Pattern Recognition 47 (2014) 3327–3342

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

# Face hallucination with imprecise-alignment using iterative sparse representation

Yan Liang [a,e], Jian-Huang Lai [b,*], Pong C. Yuen [c], Wilman W. Zou [c], Zemin Cai [d]

[a] School of Mathematics and Computational Science, Sun Yat-sen University, Guangzhou, China
[b] School of Information Science and Technology, Sun Yat-sen University, Guangzhou, China
[c] Department of Computer Science, Hong Kong Baptist University, 224 Waterloo Road, Kowloon, Hong Kong, China
[d] Department of Electronic Engineering, College of Engineering, Shantou University, Shantou, China
[e] Media Technology Laboratory, Huawei Technologies Co., Ltd., Shenzhen, China

ARTICLE INFO

*Article history:*
Received 13 July 2013
Received in revised form
29 January 2014
Accepted 24 March 2014
Available online 4 April 2014

*Keywords:*
Face hallucination
Mis-alignment
EM algorithm
Sparse representation

ABSTRACT

Existing face hallucination methods assume that the face images are well-aligned. However, in practice, given a low-resolution face image, it is very difficult to perform precise alignment. As a result, the quality of the super-resolved image is degraded dramatically. In this paper, we propose a near frontal-view face hallucination method which is robust to face image mis-alignment. Based on the discriminative nature of sparse representation, we propose a global face sparse representation model that can reconstruct images with mis-alignment variations. We further propose an iterative method combining the global sparse representation and the local linear regression using the Expectation Maximization (EM) algorithm, in which the face hallucination is converted into a parameter estimation problem with incomplete data. Since the proposed algorithm is independent of the face similarity resulting from precise alignment, the proposed algorithm is robust to mis-alignment. In addition, the proposed iterative manner not only combines the merits of the global and local face hallucination, but also provides a convenient way to integrate different strategies to handle the mis-alignment problem. Experimental results show that the proposed method achieves better performance than existing methods, especially for mis-aligned face images.


## 1. Introducton

Super-resolution is a process to estimate a high-resolution (HR) image from its low-resolution (LR) counterpart. Super-resolution of face images, also called face hallucination, has been an independent branch in super-resolution research field since the pioneering work by Baker and Kanade [1]. It has been extensively and actively studied in recent years because of numerous important face-based applications such as recognition of face images from surveillance video by human and machine. Psychological researchers have revealed that the fine-scale information of faces is indispensable for face recognition by human [2], which strongly supports the importance of face hallucination. Different from domain-general image super-resolution, face hallucination methods develop domain-specific [3,4] prior with strong cohesion to the face domain concept. Many face hallucination methods were developed in the last decade and can be categorized in two approaches: global and local. The basic idea of local methods is based on local patch similarity in images located at the same position. Baker and Kanade [1] first proposed a local face hallucination method, which constructs high frequency components of a HR face image by finding the nearest "parent structure" at each pixel through training. Ma et al. [5] hallucinated the HR face image patch by linearly combining the image patches at the same position of each training image. Another kind of local method [6,7] were also developed recently. This method proposed a face similarity preserving process which selects a few face images in the database which are the most similar to the LR input image as training samples. In particular, Hu et al. [6] proposed to warp the selected training HR face images using optical flow to further ensure the similarity between the training samples and the test sample. Then, for each facial patch, a local structure prior is used to constrain the super-resolution results.

In contrast, the global methods consider each face image as a data point in the face image space and constrain the global face representation under compact space priors. Principal component analysis (PCA) has been proposed as global model constraint [8–10]. Zhuang et al. [11] proposed the locality preserving hallucination method by locality preserving projections before the radial basis function

* Corresponding author. Tel.: + 86 20 84110175.
*E-mail addresses:* liangyan_709@163.com (Y. Liang), stsljh@mail.sysu.edu.cn (J.-H. Lai), pcyuen@comp.hkbu.edu.hk (P.C. Yuen), wwzou@comp.hkbu.edu.hk (W.W. Zou), zmcai@stu.edu.cn (Z. Cai).

http://dx.doi.org/10.1016/j.patcog.2014.03.027
0031-3203/


**Fig. 1.** Different local appearances on the same facial position: (a) the face images are well-aligned and (b) the face images are mis-aligned.

