COOPERATIVE FACE HALLUCINATION USING MULTIPLE REFERENCES

Chih-Chung Hsu[1], Chia-Wen Lin[2], Chiou-Ting Hsu[3], and Hong-Yuan Mark Liao[4]

[1, 2]Department of Electrical Engineering, National Tsing-Hua University
[3]Department of Computer Science, National Tsing-Hua University
101, Section 2, Kuang-Fu Road, Hsinchu 30013 Taiwan
{d9661805@oz, cwlin@ee, cthsu@cs}.nthu.edu.tw
[4]Institute of Information Science, Academia Sinica
128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan
liao@iis.sinica.edu.tw

ABSTRACT

This paper proposes a cooperative example-based face hallucination method using multiple references. The proposed method first uses clustering and residual prototype faces construction to improve the performance of hallucinating a single low-resolution (LR) face to obtain a high-resolution (HR) counterpart. In the case that multiple LR face images for a person are available, a unique feature of the proposed method is to cooperatively enhance the qualities of hallucinated HR images by taking into account the multiple input face images jointly as prior models. Experimental results demonstrate that the proposed cooperative method achieve significant subjective and objective improvement over single-prior schemes.

Index Terms— Face hallucination, multi-prior references, super-resolution, prototype faces.

1. INTRODUCTION

In recent years, face hallucination [1]-[4] has become an attractive technique in super-resolution of face photos because it has many applications such as security in surveillance video, face recognition, facial expression estimation, face age estimation, and image/video editing. In these applications, the resolution requirement for input face images is typically high, for which simple interpolation schemes. Example-based super-resolution (SR) schemes [1][5] have proven to be able to obtain significantly finer details of images with a limited scaling factor compared to interpolation-based schemes, provided that a comprehensive set of training images are used to collect prior knowledge of the structures and patterns of images using machine learning techniques.

The problem of super-resolution for face images is, however, different from that for generic images because face images have a unified structure which people are very familiar with. Even only few reconstruction errors occurring in a face image can cause visually annoying artifacts. For example, geometry distortion in the mouth and eyes on a reconstructed face image may only reduce the image's objective quality slightly, while the subjective quality of the reconstructed face can be degraded significantly. Therefore, both the global face shape and textures and local geometric structures (e.g., mouth, nose, and eyes) need to be treated carefully in face hallucination [2].

There are two types of the face hallucination methods: model-based methods [2] and example-based methods [4][6][7]. The model-based method proposed in [2] uses a two-step approach that captures the global and local geometrical features of a face using a parametric Gaussian model and a non-parametric local model based on Markov random field (MRF), respectively. The method then uses a statistical approach to approximate the output HR face image under different situations. The example-based methods proposed in [3][4] decomposes an input LR face into many prototype faces (i.e., eigenfaces) as a prior model using principle components analysis (PCA). The method in [4] then applies a recursive error back-propagation method to reconstruct the corresponding HR face. The method proposed in [6] is similar to this technique but uses a different basis decomposition method.

Most existing face hallucination methods only focused on hallucinating the resolution of a single face image. They did not make use of multiple prior models when multiple LR input images are available (e.g., images taken from similar or different situations or devices). Only few works took into account multiple prior models to enhance the performance of face hallucination. For example, the example-based method proposed in [7] reconstructs a HR face image using multiple temporally neighboring LR faces from a video sequence. Nevertheless, if the multiple LR input face images are not temporally consecutive, the quality of reconstructed face image may drop significantly. The reason is the pictures are usually obtained under different situations, making the variation on the faces too large to handle using the MRF model adopted in this method.

This work aims to enhance the performance of face hallucination by cooperatively hallucinating multiple input face images by taking into account the multiple prior models jointly. Our framework provides the flexibility of allowing the multiple priors to be taken under different situations (e.g., different poses, ages, and environments). In addition, we also provide an efficient approach to enhance the visual quality while hallucinating a single face using clustering-based training and residual face refinement.

The rest of this paper is organized as follows. In Sec. 2, we briefly review the example-based method proposed in [4]. Sec. 3 presents the proposed cooperative face hallucination method. In

Sec. 4, the experimental results of the proposed scheme are demonstrated. Finally, Sec. 5 concludes this paper.

