POSITION-BASED FACE HALLUCINATION METHOD

Xiang Ma[1], Junping Zhang[2] and Chun Qi[1]

[1]School of Electronics& Information Engineering, Xi'an Jiaotong University, Xi'an, China
[2]Department of Computer Science and Engineering, Fudan University, Shanghai, China

ABSTRACT

In this paper, we propose a novel face hallucination method to reconstruct a high-resolution face image from a low-resolution observation based on a set of high- and low-resolution local training image pairs. Instead of basing on probabilistic or manifold learning models, the proposed method synthesizes the high-resolution image patch using the same position image patches of training image pairs. A cost function is formulated to obtain the optimal weights of the training image position-patches and the high-resolution patches are reconstructed using the same weights. The final high-resolution facial image is formed by integrating the hallucinated patches. Experiments show that the proposed method without residue compensation generates higher-quality images than some methods.

Index Terms— Face hallucination, super-resolution, position-patch

1. INTRODUCTION

In recent years, a number of super-resolution techniques have been proposed, which are roughly separated into two classes: multiple-frame super-resolution and single-frame super-resolution. In our method, we focus on single-frame face image super-resolution, which is also known as "face hallucination". Face hallucination is to infer a high-resolution face image from a low-resolution one based on a set of high- and low-resolution training image pairs, which incorporates image super-resolution techniques into facial image synthesis.

The common face hallucination algorithms usually perform two steps: the first step generates global face image which keeps the main characteristics of the ground truth image using probabilistic method in maximum a posteriori (MAP) frame or manifold learning method such as locally linear embedding (LLE); the second step produces residual image to compensate the result of the first step.

Baker et al. [2] firstly developed a hallucination method under a Bayesian formulation and proposed the term "face hallucination". In the method, it generates the high frequency details from a parent structure with the assistance of training samples. Liu et al. [3] developed 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 use complicated probabilistic models and are based on an explicit down-sampling function, which is sometimes unavailable in practice. Inspired by LLE, a well-known manifold learning method, Chang et al [4] developed the Neighbor Embedding algorithm based on the assumption that small patches in the low- and high-resolution images form manifolds with similar local geometry in two distinct spaces. Motivated by Chang's work, a number of face hallucinations were developed basing on Neighbor Embedding or using neighbor patch. Wang et al. [1] suggested the use of PCA to represent the structural similarity of face images, which can hardly maintain the global smoothness and visual rationality, especially on the locations around the face contour and the margin of the mouth. 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 a global high-resolution face. The details of the synthetic high-resolution face are further improved by residue compensation based on Neighbor Embedding.

Different from the common face hallucination, the proposed method does not incorporate the dimensionality reduction methods into the operation process and the additional residue compensation is no longer necessary in our method because the non-feature information is not lost and the reconstruction weights contain not only feature information but also non-feature information. The image position-patches are used for the reconstruction of the high-resolution image, and the position-patches need not to be searched in all the training image patches. Compared with neighbor patches that are widely used in face hallucination, position-patches lead to more satisfactory results.

The rest of the paper is organized as follows: Section 2 describes the proposed approach in detail. Section3 shows convincible experiments and analysis, and Section4 concludes the paper.

2. PROPOSED FACE HALLUCINATION METHOD

Each low- and high-resolution training face image of Y_L

and Y_H can be represented as a set of small overlapped image patches and considered as patch matrix $\{Y_L^{mP}(i,j)\}_{p=1}^{N}$、$\{Y_H^{mP}(i,j)\}_{p=1}^{N}$, where N is the number of the patches in image，$m=1,2,...,M$, M is the number of training image pairs, $Y_L^{P}(i,j)$ and $Y_H^{P}(i,j)$ are the patches located at the i th row and the j th column in the patch matrix. The term (i,j) denotes the position information of each patch. Except the patches on the border of the image, all patches overlapped horizontally and vertically with each other. $Y_L^{P}(i,j)$ and $Y_H^{P}(i,j)$ are the local training image pairs at the position (i,j).

