Intl. Workshop on Automatic Face and Gesture Recognition, Zurich, Switz., (June 1995)

Face Localization via Shape Statistics

M.C. Burl†, T.K. Leung‡, and P. Perona†§

† California Institute of Technology, 116-81, Pasadena, CA 91125, USA
‡ University of California at Berkeley, 387 Soda Hall, Berkeley, CA 94720, USA
§ Università di Padova, Italy
{burl,perona}@systems.caltech.edu, leungt@robotics.eecs.berkeley.edu

Abstract

In this paper, a face localization system is proposed in which local detectors are coupled with a statistical model of the spatial arrangement of facial features to yield robust performance. The outputs from the local detectors are treated as *candidate* locations and constellations are formed from these. The effects of translation, rotation, and scale are eliminated by mapping to a set of *shape* variables. The constellations are then ranked according to the likelihood that the shape variables correspond to a face versus an alternative model. Incomplete constellations, which occur when some of the true features are missed, are handled in a principled way.

1 Introduction

The problem of face recognition has received considerable attention in the literature [11, 24, 21, 4, 19, 17, 22, 10]; however, in most of these studies, the faces were either embedded in a benign background or were assumed to have been pre-segmented. For any of these recognition algorithms to work in real-world applications, a system is needed that can reliably locate faces in cluttered scenes and with occlusions.

Recent studies have begun to address the problem of face localization. Burel and Carel [5] proposed a method using multi-resolution analysis and learning from examples (multi-layer perceptron) to search for faces in an image. Yang and Huang [23] have described a system that uses a hierarchical knowledge-based method to locate faces. Also, Amit [1] has developed a system for aligning X-ray images of hands that is similar in some respects to the system we propose for localizing faces. Our algorithm improves upon these other systems in two primary respects: (1) we are able to explicitly handle occlusions, and (2) we are able to exploit the statistical structure of face images in a principled way.

Our system consists of the following steps. First, a set of local detectors is applied to the image to identify candidate locations of features such as the eyes, nose, and nostrils. To enforce the proper spatial arrangement of features, we form *constellations* from the pool of candidate feature locations and determine which constellations are the most face-like. The representation and ranking of the constellations is accomplished using the *statistical theory of shape*, which was developed by Kendall [13, 14, 15], Bookstein [2, 3], and others [6, 18, 15]. A key result that we use was obtained by Dryden and Mardia [6] who derived the exact density of the shape variables for the case when the original figure space variables obey a multivariate Gaussian distribution.

2 Local Feature Detectors

The initial step in our face localization algorithm is to identify candidate locations for various facial features using simple detectors. Any detector could be used, but in our experiments we have chosen to filter the incoming image with a set of multi-orientation, multi-scale Gaussian derivative filters [12, 9]; then detectors for specific facial features are synthesized by comparing the set of filter responses at a given location to a template set of responses. This comparison is done in a rotation and scale invariant way. A detection is declared if the degree of match τ exceeds a threshold τ_0.

This algorithm was tested on a well-controlled database containing 180 images of 18 subjects. The subjects were imaged at a distance of two meters against a plain white background. All views were quasi-frontal and were collected under the same lighting conditions. The performance of three detectors synthesized for different facial features is shown in Figure 1 using receiver operating characteristics (ROC curves) [20]. Each ROC curve shows the trade-off between the probability of detecting the target feature and the average number of false alarms per image as the detection threshold τ_0 is varied. For low values of τ_0, the target feature is detected with high probability, but there are a significant number of false alarms. For higher values, the number of false alarms is reduced, but so too is the probability of detecting the true feature.

Notice that to detect the left eye with probability 90%, the average number of false alarms per image is five. If we insist on at most two false alarms per image, the detection probability will drop to 73%. These results are *optimistic* because the faces in this database were imaged against a plain white background; with a cluttered background, the number of false alarms would certainly increase. Our basic conclusion is that *feature detectors based on the local brightness information are simply not reliable enough to provide consistent face localization.* We believe this conclusion applies not only to our particular choice of detectors (Gaussian derivative filters) but to feature detectors in general.

The unreliability of the local feature detectors leads to two problems: false alarms and, more seriously, missing features. In the sequel, we propose a rigorous probabilistic framework for handling these problems by coupling the output of the feature detectors with a statistical model of the spatial arrangement of facial features.

