Image and Vision Computing 60 (2017) 171–184

Contents lists available at ScienceDirect

Image and Vision Computing

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

A novel correspondence-based face-hallucination method ☆

Zhuo Hui [a,*], Wenbo Liu [a], Kin-Man Lam [b]

[a] Department of Electrical and Computer Engineering, Carnegie Mellon University, United States
[b] Department of Electronics and Information Engineering, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Article history:
Received 25 May 2016
Received in revised form 17 December 2016
Accepted 2 January 2017
Available online 9 January 2017

Keywords:
Face hallucination
Face super-resolution
Corresponding-based method
Local kernel

ABSTRACT

This paper addresses the problem of estimating high-resolution (HR) facial images from a single low-resolution (LR) input. We assume that the input LR and estimated HR images are under the same view-point and illumination condition, i.e. the setting of image super-resolution. At the core of our techniques is that the facial images can be decomposed as a texture vector, characterized in terms of the appearance, and a shape vector, characterized in terms of the geometry variations. This enables a two-stage successive estimation framework that is geometry aware and obviates the needs in sophisticated optimizations. In particular, the proposed technique first solves for appearance of the HR faces form the correspondence derived between an interpolated LR face and its corresponding HR face. Given the texture of the HR faces, we incorporate optical flow to solve the local structure at sub-pixel level for the HR faces; here, we use additional geometry inspired priors to further regularize the solution. Experimental results show that our method outperforms other state-of-the-art methods in terms of retaining the facial-feature shape and the estimation of novel features.

1. Introduction

Face-hallucination techniques seek to estimate high-resolution (HR) facial images from a single face training sample. As to be expected, face-hallucination is an integral part of most practical face recognition and detection task, since facial images captured by video cameras are often blurred and of low resolution. While there are numerous techniques for performing face-hallucination, the vast majority of them are devoted to generic images [1–3]. Facial images have received the bulk of the attention precisely due to structural geometry and similar appearance where the methods [1–3] relying on assumption of generic images may not fully utilize. In this paper, we present the correspondence based method for recovering the HR facial images from single LR input. In the absence of additional assumptions or constraints, face hallucination is highly ill-posed and intractable since it is entirely possible that the HR image is constructed by its own unique configuration of the pixel values. As a consequence, the vast majority of prior techniques that perform face-hallucination method rely either on exemplars, or require knowledge of the prior to regularize the solution space, or make assumptions that individual faces can be well described by an estimated low dimensional space. While prior works deal with face hallucination from different perspectives, a common assumption made is that the HR face is a linear combination of a few unknown references. Specifically, we can represent the LR input as the linear combination of a few LR exemplars in the training set and apply the same linear relationship to the corresponding HR samples to reconstruct the HR face. The assumption significantly restrict the solution space and as a consequence, has been used in context of face hallucination [4–8]. The drawback of these approaches is that the operation highly depends on the vectors assumption which may be completely violated since the vector operations do not have a well-defined meaning for images. To account for this, Vetter and Troje [9] introduced the 3D model and decomposed an image into a texture vector, characterized by the reflectance of the face, and a shape vector, characterized by the geometry variation on the face. In particular, this technique computed the correspondence from a sequence of input images and estimate the 3D position for each scene point. Given the 3D shape, the texture can be solved via an inverse linear problem. While the technique produces precise face hallucination estimates, it requires the camera calibration information as well as the 3D estimation, both of which are not desirable in many practical applications. The technique proposed in this paper relies on the core principle of texture and shape decomposition without actually estimating the 3D model and obviating the camera calibration. The assumption we

☆ This paper has been recommended for acceptance by Xi Peng.
* Corresponding author.
E-mail address: huizhuo1987@gmail.com (Z. Hui).

http://dx.doi.org/10.1016/j.imavis.2017.01.002
0262-8856/

make is that the patches at the same spatial position denote the same type of facial feature. This provides the patch based correspondence between faces that can be used to restrict the solution space for both texture and shape vectors. We show that the texture and shape vectors can be estimated via a two-stage coarse-to-fine reconstruction. Given the patch based correspondence, we first solve for texture vectors by freezing the shape vector. Subsequently, we solve the shape vector from the estimated texture vectors. We perform the iteration for the shape and vectors until the convergence can be reached.

Contributions. We make the following contributions. 1. Correspondence based representation. In contrast to traditional pixel-based representation, the proposed method considers not only the changes in pixel intensity, but also the sub-pixel spatial displacements, making our proposed algorithm robust and reliable in the reconstruction of facial features. 2. Small training set. We leverage the spaces spanned by the shape and the texture vectors to restrict the solution space, making HR faces can be well estimated from a small training set.

2. Related work

Global approach. In [5], Chakrabarti et al. proposed using a PCA-based prior to reconstruct faces globally. Wang and Tang [4] proposed a holistic face-hallucination method which employs principal component analysis (PCA) to represent testing images as a linear combination of the reference samples in the training set. The PCA technique models the global variations of facial appearance in the eigenspace. This kind of method is more useful when the resolution of the facial images is very low. However, the performance of global approaches depends on the training set used, and may degrade greatly if the number of samples in the training set is insufficient.

