

## A new face recognition system based on Kernel maximum between-class margin criterion (KMMC)

Minghua Wan<sup>1,2, a\*</sup>, Qiaoli Wang<sup>1,b</sup>, Hui Liu<sup>1,c</sup> and Xun Li<sup>1,d</sup>

<sup>1</sup>Key Laboratory of Nondestructive Testing (Nanchang Hangkong University), Ministry of Education, Nanchang 330063, China

<sup>2</sup>Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education (Nanjing University of Science and Technology), Nanjing 210094, China

<sup>a</sup>wmh36@sina.com, <sup>b</sup>1446252380@qq.com, <sup>c</sup>772507808@qq.com, <sup>d</sup>975978961@qq.com

**Keywords:** Face recognition; Feature extraction; Small sample size; Kernel maximum between-class margin criterion

**Abstract.** To avoid small sample problem in pattern recognition, the paper uses KMMC (Kernel maximum between-class margin criterion method) as the basic extraction method for face recognition, which is based on the maximum difference of between-class scatter and within-class scatter in feature space. The objective of KMMC is to seek an optimal set of discriminant vectors as the projection axis to do some projection transformation, and to make the between-class scatter of feature space sample maximum, the within-class scatter minimum, theoretically solved the problems that can not be solved due to singularity of within-class scatter and demonstrates its efficiency of feature extraction furthermore. The test results show the validity of this method on ORL database. At last, it designed and implemented face recognition system based on KMMC by using Matlab7.1.

### Introduction

At present, there are many methods for facial feature extraction. The dimension of face image is commonly very high, but the distribution of face images in high dimensional space is so sparse and the computational complexity is so high that it's not conducive to the classification. So, people often project the image onto a low dimensional subspace to discriminant analysis. There exists small sample problems as (Linear Discriminant Analysis, LDA) [1,2] method applied to face recognition feature extraction. Although the methods solving the small sample size problem have been proposed, however, there is no fundamental change in the presence of the small sample size problem. In recent years, Li etc. proposed a new algorithm based on maximum margin criterion (maximum margin criterion, MMC [3]) which solved the small size sample problem of LDA method fundamentally.

In 1998, Scholkopf [4] proposed the kernel principal component analysis (Kernel Principal Component Analysis, KPCA) theory and algorithm, by drawing from kernel method of SVM (Kernel Method) and applying the earliest theory to the feature extraction. Then, Mika [5], Baudat and Anouar [6] as well as Roth and Steinhage[7] put forward the kernel Fisher discriminant analysis(Kernel Fisher Discriminant Analysis, KFDA) method, using kernel methods to further extent the Fisher linear discriminant analysis to the nonlinear case from different aspects.

Therefore, this paper proposed the kernel maximum margin criterion (KMMC) based on the kernel method and the MMC algorithm, and applied it to the face recognition system. Finally, the results of the experimental in the ORL face database verify its validity.

### Maximum Margin Criterion (MMC) Method

#### The Propose of MMC

In the MMC algorithm, given N dimensional C type training samples,  $X = \{x_1, x_2, \dots, x_N\}$  a total of N samples.  $S_b$ ,  $S_w$  and  $S_t$  denote the between class scatter matrix, the within class scatter matrix and

the total scatter matrix of the training sample. By its definition,  $S_b$ 、 $S_w$  and  $S_t$  are all non negative definite matrix, and meet  $S_t = S_b + S_w$ , among them:

$$S_b = \sum_{i=1}^c l_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T, \quad S_w = \sum_{i=1}^c \sum_{j=1}^{c_i} (x_i^j - \bar{x}_i)(x_i^j - \bar{x}_i)^T \quad (1)$$

$l_i = N_i/N$  is the prior probability of class  $i$ ,  $\bar{x}$  represents the average of all samples,  $\bar{x}_i$  says the average of class  $i$  of the training sample,  $N_i$  is the number of training samples of class  $i$ ,  $N$  is the number of all samples.

The MMC criterion function is defined as follows:

$$\max J(V) = \text{tr}(S_b' - S_w') \quad (2)$$

Among them  $S_b' = V^T S_b V$ ,  $S_w' = V^T S_w V$ . It can obtain the following relationship for the constraint conditions for  $v_i^T v_i = 1$ . By using the Lagrange multiplier can obtain:

$$(S_b - S_w)v = \lambda_i v \quad (3)$$

where  $\lambda_i$  is the eigenvalue of  $S_b - S_w$ , and  $v_i$  is the corresponding eigenvector.

