
Kernel Collaborative Representation Classification Based on Adaptive Dictionary Learning

Zheng-ping Hu^{1,2}, Yi Liu^{2,3}, Xuan Zhang^{2,3}, Yang-hua Yin^{2,3}, Rui-xue Zhang^{2,3}, De-gang Sun¹

¹Department of Electronic Information and Engineering, Shandong Huayu University of Technology, Dezhou, China

²Department of Information and Engineering, Yanshan University, Qinhuangdao, China

³Hebei Key Laboratory of Information Transmission and Signal Processing, Qinhuangdao, China

Email address:

hzp_ysu@163.com (Zheng-ping Hu), 2360097113@qq.com (Yi Liu)

To cite this article:

Zheng-ping Hu, Yi Liu, Xuan Zhang, Yang-hua Yin, Rui-xue Zhang, De-gang Sun. Kernel Collaborative Representation Classification Based on Adaptive Dictionary Learning. *International Journal of Intelligent Information Systems*. Vol. 7, No. 2, 2018, pp. 15-22.

doi: 10.11648/j.ijis.20180702.11

Received: May 24, 2018; **Accepted:** August 30, 2018; **Published:** September 26, 2018

Abstract: In recent years, with the progress of technology, face recognition is used more and more widely in various fields. The classification algorithm based on sparse representation has made a great breakthrough in face recognition. However, face images are often affected by different poses, lighting, and expression changes, so test samples are often difficult to represent with limited original training samples. Due to the conventional dictionary learning methods lacking adaptability, we propose a kernel collaborative representation classification based on adaptive dictionary learning. In this paper, the coarse to fine sparse representation is related to the adaptive dictionary learning problem. First, the labeled atom dictionary is learned from each kind of training samples by sparse approximation. Based on this assumption, we use an efficient algorithm to generate an adaptive dictionary that is related with the test sample. Then, based on the adaptive class dictionary, the kernel collaborative representation is used to realize the inter class competition classification. The kernel function is combined with the coarse to fine sparse representation to extract the non-linear factors such as facial expression change, posture, illumination, occlusion and so on. The kernel collaborative representation is used to realize the inter class competition classification. The main advantage of this approach is to combine coarse to fine kernel collaborative representation with dictionary learning to generate adaptive dictionaries that approximate to the test image. Experimental results demonstrate that the proposed approach outperforms some previous state-of-the-art dictionary learning methods and sparse coding methods in face recognition.

Keywords: Pattern Recognition, Dictionary Learning, Kernel Space, Collaborative Representation

1. Introduction

Recently, sparse coding and dictionary learning are one of the most successful applications of pattern recognition and computer vision, and dictionary learning based pattern recognition has been widely concerned. Due to some face images are affected by different postures, illumination and expression changes [1, 2], it is usually difficult for test samples to be represented by original training samples with limited number. Based on this problem, dictionary learning can effectively use a large number of data to model the attitude, illumination, and facial expression information of the corresponding changes. Therefore, the test sample can be better represented by the atoms from the optimized dictionary.

Wright et al. [3] have demonstrated that the test image can be approximated by a sparse linear combination of the training images. The sparse coefficients corresponding to most of the training samples are zero [4]. Finally, we determine which class the test samples belong to according to the minimum reconstruction error of different classes. It tries to find the potential contribution of different training samples to test images. All the training sample images are used as the shared dictionary D . When the changes in the face image are small or the training samples are full, the shared dictionary can fully capture the main features of face image, so the dictionary can represent and classify the test sample. Sparse representation based classification algorithm has made great breakthroughs in face recognition [5-8]. However, due to various noises in the real application, training samples usually have large intra

class changes. It is usually difficult for test samples to be represented by original training samples with limited number of changes. At present, the research of face recognition based on sparse representation is mainly focused on two points: (1) different sparse constraints; (2) dictionary optimization with different constraints.

For sparse constrained problems, the sparse representation based on ℓ_0 -norm convex programming algorithm can get preferably recognition results [9]. However, minimizing ℓ_0 -norm is a NP problem with large amount of computation. The traditional sparse representation algorithm model uses ℓ_1 -norm as a convex optimization function approximating to ℓ_0 -norm to solve this kind of sparsity problem [10]. However, there are still problems with the complexity of the algorithm and the relatively high time cost. Zhang L [11] proposed a collaborative representation with the ℓ_2 -norm. Considering that the training dictionary is sufficient, the recognition performance is similar to the ℓ_1 -norm. and it has a closed analytic solution compared to the sparse representation of the L1 norm, thus it has a lower computation Complexity. Research shows that in sparse representation classification performance, inter-class collaboration has more important contribution than sparsity. The authors in [12] gave a probabilistic explanation for collaborative representation and proposed probabilistic cooperative sparse representation. Zheng J [13] proposed iterative constraint group sparse and adaptive weight learning face recognition. It obtains more structural information and discriminant information compared with other methods based on linear regression. The methods above are all based on the same basic assumption that the test samples can be represented by a linear combination of training samples. However, the original face images are interfered by the nonlinear factors such as facial expression change, posture, illumination, occlusion and so on. Thus, this linear hypothesis does not exploit the nonlinear relationship between the training samples, which may reduce face recognition robustness. To classify faces more effectively, Xu Y [14] used nonlinear functions to transform the original space to the feature space where the test image can be approximated by a sparse linear combination of the training images. Based on this assumption, kernel method is combined with sparse representation [15, 16] and low rank representation [17, 18] to solve classification and approximation problems. Dong Wang et al. proposed the Kernel Collaborative Representation (KCR) [19] to select face function and combine Hamming Kernel and LBP features for face recognition, and the experimental results show that the method has a good recognition effect especially when the training sample is not sufficient.

