Digital Signal Processing 31 (2014) 28–34

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Parameter estimation in sparse representation based face hallucination

Zhongyuan Wang *, Ruimin Hu, Zhenfeng Shao, Zhiqiang Hou

NERCMS, School of Computer, Wuhan University, Wuhan, 430072, China

ARTICLE INFO

Article history:
Available online 15 May 2014

Keywords:
Face hallucination
Sparse representation
Patch size
Regularization parameter

ABSTRACT

Owning to the excellent ability to characterize the sparsity of natural images, ℓ_1 norm sparse representation is widely applied to face hallucination. However, the determination on two key parameters such as patch size and regularization parameter has not been satisfactorily resolved yet. To this end, we proposed a novel parameter estimation method to identify them in an analytical way. In particular, the optimal patch size is derived from the sufficient condition for reliable sparse signal recovery established in compressive sensing theory. Furthermore, by interpreting ℓ_1 norm SR as the corresponding maximum *a posteriori* estimator with Laplace prior constraint, we obtain an explicit expression for regularization parameter in statistics of reconstruction errors and coefficients. Our proposed method can significantly reduce the computational cost of parameter determination while without sacrificing numerical precision and eventual face hallucination performance. Experimental results on degraded images in simulation and real-world scenarios validate its effectiveness.

1. Introduction

Face super-resolution, or face hallucination, refers to the technique of estimating a high-resolution (HR) face image from low-resolution (LR) face image sequences or a single LR one. Due to restricted imaging conditions in many scenarios, it is hard to capture HR face images, and thus face hallucination is extensively used for pre- and/or post-processing in video applications, such as video surveillance and face recognition. In their pioneering work on face hallucination [1], Baker and Kanade employed a Bayesian approach to infer the missing high-frequency components of an input LR image from a parent structure with HR/LR training samples, leading to a large magnification factor with relatively good results. Liu et al. [2] proposed a two-step statistical modeling approach that integrates global structure reconstruction with local detail refinement.

Following [1,2], learning-based face hallucination approaches have gained great popularity in recent years. The main idea is to estimate an HR face image from a single LR face image with the help of a training set of HR and LR image pairs. For example, singular value decomposition [3] and morphological component analysis [4] are respectively used to learn mapping coefficients from LR–HR training pairs. Owning to the excellent performance grasping salient properties of natural images, sparse representation (SR) has attracted more attention in face hallucination. Yang et al. [5] are the first to introduce SR to image super-resolution, where images are approximated by an over-complete dictionary for adaptive sparse image decompositions. This work spurred much follow-up research on SR based face hallucination. In [6], authors presented a dual-dictionary learning method to recover more image details, in which not only main dictionary but also residual dictionary are learned by sparse representation.

In addition to the generic sparsity prior, some specific image priors are further exploited to boost performance in SR based image restoration. Dong et al. [7] proposed non-locally centralized sparse representation to explore the image nonlocal self-similarity. For a class of highly structured objects, such as human faces, the prior of facial positions is of importance and can be utilized to retain the holistic structure of face images. Following this idea, Ma et al. [8] proposed a sparse representation and position prior based face hallucination method. This method classifies high- and low-resolution atoms to form local dictionaries according to the different regions of human face and then uses different local dictionaries to hallucinate the corresponding regions of a face. Generally, SR based methods may select very distinct patches that are far from the input patch to favor sparsity and consequently result in dissimilarity in terms of Euclidean distance. To address this problem, literature [9] introduced a similarity constraint into sparse representation to promote accuracy and stability.

Sparse representation is effective in face hallucination problem when sufficient observations are available, but there are at least two questions we need to answer. As usually done, face hallucination is performed on the basis of small image patches instead of

* Corresponding author.
E-mail address: wzy_hope@163.com (Z. Wang).

http://dx.doi.org/10.1016/j.dsp.2014.05.005
1051-2004/© 2014 Elsevier Inc. All rights reserved.

a whole image. Thus, firstly, how to choose the optimal patch size given an observed image is worth investigating. Secondly, SR uses a regularization parameter λ to reach a reasonable balance between regression error and sparsity penalty, yet how to efficiently determine λ is unresolved.

To the best of our knowledge, very few face hallucination literatures address patch size choice in a reasoning way. Wang [10] presented an algorithm to detect the lower bound of patch size by measuring the range of the local features in the texture. Kwatra et al. [11] proposed a graph-cut based method, in which the whole input texture is used as a patch and is cut when stitching to other patches. However, these two methods are primarily proposed for texture synthesis in pixel domain (e.g., image in-painting) rather than learning-based super-resolution.

In contrast, classic parameter identification techniques have ever been used to find regularization parameter in super-resolution, such as generalized cross validation (GCV) [12] and L-curve [13]. L-curve is a tool for showing the parametric plot of the error versus penalty with the regularization parameter λ varied. The value of λ at the corner of the L-curve is the desired outcome. GCV produces asymptotically optimum value by iteratively solving a constrained minimization problem. GCV and L-curve can provide high quality solutions, whereas they are computationally expensive.

