR-Faces	120 × 170	50–150	No	65	No	No	No	Yes	No
Kurita et al. (2003)	NeuralNetworks	AR-Faces	18 × 25	93–930	No	79	No	No	No	Yes	No
Adini et al. (1997)	ROF	AR Faces	256 × 256	50–150	No	81	Yes	No	No	Yes	No
Lanitis and Taylor (2000)	Age Functions	PropertyDB		80–85	No	71	Yes	Yes	No	No	Yes

4. 3D face recognition

As shown in Face Recognition Vendor Test 2002 (Phillips et al., 2002), the vast majority of face recognition methods based on 2D image processing using intensity or color images, reached a recognition rate higher than 90% under lighting controlled conditions, and whenever subjects are consentient. Unfortunately in case of pose, illumination and expression variations the system performances drop, because 2D face recognition methods still encounter difficulties.

In a recent work, Xu et al. (2004) compared intensity images against depth images with respect to the discriminating power of recognizing people. From their experiments, the authors concluded that depth maps give a more robust face representation, because intensity images are heavily affected by changes in illumination.

Generally, for 3D face recognition is intended a class of methods that work on a three-dimensional dataset, representing both face and head shape as range data or polygonal meshes. The main advantage of the 3D based approaches is that the 3D model retains all the information about the face geometry. Moreover, 3D face recognition also growths to be a further evolution of 2D recognition problem, because a more accurate representation of the facial features leads to a potentially higher discriminating power. In a 3D face model, facial features are represented by local and global curvatures that can be considered as the real signature identifying persons. The 3D facial representation seems to be a promising tool coping many of the human face variations, extra-personal as well as intra-personal.

Two main representations are commonly used to model faces in 3D applications that are 2.5D and 3D images (see Fig. 6). A 2.5D image (range image) consists of a two-dimensional representation of a 3D points set (x, y, z), where each pixel in the X–Y plane stores the depth value z. One can think of a 2.5D image as a grey-scale image, where the black pixel corresponds to the background, while the white pixel represents the surface point that is nearest to the camera. In particular, a 2.5D image taken from a single viewpoint only allows facial surface modelling, instead of the whole head. This problem is solved by taking several scans from different viewpoints, building a 3D head model during a training stage. On the contrary, 3D images are a global representation of the whole head, and the facial surface is further related to the internal anatomical structure, while 2.5D images depend on the external appearance as well as environmental conditions.

The simplest 3D face representation is a 3D polygonal mesh, that consists of a list of points (vertices) connected by edges (polygons). There are many ways to built a 3D mesh, the most used are combining several 2.5D images, properly tuning a 3D morphable model or exploiting a 3D acquisition system (3D scanner). A further difference between 2.5D and 3D images is that last ones are not affected by self-occlusions of the face, when the pose is not full-frontal.

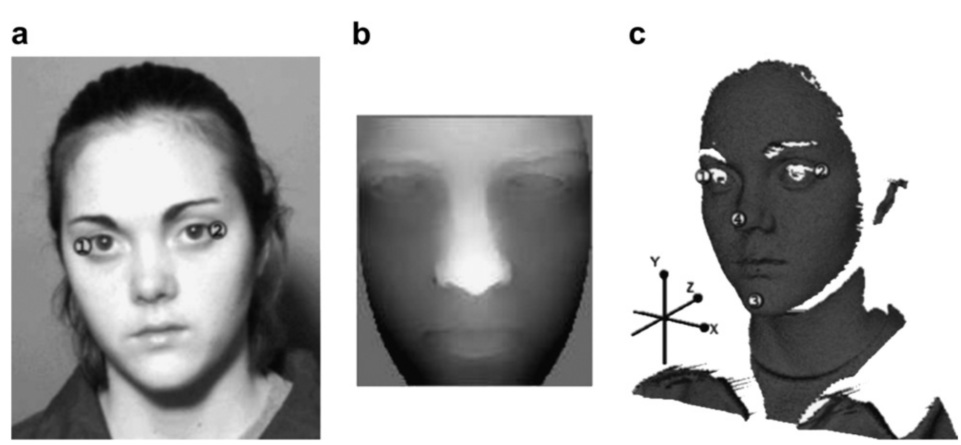

Fig. 6. (a) 2D image, (b) 2.5 image and (c) 3D image (Chang et al., 2004).

4.1. 3D face data acquisition

3D face recognition based technique should posses several proprieties such as robustness with respect to lighting variations as well as position, rotation and scaling of the original model within an absolute reference frame. Unfortunately, 3DFR technique does not achieve all these goals completely.

For example, to align a 3D polygonal mesh within an absolute reference frame could be computationally expensive and existing methods are not always convergent. In addition, the assertion that 3D data acquisition (laser and structured light scanners) is light independent is not completely true; 3D sensors could be affected by strong light source or by reflective surfaces, so it can be asserted that different light sources could generate quite different 3D data sets.

Nevertheless, the use of 3D data can provide some improvements with respect to 2D images, increasing robustness with respect to viewpoints and illumination variations. Then it gives the potential for greater accuracy and a more suitable description of facial features.

There are two main ways to generate a 3D facial model from a human face. The first one is by means of a 3D scanner, which returns a set of points approximating the facial surface or by capturing range images representing the face depth. The 3D acquisition of the face surface is not so far from the scannerization of a generic 3D object. Indeed, the face is acquired in sequential steps:

(1) 3D data are aligned according to reference system of the camera, so that the z-axis is along the optical axis;
(2) 3D data coming from different point of views undergo to a fusion process;
(3) 3D data are optimized according to some given criteria. Finally, the 3D polygonal mesh is built from the cloud of 3D points by using a mesh generation algorithm.

Concerning the available technologies in the field of 3D data acquisition, three main solutions can be mentioned. First of all, the stereoscopic camera system (i.e., Geometrix, Geometrix System, 2005): by taking snapshots of the object, it reconstructs its original 3D shape, by means of a triangulation process, matching corresponding points in both pictures. Secondly, a structured light is used to scan the object; in particular, distortions of different light patterns (grids, stripes, elliptical patterns) are used to deduce the 3D shape of the object. At last, a laser range finding system is also available. It projects a laser light onto the facial surface, while a digital camera computes the position of points along the laser stripe, in three dimensions. All these methods are able to scan both the 3D shape and the color skin (texture map) at the same time.

