DE GRUYTER OPEN

BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **14**, No 3

Sofia • 2014 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2014-0031

Face Recognition Based on Wavelet Kernel Non-Negative Matrix Factorization

Lin Bai, Yanbo Li, Meng Hui

School of Electronics and Control Engineering, Chang'An University, Xi'An 710064, Shaanxi, China
Emails: bai1981@sina.com.cn hmeng@chd.edu

Abstract: *In this paper a novel face recognition algorithm, based on wavelet kernel non-negative matrix factorization (WKNMF), is proposed. By utilizing features from multi-resolution analysis, the nonlinear mapping capability of kernel non-negative matrix factorization could be improved by the method proposed. The proposed face recognition method combines wavelet kernel non-negative matrix factorization and RBF network. Extensive experimental results on ORL and YALE face database show that the suggested method possesses much stronger analysis capability than the comparative methods. Compared with PCA, non-negative matrix factorization, kernel PCA and independent component analysis, the proposed face recognition method with WKNMF and RBF achieves over 10 % improvement in recognition accuracy.*

Keywords: *Face recognition, non-negative matrix factorization, RBF network, kernel method.*

1. Introduction

Face recognition has received ever-increasing attention for applications, such as identity authentication, information security, video surveillance, human-computer interface, and so on. However, the major challenge of face recognition is that the captured face images often lie in a high dimensional feature space. Generally, the dimensions of these spaces are too high to allow effective and efficient face recognition [1, 2].

During the past decades, many useful techniques of dimensionality reduction have been developed. The most well known techniques for dimensionality reduction are the Principal Component Analysis (PCA) [3, 4], and Linear Discriminant Analysis (LDA) [5]. Many face recognition methods, such as Eigenfaces [6] and Fisherfaces [7], are built on these two techniques or their variants. PCA generates a set of orthonormal basis vectors aiming at maximizing the variance over all samples. It computes the eigenvectors of the sample covariance matrix and approximates the original data by linear combination of the leading eigenvectors. PCA is optimal in terms of representation and reconstruction, but not in discriminating one face class from others.

Unlike the PCA method which is unsupervised, LDA is a supervised dimensionality reduction method. LDA seeks for an embedding transformation, such that the between-class scatter is maximized and the within-class scatter is minimized. The optimal transformation (projection) of LDA can be computed by applying an eigen-decomposition on the scatter matrices of the given training data. As for pattern classification, it is generally believed that LDA-based algorithms outperform PCA-based algorithms. However, one major drawback of LDA is that it suffers from a small sample size or is under-sampled when the number of samples is smaller than the dimensionality of samples [8, 9].

It was shown by Lee and Seung [10, 11] that positivity or non-negativity of a linear expansion is a very powerful constraint that also seems to yield sparse representations. Their technique, called Non-negative Matrix Factorization (NMF), was shown to be a useful technique in approximating high dimensional data, where the data are comprised of nonnegative components. However, NMF and many of its variants are essentially linear, and thus they cannot discover the nonlinear structures hidden in the face data. Besides, they can only deal with data with attribute values, while in many applications we do not know the detailed attribute values and only relationships are available. Thus, NMF cannot be directly applied for relation data. Furthermore, one requirement of NMF is that the values of the data must be non-negative, however in many real world problems the non-negative constraints cannot be satisfied.

Since the middle of 1990-ies, the kernel method has been successfully applied. Many nonlinear feature extraction methods based on the kernel method have been proposed [12-15].

In this paper a novel algorithm is proposed for face recognition by using Wavelet Kernel Non-negative Matrix Factorization (WKNMF), which can overcome the above limitations of NMF. Face recognition is achieved by combining WKNMF feature extraction and radial basis function. The proposed method is evaluated on ORL and YALE face database. Compared with other state-of-the-art algorithms, the classification accuracy of the method proposed can be increased by 10 %. The outline of this paper is as follows. Section 2 presents the proposed feature extraction based on WKNMF. Experimental results are reported in Section 3. Finally, conclusions are given in Section 4.

2. Methodology

2.1. Non-negative matrix factorization

NMF imposes the non-negativity constraints in learning the basis images. Both the values of the basis images and the coefficients for reconstruction are all non-negative. The additive property ensures that the components are combined to form a whole in a non-negative way, which has been shown to be the part based on representation of the original data. However, the additive parts learned by NMF are not necessarily localized [8, 9].

