Face Recognition with MRC-Boosting

Xun Xu and Thomas S. Huang
Beckman Institute
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
{xunxu, huang}@ifp.uiuc.edu

Abstract

In this paper, a novel classification algorithm called MRC-Boosting is proposed. Through aggregating Maximal-Rejection-Classifier features under boosting framework, this algorithm can deal with complicated two-class classification problem, especially the category called target detection problem where a target class should be discriminated from the surrounding clutter class. MRC-Boosting is efficient since unlike many other boosting based algorithms, at each iteration the optimal feature is computed in closed-form, with neither exhaustive search nor time-consuming numerical optimization. Furthermore, a variant of MRC-Boosting is derived and applied to face recognition. This variant MRC-Boosting algorithm is able to utilize large amount of training samples efficiently, overcoming the difficulty faced by other algorithms like AdaBoost. The effectiveness of the proposed algorithm is validated by face recognition experiments on CMU-PIE database.

1 Introduction

As a problem with theoretical importance and wide applications, automatic face recognition has been actively investigated in the computer vision and pattern recognition community for a long time. Since the well-known Eigenface method [11][13], many methods have emerged during the past decades, among which an influential one is the Bayesian method proposed by Moghaddam and Pentlend [6]. The essential idea of Bayesian method is converting face recognition to a two-class discrimination problem for face variation, i.e. classifying the difference between two face images as intra-personal or extra-personal. This framework directly models the variation of face images which is most critical for recognition, although it is simple and easy to implement, it outperforms many conventional methods. The original Bayesian method assumes Gaussian distribution for both the intra-personal and extra-personal differences. However, due to the complicated variation that human faces may have, this assumption may not be valid, thus the Bayesian classifier learned assuming Gaussian distribution for the two classes may not be good enough.

Following such a framework of recognizing faces by discriminating differences, recently more sophisticated classifiers such as SVM [4] and AdaBoost [3][15] have been employed to solve this problem. AdaBoost is a simple way to build a strong classifier from weak classifiers. However, each iteration of AdaBoost involves searching a large pool of candidate weak classifiers, which is very computationally expensive. On the other hand, this scheme also restricts the best classifier that can be sought (since the chosen classifiers must come from the pool), if the pool does not contain classifiers that are discriminative for the classes under consideration, AdaBoost cannot perform well. Moreover, in the works applying AdaBoost to face recognition [3][15], another difficulty faced by this algorithm is the huge number of training samples. In both [3] and [15], resampling strategy is employed to select a small part of samples for training at a time. Although directly utilizing the whole training sample set is preferable, it is infeasible for AdaBoost to do so.

In this paper, we are going to show that the face difference discrimination problem belongs to the category called target detection, where a target class should be separated from the surrounding clutter class. An effective algorithm to tackle problems of this category, namely MRC-Boosting, is proposed, which aggregates Maximal-Rejection-Classifier [1] features in boosting framework. The merit of this algorithm is that at each iteration, it computes the most discriminative feature in closed-form, i.e. there is no exhaustive search (like AdaBoost) or numerical optimization (like KL-Boosting [5]) involved. Furthermore, a variant of MRC-Boosting for face recognition is derived, which is able to directly utilize the whole training sample set in an efficient way, thus overcome the difficulty faced by other algorithms such as AdaBoost.

In section 2, we first briefly review Maximal-Rejection-Classifier and generalize it to weighted case, then propose MRC-Boosting algorithm. In section 3, face recognition is analyzed as a target detection problem, and a variant of MRC-Boosting which is very efficient for this problem is then derived. Section 4 will present experiments on CMU-PIE database which validate the effectiveness of the proposed method. Following is section 5 where the proposed method is compared to related methods and section 6 concludes.

2 MRC-Boosting

2.1 Maximal-Rejection-Classifier (MRC)

Classification is a core problem of pattern recognition, for which linear classifiers are one of the simplest and fastest solution. Well-known Fisher Linear Discriminant (FLD) assumes that the two classes to be discriminated are linearly separable (or nearly so). But for many practical problems, it may not be the case. One important category of classification problem is "target detection", where two classes to be discriminated are called the "target" and the "clutter". In the feature space, the samples from the target class are surrounded by clutter samples. The goal is to separate (detect) the target from the clutter. Clearly, in this case the two classes are completely not linearly separable. Elad et al [1] proposed an effective method, called Maximal-Rejection-Classifier (MRC) to solve this problem. MRC is a linear-based classifier which is able to deal with non-linearly separable two class problem. The formulation of this method is similar to that of FLD, the difference lies in the criterion function. Instead of minimizing the within-class scatter while maximizing between-class scatter as FLD does, MRC tries to find the projection vector which minimizes target scatter while maximizes clutter scatter. Formally, it seeks a vector $\mathbf{w}$ minimizing a functional:

