A Modification of Kernel-based Fisher Discriminant Analysis for Face Detection

Takio Kurita
Neuroscience Research Institute, AIST
1-1-1 Umezono, Tsukuba, 305-8568, Japan
takio-kurita@aist.go.jp

Toshiharu Taguchi
Saitama University
255, Shimo-Okubo, Saitama, 338-8570, Japan
taguchi@me.ics.saitama-u.ac.jp

Abstract

This paper presents a modification of kernel-based Fisher Discriminant Analysis (FDA) for face detection. In face detection problem, it is important to design a two-category classifier which can decide whether the given input sub-image is a face or not. There is a difficulty to train such tow-category classifiers because the "non face" class includes many images of different kinds of objects and it is difficult to treat them as a single class. Also the dimension of the discriminant space constructed by the usual FDA is limited to 1 *for tow-category classification. To overcome these problems of the usual FDA, the discriminant criterion of the usual FDA is modifed such that the covariance of the "face" class is minimized while the differences between the center of the "face" class and each training sample of the "non face" class are maximized. By this modification we can obtain a higher dimensional discriminant space which is suitable for "face" and "not face" classification. It is shown that the proposed method could outperform the support vector machine (SVM) by experiments of "face" and "non face" classification using the face images gathered from the available face database and the many face images on the Web.*

1. Introduction

Object detection is an important and fundamental problem in computer vision, and there are many applications including face detection, person detection, and so on. In view-based object detection, machine learning techniques have been used to design a classifier from the given training examples. For face detection, the positive and negative examples are trained by using a probabilistic or statistic models, such as the neural network[1, 2, 3, 4], principal components analysis (PCA)[5, 6], linear discriminant analysis[7, 8] and so on.

In the last years, a number of powerful kernel-based learning machines, e.g., support vector machines (SVMs)[9, 10], kernel Fisher discriminant analysis (KFDA)[11, 12, 13], and kernel principal component analysis (KPCA)[14, 15] have been proposed. Successful applications of kernel-based algorithms have been reported for various fields including object recognition, text categorization, time-series prediction, etc.[16]. Especially support vector machines (SVM))have been successfully applied to face detection problem [17, 18, 19] .

In this paper, we present a modification of kernel-based Fisher discriminant analysis (KFDA) to detect frontal views of faces in gray-scale images. In face detection, it is important to design a tow-category classifier which can decide whether the given input sub-image is a face or not. The classifier is usually trained by using training examples of "face" images and "non face" images. If we straightforwardly apply Fisher discriminant analysis to this problem, the discriminant space is constructed such that two classes; "face" and "non face" classes, are equally treated. The "face" class includes only "face" images and they are similar with each other. It can be treated as a single class. However, the "non face" class includes many images of different kinds of objects. It can not be treated as a single class. It is difficult to model them as samples from a single distribution. This means that the constructed discriminant space may be unsuitable for "face" and "non face" classification even if we use highly nonlinear discriminant analysis such as kernel-based FDA. Also the dimension of the feature vectors constructed by the usual FDA is limitted to 1 for tow-category classification. This means that we can not obtain high dimensional discriminant space. To overcome these drawbacks of the usual FDA, the discriminant criterion of the utusal FDA is modified such that the covariance of the "face" class is minimized while the differences between the center of the "face" class and each training sample of the "non face" class are maximized. By this modification of the usual FDA we can obtain a higher dimensional discriminant space which is suitable for "face" and "non face" classification.

In the kernel-based methods, proper selection of the kernel size and the regularization parameter is very important to achieve better genelarization ability. In this paper a data set for cross validation is prepared and used to determine the proper parameters by directly evaluating the error rates of the data. We think that this is the simplest and the most practical approach to determine the proper parameters of the kernel-based leaining method if we can gather additional samples for cross validation.

To evaluate the proposed method, several experiments on "face" and "non face" classification were performed by using the face images gathered from the face databases (MIT face databse[20] and AIST face database[21]) and many face images on the Web. The results shows that the proposed method can outperform the support vector machine (SVM). It is also shown that the selection of the kernel size and the regularization parameter is important and the proper parameters can be easily determined by simply preparing the data set for cross validation.

