PERGAMON

Pattern Recognition 33 (2000) 1525–1539

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

# Real-time face location on gray-scale static images

Dario Maio[a,b,*], Davide Maltoni[a,b]

[a]*Corso di Laurea in Scienze dell'Informazione, Università di Bologna, via Sacchi 3, 47023 Cesena, Italy*
[b]*DEIS - CSITE-CNR, Facoltà di Ingegneria, Università di Bologna, viale Risorgimento 2, 40136 Bologna, Italy*

Received 14 October 1998; accepted 19 May 1999

**Abstract**

This work presents a new approach to automatic face location on gray-scale static images with complex backgrounds. In a first stage our technique approximately detects the image positions where the probability of finding a face is high; during the second stage the location accuracy of the candidate faces is improved and their existence is verified. The experimentation shows that the algorithm performs very well both in terms of detection rate (just one missed detection on 70 images) and of efficiency (about 13 images/s can be processed on Hardware Intel Pentium II 266 MHz).


*Keywords:* Face location; Directional image; Generalized Hough transform; Ellipse detection; Elliptical fitting; Deformable template

## 1. Introduction

Automatic face location is a very important task which constitutes the first step of a large area of applications: face recognition, face retrieval by similarity, face tracking, surveillance, etc. (e.g. Ref. [1]). In the opinion of many researchers, face location is the most critical step towards the development of practical face-based biometric systems, since its accuracy and efficiency have a direct impact on the system usability. Several factors contribute to making this task very complex, especially in the case of applications requiring to operate in real-time on gray-scale static images. Complex backgrounds, illumination changes, pose and expression changes, head rotation in the 3D space and different distances between the subject and the camera are the main sources of difficulty.

Many face-location approaches have been proposed in the literature, depending on the type of images (gray-scale images, color images or image sequences) and on the constraints considered (simple or complex background, scale and rotation changes, different illuminations, etc.). Giving a brief summary of the conspicuous number of works requires a pre-classification; unfortunately, due to the large amount of different techniques used by researchers this task is not so easy. While we are aware of the unavoidable inaccuracies, we have tried to make a tentative classification:

- Methods based on template matching with static masks and heuristic algorithms which use images taken at different resolutions (multiresolution approaches) [2,3].
- Computational approaches based on deformable templates which characterize the human face [4] or internal features [5–8]: eyes, nose, mouth. These methods can be conceived as an evolution of the previous class, since the templates can be adapted to the different shapes characterizing the searched objects. The templates are generally defined in terms of geometric primitives like lines, polygons, circles and arcs; a fitness criterion is employed to determine the degree of matching.
- Face and facial parts detection using dynamic contours or snakes [6,9–11]. These techniques involve a constrained global optimization, which usually gives very accurate results but at the same time is computationally expensive.
- Methods based on elliptical approximation and on face searching via least-squares minimization [12],

* Corresponding author. Tel.: + 39-(051)-2093547; fax: + 39-(051)-2093540.
*E-mail address:* dmaio@deis.unibo.it (D. Maio).

0031-3203/00/$20.00 
PII: S0031-3203(99)00130-2

incremental ellipse fitting [13] and elliptic region growing [14].
- Approaches based on the Hough transform [7] and the adaptive Hough transform [15].
- Methods based on the search for a significant group of features [triplets, constellations, etc.] in the context considered: for example, two eyes and a mouth suitably located constitute a significant group in the context of a face [7,16–19].
- Face search on the eigenspace determined via PCA [20] and face location approaches based on the information theory [21,22].
- Neural Network approaches [23–30]. The best results have been obtained by using feed forward networks to classify image portions normalized with respect to scale and illumination. During the training, examples of face objects and non-face objects are presented to the network. The high computational cost, induced by the need to process at different resolutions all the possible positions of a face in the image, is the main drawback of these methods.
- Face location on color images through segmentation in a color-space: YIQ, YES, HSI, HSV, Farnsworth, etc [27,31–36]. Generally, color information greatly simplifies the localization task: a simple spectrographic analysis shows that the face skin pixels are usually clustered in a color space, and then an ad hoc segmentation allows the face to be isolated from the background or, at least, to drastically reduce the amount of information which must be processed during the successive stages.
- Face detection on image sequences using motion information: optical flow, spatio-temporal gradient, etc. [27,33,37].

Since in several applications it is mandatory (or preferable) to deal with static gray-scale images we believe it is important to develop a method which does not exploit additional information like color and motion. For example, most of the surveillance cameras nowadays installed in shops, banks and airports are still gray-scale cameras (due to the lower cost), and the electronic processing of mug-shot or identity card databases could require to detect faces from static gray-scale pictures printed on paper. Unfortunately, if we discard color and motion-based approaches, the most robust methods are generally time consuming and cannot be used in real-time applications.

The aim of this work is to provide a new method which is capable of processing gray-scale static images in real time. The algorithm must operate with structured backgrounds and must tolerate illumination changes, scale variations and small head rotations.

Our approach (Fig. 1(a)) is based on a location technique which starts by approximately detecting the image positions (or *candidate positions*) where the probability to find a face is high (module AL) and then for each of them improves the location accuracy and verifies the presence of a true face (module FLFV). Actually, most of the applications in the field of biometric systems require detection of just one object in the image (i.e. the foreground object): under this hypothesis, a more efficient implementation of our method is reported in Fig. 1(b), where at each step the module AL passes only the most likely position to FLFV and FLFV continues to require a new position until a valid face is detected or no more candidates are available. It should be noted that, even in this case, the system could be used to detect more faces in an image,

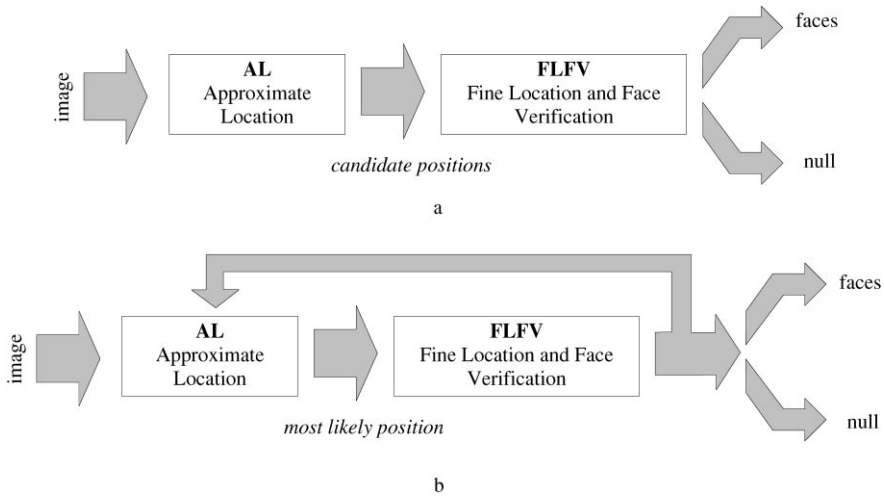


Fig. 1. Two different functional schemes of our approach.

