Face Hallucination Based on Neighbor Embedding via Illumination Adaptation

Sijie Song, Yanghao Li, Zhihan Gao, Jiaying Liu*
Institute of Computer Science and Technology, Peking University, Beijing, P.R.China, 100871

***Abstract*—In this paper, we present a novel face hallucination method by neighbor embedding considering illumination adaptation (NEIA) to super-resolve faces when the lighting conditions of the training faces mismatch those of the testing face. For illumination adjustment, face alignment is employed through dense correspondence. Next, every training face is composed into two layers to extract both details and highlight components. By operating the two layers of each face respectively, an extended training set is acquired by combining the original and adapted faces compensated in illumination. Finally, we reconstruct the input faces through neighbor embedding. To improve the estimation of neighbor embedding coefficients, nonlocal similarity is taken into consideration. Experimental results show that the proposed method outperforms other state-of-the-art methods both in subjective and objective qualities.**

I. Introduction

Face hallucination refers to a domain-specific super-resolution (SR) problem. It aims to reconstruct a high resolution (HR) human face automatically from one or a set of low resolution (LR) faces. The technique benefits a lot of fundamental applications such as face recognition, face detection and image compression. However, general super resolution algorithms might fail to recover high-frequency facial details due to the negligence of the structure features of human faces. There are still a lot of challenges remaining in face hallucination.

In order to recover the facial details more effectively and improve the visual qualities of hallucinated faces, a number of methods have been proposed in the past decades. The work of learning-based face hallucination can be traced back to the Baker and Kanada's work [1], in which a probabilistic framework was first proposed to model the relationship between LR and HR image patches. The target HR image is inferred with the help of a training set based on a Bayesian formulation. Following the pioneering work, other methods are proposed to hallucinate faces. Among them, a general super resolution method [2] based on neighbor embedding with a manifold assumption has shown impressive performance when applied to face hallucination. LR and HR patches are assumed to share similar local geometry and neighborhood relationship. The HR patches are generated by linearly combining similar patches in the training set with weights of the training patches estimated in the LR feature space.

Then, some two-step approaches were proposed to combine both global and local methods to improve the results of previous methods. These methods first reconstruct a smooth face by global methods and then compensate the residue with local patch-based methods to further enhance the local details. Motivated by a Markov Random Field (MRF) based image SR framework, Liu *et al.* [3] hallucinated LR face images by decomposing face appearance into a global parametric model and a local Markov network model with a large collection of HR face images.

Recently, Ma *et al.* [4] proposed a position-prior approach to save the computational complexity and produce good super-resolved results. The HR faces are generated by linearly combining image patches at the same position of each training image. Yang *et al.* [5] introduced a face hallucination algorithm by employing the most similar facial components in the training faces. The drawback of this method is that there are significant distortions in results when no highly similar components are found in the training set.

Nevertheless, few of the existing approaches address the issue that non-equal testing and training set might result in degraded results. For example, when the testing face undergoes significant illumination changing like highlight or shading comparing with the training faces, the facial details cannot be generated effectively by existing learning-based approaches. From the theoretical view, it is mainly because the inconsistency of high-frequency statistics between patches under different illumination conditions. So the high-frequency signals of the ill-posed image cannot be well recovered.

In our work, we develop a novel face hallucination method based on neighbor embedding to relieve the inconsistency between training and testing faces in terms of illumination. Considering the diverse distribution of highlight on human faces caused by the position of the light source, we first precisely align the training and testing faces through dense correspondence with the help of SIFT flow. Decomposed into two layers, each training face is modified according to the energy variation of the testing face to introduce both high-frequency details and amount of highlight to the original training faces. Then, the training set is extended by combining original and light-adapted training faces. We finally reconstruct the testing face via neighbor embedding, which takes the nonlocal similarity into account to obtain good estimates of neighbor embedding regression coefficients.

The rest of the paper is organized as follows. In Sec.II, the basic neighbor embedding is reviewed. Sec.III focuses on

*Corresponding author
This work was supported by National Natural Science Foundation of China under contract No. 61472011 and National High-tech Technology R&D Program (863 Program) of China under Grant 2014AA015205.

