Neurocomputing 134 (2014) 92–99

Contents lists available at ScienceDirect

Neurocomputing

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

A novel sparse representation based framework for face image super-resolution

Guangwei Gao, Jian Yang*

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

ARTICLE INFO

Article history:
Received 2 June 2012
Received in revised form
25 October 2012
Accepted 31 December 2012
Communicated by Chennai Guest Editor
Available online 3 August 2013

Keywords:
Face image
Super-resolution
Sparse representation
MAP
Face recognition

ABSTRACT

In this paper, we present a new face image super-resolution framework using the sparse representation (SR). Firstly, a mapping function between the embedding geometries in respective image space is estimated from a training set. For super-resolution, we first seek a sparse representation for each low-resolution (LR) input, and then the representation coefficients are mapped to generate the corresponding representation coefficients in high-resolution (HR) space. Finally, the mapped coefficients are used to reconstruct the initial estimation of the target HR image. To obtain the HR images with higher fidelity, the maximum a posteriori (MAP) formulation is introduced. The effectiveness of the proposed method is evaluated through the experiments on the benchmark face database, and the experimental results demonstrate that the proposed method can achieve competitive performance compared with other state-of-the-art methods.

1. Introduction

In the applications such as video surveillance and object tracking, owing to the long distance between the objects and the camera, the obtained images are often with severe degradations (e.g., low-resolution, etc.) and indiscernible by humans. Such images make great challenges in automatic face recognition and identification.

Quite a number of classic face recognition methods and systems [1–9] have been developed to perform well under controlled conditions, where face images both in training set and the testing set have good quality (e.g., high-resolution). Nevertheless, Wang [10] and Gunterk [11] showed in their research that the recognition accuracy of current face recognition systems significantly drop off as the resolution of facial images decreases (In Gunterk's experiments, the decrease was from 79% to 44%). However, to our joy, with the super-resolution reconstruction, the recognition rate can be improved significantly, preserving a high and constant level or even getting close to that of the original high-resolution face images. Based on the above observations, it has a great necessity to make some enhancement on the image resolution. Face image super-resolution, also called face hallucination [12], which aims to render a high-resolution (HR) image from its corresponding low-resolution (LR) one, provides a solution to these problems.

The simplest way to increase image resolution is direct interpolation [13–15], such as bilinear, bicubic and cubic spline interpolators. However, they often reconstruct HR images with jaggy and zipping artifacts since no new information is added in the procedure. Iterative back-projection (IBP) [16], which minimizes the reconstruction error by iteratively back-projecting the reconstruction error into the reconstructed image, is presented. The advantage of IBP is that it is understood intuitively and easily. However, due to the fact that the reconstruct errors are back projected in IBP, it also produces many jaggy and ringing artifacts. To overcome this drawback, Dong [17] presented a novel non-local iterative back-projection algorithm for image enlargement, which incorporates the non-local information into the IBP process so that the reconstruction errors can be reduced.

Wang [18] presented a face hallucination method using eigentransformation, which views hallucination as a transformation between different image styles. They used Principle Component Analysis (PCA)[19] to code the input face image as a linear combination of the LR face images in the training set, and the HR image is rendered through replacing the LR training images with corresponding HR ones, retaining the same representation coefficients. The similar idea is used in [20]. However, due to the fact that the PCA is based on the optimal representation criterion in the sense of mean-square error, and the PCA, which is based on L2-norm, is sensitive to noise.

Very recently, a pioneer work was reported by Wright [21], where the sparse representation (SR) is employed for feature extraction and classification, and series of methods based on SR are developed for image super-resolution [22–23]. Ref. [22] seeked

* Corresponding author at: School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China.
Tel.: +86 25 8431 7297.
E-mail address: csjyang@njust.edu.cn (J. Yang).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.12.059

a sparse representation for each patch of the low-resolution input, and then used the coefficients of this representation to generate the high-resolution output. The similar intuitive is used in [23] which is based on position-patch.

Motivated by the work in [18] and the SR mentioned above, we propose a new face image super-resolution framework using SR. Differing from the method in [18], we use SR to calculate the representation coefficients while the obtained embedding geometry is not directly used in the following step. Firstly, a mapping function of the coefficients (i.e. embedding geometry) is estimated from the training set, which is extremely different from the existing methods. The mapping function can reveal the intrinsic relationship between the two styles' domains. Then the LR training samples are used as the dictionary to code the input LR image as a sparse linear combination via l1-norm minimization. The calculated coefficients are mapped by the computed mapping function to generate the corresponding representation in the HR image space. Finally, the mapped representation coefficients are used to reconstruct the preliminary estimation of the target HR image by replacing the LR training samples with corresponding HR counterparts. In order to obtain the HR images with higher fidelity, the maximum a posteriori (MAP) formulation is introduced so as to derive the final reconstruction result. As a prior, this preliminary estimation not only constrains the target HR face to reside in the image space spanned by the HR training samples, but also enforces a specific embedding geometry learned from the LR image space for the target face. The proposed method can automatically select the samples in the training dictionary which make a great contribution to the reconstruction precision of the HR image. Our experiments clearly validate the performance of the proposed method.