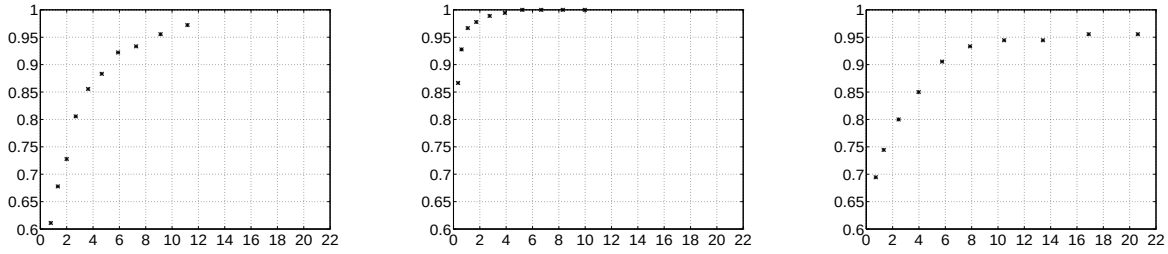

Figure 1: Probability of detecting true feature vs. the average number of false alarms per image.

3 Constellations

The output of the i-th feature detector ($i = 1, \ldots, N$) is a set S_i of candidate locations for the feature. A *special point* that allows for missing features is appended to each of these sets. A *constellation* $\mathbf{z}$ is then defined as any ordered N-tuple of points in which the i-th point is drawn from S_i. Constellations containing the special point will be referred to as *incomplete constellations*.

Let $\mathbf{Z}$ be the set of all possible constellations that can be formed from the S_i's and let m_i be the number of points in S_i (excluding the special point). Then, the number of possible constellations is $|\mathbf{Z}| = \prod_{i=1}^{N}(1 + m_i)$.

Assuming the local feature detectors perform well, $|\mathbf{Z}|$ will not be too large, and a face localization algorithm could be implemented that enumerates each constellation and ranks it according to how face-like the arrangement of features is. For the "easy database" described in the previous section, this approach may be computationally reasonable. However, if the detectors perform significantly worse (as would be expected in a cluttered environment) or if more features are used, the total number of candidate constellations will greatly increase. For example, in Figure 2, the total number of constellations is $\sim 10^6$. Fortunately, we will see in Section 6 that full enumeration is unnecessary, and an intelligent search through the constellations can be performed. First, however, we need to discuss how constellations are represented and compared.

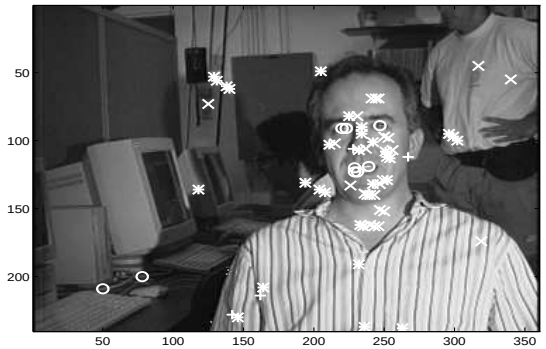
Figure 2: Candidate feature locations produced by the detectors: left eye-o, right eye-o, nose/lip-*, left nostril-+, right nostril-x.

4 Shape

To determine how face-like a particular constellation is, the representation of $\mathbf{z}$ as an ordered N-tuple of points in the plane is inconvenient because the useful information is clouded by the effects of translation, rotation, and scale. The information we really want is "what remains after differences due to [these effects] have been factored out", which is precisely what Kendall [13] has defined as *shape*.

In our previous work on face localization [16], a constellation of N points was represented using the set of $d = N(N-1)/2$ mutual distances between the feature points. This representation automatically guaranteed invariance with respect to translation and rotation. In addition, invariance to scaling was obtained by estimating the scale via maximum likelihood and then dividing it out. Defining probability distributions over this space of scaled mutual distances, however, was not ideal since the space is degenerate — many points in $\mathcal{R}^d$, where the vectors of mutual distances lie, do not correspond to valid planar constellations.

