Neurocomputing 212 (2016) 3–11

Contents lists available at ScienceDirect

Neurocomputing

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

Single image super-resolution via blind blurring estimation and dictionary learning

Xiaole Zhao [a], Yadong Wu [a,*], Jinsha Tian [a], Hongying Zhang [b]

[a] School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, China
[b] School of Information Engineering, Southwest University of Science and Technology, Mianyang, China

ARTICLE INFO

Article history:
Received 6 November 2015
Received in revised form
2 February 2016
Accepted 15 February 2016
Available online 28 June 2016

Keywords:
Super Resolution (SR)
Non-Local Self-Similarity (NLSS)
Blind Blurring Estimation (BBE)
Dictionary learning

ABSTRACT

Learning-based methods have been becoming the mainstream of single image super resolution (SR) technologies. This kind of methods makes it effective to generate a high resolution (HR) image from a single low resolution (LR) image. There exists, however, two main problems with these methods: the quality of training data and the computational demand. We propose a novel framework for single image SR tasks in this paper, which consists of blur kernel estimation (BKE) and dictionary learning. BKE is utilized for improving the quality of training samples and realized by minimizing the dissimilarity between cross-scale patches iteratively. Couple dictionaries are trained by improved training samples before sparse recovery. More important is that a selective patch processing (SPP) strategy is adopted in BKE and sparse recovery, which brings more accurate BKE results and immensely reduces time consumption of the entire process. The experiments show that the proposed method produces more precise BKE estimation and better SR recovery than several typical SR algorithms at a higher efficiency.

1. Introduction

Image super-resolution (SR) is a cluster of technologies recovering a super-resolved image from a single image or a sequence of images of the same scene, which is a basic operation of many subsequent image manipulation (such as feature extraction [1] and image fusion [2]). In many practical applications, however, it is not easy to obtain an adequate number of LR observations. Therefore, single image super-resolution has attracted great attentions in recent years.

Machine learning based methods are promising technologies for SR problem, and it has become the most popular topic in single image SR field. Freeman et al. [3] proposed example-based learning method firstly. The algorithm predicted HR patches from LR patches by solving Markov Random Field (MRF) model via belief-propagation algorithm. Then, Sun et al. [4] enhanced discontinuous features such as edges and corners etc. by primal sketch priors, which extended example-based method further. To improve execution efficiency, Chang et al. [5] proposed nearest neighbor embedding (NNE) method motivated by the philosophy of locally linear embedding (LLE) [6]. They assumed LR patches and HR patches have similar space structure. The coefficients of LR patch can be solved through least square problem, which are then applied to HR patches directly. NNE utilizes a few of training data to represent a test sample and reduces the computation time dramatically. However, fixed number of NNs may cause over-fitting and/or under-fitting phenomenon [7] which is similar to other typical learning issues (e.g. Boosting [8]). Aiming at this problem, Yang et al. [9] proposed an effective method based on sparse representation and compress sense theory, which selects the number of NNs adaptively and solved the fitting problem effectively.

There exist two other issues about the original sparse representation methods: the compatibility between training samples and testing samples, and the mapping relation between LR feature space and HR feature space. Firstly, patches sampled from the same image generally have higher compatibility because of the same illumination and device parameters etc. Glasner et al. [10] exploited image patch non-local self-similarity (NLSS) within and cross image scale for single image SR tasks, which makes an effective solution for the compatibility problem between training and testing samples. Secondly, patch-based methods usually need to extract LR/HR features from training samples. The relation between these training features were learned via certain methods and then reflected in testing samples. But LR patch space and HR patch space, actually, are tied by some mapping function, which could be unknown, complicated and not necessarily linear [11]. The direct mapping, e.g. [9,26], may not reflect this unknown relation correctly. Yang et al. [12] proposed another joint dictionary training approach to learn the duality relation between LR/HR patch spaces. The method essentially concatenated the two patch/feature spaces and converted the

* Corresponding author.
E-mail address: wyd028@163.com (Y. Wu).

http://dx.doi.org/10.1016/j.neucom.2016.02.079
0925-2312/

problem to the standard sparse representation in a single feature space. Further, they explicitly learned the sparse coding problem across different feature spaces, which is so-called coupled dictionary learning (CDL) [11]. He et al. [13] proposed another appealing beta process joint dictionary learning (BPJDL) for CDL based on a Bayesian method using a beta process prior. The above-mentioned approaches mainly focus on learning unknown relation between LR/HR feature spaces more accurately. When concentrating on single image SR task, these methods still need to take the characteristic of training samples (e.g. blurring and noise etc.) into account for better execution and performance.

The most of current SR technologies typically assume that the blur kernel is known under the condition of single input image (non-blind). For example, low-pass filters (LPF) such as a Gaussian or a bicubic kernel is usually used to replace the unknown blur kernel or the point spread function (PSF) of the imaging device. But Michaeli et al. [14] pointed out that the performance of SR methods significantly deteriorates when the assumed blur kernel deviates from the true one. Obviously, imprecise blur kernel will lead to the low-quality training data and further affect SR results. Another nonparametric blind super resolution technology was proposed by Wen Ze Shao and Michael Elad [25] based on the philosophy of optimizing the estimated blur kernel and an intermediate super-resolved image jointly, which is a bi-L_0-L_2-norm regularization optimization problem. But it is seriously time-consuming as solving L_0/L_1 norm constrained optimization problems during SR recovery.

Our primary observation is two-fold: 1). BKE is a process of solving a convolution kernel. But the convolutional influence is small in smooth area. See Fig. 1(a); 2). Natural images generally consist of continuity and discontinuity patterns. The existing methods, which aimed to resolve the discontinuous patterns of natural images, cannot outperform traditional interpolation technologies within smooth area. See Fig. 1(b). In this paper, we present a novel single image SR method according to our previous observation. The proposed approach estimated the true blur kernel based on minimizing the dissimilarity between cross-scale patches firstly. Then, LR/HR coupled dictionaries were trained through input image downsampled by estimated blur kernel, which improved the quality of training samples. The more important is that a SPP strategy measured by average gradient amplitude [15] is employed in both BKE and SR recovery. As shown by our experiments, SPP not only reduced the time consumption drastically, but also improved the quality of BKE and SR recovery.

Our paper is structured as follows. In Section 2 we focus on previous works that are closely related to the proposed method. In Section 3 we introduce our improved blur kernel estimation technique and SR recovery method. The experiments are conducted in Section 4 with some discussions. Section 5 concludes the paper.

