# Face Identification by Deformation Measure

Bertrand LEROY[1]        Isabelle L. HERLIN[1]        Laurent D. COHEN[2]

[1]INRIA - Projet AIR , Rocquencourt - B.P. 105, 78153 Le Chesnay Cedex, France.
[2]Université Paris IX Dauphine - CEREMADE,
Place du Maréchal de Lattre De Tassigny - 75775 Paris Cedex 16.
E-mail: Bertrand.Leroy@inria.fr,Isabelle.Herlin@inria.fr,
Laurent.Cohen@dauphine.fr

## Abstract

*This paper studies the problem of face identification for the particular application of an automatic cash machine withdrawal: the problem is to decide if a person identifying himself by a secret code is the same person registered in the data base. The identification process consists of three main stages. The localization of salient features is obtained by using morphological operators and spatio-temporal information. The location of these features are used to achieve a normalization of the face image with regard to the corresponding face in the data base. Facial features, such as eyes, mouth and nose, are extracted by an active contour model which is able to incorporate information about the global shape of each object. Finally the identification is achieved by face warping including a deformation measure.*

## 1. Introduction

Research studies related to face recognition from 2-D frontal views may be shared in two main classes according to whether they refer to a feature based approach or to a pixel approach. The first class consists in extracting the position of several facial features in order to restrict the data space. Several statistical methods, generally classical algorithms of pattern recognition, are then used to identify faces using these measures [13]. The pixel approach is a global one that requires few knowledge about face geometry. Some studies are linked to pattern matching: the measures used to identify faces are correlations between template regions. Other methods use principal component analysis applied on pixel's grey level values [12].

The application, described in this paper, is a system similar to an automatic cash machine withdrawal where the user identifies himself by using a credit card with a secret code. It is possible to enhance the security of such a transaction by verifying that the observed face is identical to a stored face of the same person or to compare the face to information (points, curves, quantitative values, . . . ) obtained by learning and stored in the data base.

We propose an identity check system using warping between faces as a measure of likeness. Knowing the assumed identity of a face (supposed by the secret code), the system compares this face with that one in the database corresponding to the same person. This is achieved by warping one face in order to fit the other one.

Since we have chosen to use the geometry and the locations of facial features (such as the mouth, nose and eyes) to achieve warping, each person of the database is represented by this information.

The face identification scheme consists of three main stages. In the first step, in section 2, facial features are located and the image is normalized so that features location match those of the face corresponding to this person in the database. In section 3, the features boundaries are extracted using a specific active contour model. This model is based on Fourier descriptors and is able to incorporate information about the global shape of each object. The last stage is described in section 4 and consists in measuring the warping between the normalized face and the corresponding face in the database.

## 2. Face normalization

In the context of our application, the image acquisition is achieved with a fixed lightning and is acquired from a near frontal view throughout the temporal sequence. The normalization process is based on this constraint.

The first step of our facial features localization process concerns the eyes. The nose and the mouth will be further detected using some facial geometry hypothesis.

The presence of specular spots (usually in the neighborhood of the irises) due to the direct lightning allows to locate the eyes using a "peak" morphological operator which has

the effect of bringing out the local maxima of the grey level values function. A rough estimation of the eyes position is consequently obtained by looking for the two regions corresponding to the highest values of the "peak" image. Since the lightning used for the acquisition is frontal, these estimations correspond approximately to the irises positions.

In order to locate the nose and the mouth, we then use a spatio-temporal information by computing the sum of the spatio-temporal gradient norm over the sequence:

$$I_{sum} = \sum_{i=1}^{N} \sqrt{(\frac{\partial I}{\partial x})^2 + (\frac{\partial I}{\partial y})^2 + \mu(\frac{\partial I}{\partial t})^2},$$

where $I(x, y, t)$ represents the grey level value, $x$ and $y$ the spatial components, $t$ the temporal component and $N$ the number of images of the sequence. $\mu$ is a normalization term depending on the temporal sampling of the sequence and on the motion of the different features during the sequence.



