ARTICLE IN PRESS

Optics Communications ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Optics Communications

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

Optimized multiple linear mappings for single image super-resolution

Kaibing Zhang [a], Jie Li [b], Zenggang Xiong [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 Computer and Information Science, Hubei Engineering University, Xiaogan 432000, China

A R T I C L E I N F O

Keywords:
Example learning
EM-algorithm
Image super-resolution (SR)
Multiple linear mappings

A B S T R A C T

Learning piecewise linear regression has been recognized as an effective way for example learning-based single image super-resolution (SR) in literature. In this paper, we employ an expectation–maximization (EM) algorithm to further improve the SR performance of our previous multiple linear mappings (MLM) based SR method. In the training stage, the proposed method starts with a set of linear regressors obtained by the MLM-based method, and then jointly optimizes the clustering results and the low- and high-resolution subdictionary pairs for regression functions by using the metric of the reconstruction errors. In the test stage, we select the optimal regressor for SR reconstruction by accumulating the reconstruction errors of m-nearest neighbors in the training set. Thorough experimental results carried on six publicly available datasets demonstrate that the proposed SR method can yield high-quality images with finer details and sharper edges in terms of both quantitative and perceptual image quality assessments.

1. Introduction

The objective of single image super-resolution (SR) is to produce a high-resolution (HR) image by using only one low-resolution (LR) image as input [1]. The SR technique has gained great attention in literature due to its capacity of generating a high-quality image with fine details that a low-cost imaging system cannot directly obtain. Roughly speaking, the existing single image SR approaches can be divided into three categories: interpolation-based SR methods, reconstruction-based SR methods, and example learning-based SR methods.

Interpolation-based SR methods [2–5] typically employ a fixed base function or an analytical interpolation kernel to estimate millions of the unknown pixels in the HR grids. This group of methods is simple but efficient, and ready for real-time applications. However, most approaches tend to generate noticeable artifacts along salient edges and blurring high frequency details in texture areas, leading to unacceptable perceptual quality for many applications.

The second group of SR techniques is referred to as reconstruction-based techniques, which strive for integrating a certain priori knowledge (represented as one or more regularization terms) into the process of SR reconstruction so as to obtain a stable solution [6–12]. Usually, the edge-directed priors such as the edge prior [6,7], the gradient profile prior [8,9], and the total variation [10], are popularly used for maintaining sharp edges. Another effective prior that exploits the self-similarity redundancy inside the input LR image itself [11] and [12], shows promising SR capacity under the reconstruction-based framework. While the reconstruction-based approaches are effective at producing sharp edges and suppressing annoying artifacts, they are still cumbersome for adding novel details, especially in the case of a large magnification factor (e.g., larger than ×2).

Example learning-based SR methods typically use a set of LR–HR image pairs as prior to establish the mapping relationship between the LR and HR images. In terms of how the input, output, and mapping relationship are established, the representative learning models can be further divided into four subclasses, i.e., k-nearest neighbor (k-NN) learning-based [13,14], manifold learning-based [15,16], dictionary learning-based [17–19], and regression-based SR methods [20–28].

The seminal k-NN learning-based SR method was proposed by Freeman et al. [13], where the belief propagation algorithm is applied to train a Markov network for image SR reconstruction. Sun et al. [14] followed up Freeman's method by introducing the primal sketch prior to alleviate the blurred effects on details of edges, ridges, and corners. Although the above methods can add lots of novel details into an input LR image, one of the most challenging problems is their intensive computational cost. Chang et al. [15] proposed neighbor embedding (NE)-based learning for SR, assuming that the two manifolds of the LR

* Corresponding author.
E-mail address: xbgao@mail.xidian.edu.cn (X. Gao).

http://dx.doi.org/10.1016/j.optcom.2017.06.102
Received 14 February 2017; Received in revised form 22 June 2017; Accepted 30 June 2017
Available online xxxx

Please cite this article in press as: K. Zhang, et al., Optimized multiple linear mappings for single image super-resolution, Optics Communications (2017), http://dx.doi.org/10.1016/j.optcom.2017.06.102.

feature space and the corresponding HR feature space are locally in similar local geometries. This method, to a certain degree, breaks the limitation of huge samples suffered in [13] and [14], and therefore can achieve better reconstruction quality with a relatively smaller number of training examples. The k-NN and NE-based SR algorithms need to search a huge reference dataset for similar patterns in order to optimally represent complicated structures in generic images, so the SR lacks the efficiency for practical applications.