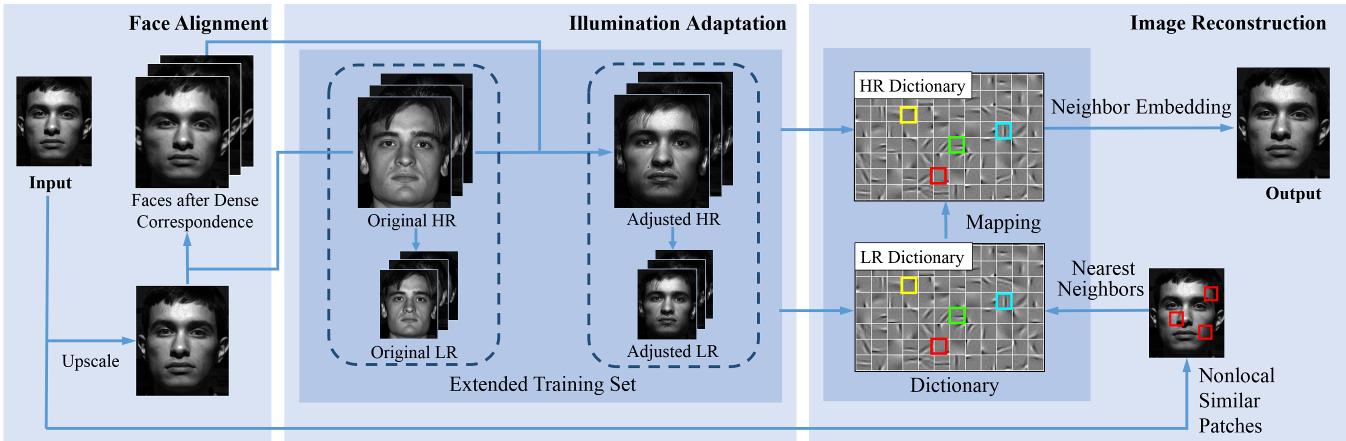


Fig. 1. The Framework of the Face Hallucination Based on Neighbor Embedding via Illumination Adaptation.

the illumination adaptation and face reconstruction process. Experimental results compared with other face hallucination methods are presented in Sec.IV. Finally, the conclusion is drawn in Sec.V.

II. Overview of Neighbor Embedding

The traditional neighbor embedding [2] usually reconstructs the images using coupled dictionaries. HR and LR images are assumed to form manifolds with similar local geometry. Given one input LR image, the goal is to reconstruct HR patches as a weighted average of neighbors using the same coefficients in the LR space. Then, the target HR image is generated by integrating the HR patches according to their positions. To formulate this problem, let $\mathcal{X} = \{x_s^i\}_{i=1}^{\mathcal{N}}$ and $\mathcal{Y} = \{y_s^i\}_{i=1}^{\mathcal{N}}$ be the LR and HR patch dictionaries, respectively. $\mathcal{N}$ is the dictionary size. After separating the input LR image X_t into small patches, for each LR patch x_t, its K nearest neighbors $N_l = [x_s^{i_1}, x_s^{i_2}, ..., x_s^{i_K}]$ in the training set $\mathcal{X}$ are obtained through retrieval algorithms such as K-nearest neighbor (K-NN). To minimize the reconstruction error, the optimal reconstruction coefficient α is estimated by:

$$\min_{\alpha} \|x_t - N_l \alpha\|_2^2 \quad s.t. \ \mathbf{1}^T \alpha = 1, \tag{1}$$

which is a constrained least squares problem. The coefficient α can be calculated by solving a linear system equation $G_l \alpha = 1$, subject to $\mathbf{1}^T \alpha = 1$, where $G_l = (x_t \mathbf{1}^T - N_l)^T (x_t \mathbf{1}^T - N_l)$. The corresponding HR patch y_t is given by applying the same reconstruction weights to corresponding neighbor HR patches $N_h = [y_s^{i_1}, y_s^{i_2}, ..., y_s^{i_K}]$ in the HR domain as follows:

$$y_t = N_h \alpha. \tag{2}$$

Finally the target image is reconstructed by integrating hallucinated HR patches. The overlap portions of patches are averaged among different patches.

However, face hallucination with this approach does not consider the characteristics of the testing face to process the training set. The inconsistency in illumination between training and testing faces affects the visual quality of reconstructed images. Thus, we propose a face hallucination method with illumination adaptation to compensate and extend the training set.

III. Neighbor Embedding via Illumination Adaptation

Given a testing image under significant illumination variations comparing with the training faces, the proposed method can be roughly separated into three stages as indicated in Fig.1: face alignment through dense correspondence, double-layer adaptation of illumination and image reconstruction. Before alignment, we simply interpolate the input LR image by the Bicubic method the size of the HR training image.

