Face Detection with Clustering, LDA and NN

Hiroyuki Kobayashi and Qiangfu Zhao

Abstract— **In this paper, we study neural network (NN) based face detection. The main purpose is to reduce the complexity of the NN detector and thus speedup training and detection through linear dimensionality reduction. Neither principle component analysis (PCA) nor linear discriminant analysis (LDA) is good for this purpose. PCA often reduces descriptive and discriminative information together. On the other hand, LDA maps all data into a $(C-1)$-Dimensional feature space, where $C = 2$ for face detection. Since face detection is highly non-linear, classification in 1-dimensional space is clearly not enough. In this paper, we propose a new method for face detection. In this method, the problem is first changed to a multi-class problem using clustering. A modified LDA (m-LDA) is then proposed to extract useful features. The NN is used to make the final decision. Here, m-LDA is used to minimize the within-cluster variance between all "face" clusters; and maximize the between-cluster variance between all "face" clusters and all "non-face" data points. The feature space so obtained has a dimensionality less than that of the original problem. To validate the proposed method, we conducted several experiments with four methods, namely the proposed method, NN, PCA+NN, and LDA+NN. Results show that the proposed method can provide lower false positive and false negative errors for test images.**

I. INTRODUCTION

In resent years, images containing human faces are becoming more and more important in different security systems, and this is one of the main reasons why human face related topics, say, face detection, face recognition, expression recognition, emotion recognition, and so on, have attracted great attention of many researchers. Face detection is to detect the positions and/or sizes of human faces and segment (if necessary) them from given images (photographs or video images). Face detection is the essential step for face recognition and other processing.

It is known that face detection is a very hard problem. In fact, different expressions, different lighting conditions, different colors, etc., can create infinitely many different face patterns. Since non-face patterns are even more complex, the decision boundary between face class and non-face class is highly non-linear. Therefore, pure linear approaches such as principal component analysis (PCA) and linear discriminant analysis (LDA) are not suitable for face detection.

So far, many approaches for face detection have been proposed in the literature (see [1] and references therein). Among them, kernel based non-linear approaches such as support vector machine (SVM) and kernel LDA (k-LDA) seem to be the best in the sense that they can provide the lowest errors rates [2]-[5]. However, kernel based systems are usually complex and time consuming because they employ a great number of training examples for making decisions. Neural network (NN) based approach is a good compromise [6]. Usually, an NN detector can provide a good performance with a much simpler structure compared with SVM or k-LDA. However, for appearance based face detection, since all pixels in the sub-image under consideration are used as the inputs of the network, the network structure is still very complex and the computational cost is high both for training and for detection.

One way for simplifying the NN structure and reducing the computational cost of the NN based approach is to combine NN with linear approaches. The basic idea here is to reduce the dimensionality of the input vector using some linear approach, and then detect the faces using the NN. PCA [7], and LDA [8] are two typical methods for dimensionality reduction. However, they are not good for our purpose. In fact, PCA often reduces important discriminative information, and may result in poor performance for face detection. On the other hand, LDA maps all patterns into a $C-1$ dimensional space, where C is the number of classes. Since face detection is a two-class problem (i.e., $C = 2$), LDA maps all patterns into a 1-D space. As pointed out earlier, the decision boundary between face and non-face classes is highly non-linear, and classification in the mapped 1-D space is clearly not enough at all.

To solve the above problem, we can adapt some clustering method to divide all face and non-face patterns into C clusters. For example, we may divide all face patterns into k_1 clusters, and all non-face patterns into k_2 clusters ($C = k_1 + k_2$) [9]. Then, we can consider the face detection problem as a C-class classification problem. By so doing, a sub-image can be compressed using the LDA to a $C-1$ dimensional vector, which is then used as the input of the NN detector.

