Available online at www.sciencedirect.com
SCIENCE DIRECT®

Pattern Recognition 39 (2006) 2218–2222

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Rapid and brief communication

Face recognition using common faces method

Yunhui He[a,b,*], Li Zhao[a], Cairong Zou[a]

[a]*Department of Radio Engineering, Southeast University, Nanjing 210096, China*
[b]*Department of Information and Communications, Nanjing University of Information Science and Technology, Nanjing 210044, China*

Received 14 November 2005; accepted 24 April 2006

Abstract

In this paper, we propose a face recognition method called the commonface by using the common vector approach. A face image is regarded as a summation of a common vector which represents the invariant properties of the corresponding face class, and a difference vector which presents the specific properties of the corresponding face image such as face appearance, pose and expression. Thus, by deriving the common vector of each face class, the common feature of each person is obtained which removes the differences of face images belonging to the same person. For test face image, the remaining vector with each face class is derived with the similar procedure to the common vector, which is then compared with the common vector of each face class to predict the class label of query face by finding the minimum distance between the remaining vector and the common vector. Furthermore, we extend the common vector approach (CVP) to kernel CVP to improve the performance of CVP. The experimental results suggest that the proposed commonface approach provides a better representation of individual common feature and achieves lower error rates in face recognition.

Keywords: Face recognition; Common vector approach; Kernel method

1. Introduction

The face recognition technique has been intensively studied over the past few decades due to its potential applications. Since the face images are very sensitive to the variations such as face appearance, pose and expression variations, the variations between the image of the same face are often larger than image variations due to change in same face class. Besides, the face recognition task often encounters the so-called small sample size problem since the number of the samples is usually smaller than the dimensionality of the sample, which leads to the ill-posed problem in the traditional Fisher linear discriminant analysis (FLDA) for facial feature extraction. These problems make the face recognition a difficult task. To address these problems, many methods are proposed mainly focusing on the facial feature extraction to obtain the optimal representation of face image which is more compact in lower subspace, and therefore more suitable for face classification [1,2].

Recently, a pattern recognition method called common vector approach (CVP) was applied to isolated word recognition [3]. The environmental effects and personal differences are removed by deriving a common vector from a spoken word which represents common properties of the spoken word. In Ref. [4], it was further proved that the common vectors are related to the principle component analysis (PCA) and can be obtained with the eigenvectors corresponding to the zero eigenvalue of the covariance matrix. In this paper, we propose a face recognition method based on the CVP. By using the CVP, a common vector (we called commonface) for each face class is derived from the face image belonging to the same face class, which aims to eliminate the undesired variations in the same face class such as the face appearance, pose and expression variations. The commonface for each face class is independent of the face images arbitrarily selected from the respective face class, and thus the common invariant properties of each person can be presented by using commonfaces. Furthermore, we extend the CVP to nonlinear case by using kernel trick which nonlinearly maps face image to high-dimensional feature

* Corresponding author. Tel.: +86 25 83793791.
E-mail address: yunhuihe@163.com (Y. He).

0031-3203/$30.00 © 2006 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2006.04.037

space and then classification is performed in feature space. Finally, the angle between the remaining vector of query face and common vectors is used as the decision criteria to classify the query face in the experiments.

2. Face recognition using common vector approach

Assume that there are C classes, and N_c face image samples in class c. Let $x_m^i \in R^d$ be mth sample in ith class. The matrix B_i whose columns span a difference subspace $L(B_i)$ for ith class is defined as follows:

$$B_i = [b_1^i, b_2^i, \ldots, b_{N_i-1}^i], \tag{1}$$

where $b_k^i = x_{k+1}^i - x_1^i$, $k = 1, 2, \ldots, N_i - 1$. The x_1^i is called reference vector which can be randomly selected from ith class and here the first sample is selected.

