Face Super-Resolution based on Singular Value Decomposition

Muwei Jian and Kin-Man Lam
Centre for Signal Processing, Department of Electronic and Information Engineering,
The Hong Kong Polytechnic University, Kowloon, Hong Kong.
E-mail: {10902666r, enkmlam}@polyu.edu.hk

Abstract— **In this paper, a novel face image super-resolution approach based on singular value decomposition (SVD) is proposed. We prove that the singular values of an image at one resolution have approximately linear relationships with their counterparts at other resolutions. This makes the estimation of the singular values of the corresponding HR face images more reliable. From the signal-processing point of view, this can effectively preserve and reconstruct the dominant information in the HR face image. Interpolating the two other matrices obtained from the SVD of a LR face image does not change either the primary facial structure or the pattern of the face image. Furthermore, the mapping scheme for interpolating the matrices can be viewed as a "coarse-to-fine" estimation of HR face images, which uses the mapping matrices learned from the corresponding reference image pairs. Experimental results show that the proposed super-resolution scheme is effective and efficient.**

I. INTRODUCTION

Face super-resolution (SR) is useful for face recognition, in particular when the face images concerned are of low resolution. This situation always happens when an image or video is captured under outdoor or uncontrolled conditions.

Constructing a high-resolution (HR) image from its low-resolution (LR) inputs is also called face hallucination [1], which has become one of the most important fields for LR face recognition. Face hallucination was firstly proposed by Baker et al. [1], and has drawn many researchers' attention since then. A pixel-wise super-resolution (SR) method was proposed, which uses the Laplacian pyramid and the Gaussian pyramid to decompose an image into a pyramid of features in order to generate a HR face image. Later, a new edge-directed interpolation method was proposed in [2]. Freeman et al. [3] proposed a nonparametric patch-based prior along with the Markov random field model to produce the desired HR images. In [4], temporal correspondence and a prior model are combined to hallucinate faces. Many researchers [5, 6] have further developed patch-based SR frameworks. A sparse-coding method [7] was proposed to represent a LR input patch as a combination of its raw neighboring image patches, and the target HR patch is generated directly by using the same combination coefficients to the corresponding neighboring HR patches. The algorithms proposed in [5, 6, 7] also used the same approach, in which a number of similar neighbors to the LR input patches are searched from a training dataset, and then a specific method is adopted to reconstruct the corresponding HR images. In [8], Wang et al. proposed a holistic face-hallucination method, which employs Principal Component Analysis (PCA) to represent a LR input image as a linear combination of LR training samples. The HR image to be reconstructed is then estimated using the same linear combination to the corresponding HR training samples. Park et al. [9] utilized the PCA-based SR framework [8] to develop an example-based face-hallucination method. As the PCA method considers global-structure information about facial images, it is less suitable for use in patch-based approaches [12]. In [10], a hybrid method was proposed based on global and local constraints to apply face hallucination to unregistered images. In [11], a novel example-based image SR method was proposed, in which a class-specific predictor is designed for each class of patches so as to improve the accuracy of estimating the high-frequency content. Recently, a new face-hallucination framework – namely, from local-pixel structure to global image SR (LPS-GIS) – was developed in [12] and is based on the assumption that two similar face images should have similar local-pixel structures. This new framework uses an input LR face image to search a face database for similar example HR faces in order to learn the local-pixel structures for the target HR face.

In contrast to these previous works, we propose a novel and efficient face-hallucination scheme based on a mapping model. Since we can observe and prove that the singular values of an image at one resolution have approximately linear relationships with their counterparts at other resolutions, the estimation of the singular values of the corresponding HR face images becomes more accurate. These relationships of the singular values for images at different resolutions can effectively preserve and reconstruct the dominant information in the HR face image. We also propose a mapping scheme to interpolate the two other matrices obtained in the singular value decomposition (SVD) of the LR input image. The respective mapping matrices are learned from reference image pairs, which can generate facial details more accurately. Experimental results show that our algorithm is effective.

The rest of the paper is organized as follows. In Section II, we will present our proposed face-hallucination scheme. Experimental results are presented in Section III. The paper closes with a conclusion and discussion in Section IV.