Although very good results were reported in [9], experimental results on the databases used in this study show that the above clustering based LDA+NN approach is not as good as the NN approach (see the results to be given in Section 4). The main reason we think is that C clusters are not enough to cover all kinds of face and non-face patterns. Note that the correlation between all face patterns should be much higher than that between all non-face patterns, it is reasonable to use a much larger number of clusters for non-face patterns. That is, k_2 should be much larger than k_1. As an extreme case, in this paper, we consider to use C clusters for the face patterns, and **assign a cluster to each non-face pattern**. The effectiveness of this approach is verified

H. Kobayashi and Q. Zhao are with School of Computer Science and Engineering, The University of Aizu, Japan, Tsuruga, Ikkimachi, Aizuwakamatsu-city, Fukushima prefecture, 965-8580 Japan {m5091127,qf-zhao}@u-aizu.ac.jp

1-4244-0991-8/07/$25.00/©2007 IEEE

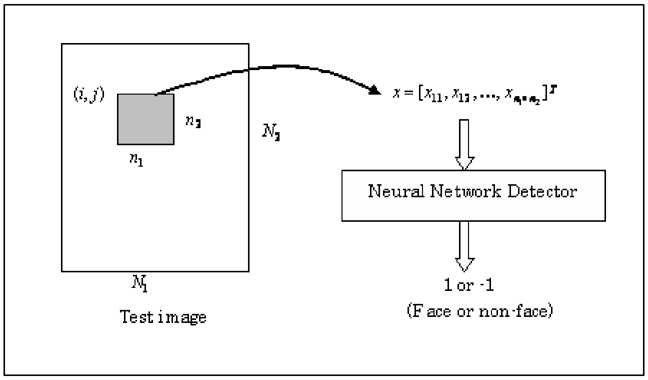

Fig. 1. Illustration of NN based face detection from still images

through experiments.

The rest of this paper is organized as follows. In the next section, we provide a brief introduction to NN based face detection and some dimensionality reduction methods. Section III proposes the new approach for face detection. Basically, this is also an NN based approach, but uses a modified LDA for feature extraction. Section IV gives the experimental results, and Section V is the conclusion.

II. PRELIMINARIES

A. NN Based Face Detection

In this paper, we consider only the case of face detection from still images. Fig. 1 shows the basic process of face detection based on NN. The input vector of the NN consists of all pixel of the sub-image scanned row by row. If the sub-image is a face, the NN detector should output 1; otherwise, the NN should output -1. To detect all possible faces, we must move the sub-image from upper-left to bottom-right, and pixel by pixel. We must also change the size of the sub-image from the minimum possible face size to the maximum one. Clearly, it is not a good strategy to change the number of inputs of the NN detector. Instead, we change the size of the test image. In this paper, the sub-image size is fixed to 19×19, and the test image is scaled down by 1.2 each time when all possible sub-images are tested.

B. Training of the NN Detector

To train the NN detector, we must prepare a set of training data with known labels. In our experiments, each training datum is a 19×19 sub-image. Therefore, the dimensionality of the input vector of the NN is 361. The NN model we used is the multilayer feedforward NN. The learning algorithm used to train the NN is the well-known back-propagation (BP) algorithm. We have used the function provided in MatLab in all experiments. In our experiments, we just used one hidden layer, and fixed the number of hidden neurons to the number of inputs.

C. Pre-processing

To reduce the effect of lighting condition, the specification of camera and so on, it is also necessary to perform some pre-processing before face detection. In our experiments, we used a 2×2 median filter first to remove noises in the test image. Before sending a sub-image to the input of the NN detector, the following processing is conducted:

1) Illumination gradient correction: This operation computes a best-fit value of brightness and subtracts this value from all pixels in the sub-image. It is known that for face patterns, illumination correction can reduce heavy shadows caused by extreme lighting angles [9].
2) Histogram equalization: The face patterns in an image may not be clear due to changes in illumination brightness and differences in camera response curves. Histogram equalization is a kind of image transformation that can flatten the histogram of the image, and compensate the above mentioned changes and differences.