In the next section, kernel-based FDA for "face" and "non face" classification is explained. The experimental results are shown in section 3.

2. Kernel-based FDA for Face and Non Face Classification

2.1. Fisher Discriminant Analysis

Let $\boldsymbol{x} = (x_1, \dots, x_M)^T$ be a M dimensional feature vector. Suppose that we have K classes $\{C_k\}_{k=1}^{K}$. Then linear Fisher discriminant analysis constructs a dimension reduction linear mapping from the input vector $\boldsymbol{x}$ to a new feature $\boldsymbol{y} = (y_1, \dots, y_L)^T$ as

$$\boldsymbol{y} = A^T \boldsymbol{x}, \tag{1}$$

where $A = [a_{ij}]$ is the coefficients matrix. The discriminant criterion

$$J = \boldsymbol{tr}(\hat{\Sigma}_W^{-1} \hat{\Sigma}_B) \tag{2}$$

is used to evaluate the performance of the discrimination of the new features $\boldsymbol{y}$ and is maximized, where $\hat{\Sigma}_W$ and $\hat{\Sigma}_B$ are the within-class covariance matrix and the between-class covariance matrix defined on the new features $\boldsymbol{y}$.

The optimal coefficient matrix A is then given by solving the following eigen-equation

$$\Sigma_B A = \Sigma_W A \Lambda \quad (A^T \Sigma_W A = I), \tag{3}$$

where Λ is a diagonal matrix of eigenvalues and I denotes the unit matrix. The matrices Σ_W and Σ_B are the within-class covariance matrix and the between-class covariance matrix of the input feature vectors $\boldsymbol{x}$ and they are computed as

$$\Sigma_W = \sum_{k=1}^{K} \omega_k \Sigma_k,$$

$$\Sigma_B = \sum_{k=1}^{K} \omega_k (\bar{\boldsymbol{x}}_k - \bar{\boldsymbol{x}}_T)(\bar{\boldsymbol{x}}_k - \bar{\boldsymbol{x}}_T)^T, \tag{4}$$

where ω_k, $\bar{\boldsymbol{x}}_k$, $\bar{\boldsymbol{x}}_T$, and Σ_k denote *a priori* probability of class C_k, the mean vector of class C_k, and the total mean vector, and the covariance matrix of the class C_k, respectively.

The j-th column of A is the eigenvector corresponding to the j-th largest eigenvalue. Thus, importance of each elements of the new features $\boldsymbol{y}$ are evaluated by the corresponding eigenvalues.

2.2. Kernel-Based Fisher Discriminant Analysis

Kernel-based Fisher discriminant Analysis (KFDA) is the method which uses the model defined by a linear combination of some specified kernel bases as

$$\boldsymbol{y} = \sum_{i=1}^{N} \boldsymbol{a}_i K(\boldsymbol{w}_i, \boldsymbol{x}), \tag{5}$$

where K is the kernel function, and $\boldsymbol{a}_i$ is the coefficient vector for the i-th kernel base. As the kernel function, an isotropic Gaussian function $K(\boldsymbol{w}_i, \boldsymbol{x}) = \exp[-\frac{||\boldsymbol{w}_i - \boldsymbol{x}||^2}{2\sigma^2}]$ is typically used. The location of the kernel base $\boldsymbol{w}_i$ is fixed to one of the training samples and the number of kernel bases N equals to the number of training samples.

Let $\boldsymbol{k}(\boldsymbol{x}) = (K(\boldsymbol{w}_1, \boldsymbol{x}), \dots, K(\boldsymbol{w}_N, \boldsymbol{x}))^T$ be a kernel bases vector for the feature vector $\boldsymbol{x}$. Then the equation (5) can be written as

$$\boldsymbol{y} = A^T \boldsymbol{k}(\boldsymbol{x}) \tag{6}$$

where $A^T = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_N]$ is the coefficients matrix.

