[25] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro, "MCYT baseline corpus: A bimodal biometric database," in *IEE Proc. Vision, Image Signal Process.*, vol. 150, 2003, pp. 395–401.

[26] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of decision task performance," in *Proc. ESCA Eur. Conf. Speech Comm. Tech., EuroSpeech*, 1997, pp. 1895–1898.

Hallucinating Face by Eigentransformation

Xiaogang Wang and Xiaoou Tang

***Abstract*—In video surveillance, the faces of interest are often of small size. Image resolution is an important factor affecting face recognition by human and computer. In this paper, we propose a new face hallucination method using eigentransformation. Different from most of the proposed methods based on probabilistic models, this method views hallucination as a transformation between different image styles. We use Principal Component Analysis (PCA) to fit the input face image as a linear combination of the low-resolution face images in the training set. The high-resolution image is rendered by replacing the low-resolution training images with high-resolution ones, while retaining the same combination coefficients. Experiments show that the hallucinated face images are not only very helpful for recognition by humans, but also make the automatic recognition procedure easier, since they emphasize the face difference by adding more high-frequency details.**

***Index Terms*—Eigentransformation, face hallucination, face recognition, principal component analysis, super-resolution.**

I. Introduction

In video surveillance, the faces of interest are often of small size because of the large distance between the camera and the objects. Image resolution becomes an important factor affecting face recognition performance. Since many detailed facial features are lost in the low-resolution face images, the faces are often indiscernible. For identification, especially by humans, it is useful to render a high-resolution face image from the low-resolution one. This technique is called face hallucination or face super-resolution [1].

The simplest way to increase image resolution is a direct interpolation of input images with such algorithms as nearest neighbor or cubic spline. However, the performance of direct interpolation is usually poor since no new information is added in the process. A number of super-resolution techniques have been proposed in recent years [2]–[13]. Most try to produce a super-resolution image from a sequence of low-resolution images [2], [3]. Some other approaches [5], [6], [8]–[10], [12], [14], [15] are based on learning from the training set containing high- and low-resolution image pairs, with the assumption that high-resolution images are Markov random fields (MRFs) [5], [9], [13]. These methods are more suitable for synthesizing local texture, and are usually applied to generic images without special consideration of the property of face images.

Baker and Kanade [1], [11], [16] developed a hallucination method based on the property of face images. Abandoning the MRF assumption, it infers the high-frequency components from a parent structure by recognizing the local features from the training set. Liu*et al.* [4] developed a two-step statistical modeling approach integrating global and local parameter models. Both of the two methods use complicated probabilistic models and are based on an explicit resolution reduction function, which is sometimes difficult to obtain in practice.

Since face images are well structured and have similar appearances, they span a small subset in the high dimensional image space [17], [18]. In the study by Penev and Sirovich [19], face images are shown to be well reconstructed by Principal Component Analysis (PCA) representation with 300–500 dimensions. Zhao*et al.* [20] showed that the dimensionality of face space is insensitive to image size. Moghaddam [21] downsampled face images to 12 by 21 pixels and still achieved 95% recognition accuracy on $1800+$ face images from the FERET database. These studies imply that facial components are highly correlated and the high-frequency details of face images may be inferred from the low-frequency components, utilizing the face structural similarities.

Instead of using a probabilistic model, we propose a face hallucination method using PCA to represent the structural similarity of face images. The algorithm treats the hallucination problem as the transformation between two different image styles. This method is closely related to the work in [22], [23], in which a style transformation approach was applied to transform a photo into a sketch. In a similar way, we could transform face images from low-resolution to high-resolution based on mapping between two groups of training samples without deriving the transformation function [24]. The hallucinated face image is rendered from the linear combination of training samples. Using a small training set, the method can produce satisfactory results.

Hallucination can effectively improve the resolution of a face image, thus making it much easier for a human being to recognize a face. However, how much information has been extracted from the low-resolution image by the hallucination process and its contribution to automatic face recognition has not been studied before. In our method, PCA is applied to the low-resolution face image. In the PCA representation, different frequency components are uncorrelated. By selecting the number of eigenfaces, we could extract the maximum amount of facial information from the low-resolution face image and remove the noise. We also study the face recognition performance using different image resolutions. For automatic recognition, a low resolution bound is found through experimentation. We find that hallucination may help the automatic recognition process, since it emphasizes the face difference by adding some high-frequency details.

II. Hallucination by Eigentransformation

A. Multiresolution Analysis

Viewing a two-dimensional (2-D) image as a vector, the process of getting a low-resolution face image from the high-resolution one can be formulated as

$$\overline{I}_l = H\,\overline{I}_h + \overline{n}\ . \tag{1}$$

Here, $\overline{I}_h$ is the high-resolution face image vector to be rendered, with length N as the total pixel number. $\overline{I}_l$ is the observed low-resolution face image vector with length s^2N, where s is the downsampling factor

Manuscript received March 15, 2003; revised July 13, 2004. This work was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project no. CUHK 4190/01E, CUHK 4224/03E, AOE/E01/99, and N_CUHK409/03). This paper was recommended by Guest Editor D. Zhang.

The authors are with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: xgwang1@ie.cuhk.edu.hk; xtang@ie.cuhk.edu.hk).

Digital Object Identifier 10.1109/TSMCC.2005.848171

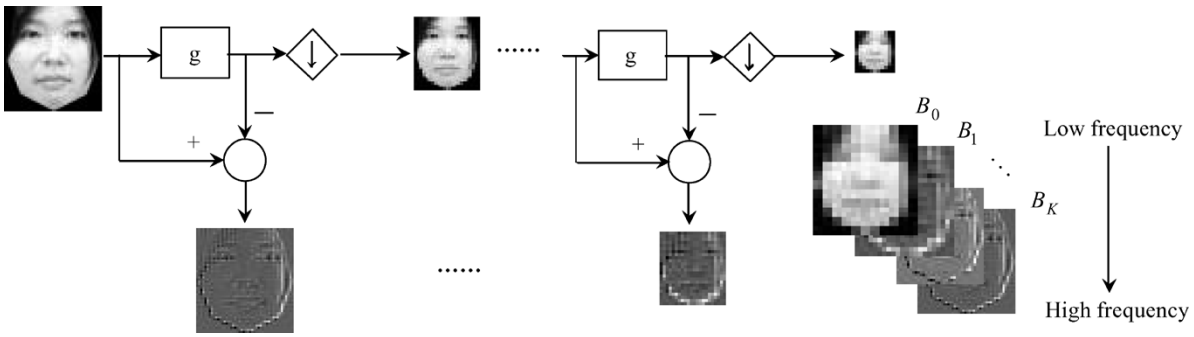

Fig. 1. Multiresolution analysis in spatial domain, where g is the smoothing function, and $B_0, \ldots, B_K$ are different frequency bands.