The low-resolution image input X_L is also represented as a set of small overlapped image patches. A two-dimensional face image or patch is represented as a column vector of all pixel values. Because of the structural similarity, face image can be reconstructed from the optimal linear combination of the training face images [1]. We expect each patch $X_L^{P}(i,j)$ in the input low-resolution face image $\{X_L^{P}(i,j)\}_{p=1}^{N}$ can be represented by:

$$X_L^{P}(i,j) \cong \sum_{m=1}^{M} w_m(i,j)Y_L^{mP}(i,j) = \tilde{X}_L^{P}(i,j)$$

(1)

where $\tilde{X}_L^{P}(i,j)$ is the reconstructed image patch estimated with the linear combination of the low-resolution training image patches, $w_m(i,j)$ represents the contribution of each training image patch located in position (i,j) to the reconstruction of the input face image's patch located in the same position. It meets the following equation:

$$\sum_{m=1}^{M} w_m(i,j) = 1$$

(2)

For training image, the image patches locating in position (i,j) such as $Y^{mP}(i,j)$ are defined as position-patches of the image patch $X^{P}(i,j)$.

For each patch $X_L^{P}(i,j)$, the reconstruction error θ is defined as follows:

$$\theta = \left\| X_L^{P}(i,j) - \tilde{X}_L^{P}(i,j) \right\|^2$$

(3)

The optimal reconstruction weights are based on the minimization of the reconstruction error θ:

$$w(i,j) = \arg\min_{w_m(i,j)} \left\| X_L^{P}(i,j) - \sum_{m=1}^{M} w_m(i,j)Y_L^{mP}(i,j) \right\|^2$$

(4)

where $w(i,j)$ is a M-dimensional weight vector by stacking each reconstruction weight $w_m(i,j)$.

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

Equation (4) is a constrained least squares problem which has the following solution:

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

(5)

It's known that the acquisition process can be expressed as [2], [3]:

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

(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:

$$I_L = HI_H + n$$

(7)

where H is a matrix. Equation (7) combines a smoothing and a down-sample step.

Replacing each low-resolution image patch $Y_L^{1P}(i,j)$, $Y_L^{2P}(i,j)$, ... , $Y_L^{MP}(i,j)$ by its corresponding high-resolution patch $Y_H^{1P}(i,j)$, $Y_H^{2P}(i,j)$, ... , $Y_H^{MP}(i,j)$ in (1), the result is denoted as $\tilde{X}_H^{P}(i,j)$, we have

$$\sum_{m=1}^{M} w_m(i,j)Y_H^{mP}(i,j) = \tilde{X}_H^{P}(i,j)$$

(8)

Equation (8) shows that $\tilde{X}_H^{P}(i,j)$ is the linear combination of the high-resolution training image patches, so it should be approximately face-like at a high-resolution level.

From (1), (7) and (8), without consideration of noise disturbance, we have:

$$H \cdot \tilde{X}_H^{P}(i,j) = \sum_{m=1}^{M} w_m(i,j) \cdot H \cdot Y_H^{mP}(i,j)$$

$$= \sum_{m=1}^{M} w_m(i,j) \cdot Y_L^{mP}(i,j) \cong X_L^P(i,j)$$

$$(9)$$

Equation (9) shows that the degradation of $\tilde{X}_H^P(i,j)$ is close to the low-resolution patch $X_L^P(i,j)$.

All the reconstructed patches $\tilde{X}_H^P(i,j)$ are integrated to form the final global image according to the original position (i,j). The pixels of the overlapped regions in the final result are obtained by averaging the pixels value in the overlapped regions between two adjacent patches.

The proposed face hallucination method is summarized as follows:

Step1: Denote the low resolution image input, low resolution training image and high resolution training image in patches respectively as $\{X_L^P(i,j)\}_{p=1}^N$, $\{Y_L^{mP}(i,j)\}_{p=1}^N$ and $\{Y_H^{mP}(i,j)\}_{p=1}^N$.

Step2: For each low-resolution patch $X_L^P(i,j)$ in low-resolution image input:

(a) Compute the reconstruction weights $w(i,j)$

(b) Reconstruct the patch $\tilde{X}_H^P(i,j)$

Step3: Concatenate and integrate the hallucinated patches to form a facial image $\{\tilde{X}_H^P(i,j)\}_{p=1}^N$ according to their original positions, which is the final result.

3. EXPERIMENTAL RESULTS

Our face hallucination method was performed on the CAS-PEAL Face Database [6]. We randomly selected 290 normal expression images of different persons on the same light condition. We aligned these face images manually and marked the locations of 3 points: the centers of the eyeballs and the center of the mouth. According to demand, we cutout the interesting region of the faces and unified the images to the size of 128×96.