### Kernel maximum margin criterion algorithm (KMMC algorithm)

Suppose  $C$  is the number of pattern classes,  $X_j^i \in R^n$  is the  $j$  training samples of class  $i$  of the original input space, where  $i=1,2,\dots,C; j=1,2,\dots,N_i$  denotes the number of training samples in class  $i$ ,  $N_1+N_2+\dots+N_C=N$ . After nonlinear mapping  $\phi$  transform,  $\{\phi(X_1^1), \phi(X_2^1), \dots, \phi(X_{N_1}^1), \dots, \phi(X_1^C), \dots, \phi(X_{N_C}^C)\}$  said the  $N$  training samples of the high dimensional feature space  $H$ . The within class scatter matrix  $S_w^\phi$ , the between class scatter matrix  $S_b^\phi$  and the total scatter matrix  $S_t^\phi$  of training samples of the high dimensional feature space  $H$  are defined as

$$S_b^\phi = \sum_{i=1}^C \frac{N_i}{N} (m_i^\phi - m^\phi)(m_i^\phi - m^\phi)^T, \quad S_w^\phi = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} (\phi(X_j^i) - m_i^\phi)(\phi(X_j^i) - m_i^\phi)^T \quad (4)$$

$$\text{Among them, } S_t^\phi = S_b^\phi + S_w^\phi; \quad m^\phi = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} \phi(X_j^i); \quad m_i^\phi = \frac{1}{N_i} \sum_{j=1}^{N_i} \phi(X_j^i).$$

The maximum margin criterion (MMC) function of the feature space  $H$  is

$$\max J^\phi(V) = \sum_{j=1}^d v_j^T (S_b^\phi - S_w^\phi) v_j \quad (5)$$

Among them  $V = [v_1, v_2, \dots, v_d] \in R^{L \times d}$ ,  $v_i \in R^L$  is the any non-zero vector in the feature space.

If suppose all the  $N$  training samples of the high dimensional feature space  $H$  as  $\phi(X_1), \phi(X_2), \dots, \phi(X_N)$ , the vector  $V$  is located in the space of the  $N$  training sample vector  $\phi(X_1), \phi(X_2), \dots, \phi(X_N)$  of the  $H$  high dimensional feature space, namely

$$v = \sum_{i=1}^N a_i \phi(X_i) = \phi \alpha$$

(6)

Project the  $\phi(X_k^i)$  to v can obtain,

$$\begin{aligned} v^T a(X_k^i) &= \alpha^T \phi^T a(X_k^i) = \alpha^T \left[ \phi(X_1)^T \phi(X_k^i), \dots, \phi(X_N)^T \phi(X_k^i) \right]^T \\ &= \alpha^T \left[ k(X_1, X_k^i), \dots, k(X_N, X_k^i) \right]^T = \alpha^T \zeta_k^i \end{aligned} \quad (7)$$

**Definition 1** Set  $\zeta_k^i = \left[ k(X_1, X_k^i), \dots, k(X_N, X_k^i) \right]^T$ ,  $\zeta_k^i$  is called the kernel sample vector (corresponding to the original input sample  $X_k^i$ ), and called the corresponding vector  $\alpha = [a_1, a_2, \dots, a_N]^T \in R^N$  as kernel discriminant vectors. If project the mean vector  $m_i^\phi$  and the overall mean vector  $m^\phi$  of the feature space H to the potential optimal discriminant vector V. Then

$$w^T S_b^\phi w = \alpha^T K_b \alpha, w^T S_w^\phi w = \alpha^T K_w \alpha \quad (8)$$

Through the Eq. 8 can get the equivalent form of the maximum margin criterion Eq. 3 in feature space H

$$\max J^\phi(v) = \sum_{j=1}^d \alpha_j^T (K_b - K_w) \alpha_j \quad (9)$$

Eq. 9 is called as the kernel maximum margin criterion (KMMC) function, vector  $\alpha_i \in R^N$  is called optimal kernel discriminant vectors, by the following theorem  $\alpha_1, \alpha_2, \dots, \alpha_d$  we can determine a set of optimal kernel discriminant vector.

**Theorem 1** The optimal kernel discriminant vector kernel maximum margin criterion set  $\alpha_1, \alpha_2, \dots, \alpha_d$  from the characteristic equation of  $(K_b - K_w)X = \lambda X$  of the first d largest eigenvalues corresponding to the vector based on unit features.

## Experimental results and analysis

**The results of KMMC algorithm.** In this chapter, we carried out an experiment in ORL face database to compare the classification and recognition performance of KMMC algorithm and PCA, LDA and MMC algorithm. All the algorithms used Euclidean distance and the nearest neighbor classifier. In addition to the PCA algorithm, other algorithms adopted the PCA algorithm to do the pretreatment to get results faster in this experiment. The experiment environment: Dell PC, CPU: Inter Athlon (tm) 64 Processor, memory: 1024M, Matlab 7.01.

The ORL standard face database (<http://www.uk.research.att.com/facedatabase.html>) consists of 40 persons; each person has ten gray level images which size is 112×92. In this experiment, all the images are processed into gray level with size of 56×46. Fig. 1 is ten images of one person in ORL face database.



Fig. 1 Ten images of one person in ORL face database

The experiment randomly selected  $L$  images of each person for training, and used the remaining to do the test. The first step of PCA transformation in the PCA, LDA, MMC and KMMC can keep the image energy up to 98%, which means the dimension is reduced to 50. Each image selected would be preformed for 10 times. At last, we get the maximum average recognition rates of different algorithms and the corresponding feature dimension in Table 1.