II. FACE-HALLUCINATION SCHEME BASED ON SVD

A. Singular Value Decomposition of Face Images

An image I with m rows and n columns, assuming that $m \geq n$, can be viewed as a real matrix. By using SVD, I can be written as the product of a left matrix U, a $n \times n$ diagonal matrix W with positive or zero elements, and the transpose of a right matrix V, i.e.

$$I = UWV^T, \tag{1}$$

where $U^T U = V^T V = E$, and E is the unit matrix. The matrix U is a $m \times n$ column-orthogonal matrix, while V is a $n \times n$ orthogonal matrix. W is a diagonal matrix whose elements w_i on the diagonal are called singular values (the square root of the eigenvalues), i.e.

$$W = diag(w_1, w_2, ..., w_i ...., w_n). \tag{2}$$

The singular value vector s of I is defined as follows:

$$s = [w_1, w_2, ..., w_i ...., w_n]^T, \tag{3}$$

where $1 \leq i \leq n$, w_i is the i^{th} singular value of I in the singular value vector s such that $w_i \geq w_{i+1}$, and where the singular values usually decrease dramatically.

We found that the first several eigenvectors are sufficient to account for almost all of the information contained in a face image. This observation is also true for texture images [13, 14]. Since the singular values decrease rapidly and the first few eigenvectors can account for most of the information, the original diagonal matrix W can be approximated by retaining the first k largest-magnitude singular values, as follows:

$$\hat{W}_I = diag(w_1, w_2, ..., w_k, 0, ..., 0). \tag{4}$$

We have:

$$\sum_{i=1}^{n} w_i^2 \approxeq \sum_{i=1}^{k} w_i^2, \tag{5}$$

if the singular values discarded have a small magnitude. With $\hat{W}_I$, an approximated image $\hat{I}$ can be generated, which contains almost the same information as the original image I. The image $\hat{I}$, which can be viewed as a matrix, can be expressed as follows:

$$\hat{I} = U\hat{W}_I V^T. \tag{6}$$

B. The Diagonal Matrix W at Different Resolutions

It can be observed that when a LR image I_2 is interpolated or super-resolved to produce a new HR image I_1 with a magnification factor of α, the first k main singular values in the singular-value vector $s_1 = [w_{1,1}, w_{1,2}, ..., w_{1,k}]^T$ of the new HR image I_1 can be approximated as α times the corresponding first k main singular values in the singular-value vector $s_2 = [w_{2,1}, w_{2,2}, ..., w_{2,k}]^T$ of the original image I_2. Hence, we have

$$s_1 \approxeq \alpha\, s_2. \tag{7}$$

According to a linear algebra theory [17], if a matrix A has singular values $w_1, w_2, ..., w_i ...., w_n$, then

$$\| A \|_F^2 = \sum_{i=1}^{n} w_i^2, \tag{8}$$

where $\| A \|_F$ is the Frobenius norm of matrix A, which is defined as the square root of the sum of the squares of all its entries. The following is a brief proof of (7):

Proof: Suppose that I_2 is interpolated to produce a new HR image I_1 with a magnification factor of α, we have:

$$\begin{aligned} \| I_1 \|_F^2 &\overset{[1]}{\approxeq} \| \alpha I_2 \|_F^2 = \alpha^2 \| I_2 \|_F^2 \\ &\overset{[2]}{=} \alpha^2 \sum_{i=1}^{n} w_{2,i}^2 \\ &\overset{[3]}{\approxeq} \alpha^2 \sum_{i=1}^{k} w_{2,i}^2 = \sum_{i=1}^{k} (\alpha w_{2,i})^2 = \sum_{i=1}^{k} (w_{1,i})^2 \end{aligned}$$

$$\therefore\ s_1 \approxeq \alpha\, s_2.$$

[1] For every pixel in I_2, an interpolation method, with a magnification factor of α, produces $\alpha \times \alpha$ neighboring pixels in the HR image I_1, with similar pixel values to that in the original LR image I_2. For instance, the nearest-neighbor interpolation generates $\alpha \times \alpha$ neighbors of equal values, and the bilinear and bicubic interpolation methods produce $\alpha \times \alpha$ similar pixels.