$$\mathbf{w} = \arg\min_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{R}_X \mathbf{w}}{\mathbf{w}^T[\mathbf{R}_X + \mathbf{R}_Y + (\mathbf{m}_X - \mathbf{m}_Y)(\mathbf{m}_X - \mathbf{m}_Y)^T]\mathbf{w}} \qquad (1)$$

where $\mathbf{m}_X$ and $\mathbf{R}_X$ are the mean and covariance matrix of the target class X respectively, $\mathbf{m}_Y$ and $\mathbf{R}_Y$ are those of clutter class Y. In the case that both X and Y have zero means, (1) can be equivalently written as:

$$\mathbf{w} = \arg\min_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{R}_X \mathbf{w}}{\mathbf{w}^T \mathbf{R}_Y \mathbf{w}} \qquad (2)$$

Note that the $\mathbf{R}_X$ in the denominator is also dropped, it is easy to verify that this does not affect the solution. Just like FLD, the functional to be minimized is a generalized Rayleigh quotient, and the optimal $\mathbf{w}$ can be found through solving a generalized eigenvalue problem and picking the eigenvector associated with the smallest eigenvalue. Intuitively, this formulation makes sense since it tries to find a projection vector on which the target "shrinks" as much as possible, and the clutter is pushed apart as possible, so that the overlapping between the target and the clutter is minimized. Note that in the target detection scenario, it is impossible to separate the two classes on any projection vector (thus FLD cannot work well), the vector sought by MRC is the optimal one in the sense that it may achieve minimal classification error rate.

After finding the optimal projection vector $\mathbf{w}$, samples from both classes are projected to $\mathbf{w}$, then *two* threshold values T_1 and T_2 are picked so that as many as possible clutter samples are rejected while retaining all target samples, see Figure 1. Note that since two thresholds are used, the decision region for the target using one MRC, i.e. $\{\mathbf{x} : T_1 \leq \mathbf{w}^T\mathbf{x} \leq T_2\}$, is the region between two parallel hyperplanes, therefore MRC is not a linear classifier, but linear-based [1].

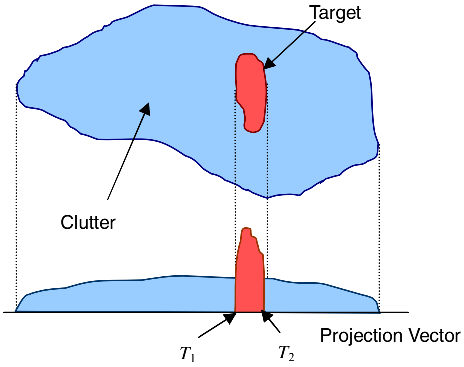


Figure 1. Maximal-Rejection-Classifier

2.2 Weighted MRC

Original MRC assumes the contribution of each sample is equal, it can be generalized to the case that each sample carries some weight thus the sample set represents the underlying distribution of the two classes. Supposing we have a target sample set $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{N_X}\}$ with weights $\{w_1^X, w_2^X, \cdots, w_{N_X}^X\}$ for each $\mathbf{x}_i \, (i = 1, 2, \ldots, N_X)$ respectively, also a clutter sample set $\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_{N_Y}\}$ and weights $\{w_1^Y, w_2^Y, \cdots, w_{N_Y}^Y\}$. Since the weights represents a probability distribution, we have $\sum_{i=1}^{N_X} w_i^X + \sum_{j=1}^{N_Y} w_j^Y = 1$. With the weights, the covariance matrices of two classes are estimated by $\mathbf{R}_X = \sum_{i=1}^{N_X} w_i^X \mathbf{x}_i \mathbf{x}_i^T$ and $\mathbf{R}_Y = \sum_{j=1}^{N_Y} w_j^Y \mathbf{y}_j \mathbf{y}_j^T$. For the simplicity of presentation, we still assume that both classes are of zero-mean. With such definitions, the expression given by (2) is directly applicable to find a weighted MRC.