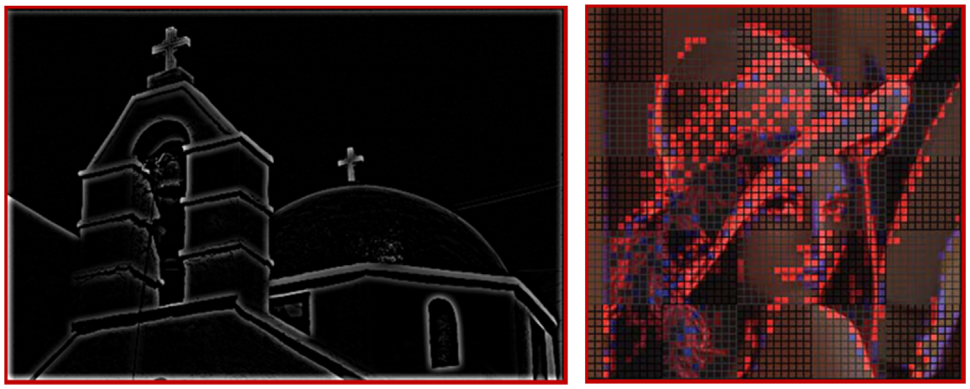
2. Related work

2.1. Internal statistics of a single natural image

One of the most useful internal statistical attributes of natural image patches is the patch recurrence within and cross scale, which is exploited by Glasner et al. in 2009 [10]. Image patch recurrence is also known as image patch redundancy or non-local self-similarity (NLSS). This means a small image patch tends to recur many times no matter within or cross different scales. NLSS in natural images can be exploited as useful prior knowledge for various image restoration tasks, such as super resolution [10,15–18], image denoising [19], deblurring [20] and inpainting [21] etc. However, it is not necessary to consider the redundancy of all patches most of the time according to our previous observations. Zontak et al. [15] further quantified the property by relating it to the spatial distance from the patch and the mean gradient magnitude of the patch. The three main conclusions can be perceived according to [15]: (1) smooth patches recur very frequently, whereas highly structured patches recur much less frequently; (2) a small patch tends to recur densely in its vicinity and the frequency of recurrence decays rapidly as the distance from the patch increases; (3) patches of different gradient content need to search for nearest neighbors at different distances. These conclusions are very useful when we employ discriminatory processing strategies for different patches.

2.2. Cross-scale blur kernel estimation

Michaeli et al. [14] utilized the NLSS property to estimate the optimal blur kernel and cross-scale patch redundancy was maximized iteratively. More specifically, the initial kernel was assumed to be a delta function and used to down sample the input image. For each small patch in input image, they found a few NNs in down-sampled version of input image. Each NN for the small patch corresponds to a large patch in input image, and all NNs for small query patch constitute a set of LR/HR patch pairs. All the patch pairs construct a set of linear equations which can be solved



(a) (b)

Fig. 1. Observations on selective patch processing. (a) The residual image of "Tower" generated by the original image subtracting blurred image which obtained by convolution with a motion kernel. (b) The different SR results performed by bicubic interpolation and sparse recovery (**from** [8]). Red color denotes the patches where sparse recovery beats bicubic interpolation. Blue color represents patches where bicubic interpolation is superior. Gray color indicates that the two perform on par with each other. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

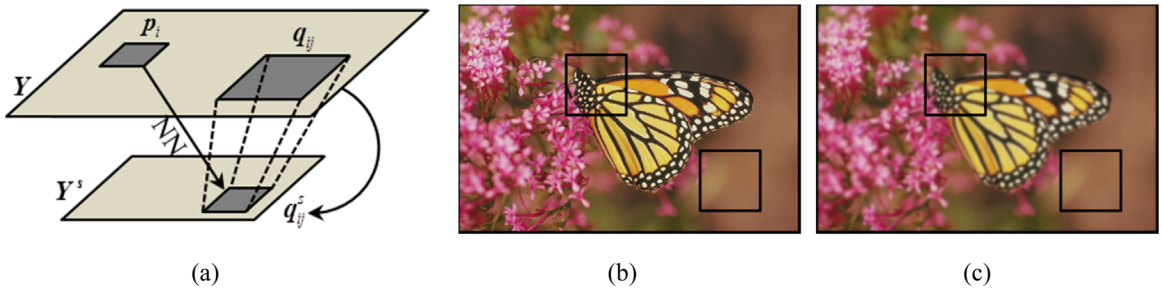


Fig. 2. Descriptions of cross-scale patch redundancy and blurring effect on different patches in natural images. (a) Cross-scale patch redundancy (see context), where $\boldsymbol{k}$ is blurring kernel. (b) Structured patch and smooth patch in **clean** "Monarch". (c) Structured patch and smooth patch in **blurred** "Monarch". Here, (c) is the result of convolving (b) with a 11×11 Gaussian kernel with $\sigma = 1.5$. Obviously, structured patch (upper left) is very different after being blurred with Gaussian kernel, whereas smooth patch (lower right) keeps almost unchanged visually.

by using weighted least-squares to obtain an updated kernel. These steps are repeated until the root mean squared error (RMSE) between patches in the input image and their NNs in down-sampled version does not increase further. Fig. 2(a) shows the cross-scale blur kernel estimation process.

The method of [14] is actually to solve a convolution kernel that makes cross-scale patch as redundant as possible. It depends on the observation that HR images possess more patch recurrence than LR images to solve optimal blur kernel by maximizing cross-scale patch recurrence. The process was interpreted as a maximum a posteriori (MAP) estimation process in Ref. [14]. However, as we can see in Fig. 2(b) and (c), the effect of blurring in smooth area is unconspicuous. Fig. 1(a) also shows the same issue. This phenomenon can be explained easily according to the definition of convolution. Therefore, smooth patches almost contribute nothing in blur kernel estimation with respect to structured patches.

2.3. Coupled dictionary learning

The current research on digital image and signal shows that sparse representation can recover latent signal more accurately relative to traditional approaches. This is mainly due to its capacity of adaptively computing the number of reconstructing samples and the coefficient of linear combination. In SR tasks, however, it generally involves two feature spaces: LR feature space $\mathbf{F}_L$ and HR feature space $\mathbf{F}_H$. There exists a certain unknown relation M: $\mathbf{F}_L \rightarrow \mathbf{F}_H$ between the two feature spaces. Generally, M is unknown and non-linear. So the traditional practice of applying the sparse representation of LR feature space directly to HR feature space is not applicable.

The goal for coupled dictionary learning is to find a coupled dictionary pair $\boldsymbol{D}_L$ and $\boldsymbol{D}_H$ for feature space $\mathbf{F}_L$ and $\mathbf{F}_H$ respectively, so that we can use the sparse representation of any testing signal $\boldsymbol{y} \in \mathbf{F}_L$ in terms of $\boldsymbol{D}_L$ to recover the corresponding latent HR signal $\boldsymbol{x} \in \mathbf{F}_H$ in terms of $\boldsymbol{D}_H$ directly. Under the condition of L_1 norm, it can be expressed as following equations for any coupled signal pair formally:

$$\boldsymbol{\pi}_i = \arg\min_{\boldsymbol{\alpha_i}} \|\boldsymbol{y}_i - \boldsymbol{D}_L\boldsymbol{\alpha}_i\|_2^2 + \lambda\|\boldsymbol{\alpha}_i\|_1, \qquad \forall\, i = 1, \ldots, N_L \tag{1}$$

$$\boldsymbol{\pi}_i = \arg\min_{\boldsymbol{\alpha_i}} \|\boldsymbol{x}_i - \boldsymbol{D}_H\boldsymbol{\alpha}_i\|_2^2, \qquad \forall\, i = 1, \ldots, N_H \tag{2}$$

where N_L and N_H are the numbers of the atoms in $\boldsymbol{D}_L$ and $\boldsymbol{D}_H$ respectively, and generally they are equal to each other. $\boldsymbol{\pi} = \{\boldsymbol{\pi}_i | i = 1, 2, \ldots, N_H\}$ is the set of sparse representation on coupled feature spaces. The sparse representation coefficient vector $\boldsymbol{\alpha}_i$ is utilized for the ith signal pair, and λ is the regularization parameter as tradeoff between the reconstruction error and the L_1 norm regularization terms.