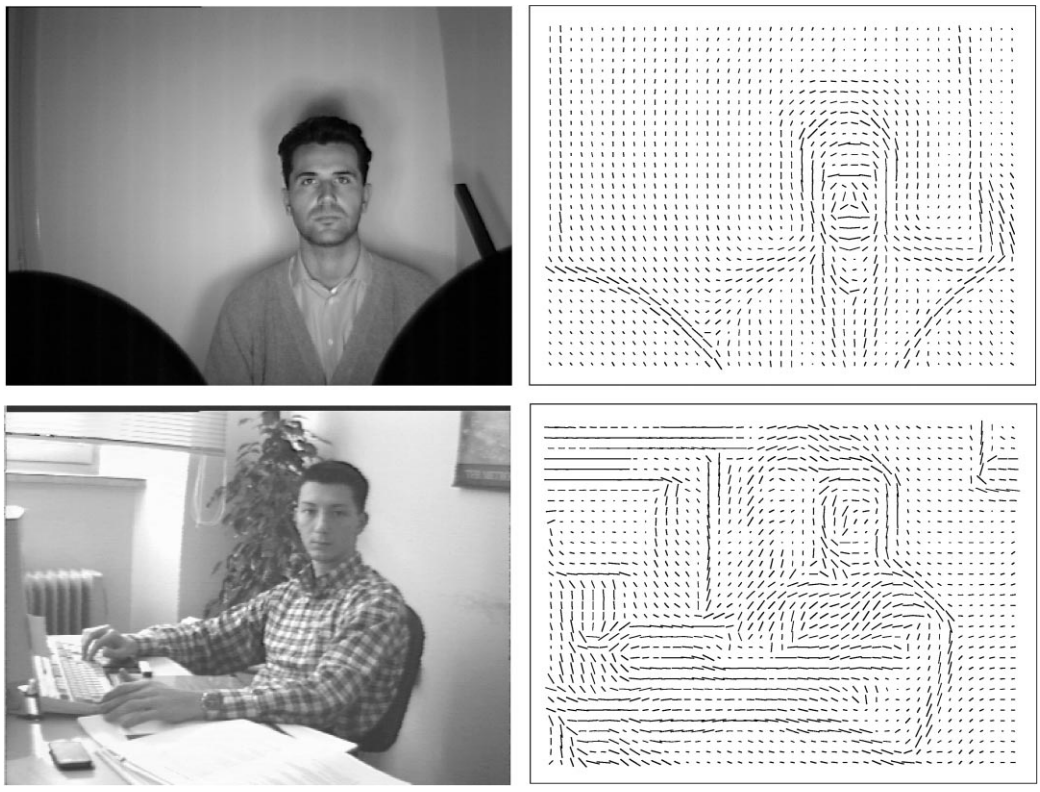

Fig. 2. Two images and the corresponding directional images. The vector lengths are proportional to their moduli.

assuming that the iterative process is not prematurely interrupted.

Although AL and FLFV have been implemented in a very different manner, both these modules work on the same kind of data: that is the directional image extracted by the starting gray-scale image.

In Section 2 the directional image is defined and some comments about its computation are reported. Section 3 describes the module AL which is based on the search of elliptical blobs in the directional image by means of the generalized Hough transform. In Section 4 we present the dynamic-mask-based technique used for fine location and face verification (module FLFV) and in Section 5 we discuss how to practically combine AL and FLVF in order to implement the functional schema of Fig. 1(b). Section 6 reports the results of our experimentation over a 70 image database; finally, in Section 7, we present our conclusions and discuss future research.

## 2. Directional image

Most of the face location approaches perform an initial edge extraction by means of a gradient-like operator; few methods also exploit other additional features like directional information, intensity maxima and minima, etc. Our technique strongly relies on the edge phase-angles contained in a *directional image*.

A directional image is a matrix defined over a discrete grid, superimposed on the gray-scale image, whose elements are in correspondence with the grid nodes. Each element is a vector lying on the $xy$ plane. The vector direction[1] represents the tangent to the image edges in a neighborhood of the node, and its modulus is determined as a weighted sum of the *contrast* (edge strength) and the *consistency* (direction reliability) (Fig. 2). In this work the directional image is computed by means of the method proposed by Donahue and Roklin [38]. Each directional image element is calculated over a local window where a gradient-type operator is employed to extract several directional estimates (2D sub-vectors), which are averaged by least-squares minimization to control noise. This technique is more robust than the standard operators used for computing the gradient phase angle and enables the contrast and the consistency

[1] The vector direction is unoriented and lies in the range $[-90°, +90°[$.

to be calculated with a very small overhead. In particular, the contrast is derived by the magnitude of the 2D sub-vectors and the consistency is in inverse proportion with the residual resulting from the least-squares minimization [38] (in fact, an high reliability corresponds to a low residual, that is, all the 2D sub-vectors are nearly parallel).

## 3. AL - approximate location

The analysis of a certain number of directional images suggested the formulation of a simple method for detecting faces. In particular, we noted that when a face is present in an image the corresponding directional image region is characterized by vectors producing an elliptical blob. For this reason, the module AL is based on the search for ellipses on the directional image. Several techniques could be used for this purpose, for example multiresolution template matching [39] and least-squares ellipse fitting [12]. We adopted a new approach based on the generalized Hough transform [40] which performs very well in terms of efficiency. The method is capable of detecting the approximate position of all the ellipses within a certain range of variation defined according to pre-fixed scale and rotation changes. The basic idea is to perform a generalized Hough transform by using an elliptical annulus **C** as template. Actually, the directional information allows the transform to be implemented very efficiently, since the template used for the accumulator array updating can be reduced to only two sectors of the elliptical annulus.

Formally:

Let $a$ and $b$ be the lengths of the semi-axes of an ellipse used as reference, and let $\rho_r$ and $\rho_e$ be, respectively, the reduction and expansion coefficients defining the scale range (and hence the elliptical annulus **C**): $a_{min} = \rho_r \cdot a$, $b_{min} = \rho_r \cdot b$, $a_{max} = \rho_e \cdot a$, $b_{max} = \rho_e \cdot b$ (Fig. 3). Let **D** be the directional image and let **A** be the accumulator array, then the algorithm can be sketched as:

```
Reset A;
∀ vector d∈D
  {[x0, y0] = origin(d);
    φ = direction(d);
    σ = modulus(d);
    T = current_template([x0, y0], φ);
    ∀ pixel [x, y]∈T
      {A[x, y] = A[x, y] + σ · weight_T([x, y]);}
  }
```

The high-score **A** cells are good candidates for ellipse centers.

The *direction*(**d**) and the *modulus*(**d**) of the directional elements **d** are calculated off-line as described in Section 2; *current_template*($[x_0, y_0]$, $\varphi$) determines the current template **T** as a function of the direction $\varphi$ of the vector centered in $[x_0, y_0]$. The points $[x_1, y_1]$ and $[x_2, y_2]$ in Fig. 3 are the only two points where an ellipse tangent to **d** in $[x_0, y_0]$ with semi-axes $a$, $b$ could be centered. Since we are interested in all the ellipses whose semi-axes are in the range $[a_{min} \ldots a_{max}, b_{min} \ldots b_{max}]$, we must take into account all the points lying on the two segments determined by the intersection between the straight line defined by $[x_1, y_1]$, $[x_2, y_2]$ and the elliptical annulus **C**. Finally, if we assume a maximum angular variation[2] $\theta$ on the directional information, the geometric locus **T** of the possible centers becomes:

$$\mathbf{T} = \left\{ [x, y] | \rho_r^2 \leqslant \left(\frac{x - x_0}{a}\right)^2 + \left(\frac{y - y_0}{b}\right)^2 \leqslant \rho_e^2, \right.$$

$$\left. \Delta angle\left( arctg\left(\frac{y - y_0}{x - x_0}\right), \phi \right) \leqslant \frac{\theta}{2} \right\}.$$

where $\Delta angle(\alpha, \beta)$: $[-90°, +90°[ \times [-90°, +90°[ \rightarrow [0°, 90°]$ returns the smaller angle determined by the

