FACE DETECTION USING A MIXTURE OF FACTOR ANALYZERS

Ming-Hsuan Yang[⋆] *Narendra Ahuja*[†] *David Kriegman*[⋆]

Department of Computer Science[⋆] Department of Electrical and Computer Engineering[†]
University of Illinois at Urbana-Champaign, Urbana, IL 61801
E-mail: {myang1, n-ahuja, kriegman}@uiuc.edu
Web Page: http://vision.ai.uiuc.edu

ABSTRACT

We present a probabilistic method to detect human faces using a mixture of factor analyzers. One characteristic of this mixture model is that it concurrently performs clustering and, within each cluster, local dimensionality reduction. 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. In order to fit the mixture model to the sample face images, the parameters are estimated using an EM algorithm. Experimental results show that faces in different poses, with different facial expressions, and under different lighting conditions are accurately detected by our method.

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 nonrigid 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 a method that uses a mixtures of factor analyzers 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. Factor analysis (FA), a statistical method for modeling the covariance structure of high dimensional data using a small number of latent variables, has some analogues 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 maximizes 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 [9]. A mixture model of factor analyzers has recently been extended [6] and applied to face recognition [5]. 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.

To capture the variations in face patterns, we use a set of 1,681 face images from Olivetti [17], UMIST [7], Harvard [8], Yale [2] and FERET [13] databases. Our method has been tested using the databases in [16] [19] to compare their performances with other methods. Our experimental results on the data sets used in [16] [19] (which consist of 145 images with 619 faces) show that the proposed method performs as well as the reported methods in the literature, yet with fewer false detects. To further test our method, 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 method is able to detect most of these faces regardless of their poses, facial expressions and lighting conditions. Furthermore, our face detector has 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 [20] for a comprehensive survey on face detection. Sung and Poggio [19] 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 [14], 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 [10] propose a probabilistic method that is based on density estimation in a high dimensional space using an eigenspace decomposition. In [16], 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 [18]. Their method gives some preliminary results on profile face detection. Finally, hidden Markov models [15], higher order statistics [15], and support vector machines (SVM) [11] [12] 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

We fit a 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] [6].

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:

$$E[z|x] = \beta x \quad (2)$$

$$E[zz^T|x] = I - \beta\Lambda + \beta x x^T \beta^T \quad (3)$$

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

3.2. Mixture Model

In this section, we consider a mixture of m factor analyzers (indexed by $j, j = 1, \ldots, m$) 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:

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

where

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

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

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, $\pi_j = P(j)$. The latent variables in this model are the factors z and the mixture indicator variable j, where $j = 1$ 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.

$$E[jz|x^{(t)}] = E[j|x^{(t)}]E[z|j, x^{(t)}] \quad (7)$$

$$E[jzz^T|x^{(t)}] = E[j|x^{(t)}]E[zz^T|j, x^{(t)}] \quad (8)$$

Defining

$$h_{ij} = E[j|x^{(t)}] \propto P(x^{(t)}, j) = \pi_i \mathcal{N}(x^{(t)} - \mu_j, \Lambda_j \Lambda_j^T + \Psi) \quad (9)$$

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

$$E[jz|x^{(t)}] = h_{ij}\beta_j(x^{(t)} - \mu_j) \quad (10)$$

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

$$E[jzz^T|x^{(t)}] = h_{ij}(I - \beta_j\Lambda_j + \beta_j(x^{(t)} - \mu_j)(x^{(t)} - \mu_j)^T\beta_j^T) \quad (11)$$

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

- **E-step**: Compute $E[j|x^{(t)}]$, $E[z\,|x^{(t)}\,, j]$ and $E[zz^T|\, x^{(t)}, j]$ 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. EXPERIMENTAL RESULTS