2.3 MRC-Boosting

Boosting is an effective framework to construct a strong classifier by combining weak component classifiers. The most popular boosting algorithm, AdaBoost, iteratively adds weak classifiers into the ensemble, focusing on most informative samples which are not classified well by previously added classifiers. This is done by giving each sample a weight and reweighting each sample at each iteration according to whether it is correctly classified or not at previous iteration. With AdaBoost, the training error of the classifier ensemble can be made arbitrarily low. AdaBoost is a simple and effective way to build a powerful classifier, but its performance and efficiency largely depends on the what weak classifiers are employed. In most literatures which apply AdaBoost to computer vision problems [14][3][15], the “optimal” weak classifiers are found through searching a huge pool (e.g. 52,374 in [3], 60,480 for [15]) of candidate classifiers, and selecting the one with minimal error rate. This is not only time consuming, but also suboptimal, since the candidate pool restricts the classifiers that can be used. Liu and Shum [5] proposed KL-Boosting algorithm which is able to find an optimal linear feature at each iteration, thus a strong classifier can be built with much fewer weak classifiers compared to AdaBoost. However, taking KL divergence between the two classes’ marginal distributions as the goal function, KL-Boosting algorithm has to employ numerical optimization to find such optimal classifiers which is also computationally expensive. Tu et al [12] proposed Fisher-Boosting method employing FLD as weak classifiers, which can be calculated efficiently. However, FLD can not work well for highly non-linearly separable classes such as those in the target detection problem. Therefore, it is not likely to build an effective classifier using Fisher-Boosting for such problems. On the contrary, MRC has the capability of dealing with such non-linearly separable cases. Furthermore, weighted MRC’s can be aggregated in a way similar to AdaBoost to construct a strong classifier which is able to tackle complex, target detection like classification problems. This leads to the MRC-Boosting algorithm, which can be used to discriminate the target (denoted by class label $+1$) from the clutter (denoted by -1), as shown in Figure 2.

Note that original MRC [1] can only construct a convex (more precisely, a parallelogram polytope) decision region for the target class since the component MRC’s are simply combined using AND rule. Whereas in MRC-Boosting, the component MRC’s are aggregated through weighted voting, so that more sophisticated decision bound can be formed.

Given: $\left\{ (\mathbf{x}_i, c_i), i = 1, 2, \cdots, N : \mathbf{x}_i \in \mathrm{R}^d, c_i \in \{+1, -1\} \right\}$

1. Initialize: $w_i = \begin{cases} 1/2N^+ & , c_i = +1 \\ 1/2N^- & , c_i = -1 \end{cases}$, where N^+ and N^- are the numbers of the target and clutter samples, respectively. The maximal number of weak classifiers K.
2. For $k = 1, 2, \cdots, K$
    - (a) Find optimal weighted MRC vector $\mathbf{w}$ through solving (2), where $\mathbf{R}_X = \sum\limits_{i:c_i=+1} w_i \mathbf{x}_i \mathbf{x}_i^T$ and $\mathbf{R}_Y = \sum\limits_{i:c_i=-1} w_i \mathbf{x}_i \mathbf{x}_i^T$.[1]
    - (b) Obtain a weak classifier:

      $$f_k(\mathbf{x}; \mathbf{w}, T_1, T_2) = \begin{cases} +1 & , T_1 \le \mathbf{w}^T \mathbf{x} \le T_2 \\ -1 & , else \end{cases}$$

      T_1 and T_2 are determined by minimizing classification error $\varepsilon_k = \sum\limits_{i=1}^{N} w_i I\left(f_k(\mathbf{x}_i) \ne c_i\right)$.
    - (c) Updating weights: $w_i \leftarrow \frac{1}{Z_k} w_i \exp\left[-\alpha_k c_i f_k(\mathbf{x}_i)\right]$, where $\alpha_k = \frac{1}{2} \ln \frac{1-\varepsilon_k}{\varepsilon_k}$ and Z_k is a normalization factor to ensure $\sum_{i=1}^{N} w_i = 1$.

Output: Strong classifier $F(\mathbf{x}) = sgn\left[G(\mathbf{x})\right]$ where the classification function is $G(\mathbf{x}) = \sum_{k=1}^{K} \alpha_k f_k(\mathbf{x})$.

Figure 2. MRC-Boosting algorithm

3 Efficient face recognition with MRC-Boosting

3.1 Face recognition as target detection

Bayesian method for face recognition [6] is among the most influential ones proposed in the recent years. The essential idea of Bayesian method is converting face recognition to a two-class discrimination problem for face variation, i.e. classifying the difference between two face images as intra-personal or extra-personal. In [6], the two categories are modeled using Gaussian distribution, and a Bayesian classifier is constructed. With Bayesian method, face recognition is performed with MAP or Maximum-Likelihood principle, and this can be efficiently implemented as an enhanced eigenface algorithm. Since this method directly models the variation of face images which

---

[1] For the simplicity of presentation, here we assume zero-mean for both classes, and the general case contains terms related to the weighted means and is also easy to compute.

is most critical for recognition, although it is simple and easy to implement, it outperforms many conventional methods. The main drawback of Bayesian method is that it assumes Gaussian distribution for both the intra-personal and extra-personal differences. However, due to the complex variation that people's faces may have under different poses, lighting conditions and expressions etc, the actual probability distribution of the differences is more complicated than Gaussian. Therefore, the simplified Gaussian assumption restricts the performance of Bayesian method.