D. Post-Processing

Given a large test image, the NN often detects many faces (of the same person) at multiple nearby positions or scales. This can result in many false positive errors. To reduce such kind of errors, we adopt the merging overlapping detection method given in [6]. The basic idea of this heuristic method is to find the number of faces (as the results of detection) for different positions and scales around a given position(i, j). If the number of faces is larger than a given threshold T(in this paper, T=2 or 3), we consider that there is a face at this position. This simple method can reduce the number of false detections greatly.

E. Linear Discriminant Analysis (LDA)

Let us consider a C-class pattern recognition problem. Suppose that each class contains N_i training examples, x_{ij} is the jth example of the ith class, and d is the dimensionality of the training data. The within-class scatter matrix S_w and the between-class scatter matrix S_b are defined as follows:

$$S_w = \sum_{i=1}^{C}\sum_{j=1}^{N_i}(x_{ij} - \mu_i)(x_{ij} - \mu_i)^T \tag{1}$$

$$S_b = \sum_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{2}$$

where μ is the mean of all data and μ_i is the mean of all examples in the i-th class. In LDA, the discriminant criterion J is defined as follows:

$$J = \frac{|A^T S_b A|}{|A^T S_w A|} \tag{3}$$

The problem is to find the optimal matrix A to maximize the above criterion. The optimal matrix can be obtained by solving the eigenvalue problem of (3). Note that the rank of S_b is at most $C - 1$ because it is the sum of C matrices of rank one or less. Thus, the upper bound on the number of non-zero eigenvalues is $C - 1$ [15]. That is, using the transformation $y = A^T x$, a training example $x \in R^d$ is mapped to $y \in R^{C-1}$, with $d \gg C - 1$.

F. LDA with Clustering

Since face detection is a highly non-linear 2-class problem, LDA cannot get good results because it maps all data into a 1-Dimensional feature space. One simple way to solve this problem is to increase the number of classes first using some clustering algorithm, and then adopt LDA. In this study, we use the well known K-means algorithm for clustering [16].

In face detection, we can use K-means to cluster the face patterns and the non-face patterns separately. If K-means is applied to all patterns together, a cluster may contain both face and non-face patterns. This problem can be solved by using supervised learning such as learning vector quantization (LVQ)[10]. In our study, we adopt K-means because the batched approach often converges faster.

III. A MODIFIED LDA FOR FACE DETECTION

As stated in the previous section, LDA cannot be used for dimensionality reduction directly for face detection because the dimensionality of the mapped feature space is only one, and this is clearly not enough. To solve this problem, we can use K-means to divide the data into C clusters, and then map the original data into a $C-1$ dimensional space using LDA. One important thing in this approach is the determination of the number of clusters. Since the human faces are often similar with each other, a limited number of clusters might be enough to represent the faces patterns. This is especially true if we consider only frontal faces. For non-face patterns, however, anything that is not a face is non-face. Thus, it is almost impossible to limit the number of clusters of non-face patterns.

Based on the above consideration, in this paper, we divide only face patterns into C clusters with K-means, and consider all patterns outside these clusters as negative patterns. To apply LDA, each non-face pattern is considered as a cluster center. In fact, LDA is used to minimize the within-cluster variance of all face clusters; and maximize the between-cluster variance between all face clusters and all non-face clusters (note that each non-face pattern is a cluster). Specifically, suppose that the number of patterns belonging to the ith cluster of face patterns is N_i, and the number of non-face (negative) patterns is N_n, the within-class scatter matrix S_w and the between-class scatter matrix S_b are modified as follows:

$$S_w = \sum_{i=1}^{C}\sum_{j=1}^{N_i}(x_{ij}-\mu_i)(x_{ij}-\mu_i)^T \tag{4}$$

$$\begin{aligned} S_b =& \sum_{i=1}^{C} N_i(\mu_i-\mu_f)(\mu_i-\mu_f)^T \\ &+\sum_{k=1}^{N_n}(x_k^n-\mu_f)(x_k^n-\mu_f)^T \end{aligned} \tag{5}$$

where x_k^n is the kth non-face pattern, and μ_i and μ_f are, respectively, the mean of the ith cluster of face patterns and the mean of all face patterns, and they are defined by

$$\mu_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_{ij}, \quad \mu_f = \frac{1}{C}\sum_{i=1}^{C}\mu_i \tag{6}$$