We use a set of 1,411 faces images from Olivetti [17], UMIST [7], Harvard [8], Yale [2] and FERET [13] data sets. to capture the variations in face patterns. Each image is manually cropped and normalized such that all the images are aligned and the size of each image is 20×20. We fit a mixture model of factor analyzers to these face samples using the EM algorithm described in Section 3.2 and obtain a distribution of face images as equation (4). To detect faces, each input image is scanned with a rectangular window in which the probability of their being face pattern is calculated. A face is detected if the probability of being face pattern is above a predefined threshold. In order to detect faces of different scales, each input image is subsampled by a factor of 1.2 and scanned through for 10 iterations. We test the resulting mixture model on both the training face images and test sets of images used by Sung [19] and Rowley [16]. Figure 1 shows the results of some test images (See http://vision.ai.uiuc.edu/ mhyang/ mfa. html for more results). Note that most profile faces and faces with shadows are detected by our method.

It is difficult to evaluate the performance of different methods even though they use the same benchmark data sets because different criteria (e.g. training time,

Figure 1: Sample experimental results using our method on images from two benchmark data sets. Every detected face is shown with an enclosing window.

number of training examples involved, execution time, number of scanned windows in detection) can be applied to favor one over another. Also, one can tune the parameters of one's method to increase the detect rates while increasing also the number of false detects. The methods using neural networks [16], and naive Bayes [18] report several experimental results based on different sets of parameters. Table 1 summarizes the best detect rates and corresponding false detects of these methods in two test sets. Experimental results on test set 1, which consists of 130 images (483 faces) excluding 5 images of hand drawn faces, show that our method has detection performance comparable to [16] and [18]. Test set 2 consists of 20 images (136 faces) with different poses, expressions and faces with shadows. Our method performs equally well in detecting these faces. It is not clear how other methods perform in detecting profile faces, face with expressions, and faces with shadows.

5. CONCLUSION

We have described a probabilistic method to detect human faces regardless of their poses, facial expressions and effects of lighting conditions. Our method fits a mixture of factor analyzers to estimate the density function of face images. Experimental results show that our method has performance comparable to some of the best algorithms currently available in detecting upright frontal faces and can detect faces in different poses and facial expressions regardless of lighting conditions.

The contributions of this paper are summarized as follows. First, we introduce a projection method that performs better than PCA. Consequently, the classification in the linear subspace is better. Second, we apply a mixture model such that the linear subspaces can better capture the variations of face patterns. Although some methods [10] [19] have applied mixture models, they use PCA for projection which is not optimal for classification in subspaces. On the other hand, it is not clear how SVM performs in face detection since the study in [11] has applied SVM on a rather small test set with 136 faces. It will be of great interest to compare our method 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.

Table 1: Experimental results on face detection

	Proposed Method		Rowley [16]		Schneiderman [18]	
	Detection Rate	False Alarms	Detection Rate	False Alarms	Detection Rate	False Alarms
Test set 1 (125 images, 483 faces) [16]	92.3%	82	92.5%	862	93.0%	88
Test set 2 (20 images, 136 faces) [19]	89.4%	3	76.8%	5	89.8%	3

Acknowledgments

N. Ahuja and M.-H. Yang were supported by ONR and Advanced Telecommunication Research International. D. J. Kriegman was supported by NSF under an NYI, IRI-9257990 and by ARO grant DAAG55-98-1-0168.

6. 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 Trans. Pattern Anal. Mach. Intell.*, 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] 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.

[6] Zoubin Ghahramani and Geoffrey E. Hinton. The em algorithm for mixtures of factor analyzers. Technical Report CRG-TR-96-1, Department of Computer Science, University of Toronto, 1996.

[7] Daniel B. Graham and Nigel 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.

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

[9] 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.

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

[11] 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.

[12] Constantine Papageorgiou, Michael Oren, and Tomaso Poggio. A general framework for object detection. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 555–562, 1998.

[13] P. Jonathon 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.

[14] 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.

[15] 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.

[16] Henry Rowley, Shumeet Baluja, and Takeo Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

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

[18] 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.

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

[20] Ming-Hsuan Yang, Narendra Ahuja, and David Kriegman. A survey on face detection methods. 1999. Working paper, available at http: // vision.ai.uiuc.edu/ mhyang/ papers/ survey.ps.gz.