Active vision-based face authentication

M. Tistarelli*, E. Grosso

Computer and Systems Science (DIST) Computer Vision Laboratory, Department of Communication, University of Genoa, Via Opera Pia 13-16145 Genoa, Italy

Received 29 September 1998; received in revised form 30 June 1999; accepted 13 July 1999

Abstract

The use of biometric data for automated identity verification, is one of the major challenges in secure access control systems. In this paper, several issues related to the application of active vision techniques for identity verification, using facial images, are discussed and a practical system (developed within an European research project), encompassing the active vision paradigm, is described.

The system, originally devised for banking applications, uses a pair of active tracking cameras to fixate the face of the subject and extract space-variant images (namely “fixations”) from the most relevant facial features. These features are automatically extracted with a two-level algorithm which uses a morphological filtering stage for a coarse localization, followed by an adaptive template matching.

A simple matching algorithm, based on a space-variant representation of facial features, is applied for identity verification and compared with a technique based on the Principal Component Analysis.

Several experiments on identity verification, performed on real images, are presented. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Face authentication; Biometric data; Active vision techniques; Principal component analysis

1. Introduction

The automatic verification of a person’s identity is a very interesting issue both in social and industrial environments. As an example you may consider: surveillance, law-enforcement, secure access control, smart interfaces and electronic commerce (World Wide Web).

Many approaches have been proposed for the authentication/recognition of a person’s identity, like fingerprint recognition, retinal scan, voice recognition, hand geometry, signature comparison, iris recognition and also facial image analysis [1]. All of these methods are characterized by intrinsic limitations and many require an active cooperation from the user, which implies a low social acceptability.

In principle, the analysis of face images seems to be the best way to perform identity authentication and also the most acceptable for people: this is what every human being does everyday in life. On the other hand, many difficulties arise from the enormous dimensionality of the search space when dealing with natural images. While a great progress has been made in the design an development of new methodologies for image-based person authentication/recognition to deal with pose, shadows or other ambiguities which may arise, only a few researchers devoted their efforts to limit or circumvent the complexity of vision-based recognition/verification systems.

In the last decade, the maturity of computer vision and the availability of low cost powerful computing hardware, pushed many researchers to develop novel techniques for human face recognition [2]. Pentland and Turk developed an algorithm for face recognition which is based on the principal component analysis (PCA) applied to an image database of human faces (the “eigenfaces”). The aim is to reduce the dimensionality of the search space by eliminating the redundancies in the image database [3]. This methodology has been refined over time also including face representations based on facial features [4]. The same concept underlying the “eigenface” approach was pushed forward by Belhumeur et al. [5] but also trying to avoid the dependence on facial illumination. This is accomplished by applying the Fisher’s linear discriminant analysis to detect facial regions with strong illumination changes (the “fisherfaces”). The same concept is to detect regions with strong illumination changes (the “fisherfaces” approach). Also other researchers tried to circumvent the intrinsic ambiguities arising from changes in pose and illumination of the human face. The solutions proposed span from purely geometric methods [6] to approaches based on a probabilistic representation of face appearance [7]. Edwards et al. [8] developed a statistical model of both the shape and the gray-level appearance of faces (the “active appearance model”) as a basis for facial image analysis. Three-dimensional
image data has been also successfully exploited for personal authentication. The added dimensionality provides more information to characterize the individuals, but also increases the amount of data to be processed, consequently adequate strategies must be adopted to make the authentication process more efficient [9–11].

Brunelli and Poggio [12] considered the trade between iconic and feature-based recognition methods reaching the conclusion that both approaches can be valid for a given application. More recently the support vector machine (SVM) theory [13], developed as a “general purpose” classifier, has been successfully applied to three-dimensional (3D) object recognition. The SVM classifier allows to represent the database classes as set of points in a higher dimensional space, where the class boundaries are explicitly defined as hyperplanes. The object is simply represented by the hyperplane parameters and the recognition, or inter-class matching, is simply reduced to a point-to-plane distance computation [14–16]. Even though the application of SVM to face authentication has not been attempted yet, it is envisaged that this might be a very powerful technique to distinguish between client and impostor claims.

