Pattern Recognition 47 (2014) 1261–1270

Contents lists available at ScienceDirect

Pattern Recognition

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

Face hallucination based on sparse local-pixel structure ☆

Yongchao Li [a,1], Cheng Cai [a,*], Guoping Qiu [b,c], Kin-Man Lam [d]

[a] Department of Computer Science, College of Information Engineering, Northwest A&F University, Xi'an, China
[b] School of Computer Science, University of Nottingham, United Kingdom
[c] International Doctoral Innovation Centre, The University of Nottingham, Ningbo, China
[d] Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Article history:
Received 21 January 2013
Received in revised form
28 June 2013
Accepted 16 September 2013
Available online 25 September 2013

Keywords:
Face hallucination
Sparse local-pixel structure
Super-resolution
Sparse representation

ABSTRACT

In this paper, we propose a face-hallucination method, namely face hallucination based on sparse local-pixel structure. In our framework, a high resolution (HR) face is estimated from a single frame low resolution (LR) face with the help of the facial dataset. Unlike many existing face-hallucination methods such as the from local-pixel structure to global image super-resolution method (LPS-GIS) and the super-resolution through neighbor embedding, where the prior models are learned by employing the least-square methods, our framework aims to shape the prior model using sparse representation. Then this learned prior model is employed to guide the reconstruction process. Experiments show that our framework is very flexible, and achieves a competitive or even superior performance in terms of both reconstruction error and visual quality. Our method still exhibits an impressive ability to generate plausible HR facial images based on their sparse local structures.

1. Introduction

The idea of super-resolution (SR) was first presented by Tsai and Huang [1], and significant progress has been made with it over the last few decades. Since SR is an ill-posed problem, prior constraints are necessary to attain a good performance. Based on the different approaches to attaining these prior constraints, SR methods can be broadly classified into two categories: one is the conventional approach, which is also widely known as multi-image SR [2–5] or regularization-based SR, and which reconstructs a HR image from a sequence of LR images of the same scene. These algorithms mainly employ regularization models to solve the ill-posed image SR, and use smooth constraints as the prior constraints, which are defined artificially. The other approach is single-frame SR [6–11], which is also called learning-based SR or example-based SR. These methods generate a HR image from a single LR image with the information learned from a set of LR–HR training image pairs. These algorithms attain the prior constraints between the HR images and the corresponding LR images through a learning process. Many example-based or learning-based algorithms [6–16] have been proposed in the field of image processing. Also in SR, Qiu [13] and Baker and Kanade [17] have demonstrated that the smooth prior constraints used in many regularization-based methods will become less effective at solving the SR problem as the zooming factor increases, while example-based approaches have the potential to overcome this problem using advances in machine learning and computer vision. In this paper, we focus on the single-image SR problem.

Fig. 1 shows a general framework of example-based SR: the input LR image is first interpolated, using the conventional methods, to the size of the target HR image, and the input interpolated LR image – a blurry image lack of high-frequency information – is then used as the initial estimation of the target HR. The input LR image is also divided into either overlapping or non-overlapping image patch, and the example-based framework will use the image patches to find out the most matched examples by searching a training dataset of LR–HR image pairs. The selected HR examples are then employed to learn the HR information as the prior constraints. Finally, the learned HR information and the input interpolated image are combined to evaluate the target HR image.

The idea of face hallucination was first proposed by Baker and Kanade [18], and it was then used for SR problems in [8,11,16,19]. Example-based face hallucination is a subcategory of the

Abbreviations: SR, super resolution; HR, high resolution; LR, low resolution; PSNR, peak signal to noise ratio; SSIM, Structural Similarity Index; SRM, sparse representation models; NARM, nonlocal autoregressive model

* Corresponding author. Tel.: +86 1899 1291 338; fax: +86 029 87092353.
E-mail addresses: ychao_lee@nwsuaf.edu.cn (Y. Li), cheney.chengcai@gmail.com (C. Cai), qiu@cs.nott.ac.uk (G. Qiu), enkmlan@polyu.edu.hk (K.-M. Lam).
[1] Postal address: Room 322, College of Information Engineering, Northwest A&F University, Yangling, Xi'an, Shaanxi 712100, China.

0031-3203/$ - see front matter © 2013 The Authors. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.09.012

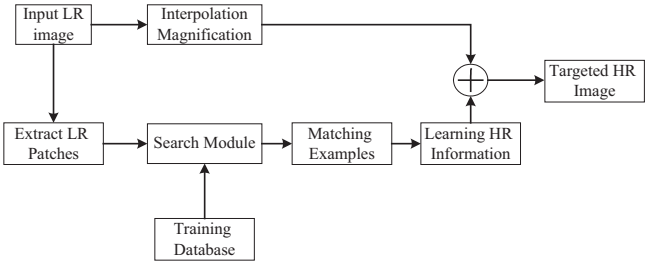

Fig. 1. A general framework of example-based SR methods.

framework shown in Fig. 1, and it is also a specific and important category of image SR. In [20], Liang conducted a good survey of face hallucination, and summarized the existing face-hallucination methods into two approaches: namely global similarity and local similarity. The theoretical backgrounds and practical results of the existing face-hallucination techniques and algorithms are compared. Based on the comparison results, the strengths and weaknesses of each algorithm are summarized, which forms a base for proposing an effective method to hallucinate mis-aligned face images.

In [16], Liu et al. argued that a successful face-hallucination algorithm should meet the following three constraints:

1. *Sanity constraint*: the target HR image should be very close to the input LR image when smoothed and down-sampled.
2. *Global constraint*: the target HR image should have the common characteristics of human faces, e.g., possessing a mouth and a nose, being symmetrical, etc.
3. *Local constraint*: the target HR image should have the specific characteristics of the original LR face image, with photorealistic local features.

