

Pairwise Face Recognition

Guo-Dong Guo, Hong-Jiang Zhang, and Stan Z. Li
Microsoft Research China, 5F, Beijing Sigma Center, P. R. C.

Abstract

We develop a pairwise classification framework for face recognition, in which a C class face recognition problem is divided into a set of $C(C - 1)/2$ two class problems. Such a problem decomposition not only leads to a set of simpler classification problems to be solved, thereby increasing overall classification accuracy, but also provides a framework for independent feature selection for each pair of classes. A simple feature ranking strategy is used to select a small subset of the features for each pair of classes. Furthermore, we evaluate two classification methods under the pairwise comparison framework: the Bayes classifier and the AdaBoost. Experiments on a large face database with 1079 face images of 137 individuals indicate that 20 features are enough to achieve a relatively high recognition accuracy, which demonstrates the effectiveness of the pairwise recognition framework.

1. Introduction

Face recognition technology can be used in a wide range of applications such as identity authentication, access control, and surveillance. Interests and research activities in face recognition have increased significantly over the past decade [18] [25] [17]. Two issues are central for face recognition, *i.e.*, what features to use to represent a face, and how to classify a new face based on the chosen representation.

For a given face representation, we are interested in how to do classification. When the face database becomes large, some traditional classification methods may deteriorate rapidly or may not be working any more. One solution is to change the original complex problem into a set of smaller and simpler ones to solve. For this consideration, we develop a pairwise classification framework to solve the multi-class face recognition problem. The motivation of pairwise comparisons also comes from the character discrimination experiments [11], which demonstrate that features useful to distinguish letter 'E' from 'F' may differ from those distinguishing 'E' from 'R'. The pairwise architecture is to decompose the complex face recognition problem on a large database, into a simple discrimination be-

tween just two persons in each step, and explores the actual difference between two persons independently, *i.e.*, identifies what features are most discriminant for a specific pair of individuals. In our approach, instead of choosing one feature space for the whole problem, as is conventionally done, a pair specific feature selection strategy is developed. So the features suitable for discrimination between one pair of classes may not be suitable for distinguishing some other pairs of classes.

Under this pairwise recognition framework, we examine two kinds of classifiers, the probabilistic approach and the large margin classifier. Since both kinds of classifiers have reported high accuracy for general pattern recognition, we try to find if the pairwise framework can further improve the face recognition accuracy. In Section 2, we simply review current subspace analysis methods for face representations. Section 3 describes the Bayes classifier and the AdaBoost algorithm. We present the pairwise classification framework in Section 4, and the experimental results are given in Section 5. We also discuss some related issues in Section 6 and finally give the conclusions.

2. Face Representations

Principal Component Analysis (PCA), also known as the Karhunen-Loeve expansion, is a classical technique for signal representation [9]. Sirovich and Kirby [21] applied PCA for representing face images. Turk and Pentland [23] developed a well known face recognition method, known as **eigenfaces**. While PCA pursues a low dimensional representation of the faces, it is not necessarily with good discrimination capability between different faces. Belhumeur *et al* [2] developed an approach called **Fisherfaces** by applying first PCA for dimensionality reduction and then using FLD (Fisher Linear Discrimination) for discriminant analysis. However, the robustness of the FLD procedure depends on whether or not the within class scatter can capture enough variations for a specific class. When the training sample size for each class is small, the FLD procedure leads to overfitting, and hence with poor generalization to new data [12] [13].

Another recently proposed method for face feature extraction is Independent Component Analysis (ICA) [1],

which separates the high-order moments of the input in addition to the second-order moments [3]. However, it is not clear that how much is the non-Gaussianity of the face images and how useful it is for face recognition. Moghaddam [15] compared the PCA and ICA methods for face recognition on "FERET" face database, and found that both of them gave the same recognition accuracy.

Apart from above linear transformations for low dimensional face representation, there are also some non-linear approaches to extract face features, such as the nonlinear PCA [10] or kernel PCA [20], and nonlinear FLD [14]. A thorough comparison of these methods for face feature extraction on a standard database with fair conditions is necessary. Here, we focus on the classification problem. The goal is to improve the face recognition accuracy with less features for a given feature set. We choose to use the PCA method for face feature extraction in our experiments.

3. Classifiers

For pattern recognition, the Bayes classifier yields the minimum error rates when the underlying probability density function (pdf)'s are known [6]. On the other hand, large margin classifiers are proposed recently in machine learning society, such as Support Vector Machine (SVM) [26] and AdaBoost [4]. Because the AdaBoost algorithm has not been used before for face recognition, we try to evaluate it under the pairwise classification framework.

