Face Detection Using Mixtures of Linear Subspaces

Ming-Hsuan Yang Narendra Ahuja David Kriegman
Department of Computer Science and Beckman Institute
University of Illinois at Urbana-Champaign, Urbana, IL 61801
Email: {myang1, n-ahuja, kriegman}@uiuc.edu

Abstract

We present two methods using mixtures of linear subspaces for face detection in gray level images. One method uses a mixture of factor analyzers to concurrently perform clustering and, within each cluster, perform local dimensionality reduction. The parameters of the mixture model are estimated using an EM algorithm. A face is detected if the probability of an input sample is above a predefined threshold. The other mixture of subspaces method uses Kohonen's self-organizing map for clustering and Fisher Linear Discriminant to find the optimal projection for pattern classification, and a Gaussian distribution to model the class-conditional density function of the projected samples for each class. The parameters of the class-conditional density functions are maximum likelihood estimates and the decision rule is also based on maximum likelihood. A wide range of face images including ones in different poses, with different expressions and under different lighting conditions are used as the training set to capture the variations of human faces. Our methods have been tested on three sets of 225 images which contain 871 faces. Experimental results on the first two datasets show that our methods perform as well as the best methods in the literature, yet have fewer false detects.

1 Introduction

Images of human faces are central to intelligent human computer interaction. Much research is being done involving face images, including face recognition, face tracking, pose estimation, expression recognition and gesture recognition. However, most existing methods on these topics assume human faces in an image or an image sequence have been identified and localized. To build a fully automated system that extracts information from images of human faces, it is essential to develop robust and efficient algorithms to detect human faces. Given a single image or a sequence of images, the goal of face detection is to identify and locate all of the human faces regardless of their positions, scales, orientations, poses and lighting conditions. This is a challenging problem because human faces are highly non-rigid objects with a high degree of variability in size, shape, color and texture. Most recent methods for face detection can only detect upright, frontal faces under certain lighting conditions. In this paper, we present two face detection methods that use mixtures of linear subspaces to detect faces with different features and expressions, in different poses, and under different lighting conditions.

Since the images of a human face lie in a complex subset of the image space that is unlikely to be modeled by a single linear subspace, we use a mixture of linear subspaces to model the distribution of face and nonface patterns. The first detection method is an extension of factor analysis. Factor analysis (FA), a statistical method for modeling the covariance structure of high dimensional data using a small number of latent variables, has analogue with principal component analysis (PCA). However PCA, unlike FA, does not define a proper density model for the data since the cost of coding a data point is equal anywhere along the principal component subspace (i.e., the density is unnormalized along these directions). Further, PCA is not robust to independent noise in the features of the data since the principal components maximize the variances of the input data, thereby retaining unwanted variations. Hinton et al. have applied FA to digit recognition and they compare the performance of PCA and FA models [10]. A mixture model of factor analyzers has recently been extended [7] and applied to face recognition [6]. Both studies show that FA performs better than PCA in digit and face recognition. Since pose, orientation, expression, and lighting affect the appearance of a human face, the distribution of faces in the image space can be better represented by a mixture of subspaces where each subspace captures certain characteristics of certain face appearances. We present a probabilistic method that uses a mixture of factor analyzers (MFA) to detect faces with wide variations. The parameters in the mixture model are estimated using an EM algorithm.

The second method that we present uses Fisher Linear

Discriminant (FLD) to project samples from a high dimensional image space to a lower dimensional feature space. Recently, the Fisherface method has been shown to outperform the widely used Eigenface method in face recognition [2]. The reason for this is that FLD provides a better projection than PCA for pattern classification. In the second proposed method, we decompose the training face and nonface samples into several classes using Kohonen's Self Organizing Map (SOM). From these labeled classes, the within-class and between-class scatter matrices are computed, thereby generating the optimal projection based on FLD. For each subspace, we use a Gaussian to model each class-conditional density function where the parameters are estimated based on maximum likelihood [5]. To detect faces, each input image is scanned with a rectangular window in which the class-dependent probability is computed. The maximum likelihood decision rule is used to determine whether a face is detected or not.

