

©2010 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

# Face Hallucination under an Image Decomposition Perspective

Yan Liang<sup>1,3,5</sup>, Jian-Huang Lai<sup>2,3,6,\*</sup>, Xiaohua Xie<sup>1,3,7</sup>, Wanquan Liu<sup>4,8</sup>

<sup>1</sup>School of Mathematics and Computational Science, Sun Yat-sen University, China

<sup>2</sup>School of Information Science and Technology, Sun Yat-sen University, China

<sup>3</sup>Guangdong Province Key Laboratory of Information Security, China

<sup>4</sup>Department of Computing, Curtin University of Technology, Australia

<sup>5</sup>sysuly@gmail.com, <sup>6</sup>stsljh@mail.sysu.edu.cn, <sup>7</sup>sysuxiehx@gmail.com, <sup>8</sup>w.liu@curtin.edu.au

**Abstract**—In this paper we propose to convert the task of face hallucination into an image decomposition problem, and then use the morphological component analysis (MCA) for hallucinating a single face image, based on a novel three-step framework. Firstly, a low-resolution input image is up-sampled by interpolation. Then, the MCA is employed to decompose the interpolated image into a high-resolution image and an unsharp masking, as MCA can properly decompose a signal into special parts according to typical dictionaries. Finally, a residue compensation, which is based on the neighbor reconstruction of patches, is performed to enhance the facial details. The proposed method can effectively exploit the facial properties for face hallucination under the image decomposition perspective. Experimental results demonstrate the effectiveness of our method, in terms of the visual quality of the hallucinated face images.

**Keywords**- face hallucination; image decomposition; morphological component analysis; principal component analysis; residue compensation

## I. INTRODUCTION

Image super-resolution, which is to enhance the resolution of images, is an active technology at present. Especially, the super-resolution of face images, which is called face hallucination, can be applied in long-distance video surveillance and video processing. However, this is still a challenging technology though many algorithms have been proposed for this issue. Among them, the learning-based approaches have received much attention. Baker and Kanade [1] first proposed a hallucination method which constructs the high frequency components from a “parent structure” resorting to the training set. Wang and Tang [2] presented a principal component analysis (PCA)-based face hallucination algorithm to globally infer the high-resolution face image. Liu et al. [3] developed a two-step statistical modeling approach, which integrates a global model and a local model, corresponding to the common and specific face characteristics, respectively. Although complicated probabilistic models are required in Liu et al.’s method, the idea of “two-step” became more and more popular since then. Another representative two-step approach was proposed by Zhuang et al. [4], which uses the locality preserving hallucination (LPH) and neighbor reconstruction for residue compensation. However, the global modeling of

these algorithms often results in a “fake hallucination”, which doesn’t look like the ground truth face.

Inspired by the recently proposed morphological component analysis (MCA), a novel three-step framework for face hallucination is proposed in this paper, where the face hallucination is formulated as an image decomposition problem. At the first step, the low-resolution input is up-sampled by interpolation. We observe that the interpolated image can be represented as a superposition of two typical components, i.e., a high-resolution image and an unsharp masking. In the second step, we use MCA to implement such decomposition so as to obtain the global approximation of the high-resolution image. The third step compensates the detail information to the estimated high-resolution image by using the neighbor reconstruction of patches. The proposed method addresses the face hallucination problem from a novel perspective, and effectually exploits the facial properties for global modeling via the combination of PCA and MCA.

## II. PROPOSED METHODOLOGY

### A. Problem formulation

The classical resolution reduction model for single image can be represented as

$$x_l = s \downarrow (h * x_h), \quad (1)$$

which explains the relationship between the observed low-resolution images  $x_l$  and the ground truth high-resolution image  $x_h$ .  $s \downarrow$  represents the down-sampling operator by a factor of  $s$ .  $h$  represents the camera’s point spread function (PSF) acting as a blurring operator, and  $*$  denotes the convolution operation. If we can obtain an up-sampling operator  $s \uparrow$  and apply it into (1):

$$s \uparrow x_l = s \uparrow s \downarrow (h * x_h), \quad (2)$$

then, the imaging model becomes as:

$$x_{zoom} = s \uparrow x_l \approx h * x_h. \quad (3)$$

In this paper, the up-sampling is simulated by the bilinear interpolation. So  $x_{zoom}$  denotes the interpolated high-resolution image from  $x_l$ . However,  $x_{zoom}$  is always blurry and can be regarded as a blurred version of an image  $x_h$ . They can be formulated as

\*Correspondence author. Tel.: (+86)20-84110175

$$x_{zoom} = x_h + x_s, \quad (4)$$

where  $x_s$  is a high-frequency component containing both positive and negative values, called unsharp masking [10]. In this model,  $x_{zoom}$  is the superposition of high-resolution part  $x_h$  and an unsharp masking  $x_s$ . For the face hallucination problem, we hope to separate these two parts in order to obtain  $x_h$ . In the next section, we investigate how to use the MCA for such decomposition.