We compare our method with recent methods based on the same training set. These methods are Cubic-B-Spine, Zhuang's method [5], Chang's Neighbor Embedding [4] and Wang's method [1]. The optimal patch size of 32×32 is chosen in Neighbor Embedding method, whose corresponding low-resolution size is 8×8. The number of the neighbor-patches for reconstruction in Neighbor Embedding is 150. The image pairs of 150 people are used for training in Zhuang's method, Neighbor Embedding and our method. In order to achieve the better results in Wang's method, we use image pairs of 270 people for training and let variance contribution rate of PCA be 0.9999. We select

$h=135$, $K_1=8$, $K_2=5$ mentioned in [5] to optimize the results of Zhuang's method. The size of the low-resolution image patch our method used is 3×3. The patches overlapped horizontally and vertically with each other by 1 pixel.

The image of 20 people were used as test image which were blurred using a 7×7 Gaussian filter with $\sigma=0.85$, and down-sample to 32×24.

Some representative results are shown in Fig. 1.

The hallucinated results of Wang's method can hardly maintain global smoothness and visual rationality, especially on locations around face contour and margin of the mouth. Zhuang's method generates more detailed facial features than Wang's, but there is more noise at the local facial features such as face contour, nostril and eyebrow. Because LPP is adopted in the first step which loses non-feature information, and the first step based on global image give rise to low reconstruction precision. To improve the first step's results, Zhuang inevitably developed the residue compensation based on patch in the second step. The final results rely on the residue compensation. Some subtle characteristics are blurred in the results of Neighbor Embedding.

We also compute the peak signal-to-noise ratio (PSNR) of each method. All the results are shown in Fig. 2. We can see that our method has the highest PSNR values compared with other methods on all test face.

With regard to the computational complexity and the execution time of the proposed method, Zhuang's method takes about more than 10 minutes by PC with 3.0G CPU to compute an image including the generation of the residue training set, Neighbor Embedding takes about more than 4 minutes because it needs to search neighbor patches in image training and our method only takes more than 1 minute. And Wang's method maintains high computational speed because it performs in global image way.

4. CONCLUSION

The proposed method without residue compensation has the best results than three other methods. Because non-feature information contributes to super-resolution, the proposed approach does not incorporate the dimensionality reduction methods into the operation process and the non-feature information is not lost. The reconstruction weights contain not only feature information but also non-feature information in our method, which lead to satisfactory image quality. Furthermore, the high-resolution image patch is synthesized without searching step, which also reduces the computational cost.
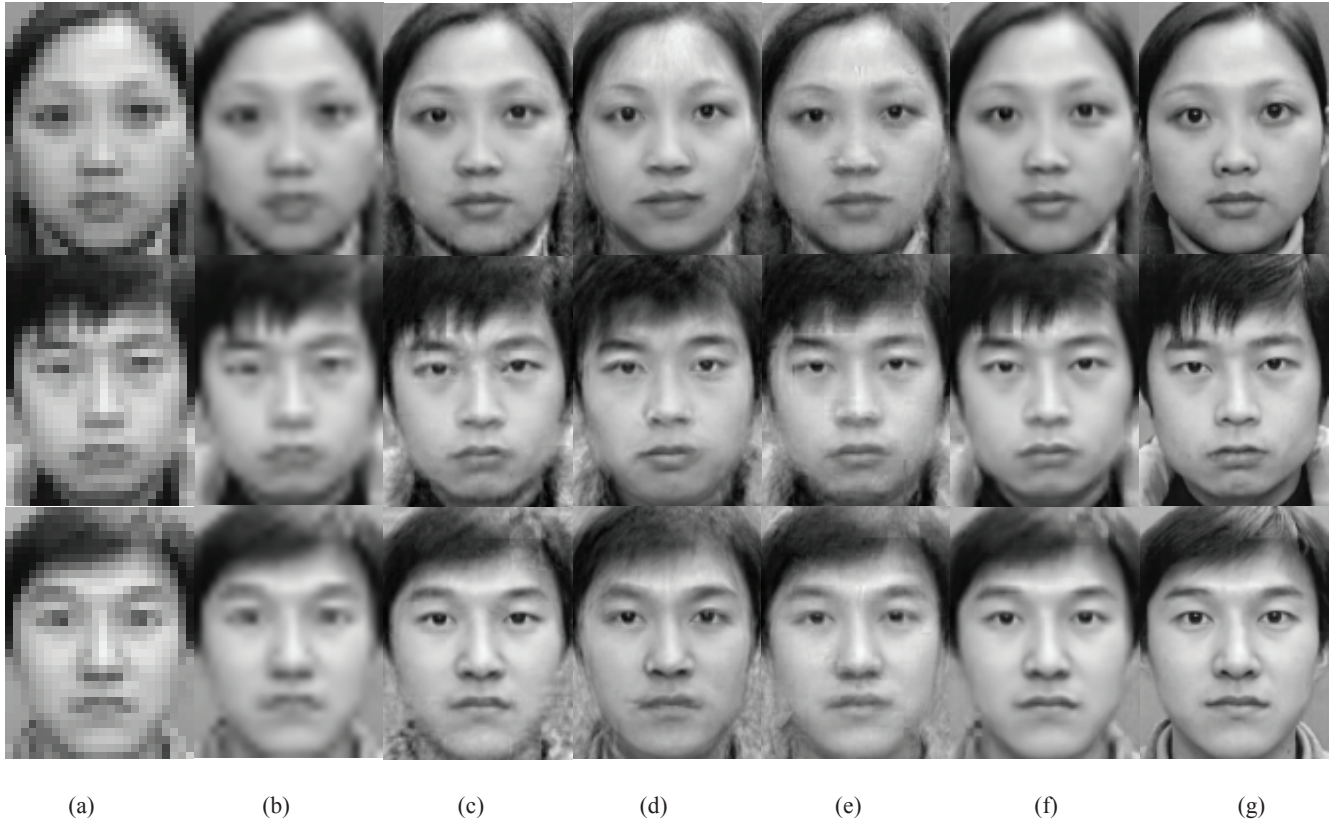
5. ACKNOWLEDGMENT