Dictionary-based SR approaches assume that a nature image patch admits a sparse representation over an over-complete dictionary. Based on this assumption, the HR image patches are synthesized by adaptively choosing the most relevant atoms for SR reconstruction with the sparsity regularity. The pioneering work was proposed by Yang et al. [18], where an external database composed of related LR–HR patches is used to jointly learn a compact dictionary pair by enforcing L_1-norm regularity prior. Under the similar framework to Yang's method, Zeyde et al. [19] improved the SR efficiency using PCA-based dimensionality reduction and Orthogonal Matching Pursuit (OMP) [29] for sparse coding. Although dictionary learning-based methods have notable advantages of reconstruction quality and memory allocation, the main weakness is their intensive computational cost of sparsity decomposition at both the training and the inference stages.

Currently, learning piecewise linear regression for single image SR has received much attention in the SR research domain, aiming at improving efficiency while maintaining quality. The main idea behind this particular SR technique is to learn a set of linear regressors from the manifold of LR patches to that of the HR patches with the manifold assumption. For example, Timofte et al. [20] proposed anchored neighbor regression (ANR) to alleviate the bottleneck of sparse coding for efficient SR reconstruction. They further extended the ANR to A+ [21] by learning the piecewise linear regressors from the training set of the local correlative neighbors of the anchored atoms. Yang et al. [22] advocated learning a set of simple mapping functions from numerous image subspaces using multivariate linear regression. Other similar follow-ups of regression-based SR approaches can be found in [23] and [24]. In addition, a deep learning-based method which uses a deep convolutional neural networks (CNN) to model the mapping relationship between the LR and HR images was presented in [25], showing promising SR performance. In [26], multiple linear mappings (MLM) are built for efficient SR by learning a set of orthogonal LR subdictionary and the inferred HR subdictionary with the assumption that the LR–HR features share the same representation.

In this paper, we extend to further improve the SR performance of our previous MLM-based SR approach by following up a jointly learning viewpoint proposed in [21]. The main motivation behind this work is that in the MLM-based method, the partition of feature space and the learning of mapping relationship are separately conducted, so the learned regression functions may be not optimal for the whole training dataset. Inspired by this motivation, in the training stage we start with the linear regressors obtained by the MLM-based method, and then use an expectation–maximization (EM) algorithm to jointly optimize the clustering results and the low- and high-resolution subdictionary pairs in terms of the metric of reconstruction errors. In the test stage, the optimal regressor is chosen by accumulating the reconstruction errors of m-nearest neighbors in the training set, which enables to alleviate mismatching when the input is close to the border error of clusters.

The remainder of the paper is organized as follows. In Section 2 we detail the proposed SR method. Section 3 presents the experimental results and assesses the SR performance by comparing with other state-of-the-art SR methods in literature. Finally, we conclude the paper and discuss the future research direction in Section 4.

2. The proposed SR method

In this section, we begin with the introduction of our previous MLM-based SR method [26] and then detail the optimized MLM (OMLM)-based method based on an EM algorithm [28]. Finally, we outline the proposed SR algorithm in the test stage.

2.1. Multiple linear mappings revisited

The main idea behind the MLM-based SR approach is to employ a set of simple linear regressors for an efficient SR reconstruction. In the learning stage, a large number of LR–HR image pairs are collected to generate a training dataset, which is comprised of two feature spaces: the LR feature space $\mathbf{X}_s = \{\mathbf{x}_s^i\}_{i=1}^{N_s}$ and the corresponding HR feature space $\mathbf{Y}_s = \{\mathbf{y}_s^i\}_{i=1}^{N_s}$, where $\mathbf{x}_s^i \in \mathrm{R}^{d_1}$ is the feature vector representing the ith LR image patch, $\mathbf{y}_s^i \in \mathrm{R}^{d_2}$ is its counterpart representing the ith HR image patch, and N_s is the number of LR–HR image pairs. To approximate the complicated nonlinear structure of the feature space spanned by a large number of images, the standard k-means clustering algorithm [30] is applied to divided the feature space of training exemplars into K coupled LR–HR feature subspaces $\{\mathbf{X}_s^k, \mathbf{Y}_s^k\}_{k=1}^{K}$, where $\mathbf{X}_s^k = \{\mathbf{x}_s^i\}_{i\in\Omega_k}$ is the kth subspace of $\mathbf{X}_s$, $\mathbf{Y}_s^k = \{y_s^i\}_{i\in\Omega_k}$ is the kth subspace of $\mathbf{Y}_s$ using the same indices as $\mathbf{X}_s^k$, and Ω_k stands for the specified index set of $\mathbf{X}_s^k$. With the obtained K coupled LR–HR feature subspaces $\{\mathbf{X}_s^k, \mathbf{Y}_s^k\}_{k=1}^{K}$, the problem of finding an optimal LR–HR subdictionary pair which shares the same representation coefficients is to minimize the data-cost function as:

$$\begin{aligned} \mathbf{B}_l^k, \{\mathbf{a}_i\} &= \arg\min_{\mathbf{B}_l^k, \{\mathbf{a}_i\}} \sum_{i\in\Omega_k} \left\| \mathbf{x}_s^i - \mathbf{B}_l^k \mathbf{a}_i \right\|_2^2, \\ \mathbf{B}_h^k, \{\mathbf{a}_i\} &= \arg\min_{\mathbf{B}_h^k, \{\mathbf{a}_i\}} \sum_{i\in\Omega_k} \left\| \mathbf{y}_s^i - \mathbf{B}_h^k \mathbf{a}_i \right\|_2^2, \end{aligned} \tag{1}$$

where $\mathbf{B}_l^k$ is the kth LR subdictionary that best represents all the feature vectors in $\mathbf{X}_s^k$, $\mathbf{B}_h^k$ is the kth HR subdictionary that best represents all the feature vectors in $\mathbf{Y}_s^k$, and $\mathbf{a}_i$ denotes their shared coefficient vector for linearly combining the LR dictionary atoms to represent $\mathbf{x}_s^i$ and the HR dictionary atoms to represent $\mathbf{y}_s^i$.

In [26], the learning structure of the LR subdictionary is imposed by an orthonormal constraint, which can be formulated as below:

$$\begin{aligned} \mathbf{B}_l^k &= \arg\min_{\mathbf{B}_l^k} \sum_{i\in\Omega_k} \left\| \mathbf{x}_s^i - \mathbf{B}_l^k {\mathbf{B}_l^k}^T \mathbf{x}_s^i \right\|_2^2 \\ &= \arg\min_{B_l^k} \left\| \mathbf{X}_s^k - \mathbf{B}_l^k \mathbf{B}_q {l^k}^T \mathbf{X}_s^k \right\|_F^2 \\ &\quad \text{s.t. } \mathbf{B}_l^k {\mathbf{B}_l^k}^T = \mathbf{I}, \end{aligned} \tag{2}$$

where $\mathbf{X}_s^k$ is the data matrix in which each column is a vector from the subspace $\{\mathbf{x}_s^i\}_{i\in\Omega_k}$, $\|\cdot\|_F$ denotes the Frobenius norm for matrices, and $\mathbf{I}$ is an m-by- m identity matrix. With the same representation coefficients, the HR subdictionary is directly inferred by minimizing the least-squares error as below:

$$\begin{aligned} \mathbf{B}_h^k &= \arg\min_{\mathbf{B}_h^k} \sum_{i\in\Omega_k} \left\| \mathbf{y}_s^i - \mathbf{B}_h^k \mathbf{a}_i \right\|_2^2 \\ &= \arg\min_{\mathbf{B}_h^k} \left\| \mathbf{Y}_s^k - \mathbf{B}_h^k \mathbf{A}^k \right\|_F, \end{aligned} \tag{3}$$

where $\mathbf{Y}_s^k$ is the data matrix in which each column is a vector from all the feature vectors in the kth subspace $\mathbf{Y}_s^k$, and $\mathbf{A}^k$ is the coefficient matrix that contains $\{\mathbf{a}_i\}_{i\in\Omega_k}$ as its columns. The optimization in Eq. (3) can be easily solved by the least squares as:

$$\mathbf{B}_h^k = \mathbf{Y}_s^k \mathbf{A}^{kT} \left(\mathbf{A}^k {\mathbf{A}^k}^T \right)^{-1}. \tag{4}$$

Finally the mapping matrix corresponding to the kth coupled LR–HR subspace can be calculated by

$$\mathbf{F}_k = \mathbf{B}_h^k \left({\mathbf{B}_h^k}^T \mathbf{B}_h^k + \lambda \mathbf{I} \right)^{-1} {\mathbf{B}_l^k}^T. \tag{5}$$

2.2. Optimized multiple linear mappings for SR

In this subsection, we start from the mapping functions obtained from the MLM-based method and follow the spirit of [28] to improve the performance of the MLM-based SR method by jointly learning the

Please cite this article in press as: K. Zhang, et al., Optimized multiple linear mappings for single image super-resolution, Optics Communications (2017), http://dx.doi.org/10.1016/j.optcom.2017.06.102.