Given a non-negative $n \times m$ matrix *V* and constant *r*, the non-negative matrix factorization algorithm finds a non-negative $n \times r$ matrix *W* and another non-negative $r \times m$ matrix *H*, such that they minimize the following optimization problem:

$$\min f(W,H)$$
$$\text{subject to } W \geq 0, H \geq 0. \tag{1}$$

This can be interpreted as follows: each column of matrix *W* contains a basis vector while each column of *H* contains the weights needed to approximate the corresponding column in *V* using the basis from *W*. Thus the product *WH* can be regarded as a compact form of the data in *V*. The rank *r* is usually chosen as $r << \min(n, m)$. Function $f(W, H)$ is a loss function. In this paper, we choose the loss function as follows:

$$f(W,H) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(V_{ij} - (WH)_{ij})^2. \tag{2}$$

Here *H* and *W* are updated iteratively by solving the multiplicative rule function as follows:

$$H_{bj} \leftarrow \frac{(W^{\mathrm{T}}V)_{bj}}{(W^{\mathrm{T}}WH)_{bj}},$$
$$W_{ib} \leftarrow W_{ib}\frac{(VH^{\mathrm{T}})_{ib}}{(WHH^{\mathrm{T}})_{ib}}. \tag{3}$$

The convergence of the process is guaranteed [9]. The initialization is performed using the positive random initial conditions for matrices *W* and *H*.

2.2. Kernel non-negative matrix factorization

Given *m* objects $\Theta_1, \Theta_2, \Theta_3, ..., \Theta_m$, with attribute values represented as an $n \times m$ matrix $\Omega = [\omega_1, \omega_2, ..., \omega_m]$, each column of which represent one of the *m* objects. We define the nonlinear map from the original input space Ω to a higher or infinite dimensional feature space Φ as follows:

$$\phi : x \in \Omega \to \phi(x) \in \Phi. \tag{4}$$

From the *m* objects we denote

$$\phi(\Omega) = [\phi(\omega_1), \phi(\omega_2), ..., \phi(\omega_m)]. \tag{5}$$

Similarly to NMF, KNMF searches for two non-negative matrix factors W_ϕ and H, such that

$$\phi(\Omega) = W_\phi H , \tag{6}$$

where W_ϕ is the basis in the feature space Φ and H is its combining coefficients, each column of which now denotes the dimension-reduced representation for the corresponding object. It is worth noting that since $\phi(\Phi)$ is unknown, it is impractical to directly factorize $\phi(\Omega)$. From (6) we obtain

$$\left(\phi(\Omega)\right)^{\mathrm{T}} \phi(\Omega) = \left(\phi(\Omega)\right)^{\mathrm{T}} W_\phi H . \tag{7}$$

A kernel is a function in the input space and at the same time the inner product in the feature space through the kernel-induced nonlinear mapping. More specifically, a kernel is defined as

$$k(x, y) = \langle \phi(x), \phi(y) \rangle = \left(\phi(x)\right)^{\mathrm{T}} \phi(y). \tag{8}$$

From (8), the left side of (7) can be rewritten as

$$\left(\phi(\Omega)\right)^{\mathrm{T}} \phi(\Omega) = \left\{ (\phi(\omega_i))^{\mathrm{T}} \phi(\omega_j) \right\}_{i,j=1}^{m} = \left\{ k(\omega_i, \omega_j) \right\}_{i,j=1}^{m} = K . \tag{9}$$

Let us denote

$$Y = \left(\phi(\Omega)\right)^{\mathrm{T}} W_\phi . \tag{10}$$

From (9) and (10) equation (7) can be changed as

$$K = YH . \tag{11}$$

Comparing (11) with (6) it can be found that the combining coefficient H is the same. Since W_ϕ is a learned basis of $\phi(\Omega)$, similarly we call Y in (11) as basis of the kernel matrix K. Equation (11) provides a practical way for obtaining the dimension-reduced representation H by performing NMF in the kernel space.