We can define the discriminant criterion J in the same way as in (3). From (5) we can see that the rank of S_b is $r = C+N_n-1$ or less because it is the sum of $C+N_n$ matrices of rank one. On the other hand, r cannot be larger than the dimensionality d of the original feature space because S_b is a $d\times d$ matrix. Thus, r is bounded by $min\{d, C+N_n-1\}$. In fact, many of the eigenvalues of $S_w^{-1}S_b$ are very small, and can be ignored without reducing the discriminative information significantly. Thus, dimensionality reduction is also possible. Assume that the eigenvalues $\lambda_i(i=1,2,...,r)$ are sorted in descending order, in our experiments, we keep the first k eigenvalues such that

$$\sum_{i=1}^{k}\lambda_i / \sum_{i=1}^{r}\lambda_i \geq 96\% \tag{7}$$

For the database we used, more than 1/3 dimensionality reduction is possible. In turn, the structure of the NN detector can also be simplified.

IV. EXPERIMENTAL RESULTS

A. The Databases

To verify the efficiency and efficacy of the proposed approach, we conducted experiments using two databases, one for training and another for testing. The training database was constructed by us as follows. First, we collected 4,115 face patterns from BioID [11], training set of MIT CBCL [12], Yale [14], and World Wide Web. These patterns are front view faces with variations in facial expressions and lighting conditions. Each face pattern is manually cropped and normalized so that the face is aligned vertically and the size is 19×19. To get a good face detector, a large number of face and non-face patterns are needed. To avoid this problem, we generated nearly 6,129 face patterns using some simple image transformations from these databases. The transformations include random scaling with a factor between 90% and 110%, and rotation between -5 and 5 degrees.

We collected non-face patterns through bootstraping [9]. We started with 4,548 non-face examples taken from the MIT CBCL database. Each time a new detector was designed, it was tested with several images taken from [13]. The falsely detected images are then added as non-face patterns to the database. This process was repeated for several times. Altogether, we got 12,000 non-face patterns.

The database used for evaluating the performance of different approaches is the CMU image database [17]. There are three data sets in this database. In this thesis, we use only the set B which was provided by K. K. Sung and T. Poggio at the AI/CBCL Lab at MIT. This set contains 23 test images, and there are 155 faces in these images.

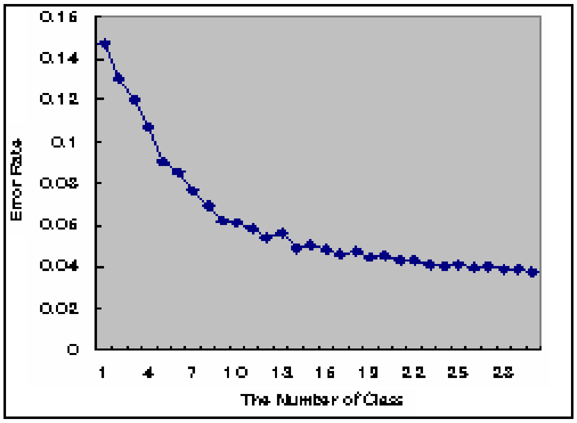

Fig. 2. Relation between the number of clusters and the error rate (for the LDA+NN method)

