Optica Applicata, Vol. XL, No. 3, 2010

Fisher's linear discriminant (FLD) and support vector machine (SVM) in non-negative matrix factorization (NMF) residual space for face recognition

CHANGJUN ZHOU, XIAOPENG WEI*, QIANG ZHANG, XIAOYONG FANG

Key Laboratory of Advanced Design and Intelligent Computing, Dalian University, Ministry of Education, Dalian, 116622, China

*Corresponding author: zhou-chang231@163.com

A novel method of Fisher's linear discriminant (FLD) in the residual space is put forward for the representation of face images for face recognition, which is robust to the slight local feature changes. The residual images are computed by subtracting the reconstructed images from the original face images, and the reconstructed images are obtained by performing non-negative matrix factorization (NMF) on original images. FLD is applied to the residual images for extracting FLD subspace and the corresponding coefficient matrices. Furthermore, features are obtained by mapping the residual image to FLD subspace. Finally, the features are utilized to train and test support vector machines (SVMs) for face recognition. The computer simulation illustrates that this method is effective on the ORL database and the extended Yale face database B.

Keywords: face recognition, Fisher linear discriminant (FLD), non-negative matrix factorization (NMF), residual image.

1. Introduction

Automatic face recognition plays an important role in our society and can be used in a wide range of applications, such as preventing unauthorized access or fraudulent use of ATMs, cellular phones, smart cards, workstations. The wide array of possible applications of face recognition has led to a continuous search for more precise algorithms and techniques [1–7]. However, since faces exhibit significant variations due to illuminations, pose and aging variations, a practical performance of automatic face recognition is dissatisfactory.

Principal component analysis (PCA) is a classical feature extraction and data representation technique widely used in the areas of pattern recognition and computer vision. It is also known as Karhunen–Loeve method, and was successfully applied to face recognition by [1, 2]. However, PCA just makes the scatter of all the samples as large as possible, and does not take the scatter of the samples in the same class into account. In order to overcome this drawback of PCA, BELHUMEUR *et al.* [3] proposed the FLD algorithm, which could make not only the scatter between classes as large as possible, but the scatter within a class as small as possible. And they have proved that FLD is insensitive to large variation in lighting direction in face recognition with the Harvard face database.

Based on the observation that the principal components corresponding to leading eigenvalues represent illumination variation rather than person identity [3, 8], and the residual images contain high-frequency components which are more insensitive to illumination change, Kim put forward a technology named "ICA in residual space" to avoid the bad recognition effect when the luminance and the pose in face were changed violently [9]. In their method, face features were extracted by ICA from residual images, which were obtained by subtracting reconstructed images with some primary components by ICA from the original face images, and were not sensitive to variation of luminance because it contained high frequency components, while luminance is the low frequency signal. To overcome the challenges arising from geometrical variations in face data, several local feature schemes, which represent a face image as the collection of facial component features, have been developed. LEE and SEUNG [6] proposed non-negative matrix factorization (NMF) to realize a parts-based representation because it allows only additive, not subtractive, combinations of basis components. Since a parts-based representation can naturally deal with partial occlusion and some illumination problems, it has received much attention recently.

In this paper, non-negative matrix describes the residual spaces of local facial components. It is well known that the first few eigenfaces represent illumination variation rather than identity, and most methods based on PCA have overcome illumination variation by discarding the projection to a few leading eigenfaces [9]. The space spanned after removing a few leading eigenfaces is called the "residual space". In these spaces, a "residual face space" is represented by a collection of non-negative features to achieve robust recognition to the slight local feature changes. Furthermore, FLD is applied to the residual images for extracting FLD subspace and the corresponding coefficient matrices. Finally, the features are utilized to train and test support vector machines (SVM) for face recognition.