To balance the reconstruction error on observation feature space $\mathbf{F}_L$ and latent feature space $\mathbf{F}_H$, Yang et al. [11] proposed an effective solution called coupled dictionary learning (CDL). CDL models the dictionary learning problem as a bilevel optimization problem to minimize the squared loss term on both feature space, namely, one of the optimization problems on the observation feature space and the other on the latent feature space. To balance the reconstruction error on observation feature space and latent feature space, CDL algorithm changes the objective function of [8]:

$$\begin{aligned} &\arg\min_{\boldsymbol{D}_H,\boldsymbol{D}_L,\boldsymbol{\pi}} \sum_{i=1}^{N_H} \left(\|\boldsymbol{D}_H\boldsymbol{\alpha}_i - \boldsymbol{x}_i\|_2^2 + \|\boldsymbol{D}_L\boldsymbol{\alpha}_i - \boldsymbol{y}_i\|_2^2 + \lambda\|\boldsymbol{\alpha}_i\|_1 \right) \\ &\text{s.t. } \|\boldsymbol{D}_H(:, i)\|_2 \le 1, \|\boldsymbol{D}_L(:, i)\|_2 \le 1 \end{aligned} \tag{3}$$

to the following form:

$$\arg\min_{\boldsymbol{D}_H,\boldsymbol{D}_L,\boldsymbol{\pi}} \sum_{i=1}^{N_H} \frac{1}{2}\left(\gamma\|\boldsymbol{D}_H\boldsymbol{\alpha}_i - \boldsymbol{x}_i\|_2^2 + (1-\gamma)\|\boldsymbol{D}_L\boldsymbol{\alpha}_i - \boldsymbol{y}_i\|_2^2\right) \quad \text{s.t. } \|\boldsymbol{D}_H(:, i)\|_2 \le 1, \quad \|\boldsymbol{D}_L(:, i)\|_2 \le 1 \tag{4}$$

where γ is the parameter balancing the reconstruction error on both feature spaces. Eq. (4) is solved by alternatively optimizing over $\boldsymbol{D}_L$ and $\boldsymbol{D}_H$ while keeping the other fixed. Stochastic gradient descent using implicit differentiation is employed to solve $\boldsymbol{D}_L$ during the iterative process.

3. The proposed approach

3.1. Selective patch processing strategy

Discriminating patches during sparse recovery was utilized for reducing time consumption by [11]. We adopting the same philosophy in our method, but there are two key differences. Firstly, Yang et.al employed SPP only for improving SR recovery efficiency, whereas we also use it to improve the quality of BKE. Our subsequent experiments will show the validity of properly abandoning useless smooth patches. Secondly, Yang et al. simply took the statistical variance of a patch as the criterion of SPP. However, the average gradient amplitude |grad| of a patch is more expressive than statistical variance according to the internal statistics of a single natural image [15]. So we utilized |grad| to differentiate patches.

If the size of patches is $w \times h$, and P_{ij} denotes a certain point located in (i, j), P_{ij}^x and P_{ij}^y represent the derivatives of x and y direction respectively. Then |grad| can be simply formulated as:

$$|\,\text{grad}\,| = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} \sqrt{\left(P_{ij}^x\right)^2 + \left(P_{ij}^y\right)^2} \tag{5}$$

We set a threshold τ_b for each query patch in BKE stage. A patch

is used to estimate blur kernel only when the |grad| of it is bigger than τ_b. Otherwise, the patches are simply dropped. Similarly, we set another threshold τ_r for each testing patch in SR recovery. If the |grad| of a testing patch is smaller than τ_r, then we use traditional interpolation to scale it. Otherwise, we utilize CDL-based sparse representation to process it.

3.2. Blind kernel estimation

We use $\boldsymbol{Y}$ to represent input LR image, and $\boldsymbol{X}$ to be latent HR image. As shown in Fig. 2(a), for each small patch $\boldsymbol{p}_i$ in the input image $\boldsymbol{Y}$, we can find a few NNs $\boldsymbol{q}_{ij}^s$ for it in down-sampled version $\boldsymbol{Y}^s$. The "parent" patches $\boldsymbol{q}_{ij}$ right above $\boldsymbol{q}_{ij}^s$ are viewed as the candidate parent patches of $\boldsymbol{p}_i$. The patch pairs $\{\boldsymbol{p}_i, \boldsymbol{q}_{ij}\}$ are used to construct a set of linear equations, which is solved by weighted least square rule. According to [14], the weight of each "parent" patch is calculated by the following formula so that good NNs contribute more than their poor counterparts:

$$w_{ij} = \frac{\exp\left(-\|\boldsymbol{p}_i - \boldsymbol{q}_{ij}^s\|^2/\sigma^2\right)}{\sum_{j=1}^{M_i}\exp\left(-\|\boldsymbol{p}_i - \boldsymbol{q}_{ij}^s\|^2/\sigma^2\right)} \tag{6}$$

where M_i is the number of NNs of each small patch $\boldsymbol{p}_i$ in $\boldsymbol{Y}$, and $\sigma$is the standard deviation of noise added on $\boldsymbol{p}_i$. s is the scale factor. Note that we apply the same symbol to express column vector corresponding to the patch for convenient expression. Maximizing the cross-scale NLSS of $\boldsymbol{Y}$ with respect to the scale factor s is intuitively equivalent to minimizing the dissimilarity between cross-scale patches. Therefore, we solve L_2 norm optimization problem for BKE:

$$arg\min_{\boldsymbol{k}} \sum_{i=1}^{N}\|\boldsymbol{p}_i - \sum_{j=1}^{M_i} w_{ij}\boldsymbol{R}_{ij}\boldsymbol{k}\|_2^2 + \lambda\|\boldsymbol{C}\boldsymbol{k}\|_2^2 \tag{7}$$

where N is the number of small query patches in $\boldsymbol{Y}$. Matrix $\boldsymbol{R}_{ij}$ corresponds to the operation of convolving with $\boldsymbol{q}_{ij}$ and down sampling by s. $\boldsymbol{C}$ is a matrix used as the penalty of non-smooth kernel. The second term of Eq. (7) is kernel prior and λ is the balance parameter as the tradeoff between the error term and kernel prior. By setting the gradient of the objective function in Eq. (7) to zero, we can get the update formula of $\boldsymbol{k}$:

$$\widehat{\boldsymbol{k}} = \left(\sum_{i=1}^{N}\sum_{j=1}^{M_i} w_{ij}\boldsymbol{R}_{ij}^T\boldsymbol{R}_{ij} + \lambda\boldsymbol{C}^T\boldsymbol{C}\right)^{-1}\cdot\sum_{i=1}^{N}\sum_{j=1}^{M_i} w_{ij}\boldsymbol{R}_{ij}^T\boldsymbol{p}_i \tag{8}$$