Based on the different dictionary optimization problems, Rubinstein R [20] described how to use mathematics and learning models to obtain the evolution of the dictionary, which showed the importance of the dictionary learning to the sparse representation model. Besides, Xu Y [21] provided a comprehensive overview of face recognition dictionary learning methods. To solve the occlusion problem and improve the robustness of face recognition, Yang M [22] used Gabor features to compress the dictionary, and proposed a

dictionary model with the sparse constraint framework. Aiming at the single sample problem, Wei C [23] proposed the auxiliary dictionary learning algorithm to expand the original dictionary and thus improve the classification performance. Hu Y S [24] used D-KSVD dictionary learning method to get a discriminative structured dictionary. In [25], a domain adaptive dictionary learning algorithm was proposed to expand the intra class diversity of the original training samples by collaboration with the source data to solve the problem of visual image classification in different source domains. Xu et al. [21] first selected train samples that are near to the test samples. Then the test image is estimated by the linear combination of the selected training samples, and the recognized result is obtained. The experiments show that the simple sparse representation face recognition algorithm (NTS) [26] can also obtain better performance. The above algorithm indicated that the quality of dictionary affected the performance of image classification.

Besides the above, face images are affected by different postures, lighting and expression changes. It is difficult for test samples to be represented by limited original training samples. For each test image, the best representation dictionary may be different. Considering the conventional dictionary learning methods suffer from the problems of lacking adaptability, we propose to construct an adaptive dictionary associated with the test sample. The labeled atom dictionary is learned from each kind of training samples by sparse approximation. Based on this scheme, we could obtain an efficient algorithm to generate an adaptive dictionary which related with the test sample. Secondly, the coarse to fine sparse representation is related to the adaptive dictionary learning problem. We fully extracted the non-linear factors such as facial expression changes, posture, lighting, and occlusion that existed in the face image. The kernel collaborative representation is used to realize the inter class competition classification.

2. Kernel Collaborative Representation Classification Based on Adaptive Dictionary Learning

2.1. Sparse Representation Based Classification

Suppose that there are L individuals and each sample can be represented as a column vector. Let $D = [A_1, A_2, \dots, A_L] \in R^{m \times n}$ be a set of L individuals, $A_i = [a_{i1}, a_{i2}, \dots, a_{in_i}] \in R^{m \times n_i}$ is a set of i th individual, where $i = 1, 2, \dots, L$, $a_{ij} \in R^m$ is the j th sample of the i th training sample and m is the dimension of the training sample. n_i is the number of i th training sample. n is the total number of samples.

The purpose of sparse representation classification is that when a training sample set is given, it can correctly identify which category the test sample belongs to. Since we can't

judge the label of test sample y , the test sample y can be represented by a linear combination of all training samples D as:

$$y = Dx \in R^m \quad (1)$$

x is sparse coefficient vector, and sparse coefficient is obtained by solving linear equation. Where $x = [0, \dots, 0, a_{i_1}, \dots, a_{i_n}, 0, \dots, 0]^T \in R^n$. The non-zero coefficient in x correspond that those associated with the i th individual. In face recognition, Face images usually satisfy $m < n$, so the linear equation $y = Dx$ is underdetermined. The solution of the equation is not unique. Sparse vectors only have a small number of elements nonzero, so the solutions are sparse. In order to solve the problem, the problem of norm optimization l_0 is adopted:

$$(\ell_0): \hat{x} = \arg \min \|x\|_0 \quad \text{s.t. } Dx = y \quad (2)$$

Among them, l_0 norm is the number of nonzero elements of a coefficient vector. However, Eq (2) is a NP-hard problem. It is difficult to accurately solve. The existing optimization theory shows that if x sufficient and sparse, the norm l_0 optimization problem can be transformed into a norm l_1 optimization problem:

$$(\ell_1): \hat{x} = \arg \min \|x\|_1 \quad \text{s.t. } Dx = y \quad (3)$$

However, the face images collected in reality are inevitably affected by noise. It is usually difficult to accurately represent test samples. Thus, the existence of error is permitted and the limit of error tolerance is defined as \mathcal{E} . Thus, Eq. (3) can be revised as:

$$(\ell_1): \hat{x} = \arg \min \|x\|_1 \quad \text{s.t. } \|Dx - y\|_2 \leq \mathcal{E} \quad (4)$$