By performing Gram–Schmidt orthogonalization procedure, the orthonormal vector set $\{z_1^i, z_2^i, \ldots, z_{N_i}^i\}$ which spans the difference subspace $L(B_i)$ is obtained. Then a sample x_k^i randomly selected from class i is projected on the orthonormal vector z_k^i $(k = 1, 2, \ldots, N_i - 1)$, and the summation of the projection is computed as follows:

$$\bar{x}^i = \langle x_k^i, z_1^i \rangle z_1^i + \langle x_k^i, z_2^i \rangle z_2^i + \cdots + \langle x_k^i, z_{N_i-1}^i \rangle z_{N_i-1}^i. \tag{2}$$

Then the common vector x_{common}^i of ith face class is derived as follows:

$$x_{\text{common}}^i = x_k^i - \bar{x}^i, \quad k = 1, 2, \ldots, N_i. \tag{3}$$

It was proved that the common vector x_{common}^i is unique and independent of the randomly selected sample x_k^i. Thus the common vector x_{common}^i can be used to represent the common invariant properties of ith face class. The face image x_m^i in training set is then regarded as a summation of common vector x_{common}^i of ith face class which represents the common invariant properties of ith face class, and a difference vector $x_{m,\text{diff}}^i$ which represents the specific properties of the face image x_m^i due to the specific pose and expression variations in this face image as follows:

$$x_m^i = x_{\text{common}}^i + x_{m,\text{diff}}^i. \tag{4}$$

Obviously, the common vector of each face class is useful information for classification purpose, and the difference vector should be removed from face image to eliminate the within-individual variations which can be regarded as noise effects and may deteriorate classification performance.

For query face x, the vector $x_{\text{remaining}}^i$ called remaining vector of ith face class is derived as

$$\begin{aligned} x_{\text{remaining}}^i = x - (&\langle x, z_1^i \rangle z_1^i + \langle x, z_2^i \rangle z_2^i \\ &+ \cdots + \langle x, z_{N_i-1}^i \rangle z_{N_i-1}^i. \end{aligned} \tag{5}$$

It was shown in Refs. [3,4] that the remaining vector is usually closer to common vector of its own face class than to the common vector of other face class, and therefore in recognition stage, the query face x can be assigned to class $\hat{c}$ by finding the minimum distance between the remaining vector and the common vectors.

The angle between the remaining vector of the query face and the common vector of each face class is used as distance criteria as follows:

$$\hat{c} = \min_i \left(\arccos \frac{\langle x_{\text{remaining}}^i, x_{\text{common}}^i \rangle}{\|x_{\text{remaining}}^i\| \, \|x_{\text{common}}^i\|} \right) \quad (i = 1, 2, \ldots, C). \tag{6}$$

The commonface method based on CVP is essentially classification method. The query face can be directly classified, and there is no need to extract the facial features in advance.

3. Kernel common vector approach

In this section, we further extend the CVP to kernel CVP by using kernel method in which the sample x is nonlinearly mapped to a high-dimensional feature space. The implicit nonlinear mapping is unknown and the computation is done by computing the inner product in feature space with a kernel function $k(x, y) = \Phi(x)^{\mathrm{T}} \Phi(y)$ (called kernel trick) [5]. To compute the common vectors in Eq. (3), the Gram–Schmidt orthogonalization procedure must be first performed in Eq. (2). However, since the nonlinear mapping $\Phi(x)$ is unknown, the orthogonalization procedure cannot be directly performed in feature space. To solve this problem, we first build the relation between the vectors $\Phi(b_1^i), \Phi(b_2^i), \ldots, \Phi(b_{N_i-1}^i)$ and orthonormal vectors $\Phi(z_1^i), \Phi(z_2^i), \ldots, \Phi(z_{N_i-1}^i)$ in feature space by using the QR decomposition of the matrix B_i

$$B_i = [b_1^i, b_2^i, \ldots, b_{N_i-1}^i] = Q_i R_i, \tag{7}$$

where the columns of Q_i are orthonormal vectors $\Phi(z_1^i), \Phi(z_2^i), \ldots, \Phi(z_{N_i-1}^i)$ in feature space, and the R_i is a upper triangle matrix.

To obtain R_i, the positive definite kernel matrix is computed as

$$K_{B,i} = B_i^{\mathrm{T}} B_i = R_i^{\mathrm{T}} Q_i^{\mathrm{T}} Q_i R_i = R_i^{\mathrm{T}} R_i, \tag{8}$$

where $K_{B,i}$ can be obtained using kernel function as $(K_{B,i})_{pq} = k(b_p^i, b_q^i)$.