2. Relevant works

2.1. NMF algorithms

Denote the training set of m face images by $V = [v_1, v_2, ..., v_m]$, $V \subset R^{n \times m}$, each column of V contains n pixel values of one face image, and each image belongs to c classes. NMF is a linear, non-negative approximate data representation [6], which refers to the decomposition of the matrix V into two matrices $W \subset R^{n \times r}$ $(r < m)$ (basis images) and $H \subset R^{r \times m}$ $(r < m)$ (encoding coefficients), such that [6]

$$V \approx WH \tag{1}$$

NMF uses the divergence of V from Y, defined as:

$$D(V \| Y) = \sum_{i,j}\left(v_{ij}\log\frac{v_{ij}}{y_{ij}} - v_{ij} + y_{ij}\right) \tag{2}$$

where $WH = Y = [y_{ij}]$. An NMF is defined as

$$\min_{H,\,W} D(V \| WH)$$

$$\text{s.t.}\, W, H \geq 0 \tag{3}$$

$$\sum_{i=1}^{n} w_{ij} = 1 \forall j$$

This optimization can be done by using multiplicative update rules [6]:

$$W_{i\alpha} = W_{i\alpha}\sum_{\mu}\frac{V_{i\mu}}{(WH)_{i\mu}}H_{\alpha\mu} \tag{4}$$

$$W_{i\alpha} = \frac{W_{i\alpha}}{\sum_{j} W_{j\alpha}} \tag{5}$$

$$H_{\alpha\mu} = H_{\alpha\mu}\sum_{i} W_{i\alpha}\frac{V_{i\mu}}{W_{i\mu}} \tag{6}$$

2.2. Fisher's linear discriminant

FLD is a famous classifying method, which could make the scatter of samples more reliable for classification. Given sample $V = [v_1, v_2, ..., v_m]$, $V \subset R^{n\times m}$, each column of V contains n pixel values of one face image, and each image belongs to c classes. The between-class scatter matrix and within-class scatter matrix are defined as [3]:

$$S_b = \sum_{j=1}^{c} N_j(\mu_j - \mu_0)(\mu_j - \mu_0)^T \tag{7}$$

$$S_w = \sum_{j=1}^{c}\sum_{V_i \in V_j}(V_i - \mu_j)(V_i - \mu_j)^T \tag{8}$$

where S_b is the between-class scatter matrix and S_w is the within-class scatter matrix; μ_j is the mean vector of class V_j, and N_j is the number of samples in class V_j; μ_0 is the mean vector of all the samples $V = [v_1, v_2, ..., v_m]$.

If S_w is nonsingular, the optimal projection matrix W_{opt} is chosen as [3]:

$$W_{\text{opt}} = \arg\max_{W} \frac{\left|W^T S_b W\right|}{\left|W^T S_w W\right|} = \left[W_1 \; W_2 \; \dots \; W_m\right] \tag{9}$$

where $\{W_i \mid i = 1, 2, ..., m\}$ is the set of generalized eigenvectors of S_b and S_w corresponding to the m largest generalized eigenvalues $\{\lambda_i \mid i = 1, 2, ..., m\}$, *i.e.*,

$$S_b W_i = \lambda_i S_w W_i, \quad i = 1, 2, ..., m \tag{10}$$

If S_w is singular, W_{opt} is chosen as [3]:

$$W_{\text{opt}}^T = W_{\text{fld}}^T W_{\text{pca}}^T \tag{11}$$

$$W_{\text{pca}} = \arg\max_{W} \left|W^T S_t W\right| \tag{12}$$

$$W_{\text{fld}} = \arg\max_{W} \frac{\left|W^T W_{\text{pca}}^T S_b W_{\text{pca}} W\right|}{\left|W^T W_{\text{pca}}^T S_w W_{\text{pca}} W\right|} \tag{13}$$

where S_t is the total scatter matrix, defined by:

$$S_t = \sum_{i=1}^{m} (V_i - \mu)(V_i - \mu)^T \tag{14}$$

2.3. Support vector machine

The SVM method achieves an optimal linear classifier in feature space, which is theoretically based on structural risk minimization (SRM) theory of statistical learning. It has been shown to provide higher performance than traditional learning machines and has been introduced as powerful tools for both pattern recognition and regression estimation problems [10]. This SVM learning possesses a linear classifier with minimum VC-dimension (machine complexity), thereby keeps low expected generalization errors. And for linear non-separable data, SVM uses a device called kernel mapping to map the data in input space to a high-dimensional feature space in which the problem becomes linear. Then the optimal decision function can be written as [11]

$$f(x) = \operatorname{sgn}\left\{ \sum_{i=1}^{m} y_i \alpha_i K(x_i, x) + b \right\} \tag{15}$$

where $K(x_i, x)$ is the kernel function. The most commonly used kernel functions are polynomial, radial basis function (RBF) and sigmoid functions, as follows [11]

– Polynomial kernel: $K(x, y) = \left(\gamma\langle x, y\rangle + b\right)^d$;

– RBF kernel: $K(x, y) = \exp\left(-\gamma\|x - y\|^2\right)$;

– Sigmoid kernel: $K(x, y) = \tanh\left(\gamma\langle x, y\rangle + b\right)$.

Previous content describes the basic theory of SVM for two class classification, and a multi-class pattern recognition system can be obtained based on the dichotomy SVMs. Usually there are two schemes for this purpose [12]. One is the one-against-all strategy to classify between each class and all the remaining; the other is the one--against-one strategy to classify between each pair. While the former often leads to ambiguous classification [13], we adopt the latter one for our face recognition system. And then, we employ RBF kernel function as a classifier in high-dimensional feature space, which leads to gratifying experimental results in the following experiments.

3. Face recognition based on residual space

Because illumination is the low frequency signal, and the residual images computed by subtracting the reconstructed images from the original face images contain the high frequency components of face, so the representation of residual images for face recognition is robust to the illuminate changes [14]. We found that NMF in the residual face space provides more efficient encoding in terms of redundancy reduction and robustness to slight local feature changes as well as illumination variation, owing to its ability to realize a parts-based representation. And then, a novel method of FLD + SVM in the NMF residual space is put forward in this paper. FLD is applied in facial residual space to gain the features, and the features are utilized to train and test SVMs for face recognition.

3.1. Preprocessing for face images

Before extracting features, we usually organize each image in the database as a long row vector, the dimension of which equals the number of pixels in the image. And then, we will do some preprocessing on the vector. It is very useful and necessary in face recognition.

The most basic preprocessing is to make X zero mean, by subtracting X by its mean. This course is called "centering" [15]. This processing is made only to simplify the algorithm. Another useful preprocessing is called "whitening" [15]. In this processing, X is multiplied by $ED^{-1/2}E^T$, where E is the orthogonal matrix of eigenvectors of $E\{XX^T\}$ and D is the diagonal matrix of its eigenvalue. By transforming linearly the observed matrix X, we obtain a new matrix $\overline{X}$ whose com-

Fig. 1. Original images (**a**), reconstructed images (**b**), and residual images (**c**).

ponents are uncorrelated and whose variances equal unity, as $E\{\overline{X}\ \overline{X}^{T}\}$. The utility of whitening resides in the fact that new mixing matrix for $\overline{X}$ is orthogonal. In the rest of this paper, we assume that all image data have been preprocessed by centering and whitening.

3.2. Residual image based on NMF

After the normalized images have been obtained, we used them to gain the residual image. We used NMF to learn the basis images of the training set by the update rules described by Eqs. (4)–(6). Furthermore, features are extracted by mapping the testing image X to basis images W_j by Eq. (16) and they are utilized to reconstruct the images by Eq. (17), and Y is the reconstructed image

$$H = W^{-1}X \tag{16}$$

$$Y = WH \tag{17}$$

While the reconstructed face image with a few independent components lose the details and look like low-pass-filtered versions, the corresponding residual image contains high-frequency components and are less sensitive to slight local feature variation, and $\overline{X}$ is the residual image

$$\overline{X} = (X - Y)^{T} \tag{18}$$

Since these residual images still contain rich information for the individual identities, face features are extracted from these residual faces instead of the original faces. Figure 1 shows some original images, corresponding reconstructed images and residual images.