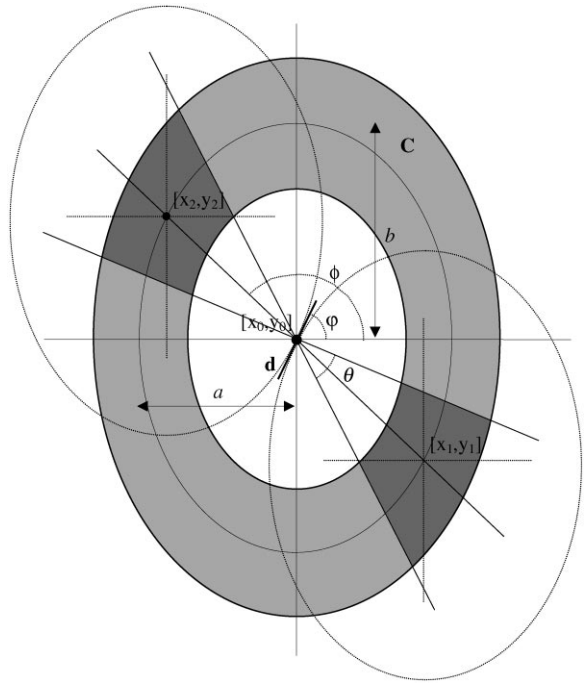


Fig. 3. The template **T** (in dark gray) is constituted by those points which are possible centers of ellipses capable of originating in $[x_0, y_0]$ a vector **d** with direction $\varphi$.

[2] The angular variation $\theta$ is introduced to compensate for both small ellipse rotations and inaccuracies in the computation of $\varphi$.

directions $\alpha$, $\beta$; the angle $\phi$ can be computed as a function of $\varphi$ by deriving the tangent vector expression by the parametric equation of the ellipse:

$$\phi = arctg\left( -\frac{b}{a \cdot tg(\varphi)} \right).$$

The function $weight_T$: $\mathbf{T} \rightarrow [0, 1]$ associates, to each point $[x, y]$ of $\mathbf{T}$, a weight which decreases with the angular distance between the straight line defined by $[x,y]$, $[x_0, y_0]$ and the direction $\phi$:

$$weight_T(x, y) = 1 - \frac{2 \cdot \Delta angle(arctg((y - y_0)/(x - x_0)), \phi)}{\theta}.$$

Fig. 4 shows a representation of a template $\mathbf{T}$ whose elements are associated to gray levels proportional to their weights (the light pixels within the sectors are associated to larger weights).

An efficient implementation of the module AL has been obtained by adopting the following tricks:

- The grid which defines the accumulator array $\mathbf{A}$ has been set equal to that defining $\mathbf{D}$. In particular the granularity used is $7 \times 7$ pixels, that is, the directional image elements are calculated every $7 \times 7$ pixels. Therefore, the size of both the directional image and the accumulator array associated to an $X \times Y$ pixel image is $\lfloor X/7 \rfloor \times \lfloor Y/7 \rfloor$.
- The directions of the elements in $\mathbf{D}$ have been discretized (256 values).
- The templates $\mathbf{T}$ have been pre-computed (in our experiments we chose $a = 34$, $b = 45$, $\rho_r = 0.6$, $\rho_e = 1.2$, $\theta = 30°$); by using relative coordinates with respect to the ellipse center, the number of different templates corresponds to the number of different directions.
- The algorithm has been implemented in integer arithmetic.

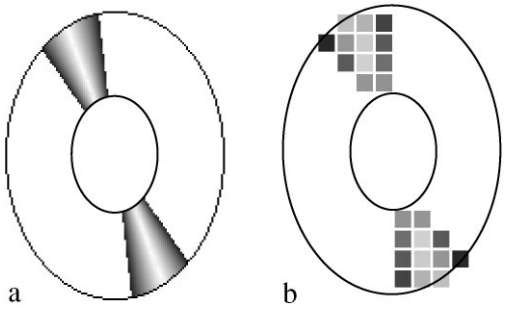


Fig. 4. (a) graphic representation of an ideal template $\mathbf{T}$, and (b) its corresponding discrete version induced by the discretization of $\mathbf{A}$.

Fig. 5 shows an example of ellipse detection as performed by AL: the starting image contains five differently shaped ellipses; the array of accumulator resulting from the Hough transform clearly exhibits a local maximum corresponding to each ellipse center.

## 4. FLFV - fine location and face verification

Different strategies can be adopted in order to improve the location accuracy and to verify whether an elliptical object is really a face or not. Some of the alternatives we explored are reported in the following:

- Improving the ellipse center location through AHT (Adaptive Hough Transform) [41,42] which requires the granularity of the hot accumulator cells to be gradually refined.
- Local optimization of the center $[x_c, y_c]$, of the semiaxes $a$ and $b$ and of the ellipse tilt angle $\xi$ through a local optimization algorithm (Steepest descent, Downhill Simplex method [43], etc.) which searches for the best-fitting ellipse in the parameter space $[x_c, y_c, a, b, \xi]$.
- Position optimization and face verification through the detection of a symmetry axis [44]; a symmetry axis

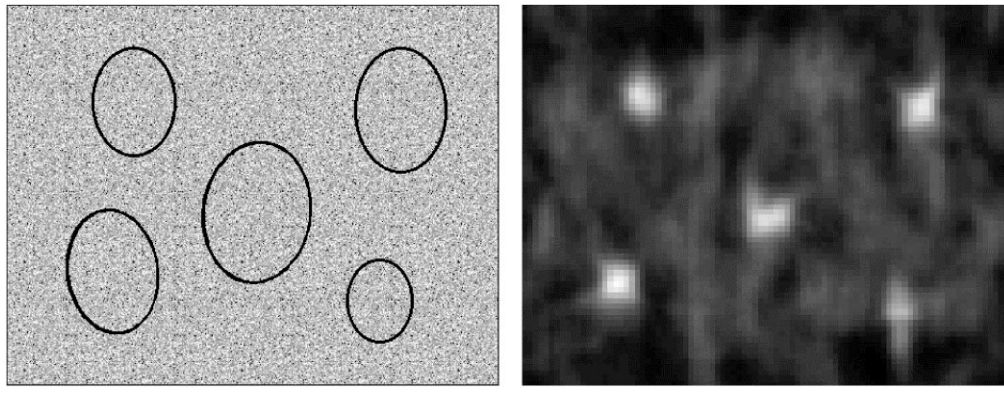

Fig. 5. An artificial image and the corresponding transform.

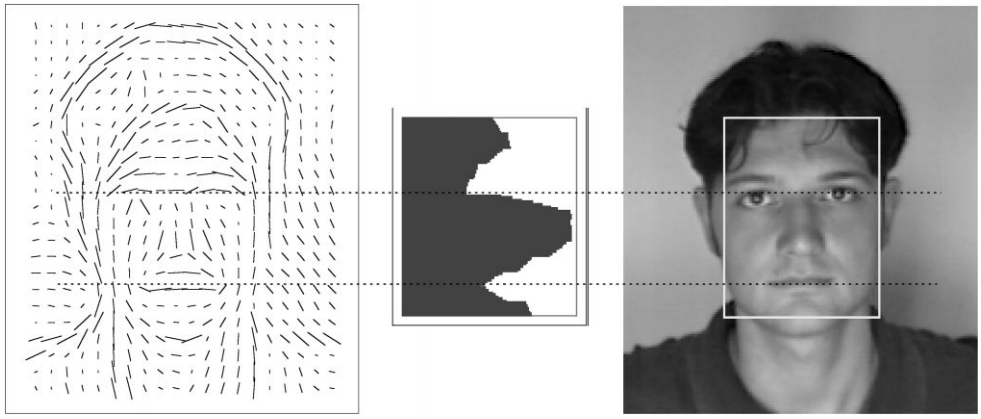

Fig. 6. The figure shows the projection, on the vertical axis, of the directional image vectors belonging to the region delimited by the white rectangle. The local minima, generated by the presence of horizontal vectors in the eye and mouth regions could be used for the registration according to the internal feature positions.

can be extracted both from the original and the directional image.