This is similar to the result of [14], which can be interpreted as maximum a posteriori (MAP) estimation on $\boldsymbol{k}$. However, our blur kernel estimation has two essential differentials with respect to [14]. Firstly, our method is driven by the idea of minimizing the dissimilarity between cross-scale patches while [14] tends to maximize the similarity directly. This may not be easy to understand. But, the latter leads Michaeli and Irani [14] to form their kernel update formula from physical analysis and interpretation of "optimal kernel". But the former leads us to obtain kernel update formula from quantitating cross-scale patch dissimilarity and directly minimizing it according to ridge regression. Secondly, the number of NNs of each small patch is not fixed which provides more flexibility during solving least square problem. Therefore, the terminal criterion cannot be the totality of NNs. We utilize the average patch dissimilarity (APD) as terminal condition of iteration:

$$\text{APD} = \left(\sum_{i=1}^{N}\sum_{j=1}^{M_i}\|\boldsymbol{p}_i - \boldsymbol{q}_{ij}^s\|_2^2\right)\cdot\left(\sum_{i=1}^{N} M_i\right)^{-1} \tag{9}$$

In order to find NNs of required patches, we need to perform broader range of search in entire image here. According to the three conclusions of [15] (See Section 2.1), finding NNs for structured patches requires larger search region. We search the whole image for each patch because SPP has dropped many invalid patches and searching in the whole image will not spend too much time.

3.3. Sparse recovery with CDL

We use the estimated kernel to down-sample the input image and extract feature from patches. Assuming $\boldsymbol{y}_p$ is a LR patch containing original image data and $\boldsymbol{y}$ is LR patch feature extracting from $\boldsymbol{y}_p$ (usually $\boldsymbol{y}$ is a vector concatenated with 1st and 2nd derivatives along with x and y direction). We need to solve the following L_1 norm constrained optimization problem:

$$\arg\min_{\boldsymbol{\alpha}}\|\boldsymbol{y} - \boldsymbol{D}_L\boldsymbol{\alpha}\|_2^2 + \varphi\|\boldsymbol{\alpha}\|_1 \tag{10}$$

where φ allows alleviating the ill-posed problems and stabilizes the solution. $\boldsymbol{y}$ corresponds to a testing LR patch extracted from interpolation version of input image. $\boldsymbol{D}_L$ is the LR dictionary trained by CDL. The solution of Eq. (10) is found by solving a L_1-regularized Lasso problem. This can be done by Sparse Learning with Efficient Projections (SLEP) [24]. After we get the coefficient vector $\boldsymbol{\alpha}$ of LR patch feature vector $\boldsymbol{y}$, HR patch feature then can be computed directly by:

$$\boldsymbol{x} = \frac{\boldsymbol{D}_H\boldsymbol{\alpha}}{\|\boldsymbol{D}_H\boldsymbol{\alpha}\|_2} \tag{11}$$

$\boldsymbol{D}_H$ is the HR dictionary trained by CDL here. Then, Recovery HR image patch $\boldsymbol{x}_p$ can be obtained by:

$$\boldsymbol{x}_p = (c\times u)\times\boldsymbol{x} + v \tag{12}$$

where $u = mean\left(\boldsymbol{y}_p\right)$ and $v = \|\boldsymbol{y}_p - u\|_2$. c is a constant and used for properly scaling the feature vector. Generally, solving Eq. (10) is severely time-demanding, and there are two directions to accelerate the SR recovery: reducing the number of patches to process and finding a fast solver for L_1 norm minimization problem Eq. (10). SPP is adopted in our method as [11]. However, as mentioned above, the criterion of selecting patches is the gradient magnitude |grad| instead of the variance of a patch according to [15].

4. Experimental results

In this section, we will give some experimental results about blur kernel estimation and SR recovery. All the experiments are performed on a Philips PC with 8.0 GB memory and running a single core of Intel Xeon 2.53 GHz CPU. We mainly compare our blur kernel estimation method with [14] in terms of kernel accuracy and efficiency, while SR performance is compared with several state-of-the-art SR algorithms.

4.1. Comparisons for estimating blur kernel

4.1.1. Preparation and parameters

We only performed detailed comparison between our BKE method and [14] because they have common background in theory basis and similar implementation. $2\times$ and $3\times$ SRs are performed in our experiments. When scale factor $s = 2$, we set the size of small query patches $\boldsymbol{p}_i$ and candidate patches $\boldsymbol{q}_{ij}^s$ of NNs to 5×5, while the size of "parent" patches $\boldsymbol{q}_{ij}$ are set to 9×9. Both $\boldsymbol{p}_i$ and $\boldsymbol{q}_{ij}^s$ do not change size, but "parent" patches are set to be 13×13 patches when perform $3\times$ SR. Noise standard deviation $\sigma$is assumed to be 5. Parameter λ in Eq. (7) is set to 0.25, and matrix $\boldsymbol{C}$

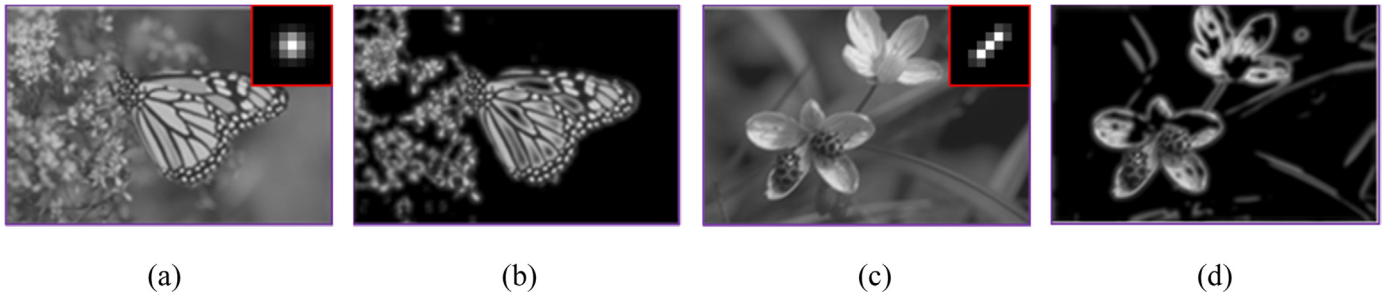


Fig. 3. Blurring on original images and selective patch process in BKE. (a) The monochrome "Monarch" image blurred by a 9×9 Gaussian kernel with a standard deviation $\sigma = 1.5$. (b) Structured content of (a) selected to estimate blur kernel. (c) The monochrome "Flower" image blurred by a 9×9 "motion" kernel with len $= 5$ and theta $= 45$. (d) Structured content of (c) selected to estimate blur kernel.

is chosen to be the derivative matrix corresponding to x and y directions of "parent" patches as [14]. The threshold of |grad| τ_b for selecting query patches is set to 10. Fig. 3 shows the structured parts used to perform BKE, and the blur kernels of "Monarch" and "Flower". It can be cleanly seen that input data are seriously blurred and inaccurate, and traditional sparse representation methods not executing BKE are inoperative in this case actually. Besides, there is only a few part of original image that can be used for BKE due to selective patch processing.