The optimal coefficient matrix A is then given by solving the eigen-equation

$$\Sigma_B^{(K)} A = \Sigma_W^{(K)} A \Lambda \quad (A^T \Sigma_W^{(K)} A = I), \tag{7}$$

where Λ is a diagonal matrix of eigenvalues and I denotes the unit matrix. The matrices $\Sigma_W^{(K)}$ and $\Sigma_B^{(K)}$ are the within-class covariance matrix and the between-class covariance matrix of the kernel bases vectors $\boldsymbol{k}(\boldsymbol{x})$. The dimension of the new feature vector $\boldsymbol{y}$ is limited to $\min(K-1, N)$.

However, this setting is ill-posed, because we are estimating the $N \times L$ coefficients of the matrix A from N samples. Thus we have to introduce some regularization technique. One of the simplest methods is to simply add a multiple of the identity matrix to $\Sigma_W^{(K)}$ as

$$\tilde{\Sigma}_W^{(K)} = \Sigma_W^{(K)} + \beta I \tag{8}$$

This makes the problem numerically more stable because the within-class covariance matrix $\tilde{\Sigma}_W^{(K)}$ becomes positive definite as for large β. This is roughly equivalent to add independent noises to each of the kernel bases.

2.3. Modification of FDA for Face and Non Face Classification

In face detection problem, it is important to design a two-category classifier which can decide whether the given input sub-image is a face or not. The classifier is usually trained by using training examples of "face" images and "non face" images. If we straightforwardly apply the kernel-based Fisher discriminant analysis to this problem, the discriminant space is constructed such that two classes; "face" and "not face" classes, are equally treated. The "face" class includes only "face" images and they are similar with each other. It can be treated as a single class. However, the "non face" class includes many images of different kinds of objects. It can not be treated as a single class. It is difficult to model them as samples from a single distribution. This means that the constructed discriminant space may be unsuitable for "face" and "non face" classification even if we use highly nonlinear discriminant analysis such as kernel-based FDA. Also the dimension of the new feature vector $\boldsymbol{y}$ constructed by the usual kernel-based FDA is limited to 1 for the case of tow-category classification. This means that we can not obtain high dimensional discriminant space by the usual kernel-based FDA.

To overcome these drawbacks of the kernel-based FDA, we modify the discriminant criterion such that the covariance of "face" class is minimized while the differences between the center of "face" class and each examples of "not face" class are maximized. In other words, we consider "face" class as one distribution in which many images of the same kind (face) are included and "not face" class as a set of individual classes each of which corresponds to each sample of the "not face" class.

Let "face" class and "not face" class are expressed by using the kernel bases as

$$\begin{aligned} C &= \{\boldsymbol{k}(\boldsymbol{x}_i) \mid i = 1, \cdots, n_f\} \\ \bar{C} &= \{\boldsymbol{k}(\boldsymbol{x}^k) \mid k = 1, \cdots, n_{\bar{f}}\}, \end{aligned} \tag{9}$$

where n_f is the number of "face" images and $n_{\bar{f}}$ is the number of "not face" images. Then the mean vector and the covariance matrix of "face" class are given by

$$\begin{aligned} \bar{\boldsymbol{k}}_f &= \frac{1}{n_f}\sum_{i=1}^{n_f} \boldsymbol{k}(\boldsymbol{x}_i), \\ \Sigma_f &= \frac{1}{n_f}\sum_{i=1}^{n_f} (\boldsymbol{k}(\boldsymbol{x}_i) - \bar{\boldsymbol{k}}_f)(\boldsymbol{k}(\boldsymbol{x}_i) - \bar{\boldsymbol{k}}_f)^T \end{aligned} \tag{10}$$

The total mean vector is given by

$$\bar{\boldsymbol{k}}_T = \omega_f \bar{\boldsymbol{k}} + \frac{n_{\bar{f}}}{N}\sum_{k=1}^{n_{\bar{f}}} \boldsymbol{k}(\boldsymbol{x}^k) \tag{11}$$

where $\omega_f = \frac{n_f}{N}$ and N is the total number of images ($N = n_f + n_{\bar{f}}$). Then the within-class covariance matrix and the between-class covariance matrix are given by

$$\begin{aligned} \Sigma_W^{(f)} &= \omega_f \Sigma_f \\ \Sigma_B^{(f)} &= \omega_f(\bar{\boldsymbol{k}}_f - \bar{\boldsymbol{k}}_T)(\bar{\boldsymbol{k}}_f - \bar{\boldsymbol{k}}_T)^T \\ &+ \frac{1}{N}\Sigma_{k=1}^{n_{\bar{f}}}(\boldsymbol{k}(\boldsymbol{x}^k) - \bar{\boldsymbol{k}}_T)(\boldsymbol{k}(\boldsymbol{x}^k) - \bar{\boldsymbol{k}}_T)^T \end{aligned} \tag{12}$$