The rest of the paper is organized as follows. In Section 2, we briefly review the eigentransformation based method. Section 3 presents the proposed SR based algorithm. The whole framework for face super-resolution is introduced in Section 4. Section 5 conducts experiments to evaluate the proposed method and the conclusion is given in Section 6.

2. Eigentransformation based super-resolution method

The main idea of the eigentransformation based super-resolution algorithm [18] is to apply PCA to the original LR training data set so as to learn the embedding geometry of the input LR image in the LR image space. The learned embedding geometry is directly used to generate the corresponding HR counterpart.

Let $X_h=[h_1, h_2\ldots h_M]\ R^{h\times M}$ be the HR training image set, $X_l=[l_1, l_2\ldots l_M]\in R^{l\times M}$ be the corresponding LR one. Denote by $m_l=(1/M)\sum_{i=1}^{M} l_i$ and $m_h=(1/M)\sum_{i=1}^{M} h_i$ the grand mean faces of the respective LR and HR training set. We can remove the mean face from each image, resulting in $X'_h=[h_1-m_h, h_2-m_h\ldots h_M-m_h]=[h'_1\ldots h'_M]$ and $X'_l=[l_1-m_l, l_2-m_l\ldots l_M-m_l]=[l'_1\ldots l'_M]$, respectively. The mathematical formulation of the eigentransformation based method can be easily expressed as follows:

$$\begin{array}{l} r_l=c_1l'_1+c_2l'_2+\cdots+c_Ml'_M+m_l=\sum_{i=1}^{M}c_il'_i+m_l \\ \quad\downarrow\quad\downarrow\quad\cdots\quad\downarrow\quad\downarrow \\ r_h=c_1h'_1+c_2h'_2+\cdots+c_Mh'_M+m_h=\sum_{i=1}^{M}c_ih'_i+m_h \end{array} \tag{1}$$

where $c=[c_1,c_2\ldots c_M]^T$ is the combination coefficients computed by eigentransformation, r_l is the LR input image, r_h is viewed as the final estimation of the corresponding target HR image.

The eigentransformation based method provides an intuitive, easy and understandable framework for face super-resolution, and can achieve a satisfactory result. However, just as the discussion in introduction, although PCA is a well data representation technique, it is sensitive to noise and can't automatically select the role that can contribute to the reconstruction process. Thus, some unnatural artifacts may appear in the reconstructed HR images by eigentransformation based method.

3. SR based face super-resolution method

Inspired by the successful application of sparse representation based classification (SRC) in face recognition [20], in this section, we will introduce the basic idea of the proposed SR based method. For the convenience of expression, we also denote by $X_h\in R^{h\times M}$ the HR training samples, $X_l\in R^{l\times M}$ the corresponding downsampled version of X_h, respectively. We aim to seek a sparse representation for each LR input face image in terms of the LR dictionary X_l.

The problem of finding the sparse solution c of a LR input image $y\in R^m$ (not included in the training set) over the corresponding given dictionary can be defined as follows:

$$\hat{c}=\arg\min_{c}\{\|y-X_lc\|_2^2+\lambda\|c\|_1\} \tag{2}$$

The corresponding HR image Y can be easily reconstructed as follows:

$$Y=X_h\hat{c} \tag{3}$$

The above idea is simple and intuitive, and has been applied in [22,23]. In the aforementioned existing method, it is supposed that the target HR image is believed to have the same embedding geometry in the HR image space as its LR counterpart in the LR image space. Nevertheless, this assumption may be unreasonable.

To demonstrate this mathematically, we suppose that c^l and c^h are the respective combination coefficient vectors of an LR input face image and its corresponding HR counterpart in the respective original image space. They are calculated over the given training data set, with different image resolutions. The given data set is composed of 1196 LR–HR face pairs. We compute c^l and c^h based on the first 500 training pairs (i.e. $M=500$), the remainder is used as the testing case. The normalized error vector e between each pair c^l and c^h is defined as follows:

$$e=\frac{c^l-c^h}{\max(c^h)-\min(c^h)} \tag{4}$$

For the given data set, we can obtain 696 values of e. Fig. 1 shows the mean value and the standard deviation of the whole e. We can see that the mean value of these error vectors is close to zero, which indicates that the combination coefficients of an input face in the respective LR and HR image space are similar to each other, only with slight fluctuation. However, from the perspective of the standard deviation which is denoted as red lines in Fig. 1, we can see that these errors should not be neglected. It may be the chief reason that results in unsatisfactory reconstruction results in our face image super-resolution experiments. Therefore, we should try to reduce these errors in Eq. (4) and hope to obtain more satisfactory results in face image reconstruction.

To this end, we introduce a mapping function P between the coupled combination coefficients and it is pre-learned from a training set. Notice that we don't know the definite formation of the mapping matrix in advance. The mathematical formulation of the relationship can be easily defined as follows:

$$\Lambda_h=P\Lambda_l, \tag{5}$$

where $\Lambda_l=[c_1^l,c_2^l\cdots c_M^l]$,$\Lambda_h=[c_1^h,c_2^h\cdots c_M^h]$. In [18,22,23], the mapping function P is predefined as an identity matrix.