Finally, the test sample y is classified into the minimum residual:

$$\text{identity}(y) = \arg \min \|y - A_i \delta_i(\hat{x}_i)\|_2 \quad (5)$$

The authors [4] proposed a classification algorithm based on collaborative representation. It was proved that the role of collaboration between classes in representing the query sample is more important than the sparse constraints. Compared with the l_1 -regularized sparse representation based classification (SRC), the l_2 -regularized CRC_RLS has very competitive FR accuracy but with significantly lower complexity. Here $\hat{\rho}$ can be obtained:

$$\hat{\rho} = \arg \min_{\rho} \left\{ \|y - D\rho\|_2^2 + \lambda \|\rho\|_2^2 \right\} \quad (6)$$

where λ is the regularization parameter. The role of the regularization term is twofold. On the one hand, it guarantees

the stability of $\hat{\rho}$, and then guarantees the sparsity of the coefficient solution. The solution of CR with regularized least square in Eq. (6) can be easily and analytically derived as:

$$\hat{\rho} = (D^T D + \lambda \cdot I)^{-1} D^T y \quad (7)$$

Let $P = (D^T D + \lambda \cdot I)^{-1} D^T$, obviously, P is independent of query y . So it can be pre-calculated as a projection matrix P . Once a query sample y comes, we project y onto P via Py . The proposed CRC with regularized least square (CRC_RLS) algorithm is summarized as follows.

- (1) Normalize the columns of D to have unit l_2 -norm.
- (2) Code y over D by

$$\hat{\rho} = Py, \quad (8)$$

Where $P = (D^T D + \lambda I)^{-1} D^T$

- (3) Compute the regularized residuals

$$r_i = \|y - A_i \cdot \hat{\rho}_i\|_2 / \|\hat{\rho}_i\|_2 \quad (9)$$

- (4) Output the identity of y as

$$\text{Identity}(y) = \arg \min_i \{r_i\} \quad (10)$$

2.2. Kernel Collaborative Representation

Kernel collaborative representation (KCR) face recognition algorithm first uses the kernel function to transform the original space to the feature space. Then KCR represents the test samples through a linear combination of all training samples in the feature space, and classifies the test samples based on the representation results. Let $[A_1, \dots, A_n]$ denote n training samples in the original space. Let Y be a test sample in the original space. The nonlinear mapping function ϕ is used to transform samples from original space to kernel space. Suppose that test samples $\phi(Y)$ in feature space can be approximately by a linear represented of training samples $[\phi(A_1), \dots, \phi(A_n)]$ in feature space:

$$\phi(Y) = \sum_{i=1}^n \beta_i \phi(A_i) \quad (11)$$

Suppose that any sample in feature space is a column vector, we can rewrite Eq. (11) into the following equation:

$$\phi(Y) = \Phi \Psi \quad (12)$$

Where $\Phi = [\phi(A_1) \dots \phi(A_n)]$, $\Psi = (\beta_1 \dots \beta_n)^T$, since Φ may not be a square matrix and ϕ is unknown, we cannot directly solve Eq. (12). But, Eq. (12) can be expressed as follows:

$$\Phi^T \phi(Y) = \Phi^T \Phi \Psi \quad (13)$$

The kernel function is defined as $k(A_i, A_j) = \phi^T(A_i)\phi(A_j)$, Eq. (13) can be converted into:

$$K_Y = K\Psi \quad (14)$$

$$\text{Where } K_Y = \begin{pmatrix} k(A_1, Y) \\ \vdots \\ k(A_n, Y) \end{pmatrix}, K = \begin{pmatrix} k(A_1, A_1) \dots k(A_1, A_n) \\ \vdots \\ k(A_n, A_1) \dots k(A_n, A_n) \end{pmatrix}, \text{ and}$$

$$\Psi = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$$

If K is a nonsingular matrix, Eq. (14) can be solved as follows:

$$\Psi = K^{-1}K_Y \quad (15)$$

If K is singular, Eq. (14) can be solved as:

$$\Psi = (K + \mu I)^{-1}K_Y \quad (16)$$

Where μ is a positive constant and I is the identity matrix. Eq. (16) can be represented as:

$$K_Y = (K_1 \dots K_n)\Psi = \beta_1 K_1 + \dots + \beta_n K_n \quad (17)$$

Where $K_i = [k(A_1, A_i) \ k(A_2, A_i) \ \dots \ k(A_n, A_i)]^T$. This indicates that the representing problem of test samples in the feature space has been converted into a new problem of representing K_Y by K_i , K_i represents the i th column of the matrix K . We refer to K_i as kernel vector of the i th training sample. It seems that different classes of training samples make different contributions to representing K_Y . We evaluate the contribution of each class and classify the test sample as follows: first, we calculate the sum of the contribution of the training samples from each class. The training samples from the k th class are represented as $A_1 \dots A_t$. Thus the test sample is linearly represented by the k th training samples $g_k = \beta_1 K_1 + \dots + \beta_t K_t$. The smaller the error $e_k = \|K_Y - g_k\|^2$ is, the greater contribution of the k th class is. We identify the class that contributes the most to the test sample Y (that is, the class that corresponds to the minimum error) and assign Y to this class. This equivalent to classify the test sample into the class that is the most similar to this category, because the smallest residual e_k means that the linear combination of the training samples of class k is closest to the test sample.

