Computers and Electrical Engineering 000 (2017) 1–12

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Single image super resolution using neighbor embedding and statistical prediction model☆

V. Abdu Rahiman[1,*], Sudhish N. George

Department of Electronics and Communication Engineering, National Institute of Technology Calicut, India

ARTICLE INFO

Article history:
Received 4 September 2016
Revised 18 December 2016
Accepted 21 December 2016
Available online xxx

Keywords:
Single image super resolution
Sparse representation
Sparse representation invariance
Neighbor embedding
Statistical prediction model

ABSTRACT

This paper proposes learning based approaches for single image super-resolution using sparse representation and neighbor embedding. Two learning based methods are proposed to recover the high-resolution (HR) image patches from the low resolution (LR) patches. The first method, named as LeNm-SRI, is a computationally efficient approach using neighbor embedding in a partitioned feature space. In this method, the training set is updated by including details extracted from different scales of LR input image. LeNm-SRI, which uses sparse representation invariance, gives acceptable results at low computational load. In the second approach, named as LeNm-RBM, a statistical prediction model is used to predict HR feature coefficients to obtain increased performance. Separate prediction models are trained for each cluster, and the model parameters are updated with each input image, to adapt to input test image. Experimental results validate the computational efficiency and performance of the proposed methods.

1. Introduction

The high-resolution (HR) image contains more details than the low resolution (LR) counterpart. HR image can be captured by using HR imaging sensors. HR imaging sensors are fabricated with a large number of sensing elements (pixels) per unit area. As the dimension of sensing element decreases, the amount of light falling on sensor element will also decrease which results in a low signal to noise ratio. Therefore highly sensitive and low noise sensors are needed which increases the cost of high-resolution image sensors [1]. An alternate approach is to use signal processing techniques to reconstruct HR images from low-resolution observations. This process of reconstructing HR images from LR observations through signal processing is called Super Resolution(SR). LR images have fewer details than HR images. Thus SR algorithms will regenerate the HR images by adding these missing details [2]. A simplified imaging model is formulated as

$$\mathbf{x} = \mathbf{DBMz} + \eta \tag{1}$$

where $\mathbf{x}$ is the LR observation, $\mathbf{z}$ is the HR image, $\mathbf{D}$ is the sampling matrix, $\mathbf{B}$ is blurring model, $\mathbf{M}$ represents the warping and η is the additive noise. Image super-resolution is the inverse problem of estimating the HR image $\mathbf{z}$ from LR observations. SR algorithms are broadly classified into single frame super-resolution and multi-frame super resolution. Multi-frame

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. E. Cabal-Yepez.

* Corresponding author.
E-mail addresses: vkarahim@gmail.com (V. Abdu Rahiman), sudhish@nitc.ac.in (S. N. George).
[1] Currently working as Assistant Professor in Electronics and Communication Engineering at Govt Engineering College Kozhikode, India

http://dx.doi.org/10.1016/j.compeleceng.2016.12.018
0045-7906/

SR algorithm utilizes multiple LR views of the same scene, and these images are combined to form HR image. Multiple LR observations may have sub-pixel shifts between each other. So the additional details in each observation are merged. Many real world applications, like high definition displays, require reconstruction of HR image from a single observation. Therefore single observation based super-resolution methods became a hot topic of research in recent years. In single frame SR algorithm, also known as single image super-resolution (SISR), HR image is estimated from a single LR observation. Broad categories of SISR are interpolation based, reconstruction based and learning based methods [3]. In interpolation-based approach, the LR image is interpolated using any image interpolation method followed by edge-directed methods or sub-pixel alignment techniques to refine the result. Reconstruction based method uses the LR patches to compute the SR image. It is done by aligning similar LR patches from the LR image. This technique exploits the self-similarity between the patches in LR image. Even if this method is superior in multi-frame image super-resolution, in single frame image super-resolution, this approach suffers from ill-conditioned image registration due to the fewer number of observations [4].

Learning based method for single image SR uses training images to learn the relation between LR and corresponding HR images. Prior information, learned from these training examples are used to recover the HR image from LR observations. Proper selection of training images results in good SR images. The performance of this approach also depends on the model used to learn the relationship between LR and HR patches in the training set. Recently a paradigm shift happened in the field of image processing with the advent of compressive sensing and sparse signal representation. LR observations are modeled as sparse signals and super-resolution can be formulated as a compressive sensing problem [5,6]. Various image processing algorithms are tried in the sparse domain, and promising results are reported in areas like image de-noising, image enhancement, image super-resolution, etc. [7,8].