Local approach. According to [10–13], the structural patterns or facial features in the HR faces can be regarded as the parent of the corresponding LR faces. Therefore, the features in HR images are estimated from the prior kernels established via learning the relationship between neighboring pixels in LR images. Recently, many algorithms have been proposed to evaluate the prior kernel. Liu et al. [6] further improved the example-based approach through the introduction of a two-stage (global-to-local) framework and the use of local constraints in HR facial-image reconstruction. In [6], the local linear-regression method is employed to extract the spatial-arrangement relation between LR and HR faces. Hu et al. [7] used a Gaussian function to characterize the local-pixel prior and to propagate the estimated prior in an image grid. However, these previous models, which employed local features to infer the corresponding HR features, lack effective methods to predict global structure patterns. In these methods, the holistic structures of facial images may be wrongly estimated, as a local patch prior is insufficient to infer a HR image, especially when the resolutions of testing images are very low.

Correspondence based approach. The traditional methods, which regard the concatenation of pixels as vectors, are incompatible with recovering facial attributes, such as the size of the eyes and the shape of the mouth. Hence, the raw images are required to be vectorized, rather than concatenating all the intensity values directly. In [9], Vetter and Troje decomposed an image into a texture vector and a shape vector, which encode the information contained in the face surface and in the spatial arrangement, respectively. The texture vector infers the changes in pixel intensities, and the shape vector indicates the spatial displacements. The texture and shape vectors are obtained through optical-flow algorithms. According to such a separation of the information in images, the raw images can be vectorized and the space spanned by the texture and the shape vectors can be regarded as linear vector space. This separation is based on the established pixel-to-pixel correspondence, in which each component in the vector space at the same position can refer to the same type of feature.

Manifold learning approach. Manifold learning refers to techniques where the high-dimensional smooth function is estimated from the data in the projected low-dimensional space. The vast majority of techniques are devoted to multi-view scenarios that the appearances of the objects typically lie close to a low-dimensional subspace. This naturally leads to the recent approach in [14] that fits a low dimensional subspace for the faces from different views, capturing the underlying structure of the space spanned by the faces. Concurrent to this work, Lu et al. [15] present a technique to extract the compact descriptors from the images under the single view point. For a given image, they fit a low dimensinal subspace to describe the pixels difference in a local region. Peng et al. [16] extend the idea by introducing the framework to reduce the effects of the errors from the projection space. In [17], Peng et al. further resolve the difficulties in optimization for the subspace clustering on the out-of-sample and large scale data. In the context of face hallucination, prior works have relied critically on the work of Wang et al. [4] where the HR face is estimated using linear combination of the basis from PCA. For the face hallucination in different views, Ma et al. [8] address the problem by utilizing PCA to characterize the collocated patches, allowing for the face hallucination via solving a least square problem. Fan et al. [18] extend the linear subspace analysis in [4] using Locally Linear Embedding (LLE) model, which captures nonlinear structure of the subspace, and employ the model to reconstruct the HR facial image. Finally, Hu et al. [19] model the global structure of the face as the sum of basis from PCA and solve for the local structure via a constrained regression problem based on local geometry from LLE.

3. Theory and algorithm

3.1. Correspondence-based representation

In our method, all face images are vectorized so that the *i*th component of the vectors derived corresponds to the same type of feature. Based on the vector spaces generated, the feature-by-feature correspondence between distinct vector spaces can be extracted. As described previously, an interpolated LR face image and its corresponding HR face images possess a similar structure. Hence, we can assume that the pixels at the same position in an interpolated LR face and the corresponding HR face refer to the same type of facial feature. Further, if we neglect the small spatial displacement, we can roughly treat the pixels at the same position as reflecting feature information of the same type. In Section 3.2, we will provide in detail the proof to derive this correspondence. Hence, in the first stage, we initially neglect the spatial displacement and roughly establish the pixel-by-pixel correspondence by using the patches at the same position in the interpolated LR and HR faces. In the second stage, we refer to the selected samples that are similar to the target face to estimate the spatial displacement neglected in the first stage.

3.2. Feature-by-feature correspondence relationship

In this sub-section, the correspondence relationship between LR and HR patches at the same position will be derived. Based on this derived relationship, we can use a two-step reconstruction to recover the HR images. The relationship between the LR and HR patches can be characterized as follows:

$$\mathbf{I}_L(x,y)_{x,y\in\Omega_i} = [G(x,y,\sigma) * \mathbf{I}_H(x,y)]_{x,y\in\Omega_i}, \quad (1)$$

where $\mathbf{I}_L$ denotes the interpolated LR faces, $\mathbf{I}_H$ denotes the corresponding HR image, G represents a Gaussian function which depends on the interpolation methods utilized, with standard deviation σ.