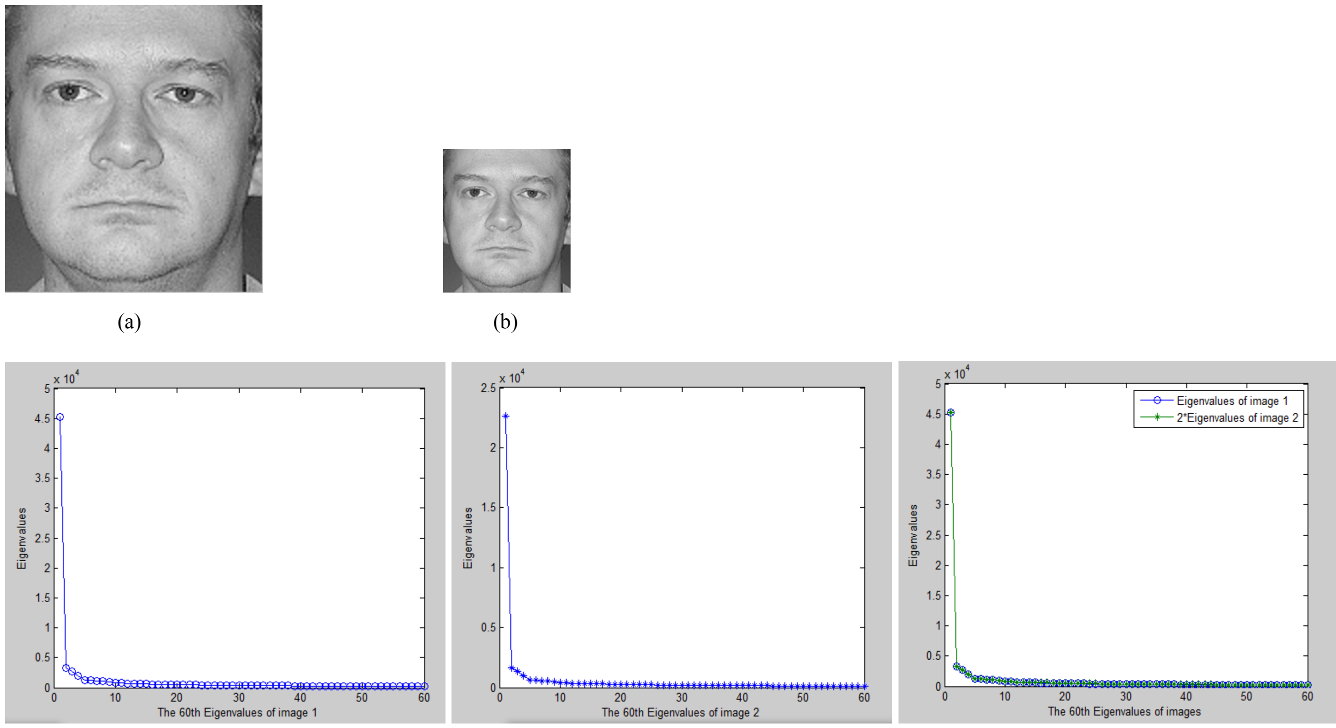
[2] According to (8), $\| A \|_F^2 = \sum_{i=1}^{n} w_i^2$.

[3] According to (5), $\sum_{i=1}^{n} w_i^2 \approxeq \sum_{i=1}^{k} w_i^2$.

For the LR image I_2, there are, at most, n singular values in the matrix W_2. The matrix W_1 of the HR image I_1 should be of size $\alpha n \times \alpha n$, and can be estimated based on W_2. A good approximation of W_1, denoted as $\hat{W}_1$, has, at most, n non-zero singular values and is given as follows:

$$\hat{W}_1 = diag(\alpha w_1, \alpha w_2, \ldots, \alpha w_n, 0, \ldots, 0). \tag{9}$$

Figs. 1(c), 1(d) and 1(e) show the linear relationship of the first k=60 main singular values for two images of the same



(a) (b)

(c) (d) (e)

Fig. 1 The eigenvectors of the two images of the same face with different resolutions: (a) An original image of size 262×300, (b) a down-sampled image of size 131×150 obtained with a factor of 2 in both the horizontal and vertical directions, (c) the first 60 eigenvalues of the image in (a), (d) the first 60 eigenvalues of the image in (b), (e) the first 60 eigenvalues of the image in (a) and twice the values of the first 60 eigenvalues of the image in (b).

person at different resolutions. The image in Fig. 1(a) has double the resolution of the image in Fig. 1(b), both horizontally and vertically. Figs. 1(c) and 1(d) show the first 60 eigenvectors of the images in Figs. 1(a) and 1(b), respectively. Fig. 1(e) illustrates the first 60 eigenvalues of the image in Fig. 1(a), as well as the first 60 eigenvalues of the image in Fig. 1(b), multiplied by two. We can see that $s_1 \cong \alpha s_2$, with the magnification factor α=2.