To capture the variations in face patterns, we use a set of 1,681 face images from Olivetti [20], UMIST [8], Harvard [9], Yale [2] and FERET [15] databases. Both methods have been tested using the databases in [18] [22] to compare their performances with other methods. Our experimental results on the data sets used in [18] [22] (which consist of 225 images with 619 faces) show that our methods perform as well as the reported methods in the literature, yet with fewer false detects. To further test our methods, we collect a set of 80 images containing 252 faces. This data set is rather challenging since it contains profile faces, faces with expressions and faces with heavy shadows. Our methods are able to detect most of these faces regardless of their poses, facial expressions and lighting conditions. Furthermore, our methods have fewer false detects than other methods.

2 Related Work

Numerous intensity-based methods have been proposed recently to detect human faces in a single image or a sequence of images. In this section, we give a brief review of intensity-based face detection methods. See [23] for a comprehensive survey on face detection. Sung and Poggio [22] report an example-based learning approach for locating vertical frontal views of human faces. They use a number of Gaussian clusters to model the distributions of face and nonface patterns. For computational efficiency, a subspace spanned by each cluster's eigenvectors is then used to compute the evidence of a face. A small window is moved over all portions of an image to determine, based on distance metrics measured in the subspaces, whether a face exists in each window. In [16], a detection algorithm is proposed that combines template matching and feature-based detection method using hierarchical Markov random fields (MRF) and maximum *a posteriori* probability (MAP) estimation. The watershed algorithm is used to segment an image at some fixed scales and to generate an image pyramid. To reduce the search, a heuristic is used to select areas where faces may appear. Layered processes are used in a MRF to reflect *a priori* knowledge about the spatial relationships between facial features (eye, mouth and the whole face) which are identified by template matching and gradient of intensity. Detection decision is based on MAP estimation. Colmenarez and Huang [3] apply Kullback relative information for maximal discrimination between positive and negative examples of faces. They use a family of discrete Markov processes to model the face and background patterns and estimate the density functions. Detection of a face is based on the likelihood ratio computed during training. Moghaddam and Pentland [12] propose a probabilistic method that is based on density estimation in a high dimensional space using an eigenspace decomposition. In [18], Rowley et al. use an ensemble of neural networks to learn face and nonface patterns for face detection. Schneiderman et al. describe a probabilistic method based on local appearance and principal component analysis [21]. Their method gives some preliminary results on profile face detection. Finally, hidden Markov models [17], higher order statistics [17], and support vector machines (SVM) [13] [14] have also been applied to face detection and demonstrated some success in detecting upright frontal faces under certain lighting conditions.

3 Mixture of Factor Analyzers

In the first method, we fit the mixture model of factor analyzers to the training samples using an EM algorithm and obtain a distribution of face patterns. To detect faces, each input image is scanned with a rectangular window in which the probability of the current input being a face pattern is calculated. A face is detected if the probability is above a predefined threshold. We briefly describe factor analysis and a mixture of factor analyzers in this section. The details of these models can be found in [1] [7].

3.1 Factor Analysis

Factor analysis is a statistical model in which the observed vector is partitioned into an unobserved systematic part and an unobserved error part. The systematic part is taken as a linear combination of a relatively small number of unobserved factor variables while the components of the error vector are considered as uncorrelated or independent. From another point of view, factor analysis gives a description of the interdependence of a set of variables in terms of the factors without regard to the observed variability. In this model, a d-dimensional real-valued observable data vector x is modeled using a p-dimensional vector of real-valued factors z where p is generally much smaller than d. The generative model is given by:

$$x = \Lambda z + u \tag{1}$$

where Λ is known as the *factor loading matrix*. The factors z are assumed to be $\mathcal{N}(0, I)$ distributed (zero-mean independent normals with unit variance). The d-dimensional random variable u is distributed $\mathcal{N}(0, \Psi)$ where Ψ is a diagonal matrix, due to the assumption that the observed variables are independent given the factors. According to this model, x is therefore distributed with zero mean and covariance $\Sigma = \Lambda\Lambda^T + \Psi$. The goal of factor analysis is to find the Λ and Ψ that best model the covariance structure of x. The factor variables z model correlations between the elements of x, while the u variables account for independent noise in each element x. The p factors play the same role as the principal components in PCA, i.e., they are informative projections of the data. Given Λ and Ψ, the expected value of the factors can be computed through the linear projections:

$$E[z|x] = \beta x \tag{2}$$

$$E[zz^T|x] = I - \beta\Lambda + \beta x x^T \beta^T \tag{3}$$

where $\beta = \Lambda^T \Sigma^{-1}$.