In practice, representative learning-based face hallucination approaches [4–9] have to use empirical parameter for patch size or λ. All patch-based methods mentioned above set an empirical value for patch size without any theoretical justification. Similarly, the choice of the regularization parameter is based on trial-and-error experiments, which is actually a manual tuning procedure. Although this empirical practice is relatively reliable, it is cumbersome and computationally intensive because of repeated manual trials.

In this paper, we take advantage of the well-known results from the compressive sensing theory to deduce the optimal patch size. As stated by compressive sensing theory, to faithfully recover sparse signal, the dimensions of observed and sparse signals as well as the sparsity should satisfy a pre-known constraint relationship. For SR based face hallucination, the dimensions of observed and sparse signals correspond to the patch size and the number of training images, respectively. Hence the proper patch size may be derived from this well-established constraint equation. Additionally, ℓ_1 norm SR can be interpreted as maximum *a posteriori* (MAP) estimator with Laplace prior imposed on solution. Under MAP framework, regularization parameter λ completely depends on the statistics of noise and coefficients, which actually implies an explicit way for determining λ. In our proposed method, the patch size and regularization parameter are analytically tractable, leading to high efficiency as well as pretty convenience. Experimental results in face hallucination task validate its effectiveness.

The remainder of this paper is organized as follows. Section 2 presents the proposed parameter estimation method in detail. Experimental results are shown in Section 3. In Section 4, we conclude the paper.

2. Proposed method

In this section, we particularly address the problem in determining patch size and regularization parameter. Fig. 1 outlines the patch-wise face hallucination framework, consisting of four parts. Among them, the training of SR coefficients and the reconstruction of HR images are based on the SR model introduced in Section 2.1, whose specific implementation details are referred to literatures [8,9]. In this paper, we mainly focus on the determination technique of patch size and regularization parameter λ.

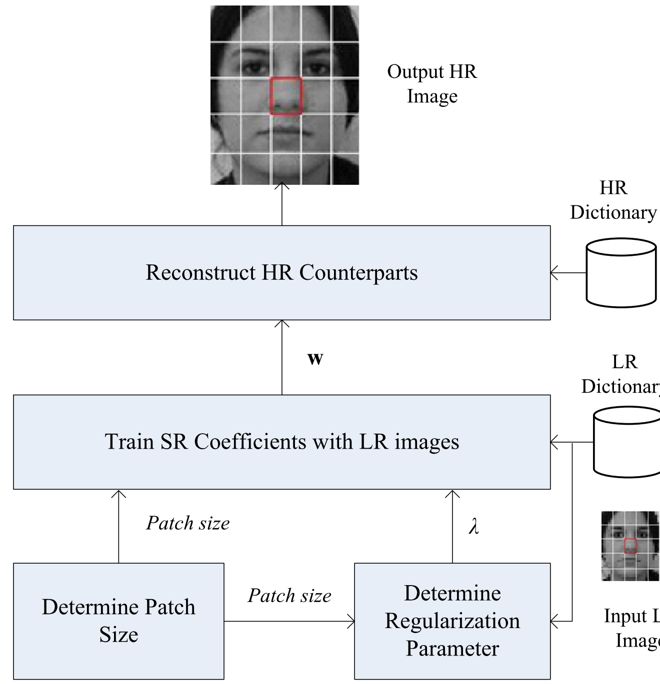

Fig. 1. Outline of the patch-wise face hallucination framework.

2.1. Preliminaries

In dictionary-learning-based face hallucination methods, image patches are represented as a linear combination of elements from an appropriately chosen over-complete dictionary, namely, $\mathbf{x} = \mathbf{Y}\mathbf{w}$. The linear combination coefficients denoted by $\mathbf{w}$ can be learned with the observed LR image over dictionary $\mathbf{Y}$. Particularly, for SR based face hallucination, the optimal coefficients are obtained by solving ℓ_1 minimization problem. To be more precise, let N be the dimension of an input sample (usually an image patch with size $\sqrt{N} \times \sqrt{N}$) and M be the number of basis samples in training set, for a given input sample $\mathbf{x} \in R^{N\times 1}$ and a training set $\mathbf{Y} \in R^{N\times M}$ with each column being an individual training sample, face hallucination via SR can be typically formulated in the following Lagrangian form:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \left\{ \|\mathbf{x} - \mathbf{Y}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1 \right\} \tag{1}$$

where $\mathbf{w} \in R^{M\times 1}$ is an unknown coefficient vector, whose entries w_m, $m = 1, 2, \ldots, M$ are associated with all bases in training set. $\lambda \geq 0$ is an appropriately chosen regularization parameter, controlling the tradeoff between the reconstruction error and the ℓ_1 norm penalty. The optimal coefficient vector $\mathbf{w}^*$ can be readily obtained by solving an SR problem.

The primary task of face hallucination is to reconstruct the HR face image from the input LR face image with the help of a training set composed of HR and LR image patches. Optimal coefficients are trained with the input LR image over LR dictionary. According to the manifold similarity paradigm in LR and HR spaces [14], the learning algorithm then maps the local geometry of LR patch space to an HR one, generating HR patch as a linear combination of HR basis patches. Regularization parameter λ and patch size should be predetermined prior to formal face hallucination. In the next subsections, we will discuss the estimate methods on them.

2.2. Patch size

Patch dimension (or patch size) is a key parameter in patch-based face hallucination. It indicates the number of pixels in a