It has been demonstrated that the amount of “distinctive information” is not uniformly distributed on the face, but rather few image regions convey the majority of the most distinctive features [17–19]. For this reason many methods have been proposed to identify “natural” facial features such as the eyes, nose and mouth [20–22]. Conversely, Walker et al. [23] define and detect salient features by means of differential invariants. It turns out that the efficiency of the recognition system can be improved by selecting a minimal set of areas on the face image to be processed and limiting the analysis to those areas.

In humans, for example, the capability to move and to plan the data acquisition process is very important to give a better description of the face but also to reduce the amount of information analyzed. This is accomplished both at the task level, by performing planned fixations, and at the sensor level, by adopting an appropriate sampling of the image. The use of active (mobile) cameras (to select the areas to be processed) and a space-variant image representation (to build a compact representation) allow to reduce the dimensionality of the data space and consequently the complexity of the system.

This paper incorporates these concepts to define a system for personal identification, originally conceived for secure access control in banking applications, which is based on four sequential modules:

- detection of the customer’s face;
- tracking of the face with two active (mobile) cameras;
- detection of the facial features within each image in a sub-set from the acquired sequence. Space-variant sub-images are extracted;
- matching of the model face (read from the personal smart card) with the extracted features.

The following sections contain a complete description of the system. In Section 2 the advantages of using active vision for face recognition are described. Section 3 reports the system set-up conceived within a banking application. In Section 4 the facial feature detection module is described. Section 5 is devoted to the identity verification algorithm. Some experiments are reported and discussed in Section 6.

Even though the paper refers to a complete system for personal identity verification, not all components are described in full detail. In particular, this paper will concentrate on the facial features detection module and the face authentication approach.

2. Recognition and active vision

Recognition is possibly the final motivation for many vision systems either artificial or natural (with different objectives, of course). One of the main reasons for the great complexity of recognition/verification tasks is the amount of information to be processed.

To achieve any visual task, all natural perceptual systems are capable of interacting with the environment and get as much information as needed, purposively controlling the flow of input data, but also limiting the amount of information acquired from the sensory system [24–26]. The anatomy of the human visual system is a clear example: despite the formidable acuity in the fovea centralis (1 min of arc) and the wide field of view (about 140 × 200° of solid angle), the optic nerve is composed of only 1 million nerve fibers. The space-variant distribution of the cells in the retina allows a formidable data flow reduction. In fact, the same resolution would result in a space-invariant sensor of about 600,000,000 pixels (or an optic nerve wider than the eye itself) [27].

Another important perceptual mechanism related to the data acquisition process is the attention mechanism. Again, as not all (visual in our case) input data is relevant for a given task, the perceptual system must be capable of making a selection of the input signal in various dimensions: “signal space”, depth, motion etc. The selection is controlled by a proper attention mechanism through ad hoc band-limiting or focusing processes.

The active vision paradigm takes into account these and other considerations related to existing perceptual systems, to realize artificial visual systems which are able to perform a given task under general assumptions [28].

2.1. Space-variant imaging

It is generally assumed that, to recognize an object, it is
necessary to provide a high resolution description of the most salient features of the object of interest. This can be accomplished either by “capturing”, in rapid succession, these parts of the scene or moving an interest window on a high resolution image [29]. On the other hand, it is not sufficient to scan the scene or the image with a high resolution window, but it is also necessary to provide some information on the area around the window. A way to meet these requirements is to adopt a space-variant sampling strategy of the image, where the central part of the visual field is sampled at a higher resolution than the periphery. In this way the peripheral part of the visual field, coded at low resolution, can still be used to describe the context in which high resolution data are located.

Many different models of space-variant image geometries have been proposed, like the truncated pyramid [29], the reciprocal wedge transform (RWT) [30], the complex logarithmic mapping (CLM) [31,32] and the log-polar mapping [33,34]. The system to be described for person authentication adopts the log-polar mapping to resample selected facial features extracted from a regular image.