3.2 Mixture Model

In this section, we consider a mixture of m factor analyzers (indexed by $f_j, j = 1, \ldots, m$) where each factor analyzer has the same number of p factors and each factor analyzer has a different mean μ_j. The generative model obeys the mixture distribution:

$$P(x) = \sum_{j=1}^{m} \int P(x|z, f_j) P(z|f_j) P(f_j) dz \tag{4}$$

where

$$P(z|f_j) = P(z) = \mathcal{N}(0, I) \tag{5}$$

$$P(x|z, f_j) = \mathcal{N}(\mu_j + \Lambda_j z, \Psi) \tag{6}$$

The parameters of this mixture model are $\{(\mu_j, \Lambda_j)_{j=1}^m, \pi, \Psi\}$ where π is the vector of adaptable mixing proportions, $\pi_j = P(f_j)$. The latent variables in this model are the factors z and the mixture indicator variable f_j, where $f_j = 1$ when the data point is generated by the first factor analyzer.

Given a set of training images, the EM algorithm [4] is used to estimate $\{ (\mu_j, \Lambda_j)_{j=1}^m, \pi, \Psi\}$. For the E-step of the EM algorithm, we need to compute expectations of all the interactions of the hidden variables that appear in the log likelihood,

$$E[f_j z|x_i] = E[f_j|x_i] E[z|f_j, x_i] \tag{7}$$

$$E[f_j zz^T|x_i] = E[f_j|x_i] E[zz^T|f_j, x_i] \tag{8}$$

Defining

$$h_{ij} = E[f_j|x_i] \propto P(x_i, f_j) = \pi_j \mathcal{N}(x_i - \mu_j, \Lambda_j \Lambda_j^T + \Psi) \tag{9}$$

and using equations (2) and (6), we obtain

$$E[f_j z|x_i] = h_{ij} \beta_j (x_i - \mu_j) \tag{10}$$

where $\beta_j \equiv \Lambda_j^T (\Lambda_j \Lambda_j^T)^{-1}$. Similarly, using equations (3) and (8), we obtain

$$E[f_j zz^T|x_i] = h_{ij}(I - \beta_j \Lambda_j + \beta_j (x_i - \mu_j)(x_i - \mu_j)^T \beta_j^T) \tag{11}$$

The EM algorithm for mixture of factor analyzers can be stated as follows:

- **E-step**: Compute $E[f_j|x_i]$, $E[z\,|f_j, x_i]$ and $E[zz^T|\,f_j, x_i]$ for all data points i and mixture components j.
- **M-step**: Solve a set of linear equations for π_j, Λ_j, μ_j and Ψ.

The mixture of factor analyzers is essentially a reduced dimensionality mixture of Gaussians. Each factor analyzer fits a Gaussian to a portion of the data, weighted by the posterior probabilities, h_{ij}. Since the covariance matrix for each Gaussian is specified through the lower dimensional factor loading matrices, the model has $mpd + d$, rather than $md(d+1)/2$ parameters dedicated to modeling covariance structure in high dimensions.

3.3 Detecting Face Patterns

To detect faces, each input image is scanned with a rectangular window in which the probability of there being a face pattern is estimated as given in equation (4). A face is detected if the probability is above a predefined threshold. In order to detect faces of different scales, each input image is repeatedly subsampled by a factor of 1.2 and scanned through for 10 iterations.

