AN INTRODUCTION TO BIOMETRICS AND FACE RECOGNITION

F. PERRONNIN*AND J.-L. DUGELAY

Eurecom Institute
Multimedia Communications Department
2229, route des Crêtes - B. P. 193
06904 Sophia-Antipolis cédex - France
E-mail: {perronni,dugelay}@eurecom.fr

We present in this paper a brief introduction to *biometrics* which refers to the problem of identifying a person based on his/her *physical* or *behavioral* characteristics. We will also provide a short review of the literature on face recognition with a special emphasis on frontal face recognition, which represents the bulk of the published work in this field. While biometrics have mostly been studied separately, we also briefly introduce the notion of *multimodality*, a topic related to decision fusion and which has recently gained interest in the biometric community.

1. Introduction to Biometrics

The ability to verify automatically and with great accuracy the identity of a person has become crucial in our society. Even though we may not notice it, our identity is challenged daily when we use our credit card or try to gain access to a facility or a network for instance. The two traditional approaches to automatic person identification, namely the *knowledge-based* approach which relies on something that you know such as a password, and the *token-based* approach which relies on something that you have such as a badge, have obvious shortcomings: passwords might be forgotten or guessed by a malicious person while badges might be lost or stolen [1].

Biometrics person recognition, which deals with the problem of identifying a person based on his/her *physical* or *behavioral* characteristics, is an alternative to these traditional approaches as a biometric attribute is inherent to each person and thus cannot be forgotten or lost and might be difficult to forge. The face, the fingerprint, the hand geometry, the iris,

*This work was supported in part by France Telecom Research

etc. are examples of physical characteristics while the signature, the gait, the keystroke, etc. are examples of behavioral characteristics. It should be underlined that a biometric such as the voice is both physical and behavioral. Ideally a biometric should have the following properties: it should be *universal*, *unique*, *permanent* and easily *collectible* [2].

In the next three sections of this introductory part, we will briefly describe the architecture of a typical biometric system, the measures to evaluate its performance and the possible applications of biometrics.

1.1. *Architecture*

A biometric system is a particular case of a *pattern recognition* system [3]. Given a set of *observations* (captures of a given biometric) and a set of possible *classes* (for instance the set of persons that can be possibly identified) the goal is to associate to each observation one unique class. Hence, the main task of pattern recognition is to distinguish between the *intra-class* and *inter-class* variabilities. Face recognition, which is the main focus of this article, is a very challenging problem as faces of the same person are subject to variations due to facial expressions, pose, illumination conditions, presence/absence of glasses and facial hair, aging, etc.

A biometric system is composed of at least two mandatory modules, the *enrollment* and *recognition* modules, and an optional one, the *adaptation* module. During enrollment, the biometric is first measured through a *sensing* device. Generally, before the *feature extraction* step, a series of pre-processing operations, such as detection, segmentation, etc. should be applied. The extracted features should be a compact but accurate representation of the biometric. Based on these features, a model is built and stored, for instance in a database or on a smart card. During the *recognition* phase, the biometric characteristic is measured and features are extracted as done during the enrollment phase. These features are then compared with one or many models stored in the database, depending on the operational mode (see the next section on performance evaluation). During the enrollment phase, a user friendly system generally captures only a few instances of the biometric which may be insufficient to describe with great accuracy the characteristics of this attribute. Moreover, this biometric can vary over time in the case where it is non-permanent (e.g. face, voice). Adaptation maintains or even improves the performance of the system over time by updating the model after each access to the system.

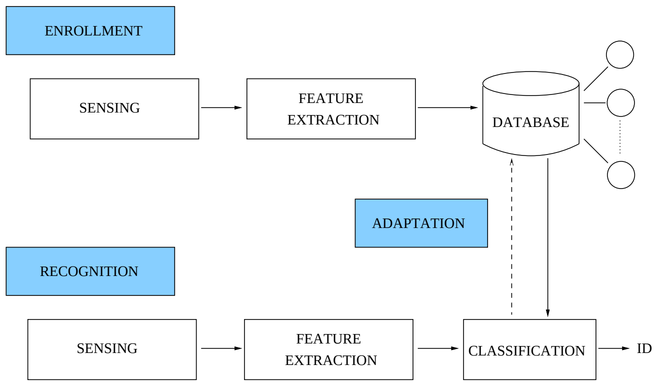


Figure 1. Architecture of a biometric system.

1.2. *Performance Evaluation*

Generally, a biometric system can work under two different operational modes: *identification* or *verification*. During identification, the system should guess the identity of person among a set of N possible identities (1:N problem). A *close-set* is generally assumed, which means that all the trials will be from people which have a model in the database and the goal is hence to find the most likely person. During verification, the user claims an identity and the system should compare this identity with the stored model (1:1 problem). This is referred as an *open-set* as persons which are not in the database may try to fool the system. One can sometimes read claims that identification is a more challenging problem than verification or vice-versa. Actually, identification and verification are simply two different problems.

As it may not be enough to know whether the top match is the correct one for an identification system, one can measure its performance through the *cumulative match score* which measures the percentage of correct answers among the top N matches. Also one could use *recall-precision* curves as is done for instance to measure the performance of database retrieval systems. The FERET face database [4] is the most commonly used database for assessing the performance of a system in the identification mode.