The analytical formulation of the log-polar mapping describes the mapping that occurs between the retina (retinal plane \((\rho, \theta)\) and the visual cortex (log-polar or cortical plane \((\eta, \xi)\)). The derived logarithmic-polar law, taking into account the linear increment in size of the receptive fields, going from the central region (fovea) towards the periphery, is given by:

\[
\begin{align*}
  x &= \rho \cos \theta \\
  y &= \rho \sin \theta
\end{align*} \quad \text{and} \quad \begin{cases}
  \eta = q \theta \\
  \xi = \ln \frac{\rho}{p_0}
\end{cases}
\] (1)

where \(a\) defines the amount of overlap among neighboring receptive fields, \(p_0\) is the radius of the innermost circle, \((1/q)\) corresponds to the minimum angular resolution of the log-polar layout, and \((\rho, \theta)\) are the polar coordinates (see Fig. 1).

At present the log-polar transformation is computed at frame rate by using special re-mapping software routines. This approach has several advantages over special hardware devices, like space-variant sensors: it is more flexible and allows the use of conventional, low-cost cameras.

3. The system set-up

The system developed for person identification is sketched in Fig. 2. Images of the subject are acquired from the left and right camera and a color-based face tracker is activated. The image coordinates of the detected face are used to direct the gaze of both cameras toward the face center and start grabbing stereo images of the face. The face images are processed to detect three facial features: the eyes and the mouth. The face images are cropped and re-sampled to obtain a log-polar space-variant representation of each facial feature. The sampled images (which will be referred to as “fixations”) are used, either to build/update a face representation or to verify the identity of the subject.

In the remainder of the paper the facial features detection algorithm and the face authentication method are described. The color-based face tracker module, which is part of the system, has not been developed by the authors, and therefore it is not described in this paper [35,36].

4. Facial features detection

The technique applied to detect the facial features (Fig. 6), relies on the application of morphological operators in order to extract contours, valleys and peaks in the gray levels.
This information is gathered to make hypotheses for the presence of specific facial features. For example, the visible part of the sclera of the eye corresponds to a peak in the gray levels while the nostrils correspond to valleys. The generalized Hough transform is applied to the contours, for the rough localization of the facial features, and a final matching with adaptive templates is performed on the peaks and valleys images to refine the localization process.

Morphological operators have been successfully used in a number of applications involving classical problems of image processing (contours extraction, segmentation, shape analysis) and image restoration (noise suppression) [37]. Denoting by $\ominus$ the dilation operator and by $\ominus$ the erosion operator the contours extraction is obtained as:

$$C_r^e_c = I_r^e_c \ominus M(h,k) = I_r$$

where $M(h,k)$ is a two-dimensional mask, $I(r,c)$ is the input image and $C_r^e_c$ is the contour image. Morphological operators are particularly suitable for the extraction of peaks and valleys and can be “tuned” to the shape of the peak/valley simply by modifying the dimension and the weight of the image mask [38]:

$$P(r,c) = I(r,c) - (I(r,c) \ominus M(h,k)) \ominus M(h,k)$$

$$V(r,c) = I(r,c) \ominus M(h,k) \ominus M(h,k) - I(r,c)$$

where $P(r,c)$ denotes the peaks image, $V(r,c)$ denotes the valleys image (see Figs. 4 and 5) and $M(h,k)$ is the contour extraction mask.

The generalized Hough transform is applied for the rough localization of the eyes. In fact, the region containing the eye is characterized by many curved contours with a strong concentricity. Two histograms are computed to locate the center of circular contour segments [39]:

$$X_o(x_o) = \sum_{x=0}^{x_{max}} \sum_{r=r_{min}}^{r_{max}} I(x,y) \, ds$$

$$Y_o(y_o) = \sum_{x=0}^{x_{max}} \sum_{r=r_{min}}^{r_{max}} I(x,y) \, ds$$

where $I(x,y)$ is the image intensity and $C[r,(x_o,y_o)]$ denotes the circle along which the integral is computed. This procedure is applied to the contours image, producing a starting point for the fitting of the eyes template.