3.3. FLD + SVM in NMF residual space for face recognition

In this method, after the residual images are obtained by performing NMF on the original images, the FLD basic images and the corresponding coefficient matrix are obtained from the residual images. Furthermore, the features are utilized to train and test SVMs for face recognition. Figure 2 shows the flow chart of FLD + SVM in NMF residual space for face recognition.

The major steps in the face recognition are as follows.

Step 1. Randomly partitioned into a training set V_{train} from face database and the rest is a testing set V_{test};

Step 2. Use NMF to learn the basis images W_{nmf} of the training set V_{train} by the update rules described by Eqs. (4)−(6);

Step 3. Obtain the encoding coefficients H_{train} based on Eq. (16), and then reconstruct image Y_{train} based on Eq. (17);

Step 4. Subtract the reconstruction image Y_{train} from the original image X_{train} to get the residual image $\overline{X}_{\text{train}}$ by Eq. (9);

Step 5. Repeat Step 4 to get the residual images $\overline{V}_{\text{train}}$;

Step 6. Perform FLD in residual images $\overline{X}_{\text{train}}$, compute FLD basic images W_{fld} and coefficient matrix B_{train} by Eqs. (11)−(13);

Step 7. Map the test set V_{test} to W_{nmf} for extract features H_{test};

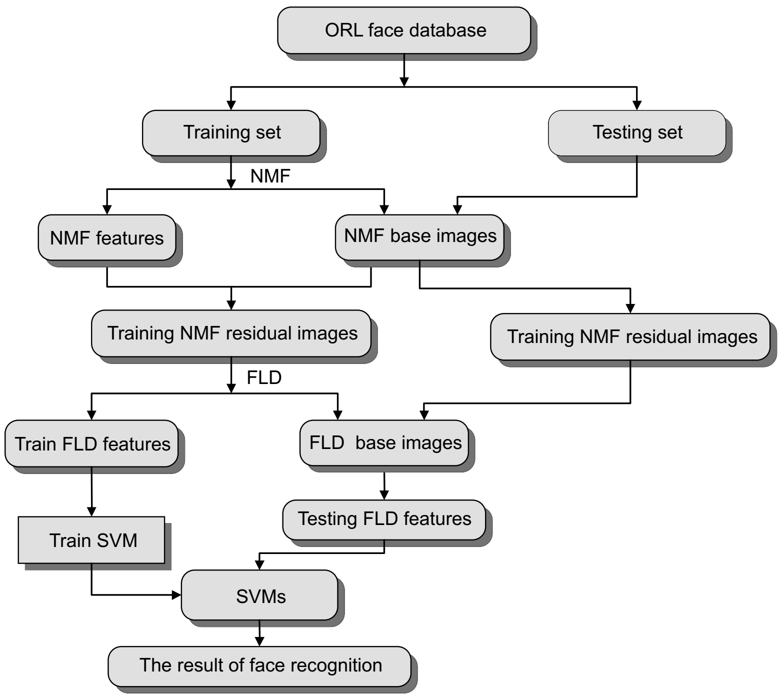

Fig. 2. The flow chart of FLD + SVM in NMF residual space for face recognition.

Fig. 3. Some sample images from two databases. ORL database (**a**), the extended Yale B database (**b**).

Step 8. Reconstruct image Y_{test} from the test set V_{test} and then get the residual images;

Step 9. Map the test set V_{test} to W_{fld} for extract features B_{test};

Step 10. Utilize the train set's features B_{train} to training SVMs;

Step 11. Utilize the test set's features B_{test} and SVMs to recognize a face.

