COMPUTER VISION AND IMAGE UNDERSTANDING
Vol. 71, No. 3, September, pp. 413–430, 1998
ARTICLE NO. IV970640

NOTE

Preprocessing of Face Images: Detection of Features and Pose Normalization*

Daniel Reisfeld† and Yehezkel Yeshurun

Computer Science Department, Tel Aviv University, Israel

Received February 20, 1996; accepted May 14, 1997

The reliability, speed, and complexity of virtually any face recognition system are substantially improved if the location and the scale of the faces are known. We propose a method for automatic and robust detection of the eyes and mouth using the context free *generalized symmetry transform* and knowledge of faces. The features are extracted from the image of the intensities gradients and are then used to normalize the face images. We show that a normalization procedure based on affine transformations whose anchor points are the locations of the eyes and mouth substantially increases the effectiveness of general purpose classification techniques in face recognition.

Other normalization procedures for avoiding the effect of background and varying light conditions are proved to be instrumental as well.

1. INTRODUCTION

Construction of automatic face processing systems is valuable for many applications in computer vision and computer graphics, as well as for understanding the principles underlying biological face recognition mechanisms. Surveys of face recognition systems are found in [2, 28, 35].

Some researchers have tried to extract facial features based on general low-level vision mechanisms or even directly from the unprocessed pixels. These include the use of principle component analysis [19, 41] or connectionist approaches with pixels as features [5, 36, 38] or with edges [25]. These methods are sensitive to changes in the background and of the viewpoint which give rise to the *curse of dimensionality* [8]. We show that the performance of standard classifiers markedly improves if an appropriate preprocessing procedure is applied.

An alternative is to use facial models [10, 20, 23, 27, 40, 42]. Locating facial features is indispensable in order to find the model parameters. Finding features is done by correlation [1] (possibly in multiscale [4]) or by using Gabor or Gaussian derivative filters combined with graph matching algorithms [21, 24] or other global consideration [3, 17, 20]. Semantic constraints are sometimes used too [6, 13]. The search space of these techniques is drastically reduced if the faces are normalized as described in the following.

Learning techniques for locating facial features include connectionist methods [22] and principle component analysis [26]. These processes are similar to the recognition of a whole face but are done in smaller scales. Again, having normalized faces simplifies the task of locating the features dramatically.

Iterative methods for locating facial features such as by deformable templates [44] depend on the initial value of the parameters. Once again, an estimate of the features location may facilitate the robustness and speed of these iterative methods. These procedures as well as the learning ones can be used to improve the results obtained by the procedure described below.

There is a biological motivation for locating the eyes and mouth and for doing that simultaneously. Of the internal features, eyes are most accurately remembered, mouths are better remembered than noses, and ears are particularly poorly perceived [7, 11, 14]. All these features are perceived much better when they are part of the whole face [39].

We proposed a method for automatic and robust detection of the eyes and mouth using the context free *generalized symmetry transform* [31, 32], followed by using plausible geometrical constrains.

These features are then used to normalize the face images. We show that a normalization procedure based on affine transformations whose anchor points are the locations of the eyes and mouth substantially increases the effectiveness of general purpose classification techniques in face recognition.

Other normalization procedures for avoiding the effect of background and varying light conditions are proved to be instrumental as well.

We presented preliminary results for localization of facial features and face normalization in [33]. We demonstrated its effectiveness for classification in [9, 15].

* Supported by Grant 4478293 by the French-Israeli MOST R&D.
† E-mail: reisfeld@math.tau.ac.il.

1077-3142/98 $25.00

2. LOCATING FACIAL FEATURES

The *generalized symmetry transform* [31, 32] is an interest operator which is motivated by the biological mechanisms of *attention* and *fixation*. We suggest using it as the first stage in face recognition as well as in other vision tasks. The transform is inspired by the intuitive notion of symmetry. It assigns a *symmetry magnitude* and a *symmetry orientation* to every pixel. The input to the transform is an edge map—the gradients of intensity at each pixel—and its output is *symmetry map*, which is a new kind of an edge map, where the magnitude and orientation of an edge depend on the symmetry associated with the pixel. Strong symmetry edges are natural interest points, while linked lines are *symmetry axes*. The symmetry transform effectively locates an interest point of an image in real time and can be incorporated in passive, as well as active, visual systems. The results of its operation are consistent with psychophysical evidence concerning symmetry as well as evidence concerning fixation points.

Although the symmetry transform is a general purpose low level transform, it effectively locates interest points in images without using *a priori* knowledge of the world. This may seem a contradiction in terms since interest is context dependent almost by definition. This criticism is further supported by the fact that the distribution of fixation points on an object changes depending on the purpose of the observer [43]. In particular, a face recognition system should take advantage of the fact that it is usually confronted with face images and the symmetry map can be further processed to locate faces and facial features. It seems that humans are highly sensitive to face stimuli and they tend to see faces in many objects that possess some elementary geometrical relations typical to faces. We will show how to turn the context independent generalized symmetry transform into a facial feature detector using some knowledge of faces.

2.1. The Generalized Symmetry Transform

The input of the symmetry transform is an edge map. Edge detectors are one of the most studied issues in computer vision, a fact that stems from their poor performance relative to edges that artists draw. There are many possibilities for edge detections suitable for producing the symmetry transform input. We use a very simple edge detector which is practically a gradient map of the original image. Therefore, a better edge detector may only increase the effectiveness of the symmetry transform and the results we obtained are due to the properties of the symmetry transform rather than being hidden in the edge detector. Moreover, it produces redundant edges rather than misses true edges. Thus, the robustness of the symmetry transform is enhanced. We briefly outline the details of the edge detector to enable reproduction of our results.

The continuous one dimensional Gaussian function is defined as

$$G_\sigma(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-t^2}{\sigma^2}}.$$

In practice we sample these functions and make them compact by neglecting their values for $|t| > 3\sigma$. We shall omit the subscript σ and we use the gradient of a blurred image with blurring parameter σ,

$$\nabla I(x, y) = \left(\frac{d}{dx} G(x)G(y) * I(x, y),\ G(x)\frac{d}{dy} G(y) * I(x, y) \right),$$

where $*$ denotes convolution.

We now consider the output of the edge detector, $E(p)$ for each $p = (x, y)$, as a pair of maps

$$E(p) = E(x, y) = \begin{pmatrix} R(p) \\ \theta(p) \end{pmatrix} = \begin{pmatrix} \log(1 + \|\nabla I(p)\|) \\ \arg\nabla(p) \end{pmatrix}.$$

Given two points, p, q such that $p - q = (\Delta x, \Delta y)$, we define the phase of the gradient at p with respect to q as with respect to point $q = (u, v)$:

$$\vartheta_q(p) = \theta(p) - \arg(\Delta x, \Delta y).$$

We define a phase weight function for each two points p, q:

$$\omega(p, q) = (1 - \cos(\vartheta_q(p) + \vartheta_p(q)))(1 - \cos(\vartheta_q(p) - \vartheta_p(q))).$$

The phase weight function is composed of two terms. The first term of ω, $(1 - \cos(\vartheta_q(p) + \vartheta_p(q)))$, peaks when the gradients at p and q are oriented in the same direction toward each other $(\vartheta_q(p) = \pi - \vartheta_p(q))$. This is consistent with the intuitive notion of symmetry. This expression decreases continuously as the situation deviates from the ideal one. The second term of ω, $(1 - \cos\cos(\vartheta_q(p) - \cos\vartheta_p(q)))$, is introduced since the first term attains its maximum whenever $\vartheta_q(p) + \vartheta_p(q) = \pi$. This includes the case $\vartheta_q(p) = \vartheta_p(q) = \pi/2$, which occurs on a straight edge, which we do not regard as interesting. The current expression compensates for this situation.

We define a distance weight function using the general two dimensional Gaussian, $G_{\sigma_x\sigma_y\theta}$:

$$D_{\sigma_x\sigma_y\theta}(p, q) = G_{\sigma_x\sigma_y\theta}(\|p - q\|).$$

To simplify the notation we shall use

$$D(p, q) = D_{\sigma_x\sigma_y\theta}(p, q).$$

The distance weight function reflects the local nature of the operations we define below. Different values for σ imply different scales, thus enabling convenient implementation of multiresolution schemes. Gaussians with elliptic level lines are useful when the transform is applied as a feature detector of elliptic regions such as eyes in human faces.

The *contribution* of the points p and q to the symmetry measure is defined as

$$C(p, q) = D(p, q)\omega(p, q)R(p)R(q).$$

The term $R(p)R(q)$ is large when there is a large correlation between two large gradients. We use gradients rather than intensities since we are mainly interested in edges that relate to object borders. For instance, a uniform intensity wall is highly symmetric but probably not very interesting. In natural scenes we prefer to use the logarithm of the magnitude instead of the magnitude itself, since it reduces the differences between high gradients, and therefore the correlation measure is less sensitive to very strong edges.

Recall that we cut a Gaussian after 3σ. The set of pairs of points that contribute to the symmetry at a point o is

$$\Gamma_\sigma(o) = \left\{ (p, q) \;\middle|\; \frac{p+q}{2} = o \bigwedge \|o - p\| < 3\sigma \right\}.$$

As usual we omit the σ subscript and simply use $\Gamma(o)$.