Naturally, one possibility to achieve higher performance is to seek a better classifier which is more accurate for this complicated two-class discrimination problem. Once such a classifier, say $F(\mathbf{d}) = sgn\left[G(\mathbf{d})\right]$, is sought, we may define the similarity between two faces $\mathbf{F}_1$ and $\mathbf{F}_2$ as the output of the classification function, i.e. $S(\mathbf{F}_1, \mathbf{F}_2) = G(\mathbf{F}_1 - \mathbf{F}_2)$. Larger $S(\mathbf{F}_1, \mathbf{F}_2)$ implies larger positive margin from the decision bound and $\mathbf{d} \equiv \mathbf{F}_1 - \mathbf{F}_2$ is more likely to be an intra-personal difference, therefore $\mathbf{F}_1$ and $\mathbf{F}_2$ are more similar in the sense that they have the same identity. Along this path, more sophisticated classifiers such as SVM [4] and AdaBoost [3][15] have been employed recently. As an simple way to aggregate weak classifiers into a strong classifier, AdaBoost [9] has received much attention in recent years and is applied successfully to face detection [14]. However, as discussed before, it has several drawbacks which restrict its efficiency and effectiveness.

In the two-class discrimination framework for face recognition, we need a good classifier which is able to separate intra-personal differences from the extra-personal ones. From a geometric point of view, in the image space both intra-personal and extra-personal differences have distributions symmetric with respect to the origin because $\mathbf{F}_1 - \mathbf{F}_2$ and $\mathbf{F}_2 - \mathbf{F}_1$ necessarily belong to the same class. Furthermore, intra-personal differences are surrounded by the extra-personal ones since most intra-personal differences have smaller magnitudes than those of extra-personal ones, as depicted in Figure 3. Therefore, this specific two-class problem belongs to the category of target detection where the MRC-Boosting algorithm proposed in subsection 2.3 is applicable.

3.2 Efficient MRC-Boosting algorithm for face recognition

Although MRC-Boosting given in Figure 2 seems to be directly applicable, there is another significant problem which should be taken into consideration. Since here the two classes to be discriminated are intra-personal and extra-personal *differences*, the number of training samples is prohibitively large. For instance, if we have 1,000 face images as training data, the total number of the differences between any two of them will be one million. In existing literatures employing AdaBoost algorithm [3][15], sub-sampling strategy has to be used to select a small part of samples for training at a time, because searching weak classifiers using all the training samples is infeasible. Clearly, utilizing the whole set of training samples is more desirable since it will lead to least biased classifiers. Here we show that with MRC-Boosting, it is able to directly take into account all the training samples in a computationally efficient way, thus overcome the problem of huge training set.

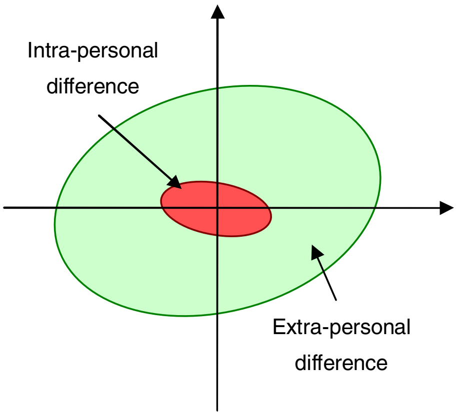


Figure 3. Face recognition as target detection: a geometric point of view

In the case of face recognition, during the training stage we are given a set of training faces $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$, with known identities $\{c_1, c_2, \cdots, c_N\}$. That is, the i-th and j-th faces belong to the same subject if and only if $c_i = c_j$. Taking difference between each pair of training faces generates N^2 differences which constitute the training sample set. As in the general case of MRC-Boosting discussed in subsection 2.3, each of these differences, $\mathbf{d}_{ij} \equiv \mathbf{x}_i - \mathbf{x}_j \quad (i, j = 1, 2, \cdots, N)$, carries a weight w_{ij}. Obviously we should have $w_{ij} = w_{ji}$ since two symmetric differences $\mathbf{d}_{ij}$ and $\mathbf{d}_{ji} = -\mathbf{d}_{ij}$ are equivalent.

