

# Single image super-resolution via blind blurring estimation and anchored space mapping

Xiaole Zhao<sup>1</sup> (✉), Yadong Wu<sup>1</sup>, Jinsha Tian<sup>1</sup>, and Hongying Zhang<sup>2</sup>

© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** It has been widely acknowledged that learning-based super-resolution (SR) methods are effective to recover a high resolution (HR) image from a single low resolution (LR) input image. However, there exist two main challenges in learning-based SR methods currently: the quality of training samples and the demand for computation. We proposed a novel framework for single image SR tasks aiming at these issues, which consists of blind blurring kernel estimation (BKE) and SR recovery with anchored space mapping (ASM). BKE is realized via minimizing the cross-scale dissimilarity of the image iteratively, and SR recovery with ASM is performed based on iterative least square dictionary learning algorithm (ILS-DLA). BKE is capable of improving the compatibility of training samples and testing samples effectively and ASM can reduce consumed time during SR recovery radically. Moreover, a selective patch processing (SPP) strategy measured by average gradient amplitude  $|\text{grad}|$  of a patch is adopted to accelerate the BKE process. The experimental results show that our method outruns several typical blind and non-blind algorithms on equal conditions.

**Keywords** super-resolution (SR); blurring kernel estimation (BKE); anchored space mapping (ASM); dictionary learning; average gradient amplitude

## 1 Introduction

Single image super-resolution has been becoming the hotspot of super-resolution area for digital images because it generally is not easy to obtain an adequate number of LR observations for SR recovery in many practical applications. In order to improve image SR performance and reduce time consumption so that it can be applied in practical applications more effectively, this kind of technology has attracted great attentions in recent years.

Single image super-resolution is essentially a severe ill-posed problem, which needs adequate priors to be solved. Existing super-resolution technologies can be roughly divided into three categories: traditional interpolation methods, reconstruction methods, and machine-learning (ML) based methods. Interpolation methods usually assume that image data is continuous and band-limited smooth signal. However, there are many discontinuous features in natural images such as edges and corners etc., which usually makes the recovered images by traditional interpolation methods suffer from low quality [1]. Reconstruction based methods apply a certain prior knowledge, such as total variation (TV) prior [2–4] and gradient profile (GP) prior [5] etc., to well pose the SR problem. The reconstructed image is required to be consistent with LR input via back-projection. But a certain prior is typically only propitious to specific images. Besides, these methods will produce worse results with larger magnification factor.

Relatively speaking, machine-learning based method is a promising technology and it has become the most popular topic in single image SR field. The first ML method was proposed by Freeman

<sup>1</sup> School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China. E-mail: X. Zhao, zxlation@foxmail.com (✉); Y. Wu, wyd028@163.com.

<sup>2</sup> School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China. E-mail: zhy0838@163.com.

Manuscript received: 2015-11-20; accepted: 2016-02-02

et al. [6], which is called example-based learning method. This method predicts HR patches from LR patches by solving markov random field (MRF) model by belief propagation algorithm. Then, Sun et al. [7] enhanced discontinuous features (such as edges and corners etc.) by primal sketch priors. These methods need an external database which consists of abundant HR/LR patch pairs, and time consumption hinders the application of this kind of methods. Chang et al. [8] proposed a nearest neighbor embedding (NNE) method motivated by the philosophy of locally linear embedding (LLE) [9]. They assumed LR patches and HR patches have similar space structure, and LR patch coefficients can be solved through least square problem for the fixed number of nearest neighbors (NNs). These coefficients are then used for HR patch NNs directly. However, the fixed number of NNs could cause over-fitting and/or under-fitting phenomena easily [10]. Yang et al. [11] proposed an effective sparse representation approach and addressed the fitting problems through selecting the number of NNs adaptively.