**Figure 1. Opposite of the sum of the spatio-temporal gradient norm calculated on an image sequence.**

As it is shown on figure 1, the image $I_{sum}$ permits to pick up regions corresponding to a weak contrast but having a high deformation, like the mouth, and regions having a high contrast like the shade due to the nose relief. The localization of the nose and the mouth in the image $I_{sum}$ consists in searching, along the mediatrix of the segment joining the two irises, the regions having the two higher values of the image $I_{sum}$. In order to minimize the computing time, the image $I_{sum}$ is computed in small windows on both sides of the mediatrix between the two irises. Figure 2 shows the results obtained by this algorithm and it can be seen that the localization of the lower part of the nose is obtained with a good accuracy. The point corresponding to the position of the mouth depends on its motion during the sequence and is not always the same physical point. When the motion is important, this point is located within the mouth and otherwise it is generally located on the boundary of the lips.

In order to compute the image $I_{sum}$, we use the first 30 images of the sequence (with a sampling rate of 25 hz). Using a shorter sequence leads to less accurate results for the temporal information will be less significant.

In next section we make use for the system of an active contour model to extract facial features boundaries: in order to use the same initialization and parameters for the active



**Figure 2. Results of the localization of the eyes, nose and mouth.**

contour model with different images acquired from one face, we normalize the image.

This is achieve by applying a similarity using the irises location as references. Since the active contour parameters and the curves used for initialization are not the same for different faces, these features will be stored in the database with the facial characteristics.

## 3. Facial feature extraction

An active contour model is an energy minimizing curve guided by internal constraints and influenced by image forces that pull it towards image edges. Since the global regularity constraints used by generic active contour models (such as "snakes" [8][3]) are included within the internal energy function, these models do not allow to refine the space of admissible curves to a specific shape class. In order to take into account the shape of facial features, several methods based on deformable templates have been developed.

Yuille, Cohen and Hallinan [13] proposed a set of deformable templates for this purpose. Each one of these templates is designed to extract a specific feature. For example the template used to extract the eye boundary consists of a circle bounded by two parabolas and possesses 11 parameters.

Craw, Tock and Bennett [6] described a model which consists of a set of 40 landmark points, each one representing the position of a particular part of the structure to be located. Several algorithms are designed to deal with a particular feature using a statistical knowledge of the landmark positions. A global algorithm controls the interaction between them. A similar method has been proposed by Cootes, Tailor and Lanitis [4] using statistically based models which iteratively move towards structures in images after a learning process on similar images.

In this section we propose a parametrical active contour model using Fourier descriptors. The aim is to design a model sufficiently general to extract the boundary of several facial features while taking into account the specific geometry of each feature. This model needs to tune few

parameters.

Actually, this model is able to extract facial features subjected to high deformation (in order to take lips motion into account) and having corners (like the mouth and the eyes).

### 3.1. Active contour model using Fourier descriptors

The use of Fourier descriptors for active contour models has been introduced by Staib and Duncan [11] in order to extract an object boundary. Their method is based on the use of probability distributions on the parameters. In order to be less sensitive to the initial parameter value, we propose a variational approach similar to the method used in the snake model [8].

An elliptic Fourier representation of a closed curve is a parametrical curve $v$ defined by:

$$v(\theta) = \begin{pmatrix} x(\theta) \\ y(\theta) \end{pmatrix} = \sum_{k=0}^{N} A_k \begin{pmatrix} \cos(\theta k) \\ \sin(\theta k) \end{pmatrix}, \quad (1)$$

where $A_k$ is a $2 \times 2$ matrix, $N$ the number of harmonics used to describe the curve and $\theta$ the angular parameterization index.