New features $\boldsymbol{y} = (\boldsymbol{y_1}, \cdots, \boldsymbol{y_L})^T$ are obtained by a linear combination $\boldsymbol{y} = A^T\boldsymbol{k}(\boldsymbol{x})$ of the kernel bases vector $\boldsymbol{k}(\boldsymbol{x})$.

To construct the discriminant space in which the covariance of "face" class is minimized and the differences between the center of "face" class and each samples of "not face" class are maximized, we can define the discriminant criterion

$$J = \boldsymbol{tr}(\hat{\Sigma}_W^{(f)-1}\hat{\Sigma}_B^{(f)}) \tag{13}$$

using the within-class covariance matrix $\hat{\Sigma}_W^{(f)}$ and the between-class covariance matrix $\hat{\Sigma}_B^{(f)}$ in the discriminant space. The optimal coefficient matrix A which maximizes the discriminant criterion J is obtained by solving the eigen value problem

$$\Sigma_B^{(f)} A = \Sigma_W^{(f)} A\Lambda \quad (A^T\Sigma_W^{(f)}A = I) \tag{14}$$

The dimension of the new feature vector $\boldsymbol{y}$ is limited to $\min(n_{\bar{f}}+1, N) = n_{\bar{f}}+1$. If we have enough training samples of non face images, we can obtain high dimensional discriminant feature space. In this case, regularization can be achieved by using the

$$\tilde{\Sigma}_W^{(f)} = \Sigma_W^{(f)} + \beta I \tag{15}$$

instead of the within-class covariance matrix $\Sigma_W^{(f)}$.

To classify the input sub-image into "face" class or "not face" class, we need some classifier which utilizes the constructed discriminant feature space. One of the simplest classifier is to check the distance between the mean vector of "face"class and the feature vector of the input sample in the discriminant space and decide the input as "face" if the distance is within a threshold Th. In the following experiments, this simple classifier is used. More sophisticated classifiers could be further improve the recognition performance.

2.4. Relation to Kernel-Based PCA

One of the most popluar methods in face recognition or face detection is "Eigenface" which is based on PCA of the face images[5, 6]. Here we point out the close relationship between the proposed kernel-based FDA and the kernel PCA.

It is well-known that the mapping from the feature space to the discriminant space constructed by Fisher discriminant analysis can be split into two stages of the mappings as

$$\boldsymbol{y} = A^T \boldsymbol{k}(\boldsymbol{x}) = A_2^T A_1^T \boldsymbol{k}(\boldsymbol{x}). \tag{16}$$

The mapping A_1 at the first stage corresponds to whitening process of the within-class covariance matrix in the mapped space. This mapping A_1 can be obtained by computing the eigen vectors of the within-class covariance matrix as

$$A_1 = \Sigma_W^{-\frac{1}{2}} = M^{-\frac{1}{2}} U^T \tag{17}$$

where the matrices U and M are the matrices of the eigen vectors and the eigen values obtained by the eigen equation

$$\Sigma_W U = UM(U^T U = I). \tag{18}$$

After the mapping of the first stage $\boldsymbol{z} = A_1^T \boldsymbol{k}(\boldsymbol{x})$, the within-class covariance matrix $\tilde{\Sigma}_W$ and the between-class covariance matrix $\tilde{\Sigma}_B$ in the mapped space are obtained by

$$\begin{aligned} \tilde{\Sigma}_W &= A_1^T \Sigma_W A_1 = \Sigma_W^{-\frac{1}{2}} \Sigma_W \Sigma_W^{-\frac{1}{2}} = I \\ \tilde{\Sigma}_B &= A_1^T \Sigma_B A_1 = \Sigma_W^{-\frac{1}{2}} \Sigma_B \Sigma_W^{-\frac{1}{2}}. \end{aligned} \tag{19}$$

The mapping at the second stage A_2 can be obtained by solving the eigen equation

$$\tilde{\Sigma}_B A_2 = A_2 \Lambda. \tag{20}$$

Namely, the principal components of the average between-class covariance matrix of the mapped space $\tilde{\Sigma}_B$ are computed in this second stage.

