

Hallucinating Faces: Global Linear Modal Based Super-Resolution and Position Based Residue Compensation

Xiang Ma¹, Junping Zhang², and Chun Qi¹

¹ School of Electronics & Information Engineering, Xi'an Jiaotong University, Xi'an, China

² Department of Computer Science and Engineering, Fudan University, Shanghai, China
maxiangmail@163.com

Abstract. A learning-based face hallucination method is proposed in this paper for the reconstruction of a high-resolution face image from a low-resolution observation based on a set of high- and low-resolution training image pairs. The proposed global linear modal based super-resolution estimates the optimal weights of all the low-resolution training images and a high-resolution image is obtained by applying the estimated weights to the high-resolution space. Then, we propose a position based local residue compensation algorithm to better recover subtle details of face. Experiments demonstrate that our method has advantage over some established methods.

Keywords: Face hallucination; Super-resolution; Residue compensation.

1 Introduction

Super-resolution addresses various applications in a variety of important sectors, as diverse as medical imaging, satellite imaging, surveillance system, image enlarging in web pages, and restoration of old historic photographs. Super-resolutions are roughly separated into two classes [8]: general super-resolution [4], which extracts a single high-resolution image from a sequence of general low-resolution images, and domain specific super-resolution [1], [2], [5], [6], [8], which extracts high-resolution image details from a restricted class of high-resolution images, such as in the face domain. In this study, we focus on face image super-resolution, which is also known as face hallucination. Face hallucination is to reconstruct a high-resolution face image from a low-resolution one based on a set of high- and low-resolution training image pairs.

Baker *et al.* [1] was the first to develop a face hallucination method under a Bayesian formulation and proposed the term face hallucination. In this method, it generates the high frequency details from a parent structure with the assistance of training samples. Liu *et al.* [2] adopted a two-step approach integrating a global parametric model with Gaussian assumption and a local nonparametric model based on Markov random field (MRF). Both of the two methods used complicated probabilistic models and were based on an explicit down-sampling function, which is sometimes unavailable in practice. Inspired by locally linear embedding (LLE), a well-known manifold learning

method, Chang *et al.* [3] proposed Neighbor Embedding algorithm based on assumption that small patches in the low- and high-resolution images form manifolds with similar local geometry in two distinct spaces. Wang *et al.* [6] suggested a face hallucination method using principal component analysis (PCA) to represent the structural similarity of face images. Zhuang *et al.* [5] developed a two-phase face hallucination. The locality preserving hallucination algorithm combines locality preserving projection (LPP) and Radial basis function (RBF) regression together to hallucinate the global high-resolution face. The details of the synthetic high-resolution face are further improved by residue compensation based on Neighbor Embedding. Jia *et al.* [8] applied hierarchical tensor algebra to face hallucination to solve different facial modalities. His super-resolution approach still integrates probability models with Gaussian assumption. Park *et al.* [9] proposed a novel example-based face hallucination method with an extended morphable face model. This method has advantages in enhancing the resolution of face image and improving the performance of face recognition.

A global linear modal based super-resolution is proposed in our study to reconstruct the face image keeping the main characteristics of the ground truth face instead of using a manifold learning or extended morphable face model. Furthermore, a residual face containing the high-frequency information was produced and piled onto the result of step one by proposed local residue compensation algorithm based on position to better the final result.

The rest of the paper is organized as follows: In Section 2, we will describe the proposed method in detail. Section 3 shows experiments results, and Section 4 concludes the paper.

2 Proposed Face Hallucination Method

2.1 Global Linear Modal Based Super-Resolution

In [8], a novel face image was represented by a linear combination of prototypes [9]. In our method, all training images are used as a linear combination to represent a new facial image input. A low-resolution face image input is denoted as I . Suppose that H^m are the high-resolution training images, whose low-resolution correspondences are represented as L^m , $m = 1, \dots, M$, where M is the number of training image pairs. Face image is represented as a column vector of all pixel values in algorithm.

We expect

$$I \cong \sum_{m=1}^M w_m L^m = P \quad (1)$$

where P is the reconstructed image estimated with the linear combination of all the low-resolution training images. w_1, w_2, \dots, w_M are the reconstruction weights of the low-resolution training images which are constrained to have a sum of one.