C. *Our Face-hallucination Scheme*

After estimating the diagonal matrix $\hat{W}_1$ using the relationship $s_1 \cong \alpha s_2$, we can use mapping matrices P_u and P_v learned from a reference image pair for face hallucination. Suppose that a reference image pair containing a HR image and a corresponding sub-sampled LR image is searched from a face dataset, which is similar to the input LR image I_l^i. The reference HR image is denoted as I_h and the corresponding reference LR image is denoted as I_l. These two images, I_h and I_l, can be expressed using SVD as follows:

$$I_h = U_1 W_1 V_1^T \text{ and } I_l = U_2 W_2 V_2^T . \qquad (10)$$

The two matrices U_2 and V_2 for I_l can be interpolated to form two new matrices U_2' and V_2', which have the same size as U_1 and V_1, respectively. Define the two mapping matrices P_u and P_v, as follows:

$$U_1 = U_2' P_u \text{ and } V_1 = V_2' P_v . \qquad (11)$$

These two matrices P_u and P_v can be calculated using pseudo-inverse as follows:

$$\tilde{P}_u = (U_2'^T U_2')^{-1} U_2'^T U_1 \text{ and } \tilde{P}_v = (V_2'^T V_2')^{-1} V_2'^T V_1 . \qquad (12)$$

As $U_2'^T U_2'$ and $V_2'^T V_2'$ are always singular matrices, P_u and P_v can be computed approximately as follows:

$$\tilde{P}_u = (U_2'^T U_2' + \lambda E)^{-1} U_2'^T U_1 \text{ and } \tilde{P}_v = (V_2'^T V_2' + \lambda E)^{-1} V_2'^T V_1 , \qquad (13)$$

where λ is a small positive integer.

To super-resolve the input LR image I_l^i to form the HR output image I_h^o, I_l^i is also decomposed by SVD using (1), and the corresponding matrices U_l^i and V_l^i can be interpolated into two new HR matrices U_h^i and V_h^i, respectively, which have the same size as the corresponding

(a) (b) (c) (d) (e)

Fig. 2 Face-hallucination results reconstructed using different methods with a magnification factor of 4 (α=4): (a) original images, (b) bicubic interpolation, (c) NEDI, (d) the eigentransformation method, and (e) our proposed method.

left matrix U_h^o and right matrix V_h^o of the HR image I_h^o, respectively. Then, the matrices U_h^o and V_h^o can be estimated as follows:

$$\hat{U}_h^o = U_h^o \widetilde{P}_u \text{, and } \hat{V}_h^o = V_h^o \widetilde{P}_v . \qquad (14)$$

Then, the HR image can be reconstructed from a reference image pair using the two mapping matrices $\tilde{P}_u$ and $\tilde{P}_v$, and the diagonal matrix $\hat{W}_1$ which is computed by using the scheme described in Section II.B. Thus, the required HR face I_h^o is obtained as follows:

$$I_h^o = \hat{U}_h^o \hat{W}_1 \hat{V}_h^{oT} . \qquad (15)$$

III. Experimental Results

In order to verify the effectiveness of the proposed face-hallucination algorithm, the dataset in [12] is used to evaluate the performance of our algorithm and to compare it to different state-of-the-art algorithms. There are 230 face images in the dataset, all of which belong to different subjects. All the facial images are aligned based on the position of the two eyes. The original HR facial images are cropped to a size of 124×108. The parameter λ in (13) is empirically set at 0.001. In the experiments, we evaluate all the methods by reconstructing the HR facial images with a magnification factor of 4. All the testing images are evaluated using the "leave-one-out" approach. Two objective quality measures, PSNR (peak signal-to-noise ratio) and SSIM (structural similarity) [15], are used to evaluate the performances of the different methods. The classical bicubic-interpolation algorithm [16], the edge-directed interpolation (NEDI) method [2], and the eigentransformation method [8] are performed for comparison.