A verification system can make two kinds of mistakes: it can reject a rightful user, often called *client*, or accept an *impostor*. Hence, the performance of a verification system is measured in terms of its *false rejection rate (FRR)* and *false acceptance rate (FAR)*. A threshold is set to the scores obtained during the verification phase and one can vary this threshold to obtain the best possible compromise for a particular application depending on the required security level. By varying this threshold, one obtains the *receiver operating curve (ROC)*, i.e. the FRR as a function of the FAR. To summarize the performance of the system with one unique figure, one often uses the *equal error rate (EER)* which corresponds to the point FAR=FRR. The M2VTS database and its extension, the XM2VTSDB [5], are the most commonly used databases for assessing the performance of a system in the verification mode.

The interested reader can also refer to [6] for an introduction to evaluating biometric systems.

1.3. *Applications*

There are mainly four areas of applications for biometrics: *access control*, *transaction authentication*, *law enforcement* and *personalization*.

Access control can be subdivided into two categories: *physical* and *virtual* access control [1]. The former controls the access to a secured location. An example is the Immigration and Naturalization Service's Passenger Accelerated Service System (INSPASS) deployed in major US airports which enables frequent travelers to use an automated immigration system that authenticates their identity through their hand geometry. The latter one enables the access to a resource or a service such as a computer or a network. An example of such a system is the voice recognition system used in the MAC OS 9.

Transaction authentication represents a huge market as it includes transactions at an automatic teller machine (ATM), electronic fund transfers, credit card and smart card transactions, transactions on the phone or on the Internet, etc. Mastercard estimates that a smart credit card incorporating finger verification could eliminate 80% of fraudulent charges [8]. For transactions on the phone, biometric systems have already been deployed. For instance, the speaker recognition technology of Nuance [9] is used by the clients of the Home Shopping Network or Charles Schwab.

Law enforcement has been one of the first applications of biometrics. Fingerprint recognition has been accepted for more than a century as a means of identifying a person. Automatic face recognition can also be very useful for searching through large mugshot databases.

Finally, personalization through person authentication is very appealing in the consumer product area. For instance, Siemens allows to personalize one's vehicle accessories, such as mirrors, radio station selections, seating positions, etc. through fingerprint recognition [10].

In the following subsections, we will provide to the reader a brief review of the literature on face recognition. This review will be split into two parts: we will devote the next section to frontal face recognition which represents the bulk of the literature on and the "other modalities", corresponding to different acquisition scenarios such as profile, range images, facial thermogram or video, will be discussed in section 3. The interested reader can refer to [11] for a full review of the literature on face recognition before 1995. We should underline that specific parts of the face (or the head) such as the eyes, the ears, the lips, etc. contain a lot of relevant information for identifying people. However, this is out of the scope of this paper and the interested reader can refer to [12] for iris recognition, to [13] for ear recognition and [14] for lips dynamics recognition. Also we will not review a very important part of any face recognition system: the *face detection*. For a recent review on the topic, the reader can refer to [15].

2. Frontal Face Recognition

It should be underlined that the expression "frontal face recognition" is used in opposition to "profile recognition". A face recognition system that would work only under perfect frontal conditions would be of limited interest and even "frontal" algorithms should have some view tolerance. As a full review, even of the restricted topic of frontal face recognition, is out of the scope of this paper, we will focus our attention on two very successful classes of algorithms: the *projection-based* approaches, i.e. the Eigenfaces and its related approaches, and the ones based on deformable models such as *Elastic Graph Matching*. It should be underlined that the three top performers at the 96 FERET performance evaluation belong to one of these two classes [4].

2.1. *Eigenfaces and Related Approaches*

In this section, we will first review the basic eigenface algorithm and then consider its extensions: multiple spaces, eigenfeatures, linear discriminant analysis and probabilistic matching.

2.1.1. *Eigenfaces*

Eigenfaces are based on the notion of dimensionality reduction. [16] first outlined that the dimensionality of the face space, i.e. the space of variation between images of human faces, is much smaller than the dimensionality of a single face considered as an arbitrary image. As a useful approximation, one may consider an individual face image to be a linear combination of a small number of face components or *eigenfaces* derived from a set of reference face images. The idea of the *Principal Component Analysis* (PCA) [17], also known as the *Karhunen-Loeve Transform* (KLT), is to find the subspace which best accounts for the distribution of face images within the whole space.

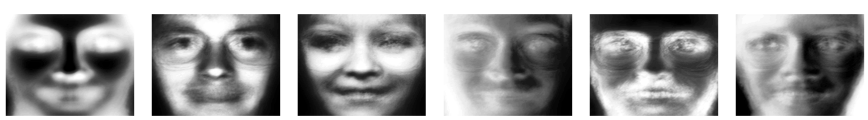
Let $\{O_i\}_{i\in[1,N]}$ be the set of reference or training faces, $\bar{O}$ be the average face and $\hat{O}_i = O_i - \bar{O}$. $\hat{O}_i$ is sometimes called a *caricature* image. Finally, if $\hat{\mathbf{O}} = [\hat{O}_1, \hat{O}_2, ...\hat{O}_N]$, the *scatter* matrix S is defined as:

$$S = \sum_{i=1}^{N} \hat{O}_i \hat{O}_i^T = \hat{\mathbf{O}}\hat{\mathbf{O}}^T \tag{1}$$