Such a class of curves does not permit to represent effectively the boundaries of objects having an irregularity point within a regular section, such as the corner of the mouth or eyes. Thus, the extraction of this kind of boundary requires the use of high frequency harmonics which have the effect of creating oscillations. In order to solve this problem we propose a new model made of two open curves joined at their ends:

$$v(\theta) = \begin{cases} v_A(\theta) & \text{if } 0 \le \theta < \pi \\ v_B(\theta) & \text{if } \pi \le \theta < 2\pi \end{cases}$$

where $v_A$ and $v_B$ respect the conditions:

$$v_A(0) = v_B(2\pi) \text{ and } v_A(\pi) = v_B(\pi).$$

In order to use orthogonal bases on the interval $[0, \pi]$, an open curve can be defined by:

$$\begin{pmatrix} x(\theta) \\ y(\theta) \end{pmatrix} = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + R_\phi \sum_{k=1}^{N} \begin{pmatrix} a_k \cos(\theta k) \\ b_k \sin(\theta k) \end{pmatrix}, \quad (2)$$

where $a_k$ and $b_k$ are the parameters model, $N$ is the number of harmonics used to describe the curve and $R_\phi$ is the rotation matrix of angle $\phi$. Thus, the phase shift between the two bases used for $x(\theta)$ and $y(\theta)$ permits to describe a half-ellipse with four parameters.

The curve modeling the boundary of the object, is obtained by minimizing a functional similar to the snake energy:

$$\begin{aligned} E(v) &= \int_0^\pi P(v_A(\theta)) + \lambda(v_A')^2 d\theta \\ &+ \int_\pi^{2\pi} P(v_B(\theta)) + \lambda(v_B')^2 d\theta. \end{aligned}$$

The first term $P$ inside each one of the integrals is an image potential equal to the opposite of the square of the image gradient ($P = -|\nabla I|^2$) and the second one is an elasticity term associated to the curve tension.

In order to extract boundaries of the mouth and eyes we use one harmonic to describe the lower boundary, $v_B$ and $N$ harmonics for the upper boundary, $v_A$. Thus $v_A$ has the form defined in equation (2) and $v_B$ is defined by:

$$v_B(\theta) = \begin{pmatrix} \sum_{i=0}^{N/2} a_{2i} \\ b_0 \end{pmatrix} + R_\phi \begin{pmatrix} \cos(\theta) \sum_{i=0}^{N/2} a_{2i+1} \\ c\sin(\theta) \end{pmatrix}.$$

To extract the lower boundary of the nose, we use a single open curve described by equation (2).
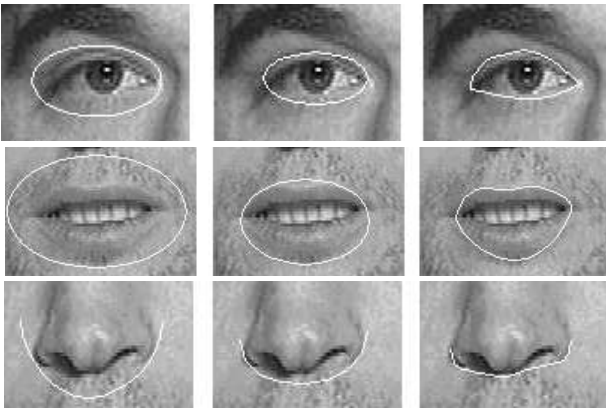
### 3.2. Results

The algorithm used to extract features is a fine to coarse scheme (see figure 3). Defining two half-ellipses as initialization, the convergence is first achieved using a single harmonic for the curve $v_A$, then the first result is used as an initialization to get a curve with two harmonics until we obtain a curve $v_A$ described by $N$ harmonics fitting the object boundary.

As the number of harmonics increases, the regularity constraints are smaller and the boundary can be described more precisely with a better stability.

## 4. Face Warping for identification

Several methods have been proposed to achieve face recognition using features location. Some studies use this information in order to calculate the correlation between the regions corresponding to a same feature in two different images [1]. The measure is then equal to the mean of the correlation values. Other methods utilize a more geometrical approach by defining a discriminating measure using physiological knowledge [7], results of principal component analysis [5] or topological graphs [9].