3.1. Bayes Classifier

The *a posteriori* probability of pattern \mathbf{x} belong to class ω_c is given by the Bayes rule:

$$P(\omega_c|\mathbf{x}) = \frac{P(\omega_c)p(\mathbf{x}|\omega_c)}{p(\mathbf{x})} \quad (1)$$

where $P(\omega_c)$ is the *a priori* probability, $p(\mathbf{x}|\omega_c)$ the conditional probability density function of ω_c , and $p(\mathbf{x})$ is the mixture density. The maximum *a posteriori* (MAP) decision is

$$\omega_c^* = \arg \max_c P(\omega_c|\mathbf{x}), \quad c = 1, 2, \dots, C \quad (2)$$

The Bayes classifier can be used for both two classes and multiple classes classifications.

Usually there are not enough samples to estimate the conditional density function for each class in face recognition. A compromise is to assume that the within class densities can be modeled as normal distributions, and all the within class covariance matrices are identical and diagonal as in [12]. They call this approach probabilistic reasoning

model (PRM). The parameters of the normal distributions are estimated as follows,

$$\mu_c = \frac{1}{N_c} \sum_{j=1}^{N_c} \mathbf{x}_j^{(c)}, \quad c = 1, 2, \dots, C \quad (3)$$

where $\mathbf{x}_j^{(c)}$, $j = 1, 2, \dots, N_c$ represents the samples from class ω_c , and

$$\Sigma_I = \Sigma_c = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_D^2\} \quad (4)$$

where D is the feature dimension. Each component σ_i^2 can be estimated by the sample variance in the one dimensional PCA subspace

$$\sigma_i = \frac{1}{C} \sum_{c=1}^C \left\{ \frac{1}{N_c - 1} \sum_{j=1}^{N_c} \left(x_{ji}^{(c)} - \mu_{ci} \right)^2 \right\} \quad (5)$$

where $x_{ji}^{(c)}$ is the i -th element of the sample $\mathbf{x}_j^{(c)}$, μ_{ci} the i -th element of μ_c , and C the number of classes.

3.2. AdaBoost

Boosting is a method to combine a collection of weak classification functions (weak learner) to form a stronger classifier. AdaBoost is an adaptive algorithm to boost a sequence of classifiers, in that the weights are updated dynamically according to the errors in previous learning [4]. AdaBoost is a kind of large margin classifiers.

Tieu and Viola [22] adapted the AdaBoost algorithm for natural image retrieval. They made the weak learner work in a single feature each time. So after T rounds, T features are selected together with the T weak classifiers.

We evaluate the AdaBoost algorithm for face recognition. For each pair of individuals, the AdaBoost is used to run for T rounds. However, in quite a few cases, two persons can be separated completely using only one feature, so the error $\epsilon = 0$, which results in $\beta = \frac{\epsilon}{1-\epsilon} = 0$ in AdaBoost. Therefore the combination coefficients $\alpha = \log \frac{1}{\beta}$ can not be defined. To solve this problem, we experimentally set $\beta = 0.01$, if $\epsilon = 0$. Another observation is that in many dimensions, the error $\epsilon > 0.5$, not as claimed in [22] that $\epsilon \leq 0.5$ always holds. When the distances to the mean values are used [22], the weak learner is simple, *i.e.*, x is classified to class 1 if $|x - \mu_1| < |x - \mu_2|$.

AdaBoost Algorithm

Input: 1) n training examples $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ with $y_i = 1$ or 0; 2) the number of iterations T .

Initialize weights $w_{1,i} = \frac{1}{2l}$ or $\frac{1}{2m}$ for $y_i = 1$ or 0, with $l + m = n$.

Do for $t = 1, \dots, T$:

1. Train one hypothesis h_j for each feature j with w_t , and error $\epsilon_j = Pr_{r_i^{w_t}}[h_j(x_i) \neq y_i]$.

2. Choose $h_t(\cdot) = h_k(\cdot)$ such that $\forall j \neq k, \epsilon_k < \epsilon_j$. Let $\epsilon_t = \epsilon_k$.

3. Update: $w_{t+1,i} = w_{t,i}\beta_t^{e_i}$, where $e_i = 1$ or 0 for example x_i classified correctly or incorrectly respectively, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$, if $\epsilon_t \neq 0$, or $\beta_t = 0.01$, if $\epsilon_t = 0$.