The optimal subspace P_{PCA} is chosen to maximize the scatter of the projected faces:

$$P_{PCA} = \arg\max_P |PSP^T| \tag{2}$$

where $|.|$ is the determinant operator. The solution to problem (2) is the subspace spanned by the eigenvectors $[e_1, e_2, ...e_K]$, also called *eigenfaces*, corresponding to the K largest eigenvalues of the scatter matrix S. It should be underlined that eigenfaces are not themselves usually plausible faces but only directions of variation between face images (see Figure 2). Each face image is represented by a point $P_{PCA} \times \hat{O}_i = [w_i^1, w_i^2, ...w_i^K]$ in



(a) (b) (c) (d) (e) (f)

Figure 2. (a) Eigenface 0 (average face) and (b)-(f) eigenfaces 1 to 5 as estimated on a subset of the FERET face database.

the K-dimensional space. The weights w_i^k's are the projection of the face image on the $k-th$ eigenface e_k and thus represent the contribution of each eigenface to the input face image.

To find the best match for an image of a person's face in a set of stored facial images, one may calculate the Euclidean distances between the vector representing the new face and each of the vectors representing the stored faces, and then choose the image yielding the smallest distance [18].

2.1.2. *Multiple Spaces Approaches*

When one has a large amount of training data, one can either pool all the data to train one unique eigenspace, which is known as the *parametric* approach or split the data into multiple training sets and train multiple eigenspaces which is known as the *view-based* approach. The latter approach has been designed especially to compensate for different head poses.

One of the first attempts to train multiple eigenspaces was made in [19]. This method, consists in building a separate eigenspace for each possible view [19]. For each new target image, its orientation is first estimated by projecting it on each eigenspace and choosing the one that yields the smallest distance from face to space. The performance of the parametric and view-based approaches were compared in [19] and the latter one seems to perform better. The problem with the *view-based approach* is that it requires large amounts of *labeled* training data to train each separate eigenspace.

More recently *Mixtures of Principal Components* (MPC) were proposed to extend the traditional PCA [20,21]. An iterative procedure based on the *Expectation-Maximization* algorithm was derived in both cases to train automatically the MPC. However, while [20] represents a face by the best set of features corresponding to the closest set of eigenfaces, in [21] a face image is projected on each component eigenspace and these individual projections are then linearly combined. Hence, compared to the former approach, a face image is not assigned in a *hard* manner to one eigenspace component but in a *soft* manner to all the eigenspace components. [21] tested MPC on a database of face images that exhibit large variabilities in poses and illumination conditions. Each eigenspace converges automatically to varying poses and the first few eigenvectors of each component eigenspace seem to capture lightning variations.

2.1.3. *Eigenfeatures*

An eigenface-based recognition system can be easily fooled by gross variations of the image such as the presence or absence of facial hair [19]. This shortcoming is inherent to the eigenface approach which encodes a *global* representation of the face. To address this issue, [19] proposed a *modular* or *layered* approach where the global representation of the face is augmented by *local* prominent features such as the eyes, the nose or the mouth. Such an approach is of particular interest when a part of the face is occluded and only a subset of the facial features can be used for recognition. A similar approach was also developed in [22]. The main difference is in the encoding of the features: the notion of eigenface is extended to eigeneyes, eigennose and eigenmouth as was done for instance in [23] for image coding. For a small number of eigenvectors, the eigenfeatures approach outperformed the eigenface approach and the combination of eigenfaces and eigenfeatures outperformed each algorithm taken separately.

2.1.4. *Linear Discriminant Approaches*

While PCA is optimal with respect to data compression [16], in general it is sub-optimal for a recognition task. Actually, PCA confounds *intra-personal* and *extra-personal* sources of variability in the total scatter matrix S. Thus eigenfaces can be contaminated by non-pertinent information.

For a classification task, a dimension reduction technique such as *Linear Discriminant Analysis* (LDA) should be preferred to PCA [24,25,26]. The idea of LDA is to select a subspace that maximizes the ratio of the inter-class variability and the intra-class variability. Whereas PCA is an *unsupervised* feature extraction method, discriminant analysis uses the category information associated with each training observation and is thus categorized as *supervised*.

Let $O_{i,k}$ be the k-th picture of training person i, N_i be the number of training images for person i and $\bar{O}_i$ be the average of person i. Then S_B and S_w, respectively the *between-* and *within-class* scatter matrices, are given by:

$$S_B = \sum_{i=1}^{c} N_i(\bar{O}_i - \bar{O})(\bar{O}_i - \bar{O})^T \quad (3)$$

$$S_W = \sum_{i=1}^{c} \sum_{k=1}^{N_i} (O_{i,k} - \bar{O}_i)(O_{i,k} - \bar{O}_i)^T \quad (4)$$