4.1.2. Accuracy comparison

Estimated BKE results from "Monarch" and "Flower" are presented in Fig. 4 to illustrate the accuracy of the recovered kernel qualitatively. As shown, both algorithms can estimate the rough shape of the ground-truth kernel, but our approach gives more accurate results on both kernel size and shape. This is true especially when the ground-truth one is a motion kernel. On the one hand, not taking fixed number of NNs for each query patch in LR images makes BKE more flexible, and reduces the adverse impact caused by over-fitting or/and under-fitting. On the other hand, Selective patch processing strategy measured by the average gradient amplitude |grad| abandoned useless smooth patches and kept more structured non-smooth patches, which relieved the computational demanding for CPUs significantly. It can be easily concluded that our method provided more accurate BKE results according to the size and shape of estimated kernels.

4.1.3. Efficiency comparison

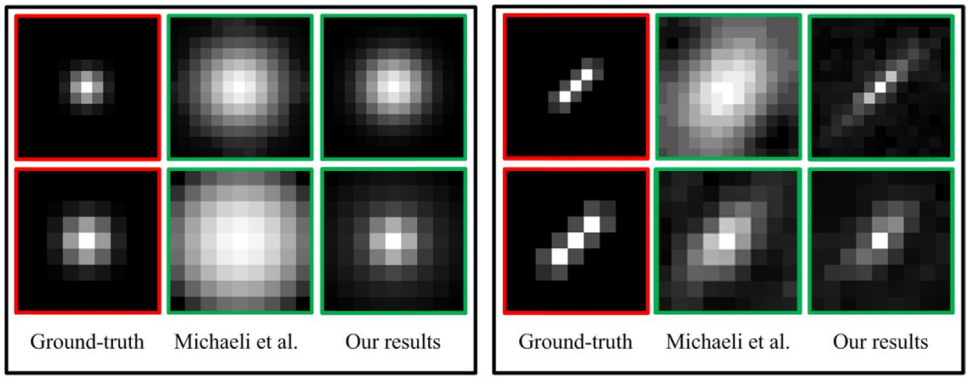
We collect the average MSE of each NN during iterations as shown in Fig. 5(a). Though we selectively use structured patches to estimate blur kernels, the average MSE of our algorithm still smaller than [14]. This is mainly attributed to the unfixed number of NNs to construct the set of linear equations. As we can see in Fig. 5(a), our BKE method converges through about 8 iterations, whereas the algorithm of [14] converges through almost 15 iterations. In fact, because we dropped a part of query patches at beginning, the time spent by each iteration is much less than [14]. After the convergence of algorithms, the average MSE between cross-scale patches of our algorithm is also smaller than Michaeli's algorithm. Therefore, the selective patch process and adaptive number of NNs produce more accurate BKE results, but need less iteration times.

We apply the same measures with [14] to compare the effect of estimated kernels on SR algorithms: The Error Ratio to Ground Truth (ERGT), which measures the ratio between the SR reconstruction error with estimated kernels and the SR reconstruction error with the ground-truth one

$$\text{ERGT}=\frac{\|\boldsymbol{X}-\overline{\boldsymbol{X}}_{\hat{\boldsymbol{k}}}\|_2}{\|\boldsymbol{X}-\overline{\boldsymbol{X}}_{\boldsymbol{k}}\|_2} \tag{13}$$

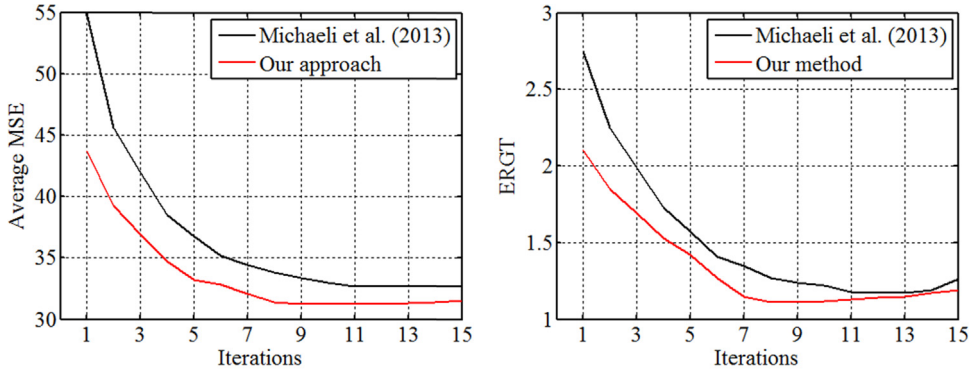
where $\overline{\boldsymbol{X}}_{\hat{k}}$ and $\overline{\boldsymbol{X}}_{k}$ are the recovered HR images with estimated kernel $\hat{\boldsymbol{k}}$ and ground-truth kernel $\boldsymbol{k}$ respectively. Apparently, if ERGT is close to 1, it indicates that the estimated kernel is nearly as good as the ground-truth kernel. Fig. 5(b) shows the results of applying both estimated kernels in the SR algorithm proposed by Yang et al. [9]. We super-resolved all the five testing images with scale factor $s = 2$, and collect the average ERGT at each iteration. In the whole iterative process, the ERGT of our algorithm is always closer to 1 than [14].

Fig. 6 shows the SR results of applying estimated kernels in the original sparse representation algorithm presented in Ref. [9]. We extract the coccinella septempunctata in the "flower" image. The ground-truth is a 13×13 (i.e. $3 \times$ SR) "motion" kernel with len $=$ 5 and theta $= 45$. The same SR algorithm with our estimated kernels gives more distinct edges and textures versus that estimated by the



(a) "gaussian" kernels (b) "motion" kernels

Fig. 4. The accuracy comparisons for BKE. (a) Gaussian kernel estimation with "Monarch" image. (b) Motion kernel estimation with "Flower" image. The size of kernels in the top row and the bottom row are 13×13 and 9×9 respectively. (a) "gaussian" kernels (b) "motion" kernels.



(a) Average MSE Comparison (b) ERGT comparison

Fig. 5. BKE efficiency Comparisons. SR algorithm is original sparse representation proposed by Yang et al. [9]. (a) The attenuation trend of average MSE as iterations. (b) The change of ERGT as presented in Ref. [14]. (a) Average MSE Comparison (b) ERGT comparison.

(a) (b) (c) (d) (e) (f)

Fig. 6. SR comparisons with different kernels. All of these SR images were recovered by original sparse representation proposed by Yang et al. [6]. (a) Reference image. (b) Bicubic interpolation. (c) Michaeli et al. [14]. (d), (e), and (f) are our results when the threshold of |grad| is 10, 20, and 30 respectively.