The second way to acquire a 3D face model is by exploiting a morphable model. The main idea is that by selecting a large set of parameters, any arbitrary face can be generated from a generic 3D face model, properly tuning these parameters, fitting the morphable model to the given image. This approach is more robust with respect to pose, rotation and scaling, because the generic 3D morphable model can be aligned to the input image according to its reference system. However, there are two main drawbacks: (1) the computational cost can be very high; (2) the accuracy of the resulting model depends on the number and quality of the selected parameters.

4.2. 3D face recognition methods

Few papers on this topic have been published even if 3D face recognition research started in last eighties. Many criteria can be adopted to compare existing 3D face algorithms by taking into account the type of problems they address or their intrinsic properties. Indeed, some approaches perform very well only on faces with neutral expression, while some others try also to deal with expression changes. An additional parameter to measure 3D models based robustness is represented by how sensitiveness they are to size variation. In fact, sometimes the distance between the target and the camera can affect the size of the facial surface, as well as its height, depth, etc. Therefore, approaches exploiting a curvature-based representation cannot distinguish between two faces with similar shape, but different size.

In order to overcome this problem some methods are based on point-to-point comparison or on volume approximation. However, the absence of an appropriate standard dataset containing large number and variety of people, whose images were taken with a significant time delay and with meaningful changes in expression, pose and illumination, is one of the great limitations to empirical experimentation for existing algorithms.

In particular, 3D face recognition systems are tested on proprietary databases, with few models and with a limited number of variations per model. Consequently, comparing different algorithms performances often turns into a difficult task. Nevertheless, they can be classified based on the type of problems they address such as mesh alignment, morphing, etc.

The goal for this section is to present a terse description of most recent 3D based face recognition algorithms. Methods have been grouped in three main categories: 2D image based, 3D image based and multimodal systems. The first category includes methods based on comparisons among intensity images, but supported by a three-dimensional procedure that increases the system robustness. The second class groups approaches based on 3D facial representation, like range images or meshes. Finally, methods combining 2D image and 3D image information fall in the third category.

4.2.1. 2D-based class

Approaches based on 2D images supported by some 3D data are identified as 2D-based class methodologies. Generally, the idea is to use a 3D generic face model to improve robustness with respect to appearance variations such as hard pose, illumination and facial expression. An example of this approach is given by Blanz and Vetter (2003). They proposed to synthesize various facial variations by using a morphable model that augments the given training set containing only a single frontal 2D image for each subject. The morphable face is a parametric model based on a vector space representation of faces. This space is constructed so that any convex combination of shape and texture vectors belonging to the space describes a human face. Given a single face image, the algorithm automatically estimates 3D shape, texture, and all relevant 3D scene parameters like pose, illumination, etc. (see Fig. 7), while the recognition task is achieved measuring the Mahalanobis distance (Duda et al., 2001) between the shape and texture parameters of the models in the gallery and the fitting model. The identification has been tested on two publicly available databases of images: CMU-PIE (Sim et al., 2003) and FERET (Phillips et al., 2000). A recognition rate of 95% on CMU-PIE dataset and 95.9% on FERET dataset is claimed. Another interesting approach using a 3D model to generate various 2D facial images is given by Lu et al. (2004). They generated a 3D model of the face from a single frontal image. From this 3D model many views are synthesized to simulate new poses, illuminations and expressions. Tests are performed by measuring dissimilarities among affine subspaces according to a given distance measure. In particular, an affine subspace contains all the facial variations synthesized for a single subject. They performed experiments on a dataset of 10 subjects building 22 synthesized images per subject with different poses, facial expressions and illuminations. The method achieves a recognition rate of 85%, outperforming the PCA-based methods on this dataset. Nevertheless, very few people are in the database, making difficult to estimate accurately the real discriminating power of the method. On the contrary, Hu et al. (2004) show that linear methods such as PCA and LDA can be further extended to cope with changes in pose and illumination by using a Nearest Neighbor approach. The dataset is gathered on 68 subjects and 41.368 bi-dimensional images under various facial expression, illuminations and poses. Their results show that using virtual face for particular poses increase the recognition rate and the highest rate reached 95% when pose is approximately frontal and LDA is used.

Creating various 2D synthetic faces could be good way to overcome the classical problems of 2D face recognition, but two important considerations have to be carefully examined: "how much realistic is a synthesized face?" and "how precise can a 3D facial reconstruction taken by one single picture be?". First of all, we have to consider that modern 3D computer graphics technologies are able to reproduce synthetic images in an excellent realistic way and with an accurate geometric precision. Secondly, we have to consider that 3D facial reconstruction from a single view image can be considered good enough, only if the experimental results show a high discriminating power.

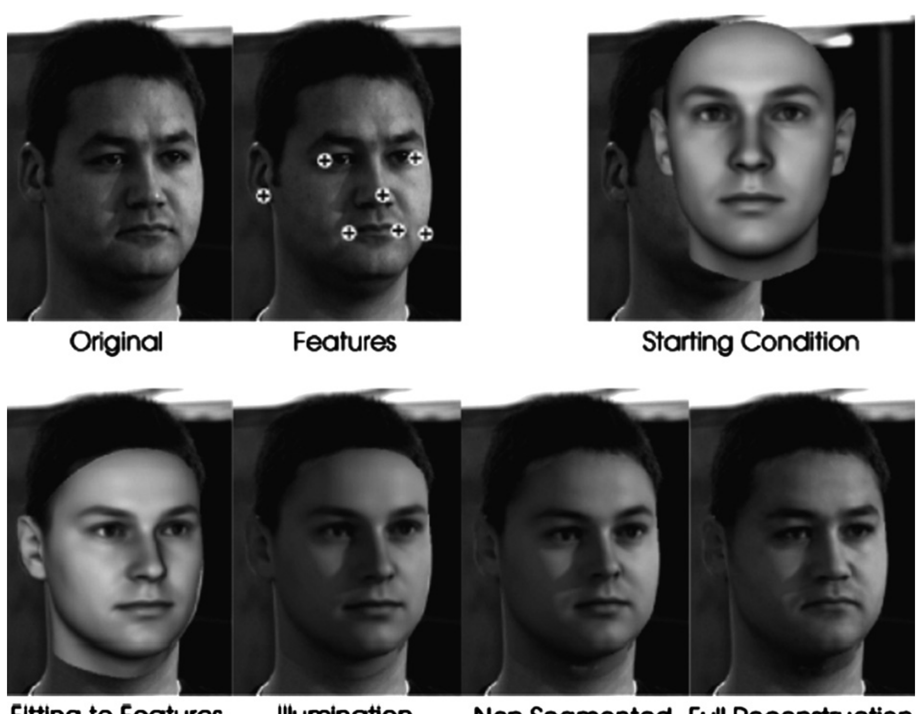

Fig. 7. Face reconstruction from a single image (Blanz and Vetter, 2003).

