

# Face Recognition Algorithm Based on Kernel Collaborative Representation

Liang Zhang, Jiwen Dong

Department of Computer Science and Technology Shandong Provincial Key Laboratory of Network based Intelligent Computing, University of Jinan, Jinan, 250022, China  
E-mail:Z1786774510@163.com

**Abstract**—Aiming at solving the problems of occlusion and illumination in face recognition, a new method of face recognition based on Kernel Principal Components Analysis (KPCA) and Collaborative Representation Classifier (CRC) is developed. The KPCA can obtain effective discriminative information and reduce the feature dimensions by extracting face's nonlinear structures features, the decisive factor. Considering the collaboration among the samples, the CRC which synthetically consider the relationship among samples is used. Experimental results demonstrate that the algorithm obtains good recognition rates and also improves the efficiency. The KCRC algorithm can effectively solve the problem of illumination and occlusion in face recognition.

**Keywords**-Face Recognition; KPCA; CRC; Illumination Problem; Occlusion Problem

## I. INTRODUCTION

Automatic face recognition (FR) is one of the most visible and challenging research focus in computer vision and pattern recognition. Although facial images have a high dimensionality, they usually lie on a lower dimensional subspace or sub-manifold [1]. Therefore, subspace learning and manifold methods have been dominantly and successfully used in FR. The classical Eigenfaces [2] and Fisherfaces [3] algorithm only consider the global scatter of training samples and they fail to reveal the essential data structures nonlinearly embedded in high dimensional space. The manifold learning [4] methods have been proposed to overcome this limitation. The success of manifold learning implies that the high dimensional face images can be sparsely represented or coded by the representative samples on the manifold. Very recently, sparse representation (SR) [5] technique is employed for robust FR. In [5], the training samples are used as the dictionary to code the input testing image as a sparse linear combination of them via  $l_1$ -norm minimization. The SR based classification (SRC) of face images is conducted by evaluating which class of training samples could minimize reconstruction error of the input testing image with sparse coding coefficients. Such a novel idea has been shown to be very effective in overcoming the problem of face occlusion, using  $l_1$ -norm minimization lead to low efficiency. SRC represents test image as the linear combination of the training samples from all classes, the use of all classes to collaboratively represent test image alleviates the problem which the small-sample-size in FR occur, especially when the number of training samples per class is small, so the collaborative representation based on classification (CRC) is proposed in [6]. The extensive

experiments prove that CRC has a very competitive classification result, with significant less complexity than SRC.

In this paper, we propose kernel CRC (KCRC) algorithm to solve the problem of face recognition. The KPCA method [7] is applied to extract the feature of the face image. KPCA method is the nonlinear promotion of Eigenfaces method. This method reduces the number of redundant feature dimensions of the image, at the same time retains the most conducive discriminative information. Finally we use CRC which considers not only local similarity between the test sample and its class but also global difference between the test sample and the other classes in kernel space, so it can solve the problems of occlusion and illumination effectively.

## II. KCRC ALGORITHM

### Feature extraction

The basic idea of KPCA method [7] is that the input linear space is projected to a high dimensional kernel space using kernel function, and then conduct principal component analysis to the kernel space, so we can obtain high-order statistical information of the face images among pixels.

We identify a grayscale image with the vector  $x \in R^m$  given by stacking its columns. We arrange the given  $n$  training images of as columns of a single matrix  $X = [x_1, \dots, x_n]$ , and let  $X_i \in R^{m \times n_i}$  denote the  $n_i$  training images of the  $i$ -th subject. Each vector  $x_i$  is projected from the input space,  $R^m$ , to a high dimensional feature space,  $R^f$ , by a nonlinear mapping function (1), we obtain the kernel training matrix  $K_x^\Phi$  which is similar to the covariance matrix in PCA [2]. Based on the idea in PCA, in kernel feature space  $R^f$  solve the eigenvalue problem as (2).

$$\kappa(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \quad (1)$$

$$M\lambda\alpha = K_x^\Phi \alpha \quad (2)$$

By solving (2), we can obtain eigenvalues  $\lambda_1 \dots \lambda_n$  and eigenvectors  $v_1 \dots v_n$ , sorting the eigenvalues in descending order and adjusting the correspond eigenvectors. According to given dimensions we extract  $t$  principle components in kernel feature space, and obtain the kernel projection space (KPS)  $W = \{\alpha_1, \dots, \alpha_t\}$ . Finally the kernel training matrix is

projected to KPS to obtain the kernel training projection coefficient, we denote it as kernel training samples  $X^\Phi$ .