The *symmetry magnitude* or *isotropic symmetry* at each point is defined as

$$M(o) = \sum_{(p,q)\in\Gamma(o)} C(p, q).$$

Note that the same measure is achieved when the gradients are oriented toward each other or against each other. The first situation corresponds to symmetry within a dark object on a light background, and the second corresponds to symmetry within a light object on a dark background. For this purpose we define a partial ordering between points as

$$(x, y) < (u, v) \longleftrightarrow x < u \bigvee x = u \bigwedge y < v.$$

q and $p < q$ contributes to the *dark symmetry* at $\frac{1}{2}(p+q)$ only if their gradients are oriented toward each other:

$$\Gamma(o) = \left\{ (p, q) \in \Gamma(o) \,|\, p < q \bigwedge \cos\vartheta_q(p) \geq 0 \bigwedge \cos\vartheta_p(q) \leq 0 \right\}.$$

We can define the *dark symmetry magnitude* or *dark isotropic symmetry* by substituting Γ with $\hat{\Gamma}$. Similarly, the *bright symmetry* can be defined for symmetry within a bright object. Note that the dark symmetry has a faster implementation than the full symmetry.

Let $P(o)$ and $Q(o)$ be the two points such that

$$C(P, Q) = \max_{(p,q)\in\Gamma(o)} C(p, q).$$

Then, we define the *symmetry direction* at point o as

$$\alpha(o) = \frac{\theta(P) + \theta(Q)}{2}.$$

The *symmetry* at point p is now defined as

$$S(o) = (M(o), \alpha(o)).$$

Note that the transform that we have defined detects reflectional symmetry. It is invariant under two dimensional rotation and translation. The transform is robust also in respect to scaling as discussed later. Moreover, it is quite effective in detecting skewed symmetry [18] as well. Since skewed symmetry results in an affine transformation of the two dimensional picture, the midpoint of a segment and parallelism are preserved. Thus the location of the symmetry edge is preserved under skewed symmetry; however, the direction of the symmetry edge should not be the average of the two directions as discussed in [32].

Sometimes it is necessary to detect points that are highly symmetric in multiple distinct orientations rather than in a principle one. We define such a symmetry as *radial symmetry*, $RS(o)$, and its value can be evaluated using the formula:

$$RS(o) = \sum_{(p,q)\in\Gamma(o)} C_{(p,q)} \sin^2\left(\frac{\theta(p) + \theta(q)}{2} - \alpha(o) \right).$$

This expression emphasizes contribution in the directions which are perpendicular to the main symmetry direction and attains its maximum in a point that is surrounded by edges. Notice that due to the continuous nature of the transform, the radial symmetry is not sensitive to gaps in the contour that surround the point o and does not require this contour to be uninterrupted.

The basic symmetry maps, the dark isotropic symmetry, and the dark radial symmetry, are demonstrated in Fig. 1. Both maps were computed with Gaussian distance weight function with $\sigma_x = 1$ and $\sigma_y = \frac{2}{3}$. These numbers are typical. We show in the following that the sensitivity to these numbers is small. The symmetry orientation map is not shown in a figure, but it is used for the secondary symmetry operations.

2.2. *Secondary Symmetry Operations*

Applying the symmetry transforms results in a triplet of maps: isotropic symmetry, M; radial symmetry, RS; and symmetry phase, α. Each triplet corresponds to a certain resolution and to a type of symmetry—full, dark, or bright. These maps can be further processed to produce more maps that are described in the following.

2.2.1. Projection. The *projection* of a map, X (isotropic symmetry, radial symmetry, or any other map that is described in this section) in a certain direction, ψ, is aimed to enhance the activity along this direction:

$$\mathrm{Proj}(X, \psi)(o) = X(o)\cos(\alpha(o) - \psi).$$

The eyes and mouth orientation is usually horizontal. Figure 2 demonstrates that the projection of the symmetry along the horizon enhances the activity along the eyes and mouth. If faces are expected to be tilted, it is useful to repeat this operation along other axes as well.

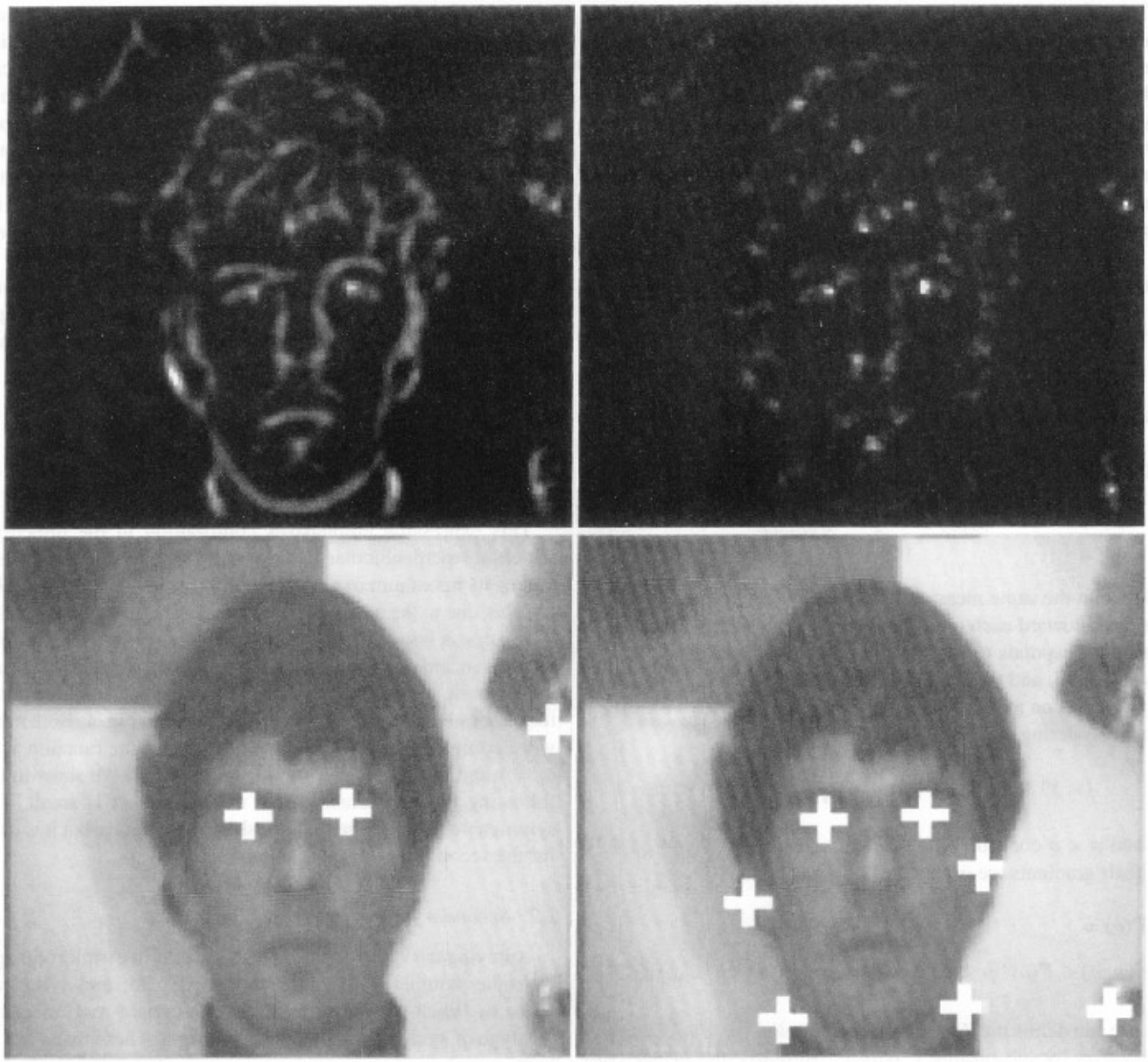

FIG. 1. The basic symmetry maps. (Top) Dark isotropic symmetry (left) and dark radial symmetry (right). (Bottom) The highest peaks of each map are marked by crosses on the original image. The face image is from [41].

2.2.2. Nonmaximal suppression. We denote the neighboring points of o whose symmetry orientations are perpendicular to ϕ by $N(o, \phi)$.

The *nonmaximal suppression* of a map, X, is namely the suppression of the activities of points which are not maximal in the direction perpendicular to their gradient:

$$\mathrm{NMS}(X)(o) = \begin{cases} X(o) & \text{if } \forall p \in N(0, \alpha(o)),\ X(o) > X(p) \\ 0 & \text{otherwise} \end{cases}$$

This process result in thin symmetry lines.

Nonmaximal suppression of the symmetry and its projection result in fine symmetry lines as demonstrated in Fig. 3.

2.2.3. Linking. Symmetry points can be grouped into symmetry lines by various methods. Straight symmetry lines can be found using the Hough transform. Symmetry clusters can be found and characterized by various techniques such as principle component analysis. We successfully used the following edge linking technique.