In the case of the proposed method the average within-class covariance matrix at the first stage can be computed form the the covariance matrix of the face images as

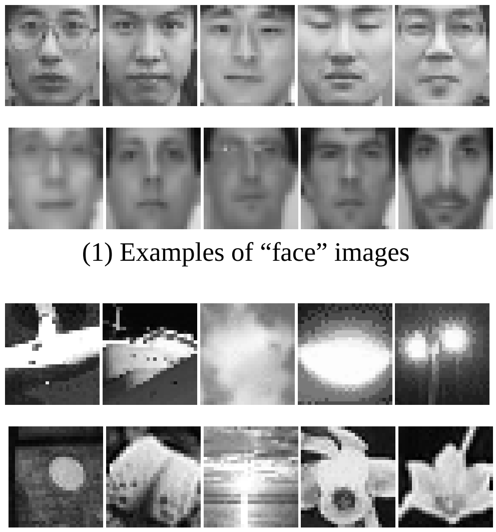
$$\Sigma_W^{(f)} = \omega_f \Sigma_f. \tag{21}$$

This means that the first stage of the mapping is closely related with the kernel-based PCA of the face images and the mapping is obtained by the kernel-based PCA. By neglecting the second stage of the proposed kernel FDA, we can obtain a method in which the whitening of the covariance matrix of the face images is used as an approximation of the proposed kernel-based FDA.

3. Experiments

3.1. Database of “Face” and “Non Face” Images

To evaluate the proposed method, we gathered “face” images and “non face” images from MIT face database[20], face image database used in [21], and many images on the Web. From each of the face images, face regions were searched by the simple nearest neighbor classifier and the size of the face regions were normalized to 30×28 pixels. By eliminating the incorrect samples from the detected images, we have a data set of 725 normalized “face” images. Also a data set of 2200 normalized “non face” images were obtained by randomly cutting the partial regions of the gathered images and normalizing to the same size. These “face” and “non face” images were divided into three sets (the training data set, the cross validation data set, and the evaluation data set). The training data set includes 100 “face” images and 200 “non face” images and is used for training the classifiers. The cross validation data set include 300 “face” images and 1000 “non face” images is used to determine the proper parameters such as the kernel size σ, the regularization parameter β, and the threshold Th of the classifier . The evaluation data set includes 325 “face” images and 1000 “non face” images and is used to evaluate the recognition rates. Examples of “face” and “non face” images used in the experiments are shown in Figure 1.



(1) Examples of “face” images

(2) Examples of “not face” images

Figure 1. Examples of ”face” and “non face” images.

3.2. Selection of the kernel size and the regularization parameter

It is known that proper selection of the kernel size is important to achieve better performance in kernel-based learning methods. Here we used the simplest and straightforward approach to determine the proper kernel size using the cross validation data set.

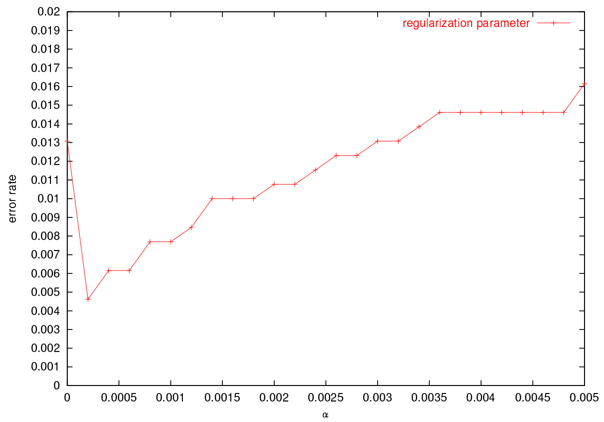


Figure 2. Relation between the regularization parameter β and the error rate.