2.3. Kernel Collaborative Representation Classification Based on Adaptive Dictionary Learning

The original sparse representation uses all the training sample sets as a dictionary, the dictionary selection is not flexible enough, and the time complexity is high. Aiming at the problem of dictionary optimization, a kernel collaborative

representation and classification based on adaptive dictionary learning is proposed. Fig. 1 shows the flow chart of kernel collaborative representation and classification based on adaptive dictionary learning. The method mainly includes two steps. Suppose there are L categories. First, the labeled atom dictionary is learned from each kind of training samples by coefficient approximation. Based on this scheme, we could obtain L fitting images as adaptive dictionary which related with the test sample. Secondly, the test sample and the adaptive class dictionary D are projected to the kernel space, and the test samples are classified by the linear combination of the label adaptive dictionary in the kernel space. Each training sample and test sample was converted into column vectors.

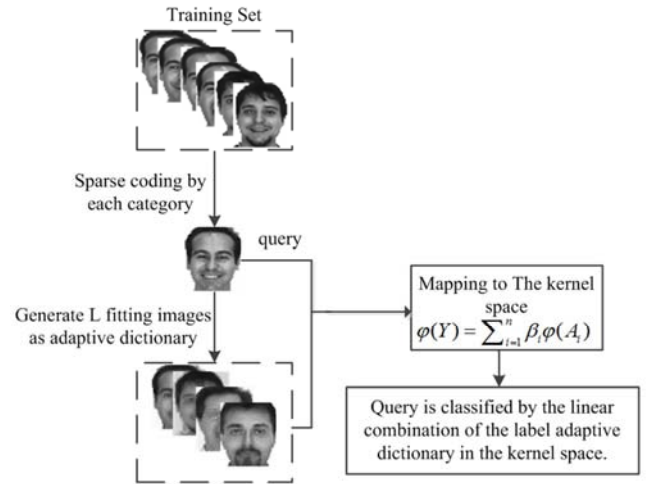


Figure 1. Flow chart for classification algorithm of kernel sparse representation based on adaptive dictionary learning.

Define $A = [A_1, A_2, \dots, A_L]$ as a set of L individuals, $A_i = [a_{i,1}, a_{i,2}, \dots, a_{i,n_i}] \in R^{m \times n_i}$ is a set of i th individual, Where M is the sample dimension and n_i is the number of training samples in each class.

The steps of the kernel collaborative representation classification based on adaptive dictionary learning are as follows:

(1) We code test sample y through each class of training samples A_i , $\hat{\rho}_i = Py$, where, $P = (A_i^T A_i + \lambda I)^{-1} A_i^T$

(2) We perform sparse approximation of test samples through each type of training sample $d_i = A_i \cdot \hat{\rho}_i$, where, $i = 1, 2, \dots, L$.

(3) Using sparsely approximated images as adaptive dictionary D , $D = [d_1, d_2, \dots, d_L] \in R^{m \times L}$.

(4) Map the test sample y and the adaptive class dictionary D to the kernel space and solve the linear system: $K_Y = K\Psi$

Where,

$$K_Y = \begin{pmatrix} k(d_1, y) \\ \vdots \\ k(d_L, y) \end{pmatrix} \Psi = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_L \end{pmatrix} \quad (18)$$

$$K = \begin{pmatrix} k(d_1, d_1) \dots k(d_1, d_L) \\ \vdots \\ k(d_L, d_1) \dots k(d_L, d_L) \end{pmatrix}$$

If K is a nonsingular matrix, it can be solved as follows:

$$\Psi = K^{-1}K_Y$$

If K is singular, it can be solved by using:

$$\Psi = (K + \mu I)^{-1}K_Y$$

(5) The test sample is represented by the class k dictionary d_k .

$$g_k = \beta_k K_k \quad (19)$$

where

$$K_k = [k(d_1, d_k) \ k(d_2, d_k) \ \dots \ k(d_L, d_k)]^T \quad (20)$$

(5) Classifying test samples using residuals:

$$\text{identity}(y) = \arg \min_k \{e_k\}, \quad e_k = \|K_Y - g_k\|^2 \quad (21)$$

The kernel space classification method has two advantages: (1) it proposes a kernel linear system and uses it to classify test samples; (2) the linear system identification method has a lower time complexity. Suppose m is the number of training samples for each class, there are L individuals, $n = Lm$ is the total number of training samples. The kernel

representation method should solve only one linear system in the form of Eq. (11), the time complexity is $O(n^3 + n^2)$. The l_1 sparse representation uses an iterative method to solve the linear system, which has a higher time complexity. Even if we do not consider the iterative process and only solve the linear system, the time complexity of the l_1 sparse representation algorithm is $O(n^3 + n^2M + nM)$, where M is the dimension of the sample vector in the original space. Thus, the kernel space representation method has a lower time complexity than the l_1 sparse representation method.