### B. Face image decomposition using MCA

In a series of recent literatures [5][6][7], the MCA was developed for separating the texture part from the piecewise smooth part in an image. The basic idea behind MCA is to associate each part to a subdictionary of atoms, and each subdictionary can sparsely generate only its intended part but not the others. For a specific decomposition task, adaptive dictionary design is still an open question.

In this paper, we employ the MCA to decompose an up-sampled image  $x_{zoom}$  into a high-resolution image  $x_h$  and an unsharp masking  $x_s$ . Accordingly, a rational design for the subdictionaries  $\Phi_1$  and  $\Phi_2$ , associated to  $x_h$  and  $x_s$ , respectively, should meet the following four conditions:

1. Each subdictionary should fully contain various important properties that the intended part consists of, so that this part can be sparsely represented by these properties.

2. The two subdictionaries should be as incoherent as possible. Assuming that the columns of the subdictionaries are normalized to unit  $l_2$ -norm, the mutual coherence of  $\Phi_1$  and  $\Phi_2$  can be defined as:

$$\mu(\Phi_1, \Phi_2) = \max_{i \neq j} |\Phi_1^T \Phi_2(i, j)|, \quad (5)$$

where  $\Phi_1^T \Phi_2(i, j)$  denotes the entry in the  $i$ -th row and the  $j$ -th column of  $\Phi_1^T \Phi_2$ . The mutual coherence is a measurement of the dependence between the atoms of the two subdictionaries. The range of mutual coherence is [0, 1]. A smaller mutual coherence stands for more

distinguishable sparse representations between two considered subdictionaries.

3.  $\Phi_1$  and  $\Phi_2$  should have fast analysis and synthesis algorithms to make MCA computationally tractable.

4. More decomposition constraints can be designed from the dictionary properties.

In this paper, we adopt the PCA to design dictionaries, since PCA can properly describe the common facial properties. Given the training high-resolution face images and unsharp maskings, by PCA, we obtain the mean values  $\bar{x}_h$  and  $\bar{x}_s$ , eigenvalues  $\{\sigma_h^2(i)\}_{i=1}^N$  and  $\{\sigma_s^2(i)\}_{i=1}^N$ , and eigenvector matrices  $P_h$  and  $P_s$ , respectively. Accordingly,  $x_h$  and  $x_s$  can be respectively represented as

$$x_h \approx \bar{x}_h + P_h \alpha_h, \quad x_s \approx \bar{x}_s + P_s \alpha_s, \quad (6)$$

where  $\alpha_h$  and  $\alpha_s$  are the coefficients of projection. We analyze as follows that  $P_h$  and  $P_s$  can be used as the subdictionaries of  $x_h - \bar{x}_h$  and  $x_s - \bar{x}_s$ , respectively. Firstly, as the bases of eigen subspaces,  $P_h$  and  $P_s$  include the most important atoms to represent the objects. Secondly, in our experiment, the mutual coherence of  $P_h$  and  $P_s$  is 0.3303, on behalf of the incoherence between them. Thirdly, they are unitary matrices and easy for the implementation of analysis and synthesis. Finally, in the MCA-based decomposition,  $\alpha_h$  and  $\alpha_s$  can be adaptively rescaled into the allowed range, so as to produce a plausible face. Since eigenvalues are the sample variances along the eigenvectors, we constrain  $\alpha_h$  and  $\alpha_s$  within a hyper-ellipsoid about the origin, with the range of axis  $[-3\sigma_3(i), 3\sigma_3(i)]$ ,  $\mathfrak{S} = h, s$  for the projection coefficients on the  $i$ -th eigenvector.

Based on the above analysis, in order to extract  $x_h$  and  $x_s$  from  $x_{zoom}$ , we need to solve the following optimization problem:

$$\begin{aligned} \min_{\alpha_h, \alpha_s} & \|\alpha_h\|_0 + \|\alpha_s\|_0, \\ \text{s. t.} & \|x_{zoom} - \bar{x}_h - \bar{x}_s - P_h \alpha_h - P_s \alpha_s\|_2 \leq \varepsilon, \\ & -3\sigma_3(i) \leq \alpha_3(i) \leq 3\sigma_3(i), \quad \mathfrak{S} = h, s, \quad i = 1 \cdots N. \end{aligned} \quad (7)$$

The parameter  $\varepsilon$  trades off the sparseness and completeness.