In this paper, two faces are compared by a deformation measure. Given a set of landmark points on the face, the measure is a quantification of the bending that has to be applied to the face, so that its landmark points match the ones of the other face. To this issue, we utilize a class of

**Figure 3. Initialization (first column) and results obtained with 1 (second column) and then 3 (third column) harmonics for the upper boundary of mouth and eyes and the lower boundary of nose.**

deformations similar to the one used for normalization by Reisfeld [10]. While his aim was to normalize faces in order to apply a classification on the image data, we propose to achieve the identification using the deformation measure value as the discriminating information.

### 4.1. Face characterization and database

The landmark points used to represent the face are located on the the features boundaries which have been extracted as described in section 3. Using all the points of the boundaries for the matching process will be useless since the information is redundant (like neighbor points on the same boundary) and some points cannot be easily extracted from the curves describing the boundaries (due to the parameterization which is not intrinsic). We made the choice of 15 points displayed on figure 4. These points characterize the position of the features as well as their global geometries.
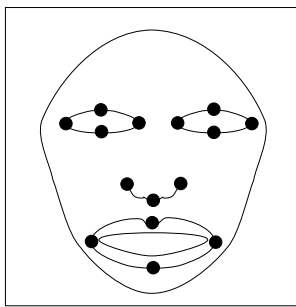


**Figure 4. points characterizing a face**

However some of the chosen landmark points are mobile since they are located on a feature which is subject to temporal deformation. Therefore, in order to construct the database we use a sequence of images of people uttering a particular sentence while blinking. Then, on each image of the sequence, the features boundaries are extracted on a normalized image (as it is shown in sections 2 and 3) and for each landmark point we calculate the distribution of its position. Normality tests showed that these distributions may be represented by Gaussian laws. Thus, a person is represented in the database by a set of 4 values:

$$\{x_i, \sigma_{xi}, y_i, \sigma_{yj}\}$$

where $(x_i, y_i)$ is the mean position of the $i^{th}$ landmark point and $(\sigma_{xi}, \sigma_{yj})$ its standard deviation on the horizontal and vertical axes.

### 4.2. Deformation measure

Bookstein [2] showed the interest of the thin plate splines for image registration purpose since they can be used to solve the problem of scattered data interpolation while minimizing the warping.

The total amount of bending of a surface $(x, y, f(x, y))$ ($f$ being a twicely differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$), can be quantified by the value of the functional:

$$I(f) = \int\int_{\mathbb{R}^2} (\frac{\partial^2 f}{\partial x^2})^2 + 2(\frac{\partial^2 f}{\partial x \partial y})^2 + (\frac{\partial^2 f}{\partial y^2})^2 dxdy \quad (3)$$

Given a set of landmark points $\{p_i = (x_i, y_i)\}$ on one face and $\{p_i' = (x_i', y_i')\}$ on a second one, the function $T = (T_x, T_y) : \mathbb{R}^2 \to \mathbb{R}^2$, that matches each point $p_i$ to $p_i'$ with respect to its standard deviation $(\sigma_{xi}, \sigma_{yi})$ and that is least bent according to the functional (3), minimizes the functional:

$$\begin{aligned} J_T &= I(T_x) + I(T_y) \\ &+ \lambda \sum_{i=1}^{N} (\frac{(T(x_i) - x_i')^2}{h(\sigma_{xi})^2} + \frac{(T(y_i) - y_i')^2}{h(\sigma_{yi})^2}). \end{aligned}$$

$\lambda$ is a positive weighting parameter and $h$ is defined by:

$$h(x) = x \text{ if } |x| > 1 \text{ and } 1 \text{ otherwise.}$$

Thus, a weight is associated to each landmark point which decreases the influence of a point if its standard deviation is large. The function $h$ is constructed so that $J_T$ is always defined and the points having small standard deviation are not too influent.