Two models have been considered to determine accurately the position of the eyes. The former is composed by a circle and a couple of parabola, with nine real parameters to be identified [22] (see Fig. 3). The latter is a very simple deformable model (an ellipse), which is described by four parameters and limited to the iris of the subject. Both models are driven towards the final position by a cost function, which is computed in the peaks/valleys images. Due to the high computational cost required to adapt the first eye model to the image, in the identity verification experiments only the simplified model has been used.

As the shape of the mouth strongly depends on the subject

![Fig. 3. The deformable template for the eye is given by a circle with radius $r$, centered in $x_c,y_c$, and by two parabola centered in $x_p,y_p$. The circle corresponds to the border between the iris and the sclera; the parabola the upper and lower bounds of the eye.](image)

![Fig. 4. Result of the morphological processing: (from left to right) original image, contours, valleys, peaks.](image)
and also on the facial expression, it is difficult to design a template which can be match against every mouth. Therefore, the position of the mouth is computed by looking for a linear pattern in the valleys image within the face region below the position of the eyes. False matches are avoided by imposing a geometrical constraint to the position of the eyes and mouth, which is to lie at the vertexes of a triangle. The exact position of the mouth is determined by computing the cross-correlation between the image and a mouth template, within a $10 \times 10$ window centered on the previously determined position. The mouth template is a gray level log-polar image obtained by averaging ten log-polar mouth images of five different subjects. The scale of the template is adjusted according to the inter-ocular distance on the image plane.\(^3\)

5. Visual identity verification

The authentication of a person's identity given biometric data from the subject is certainly a challenging problem which has been studied with a considerable effort in the last years. Dealing with access verification from image data, there are two distinct problems: face recognition (i.e. to recognize a subject's face given a database of faces) and identity verification (i.e. to identify a client from an image of the face) [3,12,20,40–42].

Dealing with real applications where a huge number of customers have access to several stations, it is quite difficult to maintain a large database of facial images (or any features) and assure an efficient retrieval of the data for recognition. As an example consider a marketing system based on the World Wide Web or the huge network of money tellers of banks worldwide. For this reason, it is more feasible to perform identity verification instead of recognition, based on a representation of the subject's appearance which is provided by the customer himself (or herself). This is possible, for example, by providing the access control system with a smart card\(^4\) reader, and storing, on the personal smart card, a representation of the person’s face.

Identity verification is related to but differs from face recognition. Therefore, techniques that have been formulated and successfully applied for the former may not give good results with the latter and vice versa.

For face recognition the population of all possible subjects (not including the impostors) is known a priori (the database), while it is unknown for verification: everyone can be an impostor. For this reason face verification is, in principle, a more difficult task than recognition.\(^5\) It is necessary to bound very well the person’s representation in the chosen feature space. On the other hand, given the

\(^3\) This measure is reliable as long as the rotation of the face around the vertical axis is small.

\(^4\) A smart card or chip card is a bank card with a small memory on it. This is used to store data about the customer’s account and also data for identification.

\(^5\) This is true whenever the space of possible subjects is bounded. In the special case where recognition is carried out with a reject option, then the subject space for recognition can also be unbounded and verification becomes a sub-problem of recognition (with a single class and reject option).
limited amount of memory currently available on a smart card (about 8 Kbytes up to a maximum of 16 Kbytes with next generation of smart cards), the face representation must be very compact.

5.1. Identity verification with the PCA

The principal component analysis technique has been originally proposed for face recognition by Sirovich and Kirby [41] and also applied by others with several variations and improvements [3, 5, 43].

In this approach a database of face images (the set of “known” faces) is coded by the “eigenimages” computed from the original set of face images, obtaining a reduced representation space.

The same procedure of Ref. [3] has been applied but using a set of log-polar fixations instead of a single image to represent each subject in the database. During the on-line recognition phase a new face vector is projected into the face space by computing the vector

\[ \mathbf{V} \]

The face is recognized according to the Euclidean distance \( e_k \) from the projection of each known individual [3]. The new face is recognized to belong to the individual corresponding to the minimum value of the error. An upper threshold is set to determine if the face image does not belong to any known individual.