Table 1
Thresholding effects on recovery accuracy and time consumption ($\tau_b = 10$).

τ_r ($\times 2$)		0	10	20	30
Lena	SSIM	0.9538	0.9527	0.9514	0.9507
	Time (s)	53.5624	44.2485	27.7148	15.6733
Monarch	SSIM	0.9428	0.9415	0.9405	0.9386
	Time (s)	63.7813	41.1298	22.8712	11.8635

Table 2
SR comparisons with several typical blind methods and proposed algorithm ($\tau_b = 15$).

Image Sets	Scale	Michaeli et al. [14]		Shao et al. [25]		Ours	
		PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
Set2	$\times 2$	28.17	0.9014	30.72	0.9377	31.87	0.9507
	$\times 3$	26.35	0.8675	28.95	0.8995	30.35	0.9296
Set5	$\times 2$	28.66	0.9167	30.34	0.9363	32.01	0.9469
	$\times 3$	27.19	0.8741	28.29	0.9045	30.47	0.9324
Set14	$\times 2$	28.31	0.9009	30.49	0.9331	31.92	0.9478
	$\times 3$	26.87	0.8693	29.03	0.8986	29.98	0.9289

method of [14]. Furthermore, the change of visual effect is not obvious as the threshold of average gradient magnitude increases, but properly increasing the threshold can reduce the number of query patches and hence shorten the time consumption of BKE effectively.

4.2. Comparisons for SR recovery

4.2.1. Objective evaluation

To simulate many actual application scenarios, all input images for each algorithm are obtained by blurring clean images with default blur kernels. In other words, input image data is set to be of low quality via some blur kernels in our experiments. This is also very different from current general experimental methodologies employed by most of non-blind algorithms and the most important equivalent condition set on our experiments.

Instead of simply using the variance of a small patch to be the criterion of patch selection, we adopted average gradient magnitude |grad| presented in Ref. [15] to filter smooth patches. Table 1 presents the effects of thresholding on SR recovery accuracy and processing time with sparse representation solved by L_1 norm minimization problem. With the threshold on |grad| increasing, the consumed time decreases rapidly whereas SSIM reduces slightly. Note that the threshold on |grad| cannot increases unlimitedly for that may reduce the number of valid patches and lead imprecise blur kernel estimation. Apparently, our method provides the least amount of time consumption due to SPP in terms of the methods of [9,22,23], which calculate L_0/L_1 form normalized optimization problem for the unnecessary smooth patches. Table 2 presents the average PSNR and SSIM comparisons between the proposed method and two typical blind SR algorithms proposed recently. Image data include Set2, Set5 and Set14, which are collected form [26]. Note the reference is the blurred images, but not the original clean image when comparison. Our approach brings big improvement over PSNR and SSIM compared with other blind SR methods. This is mainly due to the more accurate results of our BKE providing the training samples of higher quality, which can be seen from the following visual comparison results.

4.2.2. Subjective evaluation

Fig. 7 shows some visual comparisons of above-mentioned

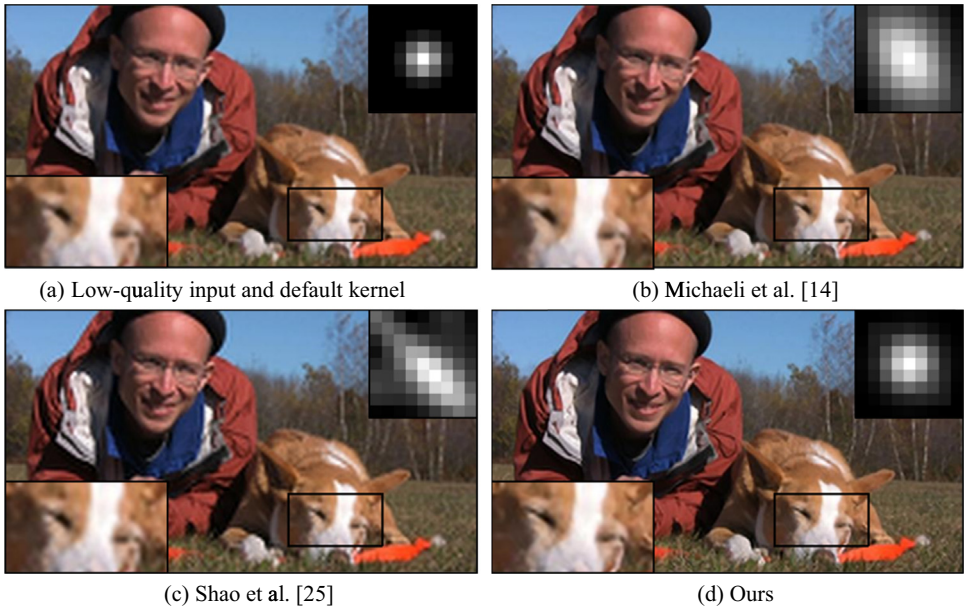


Fig. 7. SR Recovery with low-quality "dog" image from Set2 ($\times 2$). Input image generated from clean image blurred with a 9×9 "gaussian" kernel (gsize $= 5$ and sigma $=$ 1.0). (a) Low-quality input and default kernel (b) Michaeli et al. [14] (c) Shao et al. [25] (d) Ours.