4. Experiments

In order to validate the performance of our method, we carried out experiments in comparison with other schemes on two independent face databases (Fig. 3 shows some sample images from these databases), the ORL face database [16] and the extended Yale face database B [17]. Our experiments were carried out on a P4 3.0 GHz PC machine with 1 G memory, and the SVM classifier we adopted was OSU_SVM [18] which implements a Matlab interface to LIBSVM [19].

4.1. Experiment using the ORL face database

The first set of experiments was carried out using the ORL database. The ORL face database was composed of 400 images of size 112×92. There were 40 persons, 10 images for each person. The images were taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses). All the images were taken against a dark homogeneous background. The faces were in up-right position of frontal view, with slight left-right out-of-plane rotation. Each image was linearly stretched to the full range of pixel values of [0, 255]. On each benchmark, we reduced the face images size from 112×92 to 24×24 for efficiency by using the general function of Matlab, imresize.

Each set of 10 images of a person was randomly partitioned into a training set of five images and a testing set of the other five images. The training set was used for

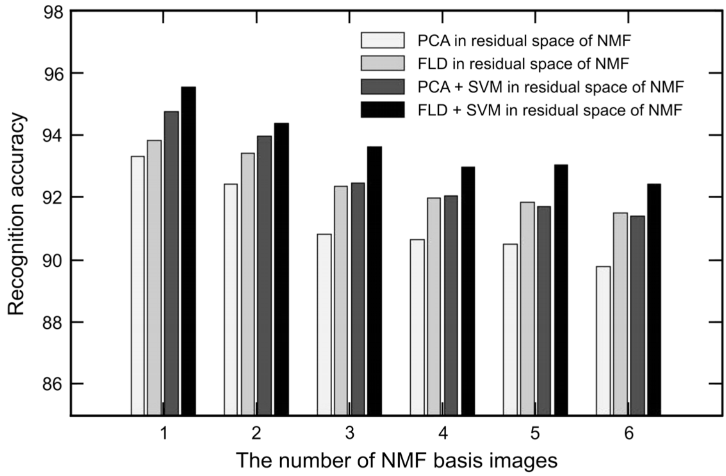

Fig. 4. Face recognition on ORL database versus different number of NMF basis images.

calculating basis images, which was for obtaining residual images and their features. Afterward, we used two ways as forenamed for face recognition by using these features. Such procedures were repeated for fifty times, and their results were fifty groups of data. For each group, we calculated the reconstruction images by NMF versus the number of basis images 1, 2, 3, 4, 5, 6 and got the residual images by subtracting the reconstruction images from the original images. And then, the features had been extracted by FLD and PCA versus the number of basis images 30 from the residual images, which were utilized to realize the face recognition based on the nearest neighbor (NN) classification and SVM classification. The results are shown in Fig. 4.

It can be seen from Fig. 4, with the ascending of the number of NMF basis images, that the recognition accuracy descends. The maximal recognition accuracy was obtained when the number of NMF basis images is 1, and so that the number "1" was selected as the number of NMF basis images. Furthermore, we also compared

T a b l e 1. Comparison of our method and other methods on the ORL database.

	Recognition accuracy [%]	
Recognition method	Max	Average
NMF [6]	92.12	89. 67
FLD [4]	97.50	93.82
FLD + SVM	97.70	94.24
ICA in residual space of ICA [9]	95.50	90.41
NMF in residual space of ICA [20]	96.00	91.25
FLD in residual space of NMF	97.80	94.38
PCA in residual space of NMF	97.00	93.31
PCA + SVM in residual space of NMF	98.20	94.75
FLD + SVM in residual space of NMF	99.00	95.56

our method with other published face recognition methods on the ORL database. The results are shown in Tab. 1.

