A Hybrid Face Recognition Algorithm Based on WT, NMFs and SVM

Pei Jiang , Yongjie Li
School of Life Science and Technology,
University of Electronics Science and Technology of China,
Chengdu 610054
E-mail: liyj@uestc.edu.cn

***Abstract*—In this paper, we proposed a new scheme for face recognition, which hybridizes wavelet transform (WT), non-negative matrix factorization with sparseness constraints (NMFs) and support vector machine (SVM) with relative difference space (RDS) method. Firstly, low frequency subband images are extracted from original face image with 2D wavelet transform. Secondly, the images with low frequency information are factorized with NMFs, which could find part-based representations of images. Then, the multi-class problem is to be converted to the binary issue by RDS method and the extracted features are classified through SVM. The experiments on ORL face dataset shows the more efficient results with the proposed algorithm.**

***Index Terms*—Face Recognition, Wavelet Transform, Support Vector Machine, Non-negative Matrix Factorization , Relative Difference Space**

I. Introduction

Face recognition is a very significant task, which can be applied in a wide range of fields, such as identity authentication, access control and video retrieval. It involves the techniques of pattern recognition, image processing, and computer vision. Compared to classical pattern recognition problems, face recognition is more difficult because there are usually many individuals with only a few images for one class, and normally, a face recognition system must recognize faces by extrapolating from the training samples. Various changes in face images also present a great challenge, and a face recognition system must be robust with respect to all kind of natural varieties of face images. Numerous face recognition papers have been published till now[1,2,3,4,5]. Among these techniques, subspace methods have been successfully applied, because they allow efficient characteristics of a low dimensional subspace preserving the perceptual quality of a very high dimensional face image. Eigenface method based on principle component analysis (PCA) is one of the most popular methods[6]. But the common Eigenface-based methods suffer from some limitations. First, PCA representation has a poor discriminatory ability though it gives a very good presentation of the images. Second, many of the basis images don't have an obvious visual interpretation. Finally, this approach focuses on extracting global face features, so the case with occlusions is difficult to be handled. A new technique called non-negative matrix factorization (NMF) for obtaining a linear representation of data has been proposed [7].

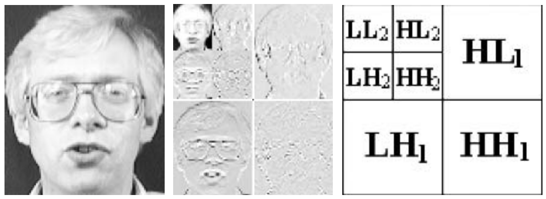
NMF is originally developed to extract basis components manifesting building parts consisting of localized features, which offers advantages in object recognition, including stability to local deformations, lighting variations and partial occlusion. It is distinguished from other methods such as PCA by its use of nonnegativity constraints. These constraints lead to a parts-based representation because they only allow additive, not subtractive combinations[8]. Non-negative matrix factorization with sparseness constraints (NMFs) is a great improvement to the standard NMF. It can control the sparseness of the expected matrix and also learn the part-based expression well. The application of NMFs will pave the way for the face recognition task meaningfully.

Support vector machine have been used successfully for pattern recognition problem [9]. SVM learns from the example images and relies on the techniques from machine learning to find the relevant characteristics of face recognition. But it is a tough work for SVM to deal with the multi-class issue. Many methods have been proposed to solve the problem and the most basic strategies are the one-vs-all and the one-vs-one approaches. An advanced algorithm named relative difference space (RDS) is proposed which can transform the multi-class problem to the binary issue [10]. The RDS-combined SVM makes it simple and convenient to accomplish the classification work.

II. Wavelet Transform On Face Image

Wavelet Transform is a popular tool in image processing and computer vision, which has nice features of space-frequency localization and multi-resolutions. The main reasons for its popularity lie in its complete theoretical framework, the great flexibility for choosing bases and the low computational complexity.

Two-dimensional WT decomposes an image into 4 ”subbands” that are localized in frequency and orientation, denoted by LL, HL, LH, and HH (Fig.1). The Low frequency components contribute to the global description while high-frequency components contribute to the details. Decomposing a face image using wavelet transform, the effect of different facial expressions can be attenuated by removing the high frequency components. Wavelet transform for face image can offer some advantages to us: the data of the low frequency subband is greatly compressed from the original, meanwhile the effect of different facial expressions can be attenuated and the main information for recognition is remained.