First we describe the *link left* procedure using a dynamic programming procedure. This procedure involves three parameters: a global threshold, $T \in [0, 1]$, a local threshold, $t \in [0, 1]$, and a

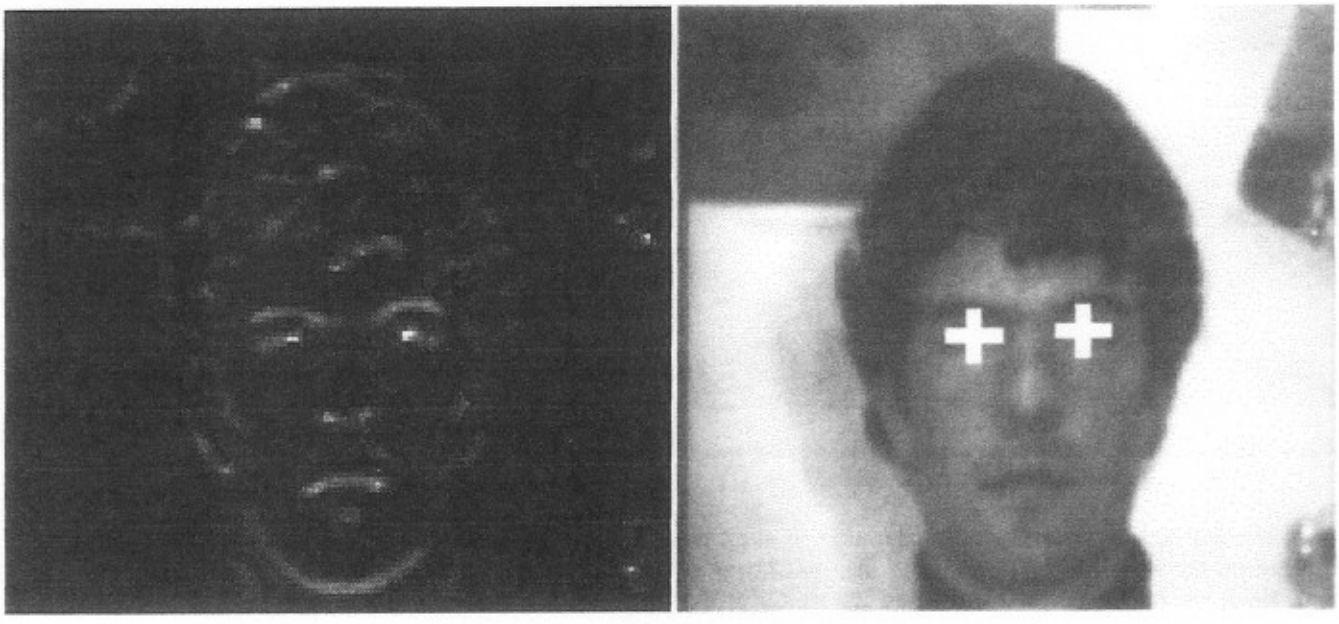

FIG. 2. Projection of the dark isotropic symmetry on the horizon (left) and its highest peaks marked on the original image (right).

neighbor weight, $w \in [0, 1]$. Let m be the maximal value in an input map X. Going from left to right, for each pixel, $o = (x, y)$, let

$$l(x, y) = \max(L(x-1, y-1), L(x-1, y), L(x-1, y+1)),$$

where L, the output of the link left procedure, is defined as

$$L(o) = \begin{cases} X(o) & \text{if } X(o) < Tm \bigvee X(o)\, tm > l(O) \\ X(o) + wl(o) & \text{otherwise} \end{cases}$$

Typically, L is applied on the logarithm of the dark symmetry with the parameters set as

$$T = 0.01, t = 0.7, w = 0.9.$$

Other values were tested with equal success.

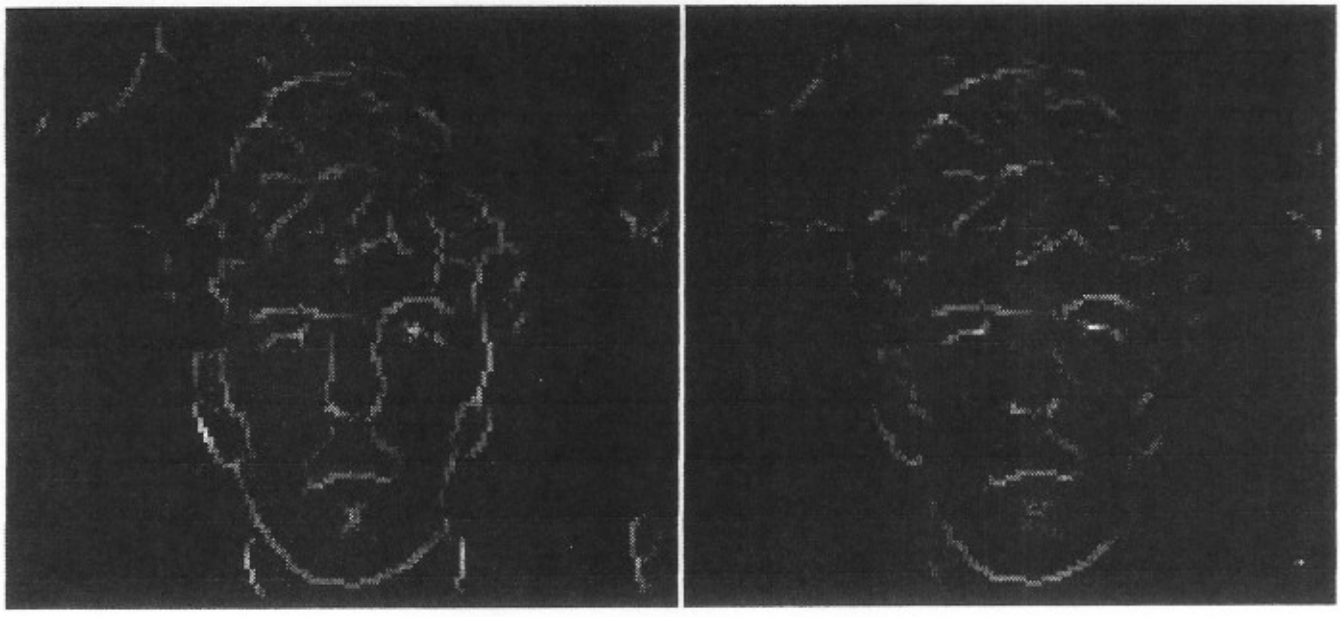

FIG. 3. Nonmaximal suppression of the dark isotropic symmetry (left) and of the projection of the dark isotropic symmetry (right).

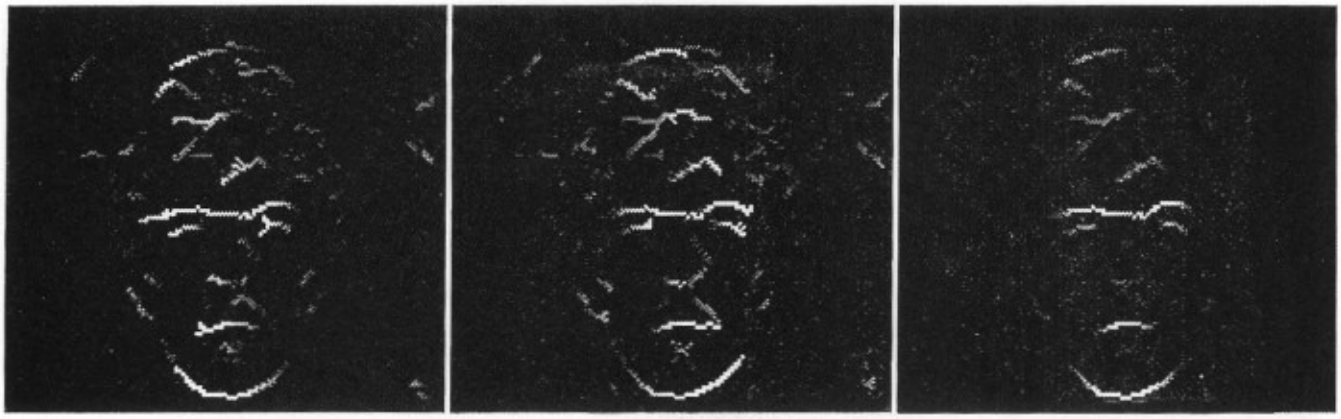

FIG. 4. Linking the result of the nonmaximal suppression on the projected dark isotropic symmetry (left of Fig. 3). (Left to right) Link left, link right, link.

It is convenient to demonstrate the operation of the link procedure on the nonmaximal suppression map as shown in Fig. 4. However, for the purpose of finding the center of the mouth we empirically found that an application on the logarithm of the projection of the dark symmetry is better (see Fig. 5). Points with high symmetry edge from both sides are enhanced. The local maxima are searched in a neighborhood to radius which is about 1/10 the image width. Other radii are equally valid.

2.2.4. Saliency. Our main purpose is to locate interesting regions in a scene. Nevertheless, the symmetry transform can also help in segmenting them. The *saliency* operation enhance edges that contribute to high symmetry values and thus are natural candidates for being boundaries of interesting regions. Given an edge map with magnitude R and a symmetry map (X, ψ), we denote by influence(p) the points

$$\text{influence}(p) = \left\{ o \mid \exists_q \, , (p,q) \in \Gamma(o) \bigvee (p,q) \in \Gamma(o) \right\}.$$

A possible way to define the *saliency* of a point P is

$$\text{Salient}(p) = R(p) \max_{o \in \text{influence}(p)} X(o).$$

We demonstrate the use of the saliency transform for segmentation using Fig. 6. The top left is the logarithm of a nonmaximal

FIG. 5. Application of link on the logarithm of the projection of the dark symmetry (left) and the highest local maxima (right).

