International Journal of Multimedia and Ubiquitous Engineering
Vol. 9, No. 10 (2014), pp. 137-144
http://dx.doi.org/10.14257/ijmue.2014.9.10.14

Face Recognition Algorithms Based on Orthogonal Sparse Preserving Projections of Kernel

Kezheng Lin*, Di Wu, Youhu Rong and Liangwei Zhuang

Department of Computer Science and Technology, Harbin University of Science and Technology, Harbin, 150080, China
Email: link@hrbust.edu.cn, wudi1989wudi@126.com

Abstract

According to the error approximation problem the sparse preserving projections (SPP) reconstruct the original sample. This paper proposes the algorithms based on orthogonal sparse preserving projections of kernel. In order to get sparse representation coefficients that contain more identification information by kernel method, it mapped samples to high-dimensional feature space to. Then, reconstructing sparse coefficient of kernel sparse representation increase the similar non neighbor sample weight, and reduce heterogeneous neighbor sample weight. Finally, the whole orthogonal constraint transformation improve the ability of sparse retain sample. The algorithm experiments were carried out on the YALE_B and ORL face database, and the recognition rate reached 96.3%, and the results verify the effectiveness and robustness of the algorithm.

Keywords: *kernel methods; sparse preserving projections; sparse coefficients representation; orthogonal constraint transformation*

1. Introduction

Face recognition as a challenging problem in the field of multidisciplinary field, it covers the image processing, pattern recognition, computer vision, psychology and other fields, and it is widely used in national security, military with an important theoretical research value and broad application prospects.

The advantages and disadvantages of face recognition algorithm mainly depend on the face feature extraction and classification method. In face feature extraction and classification method, the subspace learning method is used more widely. Classic subspace learning methods include Principal Component Analysis (PCA) and Linear Discriminant Analysis method (LDA). Many scholars have improved the subspace algorithm and obtained some achievements. In recent years, the manifold learning method has become a hot research object. But unsupervised algorithm existing problems on the choice of neighbor sample size. Local Preserving Projections (LPP)[1] , neighborhood Preserving Embedding (NPE)[2] and other methods as a linear improvement, it retain the local structure of image better through the adjacency graph matrix and will projection face to manifold structure to reflect the nature, achieved good results. A growing number of manifold learning method [3-4] constantly emerging, although these algorithms make full use of the local structure of sample information, but it is not strong robustness under the factors such as illumination, expression, posture. Therefore scholars proposed a classification method based on Sparse Representation (SRC)[5], using image constitute a complete dictionary of training sample set, and then use the dictionary to Sparse Representation of unknown test samples, so as to estimate the test sample category. The SRC method has good robustness, especially for image with shade, the recognition rate is superior to other algorithms. Recently Sparsity Preserving Projections

ISSN: 1975-0080 IJMUE

(SPP)[6] is proposed, it sparse reconstruction data through sparse representation, not only using the natural discriminant ability of sparse representation effectively, but also reducing the difficulty of neighbor parameter selection. However, SPP is a completely unsupervised algorithm, if you can't make good use of the local structure information of same sample and heterogeneous samples, sparse reconstruction sample easily lead to sample error approximation, recognition rate decline.

For the above problems, based on SPP and other algorithms [7-10], this paper proposes a face recognition algorithm based on orthogonal sparse preserving projections of kernel, through kernel mapping method [11-12] to get a more effective coefficient of sparse representation, updating homogeneous and heterogeneous sample weights when reconstructing coefficients, orthogonal constraint transformation to the whole to avoid the sample error approximation, so that improving the recognition rate.

2. Sparse Preserving Projections

The purpose of SPP algorithm is to reserve the sparse reconstructing relational of data. The projection matrix has invariance about data rotation and scale, although SPP is unsupervised method, but it contains identifying information.

SPP weight matrix is based on the improved framework of sparse representation. $\{x_i\}_1^n$ is a set of training samples and $x_i \in R^m$. The matrix of the training sample set is expressed as $X = [x_1, x_2, \dots, x_n] \in R^{m \times n}$. We hope to reconstruct x_i with less samples as possible. First of all, get x_i by minimizing the modified l_1 norm, and sparse reconstruction weight vector s_i.

$$\min \| s_i \|_1$$
$$s.t. \quad x_i = X s_i$$
$$1 = e^T s_i$$

Among, $s_i = [s_{i,1}, \dots, s_{i,i-1}, 0, s_{i,i+1}, \dots s_{i,n}]^T$ is a n dimension sparse representation vector, the value of s_i is 0, make sure x_i are removed from the data set X, $s_{ij} (j \neq i)$ is the contribution that reconstruct x_i with each of x_j; $e \in R^n$ represents a column vector that all elements are 1. Similarly, due to the presence of noise and sample observations error, we need to loosen constraints $x_i = X s_i$ to get the following constrained optimization problem.