The regularization parameter β is an important factor for generalization ability of the learned classifier. The error rates to the cross validation data set were evaluated at the different regularization parameters. The results are shown in Figure 2. It is noticed that the error rate is minimum at $\beta = 0.0002$. So the regularization parameter was set to $\beta = 0.0002$ in the following experiments.

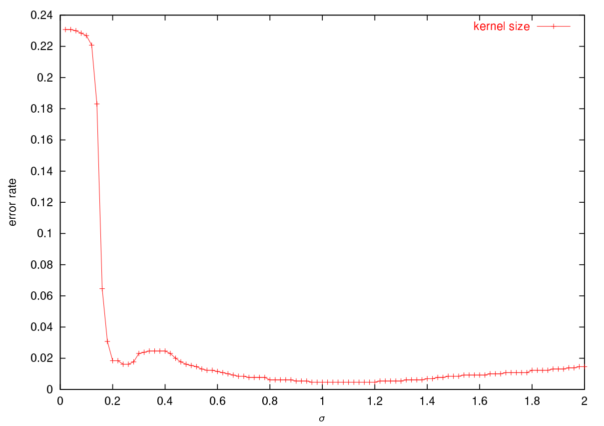


Figure 3. Relation between the kernel size σ and the error rate.

Another important parameter of the proposed method is the kernel size σ. Figure 3 shows the relation between the kernel size σ and the error rate to the cross validation data set. In this experiment, the regularization parameter β was set to $\beta = 0.0002$. It is noticed that the error rates are gradually decrease from $\sigma = 0.5$ to 1.0 and are gradually increased when the kernel size σ becomes larger than $\sigma = 1.2$. There is minimum around $\sigma = 1.0$. From this results we can understand that the proper kernel size is from 0.96 to 1.0. In the following experiments, the kernel size was set to $\sigma = 1.08$.

3.3. Experimental Comparison with SVM and Kernel-based PCA

Table 1. Comparison of the error rates.

	All samples	face	non face
Kernel FDA ($\beta = 0.0002$, $\sigma = 1.08$)	0.0083	0.0185	0.0050
Kernel FDA ($\beta = 0.0$, $\sigma = 1.08$)	0.0128	0.0400	0.0040
Kernel PCA ($\beta = 0.0$, $\sigma = 1.08$)	0.0211	0.0462	0.0130
SVM ($\sigma = 1.08$)	0.0174	0.0062	0.0210

The recognition performance of the proposed method was evaluated using the evaluation data set. Table 1 summarizes the error rates to all samples, the error rates to the "face" samples, and the error rates to the "non face" samples. The threshold Th of the proposed method was determined by using the cross validation data set. The error rates to the kernel-based FDA without regularization, namely ($\beta = 0.0$), is also shown in the second raw of the table. It is noticed that the recognition rate is improved by introducing the regularization. This shows the importance of the regularization in the kernel-based methods. For comparison the results of the whitening method explained in the section 2.4 and support vector machine (SVM) with Gaussian kernels are also shown in the table. The kernel size of the whitening method and the SVM was set to the same value with the proposed kernel-based FDA, namely $\sigma = 1.08$. From these results we can conclude that the recognition performance of the proposed kernel-based FDA is better or as good as the SVM. Also we can say that the whitening by the kernel-based PCA is comparable with the SVM but is not as good as the proposed kernel-based FDA.

Finally an example of the face detection is shown in figure 4. Whole image was scanned with small windows of different sizes and all the candidates are fed into the proposed classifier. The squares shown in the image denote the windows decided as "face" by the classifier.

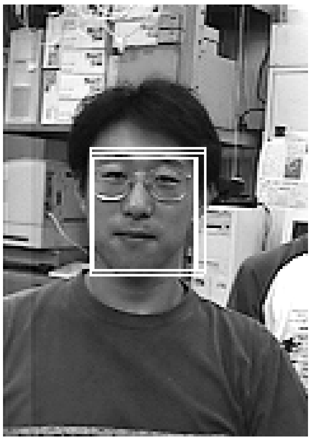


Figure 4. An Example of face detection

Acknowledgment

The authors would like to thank Dr. Kazuhiro Hotta for data acquisition and experimental supports. This work is partly supported by Special Coordination Funds for Promoting Science and Technology from the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government.