Fig. 2 shows some samples of the reconstruction results. It can be seen from Fig. 2(b) that the bicubic-interpolation algorithm produces the blurriest results. The results in Fig. 2(c) are generated using the NEDI method. However, if a face image has very low resolution, the NEDI method struggles to distinguish edges, and hence also produces blurry results. Fig. 2(d) shows the results generated using the eigentransformation method, where plausible face structures can be inferred in the HR images. Fig. 2(e) shows our method, whereby plausible HR images with a holistic structure and more details with a better visual quality can be obtained.

Table 1 tabulates the average PSNR and SSIM of the different methods. The results show that our method is superior to the others in terms of these two measurements.

IV. Conclusion and Discussion

In this paper, we have proposed a novel face-hallucination scheme based on SVD and using a simple mapping model. We have also proved that the major singular values of the same image at different resolutions are proportional to the magnification factor. This property is then used to estimate the singular-value matrix of the HR face image to be reconstructed. The left and right matrices of the HR face images are generated using a simple mapping scheme. The two corresponding mapping functions are learned from a pair of LR and HR face images searched from a dataset. These learned mapping functions can be seen as holistic constraints in the reconstruction of the HR images. Compared to typical state-of-the-art algorithms, experiments have shown that our proposed method is practicable and can produce plausible HR images with both a holistic structure and high-frequency details.

Table I

The average PSNR and SSIM of the different face-hallucination algorithms.

	PSNR(dB)	SSIM
Bicubic interpolation [16]	22.1593	0.7022
NEDI [2]	21.6398	0.7034
Eigentransformation method [8]	25.3950	0.7167
Sparse representation [7]	23.0124	0.7025
Our proposed method	25.6111	0.7402

Acknowledgment

We would like to thank Prof. Guoping Qiu in the School of Computer Science, University of Nottingham and Prof. Junyu Dong at the Department of Computer Science, Ocean University of China, Qingdao, China, Mr. Zhuo Hui and Mr. Chensheng Sun in the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University for their helpful and constructive discussions and suggestions, which definitively enhance the quality of our paper.

References

[1] S. Baker and T. Kanade, "Hallucinating faces," In *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 83-88, 2000.

[2] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, pp. 1521–1527, Oct. 2001.

[3] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based superresolution," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 56–65, Mar./Apr. 2002.

[4] G. Dedeoglu and T. Kanade. "High-zoom video hallucination by exploiting spatio-temporal regularities," In Proc. *IEEE CVPR*, vol. 2, pp. 151–158. 2004.

[5] W. Fan and D.-Y. Yeung, "Image hallucination using neighbor embedding over visual primitive manifolds," in Proc. *IEEE CVPR*, pp. 1–7, 2007.

[6] K. I. Kim and Y. Kwon, "Example-based learning for single-image super-resolution," in Proc. *30th DAGM Symp. Pattern Recognition, Munich*, Germany, pp. 456–465, 2008.

[7] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Processing*, vol. 19, no. 11, pp. 2861-2873, 2010.

[8] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev*, vol. 35, no. 3, pp. 425–434, Aug. 2005.

[9] J. S. Park and S. W. Lee, "An example-based face hallucination method for single-frame, low-resolution facial images," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1806–1816, Oct. 2008.

[10] C. Liu, H. Y. Shum, and C. S. Zhang, "Face Hallucination: Theory and Practice" *Int. J. on Conputer Vision*, 75(1), pp. 115–134, 2007.

[11] X. Li, K. M. Lam, Guoping Qiu, Lansun Shen, and Suyu Wang, "Example-based image super-resolution with class-specific predictors," *J. Vis. Commun. Image Represent.*, pp. 312–322, 2009.

[12] Y. Hu, K. M. Lam, G. Qiu, and T. Shen, "From local pixel structure to global image super-resolution: A new face hallucination framework," *IEEE Trans. Image Process.*, pp. 433-445, 2011.

[13] J. Dong and M. Chantler, "Capture and synthesis of 3D surface texture," *Int. J. on Computer Vision*, 62(1–2), pp.177-194, 2005.

[14] M. Jian and J. Dong, "Capture and fusion of 3d surface texture," *Multimedia Tools and Applications*, vol. 53, no. 1, pp. 237-251, 2011.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[16] H. S. Hou and H. C. Andrews, "Cubic spline for image interpolation and digital filtering," *IEEE Transactions on Signal Processing*, vol. 26, pp. 508–517, 1978.

[17] Lay, David C., *Linear Algebra and Its Applications*, 3rd ed., Addison Wesley, August 22, 2005, ISBN978-0-321-28713-7.