(a) (b) (c)

Fig. 1. 2-level wavelet decomposition of face image

We apply the 2D wavelet transform to the original face images to extract the low frequency subband component and decompose each image into 2 levels with Daubechies D2 wavelet[11].

III. NMFs For Feature Extraction

NMF is an algorithm to obtain a linear representation of data under non-negativity constraints. These constraints lead to a part-based representation because they allow only additive, not subtractive, combinations of the original data.

First, an image database as a $n \times m$ matrix V , where each column, corresponding to an initial face image, includes n non-negative elements characterizing the pixel value and m is the number of training images. Then we can find two new non-negative matrices (W and H) to approximate the original matrix:

$$V_{n \times m} \approx W_{n \times r} \cdot H_{r \times m} \tag{1}$$

where r is the number of base vectors, which is usually chosen as small as possible for dimension reduction, and each column of matrix W represents a basis vector while each column of matrix H means the weight used to approximate the corresponding column in V using the bases from W.

Adding sparseness constraints to NMF, Hoyer [7] proposed a new NMF algorithm named NMFs, which has a good performance to figure out the intuitive parts-based representations and is easy to interpret the resulting decompositions. The new NMF algorithm can even control the degree of sparseness explicitly and converge quickly. To make the target function that

$$E(W, H) = \| V - WH \|^2 = \sum_{i,j} (V_{ij} - (WH)_{ij})^2 \tag{2}$$

is minimized, under optional constraints:

$$\begin{aligned} sparseness(w_i) &= S_w, \forall i \\ sparseness(h_i) &= S_h, \forall i \end{aligned} \tag{3}$$

where W_i is the ith column of W and h_i is the ith row of H. Here,S_w and S_h are the desired sparseness of W and H respectively.

sparseness is defined as:

$$sparseness(x) = \frac{\sqrt{n} - (\sum |x_i|)/\sqrt{\sum x_i^2}}{\sqrt{n} - 1} \tag{4}$$

where n is the dimension of x. Figure 2 illustrates the concept and the sparseness measure. In the picture, Four vectors are shown, exhibiting sparseness levels of 0.1, 0.4, 0.7, and 0.9. Each bar denotes the value of one element of the vector. At low levels of sparseness (leftmost), all elements are roughly equally active. At high levels (rightmost), most coefficients are zero whereas only a few take significant values[7].

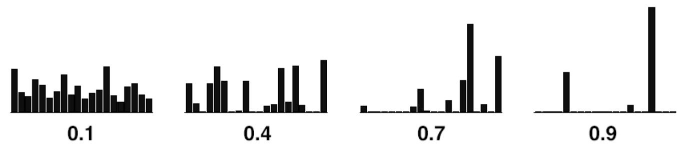


Fig. 2. Illustration of various degrees of sparseness

NMFs can control the sparseness of the expected matrix and simultaneously get the expectant characteristics from the face database.

By adjusting the sparseness of the matrix W, we could take the difference of the matrix W's elements under control, namely the number of the pixels whose intensity has been hold to zeros could be flexible. On one hand, when the matrix W is at the statement of low sparseness, the difference of the matrix elements' values is kept at the low level so that it could reflect the global relationship in the face images' elements and express the global characteristic well. On the other hand, if we adjust the sparseness of matrix W at the high level, the difference which lies in columns of W corresponding to faces' feature vectors would become large and the local relationship in face images' elements would only be retained, which makes it convenient to reveal the local characteristic of human faces.

IV. SVM With RDS Method

SVM behaves well in statistical learning in pattern recognition. It can minimize structural risk to perform well in both linear and non-linear classification, especially in the binary issues[12].

Since face recognition is a multi-class problem, the binary classification methods can not be applied directly. There are two basic strategies for solving n-class problems with SVMs. (1) In the one-vs-all approach n SVMs are trained. Each of the SVMs separates a single class from all remaining classes [13]. (2) In the one-vs-one approach $n(n-1)/2$ machines are trained. Each SVM separates a pair of classes. In the both schemes, however, it is difficult to analyze the bound of the generalization error, and the run-time cost is expensive since the final result can not be produced by one step.