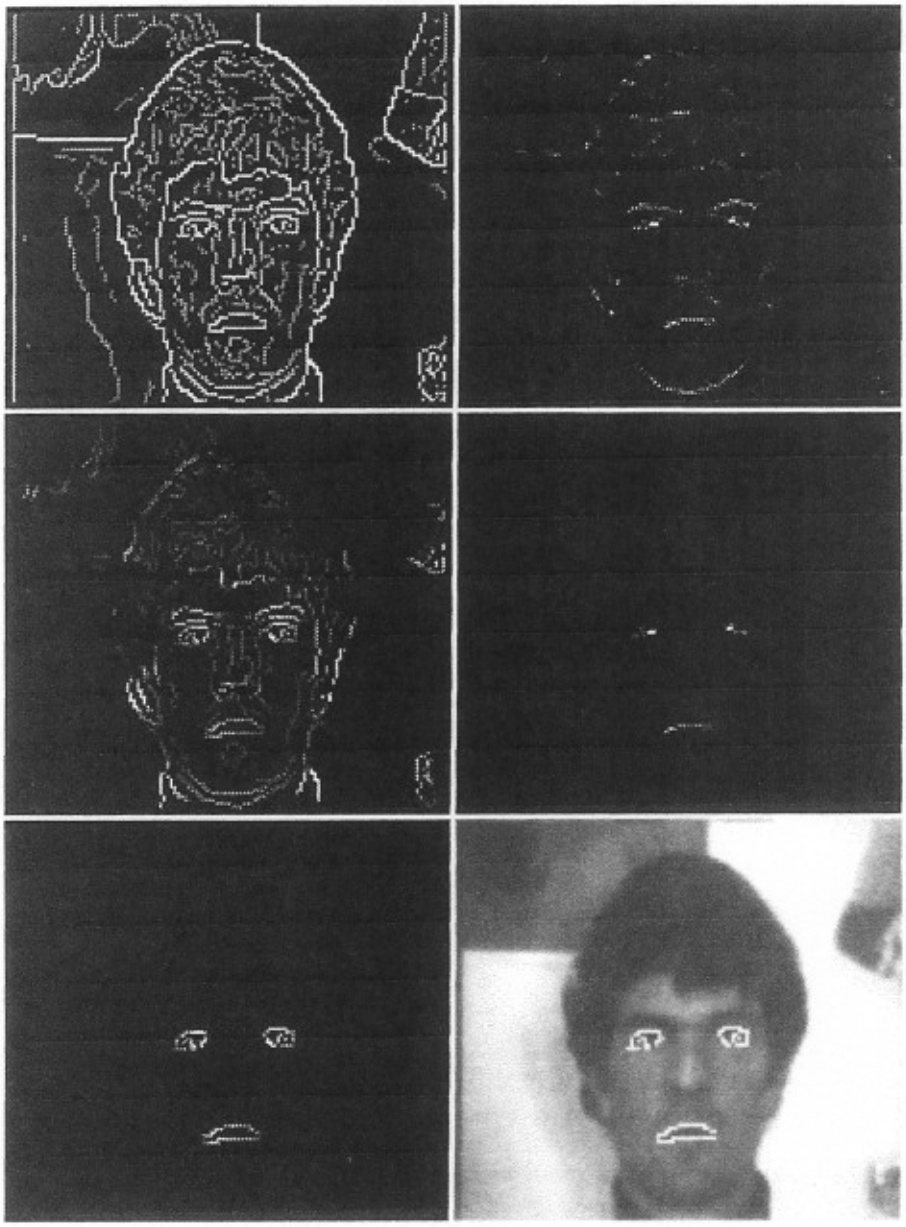

FIG. 6. Saliency used for segmentation (see text).

suppression of gradient magnitude. The use of nonmaximal gradient values are suppressed since we would like thin edges as the contours of features. Next, on the top right, the projection of the dark isotropic symmetry of the previous image is shown. In the middle left the saliency as defined above is presented. This is a considerable improvement over the original edge map, but it is not the best that can be achieved. The next image (middle right) is obtained by considering only the peaks in the projection map and the points with high values which are neighbors of the points already considered. Such a hysteresis algorithm is used in edge detectors such as the Canny one. The saliency induced by this map is now much better as can be seen at the bottom left. Finally a thresholded map superimposed on the original image is shown on the bottom right.

2.3. *Midline*

The face is symmetric in a generalized way. The symmetry transform is formalized in such a way that preference is given to gradients that are directed toward each other (or against each other) as previously discussed in length. Thus, regarding the face as a blurred bulb, the symmetry transform can be used to located a face in a picture (see [32]). However, in a fine resolution, the gradients around the midline of the face are parallel to each other, thus giving rise to low symmetry values. It is possible

FIG. 7. The best midline marked on the gradient image and the original image.

to formulate another symmetry transform that gives preference to parallel gradients. An alternative is to use cross-correlation between a candidate to the left side of a face and the reflection of the candidate for the right side.

There are few formulation for cross-correlation, where the simplest is

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_x \sigma_y}.$$

For the purpose of finding the midline a face can be regarded as a rectangle, as an ellipse (as in the cut operation in the following), or as any other figure.

To reduce shift of the midline due to illumination problems it is vital to apply the correlation on a derivative of the image in a vertical direction:

$$\Delta(x, y) = G_\sigma(x) \frac{d}{dy} G_\sigma(y) * I(x, y).$$

Typically, if the average height of a face is h, we use a derivative of Gaussian with $\sigma = 0.05h$.

If one deals with faces that fill most of the picture and which are taken with homogeneous background, then taking $\mathrm{Cov}(X, Y)$ without dividing by $\sigma_x \sigma_y$ is sufficient since the variance varies a little.

$\rho(X, Y) = 1$ when $X = Y$ (actually whenever $Y = aX + b$ for a positive constant a and any constant b). Thus high correlation values are induced by homogeneous objects such as a white background. To avoid this, we require that the multiplication of X and Y exceed a certain threshold (usually 1):

$$\hat{\rho}(X, Y) = \begin{cases} \rho(X, Y) & \text{if } \mathrm{Var}(X)\mathrm{Var}(Y) > 1 \\ 0 & \text{otherwise} \end{cases}$$

The maximum of $\hat{\rho}(X, Y)$ is demonstrated in Fig. 7.

We have found that the midline is not sensitive to the size of the window. See further discussion below.

2.4. *Basic Geometrical Properties*

The result of each of the symmetry operations described above is a list of points ranked by their magnitude. Typically, the eyes and mouth are in the first five points. In addition, candidates for the midline are ranked too. This enables an efficient choice process based on the geometrical constraint. A triplet of candidates can be generated in order to be checked by the classification process (or any special facial feature detector described in the Introduction). In the data base we used the correct triplet always came first. Since the number of points is small a simple backtracking algorithm is more effective than a learning process such as [21].

Faces are usually presented in an upright position, the facial features are rarely jumbled, and the triangles whose vertices are the eyes and mouth are usually quite similar to each other. We also used some simple geometrical preconceptions as a means to reduce the face processing time. The eyes are expected to be above the mouth, and both eyes are expected to be on a horizontal line whose angle is limited. Our method can be used for tilted faces also, as is demonstrated in Fig. 8, but in order

FIG. 8. Detection of eyes in tilted images. Face images of various tilt angles (left), and the response of the symmetry transform (right). The same parameters are used in all three images.

to avoid false matches as much as possible, and assuming that under most normal conditions faces are not greatly tilted, we have limited this angle, which is the maximal tilt tolerance, to $\pm 15°$. If the features are not found, the image can be rotated in this amount of tilt, and the features can be searched for again. Another constraint is that the eyes should be on both sides of a peak in the midline as defined above.

The eyes and mouth constitute approximately an isoscale triangle, normalizing the coordinates such that the left eye is in (0, 0) and the right eye is in (0, 1); the mouth has to be in a neighborhood of (0.5, d), where d and the size of the neighborhood is dependent on the type of faces the system expects.

There is only a single false match caused by the symmetry transform and that is its high response to the eyebrow area. However, since the geometrical relations between the eye-induced peaks and the eyebrow-induced peaks are clear, it is rather trivial to eliminate this artifact.

Integrating all these constraints can be performed using elaborated methods such as relaxation labeling [34]. Our experience shows that simple checks of all these constraints is sufficient since the search space is small.

FIG. 9. Peaks after applying geometric constraints.

Figure 9 demonstrates how these geometrical constraints enable the locating on the eyes and mouth: the highest peaks in the projection of the isotropic symmetry at the possible locations of eyes and the highest peak in the link of that projection in the possible region for the mouth.

2.5. *Robustness*

Combining the peaks of the symmetry transform with geometrical considerations is quite robust, as will be exemplified in this section. However, this robustness could be further increased if we use simple classifiers. By applying classifiers that differentiate between a "face" image and a "nonface" image [37], it is possible to automatically examine the possible configurations suggested by the preprocessing step and select only the one with the highest "face" marks.

This paper was written with *en face* pictures in mind. Slight rotations, however, are tolerated. Figure 10 shows a face which is almost in profile. In this example we are still able to find the features.