Eq. (7) can be effectively solved in an iterative thresholding manner [9], which iteratively updates the solution one part a time as a ‘‘coarse-to-fine’’ scheme. At each iteration step, the threshold results are rescaled into the allowed hyper-ellipsoid range. The flow diagram of decomposing a face image using MCA is shown in Table I.

Note that, since  $P_h$  and  $P_s$  are not overcomplete,  $x_h$  and  $x_s$  cannot be perfectly reconstructed by using the coefficients obtained from (7), and only the global approximations  $x_h^g$  and  $x_s^g$  can be obtained. In the next step, we will introduce a residue compensation algorithm to enhance the facial details.

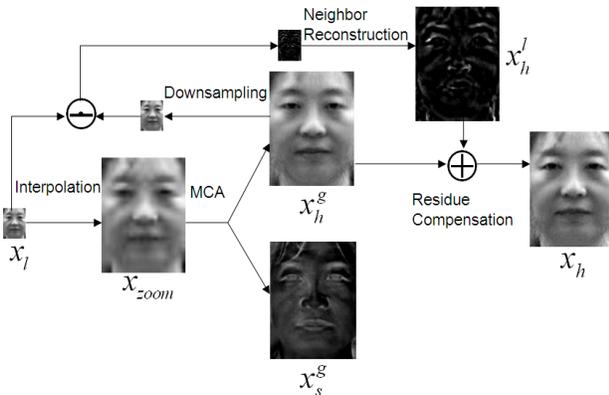


Figure 1. Diagram of the proposed framework, where the images are calibrated for display.

TABLE I. ALGORITHM FOR FACE IMAGE DECOMPOSITION

<p>1. <b>Input:</b> interpolated image <math>x_{zoom}</math>, samples mean values <math>\overline{x_h}</math> and <math>\overline{x_s}</math>, eigenvalues <math>\{\sigma_h^2(i)\}_{i=1}^N</math> and <math>\{\sigma_s^2(i)\}_{i=1}^N</math>, subdictionaries <math>P_h</math> and <math>P_s</math>, maximal number of iteration <math>I_{max}</math>, thresholds <math>minT_h</math>, <math>minT_s</math>, <math>maxT_h</math>, <math>maxT_s</math>.</p> <p>2. <b>Initiation:</b> <math>T_h = maxT_h</math>, <math>T_s = maxT_s</math>, <math>x_h = 0</math>, <math>x_s = 0</math>.</p> <p>3. <b>For</b> <math>i = 1 : I_{max}</math></p> <ul style="list-style-type: none"> <li>• Update <math>x_h</math> with <math>x_s</math> fixed:                     <ul style="list-style-type: none"> <li>-Calculate the residual <math>r = x_{zoom} - \overline{x_h} - \overline{x_s} - x_h - x_s</math>.</li> <li>-Calculate the projection of <math>x_h + r</math> onto <math>P_h</math>:                             <math display="block">\alpha_h = P_h^T (x_h + r).</math> </li> <li>-Hard threshold the coefficient <math>\alpha_h</math> with threshold <math>T_h</math> and obtain <math>\alpha_h^1</math>.</li> <li>-Rescale <math>\alpha_h^1</math> for a plausible face to obtain <math>\alpha_h^2</math>.</li> <li>-Reconstruct <math>x_h = \overline{x_h} + P_h \alpha_h^2</math>.</li> </ul> </li> <li>• Update <math>x_s</math> with <math>x_h</math> fixed:                     <ul style="list-style-type: none"> <li>-Calculate the residual <math>r = x_{zoom} - \overline{x_h} - \overline{x_s} - x_h - x_s</math>.</li> <li>-Calculate the projection of <math>x_s + r</math> onto <math>P_s</math>:                             <math display="block">\alpha_s = P_s^T (x_s + r).</math> </li> <li>-Hard threshold the coefficient <math>\alpha_s</math> with threshold <math>T_s</math> and obtain <math>\alpha_s^1</math>.</li> <li>-Rescale <math>\alpha_s^1</math> for a plausible face to obtain <math>\alpha_s^2</math>.</li> <li>-Reconstruct <math>x_s = \overline{x_s} + P_s \alpha_s^2</math>.</li> </ul> </li> <li>• Update the thresholds by                             <math display="block">T_h = T_h - \frac{maxT_h - minT_h}{I_{max} - 1}, T_s = T_s - \frac{maxT_s - minT_s}{I_{max} - 1}.</math> </li> </ul> <p><b>End.</b></p> <p>4. <b>Output</b> <math>x_h</math>.</p>
--