2. EXAMPLE-BASED FACE HALLUCINATION

Our method is built on top of the example-based framework proposed in [4]. This method exploits a whole LR face image, instead of dividing the face into small patches like in [5], to reconstruct the HR face. Let $\mathbf{I}_L$ and $\mathbf{I}_H$ respectively denote the input LR image and its HR version, a full-resolution image can be represented as

$$\mathbf{I} = (i_1, i_2, ..., i_L, i_{L+1}, i_{L+2}, ..., i_H) \quad (1)$$

where i_j represents the value of the j-th pixel, L and H denote the pixel numbers of LR and HR images, respectively. The reconstructed face image can be obtained by

$$\mathbf{I} \cong \mathbf{P} \cdot \boldsymbol{\alpha} = \mathbf{R} \quad (2)$$

where $\mathbf{P}$ denotes the prototype faces, $\boldsymbol{\alpha}$ stands for the reconstruction coefficients, and $\mathbf{R}$ denotes the reconstructed face.

The coefficients of the input LR face image corresponding to a set of LR prototype faces can be obtained by using the following least-squares projection:

$$\boldsymbol{\alpha}^* = ((\mathbf{P}_L)^{\mathrm{T}} \cdot \mathbf{P}_L)^{-1} \cdot (\mathbf{P}_L)^{\mathrm{T}} \cdot \mathbf{I}_L \quad (3)$$

where $\mathbf{P}_L$ denotes the LR prototype faces and the $\mathbf{I}_L$ denotes the input LR face image.

After computing the coefficients, a set of HR prototype faces are then used to reconstruct the HR face image. The HR face image can be approximated by the linear combination of HR prototype faces weighted by the coefficients obtained from the LR face image decomposition as illustrated in Fig. 1.

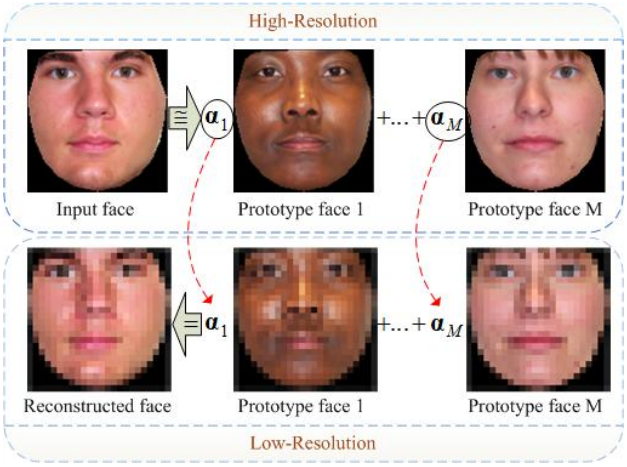

Fig. 1 Illustration of face hallucination.

The coefficient set obtained from decomposing the LR face image, however, is typically not optimal for synthesizing the HR counterpart. Therefore, in the reconstruction phase, an iterative method was proposed in [4] to obtain the best coefficients for face hallucination. Assuming that, at the t-th iteration, the HR face image $\mathbf{R}_H^{t-1}$ is reconstructed using the above-mentioned method and the corresponding LR version $\mathbf{R}_L^{t-1}$ is obtained by downscaling $\mathbf{R}_H^{t-1}$. The residual face image $\mathbf{D}_L^t$ is subsequently obtained by subtracting the current LR image $\mathbf{R}_L^t$ from the previous LR image $\mathbf{R}_L^{t-1}$. If the error value in residual face image $\mathbf{D}_L^t$ is still high, then the corresponding HR residual face image $\mathbf{D}_H^t$ can be obtained by repeating the same process. As a result, the reconstructed HR face image can be updated as $\mathbf{R}_H^{t+1} = \mathbf{R}_H^t + \mathbf{D}_H^t$ at the (t+1)th iteration. When the reconstruction error is smaller than a threshold, or the number of the iterations reaches a limit, the iteration process will be terminated, finally obtaining the optimal reconstructed HR image $\mathbf{R}_H^*$.