4.2.2. 3D-based class

This subsection explores several methodologies that work directly on 3D datasets. The first problem concerning 3D face recognition is to set up a correct alignment between two face surfaces. One possible approach to gain a correct alignment is by using an acquisition system based on a morphable model, because it is pre-aligned within a given reference frame. The work presented by Ansari and Abdel-Mottaleb (2003) could be considered as an example of this kind of methods. Starting from one frontal and one profile view image, they use 3D coordinates of a set of facial feature points to deform a morphable model fitting the real facial surface. The deformation of the model is performed in two steps. At first a global deformation is carried out to scale and to align the morphable model to the feature points extracted from the pair images. Then a local deformation is applied to bring the vertices as close as possible to feature points. The recognition task is then performed calculating the Euclidean distance between 29 features points lying on 3D facial surface on mouth, nose and eyes. Their experimental results show a recognition rate of 96.2% on a database of 26 subjects with two pairs of images, one used for training and the other for testing.

The Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992) is often used as an alternative approach aligning models. It could be used to reduce misalignment during the registration phase as well as to approximate the volume difference between two surfaces. Even though, it leads to problems with convergence when the initial misalignment of the data sets is too large, typically over 15°; it is possible countered to this limitation with a coarse pre-alignment. An approach based on Iterative Closest Point algorithm is given by Cook et al. (2004). They use ICP only to establish the correspondence between 3D surfaces in order to compensate problems due to non-rigid nature of faces. Then, once the registration is done, faces are compared by using a statistical model, namely Gaussian Mixture Model (GMM), and the distribution of the errors is then parameterized. They performed experiments on the 3D RMA database (Beumier and Acheroy, 2000) reaching a recognition rate of 97.33%. A quite similar ICP-based approach to find a point-to-point correspondence between landmarks features is given by Irfanoglu et al. (2004). They described a method to obtain a dense point-to-point matching by means of a mesh containing points that are present in all faces, so that the face alignment is trivially obtained. Then, once the dense correspondence is established, the Point Set Distance (PSD), that is a discrete approximation of the volume between facial surfaces, is used to compute the distance between two different clouds of points. In the experiments, they tested the algorithm on the 3D RMA database with a resulting recognition rate of 96.66%. Even if the ICP is a powerful tool in order to estimate the similarity between two faces, it has a serious lack. Indeed, ICP-based methods treat the 3D shape of the face as a rigid object so they are not able to handle changes in expression. Medioni and Waupotitsch (2003) proposed an ICP-based approach, that aligns two face surfaces and calculates a map of differences between the facial surfaces, then applying statistic measures in order to obtain a compact description of this map. They built 3D models of 100 subjects by using a stereo system; each subject has been acquired in 7 different poses within degrees with respect to the frontal view. The recognition rate on this dataset was 98%. As said before, a different use of the ICP algorithm is to approximate the surface difference between two faces. Indeed, the work of Lu et al. (2004) is headed in this direction. They describe both a procedure for constructing a database of 3D mesh models from several 2.5D images and a recognition method based on ICP algorithm. In order to build up the 3D meshes, features points are automatically detected on 2.5D images, searching for maximum and minimum local curvatures, so that ICP is run on these points aligning all the 2.5D images. Then, the recognition match between faces is carried out exploiting the local feature information correlated by the ICP. For the experiments, they report a recognition rate of 96.5% using a database of 113 range images for 18 subjects with different poses, facial expressions and illuminations.

A further interesting aspect dealing with 3D face recognition concerns the analysis of the 3D facial surface in order to extrapolate information about the shape. Some approaches are based on a curvature-based segmentation detecting a set of fiducial regions. Gordon (1991) presented a new method based on the idea that some facial descriptors, such as the shape of forehead, jaw line, eye corner cavities and cheeks, remain generally similar although they are taken by different range images for the same subject. This is not completely true when detection errors or changes in expression occur. His method consists in two different tasks: the former extracts a set of high level shape descriptors, for eyes, nose and head; the latter uses these descriptors to compute a set of basic scalar features corresponding to distance measurements. At last, each face images is projected in the feature space, while the Euclidean distance between feature vectors is used as a metric. The experiments of this method shows a recognition rate of 100% using a small training set of 8 subjects with three different view for each for a total of 24 faces.

Another interesting segmentation approach based on Gaussian curvature has been proposed by Moreno et al. (2003). For each 3D facial model, they detect a set of 86 different segmented regions by using an algorithm exploiting the signs of the median and Gaussian curvatures in order to isolate regions with significant curvatures (see Fig. 8). Then, this feature space is reduced in order to increase the efficiency. Finally, a feature vector is created for each subject. Experiments have been conducted on a dataset of 420 3D facial models belonging to 60 subjects, including images with light, rotation and facial expression variations, achieving a recognition rate of 78% for the best match and 92% for the five best matches. In addition, the segmentation process can be used to treat face recognition

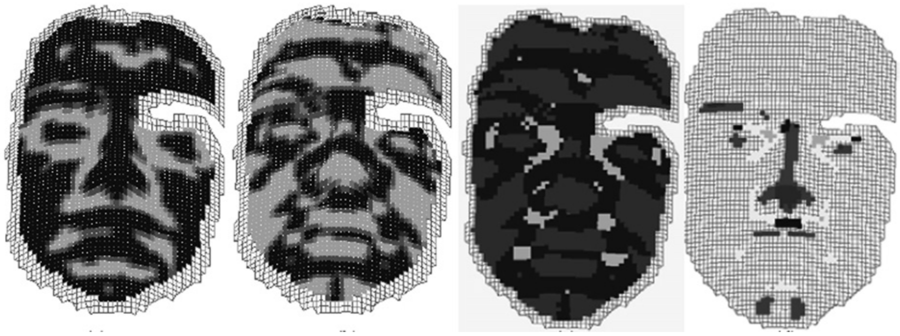

Fig. 8. Examples of 3D mesh segmentation based on local curvature (Moreno et al., 2003).

problem as a non-rigid object recognition problem to improve the robustness to facial expression variations.

Chua et al. (2000) observed that on there are regions on facial surfaces, such as nose, eye socket and forehead which undergo to much less deformation in case of expression changes (see Fig. 9). They find these "rigid" facial regions by using a Point Signature two-by-two comparison (Chua and Jarvis, 1997) among different facial expressions of the same person. Then, they store only the rigid parts in an indexed library, ranking models according to their similarity. Their experiment shows a recognition rate of 100% on a dataset of 6 subjects and 4 facial expression variations. To model facial shape is also possible by creating a mathematical framework representing local/global curvatures.