Furthermore, a two-step approach was developed for face hallucination, in which a Bayesian formulation and a nonparametric Markov network are employed to deal with face hallucination. Of all the various methods for face hallucination, Hu et al. [11] first learned local-pixel structures from the most matched example images explicitly, which are then used directly as reconstruction priors. A three-stage face-hallucination framework was proposed in [11], which is called Local-Pixel Structure to Global Image Super-Resolution (LPS-GIS). In Stage 1, k pairs of example faces which have a similar pixel structure to the input LR face are selected from a training dataset using k-Nearest Neighbors (KNN). They are then subjected to warping using optical flow, so that the corresponding target HR image can be reconstructed more accurately. In Stage 2, the LPS-GIS method learns the face structures, which are represented as coefficients using a standard Gaussian function; the learned coefficients are updated according to the warped errors. In Stage 3, LPS-GIS constrains the revised face structures, namely the revised coefficients to the input LR face, and then reconstructs the target HR image using an iterative method.

Unlike the abovementioned methods, we propose a face-hallucination framework which utilizes the sparse local-pixel structure as the prior model in the reconstruction of HR faces. The use of sparse local-pixel structure allows our method to reconstruct the details in HR faces flexibly. Furthermore, the global structure of faces is also considered, which enables the proposed method to produce plausible facial components.

As for the organization of this paper, Section 2 gives a brief introduction to the theory of sparse representation and its recent applications to super-resolution. Section 3 provides a detailed introduction to a concept called 'local-pixel structure with sparsity'. The details of our proposed framework are presented in Section 4. Section 5 presents the experiments and an evaluation of the proposed framework. Finally, the concluding remarks are given in Section 6.

2. Related works on super-resolution with sparse representation

Single-image super-resolution produces a reconstructed HR image from an input LR image using the prior knowledge learned from a set of LR–HR training image pairs, and the reconstructed HR image should be consistent with the LR input. An observed model between a HR image and its corresponding LR counterpart is given as follows:

$$\mathbf{I_l} = \mathbf{I_h}\mathbf{H}S(r) + \mathbf{N}, \tag{1}$$

where $\mathbf{I_l}$ and $\mathbf{I_h}$ denote the LR and HR images, respectively; $\mathbf{H}$ represents a blurring filter; $S(r)$ is a down-sampling operator with a scaling factor of r in the horizontal and vertical dimensions; and $\mathbf{N}$ is a noise vector, such as the Gaussian white noise. Here, we will focus on the situation whereby the blur kernel is the Dirac delta function as [11,44], i.e. $\mathbf{H}$ is the identity matrix. Thus, Eq. (1) can be rewritten as follows:

$$\mathbf{I_l} = \mathbf{I_h}\mathbf{H}S(r) + \mathbf{N}. \tag{2}$$

Therefore, the purpose of SR is to recover as much of the information lost in the down-sampling process as possible. Since the reconstruction process still remains ill-posed, different priors can be used to guide and constrain the reconstruction results. In recent years, the sparse representation model (SRM) has been used as the prior model, and has shown promising results in image super-resolution.

Sparse representation of a signal is based on the assumption that most or all signals can be represented as a linear combination of a small number of elementary signals only, called atoms, from an overcomplete dictionary. Compared with other conventional methods, sparse representation can usually offer a better performance, with its capacity for efficient signal modeling [21]. The sparse representation of signals has already been applied in many fields, such as object recognition [22,23], text categorization [24], signal classification [21], etc.

In the sparse representation, a common formulation of the problem of finding the sparse representation of a signal using an overcomplete dictionary is described as follows:

$$\hat{\omega}_0 = \min \|\omega\|_0, \quad \text{s.t.} \quad \psi = \mathbf{A}\omega, \tag{3}$$

where $\mathbf{A}$ is an $M \times N$ matrix whose columns are the elements of the overcomplete dictionary, with $M < N$, and $\psi \in R^{M\times 1}$ is an observational signal. The purpose of sparse representation is to find an $N \times 1$ coefficient vector ω, which is considered to be a sparse vector, i.e. most of its entries are zeros, except for those elements in the overcomplete dictionary $\mathbf{A}$ which are associated with the observational signal ψ. Solving the sparsest solution for (3) has been found to be *NP*-hard, and it is even difficult to approximate [25]. However, some recent results [26,27] indicate that if the vector ω in (3) is sparse enough, then the problem can be solved efficiently by minimizing the ℓ_1-norm instead, as follows:

$$\hat{\omega}_1 = \min \|\omega\|_1, \quad \text{s.t.} \quad \psi = \mathbf{A}\omega. \tag{4}$$

In fact, as long as the number of nonzero components in ω_0 is a small fraction of the dimension M, the ℓ_1-norm can replace and recover the ℓ_0-norm efficiently [22]. In addition, the optimization problem of the ℓ_1-norm can be solved in polynomial time [28,29]. However, in real applications, the data in the dictionary $\mathbf{A}$ are, in general, noisy. This will lead to the result whereby the sparse representation of an observational signal, in terms of the training data in $\mathbf{A}$, may not be accurate. In order to deal with the problem, (4) can be relaxed to a modified form as follows:

$$\hat{\omega}_1 = \min \|\omega\|_1, \quad \text{s.t.} \quad \|\psi - \mathbf{A}\omega\|_2 \le \varepsilon. \tag{5}$$