Glasses introduce specularities and cause drastic modification of the edges of the eyes. Figure 11 shows that no special treatment is needed provided that the glasses are transparent.

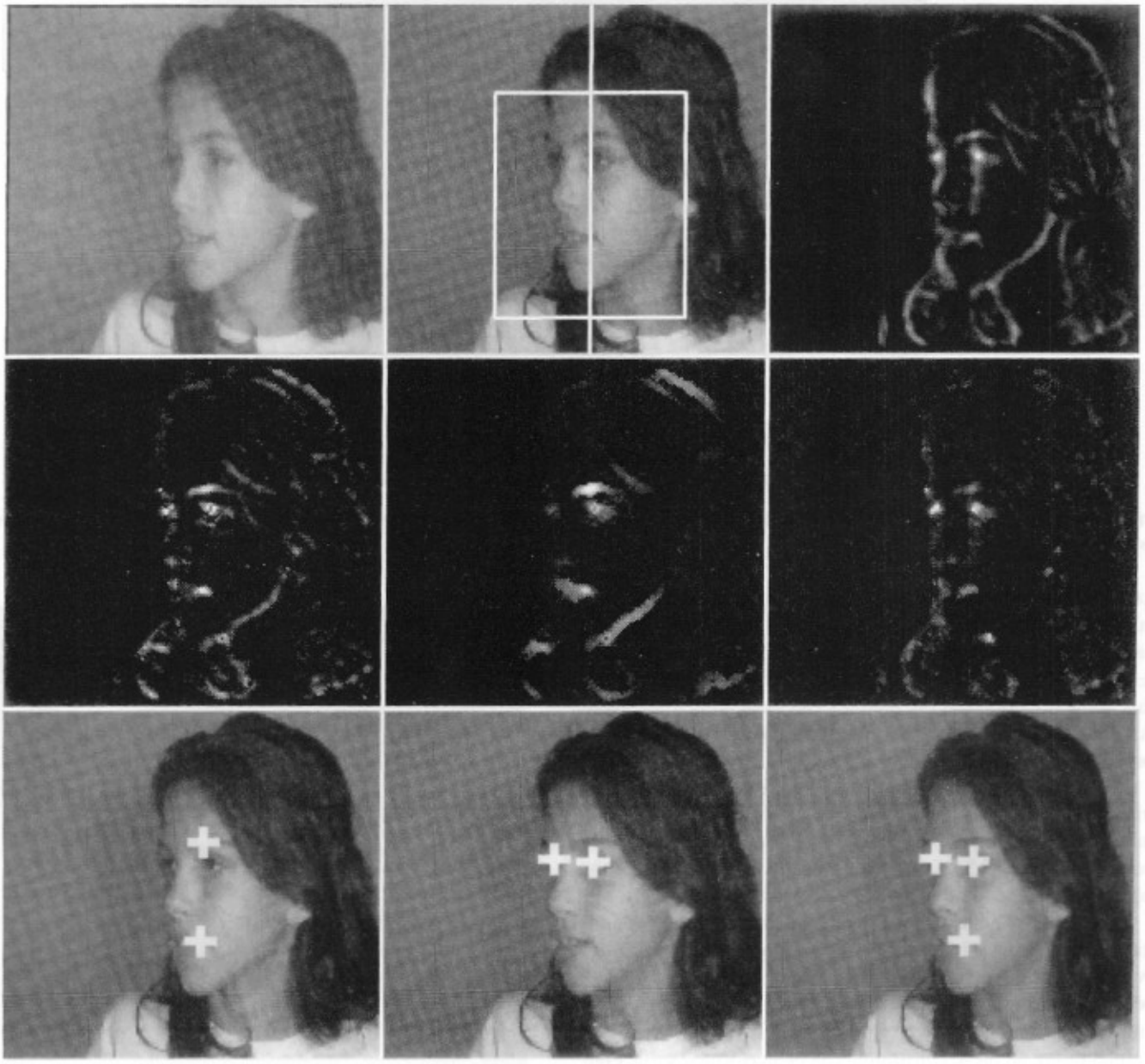

FIG. 10. The same parameters as before for 3/4 view. (Top) Original image, midline, and isotropic symmetry. (Middle) Projection of the isotropic symmetry on the horizon, link of the projection, and radial symmetry. (Bottom) Highest two peaks of the link, highest two peaks of the radial symmetry, and output after applying geometrical constraints.

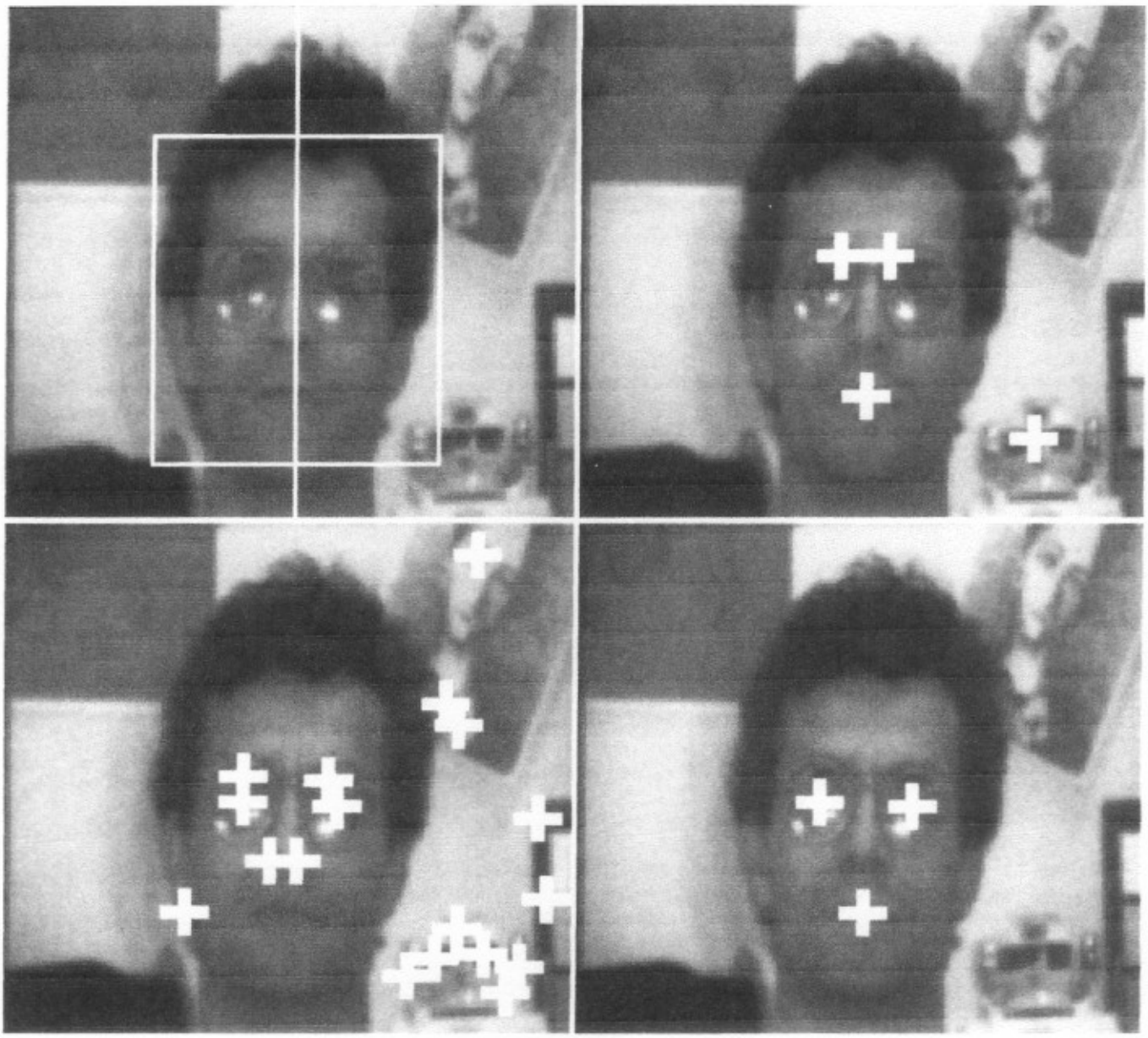

FIG. 11. Glasses and noisy background. (Top) Midline and peaks of the link of the projection of the isotropic symmetry on the horizon. (Bottom) Peaks in radial symmetry and the final results after applying geometric constraints.

This picture also has noisy background which is handled correctly. Notice that the eyebrows cause high peaks of symmetry. This is typical to many faces and should be taken care of by the geometrical constraints module.

Beards distort the edges in the region of the mouth. Nevertheless, the mouth still causes high symmetry values at the mouth as shown in Fig. 12.

The face is an extremely plastic object. Figure 13 demonstrates that the detector can deal with various facial expressions.

Next we consider the sensitivity to scale. Though the size of the faces in this paper vary considerably, exactly the same parameters of the operator are used in all cases. These figures demonstrated that the symmetry transforms can tolerate variance in scale and that a multiscale scheme can be applied along with the generalized symmetry transforms. This is also demonstrated in Fig. 14. A more complete solution which is currently tested is to use the *constrained phase congruency transform* [29] which enables the simultaneous detection of interest points and their scale.