It can be seen from the experimental results that the average recognition accuracy rate of FLD in residual space of NMF is 94.38% and that of FLD + SVM in residual space of NMF is 95.56%, while that of traditional NMF is 89.67%, that of traditional FLD is 93.82% and that of NMF in residual space of ICA is 91.25%. The recognition accuracy of our method is much higher than that of NMF and that of traditional FLD and a little higher than that of other residual space methods under the same experiment conditions. The experiment results indicate that FLD for face recognition based on NMF residual images method improve the recognition rate upon the other face recognition methods with residual space.

4.2. Experiment using the extended Yale face database B

The extended Yale face database B was first reported by GEORGHIADES *et al.* [21], and now it is composed of original image database and cropped image database. The cropped image database was reported by LEE *et al.* [17]. In this database, there were 38 persons, 65 images for each person, all image data were manually aligned, cropped, and then resized to images 168×192. In addition, because an image with ambient (background) illumination was captured in this database and some images were destroyed in download samples, we removed some images form the database and reduced the images number of each person to 60 for our experiment. And then, on each benchmark, we reduced the face images size from 168×192 to 42×48 for efficiency by using the general function of Matlab, imresize.

In the same way as on ORL face database, each set of sixty images of a person was randomly partitioned into a training set of thirty images and a testing set of the other thirty images. The face recognition progress was repeated for fifty times, the number of NMF basis image was 1 and the feature dimension of FLD was 30. Correspondingly, the basis images' numbers of PCA and ICA were 30, too. For comparison, we showed the results of our method with other published face recognition methods in the same

Table 2. Comparison of our method and other methods on the extended Yale face database B.

	Recognition accuracy [%]	
Recognition method	Max	Average
NMF [6]	90.88	88.94
FLD [4]	91.67	89.29
FLD + SVM	95.00	93.04
ICA in residual space of ICA [9]	75.25	73.07
NMF in residual space of ICA [20]	89.76	88.51
PCA in residual space of NMF	78.86	76.19
PCA + SVM in residual space of NMF	93.77	91.55
FLD in residual space of NMF	91.84	88.95
FLD + SVM in residual space of NMF	95.53	93.34

table. All the compared methods used the same training and testing data. The results are shown in Tab. 2, which indicate that FLD + SVM in residual space of NMF for face recognition is effective, and our method has advantage in recognition rate when compared with other tradition face recognition methods.

It can be seen from the experimental results that FLD + SVM in NMF residual space for face recognition is better than other methods, because the representation of "residual space" for face recognition is robust to the slight local feature changes. In reference [9], the authors proposed an "ICA in the residual face space" method for face recognition, and testified it is robust to illumination and pose variation, because of the ability of ICA to represent non-Gaussian statistics. NMF is distinguished from the other subspace methods by its use of non-negativity constraints. These constraints lead to a parts-based representation because they allow only additive, not subtractive, combinations [6]. For these reasons, the NMF residual space is more robust to the slight local features changes of face image.

In this paper, we used NMF for residual face space, FLD was adopted to extract features from this space, and SVM was classified for face recognition. From the results we can see, for the cropped image of extended Yale face database B, that there is no clear improvement between FLD + SVM in NMF residual space and FLD + SVM. However, for the ORL database, the recognition rate advancement of our method is significant. The experimental results show that the accuracy of face recognition is significantly improved by the proposed method under slight local features changes.

5. Conclusions

Based on the residual space of NMF, a novel method of FLD and SVM is put forward. It is testified that the representation of "residual space" for face recognition is robust to illumination and pose changes. In this method, the residual images are obtained by performing NMF on the original image, which is more robust to the local feature changes. And then, the FLD basic images and the corresponding coefficient matrix are obtained from the residual images. Furthermore, the features are utilized to train and test SVMs for face recognition. The experiment results on ORL face database and the extended Yale face database B show that the proposed method based on FLD in facial residual space achieves better recognition as compared with the other methods based on residual space.