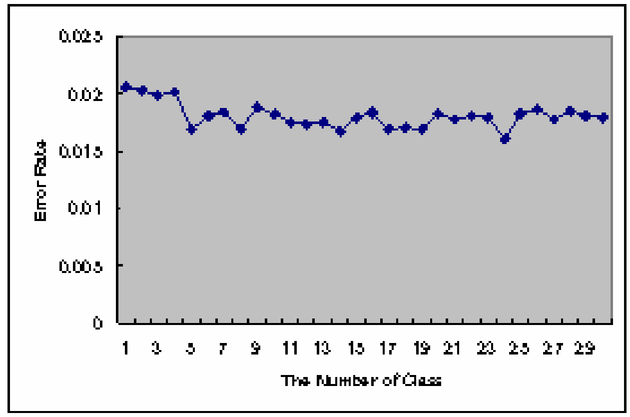

Fig. 3. Relation between the number of clusters and the error rate (for the proposed method)

Fig. 4. Center images of the face clusters

B. Methods Used for Comparison

In the experiments, the following four approached are used:

1) NN: This is the original neural network based approach. In this method, a 3 layer feed forward neural network is used directly for face detection. The number of hidden neurons equals to the number of inputs, which is $19 \times 19 = 361$ in this paper.
2) PCA+NN: This is the combination of PCA (for dimensionality reduction) and NN (for face detection). In this method, the eigenfaces are first obtained using PCA; all data are then projected to these eigenfaces; and the projected data are then recognized with NN. The dimensionality is selected such that the contribution rate of the eigenvalues is above 95%. Again, the number of hidden neurons equals to the number of inputs.
3) LDA+NN: This is a direct combination of clustering, LDA and NN. In this method, we first use K-means to divide the face and non face patterns into 30+30 clusters. The data are then mapped to a 59 dimensional feature space using LDA. The final decision is made by the NN. Once again, the number of hidden neurons equals to the number of inputs, which is 59 in this case.
4) The proposed method: In this method, we first use K-means to divide the face patterns into 24 face classes. The modified LDA is then used for dimensionality reduction. The final decision is made by the NN in the new feature space. The dimensionality of the feature space is determined such that the contribution rate of the eigenvalues is above 96%. The dimensionality is reduced from 361 to roughly 200, and this number is also the number of hidden neurons used in the NN.

C. Selection of parameters and performance comparison

As the first experiment, we conducted 10-fold cross validation on the training data. The purpose is of two-fold. First, we want to confirm which approach is the best for the available data. Second, we want to determine the number of clusters in using K-means. In face detection, the performance is often evaluated by the false positive error (accept as a face when the object is not a face) and the false negative errors (reject a face when the object is a face). The database used in cross validation contains 5,000 face patterns and 5,000 non-face patterns randomly chosen from the training set.

To use LDA+NN and the proposed method, the number of clusters is an important factor for achieving good performance. The results of LDA+NN are shown in Fig. 2, and the results of the proposed method are shown in Fig. 3. These two figures show the relation between the number of clusters and the error rate averaged over 10 runs. For simplicity, we just consider the sum of the false positive and the false negative errors without considering the cost of each type of error. In LDA+NN, the number of face clusters and

TABLE I

RESULTS OF 10-FOLD CROSS VALIDATION

	Training time [s]	Test time[s]	False negative error rate [%]	False positive error rate [%]	Number of inputs of the NN
NN	2295.2	0.1362	1.40	1.62	361
PCA+NN	222.6365	0.0687	2.07	1.38	148
LDA+NN	384.0187	0.0408	2.86	4.58	29
Proposed	368.4839	0.1053	1.73	1.48	184

TABLE II

EXPERIMENTAL RESULTS USING THE TEST IMAGES

	Detection rate [%]	False positive errors	Number of inputs of the NN
NN (T=2)	87.7	31	361
NN (T=3)	83.8	19	361
PCA+NN (T=2)	84.5	29	193
PCA+NN (T=3)	78.7	13	193
LDA+NN (T=2)	84.5	219	59
LDA+NN (T=3)	81.3	73	59
Proposed (T=2)	92.9	52	201
Proposed (T=3)	87.1	20	201

the number of non-face clusters are the same (i.e. $k_1 \equiv k_2$). In the proposed method, only the number of face clusters is considered.