In the statistics literature, this problem is handled by transforming a constellation from $\mathcal{R}^{2N}$, the $2N$-dimensional space of feature points, to a $2N-4$ dimensional space of shape variables. Essentially, one dimension is dropped for factoring out scale, one for rotation, and two for translation. Kendall [13] denotes this new shape space as Σ_2^N and shows that it can be identified with a version of the complex projective space $\mathcal{CP}^{N-2}$. Dryden and Mardia [6] have found the joint probability density junction over the shape variables under the assumption that the original feature points are positioned in the plane according to a general $2N$-dimensional Gaussian distribution. This is the fundamental result from shape statistics that we exploit in our face localization system.

Dryden and Mardia begin their derivation by assuming the figure space variables (x_i, y_i), $i = 1, \ldots, N$ are distributed according to a general $2N$-dimensional Gaussian distribution:

$$\mathbf{X} = [x_1, \ldots, x_N, y_1, \ldots, y_N]^T \sim \mathcal{N}_{2N}(\nu, \Omega) \qquad (1)$$

The effects of translation, rotation, and scale can be eliminated by transforming two points to fixed reference positions. The positions of the other points after this transformation define the shape variables. For ease of notation, we will transform the first figure point to the origin and the second figure point

to $(1,0)$. The transformation of the first point to the origin can be effected by premultiplying $\mathbf{X}$ by the $2N \times 2N$ matrix $\mathbf{L}^T$ defined below:

$$\mathbf{L}^T = \begin{bmatrix} \mathbf{I} - \underline{\mathbf{1}}\mathbf{e}_1^T & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \mathbf{I} - \underline{\mathbf{1}}\mathbf{e}_1^T \end{bmatrix} \tag{2}$$

In this equation $\mathbf{I}$ and $\underline{\mathbf{0}}$ are the $N \times N$ identity and zero matrices, respectively, $\underline{\mathbf{1}}$ is a $N \times 1$ vector of ones, and $\mathbf{e}_1$ is the vector $[1, 0, \ldots, 0]^T$. Following this transformation, we have

$$\mathbf{X}^* \triangleq [0, x_2^*, \ldots, x_N^*, 0, y_2^*, \ldots, y_N^*]^T = \mathbf{L}^T\mathbf{X} \tag{3}$$

Omitting the fixed values in $\mathbf{X}^*$ yields the reduced vector $\mathbf{X}_R^*$, which also follows a Gaussian distribution

$$\begin{aligned} \mathbf{X}_R^* &\triangleq [x_2^*, \ldots, x_N^*, y_2^*, \ldots, y_N^*]^T = \mathbf{L}_R^T\mathbf{X} & (4) \\ \mathbf{X}_R^* &\sim \mathcal{N}_{2N-2}(\mu, \Sigma) & (5) \end{aligned}$$

where $\mu = \mathbf{L}_R^T\nu$ and $\Sigma = \mathbf{L}_R^T\Omega\mathbf{L}_R$. Elimination of scale and rotation by mapping the points such that $(x_2^*, y_2^*) \to (1, 0)$ yields the shape vector

$$\mathbf{U} = [u_3, \ldots, u_N, v_3, \ldots, v_N]^T \tag{6}$$

where (for $i = 3, \ldots, N$)

$$\begin{aligned} u_i &= \left(x_i^* x_2^* + y_i^* y_2^*\right) / \left(x_2^{*2} + y_2^{*2}\right) \\ v_i &= \left(y_i^* x_2^* - x_i^* y_2^*\right) / \left(x_2^{*2} + y_2^{*2}\right) \end{aligned} \tag{7}$$

The joint probability density function (pdf) of $\mathbf{U}$ is given in the following theorem:

Theorem 1 (Shape Density [6]) *Under the multivariate Gaussian model for the figure-space coordinates (Equation 1), the joint probability density function of the shape vector* $\mathbf{U}$ *is* $p_{\mathbf{U}}(\mathbf{U})$

$$= \frac{\mathrm{q} \cdot \exp(-\mathrm{g}/2)}{(2\pi)^{\mathrm{N}-2}} \cdot \sqrt{\frac{|\Psi|}{|\Sigma|}} \cdot (\mathrm{N}-2)!(2\sigma_2^2)^{\mathrm{N}-2} \tag{8}$$