$$\min \| s_i \|_1$$
$$s.t. \quad \| X s - x_i \|_i \leq \varepsilon$$
$$1 = e^T s_i$$

To solve the weight vector s_i of each x_i, $i = 1, 2, \dots, n$. Therefore, the coefficients reconstruction matrix S is:

$$S = [s_1, s_2, \dots, s_n]^T$$

Sparse matrix S reflects the inherent sparse relations characteristics between the data, and it contains natural identification information. In order to make the original high-dimensional space relationship of the sparse features can be preserved in low dimensional space, so the objective function of sparse preserving projection can be expressed as:

$$\min_w \sum_{i=1}^{n} \| w^T x_i - w^T X s_i \|^2$$

$$s.t. \quad w^T XX^T w = 1$$

Finally, using simple linear algebra plan, the optimization problem can be solved by the following generalized eigenvalue problem solving:

$$XS_\beta X^T w = \lambda XX^T w$$

Among, $S_\beta = S + S^T - S^T S$, the optimal projection matrix is the d largest eigenvalues of $(XX^T)^{-1}(XS_\beta X^T)$ corresponding to the eigenvector.

3. Orthogonal Sparse Preserving Projections of Kernel

The face samples mapped to high-dimensional feature space by kernel method, so as to change the distribution of samples. It can map some of the data samples in the original space linear inseparable to high-dimensional feature space to become linearly separable by selecting a proper kernel function, and makes the coefficient sparse representation of data sample contains more identification information, more accurately sparse representation by the same kind of the rest sample. Thus, the algorithm uses kernel methods to map sample data to a high dimensional feature space and sparse preserving projection firstly.

Assuming sample category is M, the training sample is $A = [A_1, A_2, \ldots, A_M]$, among $A_i = [a_{i1}, a_{i2}, \ldots, a_{ik}] \in R^{m \times k}$, and k is the number of training samples for each category, A nonlinear mapped to high-dimensional feature space by kernel transformation ϕ, $X = [\phi(x_{11}), \phi(x_{12}), \ldots, \phi(x_{Mk})]$, category is divided into $X = [X_1, X_2, \ldots, X_M]$, it is the training sample matrix after nonlinear mapping.

In order to reflect more effective identification information for human face classification in the process of sample sparse reconstruction, strengthen the similar sample reconstruction weight, and reduce heterogeneous samples reconstruction weight, we optimize coefficients for sparse reconstruction.

First, strengthening the local linear characteristics between similar data by solving the least squares.

$$\min_t \| x_{ij} - X_i t \|_2 \tag{1}$$

$$s.t. \quad It = 1$$

Among, $t = [t_1, t_2, \ldots, t_{j-1}, 0, t_{j+1}, \ldots, t_k]^T$, the j th component of t is set to 0, the sample x_{ij} is removed from the sample set X_i, I represents a row vector that all elements are 1, we can get the optimal solution $\hat{t}$ by formula (1), $\hat{t} \in R^k$. It is the reconstruction weight of the similar sample, the weight information is invariant to rotation, translation, scale features, this method can be more approximate to test samples, the error is smaller.

While guarantee the similar reconstruction error minimization, we need to consider the effect of heterogeneous samples for refactoring, therefore, we should reduce the reconstruction weight of the heterogeneous sample. $\hat{t}$ will be extended to n component $\hat{h} = [O, \ldots, O, \hat{t}, O, \ldots, O] \in R^n$.

Assuming $er = x_{ij} - X_i \hat{t} = x_{ij} - X\hat{h}$, it is face samples residual after the same sample sparse reconstruction, and sparse reconstruction in heterogeneous data sets:

$$\min_s \| s \|_1 \tag{2}$$

$$s.t. \quad \| er - \hat{X}s \| < \varepsilon$$

$$Is = 0$$

The constraint condition $Is = 0$ is linear, s can be linear programed into $s = s^{+} - s^{-}$, and satisfy the following conditions.

$$s^{+} = \begin{cases} s, & s > 0 \\ 0, & s \le 0 \end{cases}, \quad s^{-} = \begin{cases} -s, & s < 0 \\ 0, & s \ge 0 \end{cases} \tag{3}$$

$\| s \|_1 = \sum_i | s_i | = \sum_i | s_i | = \sum_i (s_i^{+} + s_i^{-})$, and formula (2) can be changed as follow.