From Fig. 2 we can see that for LDA+NN, the error rate can be reduced significantly by increasing the number of clusters. Based on these results, we choose $k_1 = k_2 = 30$ in this paper. The error rate in this case is 0.0372%. From Fig. 3, we can see that for the proposed method, the error rate does not change greatly for different number of face clusters, although the error rate reaches to the minimum value (0.0132%) when the face patterns are divided into 24 clusters. Fig. 4 shows the center images of all face clusters.

Table I shows the comparison of the performance of different approaches. From this table we can see that the differences in error rates between NN, PCA+NN and the proposed method are not significant, although the proposed method seems a little bit better than others if we consider both types of errors. In addition, the proposed method is much less time consuming than the original NN approach for training because the NN used is less complex. Note that the results obtained by the LDA+NN approach are the worst, although better results were reported by other authors [9].

D. Evaluation with Test Images

To see which approach is more practical for detecting faces, we evaluated all approaches using the set B of images in CMU image database . This test set contains 23 images with 155 faces. Table II shows the experimental results on the test set. The threshold T used in the post processing is 2 and 3 (see Section II.C) for all approaches. The detection rate is the rate of correctly detecting a face when there is face. From this table, we can see that the proposed method has a better detection rate than all other approaches, and the number of false positive errors is also comparable with the NN approach. The PCA+NN approach can provide the least false positive errors, but the detection rate is relatively low. Note that in most security systems, false negative errors may cost more than false positive errors. In this sense, the proposed method is much better than other approaches. Figs. 5-8 are the detection results of all approaches. From these figures, we can see that the proposed method can detect successfully most frontal faces, and the number of false positive errors is small.

V. CONCLUSION

In this research, we have developed a system for detecting human faces with vertical frontal views from still images. In the proposed method, the problem of face detection is first changed to a multi-class problem using clustering; a modified LDA is then used to extract useful features; and the final decision is made by a feedforward neural network. The modified LDA is used to minimize the within-cluster variance between all face clusters; and maximize the between-cluster variance between all face clusters and all non-face data. The feature space so obtained has a dimensionality less than that of the original problem. The efficiency and efficacy of the proposed method has been verified through experiments.

In the next step, we would like to verify the proposed method for detecting faces with different poses and under different lighting conditions. We would also like to study face detection from moving pictures. Further, we would like to compare the appearance based approach and feature based methods (say, the method given in [18]), and try to understand which approach is more efficient and more effective under different conditions.

REFERENCES

[1] M. H. Yang, D. J. Kriegman and N. Ahuja, “Detecting faces in images: A survey,” IEEE Trans. Pattern Analysis and Machine Intelligence, 24 (1): 34-58, 2002.

[2] C. Shavers, R. Li and G. Lebby, “An SVM-based approach to face detection,” Proc. 38th Southeastern Symposium on System Theory, Tennessee Technological University, pp. 362-366, Mar. 2006.

[3] H. Jee, K. Lee and S. Pan, ”Eye and face detection using SVM,” Proc. Intelligent Sensors, Sensor Networks and Information Processing Conference, pp. 577- 580, 2004.