where

$$\begin{aligned} q &= \sum_{i=0}^{N-2} \frac{\sigma_1^{2i}}{\sigma_2^{2i}} \mathcal{L}_i^{\left(-\frac{1}{2}\right)}\{-r_1^2\}\mathcal{L}_{N-2-i}^{\left(-\frac{1}{2}\right)}\{-r_2^2\} & (9) \\ g &= \mu^T\Sigma^{-1}\mu - \xi^T\Psi^{-1}\xi & (10) \\ \Psi^{-1} &= [\mathbf{u} \vdots \mathbf{v}]^T\Sigma^{-1}[\mathbf{u} \vdots \mathbf{v}] & (11) \\ \xi &= \Psi[\mathbf{u} \vdots \mathbf{v}]^T\Sigma^{-1}\mu & (12) \\ \mathbf{u} &= [1, u_3, \ldots, u_N, 0, v_3, \ldots, v_N]^T & (13) \\ \mathbf{v} &= [0, -v_3, \ldots, -v_N, 1, u_3, \ldots, u_N]^T & (14) \\ r_k^2 &= \left(\phi_k^T\xi\right)^2 / \left(2\sigma_k^2\right) \qquad \textit{for } k = 1, 2 & (15) \end{aligned}$$

and $\sigma_1^2 \geq \sigma_2^2$ *are the eigenvalues of* Ψ *with corresponding eigenvectors* ϕ_1, ϕ_2. *The function* $\mathcal{L}_i^{(a)}(x)$ *is the generalized Laguerre polynomial of degree i:*

$$\mathcal{L}_i^{(a)}(x) = \sum_{k=0}^{i} (1+a)_i(-x)^k / \{(1+a)_k k!(i-k)!\} \tag{16}$$

where $(1+a)_0 \triangleq 1$ *and* $(1+a)_k = (a+k) \cdot (1+a)_{k-1}$.

5 Ranking of Constellations

Given two constellations $\mathbf{z}_1$ and $\mathbf{z}_2$, how do we decide which one is more face-like? Since translation, rotation, and scale are irrelevant, we transform to shape variables and rephrase the problem as a test of hypothesis for the vector observation $[\mathcal{S}(\mathbf{z}_1)\,\mathcal{S}(\mathbf{z}_2)]^T$. The first hypothesis H_1 is that $\mathbf{z}_1$ is from a face and $\mathbf{z}_2$ is not. The competing hypothesis H_2 is that $\mathbf{z}_2$ is from a face and $\mathbf{z}_1$ is not. Denoting the probability density of the shape variables conditioned on being a face by $p(\mathcal{S}(\mathbf{z})|F)$ and conditioned on not being a face by $p(\mathcal{S}(\mathbf{z})|\overline{F})$, standard results from decision theory [7, 8] show that the optimal discriminant is

$$L_* = \frac{p(\mathcal{S}(\mathbf{z}_1)|F) \cdot p(\mathcal{S}(\mathbf{z}_2)|\overline{F})}{p(\mathcal{S}(\mathbf{z}_1)|\overline{F}) \cdot p(\mathcal{S}(\mathbf{z}_2)|F)} \tag{17}$$

which can be rewritten as

$$L_* = \frac{L(\mathcal{S}(\mathbf{z}_1))}{L(\mathcal{S}(\mathbf{z}_2))} \tag{18}$$

where

$$L(\mathcal{S}(\mathbf{z})) \triangleq \frac{p(\mathcal{S}(\mathbf{z})|F)}{p(\mathcal{S}(\mathbf{z})|\overline{F})} \tag{19}$$

Equation 19 provides the proper function for ranking a constellation $\mathbf{z}$ according to how face-like it is. In words, the ranking function is just the probability that $\mathbf{z}$ corresponds to a face versus the probability it was generated by an alternative mechanism (to be discussed further below). The constellation receiving the highest ranking value L will defeat any other constellation in a head-to-head comparison to decide which is more face-like.

The Alternative Hypothesis $\overline{F}$: To implement the ranking function of Equation 19, $p(\mathcal{S}(\mathbf{z})|F)$ should be determined from Equation 8. However, it is unclear how to calculate $p(\mathcal{S}(\mathbf{z})|\overline{F})$ since the distribution conditioned on $\overline{F}$ has yet to be defined. A first guess might be to use the pdf of shape variables resulting from randomly placing N points in the image plane. The problem with this idea is that some candidate constellations may consist of $n < N$ true features and $N - n$ bad features (detector false alarms). We believe the proper approach is to expand $p(\mathcal{S}(\mathbf{z})|\overline{F})$ as follows:

$$\frac{\sum \mathrm{p}(\mathcal{S}(\mathbf{z})|\mathrm{b}_1, \ldots, \mathrm{b}_\mathrm{N}) \cdot \Pr(\mathrm{b}_1, \ldots, \mathrm{b}_\mathrm{N})}{\sum \Pr(\mathrm{b}_1, \ldots, \mathrm{b}_\mathrm{N})} \tag{20}$$

where $b_i = 0$ or 1 depending on whether feature i is a false alarm or the true feature. The summations above go over all N-tuples having at least one $b_i = 0$.