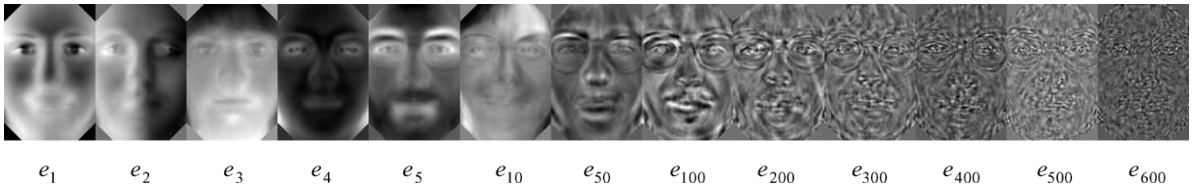

Fig. 2. Eigenface examples sorted by eigenvalues, where e_i is the ith eigenface.

$(0 < s < 1)$. H is the transformation matrix involving blurring and the downsampling process. The blurring operation can be estimated from the point-spread function of the camera. In practice, it is often simplified as a Gaussian filter. The term $\overline{n}$ represents the noise disturbance to the low-resolution face image captured by the camera. A detailed discussion on the super-resolution reconstruction constraints can be found in [11].

The hallucination problem can be discussed under the framework of multiresolution analysis. As shown in Fig. 1, a process of iterative smoothing and downsampling decomposes the face image into different bands, $B_0, \ldots, B_K$. The low-frequency component is encoded in the downsampled low-resolution image, and the difference between the original face image and the smoothed image contains the high-frequency detail. In this decomposition, different frequency bands are not independent. Some components of the high-frequency bands $B_1, \ldots, B_K$ can be inferred from the low-frequency band B_0. This is a starting point for hallucination. Many super-resolution algorithms assume the dependency as homogeneous MRFs, i.e., the pixel relies only on the pixels in its neighborhood. This is an assumption for general images. It is not optimal for the face class without considering faces structural similarities. A better way to address the dependency is using PCA, in which different frequency components are uncorrelated.

Many studies [19], [25] on face space have shown that a face image can be reconstructed from eigenfaces in the PCA representation. Like the multiresolution analysis, PCA also decomposes the face image into different frequency components. The difference is that the PCA method utilizes the face distribution to decompose face structure into uncorrelated frequency components; thus, it can encode face information more concisely. Our algorithm first employs PCA to extract as much useful information as possible from a low-resolution face image, and then renders a high-resolution face image by eigentransformation.

B. Principal Component Analysis

PCA represents face images using a weighted combination of eigenfaces. We represent a set of face images by an N by M matrix, $[\overline{l}_1, \ldots, \overline{l}_M]$, where $\overline{l}_i$ is the image vector, N is the number of image pixels, and M is the number of the training samples $(N \gg M)$. A mean face is computed as

$$\overline{m}_l = \frac{1}{M}\sum_{i=1}^{M} \overline{l}_i \,. \tag{2}$$

Removing the mean face from each image, we have

$$L = [\overline{l}_1 - \overline{m}_l, \ldots, \overline{l}_M - \overline{m}_l] = \left[\overline{l}'_1, \ldots \overline{l}'_M\right]. \tag{3}$$

A set of eigenvectors, also called eigenfaces, are computed from the eigenvectors of the ensemble covariance matrix

$$C = \sum_{i=1}^{M} (\overline{l}_i - \overline{m}_l)(\overline{l}_i - \overline{m}_l)^T = LL^T. \tag{4}$$

Directly computing the eigenvectors of C is not practical because of the large size of the matrix. Alternatively, the eigenvectors of a smaller matrix $R = L^T L$ can be first computed [26]

$$(L^T L)V_l = V_l \Lambda_l \tag{5}$$

where V_l is the eigenvector matrix and Λ_l is the eigenvalue matrix. Multiplying both sides by L, we have

$$(LL^T)LV_l = LV_l \Lambda_l. \tag{6}$$

Therefore, the orthonormal eigenvector matrix of $C = LL^T$ can be computed from

$$E_l = LV_l \Lambda_l^{-\frac{1}{2}}. \tag{7}$$

For a face image $\overline{x}_l$, a weight vector can be computed by projecting it onto the eigenfaces:

$$\overline{w}_l = E_l^T(\overline{x}_l - \overline{m}_l). \tag{8}$$

This is a face representation based on eigenfaces. A face can be reconstructed from K eigenfaces, $E_l = [e_1, \ldots, e_K]$

$$\overline{r}_l = E_l\, \overline{w}_l + \overline{m}_l \,. \tag{9}$$

Similar to other multiscale analysis methods, PCA also decomposes face images into different frequency components. Fig. 2 shows some

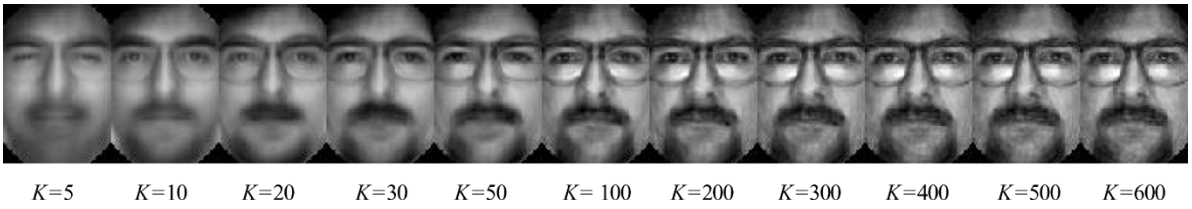

Fig. 3. Face reconstruction using a different number of eigenfaces. K is the eigenface number.

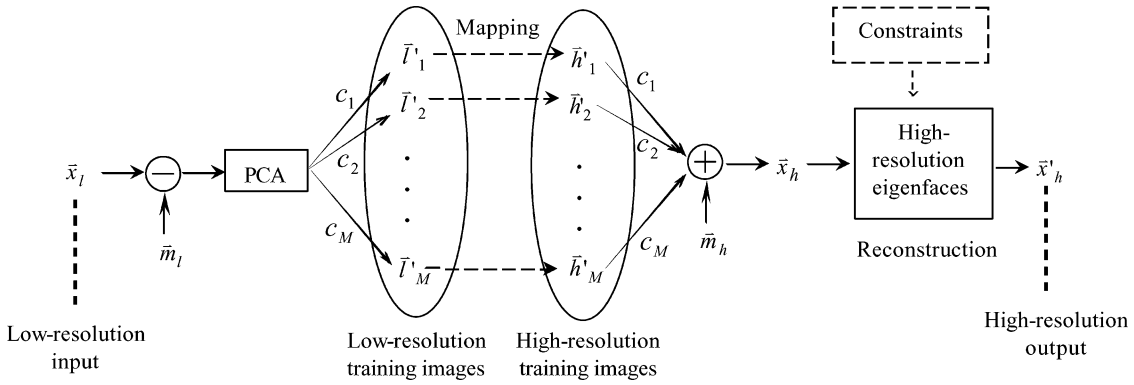

Fig. 4. System diagram using eigentransformation for hallucination.

eigenfaces that are sorted by eigenvalues. Eigenfaces with large eigenvalues are "face-like," and characterize the low-frequency components. Eigenfaces with small eigenvalues are "noise-like," and characterize the high-frequency details. PCA is optimal for the face representation because the K largest eigenfaces account for most of the energy and are most informative for the face image set. The eigenface number K controls the detail level of the reconstructed face. An example is shown in Fig. 3. As K increases, more details are added to the reconstructed face. Different from other multiscale analysis, in PCA, the frequency components are computed by decorrelation based on face structure. This uncorrelation property is important for the success of the eigentransformation algorithm.