The reconstruction error $E(w)$ is defined as follows:

$$E(w) = \|I - P\|^2 \quad (2)$$

Equation (2) represents the difference between the input facial image and its reconstructed equivalent

The optimal reconstruction weights are based on the minimization of the reconstruction error $E(w)$:

$$w = \arg \min_{w_m} \|I - P\|^2 \quad (3)$$

where w is a M-dimensional weight vector by stacking each reconstruction weights.

Let $\mathbf{Z} = (\mathbf{X} - \mathbf{Y})^T (\mathbf{X} - \mathbf{Y})$, where $\mathbf{X} = L \cdot C^T$, C is a column vector of ones and \mathbf{Y} is a matrix with its columns being the training images L^1, L^2, \dots, L^M .

Equation (3) is a constrained least squares problem which has the following solution

$$w = (\mathbf{Z}^{-1}C) / (C^T \mathbf{Z}^{-1}C) \quad (4)$$

The estimated weight w maintains the minimization of the reconstruction error $E(w)$, however, the variance of w is relatively large, which makes the reconstructed image unsmooth. To reduce the variance of w , we let $\mathbf{Z} = \mathbf{Z} + \mathbf{I} \cdot a$, where a is a large number and \mathbf{I} is a diagonal matrix of ones. Thus, the solution for w is no longer optimal but deviated.

Replacing each low-resolution image L^m by its corresponding high-resolution sample H^m in (1), the result is denoted as \tilde{P} , we have

$$\sum_{m=1}^M w_m H^m = \tilde{P} \quad (5)$$

Equation (5) shows that \tilde{P} is the linear combination of the high-resolution training image, so it should be approximately face-like at a high-resolution level.

The acquisition process can be expressed as [1], [2]:

$$I_L = \frac{1}{q^2} \sum_{k=0}^{q-1} \sum_{l=0}^{q-1} I_H (qi + k, qj + l) + n \quad (6)$$

where I_H is the ground truth high-resolution image which is q^2 times larger than I_L , q is a positive integer and n is the random noise.

For notation simplification, if I_H , I_L and n are respectively vectors, equation (6) can be rewritten as:

$$I_L = AI_H + n \quad (7)$$

where A is a matrix. Equation (7) combines a smoothing and a down-sampled step.

From (7) and (5), without consideration of noise disturbance, we have

$$\begin{aligned}
 A\tilde{P} &= \sum_{m=1}^M w_m AH^m = w_1 AH^1 + w_2 AH^2 + \dots + w_M AH^M \\
 &= w_1 L^1 + w_2 L^2 + \dots + w_M L^M = \sum_{m=1}^M w_m L^m = P
 \end{aligned}
 \tag{8}$$

From (1) and (8), we have

$$A\tilde{P} = P \cong I \tag{9}$$

Equation (9) shows that the degradation of \tilde{P} is close to the low-resolution input image I . Since the weight w is not computed from high-resolution training images, the image \tilde{P} needs further step to better recover face subtle details.

2.2 Local Residue Compensation Based on Position

In this section, we propose the residue compensation algorithm based on position. In this step, a residue image \tilde{T} is estimated and the final result H^* is obtained by adding \tilde{T} to \tilde{P} .

The input of residue compensation is obtained by subtracting the degradation version of the result \tilde{P} from the original low-resolution input face I :

$$T = I - A(\tilde{P}) \tag{10}$$

The low-resolution training image R_L^m for residue compensation can be obtained by the following equation:

$$R_L^m = L^m - A(H_1^m) \tag{11}$$

where H_1^m is the result gained by global face modal based face hallucination method using L^m as low-resolution image input, of course, the training pair number is changed from M to $M - 1$.

The high-resolution training image R_H^m for residue compensation can be obtained by the following equation:

$$R_H^m = H^m - H_1^m \tag{12}$$

The obtained training images are respectively considered as a patch matrix composed of overlapping square patches: $\{R_L^{mP}(i, j)\}_{m=1}^N$ 、 $\{R_H^{mP}(i, j)\}_{m=1}^N$, where N is the number of the patches in image. The patch located at row i and column j in the patch matrix can be denoted as $R_L^{mP}(i, j)$ or $R_H^{mP}(i, j)$. The patch $R_L^{mP}(i, j)$ and its adjacent patches are shown in Fig. 1.