Since the feature detectors essentially work independently in disjoint neighborhoods of the image, the probability $\Pr(b_1, b_2, \ldots, b_N)$ may reasonably be modeled as a product of independent terms $\Pr(b_1, b_2, \ldots, b_N) = \prod_{i=1}^{N} \Pr(b_i)$, where

$$\begin{aligned} \Pr(b_i = 0) &= (1 - \gamma_i) + \gamma_i \cdot \frac{\overline{m}_i - 1}{\overline{m}_i} & (21) \\ \Pr(b_i = 1) &= \frac{\gamma_i}{\overline{m}_i} & (22) \end{aligned}$$

Here $\overline{m}_i$ is the average number of candidates located for the i-th feature and γ_i is the probability that the true location of the i-th feature is detected.

The conditional probability $p(\mathcal{S}(\mathbf{z})|b_1, b_2, \ldots, b_N)$ can be approximated using the density in Equation 8, with the off diagonal elements of Σ that correspond to the bad features ($b_i = 0$) replaced by zeros and the diagonal elements replaced by σ_W^2 (a large value equal to the variance in position of detector false alarms).

Incomplete Constellations: Incomplete constellations that result when a feature is missed by the detectors can be handled with our algorithm. The ranking function of Equation 19 can be written as follows:

$$L(\mathcal{S}(\mathbf{z})) \quad = \quad \frac{p(\mathcal{S}(\mathbf{z_O})|\mathbf{o}, F) \cdot p(\mathbf{o}|F)}{p(\mathcal{S}(\mathbf{z_O})|\mathbf{o}, \overline{F}) \cdot p(\mathbf{o}|\overline{F})} \qquad (23)$$

where $\mathbf{o}$ denotes *which* variables were observed and $\mathbf{z_O}$ denotes the *values* of the observed variables. The marginal distributions over the observed values are obtained using the appropriate submatrices of μ and Σ in the shape density equation. The two features used to determine the transformation to shape variables must be selected from among the observed variables. The choice of which two features to use here *does* affect the value of the probability density function; however, it does not affect the ranking function, which is a ratio of densities.

The probabilities $p(\mathbf{o}|F)$ and $p(\mathbf{o}|\overline{F})$ are determined from the feature detector performance as follows:

$$\mathrm{p}(\mathbf{o}|\mathrm{F}) \quad = \quad \prod_{\mathbf{o}} \gamma_i \cdot \prod_{\overline{\mathbf{o}}} (1 - \gamma_i) \qquad (24)$$

$$p(\mathbf{o}|\overline{F}) \quad = \quad \prod_{\mathbf{o}} \frac{\overline{m}_i - 1}{\overline{m}_i} \cdot \prod_{\overline{\mathbf{o}}} \frac{1}{\overline{m}_i} \qquad (25)$$

where the products are taken over the observed and not-observed variables, respectively. Recall that γ_i is the probability of detection and $\overline{m}_i$ is the average number of false alarms per image. In practice, the probability of detection should be reduced by some factor α to account for the possibility of occlusions.

6 Intelligent Search

Now that we have a method for ranking constellations, we could simply look at every constellation $\mathbf{z} \in \mathbf{Z}$, perform the mapping to shape variables, and evaluate the ranking function $L(\mathcal{S}(\mathbf{z}))$ to find the most face-like constellations. However, the computational complexity of this brute-force approach will limit its applicability to situations in which the number of features N and the number of candidate locations for each feature m_i are small. Therefore, in this section, we define an intelligent search algorithm that significantly reduces the number of constellations that must be checked.

Observe that given the positions of two points on the face, the possible positions of all other features are highly constrained. We may use this intuition as follows. First, we form all partial constellations containing exactly two points; there are $\sum_{(i<j)} m_i m_j$ of these, each of which is considered in turn.

