Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Face hallucination based on two-dimensional joint learning

Fang Liu [a], Yu Deng [b,*]

[a] China Ship Development and Design Center, Wuhan 430064, China
[b] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Face hallucination
Two dimensional coupled constraint
Two dimensional joint learning
Maximum a posteriori

ABSTRACT

In this paper, a face hallucination method based on two-dimensional joint learning is presented. Unlike the existing works on face super-resolution algorithms that first reshape the image or image patch into 1D vector, in our study the spatial construction of the high resolution (HR) and the low resolution (LR) face image are efficiently maintained in the reconstruction procedure. Enlightened by the 1D joint learning approach for image super-resolution, we propose a 2D joint learning algorithm to map the original 2D LR and HR image patch spaces onto a unified feature subspace. Subsequently, the neighbor-embedding (NE) based super-resolution algorithm can be conducted on the unified feature subspace to estimate the reconstruction weights. With these weights, the initial HR facial image can be generated. To refine further the initial HR estimate, the global reconstruction constraint is exploited to improve the quality of reconstruction result. Experiments on the face databases and real-world face images demonstrate the effectiveness of the proposed algorithm.

1. Introduction

Due to the limitations of imaging equipment and imaging environment, the resolution of captured face image is normally low in real scenario, causing the failure of face recognition system. Therefore, it is especially necessary to recover the high-resolution (HR) face image from single or multiple low-resolution (LR) face images. However, it is difficult to obtain multiple faces of someone in real application such that many researchers mainly focus on producing the HR face image from single LR face image.

In the past several decades, various face super-resolution methods have been proposed to super-resolve single LR face image. These technologies can be roughly grouped into three categories: interpolation-based methods [1–3], reconstruction-based methods [4–6] and learning-based methods [7–11]. The interpolation-based strategies, including nearest neighbor interpolation [1], bilinear interpolation [2] and bicubic interpolation [3] etc., can generate the HR image via exploiting the natural image priors directly. Though interpolation-based methods increase the image resolution simply, the performance is often poor since no additional information is utilized [12]. The reconstruction-based methods attempt to reconstruct the HR images based on sampling theory by simulating the image degradation procedure. However, the methods are usually applicable to generic natural scenes rather than face images because the specific prior information about face image is not incorporated into the reconstruction process. What's more, the performance of reconstruction-based method will degrade rapidly when the magnification factor becomes large. The learning-based method can recover the high-frequency details of LR facial image from a training set of LR and HR image pairs. That is, the learning-based face super-resolution method, as known as face hallucination, try to model the relationship between the LR face images and the corresponding HR face images, and then use the learnt relationship to infer the HR face image for a LR probe one. Therefore, establishing a good learning model to describe the correlation is the key to generate HR face image.

Baker et al. [13] first proposed a face hallucination method based on Bayesian formulation, which can infer HR image from an input LR one by finding the nearest "parent structure" at each pixel via training. Liu et al. [14] proposed a two-step face super-resolution method by integrating a global parametric model and a local nonparametric model. After that, a variety of learning-based face super-resolution algorithms have been developed. In [15], Wang et al. adopted principle component analysis (PCA) to represent the structural similarity of face images. Following Wang's work, some classical global face models, such as locality preserving projections (LPP) [16], non-negative matrix factorization (NMF) [17] and canonical correlation analysis (CCA) [18], have been used to globally reconstruct HR face image. However, the global methods cannot efficiently reflect the local face appearance variations which can be used to recover the fine facial details. In order to generate high quality face image and maintain visual ra-

* Corresponding author.
E-mail address: dengyu86@gmail.com (Y. Deng).

http://dx.doi.org/10.1016/j.dsp.2015.09.006

tional, a large number of position patch based face hallucination methods have been proposed in recent years [19].

The position patch based methods mainly consist of three steps: 1) each LR training face and the corresponding HR one are divided into overlapped patches in terms of the patch position, respectively. Meanwhile, each input LR face image should be divided into overlapped patches in the same way; 2) the input LR patch is represented by the weighted sum of LR training patches which have the same spatial position as the input patch, and then the reconstruction weights is calculated by using some optimization approaches, such as least squares algorithm; 3) the target HR patch can be generated by replacing the LR training patches with the corresponding HR training patches. Among these position patch based face hallucination methods, one of the most representative strategies is the neighborhood embedding (NE)-based face hallucination method [20].

The nature of NE-based face super-resolution technology is that LR and HR training patches form manifolds with similar local geometry in two distinct manifold spaces. The NE-based methods have also shown impressive performance [18,21,22]. However, Su's experiments demonstrate that the local consistency assumption for LR and HR image patches (features) rarely holds due to the one-to-many mapping between the LR image and HR images. That is, the local geometric structure of LR patch (feature) manifold space may not consistent with that of the corresponding HR manifold space [23]. Thereafter, the inconsistencies often result in large reconstruction errors.