In order to apply the PCA or “eigenimages” approach for face verification, a small database is built from a set of images of the same subject, showing several characteristic poses and facial expressions. The face image of the user is compared with this “database”.

5.2. Verification by intensity matching

A much simpler technique has been tested where a collection of fixations from the face image is used to represent a subject. The matching is performed by computing the correlation between the representation of the reference subject and the one requesting the access. The algorithm is based on the following steps:

1. Given the position of selected facial features (the eyes and the mouth), log-polar fixations are extracted from the acquired image of the subject.
2. The log-polar images are warped to simulate views which are as close as possible to the pose and orientation of the reference subject’s face (generally parallel to the image plane).
3. Corresponding fixations are compared by either computing the sum of the absolute value of gray level differences, or the normalized correlation:

\[
C(I) = \frac{\sum_x \sum_y (I(x, y) - \bar{I}(x, y))(I_{model}(x, y) - \bar{I}_{model}(x, y))}{\sqrt{\left(\sum_x \sum_y (I(x, y) - \bar{I}(x, y))^2\right) \left(\sum_x \sum_y (I_{model}(x, y) - \bar{I}_{model}(x, y))^2\right)}}
\]

where \( I(x, y) \) is the customer’s image, \( I_{model}(x, y) \) is the model image, \( \bar{I}(x, y) \) and \( \bar{I}_{model}(x, y) \) are the mean intensity values. A matching score is obtained from each fixation independently.

4. The scores obtained by the log-polar fixations are combined by discarding the lowest (or highest in case of gray level difference) score and averaging the remaining values.\(^7\)

\(^6\) If the difference of gray levels is used, the gray levels of each log-polar fixation are first normalized to the range of intensity values of the corresponding facial feature of the reference subject.

\(^7\) A rigorous procedure would require the analysis of the uncertainty associated with each fixation and the fusion of the corresponding scores weighted by the computed uncertainties. Nonetheless, experimental evidence showed that even this simple integration method allows to improve the results obtained by only considering the scores registered at each fixation.
This method, also including the facial feature extraction module, has been implemented under the Khoros® development environment. The biometric representation of the user (to be stored on the smart card) is composed of:

- three log-polar images of the eyes and mouth (approximately 6 Kbytes);
- the chosen log-polar parameters (it can be coded into 256 configurations, 1 byte);
- the max and min intensity values (2 bytes);
- the value of the inter-ocular distance (2 bytes);
- the normal to the subject’s face (3 bytes);
- the optimal threshold value to verify the subject’s identity (1 byte).

5.2.1. Pose estimation
Changes in the head pose may alter the appearance of facial features and impair the authentication process [8,44]. In order to take into account small movements of the head, the images are first “warped” to obtain a view as close as possible to the reference face position. In the devised system the subject’s head is tracked with two stereo cameras verging on the same point of the face. Instead of computing the full shape model of the face [10,40,45] only the three-dimensional plane containing the eyes and the mouth of the imaged face is computed. The vector normal \( \hat{n} \) to the plane containing the eyes and the mouth, can be computed from the image coordinates of the facial features extracted from the left and right image.

As a first step, the three-dimensional coordinates of the facial features are computed:

\[
X_i = \frac{x_i}{f_i} Z_i; \quad Y_i = \frac{y_i}{f_i} Z_i
\]

\[
Z_i = \frac{B}{f_i} \left[ \sin \alpha \sin \beta \frac{y_i}{f_i} + \cos \beta \frac{x_i}{f_i} - \sin \beta \cos \alpha \right] \\
\left[ \cos \alpha \frac{x_i}{f_i} + \sin \alpha \right] - \frac{y_i}{f_i} \left[ \sin \alpha \sin \beta \frac{x_i}{f_i} + \cos \beta \frac{y_i}{f_i} - \sin \beta \cos \beta \right]
\]

where \( \alpha \) and \( \beta \) represent the difference in the "pan" and "tilt" angles of the two cameras, \( B \) is the baseline of the stereo cameras, \( f_i \) and \( f_r \) are, respectively, the focal lengths of the left and right camera expressed in pixels, \( (x_l, y_l) \) and \( (x_r, y_r) \) are the coordinates of a feature point on the left and right image, respectively.