- Projection of the image portion containing the ellipse on the symmetry axis and on its orthogonal one. Several authors [34,35,45,46] demonstrated that these projections are characterized by local intensity minima in the regions corresponding to the eyes and the mouth. The projection method can also be applied to the directional image: in this case local maxima and minima are present in the eye, nose and mouth regions due to the presence of horizontal and vertical vectors (Fig. 6).

In our preliminary work [47] the fine location and face verification sub-tasks were performed sequentially. From the experimental evidences we argued that better results could be achieved by executing both operations simultaneously. In fact, a precise face location cannot always be obtained by using an elliptical template, since particular physiognomies, hairstyles or illuminations sometimes make the face only coarsely elliptical. On the other hand, the reliability of a face verification method (for example the projection approach in Fig. 6) strongly depends on the face location accuracy. Therefore, in order to confer a greater robustness to the FLFV module, we developed a global approach which attempts to satisfy the two different aims at the same time.

The face is locally searched in a small portion of the directional image by starting from a candidate position resulting from AL; to this purpose, a mask **F**, describing the global aspect of a human face, and defined in terms of directional elements is employed. The local search is performed through an orientation-based correlation between **F** and the **D** portion of interest. Actually, after the approximate location step, **D** is locally refined (i.e. recomputed on a $3 \times 3$ pixel grid) in a neighborhood of the candidate position according to a multiresolution strategy (in the following, for sake of exposition, $\mathbf{D}'$ denotes the new directional image portion).

The mask **F** is defined as a set of directional elements $\mathbf{n}_i$, $i = 1, \ldots, n_{\max}$ each of which is characterized by an *origin*, an *unoriented direction* (in the range $[-90°, 90°[$) and a *modulus*. The element origins are determined by superimposing a grid having the same granularity of $\mathbf{D}'$ ($3 \times 3$) to the template reported in Fig. 7. In particular an element $\mathbf{n}_i$ is defined for each grid point lying within one of the template gray regions (which correspond to the salient face features). Since the template is parametrically defined according to the sizes $a$ and $b$, whereas the grid granularity is fixed, a different number $n_{\max}$ of elements is created varying $a$ and $b$.[3]

As to directions, all the elements within the mouth, eyes and eyebrows regions have horizontal direction, the nose elements have vertical direction and each element belonging to the border region has the direction of the tangent (in that point) to the external ellipse. Each of the seven regions in Fig. 7 has a global modulus which is equally partitioned among its elements. If we consider the common face stereotype, the shape and the position of the nose in Fig. 7 appear a bit unusual; in particular, the nose is too short and moved upward. Actually, this

[3] Actually, the border region constitutes an exception: in this case we fixed a priori the total number $n_b$ of elements within the region and we created an element each $360/n_b$ degree by snapping it to the grid point closest to the border ellipse. In our simulation we chose $n_b = 30$.

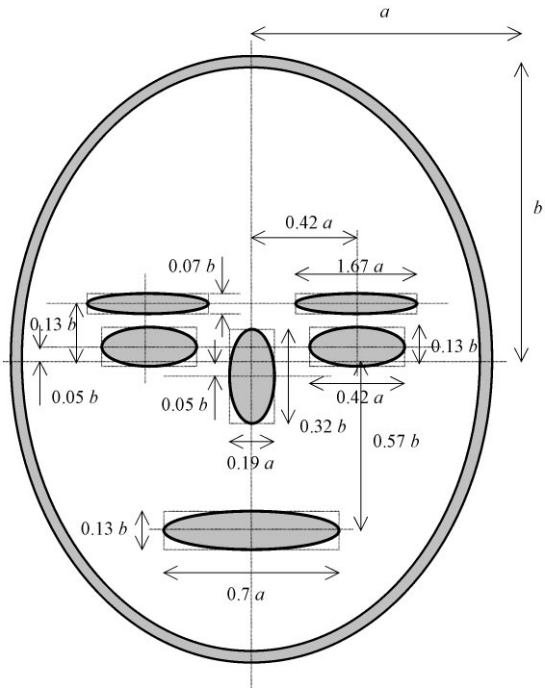


Fig. 7. The template used for the construction of the mask **F** used by the FLVF. The sizes and ratios of the template have been calculated by analyzing some face images and by keeping into account the indications reported by other authors [19,35,36]. The region global moduli are: border = 390, mouth = 110, nose = 60, eyes = 35 + 35, eyebrows = 25 + 25.

choice is motivated by the fact that in the nostril region the edge directions are not strictly vertical but can be rather chaotic.

By discretizing $a$ and $b$ in the range $[a_{\min} \dots a_{\max}, b_{\min} \dots b_{\max}]$ it is possible to pre-compute a set $\mathscr{F} = \mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_m$ of static masks (Fig. 8), where the element positions are given in relative coordinates with respect to the mask center. It should be noted that the set of masks in $\mathscr{F}$ does not explicitly model face rotation; in fact, since in our current application only small head rotation is allowed, a set of "vertical" masks proved to be adequate; anyway, more in general $\mathscr{F}$ could be expanded to cope with tilted faces by including two or more rotated copies of each mask.

The masks in $\mathscr{F}$ allow a correlation degree at each position in $\mathbf{D}'$ to be efficiently computed. Matching directional elements requires an ad hoc correlation (or distance) operator capable of dealing with the discontinuity ($-90^\circ \leftrightarrow 90^\circ$) in the definition of directions (e.g. Ref. [48]). Good results have been obtained in this work with a distance operator defined as an average sum of direction-difference absolute values. The distance between the mask $\mathbf{F}_i \in \mathscr{F}$ and the portion of $\mathbf{D}'$ centered in $[x, y]$ is computed as:

*Distance* ($\mathbf{F}_i$, $\mathbf{D}'$, $[x, y]$)
  {*Distance* = 0; ModuliSum = 0;
    ∀ element $\mathbf{n} \in \mathbf{F}_i$
      {$[x_n, y_n]$ = *origin*(**n**);
        Let $\mathbf{d} \in \mathbf{D}'$ be the element with origin
          $[x, y] + [x_n, y_n]$;
        $\varphi_n$ = *direction*(**n**); $\sigma_n$ = *modulus*(**n**);
        $\varphi_d$ = *direction*(**d**); $\sigma_d$ = *modulus*(**d**);
        Distance = Distance + $\sigma_n \cdot \sigma_d \cdot \Delta$*angle*($\varphi_n$, $\varphi_d$);
        ModuliSum = ModuliSum + $\sigma_n \cdot \sigma_d$;
      }
        Distance = Distance/ModuliSum;
  }

Let $[x_0, y_0]$ be a candidate position as resulting from the module AL, then FLFV determines the face position $[x^*, y^*]$ and the semi-axes $a^*$ and $b^*$ by minimizing the *Distance* function over a discrete state space:

$$d_{\min} = \min_{\substack{\mathbf{F}_i \in \mathscr{F} \\ x \in [x_0 - \Delta x,\ x_0 + \Delta x] \\ y \in [y_0 - \Delta y, y_0 + \Delta y]}} \{\mathit{Distance}(\mathbf{F}_i, \mathbf{D}', [x, y])\}$$

where $\Delta x$ and $\Delta y$ define a neighborhood of $[x_0, y_0]$, and $a^*, b^*$ coincide with the semi-axes of the best-fitting mask. In our simulations we set $\Delta x = \Delta y = 4$ and therefore the total number of states is $12 \times 9 \times 9 = 972$. An efficient implementation (in integer arithmetic) allowed an exhaustive strategy for determining the optimum to be adopted. In case of a large number of states the use of less expensive optimization techniques (Steepest descent, Downhill Simplex method, etc.) should be investigated.