To address this issue, some variations of NE algorithm have been proposed for image super-resolution [24–28]. In [24], a compact yet descriptive training set has been constructed from characteristic regions in images. Thus, the manifold assumption holds well in the training set. Li et al. [25] utilized locality preserving constraints (LPC) to avoid confusions through emphasizing the consistency of localities on LR and HR manifolds explicitly. Inspired by [26], Gao et al. suggested a joint learning technique to train two projection matrices simultaneously and to map the original LR and HR spaces onto a unified feature subspace, and then the nearest neighbors of each input LR image patch are selected via k-nearest neighbor (k-NN) selection method in the unified subspace to estimate the reconstruction weights. Hao et al. [27] exploited the Easy-Partial Least Squares (ES-PLS) algorithm to learning two projection matrices simultaneously, via which original HR and LR image patches are mapped onto a unified feature space. Using the two learnt projection matrices, the consistency relationship between LR representation manifold and the corresponding HR representation manifold can be well hold. Recently, Jiang et al. [28] proposed using an intermediate dictionary learning scheme to bridge the LR manifold and the original HR one. Therefore, their method can efficiently enhance the consistency between the reconstructed HR manifold and the original HR manifold. In general, these methods can faithfully capture the intrinsic relationship between LR manifold and the corresponding HR one. Extensive experiments demonstrate that the improved methods outperform the NE-related baselines.

However, all the existing NE-based face hallucination methods need reshape the image data into a 1D vector previously. The main disadvantages of 1-D vectorization process contain two parts: 1) high computation complexity; 2) the spatial structure information of image data is lost. In order to address the issues, two-dimensional principal component analysis (2D PCA) and two-dimensional canonical correlation analysis (2D CCA) have been proposed for image analysis. To the best of our knowledge, Le An et al. [11] were the first to propose a two-step 2D CCA based face hallucination method which can preserve the intrinsic 2D spatial structure of face images during the super-resolution process. However, the approach cannot well reflect the relationship between the LR training images and the HR counterparts due to the limitations of 2D CCA algorithm. Inspired by the works of Gao et al. [26] and Le An et al. [11], we propose a novel face hallucination method based on two-dimensional joint learning. In summary, our method contains three steps. At first, for each input LR image patch, the two-dimensional nearest grouping patch pair (2D GPP) is constructed by linking the nearest LR training image patches with their HR counterparts together. To select the nearest neighbors in the case of 2D data, the Frobenius norm is used as the distance metric in the k-nearest neighborhood selection algorithm. Secondly, we propose a two-dimensional joint learning algorithm to learn the projection matrices for the 2D GPP associated with each LR input patch. Next, the measurements of LR and HR patches in the corresponding 2D GPP are projected onto a unified feature subspace with the assistant of learned projection matrices. Thereafter, the local consistency between LR patch manifold and the corresponding HR patch manifold can be faithfully preserved in the unified feature subspace. The 2D neighbor embedding algorithm is exploited to estimate the optimal reconstruction weights in the unified subspace, and then the corresponding HR training patches are combined linearly with the reconstruction weights to generate HR face patches. At last, a post-processing step is introduced to compensate the high-frequency details and improve the global smoothness. The system framework of the proposed face hallucination algorithm is illustrated in Fig. 1.

Compared to the 1D joint learning for image SR via a coupled constraint in [26], the contributions of our study are as follows:

1) Instead of transforming the 2D image patch to a 1D vector, the LR-HR face image patch pairs are used to construct the 2D GPP directly, which maintain the spatial structure information of image patches well.

2) For each 2D GPP with coupled constraint, the 2D joint learning algorithm is performed to learn the projection matrices L_l, L_h, R_l and R_h, and then the original LR and HR image patch spaces are directly mapped into a unified subspace via the learned projection matrices. As a result, the local consistency between LR and HR patch manifold spaces can be preserved well in the unified subspace.

The remainder of this paper is organized as follows: Section 2 reviews briefly the 1D joint-learning method via coupled constraint for image SR and introduces the proposed 2D joint-learning algorithm. Section 3 presents the proposed face hallucination method. The extensive experiments and comparisons are given in Section 4, and this paper is concluded in Section 5.

2. 1D and 2D joint learning

2.1. 1D joint learning formation

To reduce the differences between the local manifolds of LR and HR image patches, the 1D joint learning algorithm for image SR via a coupled constraint is introduced in [26]. Given two zero mean (centered) datasets consisting of LR and HR image patches, $L = \{l_1, l_2, \ldots, l_N\}$ ($l_i \in R^m$) and $H = \{h_1, h_2, \ldots, h_N\}$ ($h_i \in R^n$), respectively, a coupled set $C = \{c_1, c_2, \ldots, c_N\}$ can be constructed by concatenating each feature vector l_i and h_i [26]:

$$c_i = \begin{bmatrix} l_i / \sqrt{m} \\ h_i / \sqrt{n} \end{bmatrix} \tag{1}$$

For each vector c_i in the coupled set C, the k nearest neighbors associated with it are selected to generate the k-nearest grouping patch pairs (GPP) as follows

$$G^i = \{c_j\}_{j \in N_k(i)} \tag{2}$$

Please cite this article in press as: F. Liu, Y. Deng, Face hallucination based on two-dimensional joint learning, Digital Signal Process. (2015), http://dx.doi.org/10.1016/j.dsp.2015.09.006