The normal vector to the face plane is computed from the coordinates of the three feature points in space:

\[
a = X_1 - X_3; \quad b = Y_1 - Y_3; \quad c = Z_1 - Z_3
\]

\[
d = X_2 - X_3; \quad e = Y_2 - Y_3; \quad f = Z_2 - Z_3
\]

\[
\tilde{N} = \begin{bmatrix} X & Y & Z \\ a & b & c \\ d & e & f \end{bmatrix}
\Rightarrow (bf - ce, cd - af, ae - db)
\]

\[
\hat{n} = \frac{\tilde{N}}{|\tilde{N}|}
\]

The normal vector \( \hat{n} \) is computed to re-project each image feature on the image plane.

Even though the planar approximation is not accurate for the whole face, it is satisfactory when considering a limited neighborhood of each facial feature. Therefore, the image warping is applied to each log-polar fixation separately. Conversely, rotations of the face on the image plane are inferred from the mutual position of the image coordinates of the two eyes, while the scale is adjusted by comparing the inter-ocular distance of the model with the subject, on the image plane. Given the scale and rotation invariance of the log-polar mapping, both the orientation and scale are corrected by simply shifting the pixels in both directions in the log-polar images.

The re-projection of the fixations is only required whenever the reference view and the current acquisition have a substantial angular difference in pose. Therefore, the warping is performed automatically only if the angular difference between the two planes exceeds a given threshold.

From Eq. (2) it is possible to compute the real inter-ocular distance of the imaged subject. This measure is compared with the same measurement from the reference subject as an additional check for identity verification.

6. Experiments

In all experiments the images used were divided into two disjoint sets: the training data set and the test data set. All the images in the test data set were used to perform impostor tests against all the images of the other subjects within the training data set. Also client tests were performed based on the available images for each subject within the test data set.

In general, the client tests are constrained by the number of views available for each subject within the test data set. In order to test the PCA approach for face verification, a database composed of a training set of images relative to the same “fixation”\(^8\) (in this case the left eye) from the same subject under several poses was built and a identity verification test performed with seven other subjects acting as impostors with two views for each subject.

In the first instance, the model was composed of only two eigenvectors. The diagram in Fig. 7a reports the distances computed for the two poses not included in the database of

\(^8\) A fixation is a sub-image extracted from a feature point on the face image. The sub-image is build by re-mapping the original image according to the log-polar law described in (1). An example of fixations extracted from a face image is presented in Fig. 6.
the correct subject (first two points) and the same poses related to the remaining seven subjects. Therefore, the first two scores represent the output of the client tests applied to the first subject, while the remaining part of the diagram represents the impostor tests.

In order to verify the scalability of the error with the number of eigenvectors, two more eigenvectors were added to the original face representation (see Fig. 7b).

A clear distinction in the output of the client and impostor tests was expected. As can be seen from Fig. 8a and b, even though in both cases the correct subject corresponds to the lower Euclidean distance, there is still one impostor showing a very similar error. This is probably due to the fact that even four eigenvectors are not sufficient to make a strong distinction between the client and one of impostors.

In Fig. 7, the results calculated by performing the same identity verification experiment as for the PCA, but obtained
with the intensity matching technique, are also shown. As can be seen, the client can be more easily identified in this case than with the PCA-based method. There may be many reasons for this. The first one is that the PCA technique is best suited to model a face given a high number of samples, which is normally required to perform recognition. On the other hand, the generalization power of the PCA can have a counter effect if too few eigenvectors are used and, as in the tests performed, if there are not enough features or they are not sufficiently distinctive to characterize the subject.