Another kind of approach to the analysis of facial shape is to create a mathematical model representative of local curvatures. This is a good way to account the 3D surface in a compact fashion using few features descriptors to characterize a face, without a wasteful complexity time. In addition, a local curvature-based representation better cope the non-rigid nature of face due to facial expressions because, though expressions changes the facial surface globally and the local curvature relations are preserved. Unluckily, this kind of representation is not able to handle information about the size of face, doing not possible to distinguish two similar faces but with different sizes. Tanaka et al. (1998) proposed an example of these approaches performing a correlation-based face recognition based on analysis of minimum and maximum principal curvature and their directions, to describe the facial surface shape. Then, these descriptors are mapped on two unit spheres, the Extended Gaussian Images (EGI). The similarity match is performed by using the Fisher's spherical approximation on the EGIs of faces. The method worked on 37 range images gathered by National Research Council of Canada (NRCC) (Rioux and Cournoyer, 1988), providing a recognition rate of 100%. On the contrary, Wang et al. (2004) presented a viewpoint-invariant technique based on a free-form representation, called Sphere-Spin-Images (SSI). The SSIs are used to describe locally the shape of the facial surface. The SSIs of a point are constructed by mapping the 3D point coordinates lying into a sphere space, centered in that point, into a 2D space. The main aim of this mapping is to represent the local shape of points by means of a histogram. To describe a face, the method selects a small set of fixed points by means of a minimum principal curvature analysis and builds a single SSI series for each subject. Then, a simple correlation coefficient is used to compare the similarity between different SSI series. They performed tests on the SAMPL dataset (Range Imagery), with 31 models of 6 different subjects, reporting a recognition rate of 91.68%. Then, a simple correlation coefficient is used to compare the similarity between different SSI series. They performed tests on the SAMPL dataset (Range Imagery), with 31 models of 6 different subjects, reporting a recognition rate of 91.68%.

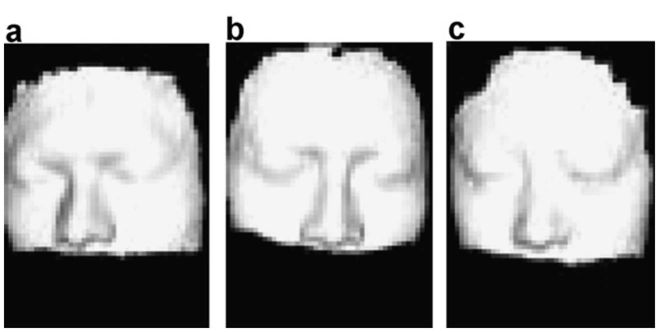

Fig. 9. Examples of 3D rigid face regions of three different subjects (Chua et al., 2000).

In Section 2.1, the Principal Component Analysis (PCA) has been mentioned as a technique largely used in 2D face recognition in order to classify face images, reducing the dimensionality of image input space. In 3D face recognition is applied treating data as a cloud of points rather than a surface and new axes that best summarize the variance across the vertices are determined. Thus, the PCA is able to work with different facial poses producing a descriptive model of the facial shape. This approach has been extended to the 3D face recognition by Hesher et al. (2002). The method apply the PCA directly to the range images, while they use the Euclidean distance to measure similarities among the resulting feature vectors. The authors state this method reached a recognition rate of 100% on a dataset of 222 range images of 37 subjects with different facial expressions. Further investigations on PCA in the 3D framework

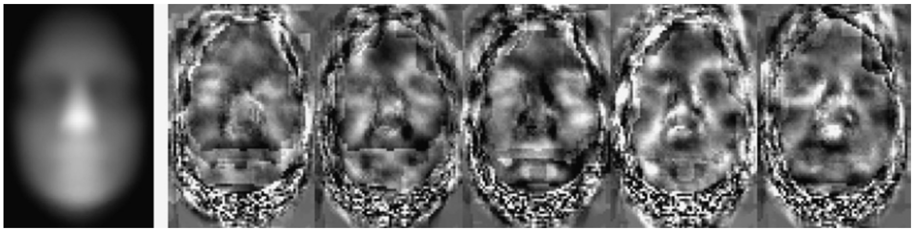

Fig. 10. A range image and the first five fishersurfaces (Heseltine et al., 2004b).

have been carried out by Heseltine et al. They presented two different works based on PCA theory, showing experimental results with several facial surface representations given by different convolution kernels and several distance metrics such as the Euclidean and cosine. The first method (Heseltine et al., 2004a) is based on a PCA-based eigensurface approach and is gathered on a data set of 330 three-dimensional mesh models available by The University of York (The 3D Face Database, 2003). It reaches a recognition rate of 87.3%. The second approach (Heseltine et al., 2004b) is an adaptation of traditional 2D Belhumeur's fisherface approach (Belhumeur et al., 1997) to 3D facial surface data (see Fig. 10). The results are gathered on a data set of 1770 three-dimensional mesh models with 280 subjects with several poses and facial expressions. The highest recognition rate it reaches is 88.7% when the surface gradients representation and the cosine distance metrics are used.

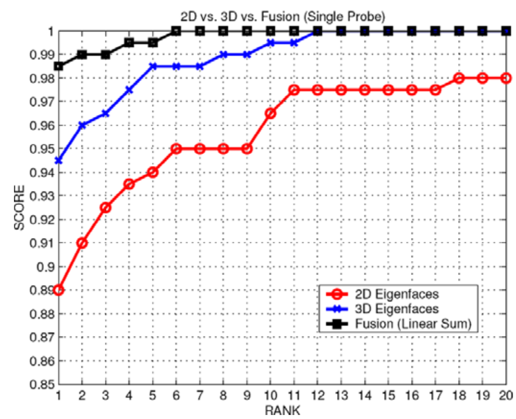

Fig. 11. PCA-based recognition experiments performed using 2D and 3D Eigenfaces.