C. Eigentransformation

For eigentransformation, we use a training set containing low-resolution face images, and the corresponding high-resolution face images. Following the previous discussion, let $[\overline{l}_1, \ldots \overline{l}_M]$ represent the low-resolution training face image set, from which eigenfaces are computed. Apply PCA to the input low-resolution face image $\overline{x}_l$. From (7) and (9), the reconstructed face image can be represented by

$$\overline{r}_l = LV_l\Lambda_l^{-\frac{1}{2}}\,\overline{w}_l + \overline{m}_l = L\,\overline{c} + \overline{m}_l \tag{10}$$

where $\overline{c} = V_l\Lambda_l^{-1/2}\,\overline{w}_l = [c_1, c_2, \ldots, c_M]^T$. Equation (10) can be rewritten as

$$\overline{r}_l = L\,\overline{c} + \overline{m}_l = \sum_{i=1}^{M} c_i\,\overline{l}'_i + \overline{m}_l\,. \tag{11}$$

This shows that the input low-resolution face image can be reconstructed from the optimal linear combination of the M low-resolution training face images. Here, $\overline{c}$ describes the weight that each training face contributes in reconstructing the input face. The sample face that is more similar to the input face has a greater weight contribution. Replacing each low-resolution image $\overline{l}'_i$ by its corresponding high-resolution sample $\overline{h}'_i$, and replacing $\overline{m}_l$ with the high-resolution mean face $\overline{m}_h$, we have

$$\overline{x}_h = \sum_{i=1}^{M} c_i\,\overline{h}'_i + \overline{m}_h\,. \tag{12}$$

$\overline{x}_h$ is expected to be an approximation to the real high-resolution face image.

$\overline{x}_h$ should meet two necessary conditions in order to adequately approximate the original high-resolution face image. First, after resolution reduction of $\overline{x}_h$, the output should produce the low-resolution input face image. Second, $\overline{x}_h$ should be face-like at the high-resolution level. The first condition can be proved easily. From (1), without considering the noise disturbance, the transformation between the high-resolution face image and the low-resolution face image can be approximated as a linear operation [11]. For the training set, we have

$$\overline{l}'_i = H\,\overline{h}'_i \tag{13}$$

$$\overline{m}_l = H\,\overline{m}_h\,. \tag{14}$$

From (11) and (12), replacing $\overline{l}'$ and $\overline{m}_l$ with (13) and (14), we have

$$\overline{r}_l = \sum_{i=1}^{M} c_i H\,\overline{h}'_i + H\,\overline{m}_h = H\left(\sum_{i=1}^{M} c_i\,\overline{h}'_i + \overline{m}_h\right) = H\,\overline{h}\,. \tag{15}$$

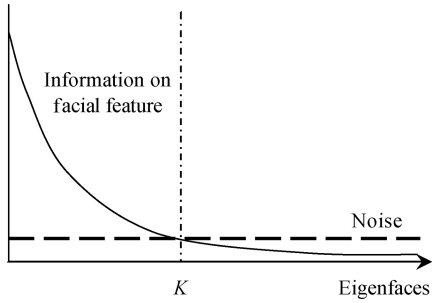

Fig. 5. Extract facial information in the PCA space of low-resolution face images.

Fig. 6. Hallucinated face images by eigentransformation. (a) Input 23 × 25; (b) Cubic B-spline; (c) hallucinated; (d) original 117 × 125.

Since $\overline{r}_l$ is an optimal reconstruction to the input low-resolution face image $\vec{x}_l$, $\overline{x}_h$ leads to a good approximation to $\vec{x}_l$ after resolution reduction. For the second condition, (12) shows that $\overline{x}_h$ is the linear combination of high-resolution face images, so it should be approximately face-like at a high-resolution level. Of course, some non-face-like distortion may be involved, since the combination coefficient c_i is not computed from the high-resolution training data. We can reduce these non-face-like distortions by reconstructing $\overline{x}_h$ from the high-resolution eigenfaces. Let E_h and $\Lambda_h = \text{diag}(\lambda_1, \ldots, \lambda_K)$ be the eigenface and eigenvalue matrixes computed from the high-resolution training images. The principal components of $\overline{x}_h$ projecting on the high-resolution eigenfaces are

$$\overline{w}_h = E_h^T(\overline{x}_h - \overline{m}_h). \tag{16}$$

The eigenvalue λ_i is the variance of high-resolution face images on the ith eigenface. If the principal component $w_h(i)$ is much larger than λ_i, non-face-like distortion may be involved for the ith eigenface dimension. To reduce the distortion, we apply constraints on the principal components according to the eigenvalues:

$$\overline{w}_h'(i) = \begin{cases} w_h(i), & |w_h(i)| \le a\sqrt{\lambda_i} \\ sign\,(w_h(i)) * a\sqrt{\lambda_i}, & |w_h(i)| > a\sqrt{\lambda_i} \end{cases}, \quad a > 0. \tag{17}$$

We use $a\sqrt{\lambda_i}$ to bound the principal components. Here, a is a positive scale parameter. The final hallucinated face image is reconstructed by

$$\overline{x}_h' = E_h\,\overline{w}_h' + \overline{m}_h\ . \tag{18}$$

The diagram of the hallucination algorithm based on eigentransformation is shown in Fig. 4. When a low-resolution image $\overline{x}_l$ is input, it is approximated by a linear combination of the low-resolution images using the PCA method, and we get a set of coefficients $[c_1, c_2, \ldots, c_M]^T$ on the training set. Keeping the coefficients and replacing the low-resolution training images with the corresponding high-resolution ones, a new high-resolution face image can be synthesized. The synthesized face image is projected onto the high-resolution

