ROBUST FACE RECOGNITION USING MINIMAX PROBABILITY MACHINE

Chu-Hong Hoi and Michael R. Lyu

Department of Computer Science and Engineering,
The Chinese University of Hong Kong,
Shatin, Hong Kong S.A.R., China
{chhoi,lyu}@cse.cuhk.edu.hk

ABSTRACT

Face recognition has been widely explored in the past years. A lot of techniques have been applied in various applications. Robustness and reliability become more and more important for these applications especially in security systems. In this paper, a new face recognition approach is proposed based on a state-of-the-art classification technique called Minimax Probability Machine (MPM). Engaging the binary MPM technique, we present a multi-class MPM classification for robust face recognition. In the experiments, we compare our MPM-based face recognition algorithm with other traditional techniques including Neural Network and Support Vector Machine. The experimental results show that the MPM-based face recognition technique is competitive and promising for robust face recognition.

1. INTRODUCTION

Face recognition has received more and more attentions from computer science and engineering societies. A variety of face recognition techniques, such as Eigenface, Fisherface, Elastic Graph Matching (EGM), Neural Network (NN) and Support Vector Machine (SVM), have been proposed in the past decade. These techniques have been applied in various applications particularly for security issues such as access control, identification, and authentication, etc [1, 2]. Although the existing algorithms are proved effective in laboratory, constructing robust face recognition techniques for practical applications is still an open and challenging problem.

In this paper, we present a robust face recognition method employing a new classification technique called Minimax Probability Machine (MPM) [3, 4]. Different from other classification techniques, MPM provides a lower bound on classification accuracy of future data by minimizing the worst-case probability of misclassification of the future data only with the mean and covariance matrix of the classes [3, 4]. This important feature is critical and helpful to construct robust face recognition techniques since it is able to estimate and guarantee the accuracy of face recognition applications.

The rest of the paper is organized as follows. Section 2 reviews conventional face recognition techniques. In Section 3, we first introduce the basic idea and formulation of binary MPM classification which was proposed in [3, 4]. Then we discuss the robust MPM boundary [4] and multi-class MPM classification employing the binary MPM. Section 4 provides experimental comparison between the robust MPM-based face recognition and the traditional methods. Section 5 gives our conclusion and future work.

2. ALGORITHMS REVIEW

In the literature, many face recognition techniques have been suggested in the past decade. Here we briefly review several well-known techniques including face feature extraction techniques and classification methods.

Feature extraction is an important step toward face recognition. The well-known approaches include Eigenface [5, 6], Fisherface [7] and Elastic Graph Matching [8], etc. The Eigenface approach transforms face images into a small set of feature images called "Eigenfaces" which are the principle components of the training set of face images [5]. Feature of a new image is obtained by projecting it into the subspace spanned by the eigenfaces [5]. Instead of the Principal Component Analysis (PCA) used in Eigenface, Fisherface suggests the projection method by Fisher's Linear Discriminant which chooses the projection directions that can project away variances in lighting and face expression and can also maintain the discriminability [7]. And the EGM technique is based on the dynamic link architecture, in which facial features are extracted by the Gabor-based wavelet transform [8].

For the classification techniques on face recognition, two kinds of well-known methods are Neural Network and Support Vector Machine. As a famous machine learning technique, NN has been widely investigated and its application on face recognition can be found in [9]. In recent years, SVM, a currently popular classification technique, has been shown with many successful applications in the patter recognition field [10]. SVM attempts to find the optimal decision

0-7803-8603-5/04/$20.00 ©2004 IEEE

boundary which separates the data points with a maximum margin based on the Structural Risk Minimization principle [10]. The research work in [11] demonstrated the success of SVMs on face recognition applications. Other face recognition techniques such as face recognition committee machines were also suggested in literature [1, 2].

3. ROBUST MPM FOR FACE RECOGNITION

3.1. MPM for Binary Classification

MPM is a very new classification technique proposed in [3, 4]. It enjoys competitive classification performance comparing with most of state-of-the-art classification techniques. The most attractive properties of MPM is that it can explicitly provide a worst-case bound on the probability of misclassification of future data when the mean and covariance matrix of the data are known [3]. The basic theory of MPM for binary classification is discussed as follows [3].

Assume two random vectors $\mathbf{x}$ and $\mathbf{y}$ represent two classes of data points with means and covariance matrices as $\{\bar{\mathbf{x}}, \Sigma_{\mathbf{x}}\}$ and $\{\bar{\mathbf{y}}, \Sigma_{\mathbf{y}}\}$, respectively, where $\mathbf{x}$, $\mathbf{y}$, $\bar{\mathbf{x}}$, $\bar{\mathbf{y}} \in \mathbb{R}^n$, and $\Sigma_{\mathbf{x}}$, $\Sigma_{\mathbf{y}} \in \mathbb{R}^{n\times n}$ both symmetric and positive semidefinite. Let $\mathbf{x}$ and $\mathbf{y}$ denote the corresponding class of the data, respectively.