Fig. 5. Detection results of the NN approach

Fig. 6. Detection results of the PCA+NN approach

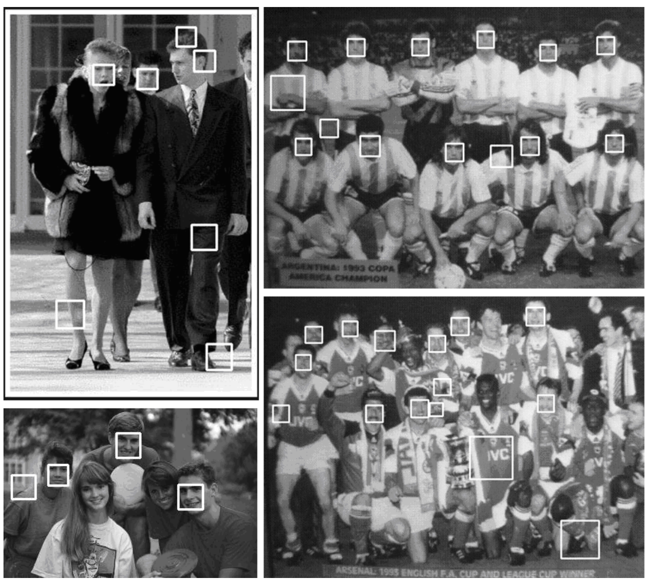

Fig. 7. Detection results of the LDA+NN approach

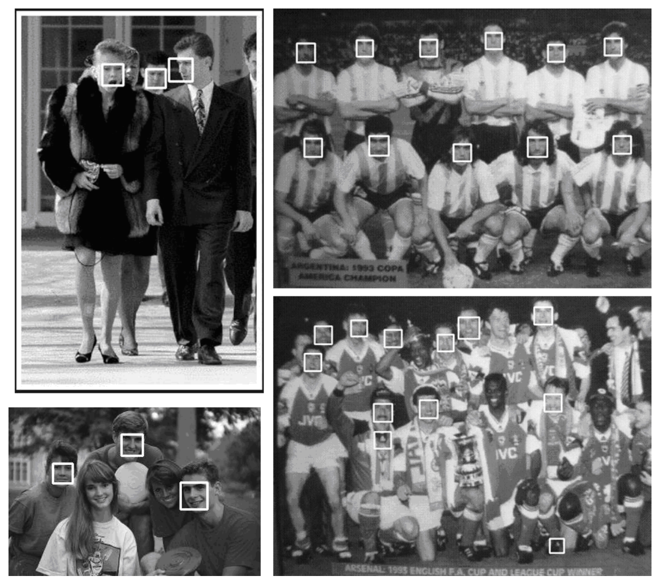

Fig. 8. Detection results of the proposed method

[4] E. Osuna, R. Freund and F. Girosi, “Training support vector machines: an application to face detection,” Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 130-136, 1997.

[5] Y. Feng and P. Shi, “Face detection based on kernel Fisher discriminant analysis,” Proc. 6th IEEE International Conference on Automatic Face and Gesture Recognition, pp.381 - 384, 2004.

[6] H. Rowley, S. Baluja, and T. Kanade. “Neural network-based face detection,” IEEE Trans. Pattern Anal. Mach. Intell., 20(1): 23-38, 1998

[7] M. Turk and A. Pentland, “Eigenfaces for Recognition,” Journal of Cognitive Neuroscience, pp. 71-86, 1991.

[8] M. H. Yang, N. Ahuja, and D. Kriegman, “Mixtures of Linear Subspaces for Face Detection,” Proc. Fourth Int. Conf. Automatic Face and Geture Recognition, pp. 70-76, 2000.

[9] K.-K. Sung and T. Poggio, “Example-based learning for view-based human face detection,” IEEE Trans. Pattern Anal. Mach. Intell., 20(1):39-51, 1998.

[10] T. Kohonen, “The Self-Organizing Map,” IEEE Proceedings, Vol. 78, No. 9, pp. 1464-1480, 1990.

[11] O. Jesorsky, K.J.Kirchberg, and R.W. Frischholz1, “Robust face detection using the Hausdorff distance,” Proc. of Third International Conference on Audio and Videobased Biometric Person Authentication, LNCS-2091, Halmstad, Sweden, pages 90-95. Springer, 2001.

[12] MIT CBCL database #1, http://cbcl.mit.edu/softwaredatasets/FaceData2.html.

[13] Computational vision archive, http://www.vision.caltech.edu/html−files/archive.html.

[14] 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.

[15] K.-K. Sung, “Learning and Example Selection for Object and Pattern Detection,” PhD thesis, MIT AI Lab, Jan. 1995 (Available as AI Technical Report 1572).

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

[17] CMU Image Database, http://vasc.ri.cmu.edu/idb/html/face/frontal_images/index.html.

[18] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” Proc. Computer Vision and Pattern Recognition, Vol. 1, pp. 511-518, 2001.