Yang et al. [9] were the pioneers to attempt super-resolution using sparse representation. They proposed a learning based SR algorithm which needs a collection of training images from which sufficiently large set of HR and corresponding LR patches are chosen. A set of over-complete dictionaries is jointly learned from these HR and LR patches. During testing/reconstruction, LR test image is divided into overlapping patches and sparse representation coefficients are found using LR dictionary. HR patches are synthesized using these coefficients and the HR dictionary. In [9], the authors used joint dictionary learning method and later the dictionary learning algorithm has been modified to coupled dictionary learning [10] which results in better SR performance in less execution time. The patch wise operation causes blocking artifacts, and it is minimized by applying iterative back projection on the super-resolved image.

Selection of training image plays a vital role in the success of training set based SR algorithms. But it is obvious that small patches of an image may have similarity across different scales of the same image. The similarity of such patches can be exploited to improve the performance of image super-resolution algorithms. Yang and Wang [11] proposed a method for SISR by utilizing the similarity between the patches of different scales of the same image. The input image is divided into overlapping patches, and these patches are categorized based on texture/smoothness of the patch. The error between interpolated patch and HR patch, in all scales of the image, are represented by a sparse model. In [11], a Support Vector Regression (SVR) model is learned from the sparse representations. The input image is interpolated, and overlapping patches are categorized based on the texture of patch. Now the error between LR and HR patch is synthesized using the error estimated by sparse recovery. Though this method utilizes similarity between different scales of the same image, it is likely that certain details which are not present in the patches from test image may be present in the HR image. These details may not be recovered if patches from the various scales of the test image are only used for training.

A modified dictionary based SR method is proposed in [12] which performs sparse representation based SR on salient patches. In this approach, different scales of LR test image is used to update the dictionary. Salient patches are determined based on the local variance of patches. Even though this method combines the benefits of self-similarity and sparse representation, it is necessary to re-train the dictionary for each test image which increases the time required for super resolution. In all the methods discussed above, an HR patch is recovered as the linear combination of most or all atoms in the dictionary. The HR patches could be better recovered if they are represented using a fewer number of closely matching patches. Chang et al. [13] proposed a method for SISR which uses the locally linear embedding (LLE) of neighboring manifolds. In their method, neighboring patches are used to represent the patches of LR image and the same coefficients, together with corresponding HR patches are then used to recover the HR image. Robust position patch-based approaches are reported in the literature for face image super-resolution [14,15]. But unlike face images, the similarity between the patches at the same position of different images is not guaranteed in natural images. So position patch-based methods are not suitable for super resolving natural images.

The SISR methods described above use sparse representation invariance assumption for recovery of SR image. Here joint dictionary learning is employed during training phase and the LR sparse coefficients are used as such with HR dictionary to recover HR image. Another direction of research is on finding a mapping from the LR coefficients $\boldsymbol{\alpha}_l$ to HR coefficients $\boldsymbol{\alpha}_h$. Wang et al. [16] suggests a linear mapping between $\boldsymbol{\alpha}_l$ and $\boldsymbol{\alpha}_h$. But all these methods use equal number of atoms in over complete LR and HR dictionaries. Therefore the low information content in LR patches may cause over-fitting with higher number of atoms in LR dictionary. Peleg et al. [17] proposed an SISR method which uses Restricted Boltzmann Machine(RBM) for predicting $\boldsymbol{\alpha}_h$ from $\boldsymbol{\alpha}_l$. The predicted value of $\boldsymbol{\alpha}_h$ is denoted as $\hat{\alpha_h}$. In this method, an under complete orthonormal dictionary is used for LR patch and an over complete dictionary is learned for HR patch. The MMSE estimator used for predicting jth coefficient of α_h is given as

$$\hat{\alpha}_{h,j} = \mathbf{c}_{hl,j}^T \alpha_l \Phi\left(b_{h,j} + \mathbf{w}_{hl,j}^T \mathbf{s}_l\right) \tag{2}$$

Please cite this article as: V. Abdu Rahiman, S. N. George, Single image super resolution using neighbor embedding and statistical prediction model, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2016.12.018