For definiteness, we begin with the partial constellation consisting of the first point $P_{1,1}$ in S_1 and the first point $P_{2,1}$ in S_2. These two points define a mapping from figure space to a set of shape variables. This mapping is applied to the two given features (to place them at fixed reference positions) and to all the candidates for the remaining features. Now, in the standardized reference space, we know where the other features should be and how much uncertainty exists in their location. The uncertainty regions (i.e., the regions where we expect to find the features) can be obtained off-line using either the analytical shape density or the empirical density measured on training faces (discussed further below).

Candidate constellations are formed only from the two given points $P_{1,1}$ and $P_{2,1}$ plus the candidate feature points that fall inside the respective uncertainty regions. (The special point is also permitted.) These constellations are scored with the ranking function and then the next pair of points $P_{1,1}$ and $P_{2,2}$ is considered. This process is repeated for all pairs of two points.

We further limit the complexity of the algorithm in two ways. First, we use a two-tiered thresholding scheme in which the pairs of points discussed above must both be *strong features*, meaning that they exceed a higher detection threshold τ_1 than the other candidate points. Second, we place an upper limit on the range of possible scales. If the distance between the two given points is too large, then candidate constellations are not formed from these. This proves important because it limits the effective size of the search areas in the figure space.

Uncertainty Regions: Figure 3 shows how the uncertainty regions can be determined empirically. For this example we used fifteen facial features that were manually located on 180 training faces. The feature definitions are shown in Figure 3a. The other two figures show the superposition of shape variables as determined from 180 training faces. In Figure 3b, the two eyes were used as reference features and mapped to standard positions; the remaining clouds of points show the variability in the other feature locations (effectively the marginal distributions). Figure 3c is similar except the left eye and nose/lip junction were mapped to reference positions. These empirical uncertainty regions can be encoded using ellipses or more coarsely as circles or rectangles.

7 Experimental Results

We have tested our face localization system on a realistic database of images collected in an ordinary computer laboratory. The database contains a total of 900 images, but we have currently analyzed only the first 150. These images show one subject seated 2–3 meters away from the camera and allowed to move freely, make facial expressions, etc. The background was complicated and continually changing due to people walking around behind the subject.

The parameters for our system were determined from a *separate* database containing 180 images of 18 different people in a studio setting. Five features were used: the two eyes, the nose/lip junction, and

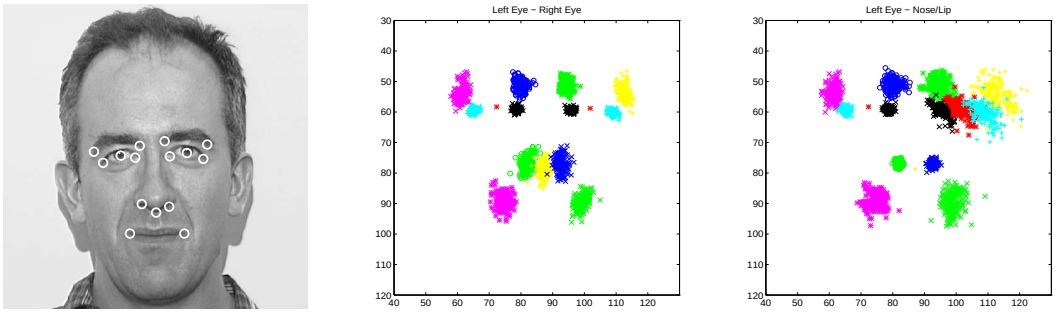

Figure 3: (a) 15 facial features on a typical face, (b) uncertainty in shape variables when the two eyes are used as references, (c) uncertainty when the left eye and nose/lip are used as references.

the two nostrils. The ROC curves obtained from this database (three of these are shown in Figure 1) were used to estimate the detector performance parameters γ_i and $\overline{m}_i$. The estimated γ_i's were reduced by an additional factor $\alpha = 0.9$ to allow for feature occlusion.

The shape parameters were also determined from the training database. In [6], Dryden and Mardia recommended using the set of observed shape vectors in a maximum likelihood procedure to infer the parameters of the shape distribution; however, they acknowledged that the "estimation can be problematic for completely general covariance structures." We implemented the ML procedure and also encountered difficulties. Hence, we decided to use the following direct method. In the studio database all the subjects were imaged from the same distance with the head basically upright, so the effects of scale and rotation are minimal. It is, therefore, reasonable to suppose that after normalization for translation, the feature positions are jointly Gaussian. Thus, the parameters of the Gaussian (ν and Ω) can be estimated directly. Although this method is not strictly correct (e.g., inter-subject scale variations are absorbed into the Gaussian), we believe it is an adequate approximation.