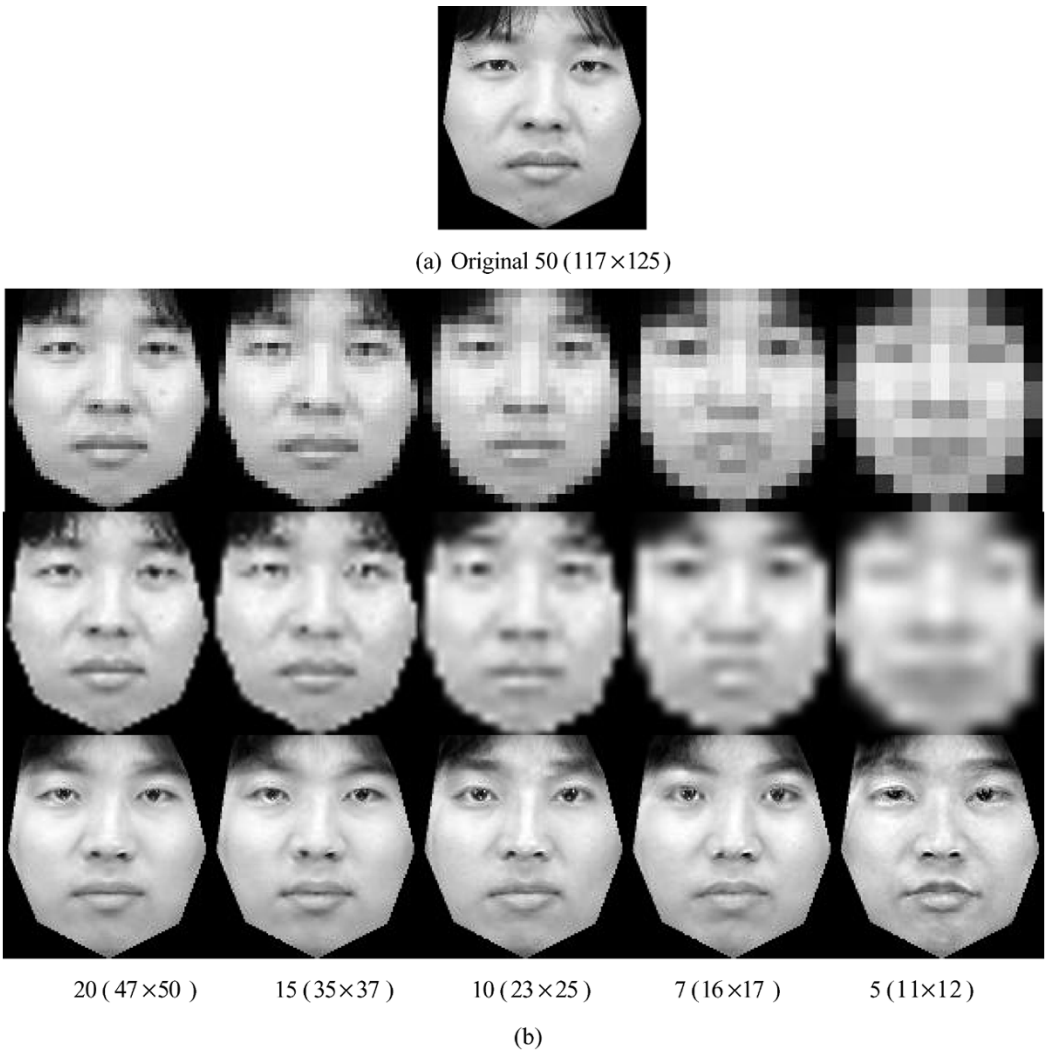

Fig. 7. Hallucinated face images using different resolutions as the input. (a) Original 50 (117 × 125). (b) First row is the input face images, whose eye center distances have been downsampled to 20, 15, 10, 7, and 5 pixels, respectively; the second row is the interpolation result using Cubic B-spline; the third row is the hallucinated face images.

eigenfaces and reconstructed with constraints on the principal components. This transformation procedure is called eigentransformation, since it uses the eigenfaces to transform the input image to the output result.

D. Discussion

In the eigentransformation algorithm, the hallucinated face image is synthesized by the linear combination of high-resolution training images and the combination coefficients come from the low-resolution face images using the PCA method. The algorithm improves the image resolution by inferring some high-frequency face details from the low-frequency facial information by taking advantage of the correlation between the two parts. Because of the structural similarity among face images, in multiresolution analysis, there exists strong correlation between the high-frequency band and low-frequency band. For high-resolution face images, PCA can compact this correlated information onto a small number of principal components. Then, in the eigentransformation process, these principal components can be inferred from the principal components of the low-resolution face by mapping between the high- and low-resolution training pairs. Therefore, some information in the high-frequency bands are partially recovered.

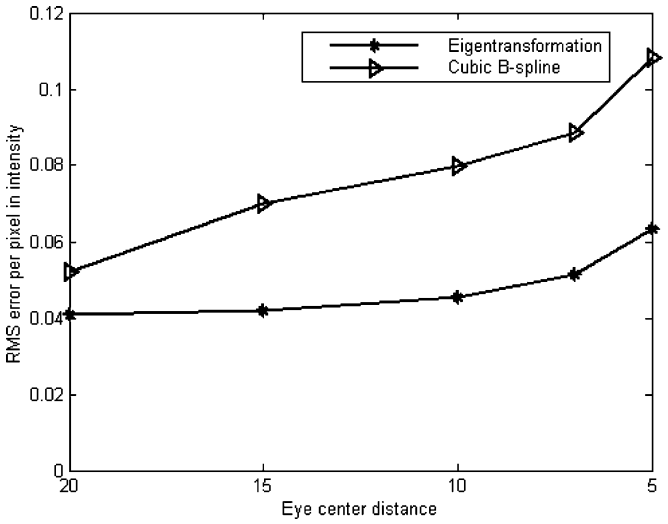

Fig. 8. RMS error per pixel in intensity using Cubic-spline interpolation and hallucination by eigentransformation. The intensity is between 0 and 1. The resolution is marked by eye center distance.

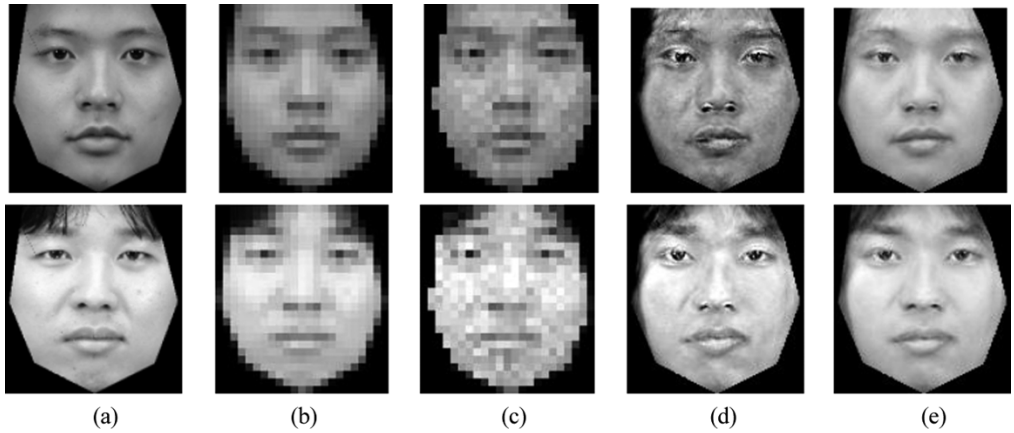

Fig. 9. Adding constraints to the principal components of hallucinated face images: (a) Original high-resolution face images; (b) low-resolution face images with $de = 10$, and size 23×25; (c) low-resolution face images with zero mean 0.03 standard deviation Gaussian noise; (d) hallucinated face images from (c) without constraints on the principal components; (e) hallucinated face images from (c) with constraints on principal components. (e) is more face-like and less noisy compared to (d), and retains most of the facial characteristics of (d).