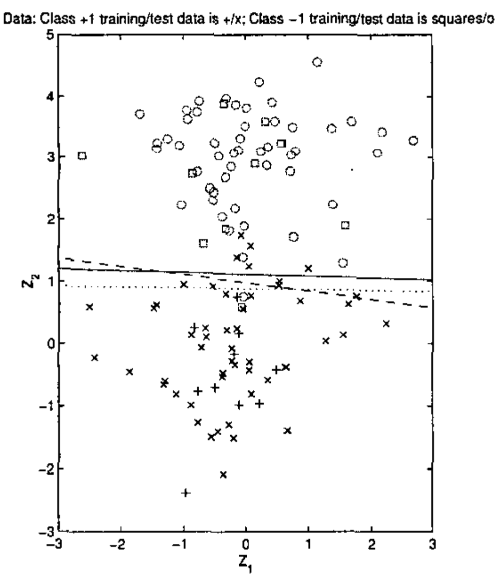

Fig. 1. Decision lines comparisons: MPM decision line (dotted red line), robust MPM (solid red line) and a linear 1-norm soft margin SVM decision line (dashed black line) [3].

Suppose that $\{\bar{\mathbf{x}}, \Sigma_{\mathbf{x}}\}$ and $\{\bar{\mathbf{y}}, \Sigma_{\mathbf{y}}\}$ obtained from the two-class data are accurate, MPM attempts to look for an optimal hyperplane

$$\mathbf{a}^T\mathbf{z} = b \quad (\mathbf{a}, \mathbf{z} \in \mathbb{R}^n, \mathbf{a} \neq \mathbf{0}, b \in \mathbb{R}) \ , \qquad (1)$$

which separates the data into two classes by minimizing the worst-case probability of misclassification of the future data. The decision boundary of MPM is illustrated in Fig. 1 given in [3]. The mathematical formulation of the original model can be written as follows [3]:

$$\max_{\alpha, b, \mathbf{a}\neq\mathbf{0}} \ \alpha \quad \text{s.t.} \quad \inf P_r\{\mathbf{a}^T\mathbf{x} \geq b\} \geq \alpha$$
$$\inf P_r\{\mathbf{a}^T\mathbf{y} \leq b\} \geq \alpha \ , \quad (2)$$

where α represents the lower bound of the accuracy for the classification of future data points, namely, the worst-case accuracy. Future points $\mathbf{z}$ for which $\mathbf{a}^T\mathbf{z} > b$ are then classified as the class $\mathbf{x}$; otherwise, they are judged as the class $\mathbf{y}$. This derived decision hyperplane is claimed to minimize the worst-case (maximum) probability of misclassification, or the error rate, for the classification of future data points.

After introducing the Lagrangian Multiplier, the optimization problem is turned into the following:

$$\max_{\kappa, \mathbf{a}} \ \kappa \quad \text{s.t.} \quad -b + \mathbf{a}^T\mathbf{x} \geq \kappa\sqrt{\mathbf{a}^T\Sigma_{\mathbf{x}}a}$$
$$b - \mathbf{a}^T\mathbf{y} \geq \kappa\sqrt{\mathbf{a}^T\Sigma_{\mathbf{y}}a} \ . \quad (3)$$

After eliminating κ, the optimization problem becomes:

$$\min_{\mathbf{a}} \sqrt{\mathbf{a}^T\Sigma_{\mathbf{y}}\mathbf{a}} + \lambda\sqrt{\mathbf{a}^T\Sigma_{\mathbf{x}}\mathbf{a}} \qquad (4)$$
$$s.t. \quad \mathbf{a}^T(\bar{x} - \bar{y}) = 1 \ . \qquad (5)$$

This is a non-linear optimization problem and can be solved by convex programming technique. Detailed descriptions of MPM can be found in [3, 4].

The major difference between MPM and other classification techniques is that MPM minimizes the worst-case (maximum) probability of misclassification of future data points and provides a lower bound estimation of classification accuracy of future data.