Once the best-fitting position has been determined, the face verification sub-task can be simply performed by comparing the distance $d_{\min}$ with a pre-fixed threshold.

Fig. 9 shows the results of the intermediate steps of a face location example.

## 5. Combining AL and FLFV

Depending on the application requirements, there are several ways of adjusting and combining AL and FLFV modules. Since at this stage our aim is to develop a method capable of efficiently detecting the foreground face, we adopted the functional schema of Fig. 1.b. In particular, the algorithm searches for just one face in the image; it returns the face position $[x_f, y_f]$ and sizes $a_f$, $b_f$, in case of detection, and null otherwise. A pseudo-code version of the whole face detection method is reported:

Compute directional image **D**;
Perform generalized Hough transform;
$[x_0, y_0]$ = get first candidate position on **A**;
$d_f = \infty$;

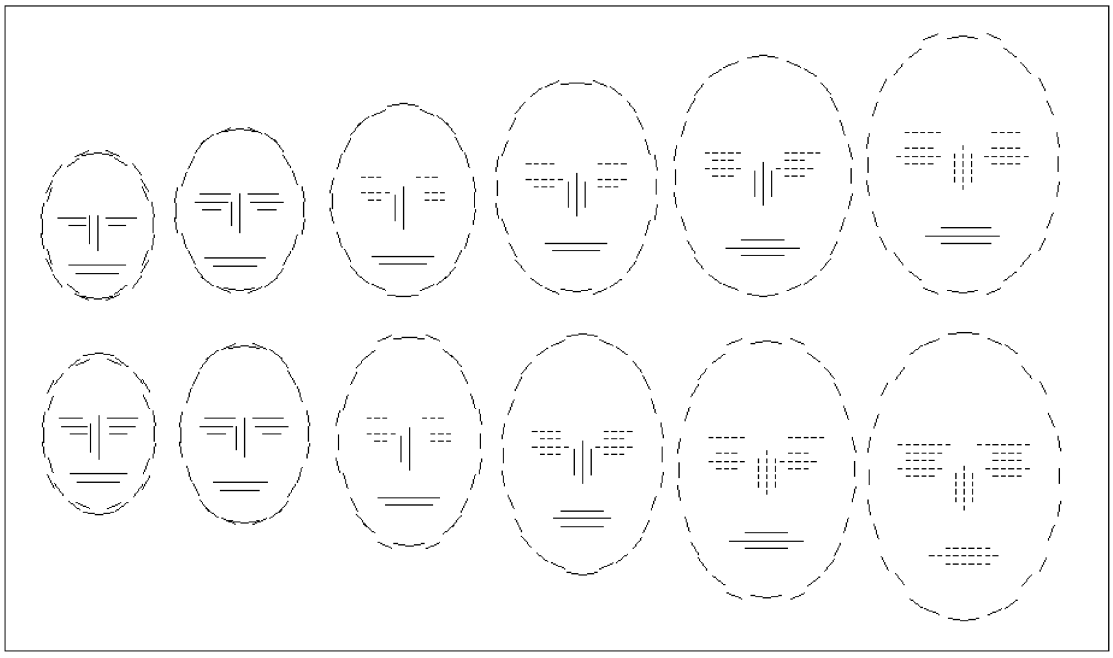

Fig. 8. A graphical representation of the 12 masks constituting the set $\mathscr{F}$ ($m = 12$) used in our experimentation (the length of the mask elements is proportional to their modulus). It should be noted that, since the region global moduli are constant, the element moduli in the smaller masks are generally larger since their number is lower. The only exception concerns the border region where the number of elements is constant and therefore their modulus is independent on the mask size.

While ($[x_0, y_0]$ is not NULL) and $d_f > T_1$
{$\mathbf{D}'$ = Refines the directional image in a neighborhood of $[x_0, y_0]$;

$$d_{\min} = \min_{\substack{\mathbf{F}_i \in \mathscr{F} \\ x \in [x_0 - \Delta x,\ x_0 + \Delta x] \\ y \in [y_0 - \Delta y, y_0 + \Delta y]}} \{Distance(\mathbf{F}_i, \mathbf{D}', [x, y])\}$$

//$[x^*, y^*]$, $a^*$, $b^*$ determine the state corresponding to $d_{\min}$.

if ($d_{\min} < d_f$){$d_f = d_{\min}$;$[x_f, y_f] = [x^*, y^*]$;
$a_f = a^*$; $b_f = b^*$;}
$[x_0, y_0]$ = get next candidate position on $\mathbf{A}$;
}
if ($d_f < T_2$) { return $[x_f, y_f]$, $a_f$, $b_f$; }
else { return null; }

Two different thresholds $T_1$ and $T_2 (T_1 < T_2)$ are used by the algorithm: $T_1$ stops the iterative search process as soon as a good match has been found ($d_f \leqslant T_1$). In case all the candidate positions have been examined (in our experiments the candidate positions are the three best local maxima of $\mathbf{A}$), the iterative process is interrupted, and a valid face is returned if the smallest distance computed is less than $T_2$. Using a pair of thresholds allows a larger number of candidate positions to be analyzed in case the current one is not sufficiently reliable, and at the same time the temporary discarded faces can be reconsidered if no more-likely candidates have been found.

## 6. Experimentation

Experimental results have been produced on a database of 70 images each of which contains at least one human face (Fig. 11). All the images ($384 \times 288$ pixels—256 gray levels) were acquired in some offices and laboratories of our department, under different illuminations (sometimes rather critical: backlighting, semidarkness, ...) and with the subject at different distances from the camera. In 10 images people wear spectacles. The subjects were required to gaze the camera. Each of the 70 images was manually labeled, by indicating with a mouse, the eye positions (**le** and **re**) and the mouth center (**mc**) of the face in foreground (Fig. 10). The following formulae were used to derive the features (center **c** and semi-axes $a$, $b$) of an ellipse approximating a face:

$$a = 1.1 \cdot \|\mathbf{le\text{-}re}\|;$$

$$b = 1.3 \cdot (\|\mathbf{le\text{-}mc}\| + \|\mathbf{re\text{-}mc}\|)/2;$$

$$\mathbf{c} = [x_c, y_c];\ \mathbf{mc} = [x_{mc}, y_{mc}];\ x_c = x_{mc} + \cos\xi \cdot 0.54 \cdot b,$$