Recently a new method based on Relative Difference space (RDS) and SVM is proposed for multi-class recognition [10]. It converts the multi-class issue to the binary problem whether the difference between two samples is within-class or between-class, and then the SVM classifiers can be directly used to work with multi-class recognition.

SVM minimizes the structural risk without incorporating problem-domain knowledge [13]. So the recognition method based on RDS and SVM is a general solution of multi-class problem.

Assume a training set as $X = \{x_1^{c1}, x_2^{c2}, \ldots, x_N^{cK}\}$, which include N sample of K classes ($K \leq N$). Here the superscript cn of the samples denotes the class ID.

Then we have to figure out the reference point ref_{ci} for each class in relative difference space (RDS). For K classes, there are K reference points. Each of those points is located at the same relative position to the distribution of its corresponding class [10].

When the reference point ref_{ci} of each class is located, the RDS transformation is :

$$RD_w = \{x_i^{ci} - ref_{ci} | i = 1, 2, \ldots, N\} \tag{5}$$

$$RD_b = \{x_i^{ci} - ref_j | c_i \neq j; i = 1, 2, \ldots, N; j = 1, 2, \ldots, K\} \tag{6}$$

where RD_w is the relative within-class set and RD_b is the relative between-class set.

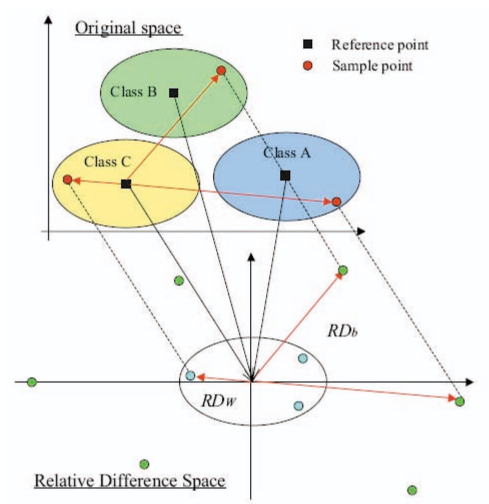


Fig. 3. The transformation of Relative Difference Space

Fig. 3 shows the demonstration of 3 classes issue converting to the binary-class one.

Thus, we get a new binary-class training set of points x_i, $i = 1, 2, \ldots, N$, where each point x_i belongs to one of two classes identified by the label $y_i \in \{-1, 1\}$. Then each point x_i in the input space can be mapped to a point $z = \Phi(x)$ of a higher dimensional space, called the feature space, where the data are separated by a hyperplane. The key property in this construction is that the mapping is subject to the condition that the dot product of two points in the feature space can be rewritten as a kernel function $K(x, y)$ [14].

The decision surface has the equation as:

$$f(x) = \sum_{i=1}^{l} \alpha_i K(x, X_i) + b \tag{7}$$

where the coefficients α_i and b are solutions of a quadratic programming problem. The polynomials kernel is applied:

$$K(x, y) = (1 + x \cdot y)^d \tag{8}$$

where d is the degree of the polynomial. In this case the components of the mapping $\Phi(x)$ are all the possible monomials of input components up to the degree d.

V. Experimental Results

Our experiments were conducted on the face database of Olivetti Research Laboratory (ORL). There are 40 subjects and each subject had 10 different facial views representing various expressions, different scale and orientations. Each image was digitized and stored as an 112×92 pixel array and 256 gray levels.

The training set is set up by a random selection of 3 samples for each person from the ORL database, which includes 120 images totally and the testing set consist of the remaining images. In the experiment, the low frequency subband images are extracted by the 2D wavelet transform on the original face images and we decompose each image into 2 level using $DaubechiesD2$; then, NMF with sparseness constraint algorithm is used on the low frequency images and the coefficient of sparseness is 0.6; finally, relative difference space based SVM is applied as the classifier. In order to test our scheme, PCA and K-nearest Neighbor (KNN) are applied to compare with our method and all of them are combined with the 2D wavelet transform.