6.1. The FERET data set

In order to assess the performances of the correlation-based technique additional experiments were performed on a subset of the FERET database [46], including 53
subjects and two frontal views for each subject. The images are all black and white, with 8 bits of resolution in intensity and different sizes. Two example views of one subject are shown in Fig. 9. The log-polar images in the example have 70 eccentricities with 86 receptive fields for each circle. The overlap factor of the receptive fields is 1.6 in both directions. It is worth noting that, the stereo image warping to adjust the head pose via Eq. (3) was not required and could not be applied because the FERET database only includes monocular images.

The full verification test has been performed by taking one image of each subject out to build a reference model and using both images of the remaining subjects as impostors. This procedure was iterated over all subjects ending with 2756 impostor claims and 53 client claims. From the tests performed 41 subjects, out of the 53 are correctly identified. From the analysis of the failing tests it has been noticed that the main reason for a misclassification (either a false rejection or a false acceptance) is the low resolution of the

---

Fig. 9. (a) Original images of one of the subjects from the FERET database. (b) Log-polar fixations extracted from the left eye. (c) Re-mapped images of the log-polar representations in (b). As it can be noticed, the log-polar representation extracted from the second image (on the right) due to the size re-scaling and gray-levels interpolation does not contain enough information to be compared with the original picture (on the left) used to represent the client.
face within the image. In this case the extracted log-polar fixations do not contain enough information (namely gray level values of pixels at high frequency bands) to ensure a faithful representation of the individual. This is also due to the low resolution of the original images which is 256 × 384 pixels.

In order to take into account the variability in resolution of the face, the images have been divided into three different classes: small face (21 subjects), medium face (28 subjects) and large face (4 subjects). The three sets are defined empirically by the size of the smaller face in the two available images.

The full verification test has been performed adapting the log-polar grid to the size of the model image, in such a way that the portion of the face covered by the log-polar sampling is always the same, independently of the size of the face. In Fig. 10 the resulting false acceptance and false rejection error curve (FA–FR) and also the receiver operator curve (ROC) are shown. The reported equal error rate (EER) is about 9%.

The same authentication experiment has been performed excluding the small face set. In this case all subjects have been correctly classified, while the resulting FA–FR curve and the ROC curve are shown in Fig. 11. Remarkably the EER drops below 3%.

This result confirms that, even though space-variant fixations are adopted to represent the individual features, the original face image must be captured at the maximal available resolution to allow a faithful representation of the subject.

In the real system, where the acquisition set-up is controlled and tailored to the application scenario, it is possible to control the size of the imaged face with respect to the camera’s field of view, avoiding low resolution faces.

6.2. System-specific data set

In order to test the performances of the correlation-based authentication algorithm on a more realistic condition, a specific database composed of 50 subjects with two views for each subject was acquired with a camera set-up similar to the one sketched in Fig. 2. The images are black and white, 512 × 512 pixels and 8 bits of resolution in intensity. The zoom of the two cameras has been set in such a way that, when the pictures are acquired, the face of the subject covers almost all the image area. The face is tracked by the cameras and images are continuously captured. The signal to determine the valid images to be retained is given by the insertion of the smart card into the card reader of the machine. The images were acquired as to have two disjoint, and also distant in time, sets for the client and impostor claims. Two example views of two subjects are shown in Fig. 13. The log-polar images in the example have 70 eccentricities with 86 receptive fields for each circle. The overlap factor of the receptive fields is 1.6 in both directions.

The full verification test has been performed by taking one image of each subject out to build a reference model and using both images of the remaining subjects as impostors. This procedure was iterated over all subjects ending with 2450 impostor claims and 50 client claims. From the tests performed all subjects are correctly identified.

In Fig. 12 the resulting false acceptance and false rejection error curve (FA–FR) and also the receiver operator curve (ROC) are shown. The reported equal error rate (EER) is about 6%.

The decrease in performance with respect to the FERET database is due to the fact that the realistic acquisition conditions introduce many more disturbing effects, especially in the background, which may degrade the performances of the matching algorithm and, more important,
the feature detection process. Moreover, noise is also introduced by the image warping which is necessary to compensate the effects of perspective, while in the FERET images this was not required because the faces are all fronto parallel with respect to the image plane.