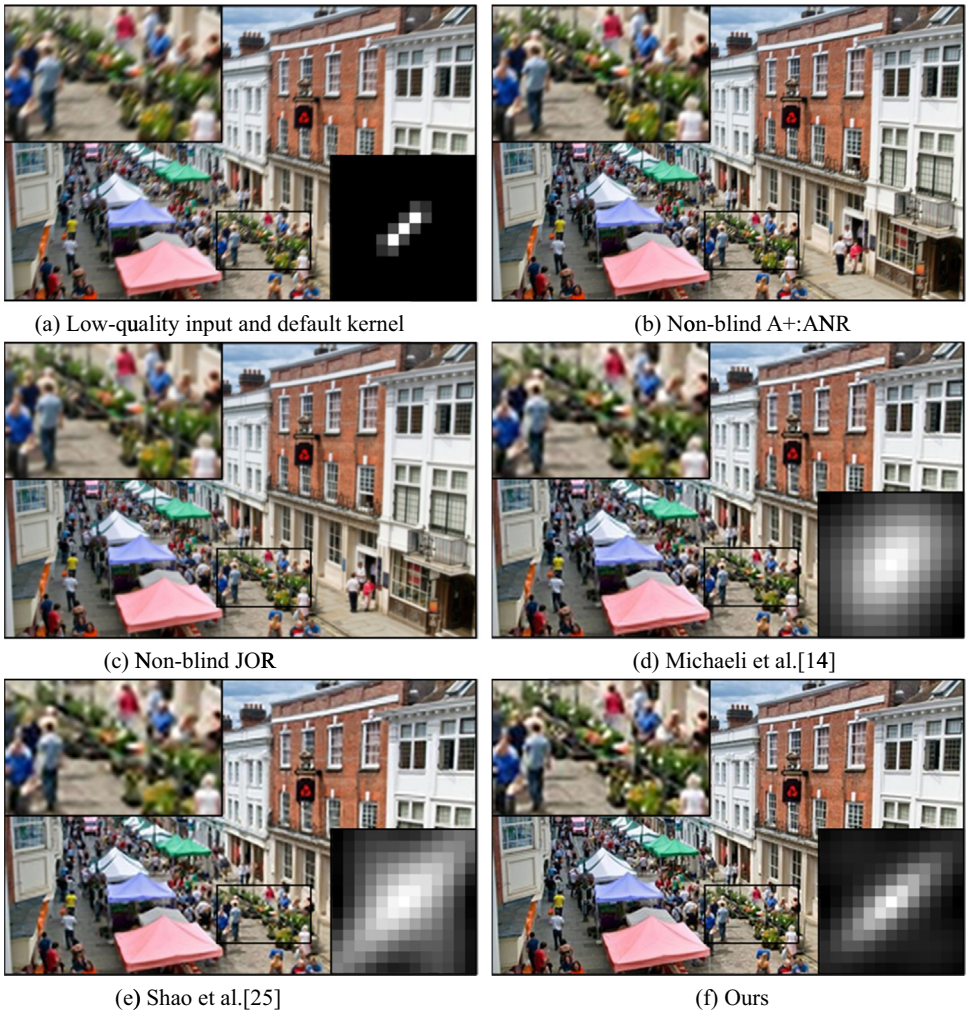


Fig. 8. SR Recovery with low-quality "high-street" image from Set2 ($\times 3$). Input image generated from clean image blurred with a 13×13 "motion" kernel (len $= 5$ and theta $= 45$). (a) Low-quality input and default kernel (b) Non-blind A+:ANR (c) Non-blind JOR (d) Michaeli et al.[14] (e) Shao et al.[25] (f) Ours.

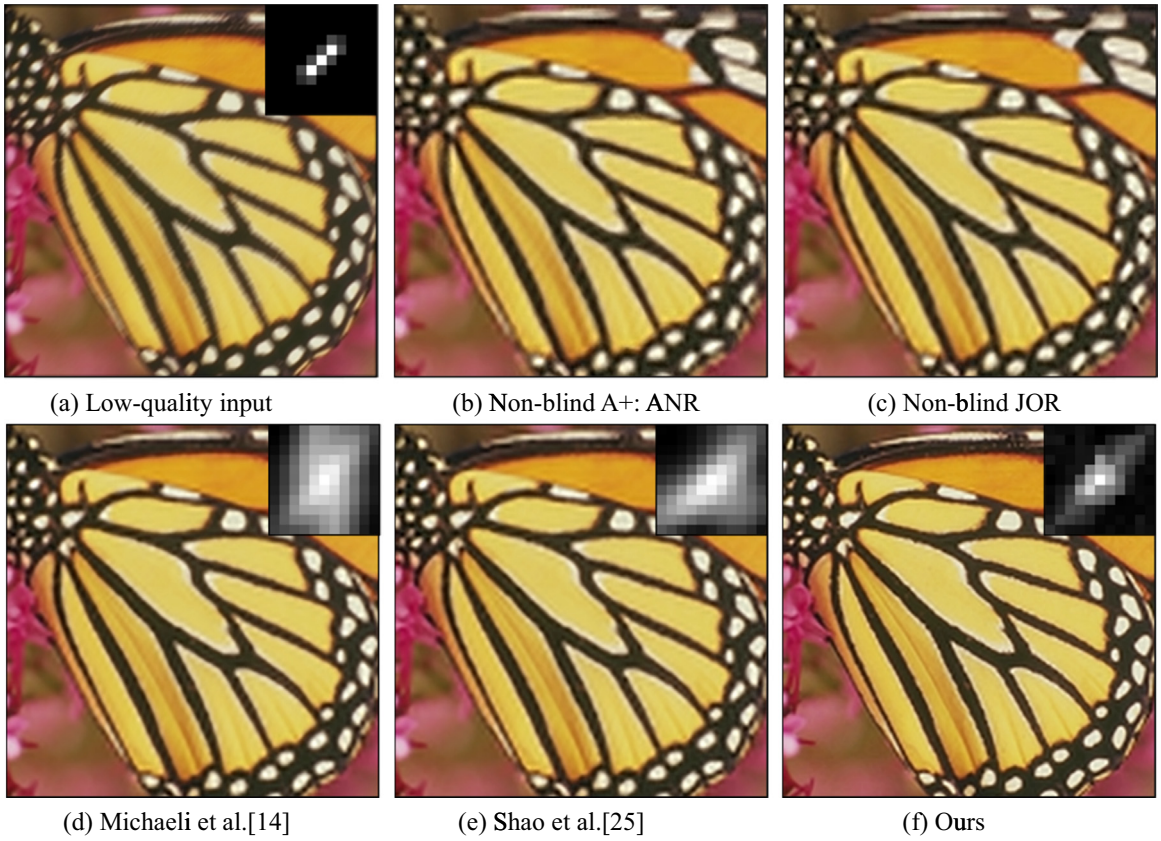


Fig. 9. SR Recovery with low-quality "butterfly" image from Set5 ($\times 2$). Input image generated from clean image blurred with a 11×11 "motion" kernel (len $= 5$ and theta $=$ 45). (a) Low-quality input (b) Non-blind A+: ANR (c) Non-blind JOR (d) Michaeli et al.[14] (e) Shao et al.[25] (f) Ours.

blind SR reconstruction approaches. For testing purpose, all input images are convolved with default kernels and then downsampled to corresponding scales. All results are generated through the same recovery method after BKE. We can see from Fig. 7 that our BKE method provides the most accurate estimated kernel (from the perspective of kernel size and shape) and more realistic SR output. Figs. 8 and 9 two non-blind SR algorithms: Adjusted Anchored Neighborhood Regression (A+ ANR) [26] and Jointly Optimized Regressors (JOR) [27], which are the most state-of-the-art algorithms proposed recently. However, they completely failed in achieving acceptable SR performance when input images are severely degraded and the used blur kernel deviates from the true one. Note Fig. 8(b) seems to be clear as Fig. 8(f), but the impact of "motion" kernel is obvious. The estimated kernel.

of our BKE method in Fig. 9 has some "steep" components. This is mainly because there is a lot of "steep" flat area in the blurred "butterfly" image, and they were dropped as smooth query patches during SPP.

5. Conclusion

We proposed a novel single image blind SR framework aiming at improving the SR effect and reducing time consumption in this paper. The algorithm mainly contains blur kernel estimation and SR recovery stages. The former is realized by minimizing dissimilarity of cross-scale image patches, which is slightly similar to the MAP estimation approach proposed by Michaeli et al. [14]. The latter relies on a coupled dictionary learning process [11] and simply solves a L_1 form constrained optimization problem on the patches reserved by SPP. The SPP utilized in these two stages depends on the criterion of average gradient amplitude |grad| instead of statistical variance. All above-mentioned processes make our SR algorithm could achieve the better SR results than several typical blind or non-blind SR approaches. However, solving L_1 norm optimization problem is still extraordinarily time-consuming. Actually, ANR and JOR are much faster than proposed method. Our further research work will be to find out other better ways of improving the SR accuracy and reducing the time consumption further.