3. COOPERATIVE FACE HALLUCINATION

Fig. 2 shows the flowchart of the proposed method. At the training stage, the prototype faces are obtained as prior models by applying PCA on the training set. In example-based face hallucination, the quality of a reconstructed image hinges on the selection of training set heavily. A small training set cannot cover sufficiently representative face images, whereas a large one has high computational complexity. On the other hand, the variety (e.g., species, ages, and genders) in the training face images also has significant impact on the quality of face hallucination. It is thus important to cluster the training face images prior to training, considering both complexity and quality.

Before face hallucination, pre-processing such as alignment using AAM, scaling, and normalization is performed on the multiple LR images to ensure that the prior models can be cooperated well when the inputs are obtained under different situations (e.g. different poses, environments and ages).

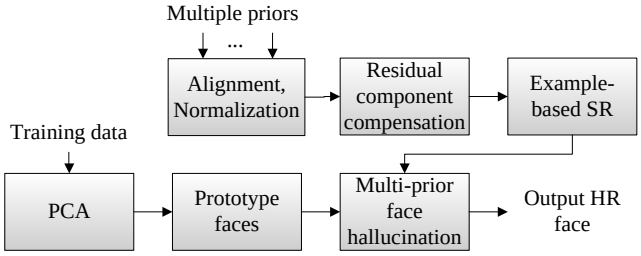

Fig. 2 The flowchart of the proposed method.

The residual image between the face image in the training set and the corresponding reconstructed face image can be calculated by a subtraction operation. In [4], the prototype faces trained from the raw training images are used to reconstruct the HR residual image $\mathbf{D}_H^t$ from LR $\mathbf{D}_L^t$. However, there is a better way to improve the performance by training the prototype residual faces using a two-phase learning.

A key issue of cooperative face hallucination is how to efficiently fuse the multiple prior models for reconstructing HR images cooperatively. In our method, the residual components between available priors and the LR face to be hallucinated are computed via the residual compensation process. Then, the example-based SR scheme proposed in [5] is used to upscale the input priors. Finally, the hallucinated face can be obtained as a weighted sum of the HR reconstructions of selected priors and the input LR face. The detailed method will be elaborated below.

3.1. Clustering-Based Training and Residual Refinement

Let $T_H = (\mathbf{I}_H^1, \mathbf{I}_H^2, ..., \mathbf{I}_H^N)$ denote the training set of HR images and $T_L = (\mathbf{I}_L^1, \mathbf{I}_L^2, ..., \mathbf{I}_L^N)$ the downscaled version of T_H. After

clustering the training sets, both HR and LR training images are divided into K subsets, denoted $T_H^k = (\mathbf{I}_H^1, \mathbf{I}_H^2, ..., \mathbf{I}_H^{N_k})$ and $T_L^k = (\mathbf{I}_L^1, \mathbf{I}_L^2, ..., \mathbf{I}_L^{N_k})$, respectively, where $\sum_{k=1}^{K} N_k = N$.

Given an input LR image $\mathbf{I}_{\text{in}}$, the closest training subset can be chosen by finding the index of the closest cluster as follows:

$$k^* = \arg\min_k \left\| \mathbf{I}_{\text{in}} - \bar{\mathbf{I}}_L^k \right\|, \quad 1 \le k \le K \tag{4}$$

where $\bar{\mathbf{I}}_L^k$ represents the mean face of T_L^k.

As a result, the reconstructed HR image can be obtained by

$$\mathbf{R}_H^* = \mathbf{P}_H^{k^*} \cdot \boldsymbol{\alpha}. \tag{5}$$

where $\mathbf{P}_H^{k^*}$ denotes the ordered set of L eigenvectors with the largest corresponding eigenvalues in $T_H^{k^*}$.

When reconstructing the HR residual face $\mathbf{D}_H^t$ from $\mathbf{D}_L^t$, the method proposed in [4] uses the prototype faces trained from the raw training images, rather than from the residual images themselves. This does not make sense as the residual images and HR raw images are very different in nature. Therefore, we propose to train a set of HR and LR prototype residual faces separately for reconstructing $\mathbf{D}_H^t$ from $\mathbf{D}_L^t$.