For example $(K_{B,i})_{12} = k(x_2^i, x_3^i) - k(x_1^i, x_3^i) - k(x_2^i, x_1^i) + k(x_1^i, x_1^i)$.

It can be seen that Eq. (7) is just the Cholesky decomposition of the positive definite kernel matrix $K_{B,i}$, and thus the upper triangle matrix R_i can be obtained by performing a Cholesky decomposition of the kernel matrix $K_{B,i}$. After R_i is obtained, in Eq. (7) the relation between vectors $\Phi(b_1^i), \Phi(b_2^i), \ldots, \Phi(b_{N_i-1}^i)$ and orthonormal vectors $\Phi(z_1^i), \Phi(z_2^i), \ldots, \Phi(z_{N_i-1}^i)$ in feature space is built, though the nonlinear mapping and orthonormal vectors in feature space is still unknown.

Corresponding to Eq. (2) in feature space, the vector $\Phi(\bar{x}^i)$ is derived using Eq. (7) as follows:

$$\Phi(\bar{x}^i) = Q_i[Q_i^{\mathrm{T}}\Phi(x_k^i)] = Q_i[R_i^{-1^{\mathrm{T}}} B_i^{\mathrm{T}}\Phi(x_k^i)] \\ = Q_i[R_i^{-1^{\mathrm{T}}} B_{\bar{x},ik}], \tag{9}$$

where $Q_i = [\Phi(z_1^i), \Phi(z_2^i), \ldots, \Phi(z_{N_i-1}^i)]$, and the vector $B_{\bar{x},ik} = B_i^{\mathrm{T}}\Phi(x_k^i)$ is computed using kernel function as $(B_{\bar{x},ik})_p = k(b_p^i, x_k^i) = k(x_{p+1}^i, x_k^i) - k(x_1^i - x_k^i)$.

Thus, common vector in feature space is derived as

$$\Phi(x_{\mathrm{common}}^i) = \Phi(x_k^i) - \Phi(\bar{x}^i) \tag{10}$$

Since the nonlinear mapping is unknown, the vector $\Phi(\bar{x}^i)$ and common vector $\Phi(x_{\mathrm{common}}^i)$ cannot be explicitly computed.

Corresponding to Eq. (5) in feature space, the remaining vector $\Phi(x_{\mathrm{remaining}}^i)$ for query face x after nonlinearly mapped is derived as follows:

$$\Phi(x_{\mathrm{remaining}}^i) = \Phi(x) - \Phi(\bar{x}) = \Phi(x) - Q_i[Q_i^{\mathrm{T}}\Phi(x)], \tag{11}$$

$$\Phi(\bar{x}) = Q_i[Q_i^{\mathrm{T}}\Phi(x)] = Q_i[R_i^{-1^{\mathrm{T}}} B_i^{\mathrm{T}}\Phi(x)] \\ = Q_i[R_i^{-1^{\mathrm{T}}} B_{x,i}], \tag{12}$$

where the vector $B_{x,i} = B_i^{\mathrm{T}}\Phi(x)$ can be computed as $(B_{x,i})_p = k(x_{p+1}^i, x) - k(x_1^i, x)$.

The angle between the remaining vector of the query face and the common vector of each individual used as the decision criterion is computed by using Eqs. (10) and (11) as follows:

$$d = \min_i \left[\arccos\left(\frac{\langle \Phi(x_{\mathrm{remaining}}^i), \Phi(x_{\mathrm{common}}^i), \rangle}{\|\Phi(x_{\mathrm{remaining}}^i)\|.\|\Phi(x_{\mathrm{common}}^i)\|} \right) \right], \tag{13}$$

where

$$\langle \Phi(x_{\mathrm{remaining}}^i), \Phi(x_{\mathrm{common}}^i) \rangle \\ = k(x, x_k^i) - k(x, \bar{x}^i) - k(\bar{x}, x_k^i) + k(\bar{x}, \bar{x}^i), \tag{14}$$

$$\|\Phi(x_{\mathrm{remaining}}^i)\|^2 = k(x, x) - 2k(x, \bar{x}) + k(\bar{x}, \bar{x}), \tag{15}$$

$$\|\Phi(x_{\mathrm{common}}^i)\|^2 = k(x_k^i, x_k^i) - 2k(x_k^i, \bar{x}^i) + k(\bar{x}^i, \bar{x}^i). \tag{16}$$