$$y_c = y_{mc} + \sin\xi \cdot 0.53 \cdot b$$

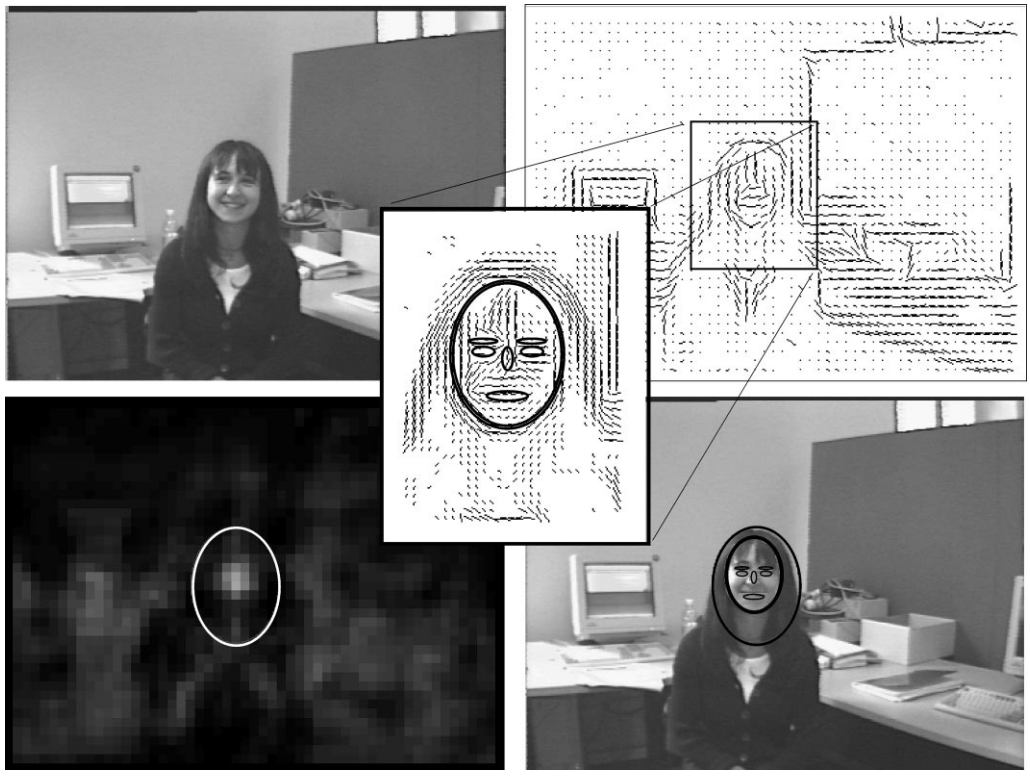

Fig. 9. The results of the intermediate steps in a face location example: directional image computation, generalized Hough transform and selection of the best candidate position (module AL), directional image local refining, determination of the best-fitting mask and position (module FLFV).

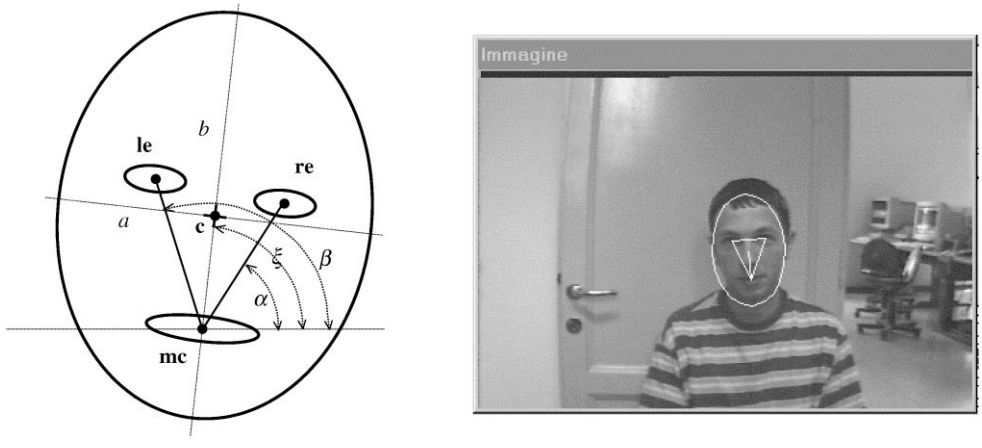


Fig. 10. The geometric construction used for defining the face ellipse (**c**, *a*, *b*) starting from the eye centers **le**, **re** and the mouth center **mc**.

where $\xi = (\alpha + \beta)/2$, with $\alpha = angle(\mathbf{ec}, \mathbf{mc})$ and $\beta = angle(\mathbf{el}, \mathbf{mc})$

*Module AL*: The approximate location module detected the correct face position as global maximum of A (i.e. first candidate position) in 65 cases (92.86%). In the remaining five images the face position was detected as the second or third candidate position. Fig. 12 shows some images and the corresponding transforms, where the

Fig. 11. The database used in our experimentation.

global maximum determined by the elliptical face shape is well visible (an ellipse of semi-axes $a$ and $b$ was superimposed on each maximum). Fig. 13 reports an example where the transform global maximum does not coincide with the face position, but it is determined by a different circular object (an apple within a poster). This false alarm, as can be seen in Fig. 14 (first row, third column), is removed by the module FLFV.

*The whole approach: AL + FLFV*: The whole face detection algorithm has been applied to the 70 images: in 69 cases the foreground face was correctly detected; only in one case no faces were found. Eleven false alarms (in 7 images) generated by AL were correctly discarded by FLFV: if we consider the 69 images, we can conclude that on average about 0.16 false alarms per image are processed and correctly discarded by FLFV before the correct face position is detected.

Fig. 13. A false alarm caused by an elliptical object in the scene. In any case, the local maximum determined by the face is well evident and constitutes the second best choice.

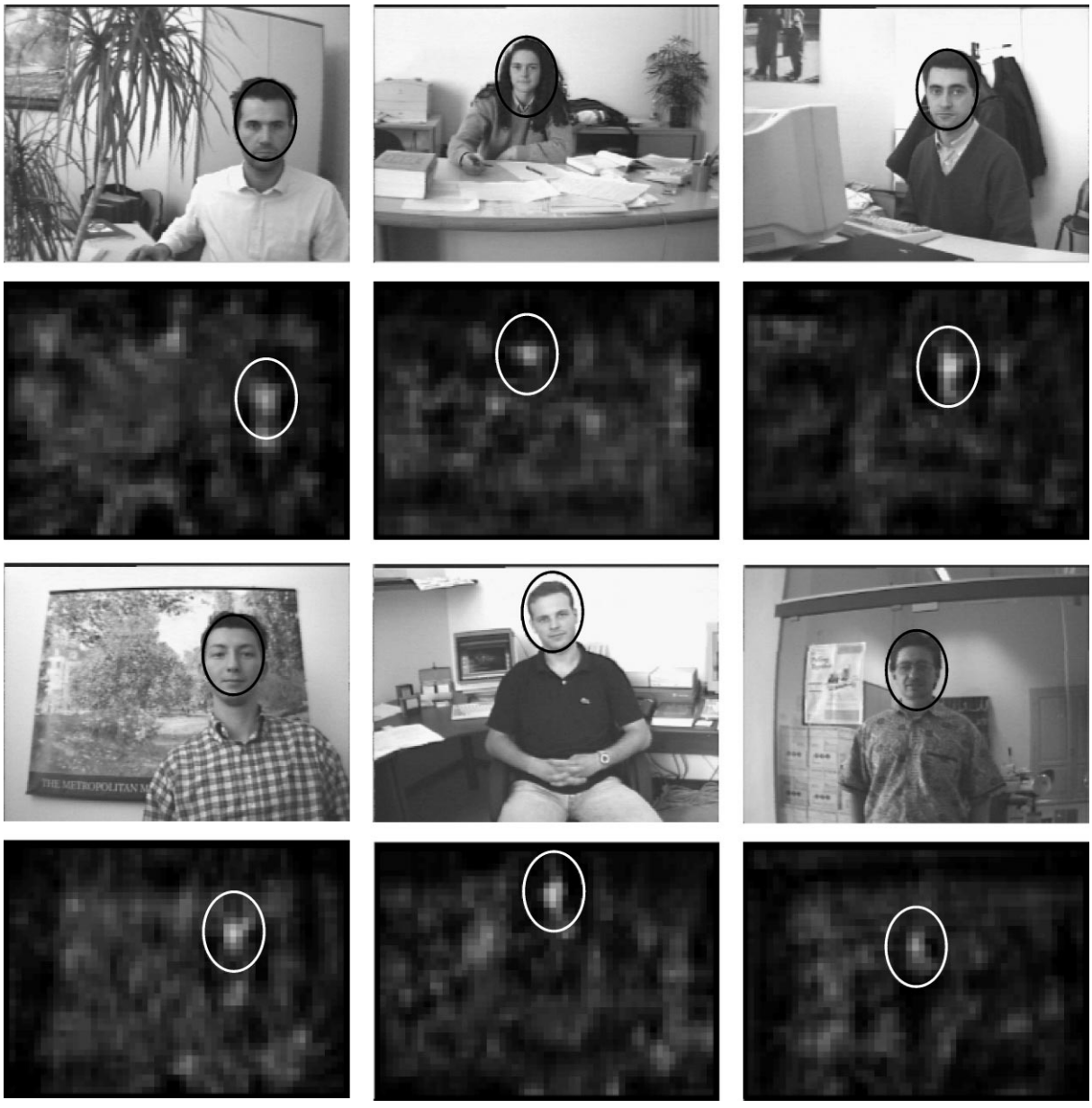

Fig. 12. Some images and the corresponding transforms as computed by the module AL.