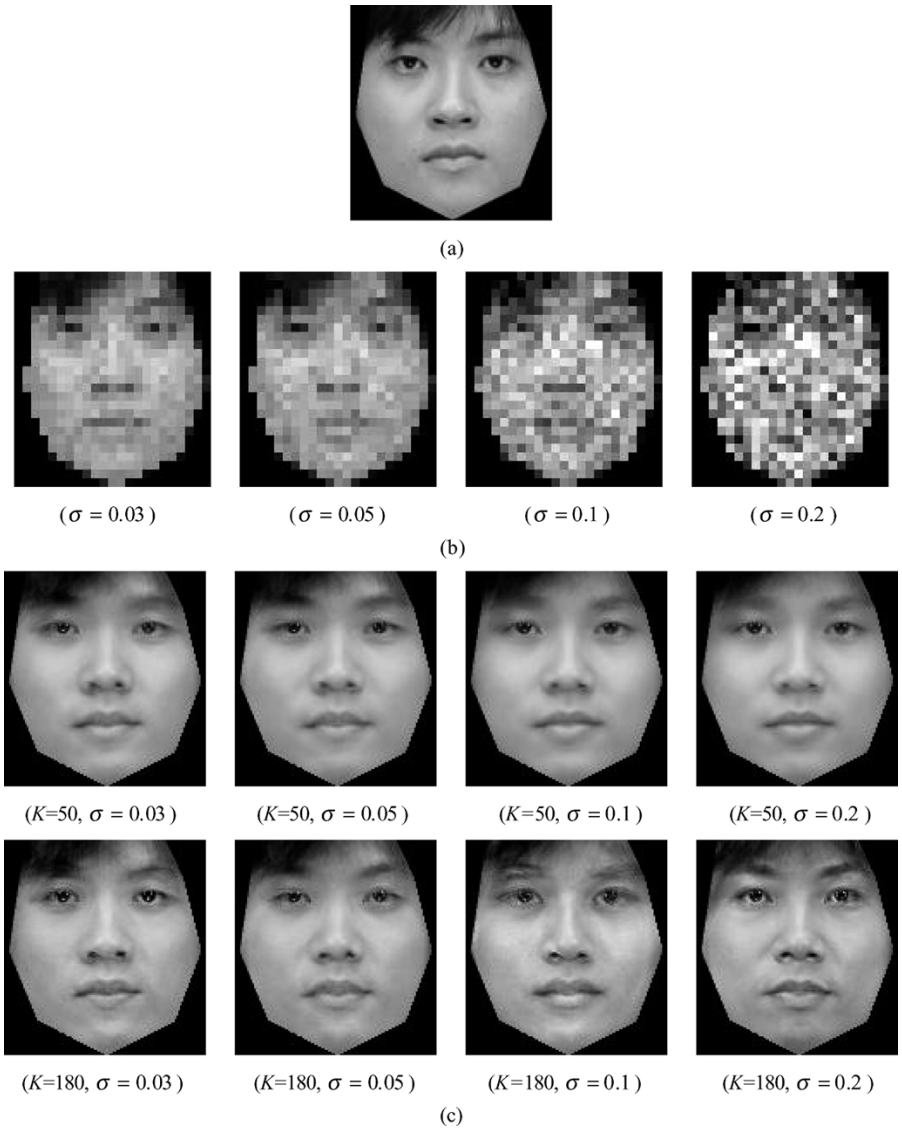

Fig. 10. Hallucinating face with additive zero mean Gaussian noise. The input face image is 23×25: (a) Original high-resolution face image; (b) low-resolution face images with noise; (c) hallucinated face images using different eigenface numbers. K is the eigenface number in eigentransformation, and σ is the standard deviation of Gaussian noise.

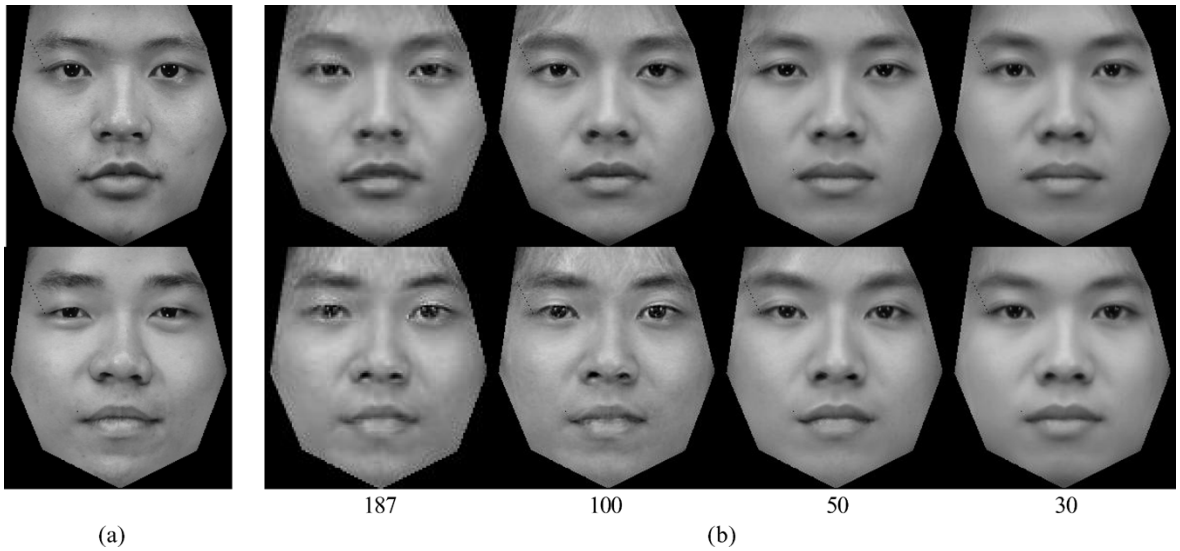

Fig. 11. Face hallucination using different training set sizes: (a) Original high-resolution face images; (b) hallucinated face images based on different training set sizes (187, 100, 50, and 30).

In practice, the low-resolution image is often disturbed by noise that has a flat distribution on all the eigenvectors. For low-resolution face images, the energy on small eigenvectors is small; thus sometimes it is overwhelmed by noise. If the face data distribution is modeled as a Gaussian distribution, the components on different eigenvectors are independent in the PCA representation. The information on these noisy components (eigenfaces after K as shown in Fig. 5) is lost, and cannot be recovered since the components on different eigenvectors are independent in the PCA representation. By selecting an optimal eigenface number K, we can extract the facial information and remove the noise. Since $\overline{r}_l$ is reconstructed from the K eigenfaces, given an optimal value of K, $\overline{r}_l$ encodes the maximum amount of facial information recoverable in the low-resolution face image.