In order to find the optimal weighted MRC projection vector at each MRC-Boosting iteration, we should compute the covariance matrix of all weighted intra-personal differences:

$$\mathbf{S}_I = \sum_{i,j:c_i=c_j} w_{ij} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T$$

and that of the weighted extra-personal differences:

$$\mathbf{S}_E = \sum_{i,j:c_i \neq c_j} w_{ij} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T$$

Direct computation of these two matrices is very expensive since the computational complexity is $\mathcal{O}\big(N^2 D^2\big)$

where D is the dimensionality of the face images. Fortunately, they can be computed in a much efficient way. We define D-by-N matrix consisting of the training face vectors $\mathbf{X} = \left[\begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_N \end{array}\right]$, N-by-N intra-personal weighting matrix:

$$\mathbf{W}_I\left(i,j\right) = \begin{cases} w_{ij} & , c_i = c_j \\ 0 & , c_i \neq c_j \end{cases}$$

and extra-personal weighting matrix:

$$\mathbf{W}_E\left(i,j\right) = \begin{cases} w_{ij} & , c_i \neq c_j \\ 0 & , c_i = c_j \end{cases}$$

It can be shown that (See Appendix A for the derivation):

$$\mathbf{S}_I = 2\mathbf{X}\tilde{\mathbf{W}}_I\mathbf{X}^T \tag{3}$$

and

$$\mathbf{S}_E = 2\mathbf{X}\tilde{\mathbf{W}}_E\mathbf{X}^T \tag{4}$$

where $\tilde{\mathbf{W}}_I = \mathrm{diag}\left(\mathbf{W}_I\mathbf{e}\right) - \mathbf{W}_I$ and $\tilde{\mathbf{W}}_E = \mathrm{diag}\left(\mathbf{W}_E\mathbf{e}\right) - \mathbf{W}_E$. Here $\mathbf{e} = \left[\begin{array}{ccc} 1, & \cdots, & 1 \end{array}\right]^T$ and $\mathrm{diag}\left(\mathbf{v}\right)$ denotes a diagonal matrix formed by the elements of vector $\mathbf{v}$. Therefore, with these expressions, $\mathbf{S}_I$ and $\mathbf{S}_E$ can be computed efficiently with complexity $\mathcal{O}\big(ND^2 + N^2D\big)$. Compared to direct computation, the saving ratio is $(N + D) : ND$. Considering a typical case $N = D = 1000$, that is $1 : 500$. It can also be noted that usually $\mathbf{W}_I$ is a highly sparse matrix, so that the computation of $\mathbf{S}_I$ is actually less than $\mathcal{O}\big(ND^2 + N^2D\big)$.

Through the efficient computation of the intra-personal and extra-personal covariance matrices, we may utilize the whole training sample set directly in a computationally feasible way. Finally, this leads to an efficient MRC-Boosting algorithm for face recognition, shown in Figure 4.

In the recognition phase, the face similarity function $S\left(\mathbf{p},\mathbf{g}\right)$ defined in Figure 4 is used to find the most similar gallery face $\mathbf{g}^*$ for a probe $\mathbf{p}$. The classifier $F\left(\mathbf{p},\mathbf{g}\right) = sgn\left[S\left(\mathbf{p},\mathbf{g}\right)\right]$ can be used for face verification.

Given: Training faces $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ and known identities $\{c_1, c_2, \cdots, c_N\}$.

1. Initialize: $w_{ij} = \begin{cases} {}^{1}\!/_{2N_I} & , c_i = c_j \\ {}^{1}\!/_{2N_E} & , c_i \neq c_j \end{cases}$ where N_I and N_E are the numbers of the intra-personal and extra-personal differences, respectively. The maximal number of weak classifiers K.
2. For $k = 1, 2, \cdots, K$
   - (a) Compute intra-personal and extra-personal covariance matrices $\mathbf{S}_I$ and $\mathbf{S}_E$, via (3) and (4), respectively.
   - (b) Find optimal weighted MRC vector $\mathbf{w}$ through solving $\mathbf{w} = \arg\min\limits_{\mathbf{w}} \frac{\mathbf{w}^T\mathbf{S}_I\mathbf{w}}{\mathbf{w}^T\mathbf{S}_E\mathbf{w}}$.
   - (c) Obtain a weak classifier:

     $$f_k\left(\mathbf{x}_i, \mathbf{x}_j; \mathbf{w}, T\right) = \begin{cases} +1 & , \left|\mathbf{w}^T\left(\mathbf{x}_i - \mathbf{x}_j\right)\right| \leq T \\ -1 & , else \end{cases}$$

     The threshold T is determined by minimizing classification error:

     $$\varepsilon_k = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} I\left(f_k\left(\mathbf{x}_i, \mathbf{x}_j\right) \neq \lambda_{ij}\right)$$

     where $\lambda_{ij} = \begin{cases} +1 & , c_i = c_j \\ -1 & , c_i \neq c_j \end{cases}$.
   - (d) Updating weights: $w_{ij} \leftarrow \frac{1}{Z_k} w_{ij} \exp\left[-\alpha_k \lambda_{ij} f_k\left(\mathbf{x}_i, \mathbf{x}_j\right)\right]$, where $\alpha_k = \frac{1}{2}\ln\frac{1-\varepsilon_k}{\varepsilon_k}$ and Z_k is a normalization factor to ensure $\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} = 1$.