4 Mixture of Linear Spaces Using Fisher Linear Discriminant

In the second mixture model, we first use Kohonen's self-organizing map [11] to divide the face and nonface samples into c_1 face classes and c_2 nonface classes, thereby generating labels for the samples. Next, Fisher projection is computed based on all $c_1 + c_2$ classes to maximize the ratio of the between-class scatter (variance) and the within-class scatter (variance). The now labeled training set is projected from a high dimensional image space to a lower dimensional feature space, and a Gaussian distribution is used to model the class-conditional density function for each class where the parameters are estimated using the maximum likelihood principle. For detection, the conditional probability of each sample given each class is computed and the maximum likelihood principle is used to decide to which class the sample belongs. In our experiments, the reason that we choose 25 face and 25 nonface classes is because of the size of training set. If the number of classes is too small, the clustering results may be poor. On the other hand, we may not have enough samples to estimate the class-conditional density function well if we choose a large number of classes.

4.1 Labeling Samples Using SOM

In applying Fisher Linear Discriminant to find a projection, we need to know the class label of each training sample. However, such information is not available in the training samples. Therefore, we use Kohonen's Self-Organizing Map [11] to divide face samples into a finite number of classes. In our experiments, we divide the face sample images into 25 classes. After training, the final weight vector for each node is the centroid of the class, i.e., the prototype vector, which corresponds to the prototype of each class. The same procedure is applied to nonface samples. Figure 1 shows the prototypical face of each class. It is clear that the sample face images with different poses and under different lighting conditions (intensity increases from the lower right corner to the upper left corner) have been classified into different classes. Note that the SOM algorithm also places the prototypes in the two dimensional feature map, shown in 1, in accordance with their topological relationships in the image space. In other words, prototype vectors corresponding to nearby points on the feature map grid have nearby locations in the high dimensional image space (e.g., nearby prototypes have similar intensity and pose).

4.2 Fisher Linear Discriminant

While PCA is commonly used to project face patterns from a high dimensional image space to a lower dimensional feature space, a drawback of this approach is that it defines a subspace such that it has the greatest variance of the projected sample vectors among all the subspaces. However, such projection is not suitable for classification since it may contain principal components which retain unwanted large variations. Therefore, the classes in the projected space may not be well clustered and instead smeared together [2] [6] [10]. Fisher Linear Discriminant is an example of a class specific method that finds the optimal projection for classification. Rather than finding a projection that maximizes the projected variance, FLD determines a projection, $z = W_{FLD}^T x$, that maximizes the ratio be-

Figure 1. Prototype of each face class.

tween the between-class scatter (variance) and the within-class scatter (variance). Consequently, classification is simplified in the projected space. Recently, it has been demonstrated that the Fisherface method outperforms the Eigenface method in face recognition [2].

Consider a c-class problem, let the between-class scatter matrix be defined as

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T \quad (12)$$

and the within-class scatter matrix be defined as

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \quad (13)$$

where μ is the mean of all samples, μ_i is the mean of class X_i, and N_i is the number of samples in class X_i. The optimal projection W_{FLD} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected sampled, i.e.,

$$W_{FLD} = \arg\max_w \frac{|W^T S_B W|}{|W^T S_W W|} = [w_1 \; w_2 \; \ldots \; w_m] \quad (14)$$

where $\{w_i | i = 1, 2, \ldots, m\}$ is the set of generalized eigenvectors of S_B and S_W, corresponding to the m largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \ldots, m\}$. However, the rank of S_B is $c - 1$ or less because it is the sum of c matrices of rank one or less. Thus, the upper bound on m is

$c-1$ [5]. See [2] for details about a method to overcome singularity problems in computing W_{FLD}.

4.3 Class-Conditional Density Function

Once W_{FLD} is computed, the now labeled training set is projected to the $c-1$ dimensional feature space, i.e., $z = W_{FLD}^T\ x$, and a Gaussian distribution is used to model each class-conditional density (CCD) function, i.e., $P(z|X_i) = \mathcal{N}(\mu_{X_i}, \Sigma_{X_i})$ where $i = 1, \ldots, c$. The parameters, $\theta_{X_i} = (\mu_{X_i}, \Sigma_{X_i})\}$ of each CCD are the maximum likelihood estimates, i.e.,

$$\hat{\mu}_{X_i} = \frac{1}{|X_i|} \sum_{z_k \in X_i} z_k \tag{15}$$

and

$$\hat{\Sigma}_{X_i} = \frac{1}{|X_i|} \sum_{z_k \in X_i} (z_k - \hat{\mu}_{X_i})(z_k - \hat{\mu}_{X_i})^T \tag{16}$$