3. Experimental Results

In order to verify the effectiveness and stability of the proposed algorithm, We adopted the Gaussian kernel function $k(x_i, x_j) = \exp[-\|x_i - x_j\|^2 / (2\sigma)]$, here σ is the parameter of the kernel function. We perform the experiments on the AR [27], FERET [28], and GT [29] face databases. To verify the effectiveness of our approach, ADL, NTS [26], SRC were selected as benchmarks. Additionally, all the comparison experiments are conducted on a PC with MATLAB R2014a software.

3.1. Experiment on AR Face Database

The AR database contains 126 people and each person has 26 images. In this paper, 50 women and 50 males were selected from the AR face database, each of which contains 14 images, including changes in hairstyle, expression, and illumination. The size of facial image is 40×50 and all the images are down-sampled to 20×25 pixels. Fig. 2 is an adaptive class dictionary obtained by fitting a training sample on an AR face database with 8 training samples per class.

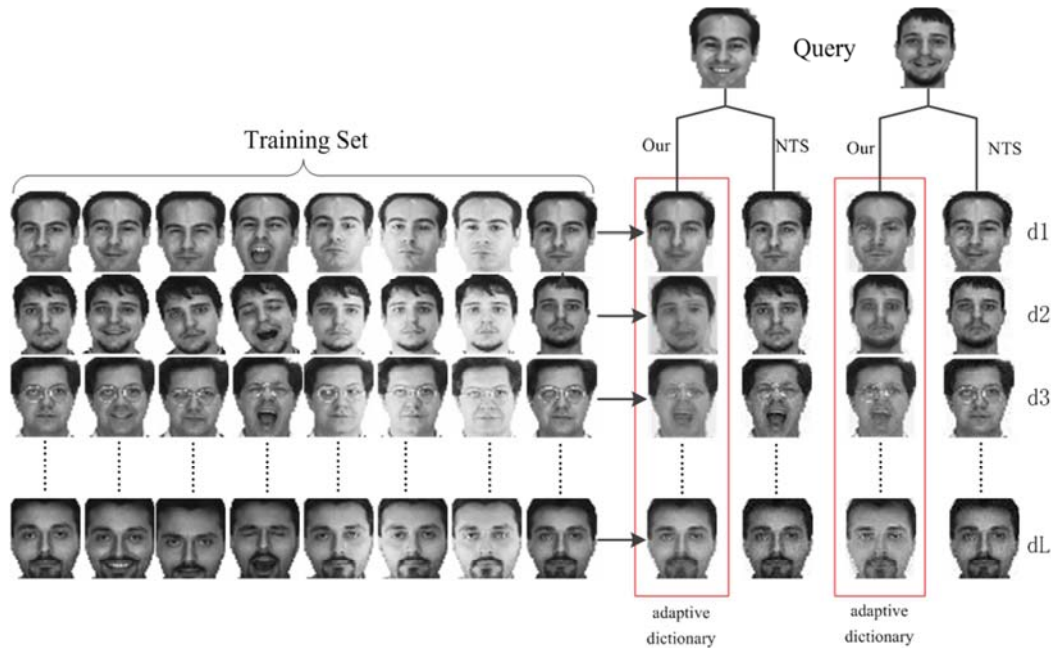


Figure 2. AR database training samples fitted to adaptive class dictionary.

In this subspace, N ($=2, 4, 6, 8, 10$) images were selected as training samples in each class, and the rest images were taken as test samples. Table 1 shows the recognition rates of kernel collaborative representation classification based on adaptive dictionary learning (KADL), the sparse representation classification based on adaptive dictionary learning (ADL), the simple fast sparse representation (NTS) and the sparse representation classification algorithm (SRC) with the number of different training samples, respectively. From Table 1, the recognition rate improved with the increase of training samples since the number of training samples increases, discriminating information increases. In addition, our method fits each type of training sample and obtains more global information than the NTS method, and has a higher recognition rate than SRC and ADL.

Table 1. Comparison of recognition rate of several methods on AR databases.

Algorithms	2	4	6	8	10
NTS	75.25%	74.9%	76.88%	91.67%	91.75%
SRC	78%	76.7%	80%	95.33%	97.25%
ADL	77.08%	77.9%	80%	98%	99.75%
KADL	77.17%	77.9%	82.25%	98.17%	99.75%

We take the effect of image dimensions into account and performed the experiments with different sample dimensions to demonstrate the robustness of the algorithm.