The kernel functions above can be computed as follows:

$$k(x, \bar{x}) = [R_i^{-1^{\mathrm{T}}} B_{x,i}]^{\mathrm{T}} Q_i^{\mathrm{T}}\Phi(x) = B_{x,i}^{\mathrm{T}} R_i^{-1} R_i^{-1^{\mathrm{T}}} B_i^{\mathrm{T}}\Phi(x) \\ = B_{x,i}^{\mathrm{T}} K_{B,i}^{-1} B_{x,i}, \tag{17}$$

$$k(x, \bar{x}^i) = [R_i^{-1^{\mathrm{T}}} B_{\bar{x},ik}]^{\mathrm{T}} Q_i^{\mathrm{T}}\Phi(x) = B_{\bar{x},ik}^{\mathrm{T}} R_i^{-1} R_i^{-1^{\mathrm{T}}} B_i^{\mathrm{T}}\Phi(x) \\ = B_{\bar{x},ik}^{\mathrm{T}} K_{B,i}^{-1} B_{x,i}, \tag{18}$$

$$k(\bar{x}, \bar{x}) = [R_i^{-1^{\mathrm{T}}} B_{x,i}]^{\mathrm{T}} Q_i^{\mathrm{T}} Q_i [R_i^{-1^{\mathrm{T}}} B_{x,i}] \\ = B_{x,i}^{\mathrm{T}} R_i^{-1} Q_i^{\mathrm{T}} Q_i R_i^{-1^{\mathrm{T}}} B_{x,i} \\ = B_{x,i}^{\mathrm{T}} K_{B,i}^{-1} B_{x,i}, \tag{19}$$

$$k(\bar{x}, x_k^i) = [R_i^{-1^{\mathrm{T}}} B_{x,i}]^{\mathrm{T}} Q_i^{\mathrm{T}}\Phi(x_k^i) = B_{x,i}^{\mathrm{T}} R_i^{-1} R_i^{-1^{\mathrm{T}}} B_i^{\mathrm{T}}\Phi(x_k^i) \\ = B_{x,i}^{\mathrm{T}} K_{B,i}^{-1} B_{\bar{x},ik}, \tag{20}$$

$$k(\bar{x}, \bar{x}^i) = [R_i^{-1^{\mathrm{T}}} B_{x,i}]^{\mathrm{T}} Q_i^{\mathrm{T}} Q_i [R_i^{-1^{\mathrm{T}}} B_{\bar{x},ik}] \\ = B_{x,i}^{\mathrm{T}} K_{B,i}^{-1} B_{\bar{x},ik}, \tag{21}$$

$$k(x_k^i, \bar{x}^i) = [R_i^{-1^{\mathrm{T}}} B_{\bar{x},ik}]^{\mathrm{T}} Q_i^{\mathrm{T}}\Phi(x_k^i) \\ = B_{\bar{x},ik}^{\mathrm{T}} R_i^{-1} R_i^{-1^{\mathrm{T}}} B_i^{\mathrm{T}}\Phi(x_k^i) \\ = B_{\bar{x},ik}^{\mathrm{T}} K_{B,i}^{-1} B_{\bar{x},ik}, \tag{22}$$

$$k(\bar{x}^i, \bar{x}^i) = [R_i^{-1^{\mathrm{T}}} B_{\bar{x},ik}]^{\mathrm{T}} Q_i^{\mathrm{T}} Q_i [R_i^{-1^{\mathrm{T}}} B_{\bar{x},ik}] \\ = B_{\bar{x},ik}^{\mathrm{T}} K_{B,i}^{-1} B_{\bar{x},ik}. \tag{23}$$

The distance criterion (13) can be further simplified using above results.