By adjusting K, the eigentransformation method can control noise distortion. It makes full use of the facial information encoded in $\overline{r}_l$ to render the high-resolution face image. We have shown that the hallucinated face image is face-like and could produce $\overline{r}_l$ after resolution reduction. Although these conditions do not guarantee that the hallucinated face image is exactly the same as the original high-resolution face image, it does provide a face-like possible solution to $\overline{r}_l$. This solution helps to infer high-frequency components from the low-frequency facial features, thus significantly improving the appearance of the face image.

Our hallucination algorithm is based on training. So the composition of the training set plays an important role. A human face may undergo many kinds of variations caused by pose, lighting, glasses, etc. These factors may greatly change the face's appearance, and cause a complicated data distribution. It is a challenge for a robust face hallucination system to render face images under all these variations. One way to handle the complicated intrapersonal and extrapersonal variations is to enlarge the training set to contain as many kinds of variations as possible. However, this may not be the most effective solution. We find that a better approach is to construct a number of smaller training subsets. Each subset contains images of small variations. Better results can be obtained using the subset that is closest to the input image than using all the subsets combined.

We have noticed that some studies [21] use face images of small size for automatic face recognition, and have achieved satisfactory results. Through experiments, we would like to explore how the face resolution affects the recognition performance, and whether there is enough information for the low-resolution face images to distinguish different faces. Given the significant improvement of the face appearance by the hallucination process, it is interesting to investigate whether the hallucination process helps automatic recognition. Since more high-frequency details are recovered, we expect the hallucination process to help the recognition performance.

III. Experiment

A. Hallucination Experiment

The hallucination experiment is conducted on a data set containing 188 individuals with one face image for each individual. Using the "leave-one-out" methodology, at each time, one image is selected for testing and the remaining are used for training. In preprocessing, the face images are aligned by the two eyes. The distance between the eye centers is fixed at 50 pixels, and the image size is fixed at 117×125. Images are blurred by averaging neighbor pixels and downsampled to low-resolution images. Here, we use the eye center distance de to measure the face resolution.

Some hallucination results are shown in Fig. 6. The input face images are downsampled to 23×25, with de equal to 10. Compared with the input image and the Cubic B-Spline interpolation result, the hallucinated face images have much clearer detail features. They are good approximations to the original high-resolution images.

1) Hallucination for Different Resolutions: We study the hallucination performance using different resolutions as input. The eye center distance is downsampled to 20, 15, 10, 7, and 5. An example is shown in Fig. 7, where (a) is the original high-resolution face image. In Fig. 7(b), the first row is the input face images with different resolutions; the second row is the result of Cubic B-Spline interpolation; and the third row is the hallucination result. Fig. 8 reports the average root mean square (rms) error per pixel in intensity for the 188 face images under different resolutions. For a very low resolution, the low-resolution and direct-interpolated face images are practically indiscernible, and the rms error of Cubic B-spline interpolation increases quickly. The performance of hallucination by eigentransformation is much better. When de is downsampled to 10, the result of eigentransformation is still satisfactory. For further lower resolutions, there are some distortions on the eyes and mouth, but the hallucinated images are still clear and face-like.

2) Robustness to Noise: In Fig. 9, we add zero mean Gaussian noise to the low-resolution face images. If no constraints are added to the principal components, the hallucinated face images in Fig. 9(d) are with noise distortion and somewhat not face-like. Adding constraints on the

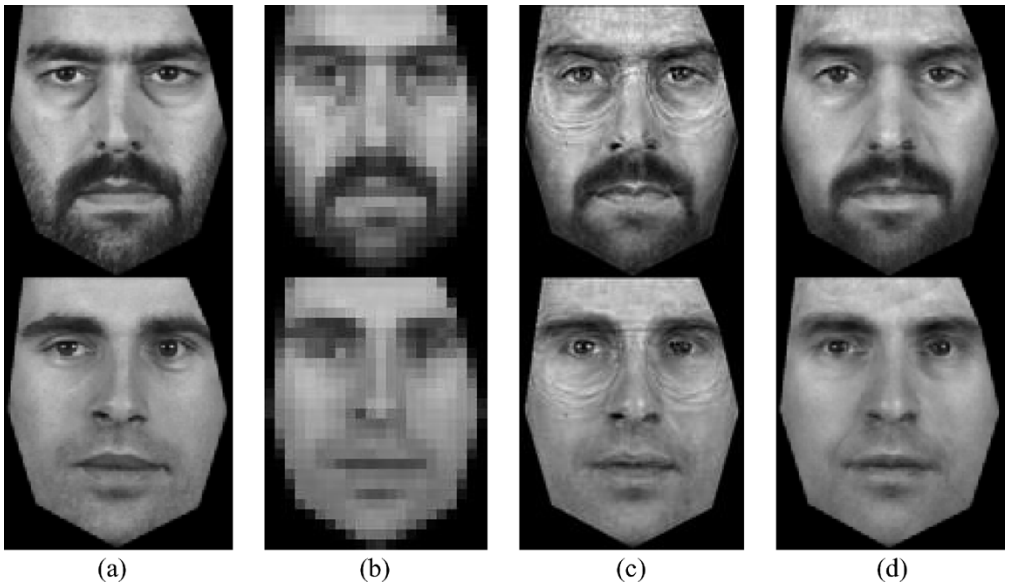

Fig. 12. Face hallucination without glasses: (a) Original high-resolution face images; (b) input low-resolution face images; (c) hallucinated face images based on the training set containing samples both with and without glasses; (d) hallucinated face images based on the training set without glasses.

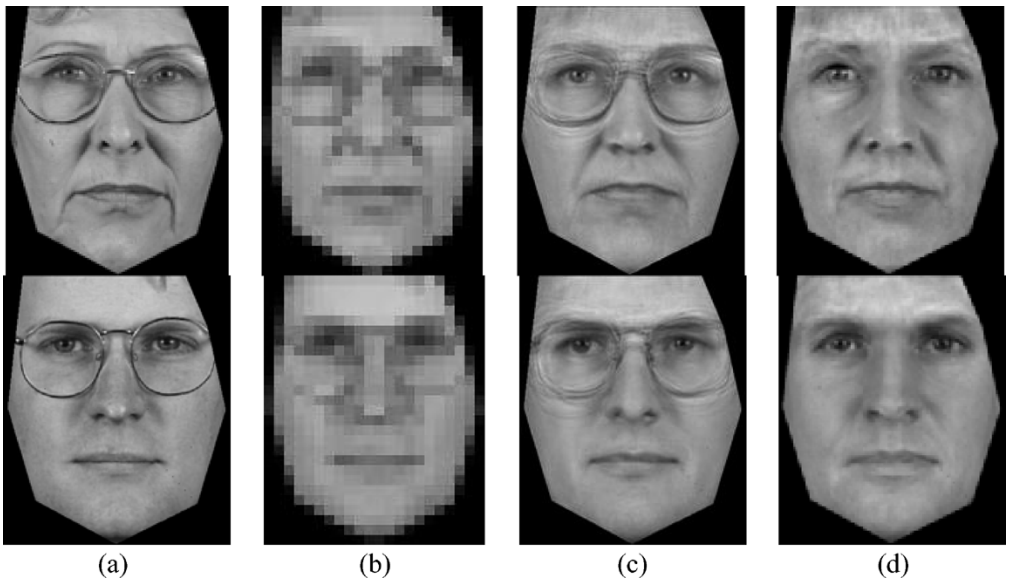

Fig. 13. Face hallucination with glasses: (a) Original high-resolution face images; (b) input low-resolution face images; (c) hallucinated face images based on the training set with glasses; (d) hallucinated face images based on the training set without glasses.

principal components using (17), the reconstructed face images remove most of the noise distortion and retain most of the facial characteristics as shown in Fig. 9(e). Here, we set the parameter a as 2.