Acknowledgements – This work is supported by the Program for Study of Science of the Educational Department of Liaoning Province (No. 05L020) and the Program for Dalian Science and Technology (No. 2005A10GX106).

References

[1] KIRBY M., SIROVICH L., *Application of the Karhunen–Loeve procedure for the characterization of human faces*, IEEE Transactions on Pattern Analysis and Machine Intelligence **12**(1), 1990, pp. 103–108.

[2] TURK M.A., PENTLAND A.P., *Eigenfaces for recognition*, Journal of Cognitive Neuroscience **3**(1), 1991, pp. 71–86.

[3] BELHUMEUR P.N., HESPANHA J.P., KRIEGMAN D.J., *Eigenfaces vs. Fisherfaces: recognition using class specific linear projection*, IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7), 1997, pp. 711–720.

[4] AN G.Y., RUAN Q.Q., *A novel mathematical model for enhanced Fisher's linear discriminant and its application to face recognition*, The 18th International Conference on Pattern Recognition, 2006.

[5] COMON P., *Independent component analysis. A new concept?*, Signal Processing **36**(3), 1994, pp. 287–314.

[6] LEE D.D., SEUNG H.S., *Learning the parts of objects by non-negative matrix factorization*, Nature **401**, 1999, pp. 788–791.

[7] WANG Y., JIA Y.D., HU C.B., TURK M., *Fisher non-negative matrix factorization for learning local features*, Asian Conference on Computer Vision, Korea, 2004.

[8] DUDA R.O., HART P.E., STORK D.G., *Pattern Classification*, Wiley-Interscience, New York, 2001.

[9] KIM T.K., KIM H., HWANG W., KITTLER J., *Independent component analysis in a facial local residue space*, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recogniton, 2003, pp. 1063–1069.

[10] BURGES C.J.C., *A tutorial on support vector machines for pattern recognition*, Data Mining and Knowledge Discovery **2**(2), 1998, pp. 121–167.

[11] CRISTIANINI N., TAYLOR J.S., *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, 2004, pp. 82–108.

[12] VAPNIK V.N., *Statistical Learning Theory*, Wiley, New York, 1998.

[13] PONTIL M., VERRI A., *Support vector machines for 3-D object recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(6), 1998, pp. 637–646.

[14] CHEN W., MENG JOO ER, SHIQIAN WU, *Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithmic domain*, IEEE Transactions on Systems, Man, and Cybernetics, Part B **36**(2), 2006, pp. 458–466.

[15] XU Y.Q., LI B.C., WANG B., *Face recognition by fast independent component analysis and genetic algorithm*, Proceedings of the Fourth International Conference on Computer and Information Technology (CIT'04), 2004

[16] SAMARIA F., HARTER A., *Parameterisation of a stochastic model for human face identification*, 2nd IEEE Workshop on Applications of Computer Vision, Sarasota, FL, 1994.

[17] LEE K.C., HO J., KRIEGMAN D.J., *Acquiring linear subspaces for face recognition under variable lighting*, IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(5), 2005, pp. 684–698.

[18] MA J., ZHAO Y., AHALT S., *OSU SVM Classier Matlab Toolbox (ver 3.00)*, http://eewww.eng.ohiostate.edu/maj/osu svm, 2002.

[19] CHANG C.C., LIN C.J., *LIBSVM: a library for support vector machines*, http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001.

[20] WANG C.J., ZHANG Q., WEI X.P., ZHOU C.J., *Face recognition based on NMF in a residual image*, Proceedings of the 4th International Conference on Impulsive and Hybrid Dynamical Systems, 2007, pp. 2125–2128.

[21] GEORGHIADES A.S., BELHUMEUR P.N., KRIEGMAN D.J., *From few to many: Illumination cone models for face recognition under variable lighting and pose*, IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(6), 2001, pp. 643–660.

Received July 1, 2009
in revised form October 26, 2009