To analyze the impact of dimensionalities on the methods, experiments were performed with different sample dimensions to demonstrate the robustness of our algorithm. Figure 3 shows the correct recognition rate change curve for different recognition algorithms when the dimension of image features changes. It can be seen that the recognition rate of the proposed KADL approach is higher than other algorithms as a whole.

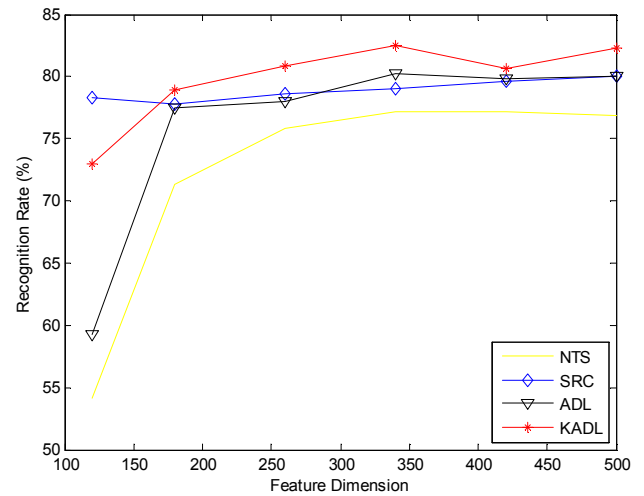


Figure 3. Recognition rate curves of different recognition methods under different dimensions.

3.2. Experiment on FERET face Database

The experiment was simulated on the FERET database. The FERET dataset contains a large number of face images, and the images from the same person's images have different ages, poses, expressions, and lighting changes. This experiment uses a cut-out FERET face database, which includes 200 individuals, each with 7 face images for a total of 1400 images. The size of all images in the database is 80×80 . All the images are resized to 32×32 . In each category, 5 images were randomly selected as training samples, and the remaining 2 samples were used as test samples. Figure 4 shows an adaptive class dictionary image obtained by fitting a training sample on a FERET face database.

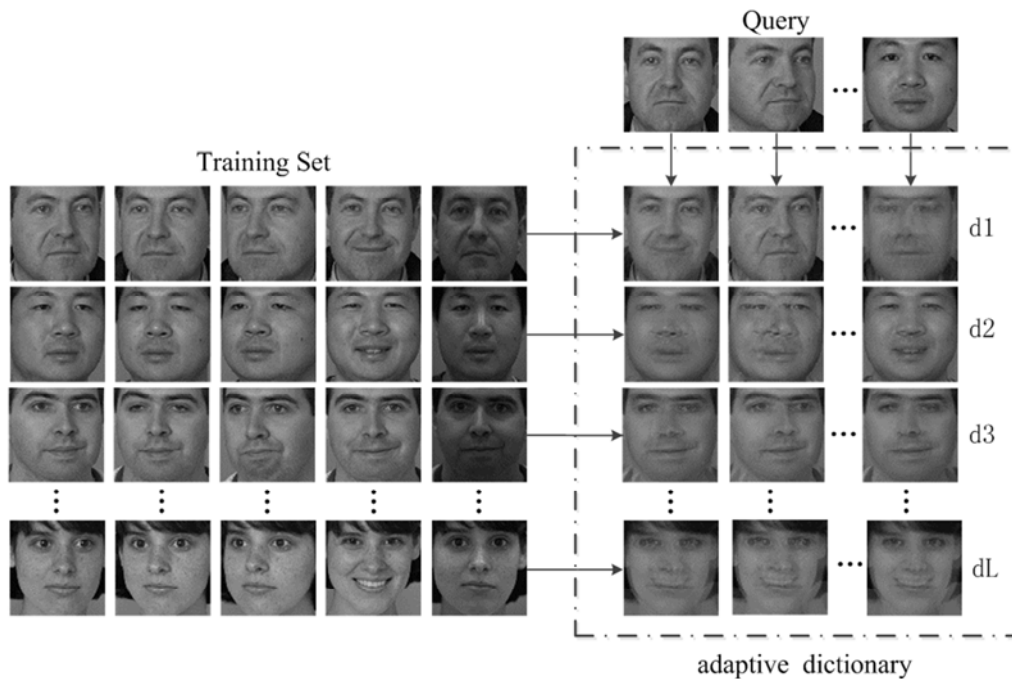


Figure 4. FERET database training samples fitted to adaptive class dictionary.

Table 2 shows the recognition rate of KADL, ADL, NTS and SRC in the FERET database.

Table 2. Recognition accuracy on FERET face database for several algorithms.

