J. Vis. Commun. Image R. 38 (2016) 175–185

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Cluster-based image super-resolution via jointly low-rank and sparse representation ☆

Ningning Han [a], Zhanjie Song [a], Ying Li [b,*]

[a] School of Science, Tianjin University, 300072, China
[b] State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, 300072, China

ARTICLE INFO

Article history:
Received 2 November 2015
Revised 25 January 2016
Accepted 28 February 2016
Available online 3 March 2016

Keywords:
Image super-resolution
Low-rank representation
Sparse representation
Subspace
Cluster

ABSTRACT

In this paper, we propose a novel algorithm for single image super-resolution by developing a concept of cluster rather than using patch as the basic unit. For the proposed algorithm, all patches are splitted into numerous subspaces, and the optimal representation problem is solved with jointly low-rank and sparse regularization for each subspace. By enforcing global consistency constraint of each subspace with nuclear norm regularization and capturing local linear structure of each patch with ℓ_1-norm regularization, effective matching functions for test and exemplar patches can be created. Accordingly, the desirable results with low computational complexity are obtained. Experimental results show that the proposed algorithm generates high-quality images in comparison with other state-of-the-art methods.

1. Introduction

Single image super-resolution (SISR) is a technology that can be used to generate a plausible and visually pleasing high-resolution (HR) image from a given low-resolution (LR) input image. However, it is a typically ill-posed problem which is dramatically under-constrained due to the insufficient observations. A key issue of SISR is how to exploit the external information or prior knowledge to predict the high-frequency details lost in the LR images.

Various algorithms with different assumptions and recovery criteria have been proposed to solve the SISR problem. According to the use of priors, the generic SISR algorithms can be categorized into four types – prediction models, edge based methods, exploiting image statistical priors, learning methods (also called example-based methods). Besides, deep learning method were applied to SISR by learning an end-to-end mapping between LR and HR patches [1].

The basic prediction model is conventional interpolation methods, of which bilinear, bicubic are most used. HR pixel intensities are generated by weighted neighboring LR pixels, which is simple and effective on smooth regions, but lead to artifacts at high-frequency regions and blurring effect along edges. Given an initial HR image, the IP method [2] iteratively generated a LR image with a predefined downsampling model and effectively compensated the divergent map in LR back to the HR image. With this method, the contrasts along the edges were better enhanced than the results obtained by bicubic interpolation. In [3], the authors presented an alias-free upsampling technique to perform image super-resolution. To improve the accuracy of prediction, He et al. proposed a framework for SISR only using the original LR image and its blurred version, where each pixel was predicted by its neighbors through the Gaussian process regression [4].

A key point in SISR is to predict the high-frequency information. Edge details, known as important primitive image structures, are usually lost in the HR image. Representative works based on edge priors include [5,6]. In [5], authors trained a gradient profile prior based on an intermediate bicubic interpolated images. However, the high-frequency texture details may not be generated due to the use of interpolated image. Similarly, Tai et al. [6] learnt a kernel regression function and utilized a post-processing filter to suppress median gradients cased by image noise, which may lead the corresponding mid-frequency details in the real image to be wrongly reduced.

For natural images, an important property is heavy-tailed distribution when applying derivative filters into the image. Some methods based on statistical priors such as [7–9] exploited image statistical distributions to regularize the LR images. However, the distributed properties are often corrupted by the noise. As a result, it is hard to seek an approximate model for actual distributions. In

☆ This paper has been recommended for acceptance by J. Mukhopadhyay.

* Corresponding author.
E-mail addresses: ning_ninghan@tju.edu.cn (N. Han), zhanjiesong@tju.edu.cn (Z. Song), yingli2011@tju.edu.cn (Y. Li).

http://dx.doi.org/10.1016/j.jvcir.2016.02.015
1047-3203/

addition, some regularization terms such as total variation [10], autoregressive model and non-local self-similarity [11], and group-sparse [12] have been used for generating HR image. Though these regularization terms are simple and effective, lacking external information to predict the rare details in the HR image is the disadvantage.

Exemplar images containing abundant visual information can be used to enrich the fine details of LR image. The learning-based approaches, assuming the high-frequency details lost in the image can be learnt from training sets, have been attracting considerable interest recently. In [13], Freeman et al. proposed an example-based super resolution method where the LR to HR prediction was learned via a Markov random field (MRF). In [14], the high-frequency details of the underlying LR image were estimated by kernel ridge regression (KRR). In addition to pixel values of patches, representing images in transferred domain, such as wavelet coefficients [15] or image contourlet [16] have also been developed. Jiji et al. [15] used wavelet transform to decompose the observed image as well as the training images. Wavelet coefficients of the super resolved image were learned from the coefficients of images in the database. The HR image was estimated under a MAP frame work using the learned wavelet prior. The authors also suggested reconstructing the HR image using the contourlet transform [16]. Two recently typical examples are neighbor embedding (NE) methods [17] and dictionary learning methods. NE were inspired by the locally linear embedding, under the assumption of manifold local similar geometries, and the HR image patches were constructed by linearly combining the corresponding HR counterparts. In [18], the authors introduced a new manifold learning method and explained its relationship with SISR. Considering that NE was not always true for complicated texture structures, Zhang et al. [19] proposed an unsupervised Gaussian mixture model and used a supervised neighbor embedding to estimate HR patches. Since the NE carried out two independent processes to synthesize HR patches, the separate processes were not optimal. Sparse neighbor embedding with predetermined neighbor and robust-SL0 algorithm was suggested in [20]. Noted that similar patches spanned low-rank structures, a novel single image super-resolution method based on the low-rank matrix recovery (LRMR) and NE were proposed in [21].