Suppose that the patch covers $n \times n$ pixels. To low-resolution image training set, if n is an odd number, the patch $R_L^{mP}(i, j)$ overlaps with its adjacent patches by size $[(n-1)/2] \times [(n-1)/2]$, and its corresponding high-resolution patch $R_H^{mP}(i, j)$ which covers $qn \times qn$ pixels overlaps with its adjacent patches by size $[q(n-1)/2] \times [q(n-1)/2]$; if n is an even number, the patch $R_L^P(i, j)$ overlaps with its adjacent patches by size $(n/2) \times (n/2)$, and its corresponding high-resolution patch $R_H^{mP}(i, j)$ covering $qn \times qn$ pixels overlaps with its adjacent patches by size $[qn/2] \times [qn/2]$.

T is also represented in patches: $\{T^{mP}(i, j)\}_{m=1}^N$. From (1), for each image patch located at position (i, j) , we have:

$$T^{mP}(i, j) \cong \sum_{m=1}^M w_m(i, j) \cdot R_L^{mP}(i, j) \tag{13}$$

where $w_m(i, j)$ represent the reconstruction weights of the low-resolution training image patches, which are constrained to have a sum of one.

From (1)-(5), $w_m(i, j)$ are obtained by the following equation:

$$w(i, j) = (\mathbf{Z}^{-1}C) / (C^T \mathbf{Z}^{-1}C) \tag{14}$$

where $\mathbf{Z} = (X - \mathbf{Y})^T (X - \mathbf{Y})$, $X = T^P(i, j) \cdot C^T$, C is a column vector of ones and \mathbf{Y} is a matrix with its columns being the training images $R_L^{1P}(i, j), R_L^{2P}(i, j), \dots, R_L^{MP}(i, j)$, $w(i, j)$ is a M-dimensional weight vector by stacking each reconstruction weights $w_m(i, j)$.

Replacing each low-resolution image patch $R_L^{mP}(i, j)$ by its corresponding high-resolution patch $R_H^{mP}(i, j)$ in (13), the result is denoted as:

$$\sum_{m=1}^M w_m(i, j) \cdot R_H^{mP}(i, j) = \tilde{T}^{mP}(i, j) \tag{15}$$

where $\tilde{T}^{mP}(i, j)$ represents the reconstructed image patch estimated by the linear combination of the high-resolution training image patches.

According to their original positions, all the reconstructed patches are integrated to form the final image $\{\tilde{T}^{mP}(i, j)\}_{m=1}^N$, which is also denoted in global image \tilde{T} . Pixels of the overlapping regions in the result are obtained by averaging the pixels value in the overlapping regions between two adjacent patches.

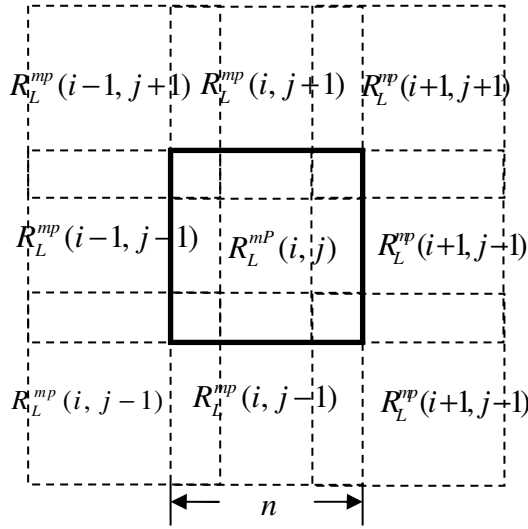


Fig. 1. Image patch $R_L^{mp}(i, j)$ with its adjacent image patches

The final result X^* is obtained by adding the \tilde{T} to the step-one result \tilde{P} .

$$X^* = \tilde{P} + \tilde{T} \tag{16}$$

where X^* is the final hallucinated image.

The proposed face hallucination method is summarized as follows:

- Step1. Approximate the low-resolution face input by the linear combination of all the low-resolution training images.
- Step2. Compute the reconstruction weight w_m and synthesize the result image \tilde{P} .
- Step3. Obtain the face residue input $\{T^{mp}(i, j)\}_{m=1}^N$ and generate the pair-wise training images $\{R_L^{mp}(i, j)\}_{m=1}^N$ and $\{R_H^{mp}(i, j)\}_{m=1}^N$ for residue compensation.
- Step4. For each patch $T^{mp}(i, j)$ located at (i, j) :
 - (a) Compute the reconstruction weight $w_m(i, j)$.
 - (b) Reconstruct the patch $\tilde{T}^{mp}(i, j)$.
- Step5. Integrate all the hallucinated residue patches to form a residue image \tilde{T} .
- Step6. Add the residue image \tilde{T} to \tilde{P} , the final hallucinated image X^* is produced.