### C. Residue compensation

Residue compensation [3][4] is widely used to compensate the detail information  $x_h^l$  to  $x_h^g$ . In this paper, neighbor reconstruction [4] is selected for this task. The basic idea of neighbor reconstruction is to consider each residue image as a patch matrix composed of overlapped square patches. For a given low-resolution residue patch, it can be reconstructed by a combination of k-nearest neighboring low-resolution training residue patches. Then, we can synthesize the corresponding high-resolution residue patch by replacing the low-resolution training residue patches with high-resolution training residue patches, while maintaining the same combination weights. Finally, the high-resolution residue face  $x_h^l$  is obtained by integrating these high-resolution residue patches by averaging the overlapped parts. The derived high-resolution residue face is

then added to the global face to obtain the final face hallucination result:

$$x_h = x_h^l + x_h^g. \quad (8)$$

Fig.1 shows the diagram of the proposed framework.

### III. EXPERIMENTAL RESULTS

The performance of the proposed method was evaluated in terms of the image quality of the hallucinated faces. The experiment was conducted on the CAS-PEAL-R1 face database [8], which consists of 99594 images of 1040 individuals. In the experiment, a subset of CAS-PEAL established by selecting the face images of all persons with normal condition, i.e., frontal pose, even illumination, and neutral expression, was used. This subset contains 1040 images, one image per person. All face images were simply aligned according to the locations of eyes and mouths. Each original high-resolution face image is with the size  $96 \times 128$ . By smoothing and down-sampling, a corresponding low-resolution image with size  $24 \times 32$  was obtained from each high-resolution image. For MCA, the maximum thresholds were automatically set by using the maximum absolute value of all coefficients, i.e.,  $maxT_h = \|P_h^T (x_{zoom} - \overline{x_h} - \overline{x_s})\|_\infty$  and  $maxT_s = \|P_s^T (x_{zoom} - \overline{x_h} - \overline{x_s})\|_\infty$ . Each minimum stop threshold was set as the square root of the  $t$ -th eigenvalue, so that the 99% of cumulative eigen ratios were attained by the first  $t$  principal components. Furthermore, the image quality was satisfied when the number of iteration was larger than 50.

We compared our method with some state-of-the-art methods, including Blake and Kanade's [1], Wang and Tang's [2], Liu et al.'s [3], and Zhuang et al.'s [4]. 1000 image pairs were chosen as the training set and the remaining image pairs were used for test. For the methods with multiple steps, 750 image pairs were used for training the global model, and 250 ones were used for training the residue compensation model. The processed results of different methods are shown in Fig. 2. It shows that Blake and Kanade's method gets sharp but noisy face images, and loses some global properties such as the symmetry of facial features. Wang and Tang's method can achieve clear results but may cause "ringing" effect, especially on the portions around face contour. The results of Liu et al.'s and Zhuang et al.'s method preserve both the global and local facial features well. In comparison, our results are more similar to the original high-resolution faces. To quantify the image quality, we also calculated the Mean Square Error (MSE), which is defined by

$$MSE = \sum_{i=1}^{N_1} \|I_{test,i}^h - J_{test,i}^h\|^2 / (WHN_1), \quad (9)$$

where  $I_{test,i}^h$  is the original high-resolution face,  $J_{test,i}^h$  is the hallucinated face.  $W$  is the width of the image,  $H$  is the height of the image, and  $N_1$  is the number of the test images. The MSE of each method is shown in Fig. 3. It shows that our method gets the lowest MSE.