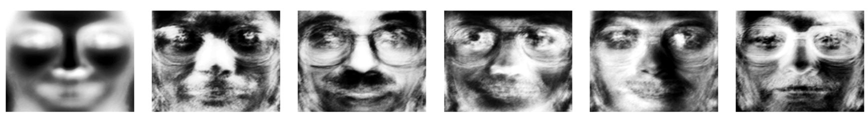
The optimal subspace P_{LDA} is chosen to maximize the between-scatter of the projected face images while minimizing the within-scatter of the projected faces:

$$P_{LDA} = \arg\max_P \frac{|PS_BP^T|}{|PS_WP^T|} \quad (5)$$

The solution to equation (5) is the sub-space spanned by $[e_1, e_2, ...e_K]$, the generalized eigenvectors corresponding to the largest eigenvalues of the generalized eigenvalue problem:

$$S_B e_k = \lambda_k S_W e_k \qquad \text{k= 1, ... K} \tag{6}$$

However, due to the high dimensionality of the feature space, S_W is generally singular and this principle cannot be applied in a straightforward manner. To overcome this issue, generally one first applies PCA to reduce the dimension of the feature space and then performs the standard LDA [24,26]. The eigenvectors that form the discriminant subspace are often referred as *Fisherfaces* [24]. In [26], the space spanned by the first few Fisherfaces are called the *most discriminant features* (MDF) classification space while PCA features are referred as *most expressive features* (MEF). It should be



(a) (b) (c) (d) (e) (f)

Figure 3. (a) Fisherface 0 (average face) and (b)-(f) Fisherfaces 1 to 5 as estimated on a subset of the FERET face database.

underlined that LDA induces non-orthogonal projection axes, a property which has great relevance in biological sensory systems [27].

Other solutions to equation 5 were suggested [27,28,29].

2.1.5. *Probabilistic Matching*

While most face recognition algorithms, especially those based on eigenfaces, generally use simple metrics such as the Euclidean distance, [30] suggests a *probabilistic similarity* based on a discriminative *Bayesian* analysis of image differences. One considers the two mutually exclusives classes of variation between two facial images: the *intra-personal* and *extra-personal* variations, whose associated spaces are noted respectively Ω_I and Ω_E. Given two face images O_1 and O_2 and the image difference $\Delta = O_1 - O_2$, the similarity measure is given by $P(\Omega_I|\Delta)$. Using Baye's rule, it can be transformed into:

$$P(\Omega_I|\Delta) = \frac{P(\Delta|\Omega_I)P(\Omega_I)}{P(\Delta|\Omega_I)P(\Omega_I) + P(\Delta|\Omega_E)P(\Omega_E)} \tag{7}$$

The high-dimensionality probability functions $P(\Delta|\Omega_I)$ and $P(\Delta|\Omega_E)$ are estimated using an eigenspace density estimation technique [31]. It was observed that the denominator in equation (7) had a limited impact on the performance of the system and that the similarity measure could be reduced to $P(\Delta|\Omega_I)$ with little loss in performance, thus reducing the computational requirements of the algorithm by a factor two.

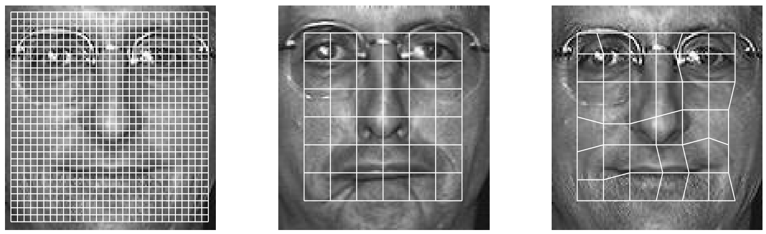
2.2. *Deformable Models*

As noted in [32], since most face recognition algorithms are minimum distance pattern classifiers, a special attention should be paid to the definition of *distance*. The distance which is generally used is the Euclidean distance. While it is easy to compute, it may not be optimal as, for instance, it does not compensate for the deformations incurred from different facial expressions. Face recognition algorithms based on deformable models can cop with this kind of variation.

2.2.1. *Elastic Graph Matching*

Elastic Graph Matching algorithm (EGM) has roots in the neural network community [33].

Given a template image $\mathcal{F}_T$, one first derives a face model from this image. A grid is placed on the face image and the face model is a *vector field* $O = \{o_{i,j}\}$ where $o_{i,j}$ is the feature vector extracted at position (i,j) of the grid which summarizes local properties of the face (c.f. Figure 4(a). Gabor coefficients are generally used but other features, like morphological feature vectors, have also been considered and successfully applied to the EGM problem [34]. Given a query image $\mathcal{F}_Q$, one also derives a vector field $X = \{x_{i,j}\}$ but on a coarser grid than the template face (c.f. Figure 4(b)). In the EGM approach, the distance between the template and query images is defined as a *best mapping* $\mathcal{M}^*$ among the set of all possible mappings $\{\mathcal{M}\}$ between the two vector fields O and X. The optimal mapping depends on the definition of the cost function $\mathcal{C}$. Such a function should keep a proper balance between the local matching of features and the requirement to preserve spatial distance. Therefore, a proper cost function should be of



(a) (b) (c)

Figure 4. (a) Template image and (b) query image with their associated grids. (c) Grid after deformation using the probabilistic deformable model of face mapping (c.f. section 2.2.3). Images extracted from the FERET face database.

the form:

$$\mathcal{C}(\mathcal{M}) = \mathcal{C}_v(\mathcal{M}) + \rho\mathcal{C}_e(\mathcal{M}) \tag{8}$$

where $\mathcal{C}_v$ is the cost of local matchings, $\mathcal{C}_e$ the cost of local deformations and ρ is a parameter which controls the *rigidity* of the elastic matching and has to be hand-tuned.