As discussed in Section II, some high-frequency detail is lost in the process of blurring and downsampling, or overwhelmed by noise. Selecting the eigenface number in eigentransformation, we could control the detail level by keeping maximum facial information while removing most of the noise disturbance. In Fig. 10, we add zero mean Gaussian noises with four different standard deviations (σ) to the low-resolution face image with de equal to 10 (size of 23×25). The image intensity is between 0 and 1. Two different eigenface number K, 50 and 180, are used in the eigentransformation. When only 50 eigenfaces are used in the eigentransformation, the hallucinated face images lose some individual characteristics. Although the edges and contours are clear, the hallucinated faces are more like a mean face. When the eigenface number is equal to 180, more individual characteristics are added to the hallucinated face images. For relatively small noise ($\sigma = 0.03, 0.05$), these characteristics are similar to the original high-resolution face image. But for large noise ($\sigma = 0.1, 0.2$), even though the hallucinated faces are still face-like, the added characteristics start to deviate from those of the original face. So when the noise is small, a larger eigenface number is more suitable, since it can characterize the face better with more individually detailed characteristics. When noise is large, a small eigenvector number is better. Although the hallucinated faces contain less individual facial characteristics, they are more face-like. In practice, we could estimate the noise effect and choose the proper detail level for hallucination.

B. Impact of Training Set

1) Training Set Size: Fig. 11 shows two examples of hallucinated face images based on a different number of training samples. There is not much difference between the results using 187 and 100 training

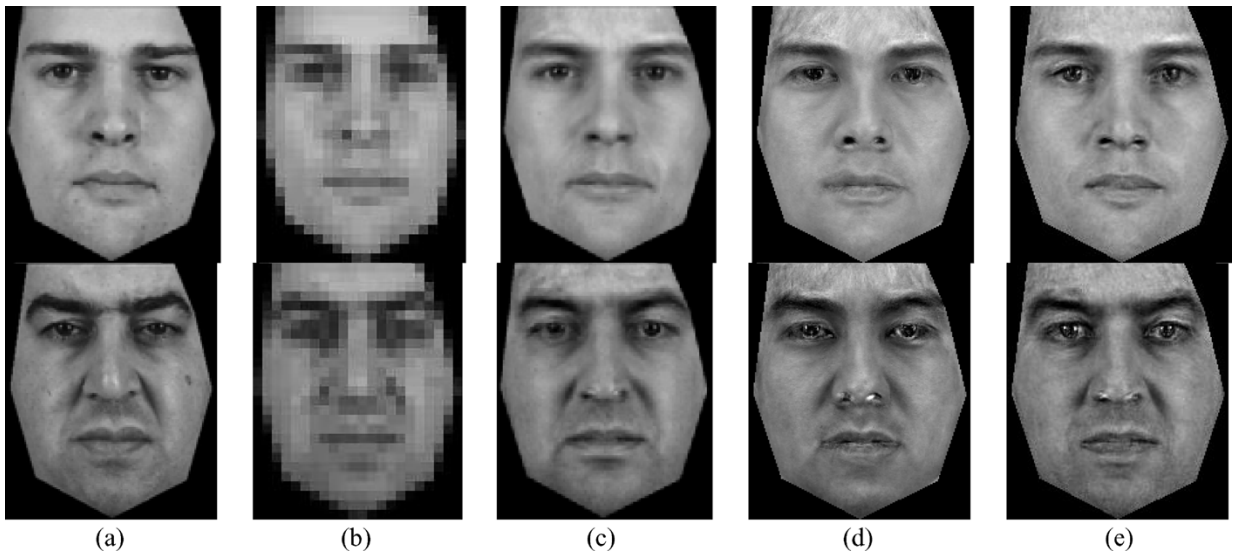

Fig. 14. Face hallucination with race variation: (a) Original high resolution occidental face images; (b) input low-resolution face images; (c) hallucinated face images based on occidental training set; (d) hallucinated face images based on oriental training set; (e) hallucinated face images based on enlarged training set containing both occidental and oriental people.

samples. This shows that our hallucination algorithm can achieve satisfactory results even based on a relatively small training set. However, when the training set is too small, many individual characteristics cannot be rendered. As shown in Fig. 11, the hallucinated face images of the two persons become very similar to each other when less than 50 images are used for training.

2) Training Set Composition: In this section, we design a set of experiments to study the impact of different training set compositions on face hallucination. We focus on two special cases: race variation and wearing glasses. The conclusion from the experiments can be easily extended to other kinds of variations such as pose and lighting. Some of the face images for experiments are selected from the XM2VTS database [27]. In Fig. 12, the face images to be rendered are not wearing glasses. When they are hallucinated from an enlarged training set containing samples both with and without glasses, the results have some noise around the eyes. If we adopt a smaller training set containing only samples without glasses, the hallucinated face images have less noise. This shows that if we can determine what type of variation the face image is undergoing, it is better to design a smaller training set with the same kind of variation. However, sometimes when the resolution of the input image is too low, we cannot discriminate whether the face is wearing glasses or not. In this case, we can hallucinate the face images using several different training subsets and choose the best result. Of course, we can also use all the results from different training sets together to identify the person. In Fig. 13, the input face images wearing glasses are properly rendered from the training set with glasses. The glasses can also be "taken off" by using the training set without glasses. Both of the two kinds hallucinated results are helpful to identify the person.

In Fig. 14, the input are occidental faces. We generate the hallucinated face images based on three training sets: occidental training set, oriental training set, and enlarged training set containing both occidental and oriental people. It is clear that the occidental training set gives the best result. This further confirms that it is important to design a training set that shares similar face characteristics with the face to be hallucinated. The face images hallucinated from the oriental training set are somewhat blurred and noisy. This may also provide some hint on judging the race of the faces. The face has a greater possibility to be the same race with the training set which can better hallucinate it.