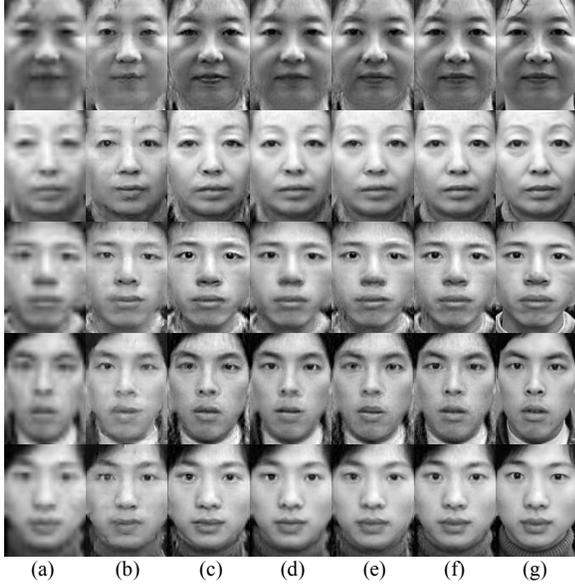


Figure 2. Face hallucination by using different methods: (a) Bilinear interpolation; (b) Blake and Kanade's; (c) Wang and Tang's; (d) Liu et al.'s; (e) Zhuang et al.'s; (f) Proposed method; (g) Original high-resolution images.

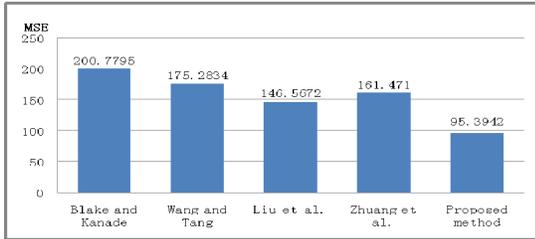


Figure 3. MSEs of different methods.

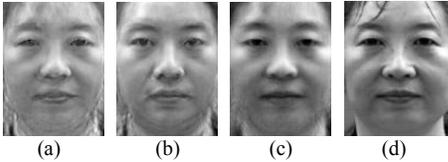


Figure 4. Comparison of the global face results: (a) Liu et al.'s; (b) Zhuang et al.'s; (c) Proposed method; (d) Original high-resolution image.

TABLE II. MSE OF THE PROPOSED METHOD.

	Step One	Step Two	Step Three
MSE	164.0701	147.3466	95.3942

In order to investigate the performance of each step of the proposed method, we also illustrate the comparison of the global faces produced by different methods in Fig. 4, and the

MSE of the proposed method step by step in Table II. They show that our method generated the best global face among the mentioned methods. Furthermore, each step has important contribution in the proposed method.

#### IV. CONCLUSION

A novel three-step face hallucination method has been presented. In this paper, the face hallucination was formulated as an image decomposition problem, that is, how to decompose an interpolated face image into a high-resolution image and an unsharp masking. Based on this model, the MCA was developed to obtain an approximate high-resolution image. Furthermore, a residue compensation algorithm was used to refine the hallucinated results. Under the image decomposition perspective, the facial properties are effectually exploited, ensuring a photorealistic hallucinated face image.

As one of the most important issue in MCA-based image decomposition, dictionary design will be a research focus in our future work.

#### ACKNOWLEDGMENT

This work was supported by NSFC (U0835005, 60633030), the 973 Program (2006CB303104), and the GuangDong Program (2008A090400013). Portions of the research in this paper use the CAS-PEAL face database collected under the sponsor of the Chinese National Hi-Tech Program and ISVISION Tech. Co. Ltd.

#### REFERENCES

- [1] S. Baker, T. Kanade, "Hallucinating faces," IEEE Int. Conf. Automatic Face and Gesture Recognition, 83-88, 2000.
- [2] X. G. Wang and X. O. Tang, "Hallucinating face by eigentransformation," IEEE Trans. SMC-C, 35(3):425-434, 2005.
- [3] C. Liu, H.-Y. Shum, and W. T. Freeman, "Face hallucination: theory and practice," IJCV, 75(1):115-134, 2007.
- [4] Y. Zhuang, J. Zhang, and F. Wu, "Hallucinating face: LPH superresolution and neighbor reconstruction for residue compensation," Pattern Recognition, 40(11):3178-3194, 2007.
- [5] J.-L. Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component analysis," Adv. in Imag. and Electron Phys., 132, 2004.
- [6] J.-L. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," IEEE Trans. Image Processing, 14:1570-1582, 2005.
- [7] J. Bobin, J.-L. Starck, J. Fadili, Y. Moudden, and D. Donoho, "Morphological component analysis: an adaptive thresholding strategy," IEEE Trans. Image Processing, 16(11): 2675-2681, November 2007.
- [8] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The CAS-PEAL large-scale Chinese face database and baseline evaluations," IEEE Trans. SMC, 149-161, 2008.
- [9] S. Sardy, A. G. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric signal denoising with wavelet dictionaries," J. Comput. Graphical Statist., 9 (2000):361-379, 2000.
- [10] R. C. Gonzalez and R. E. Woods, "Digital image processing," Pearson Education, 2002.