4.4 Detecting Face Patterns

Each input image is scanned with a rectangular window to determine whether a face exists in the window or not. The decision rule for deciding whether an input window contains a face or not is based on maximum likelihood,

$$X^* = \arg\max_{X_i} P(z|X_i) \tag{17}$$

To detect faces of different scales, each input image is repeatedly subsampled by a factor of 1.2 and scanned through for 10 iterations.

5 Experiments

For training, we use a set of 1,681 face images (collected from Olivetti [20], UMIST [8], Harvard [9], Yale [2] and FERET [15] databases) which have wide variations in pose, facial expression and lighting condition. In the second mixture method, we start with 8,422 nonface examples from 400 images of landscapes, trees, buildings, etc. Although it is extremely difficult to collect a representative set of nonface examples, the bootstrap method similar to [22] is used to include more nonface examples during training. Each face sample is manually cropped and normalized such that it is aligned vertically and its size is 20×20 pixels. To make the detection method less sensitive to scale and rotation variation, 10 face examples are generated from each original sample. The images are produced by randomly rotating the images by up to 15 degrees with scaling between 80% and 120%. This produces 16,810 face samples.

We test both methods on the three sets of images collected by Rowley [18], Sung [22] and ourselves. In our experiments, a detected face is a successful detect is if the subimage contains eyes and mouth. Otherwise, it is a false detect. The detection rate is the ratio between the number of successful detects and the number of faces in the test set. Table 1 shows the detection rates of our methods and the reported results of several detection methods on the test set in [18]. Experimental results on test set 1, which consists of 125 images (483 faces) excluding 5 images of hand drawn faces, show that our methods have comparable detection performance with other methods, yet with fewer false detects. Table 1 also shows the our experimental results on the test set of Sung and Poggio [22] which consists of 20 images excluding 3 images of line drawn faces (136 faces). Both of our methods consistently perform well and have few false detects.

Test set 3 consists of 80 images (252 faces), collected from the World Wide Web, with different poses, expressions and faces with heavy shadows. The detection rates are 86.7% and 88.2% for MFA and FLD-based methods. The number of false detects are 45 and 40, respectively. Both methods perform equally well in detecting these faces though the FLD-based method performs slightly better than the first one. Figures 2 and 3 show the results of our methods on some test images. See the web page mentioned above for more results. Notice that there is a false detect in the upper left corner of the image in Figure 2 since one window resembles a face. Also notice that our methods can detect, up to certain degree, profile faces and faces with heavy shadows. However occluded, rotated faces or faces with sunglasses cannot be detected effectively by both methods due to lack of such examples in the training sets. None of the existing detection methods cannot effectively detect these types of faces except one recent method [19] seems to able to detect rotated faces. Nevertheless, this method cannot detect occluded faces or face with heavy shadows.

6 Discussion and Conclusion

We have described methods using mixture of linear subspaces methods to detect human faces regardless of their poses, facial expressions and lighting conditions. Both methods find better projection than PCA for pattern classification, thereby facilitating detection of face and nonface patterns. The first method fits a mixture of factor analyzers to estimate the density function of face images, and the second method uses Self-Organizing Map to partition the training set into classes and Fisher Linear Discriminant to find the optimal projection for classification. Experimental results on three sets of images demonstrate that both methods perform as well as the best algorithms in detecting upright frontal faces, yet with fewer false detects.

The contributions of this paper can be summarized as follows. First, we introduce projection methods that per-

Table 1. Experimental results on images from test set 1 (125 images with 483 faces) in [18] and test set 2 (20 images with 136 faces) in [22] (see text for details).