A. *Face Alignment through Dense Correspondence*

We assume that the training and testing faces have approximately similar facial poses and expressions. Followed by a dense correspondence between the training and testing faces, they are first aligned on a coarse level. Using a face detection and landmark localization method [6], each face is annotated by 68 landmarks and can be divided into segments according to the facial components. The eyes and mouth of the testing face are roughly aligned with those of the training face via an affine transform. Guided by the segments, we warp the testing face as the training face. The alignment is finally refined by using SIFT Flow [7] to put each pixel of the testing face in correspondence with a pixel of the training face.

B. *Double-layer Adaptation of Illumination*

In this subsection, the adaptation process of training images is explained in details. We extract the illumination components by image decomposition. Referring [8], the images are decomposed into double layers with a Gaussian kernel. The details of the faces can be captured from the first layer while the second layer describes the highlight conditions. To adapt the illumination conditions of the testing face, we adopt corresponding operations on each layer of the training faces, respectively. This process performing on grayscale images is achieved in three steps.

In the first step, an HR image Y_t for training is decomposed into double layers with:

$$S_\ell[Y_t] = \begin{cases} Y_t - Y_t \otimes G(\sigma), & \ell = 1 \\ Y_t \otimes G(\sigma), & \ell = 2 \end{cases}, \tag{3}$$

where we denote the level at scale ℓ as S_ℓ, and $G(\sigma)$ is a 2D normalized Gaussian kernel with standard deviation $\sigma = 8$ in practice, the convolution operator is denoted as $\otimes$. Absolutely the same process can be applied to an LR testing image X_t.

Second, the two levels of the training face are processed separately. For the first level, we estimate its energy by the local average of the square of layer coefficients to show the variations in signals. The energy E_1 can be given by:

$$E_1[Y_t] = S_1^2[Y_t] \otimes G(\sigma). \tag{4}$$

To adjust the energy distribution of the training face as similar as the testing face, we perform the adjustment operation on the first layer as:

$$S_1[O] = S_1[Y_t] \times \sqrt{\frac{M(E_1[X_t])}{E_1[Y_t] + \varepsilon}}, \tag{5}$$

where O represents the output, $M(\cdot)$ denotes the morphing operator defined by the correspondence field in the process of alignment, and $E_1[X_t]$ can be calculated similarly as (4). In addition, we use $\varepsilon = 0.0001$ in case division by zero. To compensate the amount of the illumination, we simply substitute the second warped layer of the testing face for that of the training face as follows:

$$S_2[O] = M(S_2[X_t]). \tag{6}$$

The final processed image O is generated in the last step, which is given as:

$$O = S_1[O] + S_2[O]. \tag{7}$$

As shown in the illumination adaptation stage of Fig.1, the training faces successfully inherit the illumination conditions of the testing face, such as the highlight on the forehead. By altering the energy distribution on the first level, details caused by the light source of the testing face are well preserved onto the training faces. In the meantime, the highlight area is well transformed onto the training faces through aggregating the second warped level of the testing face directly.

C. Image Reconstruction via Nonlocally Centralized Neighbor Embedding

After modifying every training face according to the illumination conditions of the testing face, we acquire an enriched training set by combining the original and transformed training images which contain more abundant information. With the extended training set, we solve face hallucination using neighbor embedding approaches with nonlocal redundancy.

According to the traditional neighbor embedding method, it is assumed that small LR and HR patches share the same local geometrical structure. HR patches can be reconstructed as a weighted average of neighbors using the same coefficients in the LR space, which is characterized by how an LR patch can be represented by its neighbors.

Intuitively, similar patches should have similar linear representations. In our work, we employ regularization terms incorporated with nonlocal redundancy. For each LR patch x_t from the input image X_t, we attempt to acquire a good estimate β of its representation weights α with the help of nonlocal similarity. We first search its nonlocal similar patches (including x_t) from X_t through K-NN to form a set of its similar patches denoted by Ω. N_l^i is used as the nearest neighbors from the LR dictionary of patch x_t^i within Ω. Then, the K most frequent patches of $\{N_l^i\}(i \in \Omega)$ compose N_l. We reconstruct each patch x_t^i with N_l and acquire the coefficients $\{\alpha_i\}(i \in \Omega)$, which can be formulated as a ridge regression problem:

$$\min_{\alpha_i} \|x_t^i - N_l \alpha_i\|_2^2 + \lambda \|\alpha_i\|_2^2, \tag{8}$$

