Signal Processing 142 (2018) 231–243

Contents lists available at ScienceDirect

Signal Processing

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

Learning local dictionaries and similarity structures for single image super-resolution

Kaibing Zhang[a], Jie Li[b], Haijun Wang[c], Xiuping Liu[a], Xinbo Gao[b,*]

[a] College of Electrics and Information, Xi'an Polytechnic University, Xi'an 710048, China
[b] School of Electronic Engineering, Xidian University, Xi'an 710071, China
[c] School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang 471023, China

ARTICLE INFO

Article history:
Received 18 January 2017
Revised 19 April 2017
Accepted 20 July 2017
Available online 25 July 2017

Keywords:
Single image super-resolution (SR)
Local dictionary learning
Non-local similarity
Steering kernel regression
Gradient descent

ABSTRACT

The central task of reconstruction-based single image super-resolution (SR) approaches is to design an effective prior to well pose the solution to unknown up-sampled image. In this paper, we present a novel single image SR method by learning a set of local dictionaries and non-local similar structures from the input low-resolution (LR) image itself. The local dictionaries are learned by segmenting structurally different regions into different clusters and then training an individual dictionary for each cluster. With the learned dictionaries and similar information, each HR pixel in the expected HR image is estimated as the weighted average of a non-local dictionary (NLD)-based regression which assembles the local structural regularity and the non-local similar redundancies. We further transform the proposed NLD-based regression model into a unified regularization term for a *maximum a posteriori probability* (MAP) based SR framework. Thorough experimental results carried out on five publicly available datasets indicate that the proposed SR method is promising in producing high-quality images with finer details and sharper edges in terms of both quantitative and perceptual quality assessments.

1. Introduction

The objective of image super-resolution (SR) technique is to generate a high-resolution (HR) image(s) by using one or several low-resolution (LR) images from the same scene. The SR technique has a wide range of applications such as computer vision, medical and remote sensing imaging, video surveillance, and mobile devices, so it has gained a lot of attention since it was first proposed by Huang and Tsai [1]. In this paper, we mainly focus on single image SR methods. In general, the existing single image SR approaches can be divided into three categories: interpolation-based SR methods [2–5], reconstruction-based SR methods [6–12], and example learning-based SR methods [13–30].

Interpolation-based SR approaches typically utilize a base function or an interpolation kernel to estimate millions of unknown pixels in the HR grids. Despite their efficiency and simplicity, the obtained HR images often show many noticeable blurring and jaggy artifacts. As a result, the performance of this kind of methods is not ready for practical applications.

The second group of methods is referred to as reconstruction-based methods. This group of SR methods explicitly assume that the observed LR image is the outcome of an underlying HR image which is degraded by a series of degradation factors including blurring, downsampling, and noising [6–12]. Because one LR image may correspond to many different HR images, the SR reconstruction is a typical ill-posed problem. To produce a reliable HR estimate, the reconstruction-based SR approaches often need to impose a certain prior knowledge on the resulted image. With a well-designed prior regularization, the reconstruction-based methods are capable of producing pretty good results with sharper edges.

The third kind of SR methods belongs to example learning-based technique. These methods utilize statistical machine learning technique to establish the mapping relationship between the LR and HR images from a prepared training database containing a large number of LR-HR image pairs [13,31,32]. With the LR-HR image pairs as prior, lots of novel details can be effectively hallucinated and added to the LR input for improving the quality of super-resolved image. However, in the case of shortage of relevant exemplars and errors in feature matching, example learning-based SR approaches are clumsy at preserving sharp edges and prone to generate unwanted artifacts in the resultant images.

For single image SR recovery, there is only one LR observation available for reconstruction. Mathematically, the generation process of an LR image from the original HR image can be formulated

* Corresponding author.
E-mail addresses: xbgao.xidian@gmail.com, xbgao@mail.xidian.edu.cn (X. Gao).

http://dx.doi.org/10.1016/j.sigpro.2017.07.020
0165-1684/

as

$$\mathbf{y} = \mathbf{DHX} + \varepsilon, \tag{1}$$

where $\mathbf{X}$ and $\mathbf{y}$ are the HR image and the corresponding LR image, respectively, $\mathbf{H}$ denotes the blurring operation, $\mathbf{D}$ is the downsampling matrix, and $\boldsymbol{\varepsilon}$ is the additive Gaussian noise. Therefore, the task of SR reconstruction is solving the inverse problem in Eq. (1) to estimate an HR image $\mathbf{X}$ by using a given LR image $\mathbf{y}$.

Due to blurring, downsampling, and noising, one LR image may correspond to many different HR images and therefore the SR problem is severely undetermined. To make this problem well-posed, a certain prior (denoted as one or more regularization terms) is needed to incorporate into the reconstruction process. With a regularization prior, a *maximum a posteriori probability* (MAP)-based estimate to the SR problem can be expressed as below, i.e.,

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{y} - \mathbf{DHX}\|_2^2 + \gamma \cdot R(\mathbf{X}), \tag{2}$$

where the first term $\|\mathbf{y} - \mathbf{DHX}\|_2^2$ is the reconstruction error which means that the super-resolved result should be approximate to the LR input via back-projection, the second term $R(\mathbf{X})$ is a regularization term imposing a certain prior knowledge for a more stable SR estimate, and γ is a small positive constant for a tradeoff between the reconstruction error and the regularization term.