References

[1] Shang-Hung Lin and Sun-Yuan Kung and Long-Ji Lin, "Face Recognition/Detection by Probabilistic Decision-Based Neural Network," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.8, No.1, pp.114-131, 1997.

[2] Henry A. Rowley and Shumeet Baluja and Takeo Kanade, "Neural Network-Based Face Detection," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.20, No.1, pp.23-38, 1998.

[3] Kah-Kay Sung and Tomaso Poggio, "Example-Based Learning for View-Based Human Face Detection," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.20, No.1, pp.39-51, 1998.

[4] Paphael Fèraud and Olivier J. Bernier, Jean-Emmanuel Viallet and Michel Collobert, "A Fast and Accurate Face Detector Based on Neural Networks," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.23, No.1, pp.42-53, 2001.

[5] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, Vol.3, pp.71-86, 1991.

[6] B. Moghaddam and A. Pentland, "A Feature-Based Approach to Face Recognition," Proc. Fifth Int'l Conf. on Computer Vision, 1992.

[7] Kazuhiro Hotta and Takio Kurita and Taketoshi Mishima, "Scale Invariant Face Detection Method using Higher-Order Local Autocorrelation Features extracted from Log-Polar Image," Proc. of Inter. Conf. on Automatic Face and Gesture Recognition (FG'98), pp.70-75, 1998.

[8] Kazuhiro Hotta and Taketoshi Mishima and Takio Kurita, "Scale Invariant Face Detection and Classification Method Using Shift Invariant Features Extracted from Log-Polar Image," IEICE Trans. on Information and Systems, Vol.E84-D, pp.867-878, 2001.

[9] B. Schőlkopf, *Support Vector Learning*, Oldenbourg-Verlag, Munich, Germany, 1997.

[10] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

[11] S. Mika and G. Rätsch and J. Weston and B. Schölkopf and K. Müller, "Fisher Discriminant Analysis with Kernels," Proceedings of IEEE Neural Networks for Signal Processing Workshop, 1999.

[12] S. Mika and G. Rätsch and J. Weston and B. Schölkopf and A. Smola and K. Müller, "Invariant feature extraction and classification in kernel spaces," in S.A. Solla and T.K. Leen and K.-R. Müller ed. *Advances in Neural Information Processing Systems 12*, MIT Press, 2000.

[13] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," Neural Computation, Vol.12, No.10, pp.2385-2404, 2000.

[14] B. Schölkopf and A. J. Smola and K.-R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," Neural Computation, Vol.10, pp.1000-1017, 1998.

[15] S. Mika and B. Schölkopf and A. J. Smola and K.-R. Müller and M. Scholz and G. Rätsch, "Kernel PCA and De–Noising in Feature Spaces," in M. S. Kearns and S. A. Solla and D. A. Cohn ed. *Advances in Neural Information Processing Systems 11*, MIT Press, 1999.

[16] Klaus-Robert Müller and Sebastian Mika and Gunnar Rätsch and Koji Tsuda and Bernhard Schölkopf, "An Introduction to Kernel-Based Learning Algorithms," IEEE Trans. on Neural Networks, Vol.12, No.2, pp.181-201, 2001.

[17] Edgar Osuna and Robert Freund and Federico Girosi, "Training Support Vector Machines: an Application to Face Detection," Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.130-136, 1997.

[18] Constantine P. Papageorgiou and Michael Oren and Tomas Poggio, "A General Framework for Object Detection," Proc. Fifth Int'l Conf. on Computer Vision, 1998.

[19] Anuj Mohan and Constantine P. Papageorgiou and Tomas Poggio, "Example-Based Object Detection in Images by Components," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vo.23, No.4, pp.349-361, 2001.

[20] MIT face database, *ftp://whitechapel.media.mit.edu/pub/images/*

[21] Takio Kurita and Kazuhiro Hotta and Taketoshi Mishima, "Feature Ordering by Cross Validation for Face Detection," Proc. of IAPR Workshop on Machine Vision Applications, pp.211-214, 2000.