$$\min_{s^{+}, s^{-}} \sum_i (s_i^{+} + s_i^{-}) \tag{4}$$

$$s.t. \quad \| er - [\hat{X}, -\hat{X}][s^{+}, s^{-}]^T \| < \varepsilon$$

$$[I, -I][s^{+}, s^{-}]^T = 0$$

According to formula (4), the optimal solution $\hat{s}$ is obtained, $\hat{s} = [\hat{s}_1, ..., \hat{s}_{i-1}, O, \hat{s}_{i+1}, ..., \hat{s}_M]$, $\hat{s}$ is the reconstruction weight of the heterogeneous samples, we will put the reconstruction coefficient together to get final reconstruction weights $\hat{d} = \hat{s} + \hat{h} = [\hat{s}_1, ..., \hat{s}_{i-1}, \hat{t}, \hat{s}_{i+1}, ..., \hat{s}_M]$, $D = [d_1, ..., d_n]$. According to the constraint conditions $I\hat{t} = 1$, $I\hat{s} = 0$, so $I\hat{d} = 1$.

As a result, the sparse reconstruction weights projection will still be able to keep the rotation, translation, scale invariant features. Such improvements made similar sample coefficient is bigger, coefficient of heterogeneous samples tend to be 0. In order to retain the sparse reconstruction relationship between the training sample in transformation space, in projection space minimum error objective function is:

$$\min_v \sum_{i=1}^{n} \| v^T x_i - v^T X d_i \|^2 , \tag{5}$$

$$s.t. \quad v^T X X^T v = 1$$

Simplify the optimization equation:

$$\min_v v^T X (I - D - D^T + D D^T) X^T v_i ,$$

$$s.t. \quad v^T X X^T v = 1$$

Minimization problem can be equivalent to maximization problem:

$$\max_v \frac{v^T X S_\beta X^T v}{v^T X X^T v}$$

Among, $S_\beta = S + S^T - S^T S$. The optimal solution is the eigenvectors corresponding to $(X X^T)^{-1}(X S_\beta X^T)$. Because it is a symmetric matrix, the corresponding feature vector is not orthogonal, we will replace the original constraint conditions $v^T X X^T v = 1$ with $v^T v = 1$, optimization problem is converted into:

$$\max_v v^T X S_\beta X^T v$$

$$s.t. \quad v^T v = 1$$

Using the lagrange multiplier method to solve, we get optimal solution space of orthogonal sparse preserving projection as a whole, that is eigenvectors corresponding to d maximum eigenvalue of symmetric matrices $X S_\beta X^T$. This improvement not only ensures that projection vectors are orthogonal in the space, but also improve the ability of sparse reconstruction.

The specific algorithm process is shown in Figure 1.

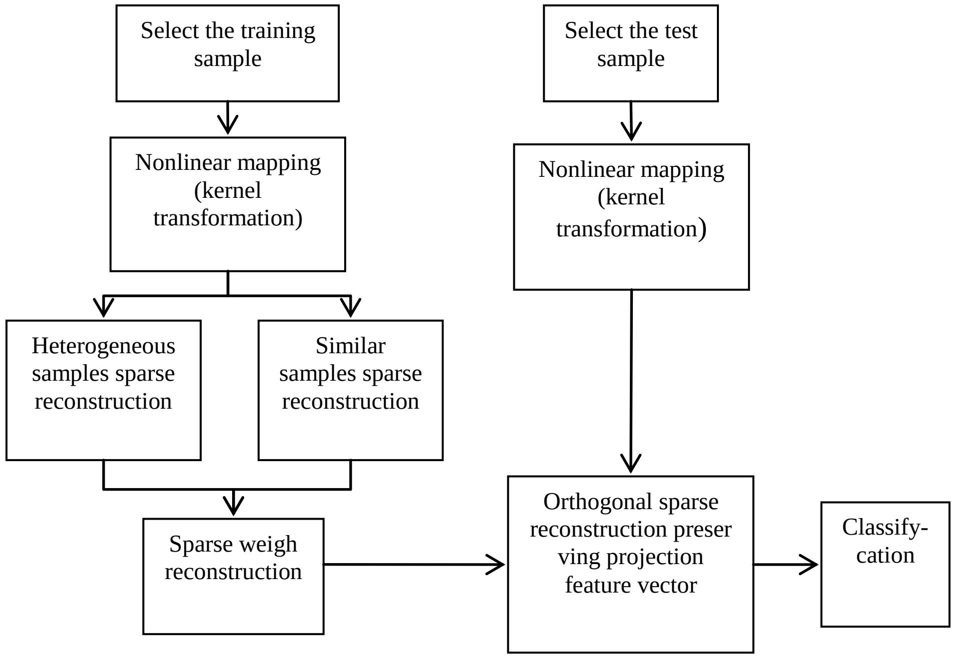

Figure 1. KODSPE Algorithm Specific processes

4. Experimental Analysis

In order to detect KODSPE algorithm is effective and feasible, we do an experiment analysis on the ORL and YALE_B face database, using matlab2010 test platform, and the nearest neighbor classifier as the classifier. We randomly selected the training samples in the experiment, and get the mean as the recognition rate by multiple independent training. At the same time, selecting different kernel function and parameters to analysis, comparing KODSPE algorithm and SPP, KSPP algorithm, so as to verify the validity of the algorithm.