Fig. 14. Some examples of face location. The thin-border ellipses denote AL outputs, whereas the best-fitting masks are reported as FLFV outputs.

The location accuracy was measured by considering, for the 69 faces correctly detected, the percentage error on the center position (c_err) and on the semi-axes length (a_err, b_err). Let $\mathbf{c}$, $a$ and $b$ be the center and the semi-axes of the foreground face in an image (as manually indicated during the database labeling) and let $[x_f, y_f]$, $a_f$ and $b_f$ be the center and the semi-axes returned by the algorithm, then:

$$c\_\text{err} = \|[x_f, y_f] - c\|/a + b$$
$$a\_\text{err} = |a_f - a|/a$$
$$b\_\text{err} = |b_f - b|/b$$

The average values obtained: c_err = 8.70%, a_err = 8.66%, b_err = 10.47% denote a good location accuracy. It is worth remarking that a face cannot be exactly described by an ellipse using the geometric model of Fig. 11 and therefore the errors measured should not be exclusively attributed to the detection algorithm. In fact, a qualitative analysis of the results confirmed that in most cases the mask positioned by the module FLFV perfectly fit the underlying face, especially in the eye and the mouth regions.

Fig. 14 shows some examples; the thin-border ellipses denote AL outputs; the false alarms produced by AL can be easily noted since no masks are associated to the corresponding ellipses. Fig. 15 shows the only images which produced a miss detection error, probably due to the lateral illumination which hides a significant part of the face, thus inducing the module AL to poorly estimate the face position.

The method discussed in this paper is capable of processing a $384 \times 288$ gray-scale image in 0.078 s on Hardware Intel Pentium II 266 MHz: in particular, 0.031 s are necessary for the directional image computation, 0.007 s for the Hough transform, 0.015 s for the directional image refining and 0.019 s for the final template matching. By increasing the last two terms by 16% (in order to take into account the average number of false alarms generated by AL) and by summing all the contributes, we obtain an average processing time of 0.078 s which corresponds approximately to 13 frames per second. We disregard the computation of the 256 templates for the Hough transform and the computation of the masks in $\mathscr{F}$, since they are performed off-line in 0.086 s.

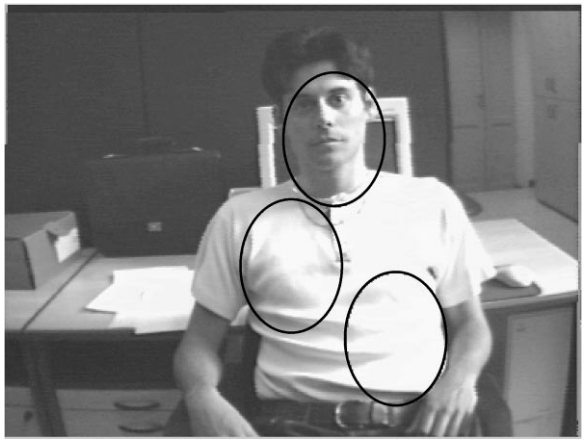

Fig. 15. The only missed detection error on our database: none of the three candidate positions have been accepted by the module FLFV; the upper candidate position is not enough close to the face to allow FLFV to converge.

## 7. Conclusions

This work proposes a two-stage approach to face location on the gray scale static images with complex backgrounds. Both the modules operate on the elements constituting the directional image, which has been proved to be very effective in providing reliable information even in the presence of critical illumination and semidarkness.

The approximate location module searches for the most likely positions in the image by means of a particular implementation of the generalized Hough transform. Great computational saving is obtained with respect to a correlation-based technique. In fact, in the former case just one directional image scan allows the "hot" positions to be extracted, whereas in the latter several elliptic templates (resembling the different elliptical shapes and sizes) should be shifted everywhere on the directional image to discover the high correlation points. Numerically, if $n_T$ is the average number of cells updated by the current template $\mathbf{T}$ and $n_D$ is the number of elements in $\mathbf{D}$ (in our implementation $n_T \approx 20$ and $n_D \approx 2000$) $O(n_T \cdot n_D)$ operations are necessary for calculating the Hough transform. If we assume that $n_T$ also denotes the average number of elements in a hypothetical elliptical template, then $O(m \cdot n_T \cdot n_D)$ operations are necessary for the template matching with $m$ templates. According to our parameters choice, a reasonable value for $m$ is 12 (please, refer to Fig. 8); hence the GHT implementation gives a considerable saving with respect to a correlation-based approach.

The fine location and face verification module analyses small-refined portions of directional image attempting to detect the exact position and size of a face together with a confidence value about its presence. This is performed by means of a set of masks resembling the human face stereotype; since the masks are defined in terms of directions, their matching is robust and very little biased by light conditioning.

Very good results have been achieved both in terms of face detection rate (just one missed detection on 70 images) and efficiency (about 13 images/sec. can be processed on Hardware Intel Pentium II 266 MHz running Window95™). Furthermore, a more efficient version

could be obtained by applying an appropriate code optimization.

As to future research, we are going to investigate how to combine the modules AL and FLFV with predictive analysis techniques (for example Kalman filtering) in the context of face tracking applications. Furthermore, we are gathering a larger database of images which will allow us to more precisely characterize the performance of our approach and to draw a ROC curve showing the false detection/missed detection tradeoff.

## References


[1] R. Chellappa, S. Sirohey, C.L. Wilson, C.S. Barnes, Human and machine recognition of faces: a survey, Tech. Report CS-TR-3339, Computer Vision Laboratory, University of Maryland, 1994.

[2] I. Craw, H. Ellis, J.R. Lishman, Automatic extraction of face-features, Pattern Recognition Lett. 5 (1987) 183–187.

[3] G. Yang, T.S. Huang, Human face detection in a complex background, Pattern Recognition 27 (1994) 53–63.

[4] I. Craw, D. Tock, A. Bennet, Finding face features, Proceedings of ECCV, 1992.

[5] A. Yuille, D. Cohen, P. Hallinan, Facial features extraction by deformable templates, Tech. Report 88-2, Harward Robotics Laboratory, 1988.

[6] C. Huang, C. Chen, Human facial feature extraction for face interpretation and recognition, Pattern Recognition 25 (1992) 1435–1444.

[7] G. Chow, X. Li, Towards a system for automatic facial feature detection, Pattern Recognition 26 (1993) 1739–1755.

[8] K. Lam, H. Yan, Locating and extracting the eye in human face images, Pattern Recognition 29 (1996) 771–779.

[9] A. Lanitis, C.J. Taylor, T.F. Cootes, T. Ahmed, Automatic interpretation of human faces and hand gesture using flexible models, Proceedings of International Workshop on Automatic Face and Gesture Recognition, Zurich, 1995, pp. 98–103.