Fig. 1. The hallucination results. (a) The input 24×32 low-resolution faces. (b) Cubic B-Spline. (c) Neighbor Embedding. (d) Wang's method. (e) Zhuang's method. (f) Our method. (g) The original 96×128 high-resolution faces.

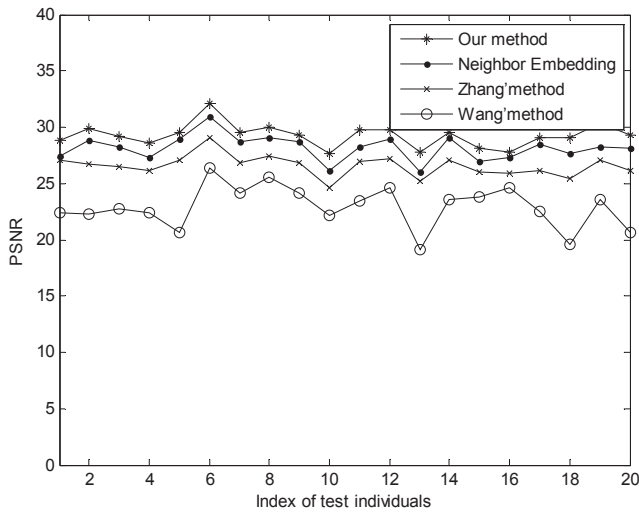Fig. 2. PSNR of different methods.

REFERENCES

[1] Xiaogang Wang and Xiaoou Tang, "Hallucinating Face by Eigentransformation," *IEEE Transactions on Systems Man and Cybernetics*, 35(3), pp. 425 – 434, 2005.

[2] S. Baker and T. Kanade, "Hallucinating Faces," In Proc. of *Inter. Conf. on Automatic Face and Gesture Recognition*, IEEE, France, pp.83-88, 2000.

[3] C.Liu, H. Shum and C. Zhang, "A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model," In Proc. of the *Inter. Conference on Image and Graphics*, IEEE, Hawaii, pp.192-198, 2001.

[4] Hong Chang, Dit-Yan Yeung, and Yimin Xiong, "Super-resolution through neighbor embedding," In Proceedings of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Washington, United States: Institute of Electrical and Electronics Engineers Computer Society, pp.1275-1282, 2004.

[5] Yueting Zhuang, Jian Zhang, and Fei wu, "Hallucinating faces: LPH super- resolution and neighbor reconstruction for residue compensation," *Pattern Recognition*, ELSEVIER, 40, pp. 3178-3194, 2007.

[6] Wen Gao, et al, "The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations," *IEEE Transactions on System Man, and Cybernetics* (Part A), 38, pp.149-161, 2008.

293