4. Normalize the weights so that they are a distribution, $w_{t+1,i} \leftarrow \frac{w_{t+1,i}}{\sum_{j=1}^n w_{t+1,j}}$.

Output the final hypothesis,

$$h_f(x) = \begin{cases} 1 & \text{if } \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where $\alpha_t = \log \frac{1}{\beta_t}$.

4. Pairwise Recognition Structure

For a C class face recognition problem, it is first decomposed into a set of $C(C-1)/2$ two-class problems. For each pair, the features are ranked by their discriminative power, and a given number of features are chosen from the top of the sorted list to discriminate between that specific pair of classes. This strategy is different from the traditional approaches which use one feature space for the whole problem. Based on the selected features, the pairwise classifiers are trained. So, there are two steps in training under the pairwise recognition framework: 1) rank and select the features for each pair of individuals; 2) train the classifiers for each pair with the selected features. In testing, when a query face image is given, it goes through two stages: 1) do pairwise classifications in each pair of classes; 2) combine the pairwise comparison results to form a final decision.

4.1. Feature Ranking

Traditionally, feature selection is defined as follows: given a set of candidate features, select a subset that performs best under some classification system [8]. In the past decade, many research concentrations are on the search algorithms for feature selection. Jain and Zongker [8] evaluated different search algorithms for feature subset selection, and found that the sequential forward floating selection (SFFS) algorithm proposed by Pudil *et al.* [16] performs the best. However, the SFFS is very time consuming when the feature dimensions are very high (for face recognition, the extracted features are usually around or more than 100). Vailaya [24] tried the SFFS to select 67 features from 600 for a two-class problem (indoor vs. outdoor images), and reported that the SFFS run for 12 days to get a result. Moreover, it is difficult to select the features in the case of small sample size [8], which is exactly the case for face recognition, typically 2-5 training samples for each class with 100 dimensions.

Because of the difficulties of classical feature selection methodology for face data, we propose another concept called "feature ranking" to distinguish from "feature selection". In feature ranking, the features in different dimensions are assumed independent, and a criterion is used to compare the discriminative capabilities of each feature along the dimensions. The feature ranking approach simplifies and speeds up the process to pick up a subset of the given features. A simple criterion,

$$r_{ij}^d = \frac{\|\mu_{id} - \mu_{jd}\|}{\sigma_d} \quad (7)$$

is used to rank the features along the dimensions d , for $d = 1, 2, \dots, D$ for a pair of classes i and j , μ_{id} (or μ_{jd}) is the mean value of class i (or j), and σ_d is the variance of the samples in dimension d . Recall that the variance of the samples within different classes but in the same dimension is assumed equal in the Bayesian approach as [12]. The larger the values of r_{ij}^d , the more discriminative the d -th feature for the classification between classes i and j .

By using this kind of ranking, the features are sorted in descending order according to their discriminability as $f_{ij}^1, f_{ij}^2, \dots, f_{ij}^D$. The user specifies N (usually $N \ll D$), the number of features to use, and the system just selects first N features $f_{ij}^1, f_{ij}^2, \dots, f_{ij}^N$ to train the classifiers and also to recognize a new face.

Feature ranking is executed for each pair of individuals. It provides some knowledge regarding the importance of certain features over others for a specific classification problem. The top features in the list are expected to have higher discrimination capacity. Using less number of features in the top list reduces the dimensionality without losing the discrimination power.

The top N features are fed into the Bayes classifier for training. In AdaBoost, the features are selected one by one according to the classification error of the weak learner in last step [22].

4.2. Combining the Pairwise Classifiers

When a query face image comes, it passes $C(C-1)/2$ comparisons. The output of the $C(C-1)/2$ classifiers construct a matrix, as shown in Fig. 1. Each element is equal to 1 or 0. $\phi_{i,j}(\mathbf{x}) = 1$ if \mathbf{x} is classified to class i in the pairwise competition between classes i and j , otherwise $\phi_{i,j}(\mathbf{x}) = 0$. All elements in the main diagonal are zeros.

The outputs of the pairwise classifiers should be combined to obtain the final decision. There are two ways to combine them: (1) by simple voting [5], or (2) by using the MAP rule on an estimate of the overall *a posteriori* probabilities obtained from the outputs of the pairwise classifiers [7]. In the voting combination scheme, a count $c(\omega_i|\mathbf{x})$ of the number of the pairwise classifiers that label \mathbf{x} into class

$$\begin{pmatrix} 0 & \phi_{1,2} & \phi_{1,3} & \cdots & \phi_{1,C} \\ \phi_{2,1} & 0 & \phi_{2,3} & \cdots & \phi_{2,C} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \phi_{C,1} & \phi_{C,2} & \phi_{C,3} & \cdots & 0 \end{pmatrix}$$

Figure 1. The pairwise classification results $\phi_{i,j}$ are listed in a $C \times C$ matrix for the C class classification problem. The values of $\phi_{i,j}$ are equal to 1 or 0. If $\phi_{i,j} = 1$, then $\phi_{j,i} = 0$.