As the number of possible mappings is extremely large, even for lattices of moderate size, an exhaustive search is out of the question and an approximate solution has to be found. Toward this end, a two steps procedure was designed:

- *rigid matching*: the whole template graph is shifted around the query graph. This corresponds to $\rho \to \infty$. We obtain an initial mapping $\mathcal{M}_0$.
- *deformable matching*: the nodes of the template lattice are then stretched through random local perturbations to reduce further the cost function until the process converges to a locally optimal mapping $\mathcal{M}^*$, i.e. once a predefined number of trials have failed to improve the mapping cost.

The previous matching algorithm was later improved. For instance, in [34] the authors argue that the two-stage coarse-to-fine optimization is sub-optimal as the deformable matching relies too much on the success of the rigid matching. The two stage optimization procedure is replaced with a probabilistic hill-climbing algorithm which attempts to find at each iteration both the optimal global translation and the set of optimal local perturbations. In [35], the same authors further drop the $\mathcal{C}_e$ term in equation (8). However, to avoid unreasonable deformations, local translations are restricted to a neighborhood.

2.2.2. *Elastic Bunch Graph Matching*

[36] elaborated on the basic idea of EGM with the Elastic Bunch Graph Matching (EBGM) through three major extensions:

- While the cost of local matchings in $\mathcal{C}_e$ only makes use of the magnitude of the complex Gabor coefficients in the EGM approach, the *phase information* is used to disambiguate features which have a similar magnitude, but also to estimate local distortions.
- The features are no longer extracted on a rectangular graph but they now refer to specific facial landmarks called *fiducial points*.
- A new data structure called *bunch graph* which serves as a general representation of the face is introduced. Such a structure is obtained by combining the graphs of a set of reference individuals.

It should be noted that the idea of extracting features at positions which correspond to facial landmarks appeared in earlier work. In [37] feature points are detected using a Gabor wavelet decomposition. Typically, 35 to 50 points are obtained in this manner and form the face graph. To compare two face graphs, a two-stage matching similar to the one suggested in [33] is developed. One first compensates for a global translation of the graphs and then performs local deformations for further optimization. However, another difference with [33] is that the cost of local deformations (also called *topology cost*) is only computed after the features are matched which results in a very fast algorithm. One advantage of [36] over [37] is in the use of the bunch graph which provides a *supervised* way to extract salient features.

An obvious shortcoming of EGM and EBGM is that $\mathcal{C}_v$, the cost of local matchings, is simply a sum of all local matchings. This contradicts the fact that certain parts of the face contain more discriminant information and that this distribution of the information across the face may vary from one person to another. Hence, the cost of local matchings at each node should be weighted according to their discriminatory power [38,39,34,35].

2.2.3. *Probabilistic Deformable Model of Face Mapping*

A novel probabilistic deformable model of face mapping [40], whose philosophy is similar to EGM [33], was recently introduced. Given a template face $\mathcal{F}_T$, a query face $\mathcal{F}_Q$ and a deformable model of the face $\mathcal{M}$, for a face identification task the goal is to estimate $P(\mathcal{F}_T|\mathcal{F}_Q,\mathcal{M})$. The two major differences between EGM and the approach presented in [40] are:

- In the use of the HMM framework which provides efficient formulas to compute $P(\mathcal{F}_T|\mathcal{F}_Q,\mathcal{M})$ and train automatically all the parameters of $\mathcal{M}$. This enables for instance to model the elastic properties of the different parts of the face.
- In the use of a shared deformable model of the face $\mathcal{M}$ for all individuals, which is particularly useful when little enrollment data is available.

3. Other "Modalities" for Face Recognition

In this section we will very briefly review what we called the "other modalities" and which basically encompass the remaining of the literature on face recognition: *profile recognition*, recognition based on *range data*, *thermal imagery* and finally *video-based* face recognition.

3.1. *Profile Recognition*

The research on profile face recognition has been mainly motivated by requirements of law enforcement agencies with their so-called *mug shot* databases [11]. However, it has been the focus of a relatively restricted number of papers. It should be underlined that frontal and profile face recognition are complementary as they do not provide the same information. A typical profile recognition algorithm first locates on the contour image points of interest such as the nose tip, the mouth, chin, etc. also called *fiducial points* and then extracts information such as the distances, angles, etc. for the matching (see [41] for an example of an automatic system based on this principle). An obvious problem with such an approach is the fact that it relies on an accurate feature extraction. Alternative approaches which alleviate this problem include (but are not limited to) the use of *Fourier descriptors* for the description of closed curves [42], the application of Eigenfaces to profiles [19] and, more recently, an algorithm based on string matching [43].