A different line of NE is sparse representation or dictionary learning. Researches on image statistics suggest that image patches can be well represented as a sparse linear combination of elements from a given over-complete dictionary. As an initial investigation, Yang et al. [22] employed sparse coding methods to perform image super-resolution. The basis of this approach is that each pair of HR and LR patches have a pre-specified correspondence. As the extension, some recent works [23–26] boosted [22] in quality and speed by changing the strategies of learning of the coupled dictionaries and characterizing the relationship between LR and HR patch spaces. Purkait et al. [27] partitioned the natural images into documents and grouped them to discover the inherent topics using probabilistic Latent Semantic Analysis (pLSA). They found the dual dictionaries of HR and LR image patch pairs for each of the topics and incorporated multiple dictionaries for a more accurate prediction rather than a single sparse dictionary. In [28,29], Radu et al. supported the use of sparse learned dictionaries in combination with neighbor embedding method and simple functions. They found the nearest neighbor based on the correlation with dictionary atoms rather than the traditional Euclidean distance. To avoid invariance assumption, which was a common manipulation in sparse representation, Peleg and Elad [30] used a statistical model by means of MMSE estimation and a feedforward neural network to obtain HR patches. Similarly, Zhu et al. [31] proposed the concept of deformation, which regarded the patch as a flexible deformation flow rather than a fixed vector for SISR. By this means, the dictionary is more expressive than traditional methods. Taking advantage of the beta process factor analysis, a series of beta process joint dictionary learning approaches such as [32–34] were proposed. In this way, the sparse representation can be decomposed to magnitudes and dictionary atom indicators so that the mapping between coupled feature spaces are consistent and accurate. Another innovative method was to cluster the LR patch space and learn a separate mapping from LR to HR space for each cluster, as proposed by Yang and Yang [35].

The recent evaluation of representative SISR techniques demonstrates the powerful ability, but there are still considerable challenges. As a powerful tool, the sparse representation approaches demonstrate perfect capabilities in SISR, however, they find the sparsest representation of each data individually, lacking global constraints on the solution space. To capture the global subspace structures, Liu et al. [36] proposed a method termed low-rank representation (LRR). In general, LRR jointly represents all the data under a global low-rank constraint, so it is better at capturing the global data structures (e.g., multiple clusters or subspaces). It has been proven that LRR can preserve exactly the membership of the samples belonging to the same subspace under mild conditions. In addition, the traditional patch based strategies recover each patch independently, ignoring the relationship among patches in the proposed image. The image structures tend to redundantly repeat themselves within and across different scales. If one divides patch spaces into different clusters, similar patches in the same cluster can form a low-dimensional subspace. It is necessary to further explore the underlying low-rank structure. As known, the training sets are relatively redundant, a trained dictionary or predicted model may be only effective for a given image, one has to train different dictionaries again. The topic that how to develop an online algorithm eliminating irrelevant data and recovering images in accordance of their characteristics is worth studying. Inspired by the LRR and structural similarity, we address these challenges in a unified framework, which implements a divide-and-conquer approach by splitting the test and training LR patches into different clusters and recovering the HR image in the domain of cluster. In specific, we collect exemplar and test patches lying on manifolds in the same space to learn an anchored regression by solving jointly a low-rank and sparse representation problem (for the resulting matrix please refer to Fig. 1(a)). In this way, the global mixture structure of each cluster and local linear structure of each patch can be exploited simultaneously in a unified framework so that the high-quality HR images are generated by developing exact mapping for the test and training images.

This paper is organized as follows. In Section 2, we motivate the need for recovering the HR image in the field of cluster with ℓ_1-norm and nuclear norm. Section 3 presents the relevant model and the iterative optimization algorithm for solving the model. Subsequently, to demonstrate its advantages over other methods, the performance of our method is evaluated in Section 4. Finally, Section 5 gives the conclusion.

2. Motivation

In this section, we provide motivation for recovering the HR image in the domain of cluster with a jointly low-rank and spare representation problem. We consider a set of 300 similar patches of size 8×8 that are extracted out of image *Lena*. For each patch, we perform a preliminary step of DC removal by subtracting the average magnitude of the patch and use the OMP algorithm to calculate the sparse representations of these patches over an trained dictionary of size 64×256. These representations are stacked into a matrix of size 256×300. Fig. 2(a) and (b) shows the image of the matrix and the singular value distributions obtained by singular