ω_i is calculated as,

$$c(\omega_i|\mathbf{x}) = \sum_j \phi_{i,j}(\mathbf{x}) \quad (8)$$

The input \mathbf{x} is assigned the class label for which the count is maximum,

$$\omega_c^* = \arg \max_i c(\omega_i|\mathbf{x}), \quad i = 1, 2, \dots, C \quad (9)$$

We use this simple combination strategy instead of the complex one [7], but still show the success of the pairwise recognition framework in next Section.

5. Experiments

The pairwise recognition framework is evaluated on a compound face database with 1079 face images of 137 persons. The Bayes classifier and AdaBoost algorithm are used for the classification of each pair of individuals. We compare the recognition rates of the pairwise approach with the probabilistic reasoning model (PRM) [12], and also the standard eigenface approach which uses the nearest center classification criterion [6].

5.1. Face Database

The face database is a collection of five databases: (1). The Cambridge ORL face database [19] which contains 40 distinct persons. Each person has ten different images. (2). The Bern database contains frontal views of 30 persons. (3). The Yale database contains 15 persons. For each person, ten of its 11 frontal view images are randomly selected. (4). Five persons are selected from the Harvard database, each with 10 images. (5). A database composed of 179 images of 47 Asian students, each with three or four images. In our face database, each face image is cropped from the original large images and resized to 128×128 pixels. There are

large variations in facial expressions and facial details, and also changes in light, face size and pose.

The face database is divided into two non-overlapping sets, one for training and the other for testing. The training data consist of 544 images: five images per person are randomly chosen from the Cambridge, Bern, Yale, and Harvard databases, and two images per person are randomly selected from the Asian students database. The remaining 535 images are used for testing.

5.2. Experimental Results

In the training stage, 120 principal components are extracted from the 544 face images by using PCA. The 120 features are normalized to a normal distribution $N(0, 1)$ independently along each dimension. For the database with $C = 137$ individuals, we train $\frac{C(C-1)}{2} = 9316$ classifiers. The mean and covariance matrices are estimated by using Eq.(3), (4), and (5). Note that sample variance in each dimension is independently estimated by the average over all classes. Then, all the pairwise classifiers are trained. The AdaBoost algorithm can sequentially select T features with each for one weak learner. In the pairwise probabilistic approach, Eq. (7) is used to rank the 120 features for each specific pair of individuals, and the top T features are used to train the classifiers. The indices of the selected features are stored for each pair of classifiers.

For a given query or test face, it goes through 9316 classifiers, each gives an output of 1 or 0. Then, a matrix of the pairwise classification results is obtained as in Fig. 1. Note that in the matrix, only the part in the upper triangle are calculated by the pairwise classifiers, while others in the bottom triangle is filled with the contrary values. For example, if $\phi_{i,j} = 1$ (or 0), then $\phi_{j,i} = 0$ (or 1). Next, the voting method of Eq. (8) is used to count the number of "1"s in each row. The biggest count corresponds to the class label the query belongs to.

Fig. 2 shows the final recognition rates with respect to the number of features to use. The results of the standard eigenfaces and the PRM [12] are also shown for comparison. In the pairwise approach, the features are selected sequentially for the probabilistic classification by using Eq. (7) for ranking, while the AdaBoost algorithm uses the comparison of the error rate of each hypothesis to select the features. In standard eigenface and the PRM approaches, the features derived from PCA are sorted in descending order according to the eigenvalues of the principal components. The higher the dimensions, the smaller the eigenvalues, as is the case of traditional approaches [12] [13]. It is obvious that under the pairwise recognition framework, the AdaBoost (labeled as PairBoost) and Bayes (labeled as PairProb) approaches have much higher recognition accuracies than the standard eigenfaces and PRM in the low