FIG. 12. A system manager with a beard. (Top) Original image and midline. (Middle) Projection of the isotropic symmetry on the horizon (left) and peaks of link of logarithm of the projection (right). (Bottom) Peaks of radial symmetry (left) and location of features after applying geometrical constraints (right).

3. NORMALIZATION

The use of the eye's location for facial image normalization dates back to 1878 [12]. The use of affine normalization substantially increases the effectiveness of general purpose pattern recognition techniques as well as face processing tasks. The first demonstration is given in Fig. 15 where principle component analysis is used as a tool for associative memory. The images on the top demonstrate the use of principle component analysis without a prior normalization. The top left picture is a new image introduced to the system. The black rectangle is a noise which is expected to be eliminated by the associative memory. The 432 images in the data base are taken from the Turk and Pentland data base [41]. The projection of the test image over the first 25 eigenvectors (center) and the first 40 eigenvectors (right) results in a very poor reconstruction, at least for human interpretation. The reconstruction improves dramatically (bottom) when the test image as well as the images in the data base

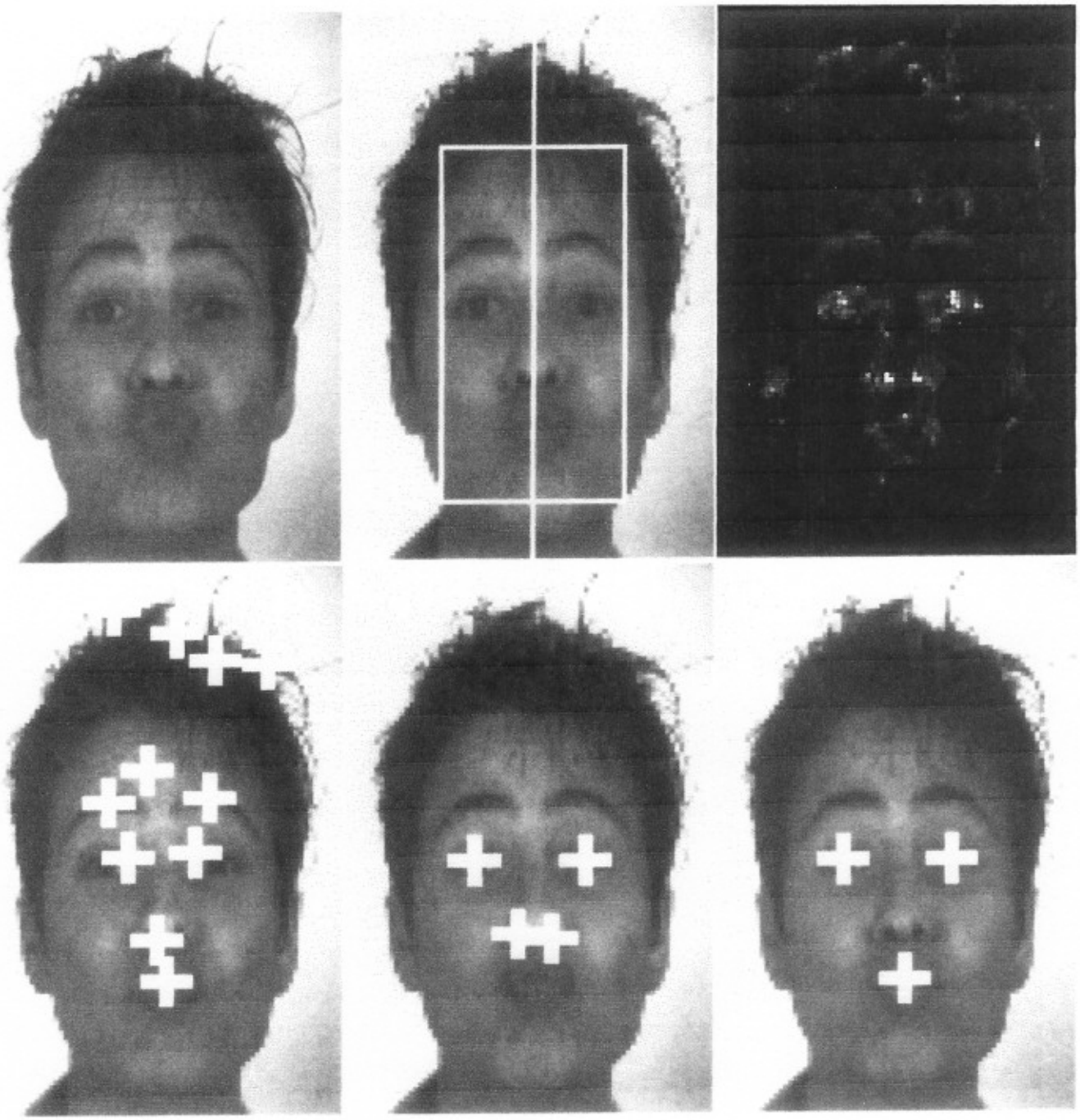

FIG. 13. Unconventional facial expression. (Top) Original image, midline, and radial symmetry. (Bottom) Peaks of links of logarithm of the projection, peaks of radial symmetry, and location of features after applying geometrical constraints.

are normalized such that the eyes and mouth are warped to a standard location using an affine transformation. It is now evident that the projection on the first 25 eigenvectors is sufficient for a human to recognize the person. Notice that mild affine transformations, as demonstrated in this figure and in numerous informal experiments which we have conducted do not change the identity of the face for a human observer.

3.1. The Normalization Procedure

We have implemented the ideas of normalization for face recognition in a system which is described in [9]. The system integrates the ideas of face normalization based on the location of the eyes and mouth, dimensionality reduction using receptive fields, and radial basis functions neural networks. The success of

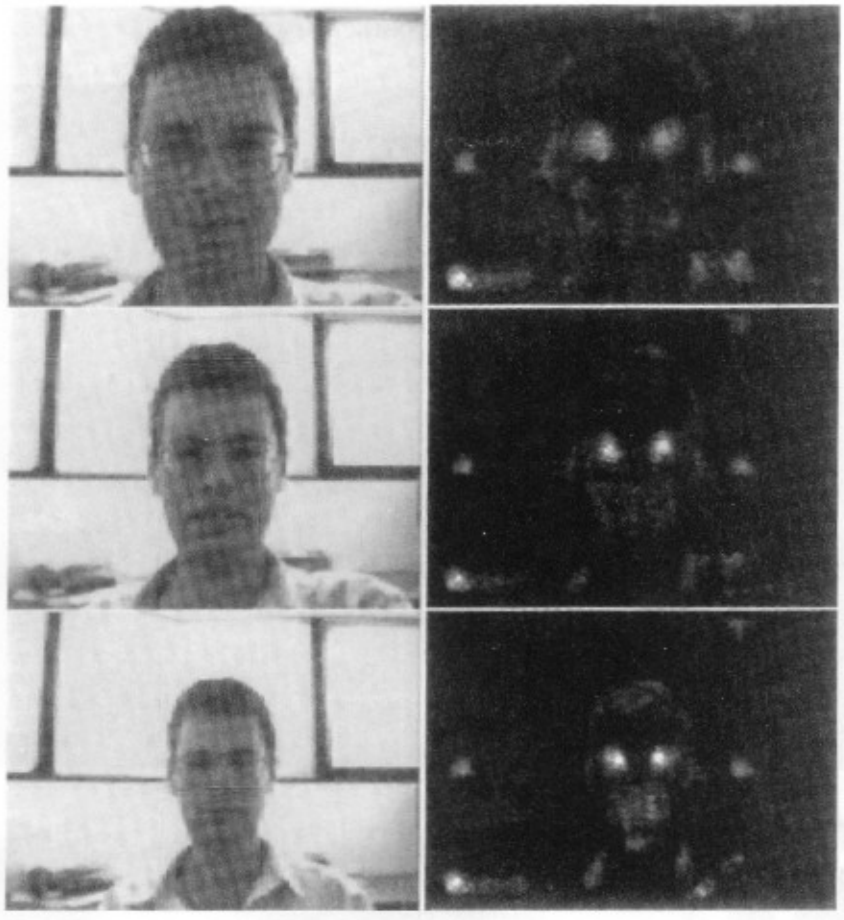

FIG. 14. Detection of eyes in various scales. Face images of various size (left) and the response of the symmetry transform (right). The same parameters are used in all three images.