Similarly to the process described in (4)~(5), the HR and LR prototype residual faces $\mathbf{P}_{D,H}^k$ and $\mathbf{P}_{D,L}^k$ are trained from the residual training set which is obtained by subtracting the training face images from their reconstructed ones.

3.2 Multi-Prior-Assisted Reconstruction

Suppose that the M LR priors are aligned via AAM (Active appearance model) [8]. The coefficient of the m-th prior image is

$$\boldsymbol{\alpha}_m^P = ((\mathbf{P}_L)^{\mathrm{T}} \cdot \mathbf{P}_L)^{-1} \cdot (\mathbf{P}_L)^{\mathrm{T}} \cdot \mathbf{I}_{L_m}^P \tag{6}$$

where $\mathbf{I}_{L_m}^P$ denotes the m-th LR prior.

Let $\boldsymbol{\alpha}_L^{\text{ref}}$ denote the coefficient vector of the input LR image, the difference between it and the m-th prior's coefficients $\boldsymbol{\alpha}_m^P$ is

$$\Delta\boldsymbol{\alpha}_m = \boldsymbol{\alpha}_m - \boldsymbol{\alpha}_L^{\text{ref}}. \tag{7}$$

In (7), the difference coefficients represent the lost detail in the current reconstructed image. Therefore, the supplementary detail from the m-th prior can be computed by

$$\Delta\mathbf{I}_{H_m} \cong \mathbf{P}_H^{k^*} \cdot \Delta\boldsymbol{\alpha}_m. \tag{8}$$

The hallucinated HR image can be further refined by including the supplements from the M priors by adding the weighted sum of the supplements into the HR face:

$$\mathbf{I}_H^* = \mathbf{I}_H + \sum_{m=1}^{M} \left(w_m \cdot \Delta\mathbf{I}_{H_m} \right) \tag{9}$$

where w_m denotes the weight of the m-th supplement, reflecting the confidence on the m-th prior. The optimal coefficients can be obtained by minimizing the following cost function

$$w_1^*, \ldots, w_m^* = \arg\min_{w_1,\ldots,w_m} \left\| \downarrow (\mathbf{P}_H \sum_{m=1}^{M} w_m \boldsymbol{\alpha}_m) - \mathbf{I}_L \right\|, \tag{10}$$

where $\downarrow$ denotes a downscaling operation. Note, if the priors are similar (e.g., taken from a video or captured with the same resolution and in similar conditions), the simplest setting is $c_m = 1/M$, leading to a low-pass refinement to smooth out high-frequency artifacts in the reconstructed face.

One problem with example-based face hallucination is that when the resolution of the LR input (or prior) is too low, the decomposed coefficients will become inaccurate. One way to solve this problem is to modify the iteration procedure described is Sec. 2 to take into account the multi-prior information. For each prior, the super-resolution method proposed in [5] is used to obtain more reliable prior information. The optimal reconstructed image $\mathbf{R}_L^*$ is replaced with $\mathbf{R}_H^*$ to ensure that the linear combination of the prototype faces will not select the incorrect basis as the coefficients are now computed in the HR domain.

For each LR prior, its LR coefficient can be used to reconstruct its corresponding HR image. After that, the residual image $\mathbf{D}_{L_m}^P$ of the m-th prior is computed by

$$\mathbf{D}_{L_m}^P = \downarrow (\mathrm{SR}(\mathbf{I}_{L_m}^P) - \mathbf{R}_H^*). \tag{11}$$

where SR(·) denotes the SR operation proposed in [5].

Assuming that the m-th prior has the minimum error in $\mathbf{D}_{L_m}^P$, the example-based SR can be used to increase the resolution of the k-th prior. The criterion function can be rewritten as

$$\mathbf{D}_L^P = \beta\mathbf{D}_{L_m}^P + (1-\beta)\mathbf{D}_L \tag{12}$$

where w denotes a weight value. A large w will lead to the reconstructed image become more close to the prior image, vice versa. Note that the input priors are refined by difference reconstructed face $\Delta\mathbf{I}_{H_m}$. Therefore, the reconstructed face is not different from the input face image significantly.