Acknowledgement

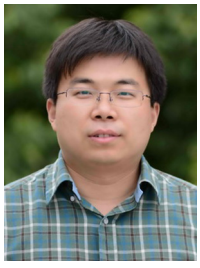
We would like to thank Mr. Wen-Ze Shao and Mr. Michael Elad for their kind help in running the blind SR method [25], which thus enables an effective comparison with [25]. This work is supported by National Natural Science Foundation of China (Grant No. 61303127), Western Light Talent Culture Project of Chinese Academy of Sciences (Grant No. 13ZS0106), Project of Science and Technology Department of Sichuan Province (Grant Nos. 2014SZ0223, 2015GZ0212), Key Program of Education Department of Sichuan Province (Grant Nos. 11ZA130, 13ZA0169), and the innovation funds of Southwest University of Science and Technology (Grant No. 15ycx053).

References

[1] Q. Miao, P. Xu, T. Liu, et al., Linear feature separation from topographic maps using energy density and the shear transform, IEEE Trans. Image Process. 22 (4) (2013) 1548–1558.
[2] Q. Miao, C. Shi, P. Xu, M. Yang, Y. Shi, A novel algorithm of image fusion using shearlets, Opt. Commun. 28 (46) (2011) 1540–1547.
[3] W.T. Freeman, T.R. Jones, E.C. Pasztor, Example-based super-resolution, IEEE

Comput. Graph. Appl. 22 (2) (2002) 56–65.
[4] J. Sun, N.N. Zheng, H. Tao, H.Y. Shum, Image hallucination with primal sketch priors. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Monona: IEEE, 2003, 2, pp. 729–736.
[5] Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. CVPR 2004: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE, 1, 2004, pp. 275–282.
[6] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[7] M. Bevilacqua, A. Roumy, C. Guillemot, M.L. Alberi-Morel, Low-complexity single-image super-resolution based on nonnegative neighbor embedding, BMVC (2012) 1–10.
[8] Q. Miao, Y. Cao, G. Xia, et al., RBoost: label noise-robust boosting algorithm based on a nonconvex loss function and the numerically stable base learners, IEEE Trans. Neural Netw. Learn. Syst. 1 (2015).
[9] J.C. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution as sparse representation of raw image patches. IEEE Conference on Computer Vision and Pattern Recognition, Anchorage AK: IEEE, 2008, pp. 1–8.
[10] D., Glasner, S. Bagon, M. Irani, Super-resolution from a single image. IEEE Conference on Computer Vision, Kyoto: IEEE, 30(2), 2009, pp. 349–356.
[11] J.C. Yang, Z.W. Wang, Z. Lin, S. Cohen, T. Huang, Coupled dictionary training for image super-resolution, IEEE Trans. Image Process. 21 (8) (2012) 3467–3478.
[12] J.C. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.
[13] L. He, H.R. Qi, R. Zaretzki, Beta Process Joint Dictionary Learning for Coupled Feature Spaces with Application to Single Image Super-Resolution. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland OR: CVPR, 2013, pp. 345–352.
[14] T. Michaeli, M. Irani, Nonparametric blind super-resolution. 2013 IEEE International Conference on Computer Vision (ICCV), Sydney VIC: ICCV, 2013, pp. 945–952.
[15] M. Zontak, M. Irani, Internal statistics of a single natural image. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence RI: IEEE, 42, 2011, pp. 977–984.
[16] C.Y. Yang, J.B. Huang, M.H. Yang, Exploiting self-similarities for single frame super-resolution. The 10th Asian Conference on Computer Vision, Queenstown New Zealand: ACM, 2011, pp. 497–510.
[17] D. Zoran, Y. Weiss, From learning models of natural image patches to whole image restoration. 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona: IEEE, 6669, 2011, pp. 479–486.
[18] J. Hu, Y.P. Luo, Single-image superresolution based on local regression and nonlocal self-similarity, J. Electron. Imaging 23 (3) (2014) 6–8.
[19] Y.Q. Zhang, J.Y. Liu, S. Yang, Z.M. Guo, Joint image denoising using self-similarity based low-rank approximations, Vis. Commun. Image Process. (VCIP) Kuhing: IEEE 7 (2) (2013) 1–6.
[20] T. Michaeli, M. Irani, Blind deblurring using internal patch recurrence. The 13th European Conference on Computer Vision (ECCV), Zurich Switzerland: Springer, 2014, pp. 783–798.
[21] C. Guillemot, O.L. Meur, Image inpainting: overview and recent advances, Signal Processing Magazine 31 (2014) 127–144.
[22] R. Timofte, V.D. Smet, L.V. Gool, Anchored neighborhood regression for fast example-based super-resolution. 2013 IEEE International Conference on Computer Vision (ICCV), (Sydney NSW: IEEE, 1920–1927, 2013.
[23] R. Zeyde, M. Elad, M. Protter, On single image scale-up using sparse representations. The 7th International Conference on Lecture Notes in Computer Science (LNCS). Avignon France: Springer, 2012, pp. 711–730.
[24] J. Liu, S. Ji, J. Ye, SLEP: Sparse learning with efficient projections, Arizona State University, 2010.
[25] W. Shao, E. Michael, Simple, accurate, and robust nonparametric blind super-resolution. The International Conference on Image and Graphics (ICIG), Tianjin China: Springer, 2015, pp. 13–16.
[26] R. Timofte, S.V. Smet, L.V. Gool, A+: adjusted anchored neighborhood regression for fast super-resolution, Lect. Notes Comput. Sci. (2014) 111–126.
[27] D. Dai, R. Timofte, L.V. Gool, Jointly optimized regressors for image super-resolution, Comput. Graph. Forum (2015) 95–104.

Xiaole Zhao, born in 1987, M.S. candidate. He received his bachelor's degree in computer science from the School of Computer Science and Technology, SWUST, China, in 2013. He is now studying in School of Computer Science and Technology, SWUST, for his master's degree. His main research interests include digital image processing, machine learning, and data mining.



Ya-Dong Wu received his B.S. degree in computer science from Zhengzhou University, China, in 2000, and his M.S. degree in control theory and control engineering from Southwest University of Science and Technology, China, in 2003.He got Ph.D. degree in computer application at University of Electronic Science and Technology of China. Now he is a full professor with the School of Computer Science and Technology, SWUST. His research interest includes image processing and visualization.

Jinsha Tian, born in 1988, M.S. candidate. She received her bachelor's degree in Hebei University of Science and Technology, China, in 2013. She is now studying in School of Computer Science and Technology, SWUST, for her master's degree. Her main research interests include digital image processing.

Hongying Zhang received her B.S. degree in applied mathematics from Northeastern University, China, in 2000, and her M.S. degree in control theory and control engineering from Southwest University of Science and Technology, China, in 2003. She got Ph.D. degree in signal and information processing at University of Electronic Science and Technology of China. Now she is a full professor with the School of Information Engineering, SWUST. Her main research interest is image processing.