The transformation $T$, that minimizes $J_T$, has the following form:

$$T_x(x,y) = a_1 + a_2 x + a_3 y + \sum_{i=1}^{N} w_i U(|T(p_i) - (x,y)|)$$

$$T_y(x,y) = b_1 + b_2 x + b_3 y + \sum_{i=1}^{N} z_i U(|T(p_i) - (x,y)|)$$

Several types of function $U$ can be used in order to obtain a minimum for $J_T$. In particular the use of $U(x) = x^2 \log(x)$ (with U(0)=0) leads to an unique solution verifying the property of the thin plate spline. The coefficients $a_1, a_2, a_3, b_1, b_2, b_3, w_1, \ldots, w_N$ and $z_1, \ldots, z_N$ can be calculated by solving a linear system as described in [2] [10].

The quantity $J(T)$ may be used to measure the likeness between two faces. The assumed identity will then be confirmed if this value is smaller than a given threshold. The threshold value is linked to the accuracy of the system that is wished for. If it is small, it can hapen that the identity of a person will be wrongly rejected; if it is large, the identification test will then be less restrictive. Lastly, the choice of the treshold value must take several parameters into account such as the image acquisition and the accuracy of the localization and extraction processes.

## 5. Conclusions

This paper presents a set of methods that are used to construct an identity check system. The localization of facial features (such as the eyes, nose and mouth) is achieved using a morphological operator as well as the spatio-temporal information. This method is robust provided that lightning is frontal. It is used to normalize the face image with regard to the corresponding face in the data base. The active contour model proposed for feature boundary extraction is a parametrically deformable model using Fourier descriptors. This model is designed so that few parameters are needed to extract irregular objects like the eyes or the mouth and the stability is increased by applying a coarse to fine scheme to the convergence process. A face is represented in the database by a set of landmark points belonging to the features boundaries as well as the standard deviations of the points calculated on a sequence of images. Finally the likeness between two faces is quantified using a deformation measure. This deformation corresponds to the bending to be applied on a face, so that its landmark points match the ones of the other face. To this issue, an enhancement of a standard method used to interpolate scattered data has been proposed which allows to associate a weight to each landmark point so that points having a small standard deviation have more effect on the bending.

## References

[1] D. J. Beymer. Face recognition under varying pose. In *IEEE Proceedings of Computer Vision and Pattern Recognition*, pages 556–761, 1994.

[2] F. L. Bookstein. Principal warps: thin-plate splines and decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.

[3] L. D. Cohen and I. Cohen. Finite element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, November 1993.

[4] T. F. Cootes, C. J. Taylor, and A. Lanitis. Multi-resolution with active shape models. In *Proceedings of the International Conference of Pattern Recognition*, pages 610–612, 1994.

[5] N. Costen, I. Craw, G. Robertson, and S. Akamatsu. Automatic face recognition: what representation? In *Proceedings of the First European Conference on Computer Vision*, 1996.

[6] I. Craw, D. Tock, and A. Bennett. Finding face features. In *Proceedings of the First European Conference on Computer Vision*, pages 92–96, 1992.

[7] T. Kanade. *Picture processing system by computer complex and recognition of human faces*. PhD thesis, Kyoto University, Department of Information Science, November 1973.

[8] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. In *IEEE Proceedings of the International Conference on Computer Vision*, pages 259–268, London, June 1987.

[9] B. Manjunath, R. Chellappa, and C. von der Malsburg. A feature based approach to face recognition. In *IEEE Proceedings of Computer Vision and Pattern Recognition*, pages 373–378, 1992.

[10] D. Reisfeld. *Generalized symmetry transforms: attentional mechanisms and face recognition*. PhD thesis, Tel-Aviv University, January 1994.

[11] L. H. Staib and J. S. Duncan. Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(11):1061–1075, November 1992.

[12] M. Turk. *Interactive-time vision: face recognition as a visual behavior*. PhD thesis, Massachusetts Institute of Technology, September 1991.

[13] A. Yuille, D. Cohen, and P. Hallinan. Facial feature extraction by deformable templates. Technical Report 88-2, Harvard Robotics Laboratory, 1988.