Output: Strong classifier $F\left(\mathbf{p},\mathbf{g}\right) = sgn\left[S\left(\mathbf{p},\mathbf{g}\right)\right]$, where $S\left(\mathbf{p},\mathbf{g}\right) = \sum_{k=1}^{K}\alpha_k f_k\left(\mathbf{p},\mathbf{g}\right)$ is the similarity measure of two faces $\mathbf{p}$ and $\mathbf{g}$.

Figure 4. Efficient MRC-Boosting algorithm for face recognition

4 Experiments

To validate the effectiveness and efficiency of the proposed MRC-Boosting algorithm, face recognition experiments are performed on CMU PIE database [10], comparing our method to others. CMU PIE database contains 40,000+ images of 68 subjects, for each person we select 168 images which cover large illumination and pose change and moderate variation in expression, constituting a very challenging face database for recognition task. Face images are cropped out from the selected images and resized to be 32x32. Samples from the 11,424 selected face images are depicted in Figure 5.

Before our experiments, all these images are normalized to be of zero-mean and unit-variance. The 168 images of each person are randomly partitioned in to 3 disjoint sets: 60 for training, 5 as gallery, and the remaining as probe faces. It should be noted that this experimental setting makes the recognition task fairly challenging due to the small gallery size and large variation in probe faces. An

Figure 5. Samples of CMU-PIE face images used in our experiments. Note that there are large variation in pose, illumination and expression.

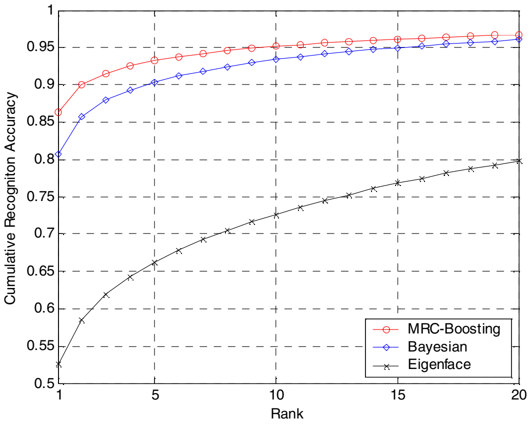


Figure 6. The average Cumulative Matching Characteristic curves for MRC-Boosting, Bayesian method, and Eigenface.

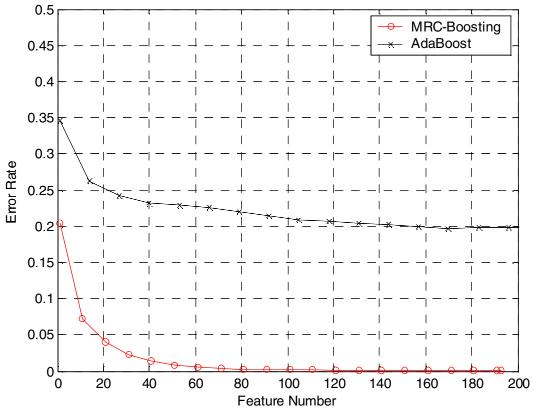
algorithm can perform well in such experiments only if it is capable of learning a model of face variation from the training data, and successfully predicting the possible variation for the limited set of gallery faces. Recognition is performed using MRC-Boosting and other three representative methods, Bayesian method [6], Eigenface [11][13], and AdaBoost [3][15]. Firstly, we compared MRC-Boosting with Bayesian method to verify whether it can do better in discriminating intra-personal and extra-personal differences, and Eigenface is used as a baseline method. MRC-Boosting is trained on the training set, 500 features are obtained in total. The dimensions of both the intra-personal and extra-personal subspaces of Bayesian method are chosen to be 125, a typical value used by [6]. The dimension of Eigenface is also set as 125. The Cumulative Matching Characteristic curves [7] of these three methods are shown in Figure 6. These curves are obtained by averaging the results from 10 experiments. In each of the experiments, the training images (60 for each person) are fixed, 5 out of the remaining 108 are randomly chosen as gallery set, and the others are used as probe images. The rank-1 recognition rate of Eigenface only reaches 52.5%, implying the challenge of this experimental setting. Bayesian method has the rank-1 recognition rate of 80.6%, superior to Eigenface because it directly models face variation which is critical for recognition. Our MRC-Boosting method achieves 86.4%, clearly outperforms Bayesian method. This result indicates that MRC-Boosting method is capable of modeling complex face variation better than Bayesian method does.