For a new data point, the dimension-reduced representation is computed as follows:

$$H_{\mathrm{new}} = \left(W_\phi\right)^{+} \phi\left(\omega_{\mathrm{new}}\right) = \left(W_\phi\right)^{+} \left(\left(\phi(\Omega)\right)^{\mathrm{T}}\right)^{+} \left(\phi(\Omega)\right)^{\mathrm{T}} \phi\left(\omega_{\mathrm{new}}\right) = Y^{+} K_{\mathrm{new}} . \tag{12}$$

Here A^{+} donates the generalized (Moore-Penrose) inverse of matrix A, and $K_{\mathrm{new}} = \left(\phi(\Omega)\right)^{\mathrm{T}} \phi\left(\omega_{\mathrm{new}}\right)$ is the kernel matrix between the *m* training instance and the new instance. Equations (11) and (12) construct the key components of KNMF when used for classification. It is easy to see that the computation of KNMF does not need any attribute values of the objects, but only the kernel matrices K and K_{new} are required.

Obviously, KNMF is more general than NMF, because the former can deal with not only attribute value data, but also relational data. Another advantage of KNMF is that it is applicable to data with negative values, since the kernel matrix in KNMF is always non-negative for some specific kernels.

2.3. Wavelet kernel non-negative matrix factorization

The purpose of building a kernel function is to project the observed data from a low dimensional space to another high dimensional space. WKNMF method uses the kernel function in non-negative matrix factorization and improves it by replacing the traditional kernel function by the wavelet kernel function. Using the features of multi-resolution analysis, the nonlinear mapping capability of the kernel non-negative matrix factorization method can be greatly improved.

Assuming that $h(x)$ is a wavelet function, the parameter α represents a stretch and β represents a pan. If there exists $x, x' \in R^N$, then we get a dot product form of the wavelet kernel function:

(13) $$K(x,x') = \prod_{i=1}^{N} h\left(\frac{x_i - \beta_i}{\alpha}\right) h\left(\frac{x'_i - \beta'_i}{\alpha}\right).$$

Under the condition of translation invariance, (13) can be rewritten as

(14) $$K(x,x') = \prod_{i=1}^{N} h\left(\frac{x_i - x'_i}{\alpha}\right).$$

In this paper Morlet wavelet function was selected as a generating function, according to the theory of translation invariance wavelet function, the kernel function is constructed as

(15) $$h(x) = \cos(1.75x)e^{(-x^2/2)}.$$

From (13), (14) and (15), a wavelet kernel function satisfying the requirements of Mercer kernel function is built as:

(16) $$K(x, x') = \prod_{i=1}^{N}\left(\cos\left(1.75\frac{(x_i - x_i')}{\alpha}\right)e^{\left(-\frac{\|x_i - x_i'\|^2}{2\alpha^2}\right)}\right).$$

By using (16) in kernel non-negative matrix factorization, we can get Wavelet kernel non-negative matrix factorization.

3. Experimental results

3.1. Experimental setting

In order to evaluate the efficiency of the algorithm proposed, an extensive experimental investigation is conducted on a subset of ORL and YALE face database. In this experiment, the face data of the experimental settings are shown in Tables 1 and 2.

The ORL face database contains 400 images of 40 individuals (each individual has 10 images). The images were captured at different times and with different variations, including expression and facial details. Each image in the database is a gray image with a size 92×112. Part images of ORL face database are shown in Fig. 1.

The Yale face database contains 165 gray scale images of 15 individuals (each individual has 11 images). Each image is a gray image and the size is 100×100.

The images demonstrate variations in the lighting condition, facial expression, and with/without glasses. Part images of YALE face database are shown in Fig. 2.

Fig. 1. Part images of the ORL face database

Fig. 2. Part images of the YALE face database

Table 1. Data of experimental setting A

Training and testing setting			
Face database	Samples	Training sample	Testing sample
ORL	Set I	200	400
YALE	Set I	90	165

Table 2. Data of experimental setting B

Training and testing setting			
Face database	Samples	Training sample	Testing sample
ORL	Set I	120	400
	Set II	160	400
	Set III	200	400
YALE	Set I	45	165
	Set II	75	165
	Set III	90	165

3.2. Experimental results

Depending on the setting of the experimental data in Table 1, the recognition rate curves with respect to the number of different sub-dimensional spaces distributed in the ORL and YALE database respectively are shown in Figs 3 and 4. From Fig. 3 and Fig. 4 we can see that the recognition rate of WKNMF is higher than in other algorithms, such as PCA, KPCA, NMF and ICA with different subspace dimensions. In ORL database, the subspace dimension of the best recognition rate is around 30, while in YALE database, the best recognition rate with subspace dimension is around 40. We can make the conclusion, that the dimension of the subspace is not as high as possible, the subspace separability mainly lies in the vicinity of a particular dimension, but this dimension will vary with changes in the face database, so that the best estimate of the dimension of the subspace is a subject of important work in future.