6.3. Failure modes and success rate

In some cases the system fails to correctly authenticate the identity of the subject. From an extensive analysis of the failing tests it has been noticed that the main reason for a misclassification (either a false rejection or a false acceptance) is the low resolution of the face within the image. In this case the extracted log-polar fixations do not contain enough information (namely gray level values of pixels at high frequency bands) to ensure a faithful representation of the individual. This is confirmed by the example in Fig. 8 where the client is correctly identified even if the face image is very small. This is probably due to the fact that the subject has big eyes and a large mouth, which cover a significant portion of the face and therefore, even after the log-polar mapping, still contain a sufficient number of pixels to assure a reasonable representation of the individual. Another source of errors is the presence of severe highlights on the eyeglasses. The system can cope for moderate specularities but not in cases such as in Fig. 14 (bottom).

Nevertheless, from these tests, the system does not seem to be seriously influenced by artifacts such as eyeglasses, change in facial expression or illumination. It is worth noting that the system is capable of performing a correct
identification of the subject, even in quite difficult situations where there is a change in scale and pose, illumination and facial expression, or if the person wears eyeglasses in one of the images, even with specular reflections.

In a real system, where the acquisition set-up can be controlled or tailored to the application scenario, it is possible to control the size of the imaged face with respect to the camera’s field of view, avoiding low resolution faces such as in Fig. 9. The dramatic improvement is demonstrated by the experiments performed on the system-specific data set. On the other hand, in order to cope for the case where the subject has specular reflections, it is sufficient to control the lighting of the acquisition station or to perform the identification process more than one time: it is unlikely for the subject to keep the in the same position for more than one snapshot.

### 7. Conclusions

The active interaction with the environment and the capability to control the flow of input data are key aspects for the solution of many vision problems. The active vision paradigm, which encompasses these and other principles, has been considered to design a system for personal identity verification based solely on biometric measurements.

Far from claiming that new insights on the active vision paradigm have been achieved, this paper presents how the application of active vision principles, such as tracking, the use of fixations and space-variant (foveated) representations, can lead to the realization of efficient vision systems operating in real world environments. The current availability of motorized cameras and low cost computing hardware, make it possible to build such a system for real applications.

The advantages of the “active” approach for identity verification have been presented, such as:

- the possibility to move and obtain a better description of the face during the authentication phase (the entire face is always captured within the image);
- a space-variant representation of the images, allows a reduction in the size of the face model and consequently also the time required for authentication.

The PCA technique has been applied straightforwardly to perform identity verification tests. The preliminary results obtained already show that the discrimination power of the PCA may require a greater memory storage capacity than what is currently available from the smart card technology. On the other hand, the direct image matching exhibits a greater stability requiring far less memory usage. A further investigation is required to determine whether using a different number of fixations, poses of the face, and/or eigenvectors, to represent a subject with the PCA technique, may improve the robustness and stability of the method.

In this paper the authentication process has been limited to a single snapshot, which is a “win or lose” scenario. In the reality of biological systems more than one guess is tried before determining the person’s identity and say whose face is the one in front. The same mechanism can be adopted by
matching more than one single image and taking as output a weighted average of the individual responses. On the other hand, it may seem reductive to use just the image intensity to characterize the subject’s appearance. Other visual features may be used as well as other matching processes combining the resulting similarity scores within a multi-expert decision maker.

Several aspects have been addressed explicitly and many are still under investigation. For example, one topic that is still to be more deeply investigated is the relation between the log-polar representation and the principal component analysis obtained from the eigenface approach. This analysis would be very important to further explore the advantages of a space-variant representation for recognition.

Acknowledgements

We thank A. Lagorio, M. Jentile, M. Perrone and S. Rimassa for their valuable contribution to the development
of this paper. The authors wish to thank the reviewers for their helpful suggestions to improve the quality of the paper. This work has been partially funded by the LTR Esprit Project 21894 “VIRSBS”.

References