dimensions ($d < 20$). In Fig. 2, the feature dimensions start from 2, where the PairBoost and PairProb give the accuracy of 64.49% and 75.89%, while the eigenfaces and PRM just have 29.16% and 29.91% respectively. This indicates the advantages of independent feature selection for each specific pair of classes, although just 2 features are used. When the feature dimensions increase to 5, the PairBoost achieves accuracy of 83.94%, even higher than the 81.12% of PairProb. Both of them are much higher than the 58.13% of eigenfaces and 60.75% of the PRM. When the feature dimensions become higher, the performance of PairBoost does not improve much, even deteriorates a little. We interpret this as the interior parameters of the AdaBoost should be adjusted more carefully for the special case of face recognition in order to get high accuracy constantly. On the other hand, the PairProb shows good performance consistently with respect to the feature dimensions. These demonstrate that the pairwise framework is powerful for the complicated face recognition problem. We also note that the PRM method does not improve much of the performance over the standard eigenfaces. Further more, the best result of PRM is 85.79% corresponding to dimension 40, which is lower than the reported recognition rate of 96% with 44 features [12] on the "FERET" database. This indirectly indicates that face recognition on our database is a little more difficult.

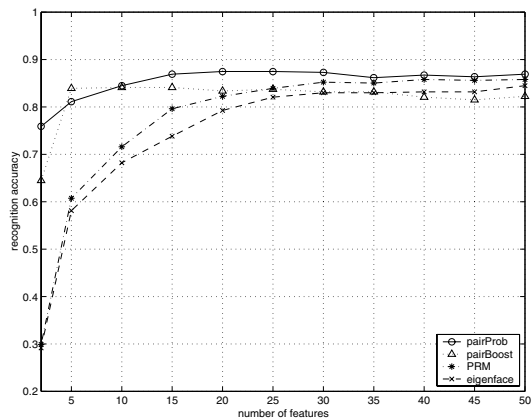


Figure 2. Face recognition performance comparison of the standard eigenface, PRM (probabilistic reasoning model), pairBoost (pairwise boosting), and pairProb (pairwise probabilistic classification).

5.3. Frequency of Feature Usage

In previous experiments, one can find that for a given number of features to use, the pairwise approach presents

better performance. In order to make it clearer, we list part of the indices of the actually used features in the pairwise comparisons in Table 1. The features used to discriminate class 1 from classes 2, 3, 4, and 5 are different. Some features such as indices 0, 4, 6 are used by class 1 to separate from classes 3 and 4, however, the discrimination between classes 3 and 4 uses different features. In addition, high dimensional features derived from PCA, such as indices 91 and 102, are selected by the system for discrimination.

Table 1. Indices of the features used for pairwise face recognition when 10 features used.

pairs	first 10 features									
1-2	8	32	47	91	1	2	11	23	42	50
1-3	0	4	6	21	42	52	91	102	1	12
1-4	0	4	6	12	14	24	42	50	52	58
1-5	0	1	4	42	50	61	90	5	8	12
3-4	9	11	17	18	19	25	26	44	51	53

Moreover, we compute the statistics of the features used by all the pairs. The left one in Fig. 3 shows the frequency of features used in the range of 120 dimensions, when just 2 features are used by each pair of classifiers. Even some high dimensional features (larger than 100) are selected for classifications. When the user specifies to use 30 features for each pair, more high dimensional features are selected, as shown on the right one in Fig. 3. Both figures show that the high dimensional features are still useful for discrimination, although the eigen values are small in these principal components. These features are usually discarded as in [15] and other approaches in face recognition.

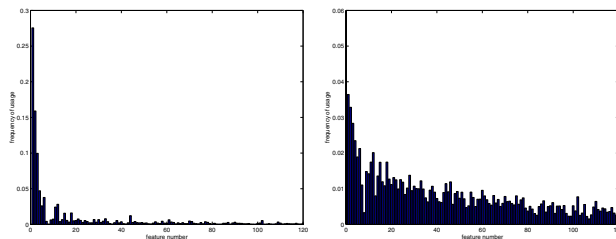


Figure 3. Frequency of usage of different features (of the 120 principal components) among the total pairs: Left – when just 2 features are used for each pair; Right – when 30 features are used.

6. Discussions

In the pairwise recognition framework, we take a simple and fast method to rank the features. Although simple, it still effectively picks up a small number of discriminative features for each pair of classes. Our main focus is the pairwise classification for face recognition. More effective method can be developed for feature selection if it properly deals with the situations of small sample size, which is expected to deliver even better results under the pairwise recognition framework.

To use AdaBoost for pairwise face recognition, we take the adapted version developed in [22]. The results are good in low dimensions but not so good overall as expected. We think more careful adjustments of the interior parameters maybe necessary in solving face recognition problems. In addition, the binary classification of each weak learner may be replaced by a probabilistic one to improve the final results which is explored currently.