PCA, NMF, KPCA (Kernel-PCA), ICA (Independent Component Analysis) and WKNMF are respectively employed for feature extraction according to the experimental data of setting B. The results of the experiments are shown in Tables 3 and 4.

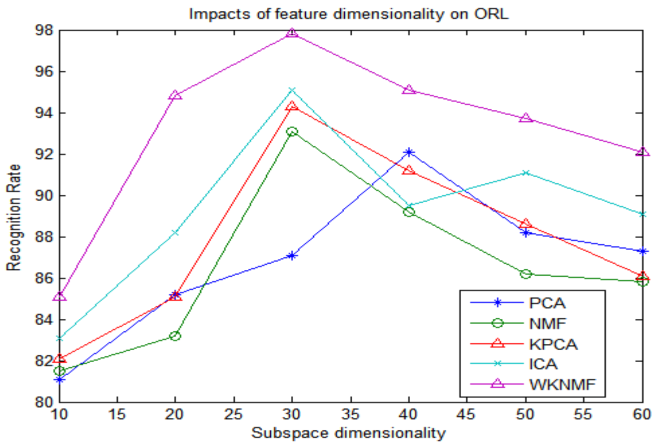

Fig. 3. Recognition rate (%) influence of the subspace dimension on ORL

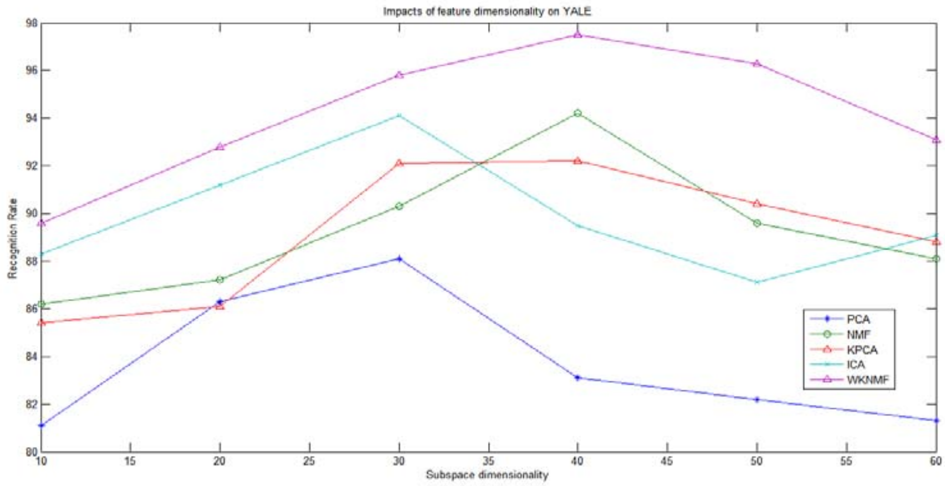

Fig. 4. Recognition rate (%) influence of the subspace dimension on YALE

We can see from Tables 3 and 4, with the same neural network classifier, that the proposed method of recognition ability outperforms significantly several other classic algorithms. The misclassified images of WKNMF+RBF of ORL database faces are shown in Fig. 5, which are mainly due to change in the age, decoration and so on in the training image face database. These problems will lead to recognition errors.

Fig. 5. Misclassified images of WKNMF+RBF on ORL face database

The WKNMF+RBF on Yale database achieves better recognition rate than the other methods. The misclassified images are shown in Fig. 6. In multiple tests on

Yale database the misclassified images have remained the same. The proposed method has some influence by the facial expression and illumination.

Fig. 6. Misclassified images of WKNMF+RBF on ORL face database

In summary, our proposed WKNMF algorithm is an efficient method for face recognition. The experimental results on ORL and YALE databases show that the proposed algorithm achieves higher recognition performance than the others.