AdaBoost is another algorithm with the potential to build a strong classifier, in order to compare the effectiveness of AdaBoost and MRC-Boosting for face recognition, experiments are performed under the same setting. AdaBoost is trained with the candidate feature pool consisting of both rectangular features and Gabor features, i.e. a superset of those used in [3] and [15]. Since the number of training samples (differences) are prohibitively large (more than 16 million), resampling strategy suggested by [3] is taken. Figure 7(a) shows the curves of both algorithms' error rate on training data with respect to the number of features. As more and more features added to the classifier ensemble, decreasing error rate can be observed for both algorithms. However, the error rate of MRC-Boosting is decreasing more rapidly than that of AdaBoost, implying that the MRC-Boosting is able to find more discriminative features at each iteration. Figure 7(b) shows the curves of the rank-1 recognition accuracy on testing data with respect to the number of features. Again, MRC-Boosting exhibits more rapid increase, and ending up with higher recognition rate than that of AdaBoost. This indicates that the feature selection mechanism of MRC-Boosting is effective. On the contrary, due to the large variation of faces involved in the experiments, it is difficult for AdaBoost to find most discriminative features from its weak classifier candidates, resulting in its low recognition rate.

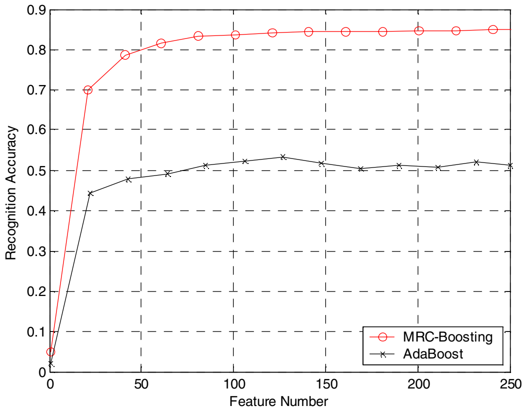
5 Discussion

5.1 Relationship with LPP

It may be noted that in subsection 3.2, the formulation of weighted MRC shares some similarity with Locality-Preserving Projections (LPP) [2]. The optimization of both methods have similar form, and both can be solved through generalized eigenvalue decomposition. The difference is that MRC seeks a projection vector best discriminating two classes (i.e. intra/extra-personal differences in the scenario



(a) Error rate on the training data vs. the number of features



(b) Face recognition accuracy vs. the number of features

Figure 7. Comparison of MRC-Boosting and AdaBoost

of face recognition), while LPP tries to find a projection vector best preserving locality relationship between samples in the high dimension space. Also, the similarity matrix S defined in [2] and our intra-personal weighting matrix $\mathbf{W}_I$ share a similar form, and so do $\tilde{\mathbf{W}}_I$ and their Laplacian matrix L. In LPP the weights in the similarity matrix are set heuristically, whereas in MRC-Boosting, the weights are dynamically adjusted through the learning process, automatically focusing on most informative face pairs.

5.2 Appearance vs. 3D model

MRC-Boosting face recognition algorithm is a pure appearance based method, it models all kinds of variation of human faces in a general framework. 3D model based methods often demonstrate better recognition performance than appearance based methods, especially when there exist large variation in pose and illumination in the face database. For example, in [8] high recognition rates are reported on CMU-PIE database by fitting a 3D morphable model, better than that of our appearance based method. However, 3D model based methods necessarily involve fitting a 3D model to *all* the face images, including every gallery and probe faces, which is an expensive procedure. Moreover, almost all 3D model fitting algorithms need manual initialization. When the size of face database is large, 3D model based methods are often unpractical. Therefore, appearance based methods are usually more efficient.

On the other hand, our MRC-Boosting face recognition algorithm can indeed be combined with 3D model based methods. We have seen that MRC-Boosting is able to model the variation of faces, actually it can also model the variation of *features*. We may apply MRC-Boosting to analyze the features obtained through 3D model fitting, thus achieve a better classification model than the nearest neighbor classifier usually used by existing 3D model based methods.

6 Conclusion

In this paper we proposed MRC-Boosting, an effective classification algorithm aggregating MRC features through the boosting framework. This algorithm is able to handle complicated two-class classification problem, especially the category called target detection problem where a target class should be discriminated from the surrounding clutter class. The training of MRC-Boosting is very efficient, at each iteration the optimal feature is computed in closed-form, which is much more efficient than searching a huge feature pool or numerical optimization as done by many other boosting based algorithms. Furthermore, we applied MRC-Boosting to face recognition, and proposed a variant of MRC-Boosting, which is able to utilize huge amount of training samples efficiently. Face recognition experiments on CMU-PIE database under a challenging setting demonstrate the effectiveness of the proposed method.