[10] R. Funayama, N. Yokoya, H. Iwasa, H. Takemura, Facial component extraction by cooperative active nets with global constraints, Proceedings of the 13th ICPR, v. B, 1996, pp. 300–304.

[11] S.R. Gunn, M.S. Nixon, Snake head boundary extraction using global and local energy minimisation, Proceedings of the 13th ICPR, v. B, 1996, pp. 581–585.

[12] S.A. Sirohey, Human face segmentation and identification, Tech. Report CAR-TR-695, Center for Automation Research, University of Maryland, 1993.

[13] A. Jacquin, A. Eleftheriadis, Automatic location tracking of faces and facial features in video sequences, Proceedings of the International Workshop on Automatic Face and Gesture Recognition, 1995, pp. 142–147.

[14] R. Herpers, H. Kattner, H. Rodax, G. Sommer, GAZE: an attentive processing strategy to detect and analyze the prominent facial regions, Proceedings of the International Workshop on Automatic Face and Gesture Recognition, 1995, pp. 214–220.

[15] X. Li, N. Roeder, Face contour extraction from front-view images, Pattern Recognition 28 (1995) 1167–1179.

[16] V. Govindaraju, S.N. Srihari, D.B. Sher, A computational model for face location, Proceedings of the 3rd ICCV, 1990, pp. 718–721.

[17] H.P. Graf, T. Chen, E. Petajan, E. Cosatto, Locating faces and facial parts, Proceedings of the International Workshop on Automatic Face and Gesture Recognition, 1995, pp. 41–46.

[18] M.C. Burl, T.K. Leung, P. Perona, Face localization via shape statistics, Proceedings of the International Workshop on Automatic Face and Gesture Recognition, 1995, pp. 154–159.

[19] S. Jeng, H.M. Liao, Y. Liu, M. Chern, An efficient approach for facial feature detection using geometrical face model, Proceedings of the 13th ICPR, v. C, 1996, pp. 426–430.

[20] B. Moghaddam, A. Pentland, Maximum likelihood detection of faces and hands, Proceedings of the International Workshop on Automatic Face and Gesture Recognition, 1995, pp. 122–128.

[21] M.S. Lew, N. Huijsmans, Information theory and face detection, Proceedings of the 13th ICPR, v. C, 1996, pp. 601–605.

[22] A.J. Colmenarez, Face and facial feature detection with information-based maximum discrimination, Nato Asi Conference on Faces, 1997.

[23] D. Valentin, H. Abdi, A.J. Otoole, G.W. Cottrell, Connectionist models of face processing—a survey, Pattern Recognition 27 (1994) 1209–1230.

[24] G. Burel, D. Carel, Detection and localization of faces on digital images, Pattern Recognition Lett. 15 (10) (1994) 963–967.

[25] K. Sung, T. Poggio, Example-based learning for view-based human face detection, A.I. Memo 1521, CBCL Paper 112, MIT, 1994.

[26] H.A. Rowley, S. Baluja, T. Kanade, Human face detection in visual scenes, Tech. Report CMU-CS-95-158R, Carnegie Mellon University, 1995.

[27] B. Schiele, A. Waibel, Gaze tracking based on face color, Proceedings of the International Workshop on Automatic Face and Gesture Recognition, 1995, pp. 344–349.

[28] N. Intrator, D. Reisfeld, Y. Yeshurun, Extraction of facial features for recognition using neural networks, Proceedings of the International Workshop on Automatic Face and Gesture Recognition, 1995, pp. 260–265.

[29] R. Feraud, A conditional ensemble of experts applied to face detection, Nato Asi Conference on Faces, 1997.

[30] F.F. Soulie, Connectionist methods for human face processing, Nato Asi Conference on Faces, 1997.

[31] H. Wu, Q. Chen, M. Yachida, An application of fuzzy theory: face detection, Proceedings of the International Workshop on Automatic Face and Gesture Recognition, 1995, pp. 314–319.

[32] Y. Dai, Y. Nakano, Face-texture model based on SGLD and its application in face detection in a color scene, Pattern Recognition 29 (1996) 1007–1017.

[33] C.H. Lee, J.S. Kim, K.H. Park, Automatic face location in a complex background using motion and color information, Pattern Recognition 29 (1996) 1877–1889.

[34] K. Sobottka, I. Pitas, Extraction of facial regions and features using color and shape information, Proceedings of the 13th ICPR, v. C, 1996, pp. 421–425.

[35] H. Sako, A.V.W. Smith, Real-time expression recognition based on features position and dimension, Proceedings of the 13th ICPR, v. C, 1996, pp. 643–648.
[36] E. Saber, A. Murat Tekalp, Face detection and facial feature extraction using color, shape and symmetry-based cost functions, Proceedings of the 13th ICPR, v. C, 1996, pp. 654–657.
[37] B. Leroy, I.L. Herlin, L.D. Cohen, Face identification by deformation measure, Proceedings of the 13th ICPR, v. C, 1996, pp. 633–637.
[38] M.J. Donahue, S.I. Rokhlin, On the use of level curves in image analysis, Image Understanding 57 (1993) 185–203.
[39] P. Seitz, M. Bichsel, The digital doorkeeper—automatic face recognition with the computer', Proceedings of the 25th IEEE Carnahan Conference on Security Technology, 1991.
[40] D.H. Ballard, Generalizing the Hough transform to detect arbitrary shapes, Pattern Recognition 3 (1981) 110–122.
[41] L.S. Davis, Hierarchical generalized Hough transform and line segment based generalized Hough transforms, Pattern Recognition 15 (1982) 277–285.
[42] J. Illingworth, J. Kittler, The adaptive Hough transform, IEEE Trans. Pattern Anal. Mach. Intell. 9 (1987) 690–697.
[43] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C, Cambridge University Press, Cambridge, 1992.
[44] D. Reisfeld, H. Wolfson, Y. Yeshurun, Detection of interest points using symmetry, Proceedings of the 3rd ICCV, 1990, pp. 62–65.
[45] R. Brunelli, T. Poggio, Face recognition: features versus templates, IEEE Trans. Pattern Anal. Mach. Intell. 15 (1993) 1042–1052.
[46] H. Wu, Q. Chen, M. Yachida, Facial Features Extraction and Face Verification, Proceedings of the 13th ICPR, v. C, 1996, pp. 484–488.
[47] D. Maio, D. Maltoni, Fast face location in complex backgrounds, Nato Asi Conference on Faces, 1997.
[48] A. Crouzil, L. Massip-Pailhes, S. Castan, A New Correlation Criterion Based on Gradient Fields Similarity, Proceedings of the 13th ICPR, v. A, 1996, pp. 632–636.

**About the Author**—DARIO MAIO is Full Professor at the Computer Science Department, University of Bologna, Italy. He has published in the fields of distributed computer systems, computer performance evaluation, database design, information systems, neural networks, biometric systems, autonomous agents. Before joining the Computer Science Department, he received a fellowship from the C.N.R. (Italian National Research Council) for participation to the Air Traffic Control Project. He received the degree in Electronic Engineering from the University of Bologna in 1975. He is a IEEE member. He is with CSITE - C.N.R. and with DEIS; he teaches database and information systems at the Computer Science Dept., Cesena.

**About the Author**—DAVIDE MALTONI is an Associate Researcher at the Computer Science Department, University of Bologna, Italy. He received the degree in Computer Science from the University of Bologna, Italy, in 1993. In 1998 he received his Ph.D. in Computer Science and Electronic Engineering at DEIS, University of Bologna, with research theme "Biometric Systems". His research interests also include autonomous agents, pattern recognition and neural nets. He is an IAPR member.