4.2.3. 2D + 3D-based class

Multimodal approaches combine information coming from 2D image as well as 3D model of faces. Recently Chang et al. (2003) investigated on possible improvements that 2D face biometric can receive integrating the 3D also. The method, they proposed, performs separately the PCA on the intensity and range images and then combines results obtained from both strategies to get a global response from the system. The authors assert four important conclusion: "(1) 2D and 3D have similar recognition performance when considered individually, (2) Combining 2D and 3D results using a simple weighting scheme outperforms either 2D or 3D alone, (3) Combining results from two or more 2D images using a similar weighting scheme also outperforms a single 2D image, and (4) Combined 2D + 3D outperforms the multi-image 2D result" (Chang et al., 2004). Experiments have been conducted on a dataset of 275 subjects by using a single and a multiprobe set. The recognition rate is 89.5% for the intensity images and 92.8% for the range images, while the combined solution provides a global rate of 98.8% (see Fig. 11).

Bronstein et al. (2003) presented a new method based on a bending invariant canonical representation (Fig. 12), they called canonical image that models deformations resulting from facial expression and pose variations. They observe that facial expressions are not arbitrary, but they can be modelled by using isometric transformations. The canonical image stores these geometric invariants and it is built by calculating the geodesic distances between points on facial surface. The 2D face image is mapped onto the canonical image shape flattening the texture coordinates

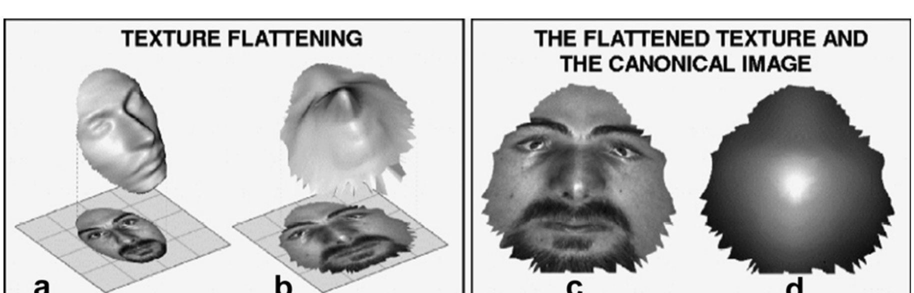

Fig. 12. Bronstein et al. (2003) facial surface representation: (a) texture mapping on the facial surface (b) and on the canonical form; (c) the resulting flattened texture and (d) the canonical image.

onto the canonical surface. The experimental results are performed on a database of 157 subjects but nothing has been said about recognition rates.

On the contrary, Tsalakanidou et al. (2003) proposed an HMM approach to integrate depth data and intensity image. The method start localizing the face with a depth and brightness based procedure, while the recognition task exploits the embedded hidden Markov model technique that is applied to 3D range images as well as 2D images. The experimental results are gathered on a very large database of 3000 range and greyscale images of 50 subjects, with various facial expressions, poses, illuminations and with/without glasses, reporting a recognition rate of 90.7% on 2D intensity images and 80% on 3D range images, while the system reaches a rate of 91.67%, when both information are combined. Papatheodorou and Rueckert (2004)

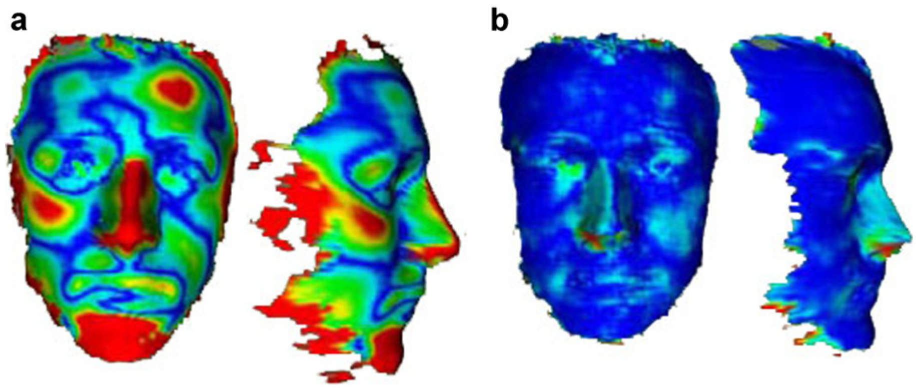

Fig. 13. Color-based representation of residual 3D distances (a) from two different subjects and (b) from the same subject (Papatheodorou and Rueckert, 2004).

Table 4
The main information about the experimental results of most of the discussed 3D methods

Method		Image/model size	No. subj.	No. images	Recogn. rate (%)	Database characteristics			
Authors	Type					Expr.	Ill.	Pose	Occl.
2D based									
Blanz and Vetter (2003)	2D-based	–	–	4,488 (PIE) + 1,940 (FERET)	95 and 95.9	Yes	Yes	Yes	No
Lu et al. (2004)	2D-based	64 × 64	10	220	85	Yes	Yes	Yes	No
Hu et al. (2004)	2D-based	–	68	41.368	<95	Yes	Yes	Yes	No
3D based									
Ansari and Abdel-Mottaleb (2003)	Set of points	–	26	52	96.2	No	No	No	No
Cook et al. (2004)	GMM (ICP)	4000 points	120	360 (3D_RMA)	97.33	Yes	No	Yes	No
Irfanoglu et al. (2004)	Set of points (ICP)	4000 points	30	90 (3D_RMA)	96.66	Yes	No	Yes	No
Medioni and Waupotitsch (2003)	ICP	–	100	700	98	No	No	Yes	No
Lu et al. (2004)	ICP	320 × 240 (18,000 points)	18	113	95.5	Yes	Yes	Yes	No
Gordon (1991)	Local shape descriptors	–	8	24	100	No	No	No	No
Moreno et al. (2003)	Local shape descriptors	2186 points	60	420 (GavabDB)	78	Yes	Yes	Yes	No
Chua et al. (2000)	Segmentation	–	6	24	100	Yes	No	No	No
Tanaka et al. (1998)	EGI	256 × 256	37	37	100	No	No	No	No
Wang et al. (2004)	Sphere-spin-images	200 × 200	6	31 (SAMPL)	91.68	No	No	Yes	No
Gao and Leung (2002)	PCA-based	10,000 points	37	220	100	Yes	No	No	No
Heseltine et al. (2004a)	PCA-based	–	100	330 (YORK)	<87.3	Yes	No	Yes	No
Heseltine et al. (2004b)	PCA-based	–	280	1770 (YORK)	<88.7	Yes	No	Yes	No
Multimodal									
Chang et al. (2004)	Multimodal	640 × 480	275	–	98.8	Yes	Yes	Yes	No
Bronstein et al. (2003)	Multimodal	2000–2500 points	157	–	–	Yes	No	Yes	No
Tsalakanidou et al. (2003)	Multimodal	571 × 752	50	3000	91.67	Yes	Yes	Yes	Yes
Papatheodorou and Rueckert (2004)	Multimodal	10,000 points	62	∼900	66–100	Yes	Yes	Yes	No

proposed a 4D registration method based on Iterative Closest Point (ICP), but adding textural information too. The data acquisition is done with a stereo camera system composed by three cameras and a pattern projector, while the measurement of facial similarity involves a 4D Euclidean distance (represented by colors[1] as shown in Fig. 13) between four-dimensional points: the three spatial coordinates more the texel intensity information. They report various results on a dataset collected from 62 subjects with 13 distinct 3D meshes and 2D texture maps considering several facial expression and poses. The percentage of correct matches and correct rejection are used as performance measures. In case of frontal pose, results show that the use of both texture and shape improves performances, while a percentage of correct recognition ranging from 66.5% to 100%, depending on several poses and expressions.