where parameter λ is the regularization term coefficient and α_i can be solved by $\alpha_i = ({N_l}^T N_l + \lambda I)^{-1} {N_l}^T x_t^i$. Then, β is calculated by the weighted average of coefficients associated with the patches in Ω:

$$\beta = \sum_{i \in \Omega} w_i \alpha_i, \; w_i = \frac{1}{W} exp\left(\frac{-\|x_t - x_t^i\|_2^2}{h}\right), \tag{9}$$

where w_i is the normalized weight, depending on the distance between x_t and x_t^i, and we use W as the normalized factor, h as a pre-determined scalar. Linearly represented by N_l, the coefficient α of patch x_t can be formulated as:

$$\min_{\alpha} \|x_t - N_l \alpha\|_2^2 + \lambda_1 \|\alpha\|_2^2 + \lambda_2 \|\alpha - \beta\|_2^2, \tag{10}$$

where λ_1, λ_2 are parameters to balance the contribution of regulation terms. Thus, the coefficient α can be solved by:

$$\alpha = (N_l^T N_l + (\lambda_1 + \lambda_2) I)^{-1} (N_l^T x_t + \lambda_2 \beta). \tag{11}$$

Finally, we employ (2) to calculate the corresponding HR patch y_t and the target face is hallucinated using these computed HR patches guided by their positions.

IV. Experimental Results

We use the the Extended Yale Face Database B [9] to evaluate the proposed method because it consists of subjects of illumination variations. One set consisting of subjects under the same lighting condition 000E+00 is utilized as the training set. And we use another four testing sets under different illumination conditions (+000E+45, +005E+10, +015E+20, +035E+15) as the testing faces. The difference between the five illumination conditions can be viewed in Fig.2.

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

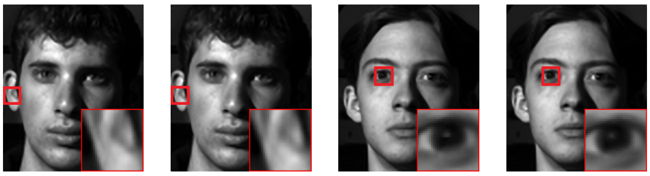
Fig. 2. Comparison of different illumination conditions. (a) +000E+00, (b) +000E+45, (c) +005E+10, (d) +015E+20, (e) +035E+15.

All the faces taken from the the Extended Yale Face Database B are at the same pose and expression. In our experiment, we first blur the original HR images with a Gaussian kernel whose width is 1.6, and then downsample the blurred images with a scaling factor 4 to obtain the input LR images. For a specific testing set, we randomly choose a face from it to adjust the illumination of training faces and generate its corresponding training set.

The effectiveness of illumination adaption (IA) can be proved from the zoomed comparison of the highlighted part as shown in Fig.3. It is obvious that the edge of the ear is clearer and the highlight in the eyeball is better recovered. With

TABLE I
AVERAGED PSNR RESULTS BY 4× ON 4 TESTING SETS

Testing Set	Bicubic	PFH [4]	SFH [5]	NEFC [10]	NENIA	NEIA	Gain vs. NEFC	Gain vs. NENIA
000E+45	32.74	30.31	26.92	33.71	33.92	**34.01**	0.30	0.09
005E+10	32.95	31.05	27.80	34.40	34.35	**34.45**	0.05	0.10
015E+20	32.53	30.35	27.07	33.88	33.87	**33.98**	0.10	0.11
035E+15	32.23	30.06	27.08	33.56	33.55	**33.69**	0.13	0.14
Average	32.62	30.44	27.22	33.89	33.92	**34.03**	0.14	0.11



(a) With IA (b) Without IA (c) With IA (d) Without IA

Fig. 3. Effectiveness of the illumination adaption (IA). (a)(c) Results of our method with IA. (b)(d) Results of our method without IA.

introducing the illumination information to the training set, facial details can be generated more effectively and obviously.

We compare the proposed algorithm (NEIA) with different methods, including the Bicubic method, the position method (PFH) [4], the structured face hallucination method (SFH) [5], the face hallucination method for facial components (NEFC) [10]. Besides, to confirm the influence of illumination adjustment, NEIA is also compared with our neighbor embedding method considering nonlocal similarity, but no illumination adaptation (NENIA). We implement algorithm in PFH [4], while others are compared with the original authors' codes.