Further research is to find the connection between the visual dissimilarity of two persons and the difference of the selected features, and determine how many features are sufficient for discrimination of a specific pair of individuals, instead of using the same number of features for all pairs. Using less features is especially useful for face retrieval [13].

To combine the pairwise classification results to get the final decision, we use a simple voting method. More complicated combination strategy like the MAP estimation [7] may further improve the recognition accuracy.

7. Conclusions

We have developed a pairwise framework for face recognition under which the original complex recognition problem is decomposed into a set of simpler ones to solve. Feature ranking is proposed for each specific pair of classes based on their discriminative abilities. Some high dimensional features (with small eigenvalues) derived from PCA are still useful for discrimination. The overall recognition rates are improved consistently for the probabilistic classifications. For the AdaBoost algorithm, further work should be done to improve its performance for face recognition.

References

- [1] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski. Independent component representations for face recognition. *Proc. of SPIE*, 2399:528–539, 1998.
- [2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. PAMI*, 19(7):711–720, 1997.
- [3] P. Comon. Independent component analysis - a new concept? *Signal Processing*, 36:287–314, 1994.
- [4] Y. Freund and R. E. Schapire. A decision-theoretic generalization of online learning and an application to boosting. *J. Comp. & Sys. Sci.*, 55(1):119–139, 1997.
- [5] J. Friedman. Another approach to polychotomous classification. *Technical report, Stanford University*, 1996.
- [6] K. Fukunage. *Introduction to Statistical Pattern Recognition*. Academic Press, second edition, 1991.
- [7] T. Hastie and R. Tibshirani. Classification by pairwise coupling. *Advances in NIPS*, 10:507–513, 1998.
- [8] A. Jain and D. Zongker. Feature selection: evaluation, application, and small sample performance. *IEEE Trans. on PAMI*, 19(2):153–158, 1997.
- [9] I. T. Jolliffe. *Principal Component Analysis*. Springer, New York, 1986.
- [10] M. A. Kramer. Nonlinear principal components analysis using autoassociative neural networks. *AIChE Journal*, 32(2):233–243, 1991.
- [11] S. Kumar, J. Ghosh, and M. Crawford. A bayesian pairwise classifier for character recognition. *Cognitive and Neural Models for Word Recognition and Document Processing, Nabeel Mursheed (Ed), World Scientific Press*, 2000.
- [12] C. Liu and H. Wechsler. Probabilistic reasoning models for face recognition. *Proc. of CVPR*, pages 827–832, 1998.
- [13] C. Liu and H. Wechsler. Robust coding scheme for indexing and retrieval from large face database. *IEEE Trans. Image Processing*, 9(1):132–137, 2000.
- [14] S. Mika, G. Ratsch, J. Weston, and K.-R. M. B. Scholkopf. Fisher discriminant analysis with kernels. *Neural Networks for Signal Processing IX*, pages 41–48, 1999.
- [15] B. Moghaddam. Principal manifolds and bayesian subspaces for visual recognition. *ICCV*, pages 1131–1136, 1999.
- [16] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11):1119–1125, 1994.
- [17] S. S. R. Chellappa, C. L. Wilson. Human and machine recognition of faces: A survey. *Proc, IEEE*, 83:705–741, May 1995.
- [18] A. Samal and P. A. Iyengar. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition*, 25:65–77, 1992.
- [19] F. S. Samaria. Face recognition using hidden markov models. *PhD thesis, University of Cambridge*, 1994.
- [20] B. Scholkopf, S. Mika, A. Smola, G. Ratsch, and K.-R. Miller. Kernel pca pattern reconstruction via approximate pre-images. *Proc. ICANN*, pages 147–152, 1998.
- [21] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. Opt. Soc. Amer., A*, 4(3):519–524, 1987.
- [22] K. Tieu and P. Viola. Boosting image retrieval. *Proc. of CVPR*, 1:228–235, 2000.
- [23] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *J. Cognitive Neurosci.*, 3(1):71–86, 1991.
- [24] A. Vailaya. *Semantic classification in image database*. Ph.D. thesis, Michigan State University, 2000.
- [25] D. Valentin, H. Abdi, and G. W. C. A. J. O’Toole. Connectionist models of face processing: A survey. *Pattern Recognition*, 27:1209–1230, 1994.
- [26] V. N. Vapnik. *Statistical learning theory*. John Wiley & Sons, New York, 1998.