3 Experiment Results

Our face hallucination method was performed on the CAS-PEAL Face Database [7]. The 290 normal expression images of different peoples under the same light condition were randomly selected and aligned manually according to the locations of 3 points: centers of left and right eyeballs and center of the mouth. We cut out the region of the face according to interesting. All faces were standardized to the size of 128×96 .

The 20 high-resolution images were blurred using a 7×7 Gaussian filter with $\sigma = 0.85$, and down-sampled to 32×24 for test. The rest of the images were used as high-resolution training images, and down-sampled to 32×24 for using as low-resolution training images.

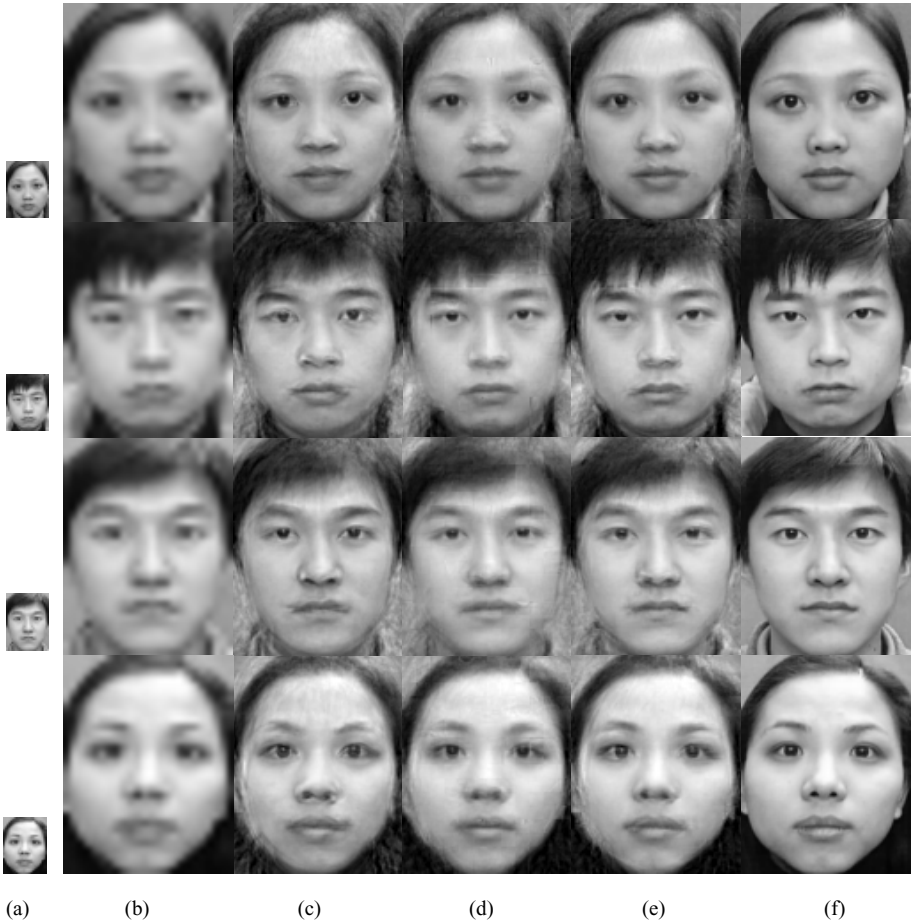


Fig. 2. The hallucination results. (a) The input 24×32 low-resolution faces. (b) Cubic B-Spline. (c) Wang's method. (d) Zhuang's method. (e) Our method. (f) The original 96×128 high-resolution faces.