3.2. Robust MPM for Error Estimation

Although MPM theoretically provide a lower bound of the classification accuracy α of the future data, the actual classification accuracy in practical testing data set may be lower than α. The reason is that the initial means and covariance matrices $\{\bar{\mathbf{x}}, \Sigma_{\mathbf{x}}\}$ and $\{\bar{\mathbf{y}}, \Sigma_{\mathbf{y}}\}$ estimated from the training data set may deviate from the true means or covariance matrices. To mitigate this problem, robust minimax probability classifier was suggested to estimate the errors in the means and covariance matrices as follows [4]:

$$\mathcal{X} = \{(\bar{\mathbf{x}}, \Sigma_{\mathbf{x}}) : (\bar{\mathbf{x}} - \bar{\mathbf{x}}^0)^T\Sigma_{\mathbf{x}}^{-1}(\bar{\mathbf{x}} - \bar{\mathbf{x}}^0) \leq \nu^2,$$
$$\|\Sigma_{\mathbf{x}} - \Sigma_{\mathbf{x}}^0\|_F \leq \rho\} \ , \qquad (6)$$

$$\mathcal{Y} = \{(\bar{\mathbf{y}}, \Sigma_{\mathbf{y}}) : (\bar{\mathbf{y}} - \bar{\mathbf{y}}^0)^T \Sigma_{\mathbf{y}}^{-1} (\bar{\mathbf{y}} - \bar{\mathbf{y}}^0) \leq \nu^2, \quad \|\Sigma_{\mathbf{y}} - \Sigma_{\mathbf{y}}^0\|_F \leq \rho\} . \qquad (7)$$

Here, the error parameters $\nu \geq 0$ and $\rho \geq 0$ are fixed. The notations $\bar{x}^0$, $\Sigma_{\mathbf{x}}^0$, $\bar{y}^0$, and $\Sigma_{\mathbf{y}}^0$ stand for the initial estimation of the means and covariance matrices, respectively. Detailed formulation of robust minimax probability classifiers can be found in [4].

3.3. Multi-Class Pattern Classification

In the above subsection, we have introduced the basic theory of binary MPM classification. Toward multi-class pattern recognition, we construct the multi-class classification technique by employing the binary MPM classifiers. In general, there are two approaches for multi-class classification based on basic binary classifiers: one-against-one and one-against-all.

In this paper, we adopt the one-against-all approach since it only needs to train n binary classifiers for n given classes while the one-against-one approach needs to train $\frac{n(n-1)}{2}$ binary classifiers. The one-against-all approach means that we train n classifiers in which each MPM classifier is trained based on the data of one class against the other $(n-1)$ classes. For each trained classifier, we can obtain its lower bound estimation of classification accuracy, denoted as α_i. In the predicting phase, we first perform a sorting of the n classifiers on a descending order of α_i. Then, the test data points are predicted based on the n sorted classifiers.

4. EXPERIMENTAL RESULTS

In our experiments, eigenfaces are used to represent the images. We compare the proposed MPM-based face recognition with other traditional techniques: Eigen (typical Eigenface approach by nearest-neighbor classifiers), SVM (Support Vector Machine), and NN (Neural Network). The kernel functions for SVM and MPM are based on Polynomial functions. The robust approach of MPM is used in the experiments [4].

4.1. Evaluation on the ORL Face Database

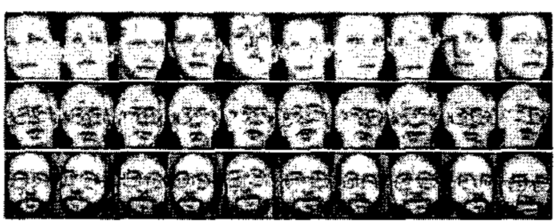

Fig. 2. A snapshot of the ORL face database for 3 people.

Subset	Eigen	SVM	NN	MPM
1	82.5%	92.5%	**97.5%**	**97.5%**
2	85.0%	**100%**	97.5%	**100%**
3	87.5%	**100%**	92.5%	92.5%
4	75.0%	**95.0%**	87.5%	**95.0%**
5	72.5%	90.0%	87.5%	**97.5%**
6	82.5%	**97.5%**	87.5%	92.5%
7	80.0%	92.5%	90.0%	**95.0%**
8	77.5%	**95.0%**	87.5%	**95.0%**
9	75.0%	**97.5%**	92.5%	**97.5%**
10	85.0%	95.0%	95.0%	**97.5%**
Avg.	80.3%	95.5%	91.5%	**96.0%**

Table 1. Experimental results on the ORL face database

The first face database for our evaluation is the ORL face database from AT&T Laboratories. It contains 400 face images, including 40 distinct persons, each with 10 faces that vary in position, rotation, scale, and different expressions. Fig. 2 shows a snapshot of the chosen face database.

In the experiment, we partition the database into 10 subsets, each contains 1 face image from each distinct person. We use the k-folder cross validation. For example, to test the 1th subset, we adopt the 2nd-7th subsets for training and 8th-10th subsets for validation. The validation is used to evaluate the performance of robust MPM and estimate the parameters [4]. 10 trials are run in the experiment. Table 1 shows the final experimental results by different methods.