4.1. Experiments on the ORL Face Database

The ORL face database contains 40 objects of different age, different gender and different races. It consists of 400 different facial expressions, shade, posture, angle of gray face image information, each object has 10 images, 92 x 112 pixels image. As shown in Figure 2.

Figure 2. Images in the ORL Face Database

First, validation KODSPE algorithm with the selection of kernel function and parameters, we selected 5 face image from each face set of ORL face database that contains 10 face image as the training sample set randomly, the training sample set is composed of 200 face images, other facial image of the database as a test sample set. As shown in Figure 3, it is the effects of kernel parameters on the face recognition rate, when we selected Gauss kernel and polynomial kernel as the nonlinear mapping respectively. The parameter of the Gauss kernel as t, polynomial parameter is d, as shown in Figure 4, when the Gaussian kernel function is selected, the optimal parameter is 5. When polynomial kernel function is selected, the optimal parameter is 2.

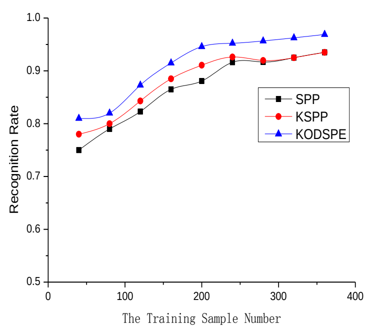

Figure 3. Different Methods Recognition Rate on ORL Face Database

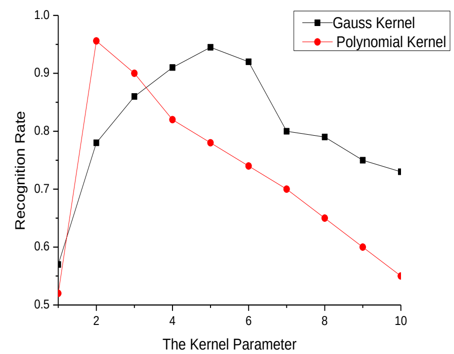

Figure 4. Recognition Rates with Kernel Parameter Changes

Then, using the optimal kernel parameter compared KODSPE with SPP, KSPP by experimental analysis. We choose a different number of face image as the training sample experiment, respectively, as shown in Figure 4, this is the contrast figure that the recognition rate of three algorithms changed with the number of training samples in the ORL face database. As shown in Table 1, it can be seen as the training sample number increases, recognition rate is higher, and in the same case KODSPE algorithm is superior to other algorithm.

Table 1. Different Training Sample Recognition Rate on ORL Face Database

	SPP	KSPP	KODSPE(polynomial)	KODPSE(gauss)
T_1(3)	82.3	84.7	87.8	87.4
T_2(4)	86.7	88.5	92.6	92.0
T_3(5)	91.0	92.4	94.6	93.8
Parameters	/	t=5	d=2	t=5

4.2. Experimental on YALE_B Face Database

YALE_B face database contains 10 people, each one has 64 pieces, a total of 640 images, its posture and illumination change images that were collected under the strictly controlled condition, so it is good to analysis the problem of illumination and posture, as shown in Figure 5.

Figure 5. Images in the YALE_B face Database

Using the same methods to experimental analysis on the YALE_B database, each face set of in the YALE_B face database contains 64 face images, we selected 20 face image as the training sample from each face set randomly, the others as test sample set. Through many experimental comparison and analysis on the YALE_B face database, we found that optimal polynomial kernel parameter d is 2, Gaussian kernel parameter is 2. then, doing a comparative experiments to the different algorithms according to the optimal kernel parameters, Figure 6 shows that is recognition rate change with the training sample number in the YALE_B face database, and as training sample number increase,the recognition rate increases, and selecting fewer facial image as the training sample set, you can get higher recognition rate compared with other algorithms, the KODSPE algorithm has better effectiveness and robustness. We selected 10, 15, 20 training samples from each face database and experiment many times randomly, as shown in Table 2, it is the recognition rate of different algorithms under the corresponding training sample. When selecting 20 training samples randomly, the recognition rate of KODSPE algorithm reaches 96.3% on the YALE_B face database. As a whole, the recognition rate of algorithm on the YALE_B face database is superior to the ORL face database, it shows that algorithm can better adapt to the influence of illumination and posture.