Algorithms	NTS	SRC	ADL	KADL
Recognition rate	62%	71%	70.75%	83.75%

In order to further verify the validity of the KADL algorithm, experiments are performed under different feature dimensions. Figure 5 shows the recognition rate curves of each algorithm when the feature dimensions of the image are different. From the data, we can see that the classification effect of our proposed algorithm is better than other methods.

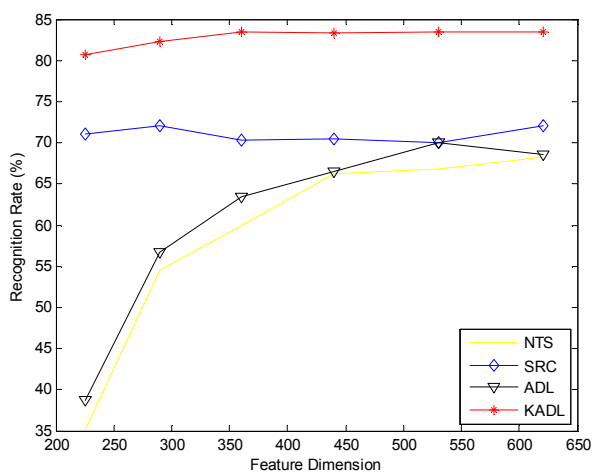


Figure 5. Recognition rate curves of different recognition methods under different dimensions.

3.3. Experiment on GT face Database

This section uses the GT face dataset to test the proposed approach. The GT dataset contains 50 persons and each person has 15 images. Some sample images in GT database are shown in Figure 6.



Figure 6. Partial sample schematic in the GT database.

In the experiment, $N (= 7, 8, 9, 10)$ images were randomly selected as training samples and the rest were used as test samples. All the images are down-sampled to 32×32 . Table 3 shows the recognition rates of KADL, ADL, NTS and SRC with different number of training samples.

Table 3. Recognition accuracy on GT face database for several algorithms.

Algorithms	7	8	9	10
NTS	61.75%	66.86%	67.67%	68.4%
SRC	68%	70.85%	73.66%	74.4%
ADL	68.25%	71.14%	73%	76%
KADL	69.5%	71.71%	73.33%	76.4%

In order to further verify the performance of KADL algorithm, we also performed experiments on different dimensions. Figure 7 shows the recognition rate curves of different methods under different dimensions. From the Figure 7, we can see that the results with the proposed KADL approach are generally higher than other methods.

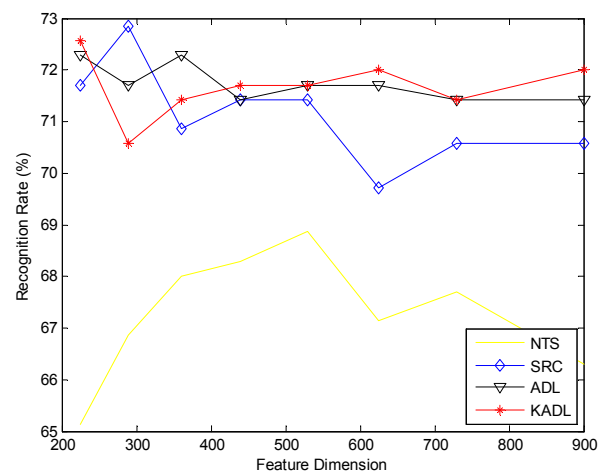


Figure 7. Recognition rate curves of different recognition methods under different dimensions.

4. Conclusions

In order to overcome the shortcomings of traditional dictionary learning mode, such as fixity and lacking of adaptability, this paper discusses adaptive dictionary learning. For each test image, a proper dictionary model is constructed according to its own characteristic, and an adaptive dictionary approximates the input pattern is generated. In order to fully extract the nonlinear factors such as facial expression change, posture, illumination and occlusion in face images, test sample and adaptive dictionary are mapped to high-dimensional feature space using kernel function, we classify the test sample in the kernel space instead of the original space. The problem of face recognition is solved by combining coarse and fine two step sparse representation with adaptive dictionary learning. A series of experiments are carried out in the AR face database, the FERET face database and the GT face database, which prove the validity of the kernel collaborative representation classification based on adaptive dictionary learning.

Acknowledgements

This work is supported by National Natural Science Foundation of China under Grant (No. 61071199), Natural