The objective results of four testing sets under different lighting conditions are measured by Peak Signal to Noise Ratio (PSNR) in Table I. Our proposed method outperforms other methods and achieves the highest PSNR results. Some subjective results are presented in Fig.4. Due to the limitation of position-prior, the results of PFH have some artifacts. SFH yields some unnatural and distorted results because of the difficulty in finding highly similar components with lighting variations. In addition, NEFC does not work well since facial components suffer illumination changes between testing and training faces, so it is not appropriate to reconstruct them, respectively. The objective results prove that illumination adaptation and neighbor embedding with nonlocal redundancy contribute to generating better super-resolved details and more natural faces.

V. CONCLUSIONS

In this paper, we develop a face hallucination method based on neighbor embedding via illumination adaptation. To relieve the inconsistency of illumination between testing and training faces, we process training faces by compensating their illumination to enrich the training set. The input face is hallucinated by neighbor embedding with nonlocal redundancy. Experimental results indicate the proposed method outperforms other methods in both objective and subjective qualities.

REFERENCES

[1] S. Baker and T. Kanade. Hallucinating faces. In *Proc. IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, 2000.

[2] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. I–I, June 2004.

[3] C. Liu, H.-Y. Shum, and W. T. Freeman. Face hallucination: Theory and practice. *Int'l Journal of Computer Vision (IJCV)*, vol. 75, no. 1, pp. 115–134, October 2007.

[4] X. Ma, J. Zhang, and C. Qi. Hallucinating face by position-patch. *Pattern Recognition (PR)*, vol. 43, no. 6, pp. 2224–2236, June 2010.

[5] C.-Y. Yang, S. Liu, and M.-H. Yang. Structured face hallucination. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 1099–1106, June 2013.

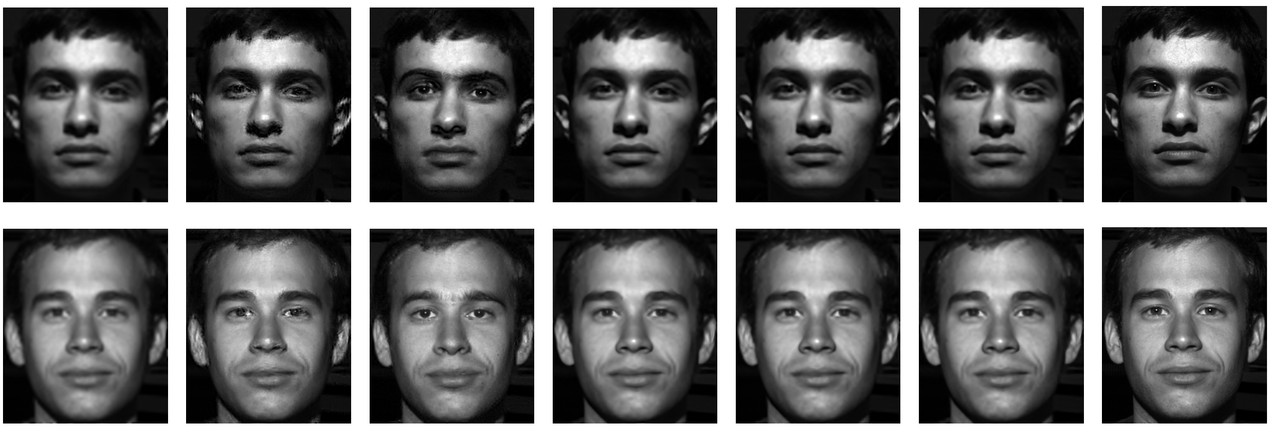
[6] D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 2879–2886, June 2012.

[7] C. Liu, J. Yuen, and A. Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 33, no. 5, pp. 978–994, May 2011.

[8] Y. Shih, S. Paris, C. Barnes, W. T. Freeman, and F. Durand. Style transfer for headshot portraits. *ACM Trans. Graphics (TOG)*, vol. 33, no. 4, pp. 148:1–148:14, July 2014.

[9] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 23, no. 6, pp. 643–660, June 2001.

[10] Y. Li, J. Liu, W. Yang, and Z. Guo. Multi-pose face hallucination via neighbor embedding for facial components. In *Proc. IEEE Int'l Conf. Image Processing (ICIP)*, 2015.



(a) Bicubic (b) PFH [4] (c) SFH [5] (d) NEFC [10] (e) NENIA (f) NEIA (g) Ground Truth

Fig. 4. The visual results by 4× on the two face images under the illumination conditions +000E+45, +005E+10, respectively.