3.2. *Range Data*

While a 2-D intensity image does not have direct access to the 3-D structure of an object, a range image contains the depth information and is not sensitive to lightning conditions (it can even work in the dark) which makes range data appealing for a face recognition system. The sensing device can be a *rotating laser scanner* which provides a very accurate and complete representation of the face as used for instance in [44,45]. However, such a scanner is highly expensive and the scanning process is very slow. In [46] the authors suggested the use the *coded light* approach for acquiring range images. A sequence of stripe patterns is projected onto the face and for each projection an image is taken with a camera. However, for shadow regions as well as regions that do not reflect the projected light, no 3-D data can be estimated which results in range images with a lot of missing data. Therefore, the authors decided to switch to a *multi-sensor system* with two range sensors acquiring the face under two different views. These two sets of range data are then merged. Although these sensing approaches reduce both the acquisition time and cost, the user of such a system should be cooperative which restricts its use. This may explain the fact that little literature is available on this topic.

In [44], the authors present a face recognition system based on range data *template matching*. The range data is segmented into four surface regions which are then normalized using the location of the eyes, nose and mouth. The volume between two surfaces is used as distance measure. In [45] the face recognition system uses *features* extracted from range and curvature data. Examples of features are the left and right eye width, the head width, etc. but also the maximum Gaussian curvature on the nose ridge, the average minimum curvature on the nose ridge, etc. In [46], the authors apply and extend traditional 2-D face recognition algorithms (Eigenfaces and HMM-based face recognition [47]) to range data. More recently, [48] point signatures are used as features for 3-D face recognition. These feature points are projected into a subspace using PCA.

3.3. *Facial Thermogram*

The facial heat emission patterns can be used to characterize a person. These patterns depend on nine factors including the location of major blood vessels, the skeleton thickness, the amount of tissue, muscle, and fat [49]. IR face images have the potential for a good biometric as this signatures is unique (even identical twins do not share the same facial thermogram) and it is supposed to be relatively stable over time. Moreover, it cannot be altered through plastic surgery. The acquisition is done with an *infrared (IR)* camera. Hence, it does not depend on the lightning conditions, which is a great advantage over traditional facial recognition. However, IR imagery is dependent on the temperature and IR is opaque to glass. A preliminary study [50] compared the performance of visible and IR imagery for face recognition and it was shown that there was little difference in performance. However, the authors in [50] did not address the issue of significant variations in illumination for visible images and changes in temperature for IR images.

3.4. *Video-Based Recognition*

Although it has not been a very active research topic (at least compared to frontal face recognition), video-based face recognition can offer many advantages compared to recognition based on still images:

- *Abundant data* is available at both enrollment and test time. Actually one could use video at enrollment time and still images at test time or vice versa (although the latter scenario would perhaps make less sense). However, it might not be necessary to process all this data and one of the tasks of the recognition system will be the selection of an optimal subset of the whole set of images which contains the maximum amount of information.
- With sequences of images, the recognition system has access to *dynamic features* which provides valuable information on the *behavior* of the user. For instance, the BioID system [14] makes use of the *lip movement* for the purpose of person identification (in conjunction with face and voice recognition). Also dynamic features are generally more secure against fraud than static features as they are harder to replicate.
- Finally the system can try to build a model of the face by estimating th 3-D depth of points on objects from a sequence of 2-D images which is known as *structure from motion* [11].

Video-based recognition might be extremely useful for *covert surveillance*, for instance in airports. However, this is a highly challenging problem as the system should work in a non-cooperative scenario and the quality of surveillance video is generally poor and the resolution is low.

4. Multimodality

Reliable biometric-based person authentication systems, based for instance on iris or retina recognition already exist but the user acceptance for such systems is generally low and they should be used only in high security scenarios. Systems based on voice or face recognition generally have a high user acceptance but their performance is not satisfying enough.

Multimodality is a way to improve the performance of a system by combining different biometrics. However, one should be extremely careful about *which modalities* should be combined (especially, it might not be useful to combine systems which have radically different performances) and *how to combine* them. In the following, we will briefly describe the possible multimodality scenarios and the different ways to fuse the information.

4.1. *Different Multimodality Scenarios*

We use here the exhaustive classification introduced in [51]:

(1) *multiple biometric systems*: consists in using different biometric attributes, such as the face, voice and lip movement [14]. This is the most commonly used sense of the term multimodality.
(2) *multiple sensors*: e.g. a camera and an infrared camera for face recognition.
(3) *multiple units of the same biometric*: e.g. fusing the result of the recognition of both irises.
(4) *multiple instances of the same biometric*: e.g. in video-based face recognition, fusing the recognition results of each image.
(5) *multiple algorithms* on the same biometric capture.

We can compare these scenarios in terms of the expected increase of performance of the system over the monomodal systems versus the increase of the cost of the system, which can be split into additional software and hardware costs.

In terms of the additional amount of information and thus in the expected increase of the performance of the system, the first scenario is the richest and scenarios (4) and (5) are the poorest ones. The amount of information brought by scenario (2) is highly dependent on the difference between the two sensors. Scenario (3) can bring a large amount of information as, for instance, the two irises or the ten fingerprints of the same person are different. However, if the quality of a fingerprint is low for a person, e.g. because of a manual activity, then the quality of the other fingerprints is likely to be low.

The first two scenarios clearly introduce an additional cost as many sensors are necessary to perform the acquisitions. For scenario (3) there is no need for an extra sensor if captures are done sequentially. However, this lengthens the acquisition time which makes the system less user-friendly. Finally, scenarios (1) and (5) induce an additional software cost as different algorithm are necessary for the different systems.

4.2. *Information Fusion*

As stated at the beginning of this section, multimodality improves the performance of a biometric system. The word performance includes both *accuracy* and *efficiency*.