All 3D based methods introduced so far are summarized in Table 4 in addition to a small set of parameters, that can be considered meaningful for a more complete and accurate evaluation of discussed approaches. In general, the recognition rate is a widely used measure for the evaluation of face recognition methods, but it strongly depends on the number of people in the database and the number of images per subject gathered for the experimental results. In addition the key features (illumination (i), expression (e), pose (p), occlusions (o)) considered on the models in the probe and gallery set are reported, in order to take into account for the testing framework, in which 3D methods have been tested.

5. Discussion and remarks

Approaches discussed in previous sections address only parts of existing open questions and a technique able to provide best performances under any circumstances does not exist yet.

In Section 2 many strategies have been analyzed, showing that almost all methods claim satisfactory recognition rates, but only when tested on standard databases or some parts of them. On the contrary, in Section 2.1. it has been observed that linear/nonlinear approaches overcome other methods when illumination changes occur. However, this class of methods is noticeably affected by changes in pose and they perform worse when both variations are present. Similarly, methods that cope with variations in pose and illumination, such as Line Edge Map (Gao and Leung, 2002) suffer from the presence of occlusions and age variations. The lack of a wide face database modelling a real-world scenario, in terms of differences in gender and ethnic group as well as expression, illumination, pose, etc. is not negligible too. Indeed, as shown by Table 1, only FERET, AR Faces and CMU-PIE provide a substantial number of face images. Nevertheless, FERET involved a very large number of people with a satisfactory differentiation in pose, illumination, gender and time delay, but no photos are taken with natural occlusions or make-up. AR Faces, instead, provides many color images of fewer people also with natural occlusions (scarves and sunglasses), while differences in ethnic groups are not very satisfactory. Likewise, the CMU-PIE database only takes into account for changes in pose and illumination.

In the same way, 3D image analysis has the potential to grow the recognition performances of the 2D face recognition with respect to pose, illumination and expression variations, but there are many challenging problems to be still addressed, such as alignment of meshes or sensitiveness of the acquisition process. Even if alternative solutions have been proposed to overcome problems such as the slowness and convergence of the ICP-based methods (for example the EGI), they introduces other drawbacks as well (the lost of spatial information in the case of EGI).

Another important issue for face recognition is the detection of occlusive objects, like glasses and facial hair. At present, none of 3D face recognition methods try to deal with occlusions although the 3D segmentation process have the potentiality to work better than into a 2D special representation. Finally, a smart integration between the texture image and facial surface is necessary. The most part of 2D + 3D approach work on 2D and 3D data separately, merging final rates together overlapping the same information taken from two differences sources. For this purpose, a proper use of data sources is needed, choosing the more appropriate representation time by time (i.e., the facial hair detection is more simply on 2D images but the segmentation process is more powerful if done on the 3D facial surface).

Therefore, the current state of the art makes clear that, by themselves, none of the existing techniques is able to cope with all kind of distortions that a face can undergo, suggesting that only a proper combination of the current methods can be robust enough to be applied in a real-world scenario. Three main types of multimodal systems can be conjectured, depending on the involved biometrics. The lowest level of multimodality is represented by a system integrating several 2D image based face recognition subsystems. The real advantage of this system is that all the strengths of the face biometric even hold, making it also suitable in case of outdoor conditions. However setting up this framework presents two great difficulties. First of all, the way to assess what kind of distortions a face present, in order to select a correct class of sub-systems to process the input. Secondly, if more sub-systems have been activated, it is very difficult to state which fusion strategy provides better performances, in terms of global system response.

A more complex type of multimodality is to combine 2D algorithms with 3D ones. Section 4.2.3 provides some examples of experimented systems in literature. Indeed recent works confirm that combining 2D and 3D approaches leads to results that overcome both 2D and 3D, in terms of recognition rate. Nevertheless, this kind

[1] For interpretation of color in figures, the reader is referred to the Web version of this article.

of systems inherits common problems of 3D processing, in particular, concerning the acquisition process. At last, a quite comprehensive scheme is represented by those systems integrating several kind of biometrics. Indeed, many systems combine together fingerprints and voice, sometimes, a badge also. However not all possible combinations among existing biometrics are suitable; some of them (i.e., face iris) can reduce the flexibility of the system to real-world applications. Therefore, some works combining face and ear have been proposed. Not much has been done in combining face and voice, instead. Once again, the main check on the research in this direction is the lack of a proper multimodal database, providing a complete as well as satisfactory testing framework.

References

Adini, Y., Moses, Y., Ullman, S., 1997. Face recognition: The problem of compensating for changes in illumination direction. IEEE Trans. Pattern Anal. Machine Intell. 19 (7), 721–732, July.

Ansari, A., Abdel-Mottaleb, M., 2003. 3-D face modeling using two views and a generic face model with application to 3-D face recognition. In: Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS 2003) Miami, FL, USA, July, pp. 37–44.

Bartlett Marian, Stewart, Movellan Javier, R., Sejnowski Terrence, J., 2002. Face recognition by independent component analysis. IEEE Trans. Neural Networks 13 (6), 1450–1464.

Belhumeur, Peter N., Hespanha, João P., Kriegman, David, 1997. Eigenfaces vs. fisherfaces: Using class specific linear projection. IEEE Trans. Pattern Anal. Machine Intell. 19 (7), 711–720.

Besl, P.J., McKay, N.D., 1992. A method for registration of 3-d shapes. IEEE Trans. Pattern Anal. Machine Intell. 14 (2), 239–256.

Beumier, Charles, Acheroy, Marc, 2000. Automatic 3D face authentication. In: Proc. Image and Vision Computing, Vol. 18(4), pp. 315–321.