4. EXPERIMENTAL RESULTS

Our training set contains 450 images with resolution 64x64 which are selected from the PAL (Productive Aging Lab) face database and are aligned using AAM. The resolution of all of the input LR images is 16x16. We used K-means clustering to divide the training images into three clusters at the training stage.

As shown in Fig. 3(c), when the training set contains many images significantly different from the input face image, the quality of the reconstructed face image using the method proposed in [4] will become poor. With the proposed pre-clustering on the training set, Fig. 3(d) shows the PSNR quality of reconstructed image is improved by about 1.4 dB. Fig. 3(e) shows the reconstructed face image obtained using residual refinement that eliminate many artifacts of the reconstructed face image in Fig. 3 (c). Fig. 3(f) shows that the proposed the cooperative multi-prior reconstruction scheme further significantly enhance both the subjective and objective qualities of reconstructed face (e.g., on the eyes).

Fig. 4 depicts a subjective comparison of different face hallucination methods for another test image. Fig. 4(c) shows the reconstructed face images using the method in [4]. We can observe from Fig. 4(d) that the proposed training set clustering procedure improves the quality of reconstructed face especially on the eyes. Fig. 4(e) shows the reconstructed face image using the multi-prior scheme proposed in [7], which is not satisfactory because this method cannot well fuse the multiple priors which are not taken from neighboring frames of a video thus are with significant variation. Fig. 4 (e) shows the quality of the reconstructed face using our proposed method is significantly better than others'.

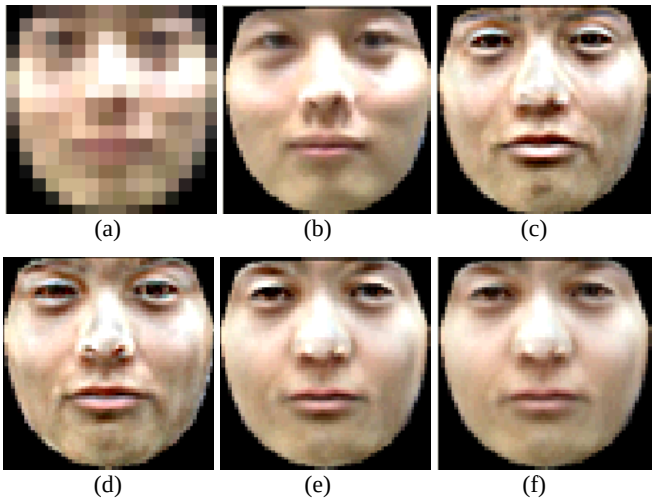

Fig. 3 Subjective quality comparison: (a) Input LR face image, (b) HR ground-truth, and the faces reconstructed using (c) [4]'s method (PSNR=19.93 dB), (d) [4]'s method with training set clustering (PSNR=21.36dB) (e) [4]'s method with training set clustering and residual face refinement (PSNR=21.71dB), and (f) our proposed method (PSNR=23.05dB).

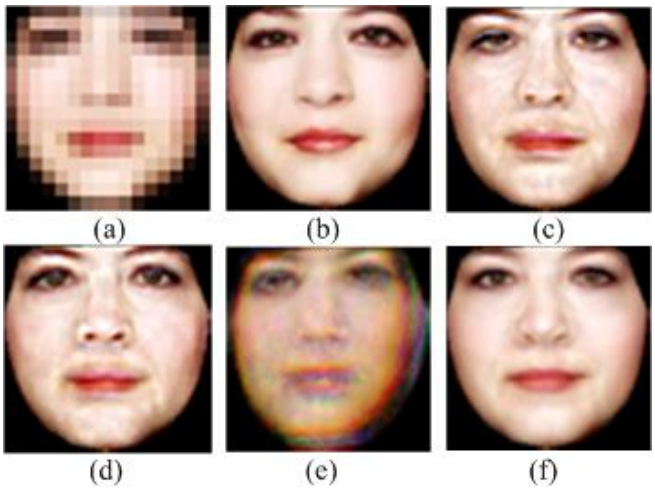

Fig. 4 Subjective quality comparison: (a) input LR face image, (b) the HR ground-truth, the images reconstructed by (c) [4]'s method, (d) [4]'s method with training set clustering, (e) [7]'s method, and (f) our proposed method.