Science Foundation of Hebei Province of China under Grant (No. F2016203422) and Hebei key laboratory of information transmission and signal processing. The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] S. Zhao, Z. Hu, A modular weighted sparse representation on Fisher discriminant and sparse residual for face recognition with occlusion, *Inform. Process. Lett.* 115 (9) (2015) 677–683.
- [2] Moeini A, Moeini H. Real-World and Rapid Face Recognition Toward Pose and Expression Variations via Feature Library Matrix [J]. *IEEE Transactions on Information Forensics & Security*, 2015, 10(5):969-984.
- [3] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2009, 31(2): 210-227.
- [4] Zhang Z, Xu Y, Yang J, et al. A survey of sparse representation: algorithms and applications [J]. *IEEE access*, 2015, 3: 490-530.
- [5] He R, Zheng W S, Hu B G, et al. A regularized correntropy framework for robust pattern recognition [J]. *Neural Computation*, 2011, 23(8):2074-2100.
- [6] Yang M, Zhang L, Yang J, et al. Robust Sparse Coding for Face Recognition [C]// *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, United states, 2011:625-632.
- [7] Y.-H. Shiao, C.-C. Chen, A sparse representation method with maximum probability of partial ranking for face recognition, in: *Proc. Int. Conf. 19th IEEE International Conference on Image Processing*, Orlando, USA, October 2012: 1445–1448.
- [8] Lu C Y, Min H, Gui J, et al. Face recognition via Weighted Sparse Representation [J]. *Journal of Visual Communication & Image Representation*, 2013, 24(2):111-116.
- [9] Hyder M M, Mahata K. An Improved Smoothed l0 Approximation Algorithm for Sparse Representation [J]. *IEEE Transactions on Signal Processing*, 2010, 58(4): 2194-2205.
- [10] Yang A Y, Zhou Z, Balasubramanian A G, et al. Fast l1-Minimization Algorithms for Robust Face Recognition [J]. *IEEE Transactions on Image Processing*, 2013, 22(8): 3234-3246.
- [11] Zhang L, Yang M, Feng X. Sparse Representation or Collaborative Representation: Which Helps Face Recognition [C]// *International Conference on Computer Vision (ICCV)*, Barcelona, Spain, 2011:471-478.
- [12] Cai S, Zhang L, Zuo W, et al. A probabilistic collaborative representation based approach for pattern classification [C]// *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, United States, 2016: 2950-2959.
- [13] Zheng J, Yang P, Chen S, et al. Iterative re-constrained group sparse face recognition with adaptive weights learning [J]. *IEEE Transactions on Image Processing*, 2017, 26(5): 2408-2423.
- [14] Xu Y, Fan Z, Zhu Q. Feature space-based human face image representation and recognition [J]. *Optical Engineering*, 2012, 51(1): 017205-1-017205-7.
- [15] Zare T, Sadeghi M T. Feature level and decision level fusion in kernel Sparse Representation based Classifiers [C]// *Telecommunications (IST)*, Tehran, Iran, 2016: 462-467.
- [16] Goswami G, Singh R, Vatsa M, et al. Kernel group sparse representation based classifier for multimodal biometrics [C]// *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, USA, 2017: 2894-2901.
- [17] Nguyen H, Yang W, Shen F, et al. Kernel low-rank representation for face recognition [J]. *Neurocomputing*, 2015, 155(9): 32-42.
- [18] Xiao S, Tan M, Xu D, et al. Robust Kernel Low-Rank Representation. [J]. *IEEE Transactions on Neural Networks & Learning Systems*, 2016, 27(11):2268-2281.
- [19] Wang D, Lu H, Yang M H. Kernel collaborative face recognition [J]. *Pattern Recognition*, 2015, 48(10): 3025-3037.
- [20] Rubinstein R, Bruckstein A M, Elad M. Dictionaries for sparse representation modeling [J]. *Proceedings of the IEEE*, 2010, 98(6): 1045-1057.
- [21] Xu Y, Li Z, Yang J, et al. A survey of dictionary learning algorithms for face recognition [J]. *IEEE access*, 2017, 5(5): 8502-8514.
- [22] Yang M, Zhang L. Gabor feature based sparse representation for face recognition with gabor occlusion dictionary [C]// *European conference on computer vision*. Springer, Berlin, Heidelberg, 2010: 448-461.
- [23] Wei C P, Wang Y C F. Undersampled face recognition via robust auxiliary dictionary learning [J]. *IEEE Transactions on Image Processing*, 2015, 24(6): 1722-1734.
- [24] Hu Y S, Li J, Yang Z, et al. Joint dimensionality reduction for face recognition based on D-KSVD [C]// *Machine Learning and Cybernetics (ICMLC)*, 2016 International Conference on. IEEE, Hongkong, Hong Kong, 2016: 483-488.
- [25] Zhu F, Shao L. Weakly-supervised cross-domain dictionary learning for visual recognition [J]. *International Journal of Computer Vision*, 2014, 109(1-2): 42-59.
- [26] Xu Y, Zhu Q. A simple and fast representation-based face recognition method [J]. *Neural Computing and Applications*, 2013, 22(7-8): 1543-1549.
- [27] CVC Technical Report, The AR Face Database, The Ohio State University, 1998.
- [28] Phillips P J, Moon H, Rauss P, et al. The FERET September 1996 database and evaluation procedure [J]. *Lecture Notes in Computer Science*, 1997, 1206:395-402.
- [29] Bhadra A, Fischer G W. A new GT classification approach: A data base with graphical dimensions [J]. *MANUF. REV.*, 1988, 1(1): 44-49.