Blanz, V., Vetter, T., 2003. Face recognition based on fitting a 3D morphable model. IEEE Trans. Pattern Anal. Machine Intell. 25 (9), 1063–1074.

Bronstein, A., Bronstein, M., Kimmel, R., 2003. Expression-invariant 3D face recognition. Proc. Audio & Video-based Biometric Person Authentication (AVBPA). In: Lecture Notes in Computer Science, Vol. 2688. Springer, pp. 62–69.

Buddharaju, P., Pavlidis, I., Kakadiaris, I., 2004. Face recognition in the thermal infrared spectrum. In: 2004 Conf. Computer Vision and Pattern Recognition Workshop June, pp. 133–138.

Cevikalp, H., Neamtu, M., Wilkes, M., Barkana, A., 2005. Discriminative common vectors for face recognition. IEEE Trans. Pattern Anal. Machine Intell. 27 (1), 4–13.

Chang, K., Bowyer, K., Flynn, P., 2003. Face Recognition using 2D and 3D facial data. In: Proc. ACM Workshop on Multimodal User Authentication December, pp. 25–32.

Chang, K., Bowyer, K., Flynn, P., 2004. An evaluation of multi-modal 2D + 3D face biometrics. IEEE Trans. Pattern Anal. Machine Intell. (PAMI 2004).

Chen, L.-F., Mark Liao, H.-Y., Ko, M-.T., Lin, J.-C., Yu, G.-J., 2000. A new LDA-based face recognition system which can solve the small sample size problem. Pattern Recognition 33 (October), 1713–1726.

Chen, X., Flynn, P.J., Bowyer, K.W., 2003. PCA-based face recognition in infrared imagery: Baseline and comparative studies. In: Proc. IEEE Internat. Workshop on Analysis and Modeling of Faces and Gestures October, pp. 127–134.

Chua, C.S., Jarvis, R., 1997. Point signature: A new representation for 3d object recognition. Internat. J. Computer Vision 25 (1), 63–85.

Chua, C.S., Han, F., Ho, Y.K., 2000. 3D human face recognition using point signature. In: Proc. 4th IEEE Internat. Conf. on Automatic Face and Gesture Recognition (FG 2000) Grenoble, France, March, pp. 233–238.

Cook, J., Chandran, V., Sridharan, S., Fookes, C., 2004. Face recognition from 3D data using iterative closest point algorithm and gaussian mixture models. In: Proc. 2nd Internat. Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2004), Thessaloniki, Greece, 6–9 September.

Cottrell, G.W., Fleming, M.K., 1990. Face recognition using unsupervised feature extraction. In: Proc. Intell. Neural Network Conf., pp. 322–325.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, second ed. John Wiley & Sons.

Ebrahimpour-Komleh, H., Chandran, V., Sridharan, S., 2001. Face recognition using fractal codes. In: 2001 Internat. Conf. on Image Processing. Vol. 3 (October), pp. 58–61.

Gao, Y., Leung, M.K.H., 2002. Face recognition using line edge map. IEEE Trans. Pattern Anal. Machine Intell. 24 (6), 764–779.

Garcia, C., Zikos, G., Tziritas, G., 2000. Wavelet packet analysis for face recognition. Image Vision Comput. 18, 289–297.

Geometrix System, 2005. <www.geometrix.com>, 12 Mars 2005.

Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J., 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Machine Intell. 23 (6), 643–660.

Gordon, G., 1991. Face recognition based on depth maps and surface curvature. In: Proc. Geometric Methods in Computer Vision, SPIE, pp. 1–12.

Gross, R., Matthews, I., Baker, S., 2002. Eigen light-fields and face recognition across pose. In: Proc. Fifth IEEE Internat. Conf. on Automatic Face and Gesture Recognition, May, pp. 1–7.

Heseltine, T., Pears, N., Austin, J., 2004a. Three-dimensional face recognition: An eigensurface approach. In: Proc. Internat. Conf. on Image Processing (ICIP 2004), October, Singapore.

Heseltine, T., Pears, N., Austin, J., 2004b. Three-dimensional face recognition: A fischersurface approach. In: Proc. Image Analysis and Recognition: Internat. Conf. (ICIAR 2004), Porto, Portugal (September–October), pp. 684–691.

Hesher, C., Srivastava, A., Erlebacher, G., 2002. Principal component analysis of range images for facial recognition. In: Proc. Internat. Conf. on Imaging Science, Systems and Technology (CISST 2002).

Hu, Y.X., Jiang, D.L., Yan, S.C., Zhang, H.J., 2004. Automatic 3D Reconstruction for Face Recognition. In: Proc. Proceeding of IEEE International Conf. on Automatic Face and Gesture Recognition (FGR 2004), Seoul, Korea, May, pp. 843–850.

Irfanoglu, M.O., Gokberk, B., Akarun, L., 2004. 3D shape-based face recognition using automatically registered facial surfaces. In: Proc. 17th Internat. Conf. on Pattern Recognition (ICPR2004), Cambridge, pp. 183–186.

Johnson, M.L., 2004. Biometrics and the threat to civil liberties. Computer 37 (4), 90–92. April.

Kim, Tae-Kyun, Kim, Hyunwoo, Hwang, Wonjun, Kee, Seok-Cheol, Kittler, J., 2003. Independent component analysis in a facial local residue space. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1 (June), pp. 579–586.

Kirby, M., Sirovich, L., 1990. Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Trans. Pattern Anal. Machine Intell. 12 (1), 103–108.

Kouzani, A.Z., He, F., Sammut, K., 1997. Fractal face representation and recognition. In: 1997 IEEE Internat. Conf. on Systems, Man, and Cybernetics, Vol. 2 (October), pp. 1609–1613.

Kurita, T., Pic, M., Takahashi, T., 2003. Recognition and detection of occluded faces by a neural network classifier with recursive data reconstruction. In: IEEE Conf. on Advanced Video and Signal Based Surveillance, July, pp. 53–58.

Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., van der Malsburg, C., Wurtz, R.P., Konen, W., 1993. Distortion invariant object recognition in the dynamic link architecture. IEEE Trans. Comput. 42 (March), 300–311.

Lanitis, A., Taylor, J.C., 2000. Robust face recognition using automatic age normalization. In: 10th Mediterranean Electrotechnical Conf., MEleCon. Vol. 2, pp. 478–481.

Lanitis, A., Taylor, J.C., Timothy, F.C., 2002. Toward automatic simulation of aging effects on face images. IEEE Trans. Pattern Anal. Machine Intell. 24 (4), 442–455.