For the testing images  $Y = [y_1, \dots, y_n]$ , each vector  $y_i$  is projected from the input space,  $R^m$ , to a high dimensional feature space,  $R^f$ , by a nonlinear mapping function (1), we obtain the kernel testing matrix  $K_y^\Phi$ , and then project it to the KPS to obtain the kernel testing projection coefficients, we denote it as kernel testing samples  $Y^\Phi$ .

**B. Classification and recognition**

Sparse representation [5] coding a test image  $y$  over a training images  $X$  such that  $y = X\alpha$  and  $\alpha$  is a sparse vector. The sparsity of  $\alpha$  can be measured by  $l_0$ -norm, which counts the number of nonzero in  $\alpha$ . Since the combinatorial  $l_0$ -minimization is NP-hard, the  $l_1$ -minimization, as the closest convex function to  $l_0$ -minimization, is widely employed in sparse coding:  $\min_{\alpha} \|\alpha\|_1$  s. t.  $\|y - X\alpha\|_2 \leq \epsilon$ , where  $\epsilon$  is a small constant. Although  $l_1$ -minimization is much more efficient than  $l_0$ -minimization, it is still time consuming. In fact, the term  $\|y - X\alpha\|_2$  has a high impact on the final coding results because it ensure that the given test image  $y$  can be faithfully represented by the training images  $X$ . So we use CRC [6] to conduct classification in the kernel feature space.

In SRC's discuss, they assumed that there are enough training samples for each class so that the dictionary  $X_i$  is over-complete. Unfortunately, FR is a typical sample-size problem, and  $X_i$  is under-complete in general. If we use  $X_i$  to represent  $y$ , the representation error can be big, even when  $y$  is from class  $i$ . Consequently, the classification will be unstable. Fortunately, one fact in FR is that face images of different classes share similarities. Some sample from class  $j$  may be very helpful to represent the testing sample with label  $i$ . In SRC, this "lack of samples" problem is solved by taking the face images from all the other classes as the possible samples of each class. That is, it codes the testing image  $y$  collaboratively over the dictionary of all samples  $X = [X_1, \dots, X_k]$  under the  $l_1$ -norm sparsity constraint.

In our work, we use the kernel training samples  $X^\Phi$  and kernel testing samples  $Y^\Phi$  to classify. After the collaborative representation with all classes for  $y^\Phi$ , SRC classifies  $y^\Phi$  individually. For the simplicity of analysis, let's remove the  $l_1$ -norm sparsity term, and then the representation becomes a least square problem:

$$(\hat{\alpha}) = \arg \min_{\alpha} \|y^\Phi - X^\Phi \alpha\| \quad (3)$$

The associated representation  $\hat{y}^\Phi = \sum_i X_i^\Phi \hat{\alpha}_i$  is actually the perpendicular projection of  $y^\Phi$  onto the space spanned by  $X^\Phi$ . In SRC, the reconstruction error by each class

$e_i = \|y^\Phi - X_i^\Phi \hat{\alpha}_i\|_2^2$  is used for classification. It can be readily derived that

$$e_i = \|y^\Phi - X_i^\Phi \hat{\alpha}_i\|_2^2 = \|y^\Phi - \hat{y}^\Phi\|_2^2 + \|\hat{y}^\Phi - X_i^\Phi \hat{\alpha}_i\|_2^2 \quad (4)$$

Obviously, from (4) we can see that it is the amount  $e_i = \|\hat{y}^\Phi - X_i^\Phi \hat{\alpha}_i\|_2^2$  that works for classification because  $\|y^\Phi - \hat{y}^\Phi\|_2^2$  is a constant for all classes.

Denote by  $\chi_i^\Phi = X_i^\Phi \hat{\alpha}_i$  and  $\bar{\chi}_i^\Phi = \sum_{j \neq i} X_j^\Phi \hat{\alpha}_j$ . In [6] shows geometrically the representation of  $y^\Phi$  over  $X^\Phi$ . Since  $\bar{\chi}_i^\Phi$  is parallel to  $\hat{y}^\Phi - X_i^\Phi \hat{\alpha}_i$ , we can readily have

$$\frac{\|\hat{y}^\Phi\|_2}{\sin(\chi_i^\Phi, \bar{\chi}_i^\Phi)} = \frac{\|\hat{y}^\Phi - X_i^\Phi \hat{\alpha}_i\|_2}{\sin(\hat{y}^\Phi, \chi_i^\Phi)} \quad (5)$$

Where  $(\chi_i^\Phi, \bar{\chi}_i^\Phi)$  is the angle between  $\chi_i^\Phi$  and  $\bar{\chi}_i^\Phi$ , and  $(\hat{y}^\Phi, \chi_i^\Phi)$  is the angle between  $\hat{y}^\Phi$  and  $\chi_i^\Phi$ . Finally, the representation error can be represented by (6).