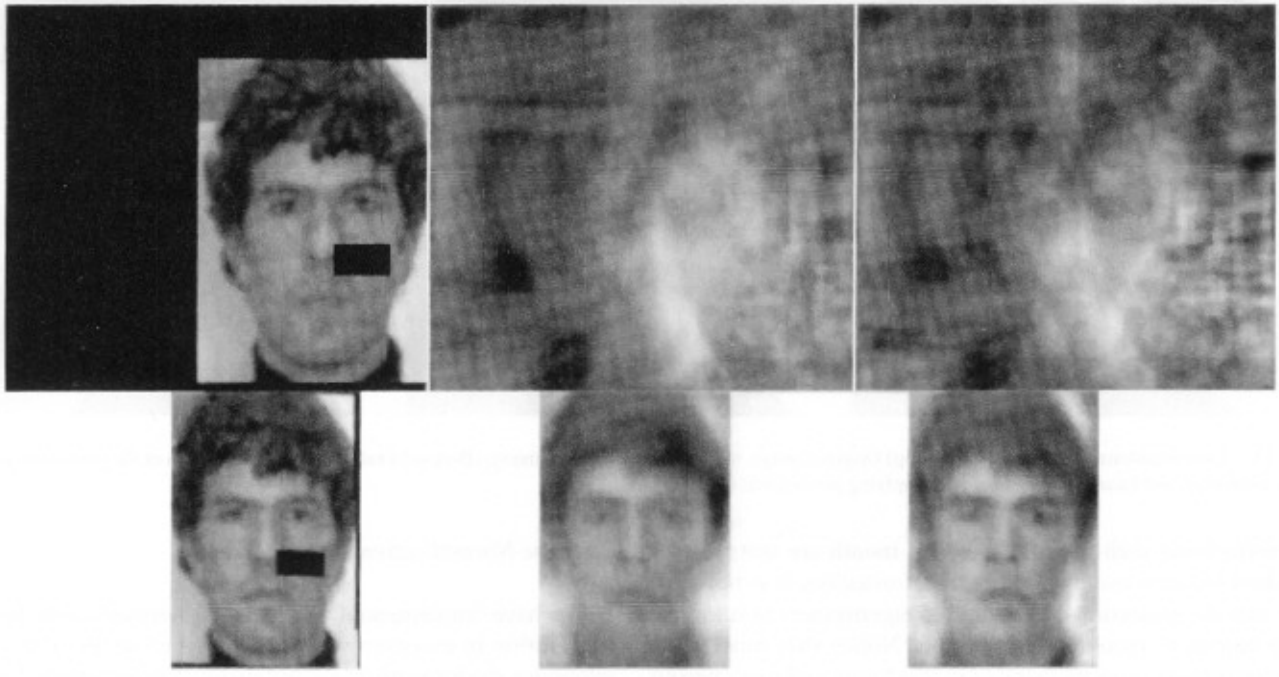

FIG. 15. Projection of a new face on a principle component space. (Top) A new image (left), projection on 25 eigenvectors (center), and on 40 eigenvectors (right). (Bottom) The same as the top but applied to normalized images by the location of the eyes and the mouth.

Cut	*Warp*	*Gradient*	*success ratio*	
no	no	no	157/416	37.74%
remove background	no	no	55/416	13.22%
remove face	no	no	112/416	26.92%
remove background	yes	yes	240/416	57.69%

FIG. 16. Nearest neighbor using one learned image for a person.

Cut	*Warp*	*Gradient*	*success ratio*	
no	no	no	165/384	42.97%
remove background	no	no	29/384	7.55%
remove face	no	no	119/384	30.99%
remove background	yes	yes	253/384	65.89%

FIG. 18. Nearest neighbor using three learned images for a person.

that system triggered the systematic test of the effect of normalization. In this study simpler classification schemes were used with improved results.

We define three transformations that facilitate normalization. A two dimensional affine transformation is uniquely determined by three points. In this section we call the affine mapping of a face image determined by the location of the eyes and mouth in the given image and standard locations *warp*. (Actually the reverse transform is evaluated.) The output image size is 40×60. The standard eye locations were chosen as (13, 27) and (27, 27), and the mouth location was set to (20, 44). All the results reported in this section were reproduced when all these numbers were doubled.

The second transformation we call *cut*: Let P_l, P_r, and P_m denote the locations of the left eye, right eye, and mouth in a given image. We define a function, $C : \mathbb{R}^2 \mapsto \mathbb{R}^2$, connected with an ellipse around the face as

$$C(P) = \begin{pmatrix} \frac{1}{1.9\|P_m - O\|} & 0 \\ 0 & \frac{1}{1.5\|P_r - P_l\|} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} (X - O),$$

where $O = (P_l + P_r)/2$ and $\theta = \arg(P_r - P_l)$.

Setting to zero each pixel, X, outside the ellipse, $C(X) > 1$, is considered removing the background, while setting to zero each pixel inside the ellipse, $C(X) \leq 1$, is considered removing the face.

The last transformation we call *gradient* and is convolving an image with a directional derivative of a Gaussian with $\sigma = 3$. Using other σ's yields similar results. It is useful for reducing the effect of lighting variations.

3.2. Classification Results

We checked the importance of the various normalization procedures—*warp*, *cut*, and *gradient*, using the Turk and Pentland database [41]. The database contains 27 face images of each of 16 different persons. Turk and Pentland used the principle component analysis procedure applied on the raw pixels to reduce the dimensionality of the images. The images differ mainly by the lighting conditions and rotation. The test images were randomly chosen. The rest were used for testing.

The background of the faces in the data base is noisy. Thus, using any general purpose feature extraction and classification techniques without eliminating the background may result in poor classification performances. Another outcome is that the system will show good results that stem from its capability to recognize the background rather the faces. In our experiments we have used the pixels as the classifier's features and thus without the *cut* operation the pixels in the background are as important as the pixels in the face.

Figures 16 and 17 present the results of applying the 1-nearest-neighbor procedure using one image of each person in the learning phase and 26 images of each person as a test. Leaving the warp and gradient aside, it shows that cutting the background as well as cutting the foreground results in a substantial decrease

Warp	*Gradient*	*success ratio*	
no	no	55/416	13.22%
no	yes	44/416	10.58%
yes	no	142/416	34.13%
yes	yes	240/416	57.69%

FIG. 17. Nearest neighbor using one learned image for a person without background.

Warp	*Gradient*	*success ratio*	
no	no	29/384	7.55%
no	yes	24/384	6.25%
yes	no	168/384	43.75%
yes	yes	253/384	65.89%

FIG. 19. Nearest neighbor using three learned images for a person without background.

Cut	*Warp*	*Gradient*	*success ratio*	
no	no	no	156/160	97.50%
remove background	no	no	115/160	71.88%
remove face	no	no	154/160	96.25%
remove background	yes	yes	147/160	91.88%

FIG. 20. Nearest neighbor using 17 random learned images for a person.

Classifier	*success ratio*	
1 NN	147/160	91.88%
3 NN	83/160	51.88%
RBF	147/160	91.88%

FIG. 22. Various classification techniques using 17 random learned images for a person without background and with warping and gradient.

in performances and that cutting the background is more damaging than cutting the foreground. This figure also demonstrates that using warp and gradient on the foreground image substantially increases performances. Notice that the chance level is $1/16 = 6.25\%$ and the results with the normalization are way above chance level although not satisfactory. The effect of the background removal is further shown in Fig. 17. The advantages of both warp and gradient can be seen clearly.

In the next experiment the number of images of the same person in the learning phase was increased to three leaving 24 images of each person in the test phase. The results are reported in Fig. 18. As can be expected the results of the full normalization (cut, warp, and gradient) are improved. It may come as a surprise that the results of the classifier on the raw images without the background is reduced to almost chance level. This is due to the fact that for each person, the images taken differ mainly in lighting conditions. Two of these images were used as tests in the previous experiment and they are the reason for the better results there. The effect of the various normalization procedures is shown in Fig. 19. The importance of the combination is evident.

Now we turn to randomly using 17 images of each person in the learning phase leaving 10 for the test. Recall that these were the conditions in the pilot system reported in the previous section. As could be expected, the results (Figs. 20 and 21) are much better than those obtained using the smaller learning set. Notice that the effect of the normalization procedure is equally important.

Warp	*Gradient*	*success ratio*	
no	no	115/160	71.88%
no	yes	118/160	73.75%
yes	no	131/160	81.88%
yes	yes	147/160	91.88%

FIG. 21. Nearest neighbor using 17 random learned images for a person without background.

Now we use the same learning and test images as in the previous experiment and check the effect of the classifier by comparing the performance of 1-nearest-neighbor, 3-nearest-neighbor, and a radial basis function classifier. As can be seen in Fig. 22 the performances are not improved by changing of the classifier.

Can we improve the classification results using principle component analysis? Figure 23 indicates a positive answer. Using 1-nearest-neighbor, optimal results are achieved for 44 eigenvectors.

Next we compare 1-nearest-neighbor, 3-nearest-neighbor, and radial basis function classification result using the optimal number of eigenvectors found for 1-nearest-neighbor (Fig. 24).

The final results presented above show a success ratio of 95%. An ongoing work using projection pursuit combined with back propagation shows a success ratio of over 99% [16]. Larger scale experiments are now needed to further explore the importance of normalization as well as classification techniques.

No. of eigenvectors	*success ratio*	
1	29/160	18.13%
5	115/160	71.88%
10	132/160	82.50%
20	142/160	88.75%
30-40	150/160	93.75%
41-43	151/160	94.38%
44-61	152/160	95.00%
62	151/160	94.38%

FIG. 23. Nearest neighbor using varying numbers of principle components of 17 random learned images for each person without background and with warping and gradient.

Classifier	*success ratio*	
1 NN	152/160	95.00%
3 NN	99/160	61.88%
RBF	152/160	95.00%

FIG. 24. Various classification techniques using 17 random learned images for a person without background and with warping and gradient for 44 eigenvectors.