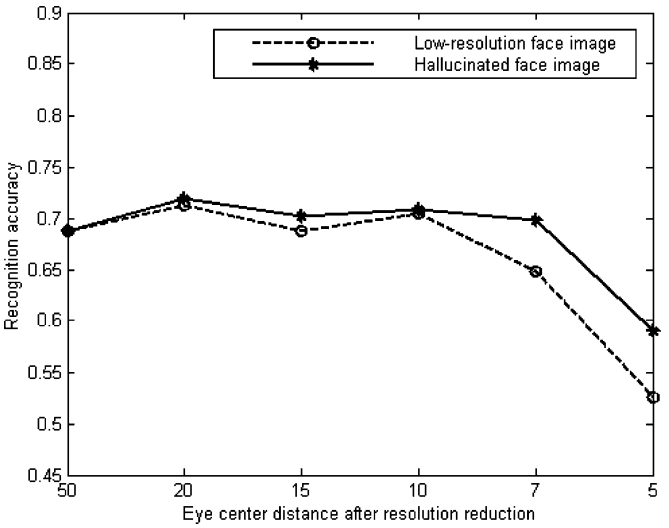

Fig. 15. Recognition accuracy using low-resolution face images and hallucinated face images based on XM2VTS database. The resolution is marked by eye center distance with 50, 20, 15, 10, 7, and 5.

C. Recognition Experiment

We study the recognition performance using low-resolution face images and hallucinated face images. Two hundred and ninety five individuals from the XM2VTS face database [27] are selected, with two face images in different sessions for each individual. One image is used as a reference, and the other is used for testing. We use direct correlation for recognition, which is perhaps the simplest face recognition algorithm. The reason for using this simple classification algorithm is that our focus is on the comparison of recognition ability of the low-resolution and hallucinated face images rather than a sophisticated classification algorithm. The recognition accuracies over different resolutions are plotted in Fig. 15. When de is reduced from 50 to 10, there is only slight fluctuation on recognition accuracy using low-resolution face images. When de is further reduced to 7 and 5, the recognition accuracy

for low-resolution face images drops greatly. Resolution with de equal to 10 is perhaps a low-bound for recognition. Below this level there may not be enough information for recognition. This is also consistent with the hallucination experiment in Fig. 7. Satisfactory hallucination results can be obtained when de is equal to or larger than 10.

We also try to explore whether hallucination can contribute to automatic face recognition. We expect hallucination to make the recognition procedure easier, since it emphasizes the face difference by adding some high-frequency details. In this experiment, the low-resolution testing image is hallucinated by reference face images, but the face image of the testing individual is excluded from the training set. As shown in Fig. 15, the hallucination improves the recognition accuracy when the input face images have very low resolutions. The improvement seems not as significant as the improvement in the face appearance. Further investigation in psychology may be needed to address this phenomenon. It seems that the human visual system can better interpret the added high-frequency details.

IV. Conclusion

Because of the structural similarity, face images can be synthesized from the linear combination of other samples. Based on this property of face images, hallucination can be implemented by eigentransformation. By selecting the frequency level in the PCA representation, our method extracts maximum facial information from the low-resolution face images and is robust to noise. The resolution and quality of face images are greatly improved over the low-resolution images. The hallucination process not only helps a human being to identify faces but also makes the automatic face recognition procedure easier. It will be interesting to study why the hallucinated image is significantly better perceived by a human being than by the automatic recognition system.

References

[1] S. Baker and T. Kanade, "Hallucinating faces," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recog.*, Mar. 2000, pp. 83–88.

[2] A. Patti, M. Sezan, and A. Tekalp, "Super-resolution video reconstruction with arbitrary sampling latices and nonzero aperture time," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1064–1076, Aug. 1997.

[3] B. Bascle, A. Blake, and A. Zisserman, "Motion deblurring and super-resolution from an image sequence," in *Proc. ECCV*, May 1996, pp. 573–581.

[4] C. Liu, H. Shum, and C. Zhang, "A two-step approach to hallucinating faces: Global parametric model and local nonparametric model," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Dec. 2001, pp. 192–198.

[5] J. D. Bonet, "Multiresolution sampling procedure for analysis and synthesis of texture images," in *Proc. SIGGRAPH 97*, Aug. 1997, pp. 361–368.

[6] J. De Bonet and P. Viola, "A nonparametric multi-scale statistical model for neutral images," in *Proc. Advances Neutral Inf. Process.*, 1997.

[7] M. Chiang and T. E. Boult, "Local blur estimation and super-resolution," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Jun. 1997, pp. 821–826.

[8] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred noisy and undersampled measured images," *IEEE Trans. Image Proc.*, vol. 6, no. 12, pp. 1646–1658, Dec. 1997.

[9] R. Hardie, K. Barnard, and E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Process.*, vol. 6, pp. 1621–1633, Jun. 1997.

[10] R. Schultz and R. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 996–1011, Jun. 1996.

[11] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

[12] S. Kim and W. Su, "Recursive reconstruction of high-resolution reconstruction of blurred multiframes images," *IEEE Trans. Image Processing*, vol. 2, no. 6, pp. 534–539, Jun. 1993.

[13] W. T. Freeman and E. C. Pasztor, "Learning low-level vision," in *Proc. IEEE Int. Conf. Computer Vision*, Sep. 1999.

[14] P. S. Penev and L. Sirovich, "The global dimensionality of face space," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, Mar. 2000, pp. 264–270.

[15] M. Elad and A. Feuer, "Super-resolution of an image sequence-adaptive filtering approach," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 387–395, Mar. 1999.

[16] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Jun. 2000.

[17] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, pp. 705–741, May 1995.

[18] W. Zhao, R. Chellappa, A. P. J. Phillips, and A. P. J. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, Dec. 2003.

[19] M. Elad and A. Feuer, "Super-resolution reconstruction of image sequences," *IEEE Trans. Pattern Anal. Mech. Anal.*, vol. 21, no. 9, pp. 817–834, Sep. 1999.

[20] W. Zhao, R. Chellapa, and P. Philips, "Subspace linear discriminant analysis for face recognition,", Tech. Rep. CAR-TR-914, 1996.

[21] B. Moghaddam, "Principal manifolds and probabilistic subspace for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 780–788, Jun. 2002.

[22] X. Tang and X. Wang, "Face photo recognition using sketch," in *Proc. of ICIP*, Sep. 2002, pp. 257–260.

[23] ——, "Face sketch recognition," *IEEE Trans. Circuits Sys. Video Technol., Special Issue on Image- and Video-Based Biometrics*, vol. 14, no. 1, pp. 50–57, Jan. 2004.

[24] X. Wang and X. Tang, "Face hallucination and recognition," in *Proc. 4th Int. Conf. Audio- and Video-Based Personal Authentication, IAPR,* University of Surrey, Guildford, U.K., Jun. 2003, pp. 486–494.

[25] M. Turk and A. Pentland, "Eigenface for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991.

[26] K. Fukunnaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York, NY: Academic, 1991.

[27] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Proc. 2nd Int. Conf. Audio- and Video-Based Personal Authentication, IAPR*, Washington D.C., Mar. 1999, pp. 72–77.