The performance of our face localization system is illustrated in Figure 4 for several images from the lab sequence. Overall, the baseline system correctly localized the face in 84% of the 150 images. As a second experiment, we added the constraint that the head orientation be within 15° of upright; however, this only increased the performance to 87%. In both cases, most of the errors occurred during one portion of the sequence in which the head was significantly rotated in depth. This failure was due to the fact that the feature detectors were designed for frontal views of the face and therefore missed the features when the head was rotated in depth.

8 Conclusions

We described an algorithm to localize quasi-frontal views of faces in cluttered images. The algorithm combines a set of local feature detectors with a statistical description of the spatial arrangement of the features. Initial evaluations on a realistic database of 150 images indicate a correct localization rate of approximately 84% (increased to 87% if the upright-head constraint is enforced). The errors occur primarily on images where the face is not viewed frontally.

We are working on a number of extensions to the system. Among these are using more features, redesigning the feature detectors, and extending the shape statistics framework to affine deformations in order to handle rotations in depth. Also, the efficiency and performance of the algorithm can be improved by incorporating additional information that is available from the feature detectors, e.g., the scale and orientation at which the feature is detected, as well as the quality of the feature match.

Acknowledgement

This work was supported by the Center for Neuromorphic Systems Engineering as a part of the National Science Foundation (NSF) Engineering Research Center Program, by the California Trade and Commerce Agency, Office of Strategic Technology, by an NSF National Young Investigator award, and by a grant from Intel.

We are very grateful to Jitendra Malik for useful discussions and Steve Evans for providing references to the shape statistics literature.

References

[1] Y. Amit and A. Kong. "Graphical Templates for Image Matching". Technical Report 373, Department of Statistics, University of Chicago, August 1993.

[2] F.L. Bookstein. "A Statistical Method for Biological Shape Comparison". *J. Theor. Biol.*, 107:475–520, 1984.

[3] F.L. Bookstein. "Size and Shape Spaces for Landmark Data in Two Dimensions". *Statistical Science*, 1(2):181–242, 1986.

[4] R. Brunelli and T. Poggio. "Face Recognition: Features versus Templates". *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(10):1042–1052, October 1993.

[5] G. Burel and D. Carel. "Detection and Localization of Faces on Digital Images". *Pattern Recognition Letters*, pages 963–967, Oct 1994.

[6] I.L. Dryden and K.V. Mardia. "General Shape Distributions in a Plane". *Adv. Appl. Prob.*, 23:259–276, 1991.

[7] R.O. Duda and P.E. Hart. *"Pattern Classification and Scene Analysis"*. Wiley, 1973.