$$e_i = \frac{\sin^2(\hat{y}^\Phi, \chi_i^\Phi) \|\hat{y}^\Phi\|_2^2}{\sin^2(\chi_i^\Phi, \bar{\chi}_i^\Phi)} \quad (6)$$

When we judge if belongs to class  $i$ , (6) shows that by using CR, we will not only consider if the angle between  $\hat{y}^\Phi$  and  $\chi_i^\Phi$  is small (i.e., if  $\sin(\hat{y}^\Phi, \chi_i^\Phi)$  is small), we will also consider if the angle between  $\chi_i^\Phi$  and  $\bar{\chi}_i^\Phi$  is big (i.e., if  $\sin(\chi_i^\Phi, \bar{\chi}_i^\Phi)$  is big). Such a "double checking" makes the classification more effective and robust.

In order to collaboratively represent the query sample using  $X^\Phi$  with low computational burden, we propose to use the regularized least square method (7).

$$(\hat{\alpha}) = \arg \min_{\alpha} \left\{ \|y^\Phi - X^\Phi \alpha\|_2^2 + \lambda \|\alpha\|_2^2 \right\} \quad (7)$$

Where  $\lambda$  is the regularization parameter. The role of the regularization term is twofold. First, it makes the least square solution stable, and second, it introduce a certain amount of "sparsity" to the solution  $\hat{\alpha}$ , yet this sparsity is much weaker than that by  $l_1$ -norm.

The solution of CR with regularized least square in (8) can be easily and analytically derived as

$$\hat{\alpha} = \left( X^{\Phi T} X^\Phi + \lambda \cdot I \right)^{-1} X^{\Phi T} y^\Phi \quad (8)$$

Let  $\alpha = \left( X^{\Phi T} X^\Phi + \lambda \cdot I \right)^{-1} X^{\Phi T}$ . Clearly,  $\alpha$  is independent of  $y^\Phi$  so that it can be pre-calculated as a projection matrix. Once a query sample  $y^\Phi$  comes, we can just simply projection  $y^\Phi$  onto  $\alpha$  via  $\alpha y^\Phi$ . This makes CR very fast.

C. *KCRC algorithm*

1. Given training images  $X = [x_1, \dots, x_n]$  and testing images  $Y = [y_1, \dots, y_n]$ .

2. Through the Gaussian kernel function  $\kappa(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ ,  $X$  and  $Y$  are projected to kernel feature space, respectively, we obtain the kernel training matrix  $K_x^\Phi$  and kernel testing matrix  $K_y^\Phi$ .

3. By solving (2) we can obtain eigenvalues  $\lambda_1 \dots \lambda_n$  and eigenvectors  $v_1 \dots v_n$ , sorting the eigenvalues in descending order and adjusting the correspond eigenvectors. According to given dimensions we extract  $t$  principle components in kernel feature space as the KPS:  $W = \{\alpha_1, \dots, \alpha_t\}$

4. Projecting  $K_x^\Phi$  and  $K_y^\Phi$  to the KPS, we obtain the kernel training samples  $X^\Phi$  and the kernel testing

Occlusion	10%	20%	30%	40%	50%
KCRC	100%	100%	100%	99.5%	98.9%

samples  $Y^\Phi$ .

5. Coding kernel testing sample  $y^\Phi$  over  $X^\Phi$  by  $\hat{\alpha} = \alpha y^\Phi$ , where  $\alpha = (X^{\Phi T} X^\Phi + \lambda \cdot I)^{-1} X^{\Phi T}$ .

6. Compute the regularized residual  $r_i = \|y^\Phi - X_i^\Phi \cdot \hat{\alpha}_i\|_2 / \|\hat{\alpha}_i\|_2$

7. Output the identity of  $y^\Phi$  as  $Identity(y^\Phi) = \arg \min_i \{r_i\}$ .

III. EXPERIMENTAL RESULTS

A. *Illumination problem*

In our experiments, we use the AR database [8]. A subset (with only illumination and expression change) that contains 50 male subjects and 50 female subjects was chosen from the AR dataset in our experiment. For each subject, the seven images from Session 1 were used for training, with other seven images from session 2 for testing. We use the original resolution (165×120) of these images, and don't crop them. The comparison of KPCA\_CRC and its competing method methods is given in Table I.

TABLE I. FACE RECOGNITION RATES ON THE AR DATABASE

Dim	54	120	300
NN	68.0%	70.1%	71.3%
LRC	71.0%	75.4%	76.0%
SVM	69.4%	74.5%	75.4%
SRC	83.3%	90.1%	93.3%
CRC_RLS	80.5%	90.0%	93.7%
KCRC	47.26%	95.57%	99.143%

We can see that SRC performs much better than all the other five methods in all dimensions except that KCRC is slightly worse than others when the dimension is 30. On

other dimensions, KCRC outperforms SRC and CRC\_RLS by about 6%.