4. CONCLUSIONS

Recognition of facial images is one of the earliest applications that has been suggested in the framework of computer vision and pattern recognition. However, practical applications have emerged only recently, due to the complexity of the problem. In this paper we have separated the problem of preprocessing from the classification and proposed a method for the automatic detection of facial features. Based on the generalized symmetry transform, we showed that the eyes and mouth in facial images can be robustly detected. Then we have used these points to normalize the images, assuming affine transformation, which can compensate for various viewing positions.

In order to estimate the efficiency of our method, we have used images taken from a public data base of face images applying standard classification methods 1-NN, 3-NN, and RBF, with and without normalization. The results obtained demonstrate the need for spatial normalization procedures. Combining the proposed method with additional normalizations of facial expression [30] can thus ensure more robust and efficient face recognition methods.

In this work the first candidates for eyes and mouth were always the right ones. In more complex data bases, however, other candidates should be considered if the normalized image is not classified as a face. Combining a simple classifier that distinguishes between a normalized face and anything else guarantees an extremely robust facial feature detector.

The results obtained here can be refined by using them to initialize the parameters of special purpose facial feature detectors and by restricting their search space as discussed in the Introduction.

Further work includes the improvement of scale flexibility by using the *constrained phase congruency* [29], integrating better classifiers [16], and testing on a larger data base.

REFERENCES

1. R. J. Baron, Mechanisms of human facial recognition, *Int. J. Robot. Res.* **15**, 1981, 137–178.
2. V. Bruce and M. Burton, Computer recognition of faces, in *Handbook of Face Processing* (A. W. Young and H. D. Ellis, Eds.), pp. 487–506, Elsevier Science, B.V. (North-Holland), 1989.
3. R. Brunelli and T. Poggio, Face recognition: Features versus templates, *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1993, 1042–1052.
4. P. Burt, Smart sensing with a pyramid vision machine, *Proc. IEEE*, **76**(8), August 1988, pp. 1006–1015.
5. G. Cottrell and M. Fleming, Face recognition using unsupervised feature extraction, in *Proc. of the International Neural Network Conf., 1990.*
6. I. Craw, H. Ellis, and J. R. Lishman, Automatic extraction of face-features, *Pattern Recog. Lett.* **5**, 1987, 183–187.
7. G. M. Davis, H. D. Ellis, and J. W. Shepherd, Face recognition accuracy as a function of mode of representation, *J. Appl. Psych.* **63**, 1978, 180–187.
8. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
9. S. Edelman, D. Reisfeld, and Y. Yeshurun, Learning to recognize faces from examples, in *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy, May 1992*, pp. 787–791.
10. I. A. Essa and A. P. Pentland, Facial expression recognition using a dynamic model and motion energy, in *Proceedings of the 5th International Conference on Computer Vision, 1995*, pp. 360–367.
11. I. Fraser and D. Parker, Reaction time measures of feature saliency in perceptual integration task, in *Aspects of Face Processing* (H. Ellis, M. Jeeves, F. Newcombe, and A. W. Young, Eds.), pp. 207–220, Martinus Nijhoff, Dordrecht, 1986.
12. F. Galton, Composite portraits, made by combining those of many different persons, into a single, resultant figure. *J. Anthropol. Inst.* **8**, 1879, 132–144.
13. V. Govindaraju, S. N. Srihari, and D. B. Sher, A computational model for face location, in *Proceedings of the 3rd International Conference on Computer Vision, Osaka, Japan, December 1990*, pp. 718–721.
14. N. D. Haig, Investigating face recognition with an image processor computer, in *Aspects of Face Processing* (H. Ellis, M. Jeeves, F. Newcombe, and A. W. Young, Eds.), pp. 310–317, Martinus Nijhoff, Dordrecht, 1986.
15. N. Intrator, D. Reisfeld, and Y. Yeshurun, Extraction of facial features for recognition using neural network, in *Int. Workshop on Automatic Face- and Gesture-Recognition, Zurich, 1995*, pp. 260–265.
16. N. Intrator, D. Reisfeld, and Y. Yeshurun, Face recognition using a hybrid supervised/unsupervised network, *Pattern Recog. Lett.* **17**(1), 1996, 67–76.
17. T. Kanade, *Picture Processing System by Computer Complex and Recognition of Human Faces*, PhD thesis, Dept. of Information Science, Kyoto University, November 1973.
18. T. Kanade and J. P. Kender, Mapping image properties into shape constraints: skewed symmetry, affine-transformable patterns, and the shape-from-texture paradigm, in *Human and Machine Vision* (J. Beck, B. Hope, and A. Rosenfeld, Eds.), pp. 237–257, Academic Press, New York, 1983.
19. M. Kirby and L. Sirovich, Application of the karhunen-loève procedure for characterization of human faces, *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(1), 1990, 103–108.
20. A. Lanitis, C. J. Taylor, and T. F. Cootes, A unified approach to coding and interpreting face images, In *Proceedings of the 5th International Conference on Computer Vision, 1995*, pp. 368–373.
21. T. K. Leung, M. C. Burl, and P. Perona, Finding faces in cluttered scenes using random labeled graph matching, In *Proceedings of the 5th International Conference on Computer Vision, 1995*, pp. 637–644.
22. R. Linnggard, D. J. Myers, and C. Nightigale (Eds.), *Neural Networks for Vision, Speech and Natural Language*, Chapman & Hall, London, 1992.
23. N. Magnenat-Thalmann, P. Primeau, and D. Thalmann, Abstract muscle action procedures for face animation, *Visual Comput.* **3**, 1988, 290–297.
24. B. S. Manjunath, R. Chellappa, and C. von der Malsburg, A feature based approach to face recognition. Technical Report CS-TR-2834, Center for Automated Research. University of Maryland, 1992.
25. R. Milward and A. O'Toole, Recognition memory transfer between spatial-frequency analyzed faces, in *Aspects of Face Processing* (H. Ellis,

M. Jeeves, F. Newcombe, and A. W. Young, Eds.), pp. 125–141, Martinus Nijhoff, Dordrecht, 1986.

26. B. Moghaddam and A. Pentland, Probabilistic visual learning for object detection, in *Proceedings of the 5th International Conference on Computer Vision, 1995*.
27. S. Platt and N. I. Badler, Animating facial expression, *Comput. Graphics* **15**(3), 1981, 245–252.
28. D. Reisfeld, *Generalized Symmetry Transforms: Attentional Mechanisms and Face Recognition*, PhD thesis, Computer Science Department, Tel-Aviv University, January 1994.
29. D. Reisfeld, The constrained phase congruency feature detector: simultaneous localization, classification and scale determination, *Pattern Recog. Lett.* **17**(11), 1996, 1161–1169.
30. D. Reisfeld, Nur Arad, and Y. Yeshurun, Normalization of face images using few anchors, In *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Jerusalem, Israel, 1994*.
31. D. Reisfeld, H. Wolfson, and Y. Yeshurun, Detection of interest points using symmetry, in *Proceedings of the 3rd International Conference on Computer Vision, Osaka, Japan, December 1990*, pp. 62–65.
32. D. Reisfeld, H. Wolfson, and Y. Yeshurun, Context free attentional operators: the generalized symmetry transform, *Int. J. Comput. Vision* **14**, 1995, 119–130.
33. D. Reisfeld and Y. Yeshurun, Robust detection of facial features by generalized symmetry, in *Proceedings of the 11th IAPR International Conference on Pattern Recognition, The Hague, The Netherlands, August 1992*, pp. A117–120.
34. A. Rosenfeld, R. A. Hummel, and S. W. Zucker, Scene labelling by relaxation operators, *IEEE Trans. Systems Man Cybernet.* **6**(6), June 1976, 420–533.
35. A. Samal and P. A. Iyengar, Automatic recognition and analysis of human faces and facial expressions: A survey, *Pattern Recog.* **25**(1), 1992, 65–77.
36. J. Stonham, Practical face recognition and verification with WISARD, in *Aspects of Face Processing* (H. Ellis, M. Jeeves, F. Newcombe, and A. W. Young, Eds.), pp. 426–441, Martinus Nijhoff, Dordrecht, 1986.
37. K. K. Sung and T. Poggio, Example-based learning for view-based human face detection, Technical Report A.I. Memo No. 1521, MIT AI Lab, December 1994.
38. T. Kohonen, E. Oja, and P. Lehtiö, Storage and processing of information in distributed associative memory systems, in *Parallel Models of Associative Memory* (G. E. Hinton and J. A. Anderson, Eds.), pp. 129–167, Elbraum, Hillsdale NJ, 1989.
39. J. W. Tanaka and M. J. Farah, Parts and wholes in face recognition, *Quart. J. Experimental Psych. A* **46**(2), 1993, 225–245.
40. D. Terzopolous and K. Waters, Physically based facial modeling, analysis, and animation, *J. Visualization Animation* **1**(2), 1990, 73–80.
41. M. Turk and A. Pentland, Eigenfaces for recognition, *J. Cog. Neurosci.* **3**(1), 1991, 71–86.
42. K. Waters and D. Terzopoulos, Modeling and animating faces using scanned data, *J. Visualization Comput. Animation* **2**(4), 1991, 123–128.
43. A. L. Yarbus, *Eye Movements and Vision*, Plenum Press, New York, 1967.
44. A. L. Yuille, P. W. Halilinman, and D. S. Cohen, Feature extraction from faces using deformable templates, *Int. J. Comput. Vision* **8**(2), 1992, 99–111.