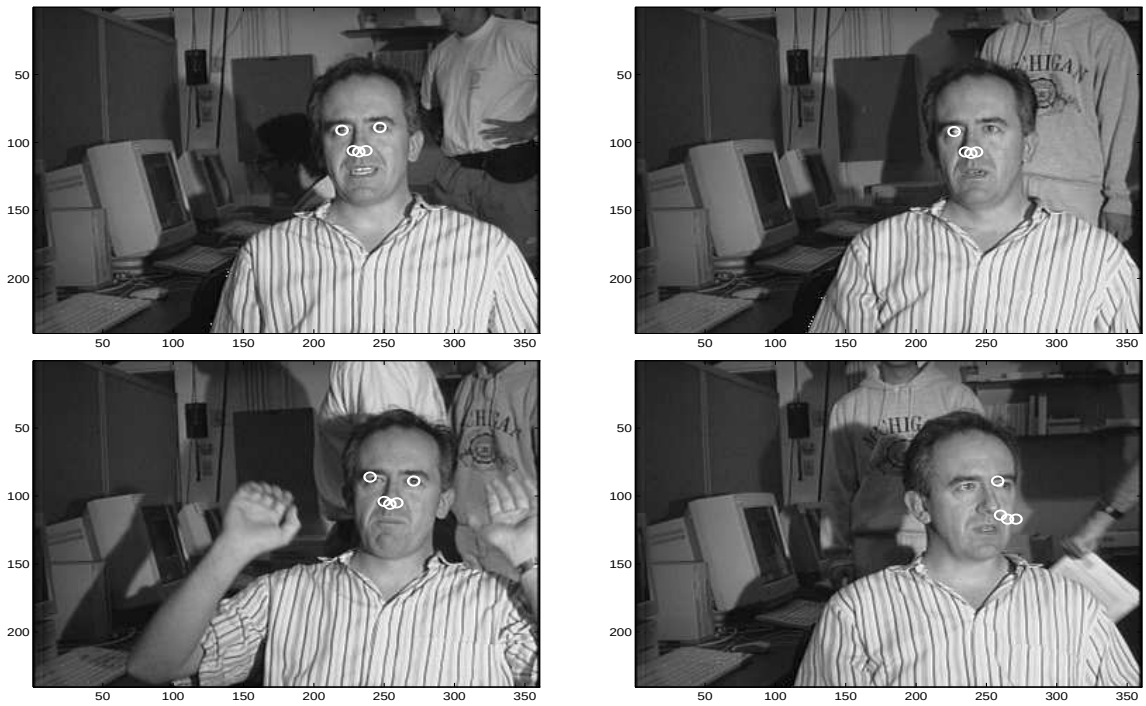

Figure 4: Algorithm performance on selected images. Upper Left: All five features were detected and the face was correctly localized. Upper Right: One eye was not detected, but our system was still able to find the face. Lower Left: The left eye was slightly mislocated, but we count this as a correct match. Lower Right: A case where the algorithm failed. Note that the head is rotated considerably in depth.

[8] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, 2nd edition, 1990.

[9] D. Jones and J. Malik. A computational framework for determining stereo correspondence from a set of linear spatial filters. In *Proc. 2^{nd} Europ. Conf. Comput. Vision, G. Sandini (Ed.), LNCS-Series Vol. 588, Springer-Verlag*, pages 395–410, 1992.

[10] M.S. Kamel, H.C. Shen, et al. "Face Recognition using Perspective Invariant Features". *Pattern Recognition Letters*, 15(9):877–883, 1994.

[11] T. Kanade. "Computer Recognition of Human Faces". *Interdisciplinary Systems Research*, 47, 1977.

[12] M. Kass. "computing visual correspondence". In *Proceedings: Image Understanding Workshop*, pages 54–60, McLean, Virginia, June 1983. Science Applications, Inc.

[13] D.G. Kendall. "Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces". *Bull. London Math Soc.*, 16:81–121, 1984.

[14] D.G. Kendall. "A Survey of the Statistical Theory of Shape". *Statistical Science*, 4(2):87–120, 1989.

[15] H. Le and D.G. Kendall. "The Riemannian Structure of Euclidean Shape Spaces: A Novel Environment for Statistics". *The Annals of Statistics*, 21(3):1225–1271, 1993.

[16] T.K. Leung, M.C. Burl, and P. Perona. "Finding Faces in Cluttered Scenes using Random Labeled Graph Matching". *Proc. 5th Int. Conf. Computer Vision*, 1995.

[17] H-Y.S. Li, Y. Qiao, and D. Psaltis. "Optical Network for Real-time Face Recognition". *Applied Optics*, 32(26):5026–5035, Sept 1993.

[18] K.V. Mardia and I.L. Dryden. "Shape Distributions for Landmark Data". *Adv. Appl. Prob.*, 21:742–755, 1989.

[19] A.J. O'Toole, H. Abdi, K.A. Deffenbacher, and D.Valentin. "low-dimensional representation of faces in higher dimensions of the face space". *J. Opt. Soc. Am. A*, 10(3), 1993.

[20] H.L. Van Trees. *Detection, Estimation, and Modulation Theory:Part 1.* John Wiley and Sons, 1968.

[21] M. Turk and A. Pentland. "Eigenfaces for Recognition". *J. of Cognitive Neurosci.*, 3(1), 1991.

[22] D. Valentin, H. Abdi, et al. "Connectionist Models of Face Processing - A Survey". *Pattern Recognition*, 27(9):1209–1230, September 1994.

[23] G.Z. Yang and T.S. Huang. "Human Face Detection in a Complex Background". *Pattern Recognition*, 27(1):53–63, January 1994.

[24] A.L. Yuille. "Deformable Templates for Face Recognition". *J. of Cognitive Neurosci.*, 3(1):59–70, 1991.