From Table 1, we can see that the MPM-based face recognition technique is competitive with other state-of-the-art techniques. The MPM-based face recognition outperforms the Eigenface simply by Euclidean Distance and the NN approach. And its performance is competitive with the SVM approach.

4.2. Evaluation on the Yale Face Database

The second database adopted in our experiments is the Yale face database from Yale University. It contains 165 face images including 15 different people each with 11 images in varied expression and lighting. The face images are grayscale and cropped to a resolution of 116×136 pixels. Fig. 3 shows a snapshot of the database.

The experimental setting is similar to the ORL database. 11 trials are conducted in the experiment. Table 2 shows the experimental results on the Yale face database. We can see that MPM-based approach is better than other approaches in most cases. It is again comparable with the SVM approach.

Moreover, the important advantage of MPM-based face recognition is that it can provide a lower bound estimation of classification accuracy of future data, which makes the system robust for better recognition performance.

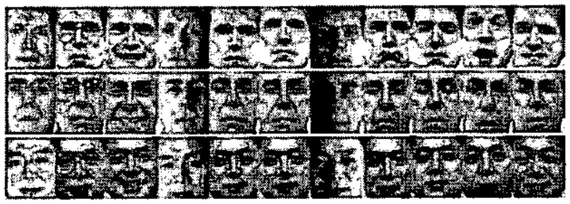

Fig. 3. A snapshot of the Yale face database for 3 People.

Subset	Eigen	SVM	NN	MPM
1: centerlight	40.0%	**93.3%**	60.0%	86.7%
2: glasses	73.3%	**86.7%**	**86.7%**	**86.7%**
3: happy	73.3%	86.7%	**93.3%**	**93.3%**
4: leftlight	26.7%	26.7%	**40.0%**	33.3%
5: noglasses	93.3%	**100%**	93.3%	93.3%
6: normal	86.7%	86.7%	**93.3%**	86.7%
7: rightlight	**26.7%**	20.0%	**26.7%**	**26.7%**
8: sad	66.7%	**93.3%**	86.7%	**93.3%**
9: sleepy	80.0%	**100%**	93.3%	93.3%
10: surprised	**73.3%**	66.7%	46.7%	**73.3%**
11: wink	93.3%	**100%**	**100%**	**100%**
Avg.	66.7%	78.2%	74.5%	**78.8%**
No lighting	75.6%	**90.0%**	86.7%	89.9%

Table 2. Experimental results on the Yale face database

5. CONCLUSION AND FUTURE WORK

In this paper, we present a robust face recognition technique employing a state-of-the-art classification technique called minimax probability machine. Different from other classification techniques, MPM provides a lower bound estimation on classification accuracy, which is important for robust classification of future data and can help for reliability evaluation. In the experiments, we show that the MPM-based face recognition is competitive with most of state-of-the-art techniques and is promising in various reliability applications.

However, an important issue is also needed to address for robust face recognition by MPM. Accurate estimation of the initial means and covariance matrices is important to impact the classification performance of MPM although robust MPM may mitigate the problem. Hence, seeking effective and reliable methods to correctly estimate the means and covariance matrices is important future work of our robust face recognition.

6. ACKNOWLEDGEMENT

Thanks G. R. G. Lanckriet for providing the Matlab source codes of Minimax Probability Machines on the web. The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. The Chinese University of Hong Kong 4182/03E).

7. REFERENCES

[1] H.-M. Tang, M. R. Lyu, and I. King, "Face recognition committee machine," in *Proc. IEEE Conf. on Acoustics, Speech, and Signal Processing*, Hong Kong, 2003, pp. 837–840.

[2] H.-M. Tang, M. R. Lyu, and I. King, "Face recognition committee machines: Dynamic vs. static structures," in *Proc. Int. Conf. on Image Analysis and Processing*, Mantova, Italy, 2003, pp. 121–126.

[3] G. R. G. Lanckriet, L. El Ghaoui, C. Bhattacharyya, and M. I. Jordan, "Minimax probability machine," in *Proc. Advances in Neural Information Processing Systems*, 2002.

[4] G. R. G. Lanckriet, L. El Ghaoui, C. Bhattacharyya, and M. I. Jordan, "A robust minimax approach to classification," *Journal of Machine Learning Research*, vol. 3, pp. 552–582, 2002.

[5] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

[6] M.A. Turk and A.P. Pentland, "Eigen faces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, 1991.

[7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[8] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. Malsburg, R. P. Würtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. on Computers*, vol. 42, pp. 300–311, 1993.

[9] J. Zhang, Y. Yong, and M. Lades, "Face recognition: eigenface, elastic matching, and neural nets," *Proc. of the IEEE*, vol. 85, pp. 1423–1435, 1997.

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

[11] P. Phillips, "Support vector machines applied to face recognition," in *Proc. Advances in Neural Information Processing Systems*, 1998.