Appendix

A The derivation of Equation 3 and 4

We consider the general case, where each difference sample $\mathbf{d}_{ij} \equiv \mathbf{x}_i - \mathbf{x}_j$ is weighted by $\mathbf{W}(i,j) = w_{ij}$. We have:

$$
\begin{aligned}
\mathbf{S} &= \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\left(\mathbf{x}_i-\mathbf{x}_j\right)\left(\mathbf{x}_i-\mathbf{x}_j\right)^T \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\left(\mathbf{x}_i\mathbf{x}_i^T+\mathbf{x}_j\mathbf{x}_j^T-\mathbf{x}_i\mathbf{x}_j^T-\mathbf{x}_j\mathbf{x}_i^T\right) \\
&= \sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\sum_{j=1}^{N} w_{ij}+\sum_{j=1}^{N}\mathbf{x}_j\mathbf{x}_j^T\sum_{i=1}^{N} w_{ij} \\
&\quad -\sum_{i=1}^{N}\mathbf{x}_i\sum_{j=1}^{N} w_{ij}\mathbf{x}_j^T-\sum_{j=1}^{N}\mathbf{x}_j\sum_{i=1}^{N} w_{ij}\mathbf{x}_i^T \\
&= 2\left(\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\sum_{j=1}^{N} w_{ij}-\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\mathbf{x}_i\mathbf{x}_j^T\right) \\
&= 2\mathbf{X}\tilde{\mathbf{W}}\mathbf{X}^T
\end{aligned}
$$

where $\tilde{\mathbf{W}} = \operatorname{diag}(\mathbf{We}) - \mathbf{W}$. Substituting $\mathbf{W}$ with $\mathbf{W}_I$ and $\mathbf{W}_E$ as are defined in subsection 3.2, Equation 3 and 4 are obtained immediately.

References

[1] M. Elad, Y. Hel-Or, and Renato Keshet, “Pattern Detection Using a Maximal Rejection Classifier”, *Pattern Recognition Letters*, Vol. 23, No. 12, pp. 1459-1471, October 2002.

[2] Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-Jiang Zhang, “Face Recognition Using Laplacianfaces”, *IEEE Transactions on Patten Analysis and Machine Intelligence*, v27, n3, Mar. 2005.

[3] Jones, M.J. and Viola, P., “Face Recognition Using Boosted Local Features”, *MERL Technical Report*. TR 2003-25, April 2003.

[4] Z. Li and X. Tang, “Bayesian face recognition using support vector machine and face clustering”, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2004.

[5] C. Liu and H.Y. Shum. “Kullback-Leibler boosting”. *IEEE Computer Society Conference of Computer Vision and Pattern Recognition*, pp. 587-594, 2003

[6] Moghaddam, B.; Jebara, T.; Pentland, A., “Bayesian Face Recognition”, *Pattern Recognition*, Vol 33, Issue 11, pps 1771-1782, November 2000

[7] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi and Patrick J. Rauss, “The FERET Evaluation Methodology for Face-Recognition Algorithms”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp 1090-1104, v22, n10, Oct. 2000.

[8] S. Romdhani, V. Blanz, and T. Vetter. “Face identification by matching a 3D morphable model using linear shape and texture error functions”. *In Proceedings of the European Conference on Computer Vision*, pages 3-19, 2002.

[9] Robert E. Schapire, Yoav Freund, Peter Bartlett and Wee Sun Lee, “Boosting the margin: a new explanation for the effectiveness of voting methods”, *In Proc. 14th International Conference on Machine Learning*, pp 322–330, Morgan Kaufmann, 1997

[10] Terence Sim, Simon Baker and Maan Bsat, “The CMU Pose, Illumination, and Expression Database”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp 1615-1618, v25, n12,Dec. 2003.

[11] L. Sirovich and M. Kirby, “Low-Dimensional Procedure for the Characterization of Human Faces”, *J. Optical Soc. Am. A*, vol. 4, pp. 519-524, 1987.

[12] Jilin Tu, Zhenqiu Zhang, Zhihong Zeng and Thomas Huang, “Face Localization via Hierarchical CONDENSATION with Fisher Boosting Feature Selection”, *IEEE-CVPR*, Vol.2, Pg. 719-724, 2004.

[13] Turk, M., Pentland, A., “Eigenfaces for Recognition”, *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991.

[14] Paul Viola and Michael Jones, “Robust Real-time Object Detection”, *ICCV 2001 Workshop on Statistical and Computation Theories of Vision*.

[15] Lei Zhang, Stan Z. Li, ZhiYi Qu, Xiangsheng Huang. “Boosting Local Feature Based Classifiers for Face Recognition”. *First IEEE Workshop on Face Processing in Video*. June 28, 2004, Washington, D.C., USA.