Fig. 5 compares the qualities of the reconstructed faces for five input faces using different methods. It is clear that the proposed method outperforms other face hallucination methods. But some individual parts (e.g., mouth, nose, and eyes) sometimes are over-smoothed due to multi-prior fusion. This can be avoided by using the local models proposed in [2].

5. CONCLUSION

In this paper, we proposed a novel cooperative face hallucination method using multiple priors. In the training phase, we proposed to pre-cluster the training set into sub-groups and to train prototype residual faces separately rather than directly using the prototype faces obtained from the raw training data. We have also proposed a multi-prior fusion method to cooperatively enhance the qualities of hallucinated HR images. Experimental results demonstrate that the proposed cooperative method achieve significant subjective and objective improvement over existing schemes.

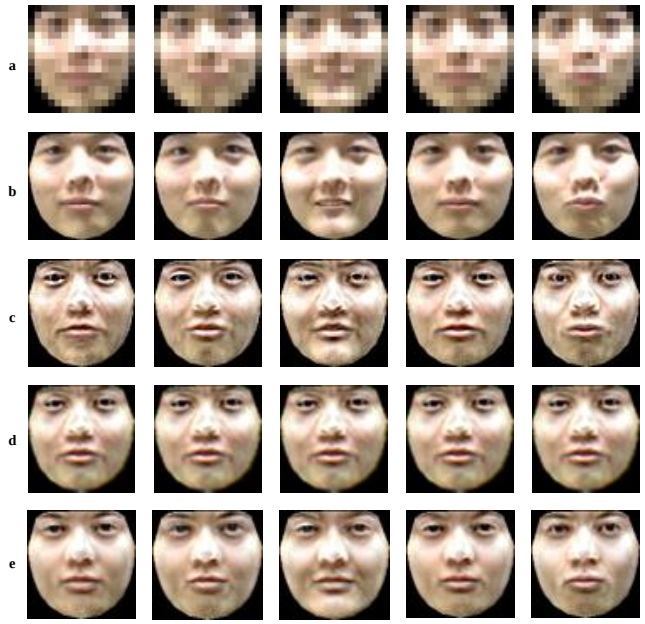

Fig. 5 Subjective quality comparison for five input face images: (a) Input LR images, (b) the ground-truths, and the hallucinated face images obtained by (c) [4]'s method, (d) [7]'s method, and (e) the proposed method, respectively.

ACKNOWLEDGEMENT

This work was supported in part by the Ministry of Economic Affairs, Taiwan, under grant 97-EC-17-A-02-S1-032 and by the National Science Council, Taiwan, under grant NSC98-2631-001-011.

REFERENCES

[1] S. Baker and T. Kanade, "Limits on superresolution and how to break them," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 9, pp. 1167-1183, Sept. 2002.

[2] C. Liu, H. Y. Shum, and W. T. Freeman, "Face hallucination: Theory and practice," *Int. J. Comput. Vision*, vol. 75, no. 1, pp. 115-134, Oct. 2007.

[3] B. K. Gunturk *et al.*, "Eigenface-domain super-resolution for face recognition," *IEEE Trans. Image Process.*, vol.12, no.5, pp. 597-606, May 2003.

[4] 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.

[5] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics & Applications*, vol. 22, no. 2, pp. 56-65, Mar. 2002.

[6] W. Zhang and W. K. Cham, "Learning-based face hallucination in DCT domain," in *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit.*, June 24-26, 2008.

[7] G. Dedeoglu, *Exploiting Space-Time Statistics of Videos for Face Hallucination*, Tech. Report CMU-RI-TR-07-05, Robotics Institute, Carnegie Mellon University, Apr. 2007.

[8] T. F. Cootes and C. J. Taylor, *Statistical Models of Appearance for Computer Vision*, Tech. Report, Univ. Manchester, U.K., Mar. 2004.