The assumption which is made is that different biometric systems make different types of errors and thus, that it is possible to use the complementary nature of these systems. This is a traditional problem of *decision fusion* [53]. Fusion can be done at three different levels [52] (by increasing order of available information):

- At the *abstract* level, the output of each classifier is a label such as the ID of the most likely person in the identification case or a binary answer such as accept/reject in the verification case.
- At the *rank* level the output labels are sorted by confidence.
- At the *measurement* level, a confidence measure is associated to each label.

Commonly used classification schemes such as the product rule, sum rule, min rule, max rule and median rule, are derived from a common theoretical framework using different approximations [54]. In [55], the authors evaluated different classification schemes, namely *support vector machine* (SVM), *multi layer perceptron* (MLP), *decision tree*, *Fisher's linear discriminant* (FLD) and *Bayesian classifier* and showed that the SVM- and Bayesian-based classifiers had a similar performance and outperformed the other classifiers when fusing face and voice biometrics.

In the identification mode, one can use the complementary nature of different biometrics to speed-up the search process. Identification is generally performed in a *sequential* mode. For instance, in [56] identification is a two-step process: face recognition, which is fast but unreliable is used to obtain an N-best list of the most likely persons and fingerprint recognition, which is slower but more accurate, is then performed on this subset.

5. Summary

We introduced in this paper biometrics, which deals with the problem of identifying a person based on his/her physical and behavioral characteristics. Face recognition, which is one of the most actively research topic in biometrics, was briefly reviewed. Although huge progresses have been made in this field for the past twenty years, research has mainly focused on frontal face recognition from still images. We also introduced the notion of multimodality as a way of exploiting the complementary nature of monomodal biometric systems.

References

1. S. Liu and M. Silverman, "A practical guide to biometric security technology", *IT Professional"*, vol. 3, no. 1, pp. 27-32, Jan/Feb 2001.
2. A. Jain, R. Bolle and S. Pankanti, "Biometrics personal identification in networked society", Boston, MA: Luwer Academic, 1999.
3. R. O. Duda, P. E. Hart and D. G. Stork, "Pattern classification", 2nd edition, John Wiley & Sons, Inc.
4. P. J. Phillips, H. Moon, S. Rizvi and P. Rauss, "The FERET evaluation methodology for face recognition algorithms", *IEEE Trans. on PAMI*, 2000, vol. 22, no. 10, October.
5. K. Messer, J. Matas, J. Kittler and K. Jonsson, "XM2VTSDB: the extended M2VTS database", AVBPA'99, 1999, pp. 72-77.
6. P. J. Phillips, A. Martin, C. L. Wilson and M. Przybocki, "An introduction to evaluating biometric systems", *Computer*, 2000, vol. 33, no. 2, pp. 56-63.
7. INSPASS, http://www.immigration.gov/graphics/howdoi/inspass.htm
8. O. O'Sullivan, "Biometrics comes to life", Banking Journal, 1997, January.
9. Nuance, http://www.nuance.com
10. Siemens Automotive, http://media.siemensauto.com
11. R. Chellappa, C. L. Wilson and S. Sirohey, "Human and machine recognition of faces: a survey", Proc. of the IEEE, 1995, vol. 83, no. 5, May.
12. J. Daugman, "How iris recognition works" ICIP, 2002, vol. 1, pp. 33-36.
13. B. Moreno, A. Sanchez and J. F. Velez, "On the use of outer ear images for personal identification in security applications", IEEE 3rd Conf. on Security Technology, pp. 469-476.
14. R. W. Frischholz and U. Dieckmann, "BioID: a multimodal biometric identification system", *Computer*, 2000, vol. 33, no. 2, pp. 64-68, Feb.
15. E. Hjelmas and B. K. Low, "Face detection: a survey", *Computer Vision and Image Understanding*, 2001, vol. 83, pp. 236-274.
16. M. Kirby and L. Sirovich, "Application of the karhunen-loève procedure for the characterization of human faces," *IEEE Trans. on PAMI*, vol. 12, pp. 103-108, 1990.
17. I. T. Joliffe, "Principal Component Analysis", Springer-Verlag, 1986.
18. M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *IEEE*