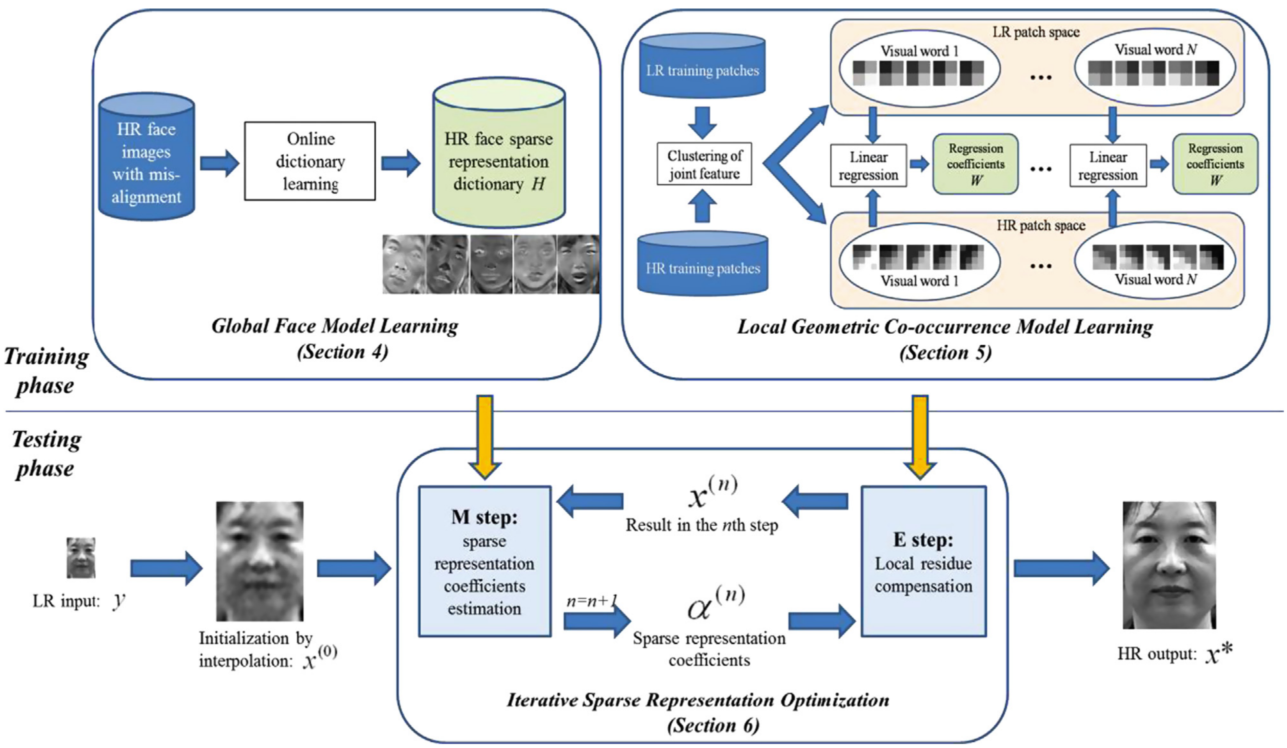


**Fig. 2.** The block diagram of the proposed method.

regression from LR to HR images. Yang et al. [12] used nonnegative matrix factorization (NMF) to linearly represent global face, and combine reconstruction constraint and representation coefficient regularization to estimate the HR face image.

Since both local and global approaches have their strength and weakness, hybrid approach [10–12] were then proposed to take the advantages of both approaches. The hybrid approach consists of two steps [10,11]: (i) hallucinating a smooth global face image by global method and (ii) compensating the residue with the HR ground truth by local methods. Yang et al. [12] zoomed the face image to a medium resolution by global methods, and then employed local methods to recover details.

Most of the existing face super-resolution methods assume that face images have been well-aligned by some facial landmarks (e.g., centers of two eyes and center of mouth). However, current state-of-the-art landmark detection algorithms locate facial landmarks with errors. As a result, the local patches from images will have more variations which we called mis-aligned variations. This effect can be illustrated in Fig. 1. We randomly get a local patch near the eye region of a face region. If all images are well-aligned, the extracted local patches are shown in Fig. 1(a). However, with the mis-aligned images, the extracted local patches are shown in Fig. 1(b). It can be easily seen that there is a much larger patch variations (lower similarities between patches) in mis-aligned case which makes similarity prior having negative effects. Existing local methods have intrinsic limitation that they are sensitive to the local feature mis-alignment even within $\pm 2$ pixels localization error on each landmark. In the following parts of this paper, landmark localization is referring to the landmark localization error introduced by the number of pixels between the located position and the ground truth. This small localization error of each landmark can severely change the geometric patterns of local image appearances as illustrated in Fig. 1(b) because three landmarks are required for alignment. In addition, to our experience, when the landmark alignment error is greater than or equal to $\pm 2$ pixels on a $32 \times 24$ LR face image, the results are not acceptable.

To overcome the problems, this paper proposes an iterative sparse representation method for face hallucination, which is robust to the mis-alignment. The block diagram of proposed method is shown in Fig. 2 and consists of training and testing phases. The training phase consists of two parts. The first part is global face sparse representation learning in which an online algorithm is used to learn a face prototype dictionary. The second part is a local geometric co-occurrence prior learning where we jointly cluster the HR and LR image residue patches, and learn the regression coefficients from LR to HR visual vocabulary. In the testing phase, when a LR face image is presented, an initial image is generated by simple interpolation. An iterative sparse representation model by using the learned dictionary is proposed to generate a HR image. The representation coefficients are