Table 3. Recognition rate of a different method on ORL data (%)

Algorithm	Set I	Set II	Set III
PCA+RBF	86.5	83.25	87.75
NMF+RBF	89.0	92.0	93.0
KPCA+RBF	89.7	92.8	94.7
ICA+RBF	87.3	88.0	92.3
WKNMF+RBF	93.4	95.2	98.8

Table 4. Recognition rate of a different method on YALE data (%)

Algorithm	Set I	Set II	Set III
PCA+RBF	63.64	74.55	86.06
NMF+RBF	67.27	78.18	89.09
KPCA+RBF	71.1	85.8	92.3
ICA+RBF	62.4	76.3	87.3
WKNMF+RBF	81.9	96.1	98.2

4. Conclusions

In this paper we propose a face recognition algorithm by using WKNMF. The idea of using WKNMF is to find a set of basis functions to represent the face image where the basis functions enable the identification and classification of intrinsic “parts” that make up the object being imaged by multiple observations. The experimental results on both ORL and YALE face databases show that WKNMF has much stronger analysis capability than the comparative algorithms. In the aspect of face recognition accuracy, the method proposed achieves over 10% improvement compared to the PCA, NMF, KPCA and ICA methods. The WKNMF balances the algorithm efficiency and performance very well.

Acknowledgments: This work was supported by the special fund for basic scientific research of central colleges, Chang’an University. The projects are No CHD2011JC170, CHD2011JC161, 2013G1321037 and CHD2011TD018.

References

1. Putra, Kusuna Gede, Chua Chin-Seng. PCA-Based Image Recombination for Multimodal 2D+3D Faces Recognition. – Image and Vision Computing, Vol. **1**, 2011, No 29, 306-316.
2. Cong, Geng, Xudong Jiang. Hook. Face Recognition Based on the Multi-Scale Local Image Structures. – Pattern Recognition, Vol. **10**, 2011, No 3, 2565-2575.
3. Gutta, S., V. Philomin, M. Trajkovic. An Investigation into the Use of Partial-Faces for Face Recognition. – In: Proc. of 5th IEEE International Conference on Automatic Face and Gesture, Vol. **5**, 2002, No 1, 33-38.
4. Gutta, S., H. Wechsler. Partial Faces for Face Recognition: Left Vs. Right Half. – In: Proc. of 10th International Conference on Computer Analysis of Images and Patterns, LNCS 2756, 2003, 630-637.
5. Koc, M., A. Barkanai. A New Solution to One Sample Problem in Face Recognition Using FLDA. – Applied Mathematics and Computation, Vol. **3**, 2011, No 8, 10368-10376.
6. Turk, M., A. Pentland. Eigenfaces for Recognition. – Journal of Cognitive Neuroscience, Vol. **3**, 1991, No 18, 71-86.
7. Belhumeur, P. N., J. P. Hespanha, D. J. Kriegman. Eigenfaces Vs. Fisherfaces: Recognition Using Class Specific Linear Projection. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **19**, 1997, No 13, 711-720.
8. Shye-Chorng, Kuo, Lin Cheng-Jian, Liao Jan-Ray. 3D Reconstruction and Face Recognition Using Kernel-Based ICA and Neural Network. – Expert Systems with Application, Vol. **3**, 2011, No 3, 5406-5415.
9. Xueming, Qian, Hua Xiansheng, Chen Ping. PLBP: An Effective Local Binary Patterns Texture Descriptor with Pyramid Representation. – Pattern Recognition, Vol. **5**, 2011, No 7, 2502-2515.
10. Lee, D. D., H. S. Seung. Learning the Parts of Objects with Nonnegative Matrix Factorization. – Nature, Vol. **401**, 1999, No 1, 788-791.
11. Lee, D. D., H. S. Seung. Algorithms for Nonnegative Matrix Factorization. – In: Advin Neur. Inform. Proc. Syst. Cambridge, MIT Press, 2000, 556-562.
12. Gao, Tao, Ma Xiang. A Novel Face Description by 2DFLDA Supervised Projection Gradient Non-Negative Matrix Factorization. – Journal of Computational Information System, Vol. **3**, 2011, No 4, 5307-5313.
13. Wang, J., M. A. Jabbar. Multiple Kernel Learning for Adaptive Graph Regularized Nonnegative Matrix Factorization. – National Association of Science and Technology for Development, Vol. **6**, 2012, No 3, 115-122.
14. Kai-Bo, Duan, Keerthi S. Sathiya. Which Is the Best Multiclass SVM Method? – An Empirical Study. – In: Proc. of 60th International Workshop on Multiple Classier Systems. Lecture Notes in Computer Science, Vol. **3541**, 2005, No 2, 278-286.
15. Gao, Tao, Ma Xiang, Wang Jiangan. Particle Occlusion Face Recognition Using Adaptively Weighted Local Gabor Filters. – Journal of Computational Information System, Vol. **5**, 2012, No 6, 5271-5277.