TABLE I
Performance Evaluation

Methods	Recognition accuracy
WT+PCA+SVM	89.1%
WT+NMFs+SVM	93.6%
WT+NMFs+KNN	91.5%

Table I illustrates the comparison of recognition accuracy of different schemes. The data is obtained at the best dimension in $40 - 50$, which means the number of the basis images.In this situation, The recognition accuracy is 93.6% by our scheme and it shows a good performance. From Fig. 4, we can also compare the effect of the different methods under the same condition and all the algorithms are preprocessed by 2D wavelet transformation. It could be observed that the recognition rate increased with the growth of the basis images' number in the range 0-45 approximately and our scheme has the most effective result.

VI. Conclusion

A new face recognition algorithm hybridizing wavelet transform, non-negative matrix factorization and support vector machine is reported. The ORL face database was employed to evaluate the performance of the proposed method. Meanwhile, the experimental results show that our approach increases the face recognition accuracy and is more robust to expression. With the 2D wavelet process, the resolution of image could be decreased and the effect of different facial expressions was attenuated. NMFs brings us an excellent way to learn the part-based feature of the face images. As a statistical algorithm in pattern recognition, SVM behaves well in binary class issue and can provide a platform in multi-class task by benefitting from the relative difference space method.

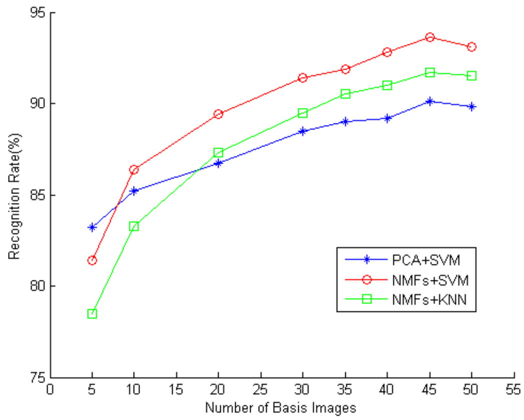


Fig. 4. Result of recognition of different methods

Acknowledgment

This work is supported by the grants from NSFC of China (Grant No.30500140 and 30730036) and a grant from the 973 Project of China (Grant No. 2007CB311001).

References

[1] Peter N. Belhumeur, Joao P. Hespanha and David J.kriefma, "Eigenfaces vs fisherfaces, recognition using class specific linear projection," IEEE Transactions On Pattern Analysis And Machine Intelligence, vol 19, no.7, pp. 711-720, July 1997.
[2] Xiaorong Pu, Zhang Yi, Ziming Zheng, Wei Zhou and Mao Ye, "Face recognition using fisher non-negative matrix factorization with sparseness constraints," Lecture Notes in Computer Science, Vol. 3497, pp. 112-117, 2005.
[3] C. Liu and H.Wechsler, "Gabor feature based classification using enhanced fisher linear discriminate model for face recognition," IEEE Transactions on image processing, vol. 11, no. 4, April, 2002.
[4] T.Cootes, "An introduction to active shape models," Image Processing and Analysis, pp. 223-248, 2000.
[5] S. Z. Li and J. Lu, "Face recognition using the nearest feature line method," IEEE Trans. Neural Networks, vol 10, pp. 439-443, Mar. 1999.
[6] G. C. Feng, P. C. Yuen, and D. Q. Dai, "Human face recognition using PCA on wavelet subband," J. Electron. Imaging, vol 9, pp. 226-233, 2001.
[7] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," Nature, 401(6755): 788-791, 1999.
[8] P.O.Hoyer, "Non-negative matrix factorization with sparseness constraints," Journal of Machine Learning research, (5): 1457-1469, 2004.
[9] J. Platt, N. Cristianini, and J. Shawe-Taylor, "Large margin DAGs for multi-class classification," Proceedings of Neural Information Processing Systems, pp. 547-553, MIT Press, 1999.
[10] Xiaoguang He, Jie Tian, Yuliang He and Xin Yang, "Face recognition with relative diference space and SVM," The 18th International Conference on Pattern Recognition, vol 3, pp. 527-531, 2006.
[11] I. Daubechies, "The wavelet transform, time-frequency localization and signal processing", IEEE Trans. Inform. Theory, vol 36, pp. 961-1005, 1990.
[12] V. Vapnik, "Statistical learning theory," John Wiley and Sons, New York, 1998.
[13] C. Cortes and V. Vapnik, "Support vector networks," Machine Learning, vol 20, no.3, pp.1-25, 1995.
[14] V. Vapnik, "The nature of statistical learning theory," Springer, 2nd edition, 1999.