4. Experiments

The proposed methods are tested using the standard AT&T database, which contains images from 40 individuals, each providing 10 different images. Five face images are randomly selected from each face class. Fig. 1 shows that the first selected face image in each face class (row 1), the common face for each face class (row 2) and the difference face (row 3) obtained by subtracting the common face from the respective face image in row 1. Fig. 2 shows that the face images with their respective difference faces for two face classes. The experiments verified that the common faces obtained using different samples in each face class is the same. This indicates that the common face is independent of the randomly selected sample. It can be seen from Figs. 1 and 2 that the difference face images represent the specific properties of the respective face images due to the variants such as face appearance, pose and expression variants. This suggests that the common face for each face class removes the undesired variants which may deteriorate the recognition performance.

In the experiments, we split the whole database into two parts. One part is used for training and the other part for testing. The samples are randomly selected in each face class for training and the remaining for testing. To reduce the computational complexity, the two-level wavelet decomposition is performed, and the low-frequency 28×23 subimage is selected. The proposed methods are compared with Fisherface [1] and PCA+NULL [2] methods. Each experiment is repeated 10 times and average recognition rate is computed. In each experiment the same samples are selected for training all methods and the same samples for testing all methods. The RBF kernel function $k(x, y) = \exp(-\|x - y\|^2/2000)$ is used in KCVP method. Table 1 shows the average recognition rate. It can be seen that KCVP achieves better results than CVP, and both KCVP and CVP outperform Fisherface method. When the number of training samples per class

Fig. 1. The face images (row 1), commonfaces (row 2) and difference faces (row 3).

Fig. 2. The face images in two classes and their respective difference faces.

Table 1
Average recognition rate (%)

Method	The number of training samples per class							
	2	3	4	5	6	7	8	9
Fisherface	82.06	89.71	91.50	92.50	94.12	92.83	94.37	95.00
PCA+NULL	86.91	92.57	94.92	95.30	96.94	95.75	96.62	98.00
CVP	82.72	89.21	92.04	93.60	95.87	96.67	96.75	98.25
KCVP	83.81	90.39	93.50	95.50	97.31	97.75	97.38	99.50

is greater than four and six, KCVP and CVP, respectively, achieve better performance than PCA+NULL method. As reported in Ref. [3], the larger the number of training samples is, the better the specific property of each person is removed. Therefore, the better performance is achieved in CVP and KCVP when the number of training samples per class increases. Since the common vector is unique for each face class and independent of the selected training samples, the 100% correct recognition rates are always obtained using CVP and KCVP for training samples. This is the one advantage of CVP and KCVP methods.

5. Conclusion

We presented a face recognition method using common vector approach which was further extended to kernel-based CVP. The high-dimensional face images can be directly classified without the pre-processing step to extract the facial feature. The common face derived from each face class is unique and independent of the selected samples which eliminates the specific variants in face images, and hence reveals the common invariant properties of each face class. Because of this, the 100% correct recognition rates are always achieved for training samples. With the number of training samples in each class increasing, the undesired variants in the face images is removed better, and therefore the better performances are achieved in the CVP and KCVP. The experimental results show that the proposed methods give a better representation of the common invariant properties of face image and therefore achieve good recognition performances.

References

[1] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[2] R. Huang, Q. Liu, H. Lu, S. Ma, Solving the small size problem of LDA, in: Proceedings of the 16th International Conference on Pattern Recognition, vol. 3, 2000, pp. 29–32.
[3] M.B. Gülmezoglu, V. Dzhafarov, M. Keskin, A. Barkana, A novel approach to isolated word recognition, IEEE Trans. Speech Audio Process. 7 (6) (1999) 620–628.
[4] M.B. Gülmezoglu, V. Dzhafarov, A. Barkana, The common vector approach and its relation to principal component analysis, IEEE Trans. Speech Audio Process. 9 (6) (2001) 655–662.
[5] B. Scholkopf, S. Mika, C.J.C. Burges, et al., Input space versus feature space in kernel-based method, IEEE Trans. Neural Networks 10 (5) (1999) 1000–1017.