B. *Occlusion problem*

1) FR with real face disguise: a subset from the AR database is used in this experiment. We conduct FR with complex disguise (disguise with different illumination and longer data acquisition interval). 300 images (3 neutral images with different illumination per subject) of non-occluded frontal views in Session 1 were used for training, while the disguised images (3 images with various illumination and sunglasses or scarf per subject per Session) in Sessions 1 and 2 for testing. Table 2 shows the result of KPCA, CRC\_CLS and KCRC. We can see that KCRC outperforms KPCA by about 14% and CRC\_RLS by about 60%.



Algorithm	KCRC	KPCA	CRC
sunglasses	99%	74.5%	30.667%
scarf	99%	75%	34%

AR database

Figure 1. The real disguised images of the

TABLE II. FACE RECOGNITION RATES ON THE AR DATABASE

2) FR with pixel corruption: a subset (with only illumination and expression change) that contains 50 male subjects and 50 female subjects was chosen from the AR dataset in our experiment. For each subject, the four images from Session 1 were used for training, with other four images from session 2 for testing. For each testing images, we replaced a certain percentage of its pixels by uniformly distributed random values within [0,255], the corrupted pixels were randomly chosen for each test image and the locations are unknown to the algorithm. Figure 2 shows the face image of AR database under the percentage of corrupted from 0 to 30%. Table 3 shows the result of recognition rates.



Figure 2. The pixel corrupted images of the AR database

TABLE III. FACE RECOGNITION RATES ON THE AR DATABASE

In this experiment, we see that KPCA and CRC\_CLS is very sensitive to occlusion, so the recognition rates very low, but KCRC is very effective.

3) FR with block occlusion: using the same subset like 2), but each test image was occluded by an unrelated image in randomly located square block from 0% to 70%. Table 4 shows the result of KCRC and KPCA.

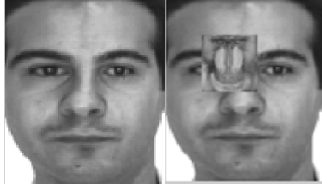


Figure 3. The block occluded images of the AR database

TABLE IV. FACE RECOGNITION RATES ON THE AR DATABASE

Occlusion	KPCA_CRC	KPCA
10%	98.25%	97.5%
20%	97.25%	96.5%
30%	95.25%	94.75%
40%	92%	93.25%
50%	87.5%	87%
60%	77%	75%
70%	42.5%	42.25%

We can see that KCRC outperforms KPCA by about 1%, when the occlusion rate lower than 40%. In our experiment, we see that CRC\_CLS is very sensitive to block occlusion.

#### IV. CONCLUSION

This paper presents KCRC algorithm, with the kernel idea, it extracts feature, and conducts classification in the kernel feature space. It proves to be robust to various types of occlusion and illumination since KCRC can extract the

most effective discriminative information for conducting classification. The experimental results demonstrate that KCRC outperform significantly previous methods, while its computational complexity is lower.

#### ACKNOWLEDGMENT

This research is supported by Natural Science Foundation of Shandong Province (ZR2010FL006).

#### REFERENCES

- [1] R. Basri and D. Jacobs, "Lambertian reflection and linear subspaces," IEEE transaction on Pattern Analysis and Machine Intelligence, vol. 25, 2003, no. 3, pp. 218-233.
- [2] M. Turk and A. Pentland, "Eigenfaces vs. Fisherfaces," Journal of Cognitive Neuroscience, vol. 3, 1991, no. 1, pp. 71-86.
- [3] P. Belhumeur, J. Hesanha and D. Kreigman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," IEEE transaction on Pattern Analysis and Machine Intelligence, vol. 19, 1997, no. 7, pp. 711-720.
- [4] J. B. Tenenbaum, V. Desilva, "A global geometric framework for nonlinear dimensionality reduction," Science, 2000, pp. 2323-2325.
- [5] J. Wright, A. Ganesh, A. Yang and Y. Ma, "Robust face recognition via sparse representation," IEEE transaction on Pattern Analysis and Machine Intelligence, 2009.
- [6] L. Zhang, M. Yang, X. Feng, "Sparse Representation or Collaborative Representation: Which Helps Face Recognition?," In International Conference on Computer Vision, 1997.
- [7] B. Schölkopf, A. J. Smola, K. R. Müller, "Kernel Principal Component Analysis," In International Conference on Artificial Neural Network, 1997.
- [8] A. Martinez and R. benavente, "The AR Face Database," CVC Tech Report No. 24, 1998.