Conf. on CVPR, 1991, pp. 586-591.
19. A. Pentland, B. Moghaddam and T. Starner, "View-based and modular eigenspaces for face recognition," *IEEE Conf. on CVPR*, pp. 84-91, June 1994.
20. H.-C. Kim, D. Kim and S. Y. Bang, "Face recognition using the mixture-of-eigenfaces method," *Pattern Recognition Letters*, vol. 23, no. 13, pp. 1549-1558, Nov. 2002.
21. D. S. Turaga and T. Chen, "Face recognition using mixtures of principal components," *IEEE Int. Conf. on IP*, vol. 2, pp. 101-104, 2002.
22. R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates", *IEEE Trans. on PAMI*, 1993, vol. 15, no. 10, pp. 1042-1052, Oct.
23. W. J. Welsh and D. Shah, "Facial feature image coding using principal components," *Electronic Letters*, vol. 28, no. 22, pp. 2066-2067, October 1992.
24. P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transaction on PAMI*, vol. 19, pp. 711-720, Jul 1997.
25. K. Etermad and R. Chellappa, "Face recognition using discriminant eigenvectors," *ICASSP*, vol. 4, pp. 2148-2151, May 1996.
26. D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. on PAMI*, vol. 18, no. 8, pp. 831-836, August 1996.
27. C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. on IP*, vol. 11, no. 4, pp. 467-476, Apr 2002.
28. L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin and G.-J. Yu, "A new lda-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol. 33, no. 10, pp. 1713-1726, October 2000.
29. J. Yang and J.-Y. Yang, "Why can lda be performed in pca transformed space?," *Pattern Recognition*, vol. 36, no. 2, pp. 563-566, February 2003.
30. B. Moghaddam, W. Wahid and A. Pentland, "Beyond eigenfaces: Probabilistic matching for face recognition," *IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 30-35, April 1998.
31. B. Moghaddam and A. Pentland, "Probabilistic visual learning for object recognition," *Int. Conf. on Computer Vision*, 1995.
32. J. Zhang, Y. Yan and M. Lades, "Face recognition: Eigenface, elastic matching, and neural nets," *Proc. of the IEEE*, vol. 85, no. 9, Sep 1997.
33. M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburd, R. Würtz and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. on Computers*, 1993, vol. 42, no. 3.
34. C. L. Kotropoulos, A. Tefas and I. Pitas, "Frontal face authentication using discriminant grids with morphological feature vectors," *IEEE Trans. on Multimedia*, vol. 2, no. 1, pp. 14–26, March 2000.
35. A. Tefas, C. Kotropoulos and I. Pitas, "Using support vector machines to enhance the performance of elastic graph matching for frontal face recognition," *IEEE Trans. on PAMI*, vol. 23, no. 7, pp. 735–746, Jul 2001.
36. L. Wiskott, J. M. Fellous, N. Krüger and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. on PAMI*, vol. 19, no.

7, pp. 775–779, July 1997.
37. B. S. Manjunath, R. Chellappa and C. von der Malsburg, "A feature based approach to face recognition," *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 373–378, 1992.
38. N. Krüger, "An algorithm for the learning of weights in discrimination functions using a priori constraints," *IEEE Trans. on PAMI*, vol. 19, no. 7, Jul 1997.
39. B. Dûc, S. Fischer and J. Bigün, "Face authentication with gabor information on deformable graphs," *IEEE Trans. on IP*, vol. 8, no. 4, Apr 1999.
40. F. Perronnin, Jean-Luc Dugelay and K. Rose, "Deformable Face Mapping for Person Identification", *ICIP*, 2003.
41. C. Wu and J. Huang, "Human Face profile recognition by computer", *Pattern Recognition*, vol. 23, pp. 255-259, 1990.
42. T. Aibara, K. Ohue and Y. Matsuoka, "Human face recognition of P-type Fourier descriptors", *SPIE Proc.*, vol 1606: Visual Communication and Image Processing, 1991, pp. 198-203.
43. Y. Gao, M. Leung, "Human face profile recognition using attributed string", *Pattern Recognition*, vol. 35, pp. 353-360.
44. G. Gordon, "Face recognition based on depth maps and surface curvature", *SPIE Proc.*, vol. 1570, pp. 234-247, 1991.
45. G. Gordon, "Face recognition based on depth and curvature features", *IEEE Conf. on CVPR*, 1992, pp. 808-810, 15-18 Jun.
46. B. Achermann, X. Jiang and H. Bunke, "Face recognition using range images", *VSMM*, 1997, pp. 129-136, 10-12 Sep.
47. F. S. Samaria, "Face recognition using hidden Markov models", *Ph.D. thesis*, University of Cambridge, 1994.
48. Y. Wang, C.-S. Chua and Y.-K. Ho, "Facial feature detection and face recognition from 2D and 3D images", Pattern Recognition Letters, 2002, vols. 23, pp. 1191-1202.
49. M. Lawlor, "Thermal pattern recognition systems faces security challenges head on", *Signal Magazine*, 1997, November.
50. J. Wilder, P. J. Phillips, C. Jiang and S. Wiener, "Comparison of visible and infra-red imagery for face recognition", *Int. Conf. on Automatic Face and Gesture Recognition*, 1996, pp. 182-187, 14-16 Oct.
51. S. Prabhakar and A. Jain, "Decision-level fusion in biometric verification", *Pattern Recognition*, 2002, vol. 35, no. 4, pp.861-874.
52. R. Brunelli and D. Falavigna, "Person identification using multiple cues", *IEEE Trans. on PAMI*, 1995, vol. 17, no. 10, pp. 955-966, Oct.
53. B. V. Dasarathy, "Decision fusion", *IEEE Computer Society Press*, 1994.
54. J. Kittler, M. Hatef, R. Duin and J. Matas, "On combining classifiers", *IEEE Trans. on PAMI*, 1998, vol. 20, no. 3, pp. 226-239,
55. S. Ben-Yacoub, Y. Abdeljaoued and E. Mayorz, "Fusion of face and speech data for person identity verification", *IEEE Trans. on NN*, 1999, vol. 10, no.5, Sept.
56. L. Hong and A. Jain, "Integrating faces and fingerprints for personal identification", *IEEE Trans. on PAMI*, 1998, vol. 20, no. 12, pp. 1295-1307.