Table 1 The maximum average recognition rates of different algorithms in ORL face database

algorithm $l$	PCA	LDA	MMC	KMMC
$l=6$	87.39(46)	89.38(38)	91.50(36)	92.42(36)
$l=5$	86.71(46)	87.23(38)	88.32(38)	89.78(36)
$l=4$	84.98(46)	86.17(38)	86.85(38)	87.21(36)
$l=3$	82.27(46)	85.09(38)	85.44(40)	86.82(36)

**System Instructions.** On the basis of the above theoretical analysis, we design and implement the face recognition system based on KMMC algorithm by Matlab7.01.

The first to introduce is the part of face detection, the interface of this part is simple which mainly processed the image in background (by adopting Adaboost algorithm to detect). Just read in an image and then detection. Results are as follows:



Fig. 2 Effect diagram of face recognition 3: Total 13/Miss 6/Error 1

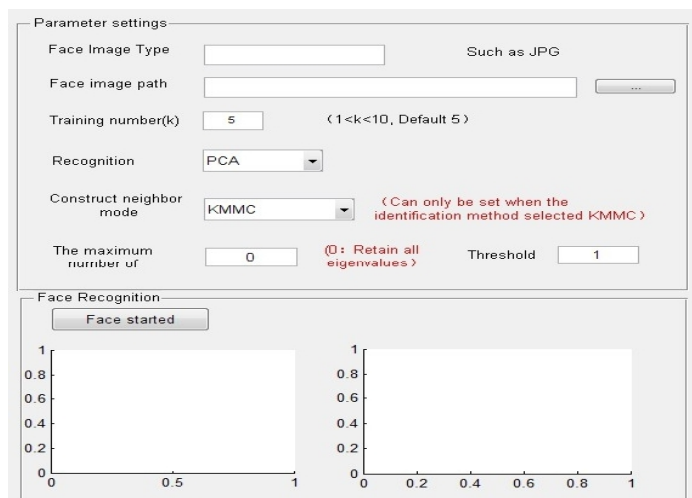


Fig.3 The interface of face recognition

The second part is the face recognition, whose interface is as Fig. 3. The first is to set the parameters: face image type is the type of face database; ace image path is the path of face database; "training

number" is the number of training samples in the face recognition algorithm, "recognition method" includes PCA,LDA,MMC and KMMC optional; "maximum number of feature" means the dimension of feature space; "threshold" is the ratio of the total information with the information contained in the selected feature vectors, the general set is 0.9. It can calculate the recognition rate of the selected face recognition algorithm after all parameters has been set.

Then we select the ORL face database, make the original image standard in both magnitude and direction, for example, the eyes were aligned in the same location. Then, we cut the face area to  $32 \times 32$  pixels to match the final image. The parameters are set as: training number is 5, recognition method is the KMMC, construct neighbor mode is Supervised, maximum number of feature is the default, the threshold is 0.9.

## Summary

KMMC algorithm can avoid the small sample problem in pattern recognition as the basic extraction method for face recognition. The experiment is in the ORL face database in which the KMMC was compared with other sub space feature extraction algorithm, such as PCA algorithm, LDA algorithm, and the results show that the KMMC algorithm can surely extract the more discriminating feature to obtain a higher recognition rate. Finally, we designed and implemented the face recognition system based on KMMC algorithm by Matlab7.01.

## Acknowledgements

This work is partially supported by the National Science Foundation of China under grant no. 61462064, 61203243, the China Postdoctoral Science Foundation under grant No. 2014T70453, 2013M530223, Jiangsu Provincial Postdoctoral Science Foundation under grant No.1301095C, the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information (Nanjing University of Science and Technology), Ministry of Education (Grant No. 30920140122006).

## References

- [1] D.L.Swets,J.Weng. Using discriminant Eigen features for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence,1996,18(8): 831-836.
- [2] H.Yu,J.Yang. A direct LDA algorithm for high-dimensional data-with application to face recognition.Pattern Recognition,2001,34(10): 2067-2070.
- [3] H.Li,T.Jiang,K.Zhang. Efficient and robust feature extraction by maximum margin criterion. IEEE Transactions on Neural Networks,2006,17 (1): 1157-1165
- [4] B.Scholkopf,S.Mika,C.Burges,P.Knirsch,K.R.Muller,Cz.Ratsch,A. Smola.Input space vs.feature space in kernel-based methods. IEEE trans.on Neural Networks,1999,10(5):1000-1017.
- [5] S.Mika,G.Rätsch,J.Weston,B.Scholkopf.A.Smola,and K.R.Muller. Constructing descriptive and discriminative non-linear features: Rayleigh coefficients in kernel feature spaces. IEEE Transaction on Pattern Analysis and Machine Intelligence,2003,25(5): 623-628.
- [6] G.Baudat,F.Anouar. Generalized discriminant analysis using a kernel approach. Neural Computation,2000,12 (10): 238s-2404.
- [7] J.V.Roth,V. Steinhage. Nonlinear discriminant analysis using kernel functions. In prioc.Of Neural Information Processing systems,Denver,Nov.1999.