Table 2. Different Training Sample Recognition Rate on YALE_B face Database

	SPP	**KSPP**	**KODSPE(polynomial)**	**KODPSE(gauss)**
T_1(10)	86.8	88.5	91.3	90.8
T_2(15)	91.6	93.2	93.7	92.4
T_3(20)	92.5	94.5	96.3	95.6
Parameters	/	t=2	d=2	t=2

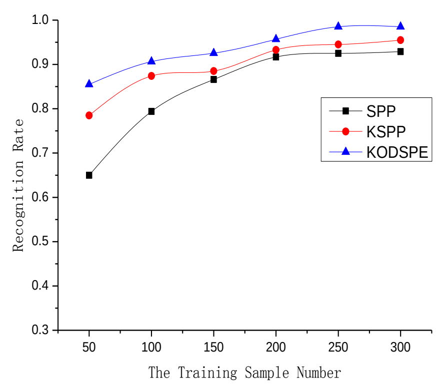

Figure 6. Different Methods Recognition Rate on the YALE_B face Database

5. Conclusion

This paper proposes the algorithms based on orthogonal sparse preserving projections of kernel (KODSPE). The algorithm used that kernel method has better linear separability in high dimensional feature space, it makes some of the inseparable data in the original space to be separable in higher dimensional linear space, and the coefficient of sparse representation contains more identification information, so as to classify and extract feature better. At the same time, reconstructing the weight of sparse coefficient can approach to the similar samples furthest, eliminate the interference of pseudo class, and overcome the problem of lack of category information, because the training sample is not enough. Finally orthogonal constraint transformation to the whole makes the sparse preserving projection matrix classify better.

Experimental results show that KODSPE algorithm has good recognition effect on ORL and YALE_B face database, it not only has a better recognition rate and feasibility compared with the traditional algorithm, but also has a very good application prospect.

Acknowledgements

The paper is supported by the National Natural Science Foundation (61103149) of China and the Technology Research Project (11551087) of Education Center in Heilongjiang Province.

References

[1] X. F. He and P. Niyogi, “Locality preserving projections”, In: Proceedings of the 2003 Neural Information Processing Systems (NIPS003). USA: The MIT Press, **(2003)**, pp. 153−160.
[2] X. F. He, D Cai, S. C. Yan and H. J. Zhang, “Neighborhood preserving embedding”, In: Proceedings of the 10th International Conference on Computer Vision (ICCV005), Beijing, China: IEEE Computer Society, (**2005**), pp. 1208−1213.
[3] W. Y. Cheng, K. Z. Lin and S. Liu, “Locally maintaining projection of typical correlation analysis”, Journal of Harbin University of Science and Technology, vol. 18, no. 5, (**2013**), pp. 65-69.
[4] Y. Jin and Q. Q. Ruan, “Kernel base orthogonal preserving projections for face recognition”, Journal of Electronics &Information Technology, vol. 31, no. 2, (**2009**), pp. 283-287.
[5] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, “Robust face recognition via sparse representation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, (**2009**), pp. 210−227.
[6] L. S. Qiao, S. C. Chen and X. Y. Tan, “Sparsity preserving projections with applications to face recognition”, Pattern Recognition, vol. 43, no. 1, (**2010**), pp. 331−341.
[7] H. Zheng and Z. Jin, “Supervised sparse neighborhood preserving embedding algorithm”, Computer Engineering, vol. 37, no. 16, (**2011**), pp. 155-157.
[8] Q. Wang and K. H. Hui, “Sparse neighbor representation for classification”, Computer Engineering and Design, vol. 4, (**2013**), pp. 1425-1431.
[9] W. N. Xiang and J. L. Zhao, “Supervised sparsity preserving projection”, Computer Engineering and Applications, vol. 47, no. 29, (**2011**).
[10] C. Y. Lu, H. Min and J. Gui, “Face recognition via weighted sparse representation”, Journal of Visual Communication and Image Representation, vol. 24, no. 2, (**2013**), pp. 111-116.
[11] F, Yin, L. C. Jiao and F. Shang, “Sparse regularization discriminant analysis for face recognition”, Neurocomputing, vol. 128, (**2014**), pp. 341-362.
[12] C. Kang, S Liao and S Xiang, “Kernel Sparse Representation with Pixel-level and Region-level Local Feature Kernels For Face Recognition”, Neurocomputing, (**2014**).