The key to success of reconstruction-based single image SR methods highly depends on the reasonability of imposed prior knowledge. To address this problem, in this paper we suggest learning the image structures inside the input LR image for the desired prior regularization term. To be more specific, we first learn a set of local dictionaries by partitioning the given LR image into several sub-regions using the clustering on the local geometric structures, and then a component analysis is applied to seek a supporting basis for the local dictionary-wised regression. To achieve a more robust representation, we further extend the local dictionary-wised regression to a non-local style by exploiting the non-local similarity, aiming at boosting up the performance of local dictionary-wised regression. Finally, we convert the non-local dictionary (NLD)-based regression into a unified regularization term and incorporate it into an MAP-based SR framework for optimization. In summary, the major contributions of this work are the following three aspects:

1) We study the structural regularity about learning a set of local dictionaries from the input LR image itself for a local dictionary-wised regression. The local dictionary-wised regression is able to preserve sharper edges and maintain finer details.
2) We exploit the similar structures to extend the local dictionary-wised regression to an NLD-based regularization term. The new regularization term benefits to well pose the SR problem.
3) Extensive experimental results on several benchmarks indicate that the newly proposed SR scheme can produce state-of-art results with sharper edges and finer details without the use of any external training exemplars which are necessary for most example learning-based methods.

The confluence of the local structure regularity and the non-local similarity priors in a given input LR image benefits high-quality SR result. The similar priors have also been applied in Jiang's method [33], where a two-step SR framework, i.e., the locally regularized anchored neighborhood regression (LANR)-based detail synthesis and the non-local means (NLM)-based quality enhancement, is presented for the SR construction. By contrast, the uniqueness of our method is focusing on how to properly assembling the two priors into a unified regularization term for the reconstruction-based SR model.

The remainder of the paper is organized as follows. Section 2 reviews the previous work that is related to this paper. Section 3 provides a brief overview of the proposed method. In Section 4 we detail the proposed SR method. Section 5 presents the experimental results and assesses the SR performance by comparing with other state-of-the-art SR methods in the literature. Finally, we conclude the paper and discuss the future work in Section 6.

2. Related work

In this work, we apply steering kernel regression (SKR) [34,35], similarity structure learning [36–40], and dictionary learning-based methods [24,25] to construct a powerful regularization term to address the inverse SR problem. The review of closely related work is briefly presented as below.

Recently, non-parametric regression methods have been widely recognized and successfully applied to handle many image restoration tasks such as denoising, upscaling, and interpolation. Representatives are edge-directed interpolation [4], normalized convolution [41], bilateral filter [42], and SKR [35]. These studies strive for an effective way to represent a pixel as the weighted average of its neighbors.

Another important property of natural images is self-similarity which means that small patches tend to redundantly repeat themselves many times within the same scale as well as across different scales. This leads to many successful applications in the image processing domain. Buades et al. [39] proposed simple yet powerful image denoising method by replacing each pixel with a weighted average of its neighbors, where the weights are evaluated by using similar matching between the image patches centered around the center pixel to be estimated and the neighborhood pixels to be averaged. Protter et al. [37] and Protter and Elad [38] utilized the non-local means (NLM) filter [39] for the SR reconstruction of video sequences with general motion patterns, without the need of explicit motion estimation. Another particular method was proposed by Glasner et al. [22]. The method exploits both example learning-based and multi-frame-based methods for SR task, in which similar patches across different scales are extracted from a set of image pyramids built from the input LR image for example learning-based reconstruction while the similar patches within the same scale for multi-frame-based reconstruction. Recently, combining the merits of reconstruction- and example learning-based SR approaches has attracted much attention in the SR literature. For example, Dong et al. [24] presented a unified image restoration framework by using adaptive sparse domain selection and adaptive regularization. Another special assembling method was proposed by Wang et al. [30]. This method utilizes locally regularized anchored neighborhood regression (LRAN) to synthesize novel details and the nonlocal means filter to reduce the artifacts introduced by the LRAN-based outcome, showing the state-of-the-art results.

Both dictionary learning-based and regression-based methods are two important techniques for example learning-based SR approaches. Dictionary learning-based SR approaches are getting popular since it was first proposed by Yang et al. [16], which can adaptively choose the most relevant neighbors to linearly combine the HR image patches for SR reconstruction. A more efficient variation of [16] was proposed in [17], where a jointly learned LR-HR over-complete dictionary pair is used for more efficient sparse coding than [16]. Another single-image scale-up algorithm using sparse representation was proposed by Zeyde et al. [18], which first learns the LR dictionary and then obtains the corresponding HR dictionary with the pseudo-inverse operation. Nevertheless, the sparse coding with ℓ_0-norm or ℓ_1-norm regularized least squares applied the above methods faces with the computational bottleneck. Recently, many efficient example regression-based methods [26–29,43] have received great attention. These methods directly