Method	Test Set 1		Test Set 2	
	Detect Rate	False Detects	Detect Rate	False Detects
Mixture of factor analyzers	92.3%	82	89.4%	3
Fisher linear discriminant	93.6%	74	91.5%	1
Distribution-based [22]	N/A	N/A	81.9%	13
Neural network [18]	92.5%	862	90.3%	42
Naive Bayes [21]	93.0%	88	91.2%	12
Kullback relative information [3]	98.0%	12758	N/A	N/A
Support vector machine [13]	N/A	N/A	74.2%	20

Figure 2. Sample experimental results using mixture of factor analyzers on images from three test sets. Every detected face is shown with an enclosing window.

Figure 3. Sample experimental results using mixture of subspaces with Fisher Linear Discriminant on images from three test sets. Every detected face is shown with an enclosing window.

form better than PCA. Consequently, the classification result in the linear subspace is better. Second, we apply mixture models such that the linear subspaces can better capture the variations of face patterns. Although some methods [12][22] have applied mixture model, they use PCA for projection which is suboptimal for classification in subspaces. On the other hand, it is not clear how SVM performs in face detection since the study in [13] has applied SVM on a rather small test set with 136 faces. It will be of great interest to compare our methods with SVM on a large test set since SVM aims to find the optimal hyperplane that minimizes the generalization error under the theoretical upper bounds.

Acknowledgments

D. Kriegman was supported in part by the Army Reseearch Office under ARO Y-99-0006 and the National Eye Institute.

References

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley, New York, 1984.

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[3] A. J. Colmenarez and T. S. Huang. Face detection with information-based maximum discrimination. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 782–787, 1997.

[4] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Joruanl of the Royal Statistical Society*, 39(1):1–38, 1977.

[5] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiely, New York, 1973.

[6] B. J. Frey, A. Colmenarez, and T. S. Huang. Mixtures of local subspaces for face recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 32–37, 1998.

[7] Z. Ghahramani and G. E. Hinton. The em algorithm for mixtures of factor analyzers. Technical Report CRG-TR-96-1, Department of Computer Science, University of Toronto, 1996. Available at ftp://ftp.cs.toronto.edu/pub/zoubin/tr-96-1.ps.gz.

[8] D. B. Graham and N. M. Allinson. Characterizing virtual eigensignatures for general purpose face recognition. In H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie, and T. S. Huang, editors, *Face Recognition: From Theory to Applications*, volume 163 of *NATO ASI Series F, Computer and Systems Sciences*, pages 446–456. Springer, 1998.

[9] P. Hallinan. *A Deformable Model for Face Recognition Under Arbitrary Lighting Conditions*. PhD thesis, Harvard University, 1995.

[10] G. E. Hinton, P. Dayan, and M. Revow. Modeling the manifolds of images of handwritten digits. *IEEE Trans. Neural Networks*, 8(1):65–74, 1997.

[11] T. Kohonen. *Self Organizing Map*. Springer, 1996.

[12] B. Moghaddam and A. Pentland. Probabilistic visual learning for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.

[13] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 130–136, 1997.

[14] C. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 555–562, 1998.

[15] P. J. Phillips, H. Moon, S. Rizvi, and P. Rauss. The feret evaluation. In H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman-Soulie, and T. S. Huang, editors, *Face Recognition: From Theory to Applications*, volume 163 of *NATO ASI Series F, Computer and Systems Sciences*, pages 244–261. Springer, 1998.

[16] R. J. Qian and T. S. Huang. Object detection using hierarchical mrf and map estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 186–192, 1997.

[17] A. N. Rajagopalan, K. S. Kumar, J. Karlekar, R. Manivasakan, and M. M. Patil. Finding faces in photographs. In *Proceedings of the Sixth International Conference on Computer Vision*, pages 640–645, 1998.

[18] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

[19] H. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 38–44, 1998.

[20] F. S. Samaria. *Face Recognition Using Hidden Markov Models*. PhD thesis, University of Cambridge, 1994.

[21] H. Schneiderman and T. Kanade. Probabilistic modeling of local appearance and spatial relationships for object recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 45–51, 1998.

[22] K.-K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):39–51, 1998.

[23] M.-H. Yang, N. Ahuja, and D. Kriegman. A survey on face detection methods. 2000. To be submitted.