Li, Q., Ye, J., Kambhamettu, C., 2004. Linear projection methods in face recognition under unconstrained illuminations: A comparative study. In: Proc. 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR04).

Lin, Shang-Hung, Kung, Sun-Yuan, Lin, Long-Ji, 1997. Face recognition/detection by probabilistic decision-based neural network. IEEE Trans. Neural Networks 8 (1), 114–132.

Liu, C., 2004. Gabor-based kernel PCA with fractional power polynomial models for face recognition. IEEE Trans. Pattern Anal. Machine Intell. 26 (5), 572–581.

Lu, J., Plataniotis Kostantinos, N., Venetsanopoulos Anastasios, N., 2003. Face recognition using LDA-based algorithms. IEEE Trans. Neural Networks 14 (1), 195–200.

Lu, X., Hsu, R., Jain, A., Kamgar-Parsi, B., Kamgar-Parsi, B., 2004. Face recognition with 3D model-based synthesis. In: Proc. Internat. Conf. on Biometric Authentication (ICBA), Hong Kong, July, pp. 139–146.

Lu, X., Colbry, D., Jain, A.K., 2004. Three-dimensional model based face recognition. In: Proc. 17th Internat. Conf. on Pattern Recognition (ICPR2004), Cambridge, pp. 362–366.

Martinez, A.M., 2002. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE Trans. Pattern Anal. Machine Intell. 24 (6), 748–763.

Martinez, A.M., Kak, A.C., 2001. PCA versus LDA. IEEE Trans. Pattern Anal. Machine Intell. 23 (2), 228–233.

Medioni, G., Waupotitsch, R., 2003. Face recognition and modeling in 3D. In: Proc. IEEE Internat. Workshop on Analysis and Modeling of Faces and Gesture Recognition (AMFG'03). Nice, France, October, pp. 232–233.

Meng, J.E., Shiqian, W., Juwei, L., Hock, L.T., 2002. Face recognition with radial basis function (RBF) neural networks. IEEE Trans. Neural Networks 13 (3), 697–710.

Moreno, A.B., Sanchez, A., Velez, J.F., Diaz, F.J. 2003. Face recognition using 3D surface-extracted descriptors. In: Proc. Irish Machine Vision and Image, (IMVIP'03), September.

Okada, K., von der Malsburg, C., 2002. Pose-invariant face recognition with parametric linear subspaces. In: Fifth IEEE Internat. Conf. on Automatic Face and Gesture Recognition, May, pp. 64–69.

Pan, Z., Healey, G., Prasad, M., Tromberg, B., 2003. Face recognition in hyperspectral images. IEEE Trans. Pattern Anal. Machine Intell. 25 (12), 1552–1560.

Papatheodorou, T., Rueckert, D., 2004. Evaluation of automatic 4D face recognition using surface and texture registration. In: Proc. Sixth IEEE Internat. Conf. on Automatic Face and Gesture Recognition Seoul, Korea, May, pp. 321–326.

Perronnin, F., Dugelay, J.-L., 2003. An introduction to biometrics and face recognition. In: Proc. IMAGE'2003: Learning, Understanding, Information Retrieval, Medical, Cagliari, Italy, June.

Phillips, J.P., Moon, H., Rizvi, A.S., Rauss, P.J., 2000. The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Machine Intell. 22 (10), 1090–1104.

Phillips, P., Grother, P., Micheals, R., Blackburn, D., Tabassi, E., Bone, M., 2002. Face Recognition Vendor Test 2002: Evaluation Report.

Range Imagery, SAMPL (Signal Analysis and Machine perception Laboratory at the Ohio State University). [Online]. Available from: <http://sample.eng.ohiostate.edu>.

Riccio, D., Nappi, M., 2003. Defering range/domain comparisons in fractal image compression. In: Proc. 12th Internat. Conf. on Image Analysis and Processing, Vol. 1 (September), pp. 412–417.

Rioux, M., Cournoyer, L., 1988. The NRCC three-dimensional image data files. Tech. Report, CNRC 29077, National Research Council Canada, Ottawa, Canada.

Sahbi, Hichem, Boujemaa, Nozha, 2002. Robust face recognition using dynamic space warping. In: Biometric Authentication, Internat. ECCV 2002 Workshop Copenhagen, June, pp. 121–132.

Sherrah, J., 2004. False alarm rate: A critical performance measure for face recognition. In: Sixth IEEE Internat. Conf. on Automatic Face and Gesture Recognition. May, pp. 189–194.

Sim, T., Baker, S., Bsat, M., 2003. The CMU pose, illumination, and expression database. IEEE Trans. Pattern Anal. Machine Intell. 25 (12), 1615–1618.

Socolinsky, D.A., Selinger, A., 2004a. Thermal face recognition in an operational scenario. In: Proc. 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR04).

Socolinsky, D.A., Selinger, A., 2004b. Thermal face recognition over time. In: Proc. 17th Internat. Conf. on Pattern Recognition (ICPR04).

Tanaka, H.T., Ikeda, M., Chiaki, H., 1998. Curvature-based face surface recognition using spherical correlation – principal directions for curved object recognition. In: Proc. 3rd Internat. Conf. on Face & Gesture Recognition (FG'98), Nara, Japan, April, pp. 372–377.

Tan, T., Yan, H., 1999. Face recognition by fractal transformations. In: Proc. 1999 IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing, 1999. ICASSP'99, Vol. 6 (March), pp. 3537–3540.

The 3D Face Database, 2003. The University of York. <http://www-users.cs.york.ac.uk/~nep/tomh/>.

Tsalakanidou, F., Tzovaras, D., Strintzis, M.G., 2003. Use of depth and colour Eigenfaces for face recognition. Pattern Recognition Lett. 24 (9–10), 1427–1435.

Wang, Y., Pan, G., Wu, Z., 2004. Sphere-spin-image: A viewpoint-invariant surface representation for 3D face recognition. In: Proc. Internat. Conf. on Computational Science (ICCS'04), Lecture Notes in Computer Science, Vol. 3037 (June), pp. 427–434.

Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C., 1997. Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Anal. Machine Intell. 19 (July), 775–779.

Xu, C., Wang, Y., Tan, T., Quan, L., 2004. Depth vs. intensity: Which is more important for face recognition? In: Proc. 17th Internat. Conf. on Pattern Recognition (ICPR 2004), Vol. 4 (August), pp. 342–345.

Yu, H., Yang, J., 2001. A direct LDA algorithm for high-dimensional data with application to face recognition. Pattern Recognition 34 (October), 2067–2070.