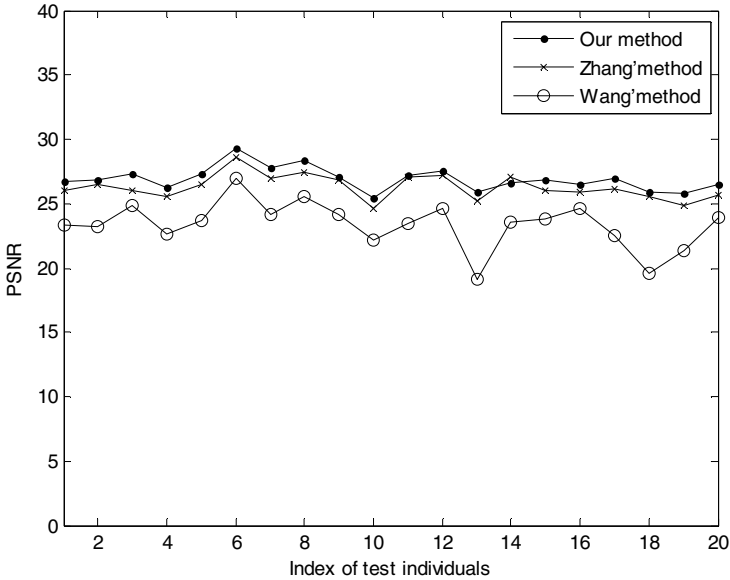


Fig. 3. PSNR values of the hallucinated results from three different methods

The size of 3×3 low-resolution patch was used in residue compensation. We compared our method with some methods based on the same training set. These methods are Cubic-B-Spline, Wang's eigentransformation method [6] and Zhuang's Locality Preserving method [5]. The image pairs of 150 people were used for training in Zhuang's method [5] and our method. In order to achieve the optimal results in Wang's method [6], we used image pairs of 270 for training and let Variance contribution rate of PCA be 0.9999; We selected $h = 135$, $K_1 = 8$, $K_2 = 5$ mentioned in [5] to optimize the results of Zhuang's method [5]. The size of the image patch in our method is 3×3 .

Some resultant images are shown in Fig. 2. To quantify the performance of the results, we also computed the peak signal-to-noise ration (PSNR) of each method shown in Fig. 3. It can see that our method has the highest PSNR values compared to the values from other methods for all test faces. Based on the comparisons, it is concluded that our method has advantage over current established methods.

4 Conclusion

A novel approach to hallucinating faces is presented in this study. Experiments show that the proposed method generates results with the best image quality compared to some established algorithms.

Acknowledgments

This work was supported in part by the grants from the National Natural Science Foundation of China under number 60641002 and in part by the National High- Tech

Research and Development Program (“863”program) of China under number 2007AA01Z176 .

The research in this paper use the CAS-PEAL-R1 face database collected under the sponsorship of the Chinese National Hi-Tech Program and ISVISION Tech. Co. Ltd.

References

- [1] Baker, S., Kanade, T.: Hallucinating Faces. In: IEEE Inter. Conf. on Automatic Face and Gesture Recognition, pp. 83–88. IEEE Press, France (2000)
- [2] Liu, C., Shum, H.Y., Zhang, C.S.: A two-step approach to hallucinating faces: global parametric model and local nonparametric model. In: Inter. Conference on Image and Graphics, pp. 192–198. IEEE Press, New York (2001)
- [3] Chang, H., Yeung, D.-Y., Xiong, Y.: Super-resolution through neighbor embedding. In: Inter. Conference on Computer Vision and Pattern Recognition, pp. 1275–1282. IEEE Press, Washington (2004)
- [4] Costa, G.H., Bermudez, J.C.M.: Statistical analysis of the LMS algorithm applied to super resolution image reconstruction. *IEEE Transactions on Signal Processing* 11, 2084–2095 (2007)
- [5] Zhuang, Y., Zhang, J., Wu, F.: Hallucinating faces: LPH super-resolution and neighbor reconstruction for residue compensation. *Pattern Recognition* 40, 3178–3194 (2007)
- [6] Wang, X., Tang, X.: Hallucinating Face by Eigentransformation. *IEEE Transactions on Systems Man and Cybernetics* 35, 425–434 (2005)
- [7] Gao, W., et al.: The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. *IEEE Transactions on System Man, and Cybernetics (Part A)* 38, 149–161 (2008)
- [8] Jia, K., Gong, S.G.: Generalized face super-resolution. *IEEE Transactions on Image Processing* 17, 873–886 (2008)
- [9] Park, J.S.S., Lee, W.: An Example-Based Face Hallucination Method for Single-Frame, Low-resolution Facial Images. *IEEE Transactions on Image Processing* 17, 1806–1816 (2008)