However, ML methods are still exposed to two main issues: the compatibility between training and testing samples (caused by light condition, defocus, noise etc.), and the mapping relation between LR and HR feature spaces (requiring numerous calculations). Glasner et al. [12] exploited image patch non-local self-similarity (i.e., patch recurrence) within image scale and cross-scale for single image SR tasks, which makes an effective solution for the compatibility problem. The mapping relation involves the learning process of LR/HR dictionaries. Actually, LR and HR feature spaces are tied by some mapping function, which could be unknown and not necessarily linear [13]. Therefore, the originally direct mapping mode [11] may not reflect this unknown non-linear relation correctly. Yang et al. [14] proposed another joint dictionary training approach to learn the duality relation between LR/HR patch spaces. The method essentially concatenates the two feature spaces and converts the problem to the standard sparse representation. Further, they explicitly learned the sparse coding problem across different feature spaces in Ref. [13], which is so-called coupled dictionary learning (CDL) algorithm. He et al. [15] proposed another

beta process joint dictionary learning (BPJDL) for CDL based on a Bayesian method through using a beta process prior. But, above-mentioned dictionary learning approaches did not take the feature of training samples into account for better performance. Actually, it is not an easy work to find the complicated relation between LR and HR feature spaces directly.

We present a novel single image super-resolution method considering both SR result and the acceleration of execution in the paper. The proposed approach firstly estimated the true blur kernel based on the philosophy of minimizing the dissimilarity between cross-scale patches [16]. LR/HR dictionaries then were trained via input image itself down-sampled by the estimated blur kernel. The BKE processing was adopted for improving the quality of training samples. Then,  $L_2$  norm regularization was used to substitute  $L_0/L_1$  norm constraint so that latent HR patch can be mapped on LR patch directly through a mapping matrix computed by LR/HR dictionaries. This strategy is similar with ANR [17], but we employed a different dictionary learning approach, i.e., ILS-DLA, to train LR/HR dictionaries. In fact, ILS-DLA unified the principle of optimization of the whole SR process and produced better results with regard to K-SVD used by ANR.

The remainder of the paper is organized as follows: Section 2 briefly reviews the related work about this paper. The proposed approach is described in Section 3 detailedly. Section 4 presents the experimental results and comparison with other typical blind and non-blind SR methods. Section 5 concludes the paper.

## 2 Related work

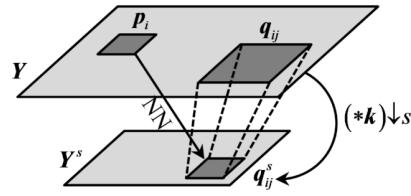
### 2.1 Internal statistics in natural images

Glasner et al. [12] exploited an important internal statistical attributes of natural image patches named the patch recurrence, which is also known as image patch redundancy or non-local self-similarity (NLSS). NLSS has been employed in a lot of computer vision fields such as super resolution [12, 18–21], denoising [22], deblurring [23], and inpainting [24] etc. Further, Zontak and Irani [18] quantified this property by relating it to the spatial distance from

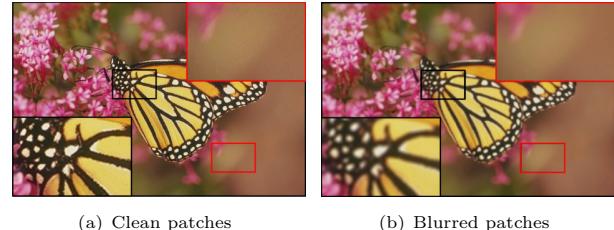
the patch and the mean gradient magnitude  $|\text{grad}|$  of a patch. The three main conclusions can be perceived according to Ref. [18]: (i) smooth patches recur very frequently, whereas highly structured patches recur much less frequently; (ii) a small patch tends to recur densely in its vicinity and the frequency of recurrence decays rapidly as the distance from the patch increases; (iii) patches of different gradient content need to search for nearest neighbors at different distances. These conclusions consist of the theoretical basis of using the mean gradient magnitude  $|\text{grad}|$  as the metric of discriminatively choosing different patches when estimating the blurring kernels.

## 2.2 Cross-scale blur kernel estimation

For more detailed elaboration, we still need to briefly review the cross-scale BKE and introduce our previous work [16] on this issue despite a part of it is the same as previous one. We will illustrate the detailed differences in Section 3.1. Because of camera shake, defocus, and various kinds of noises, the blur kernel of different images may be entirely and totally different. Michaeli and Irani [25] utilized the non-local self-similarity property to estimate the optimal blur kernel by maximizing the cross-scale patch redundancy iteratively depending on the observation that HR images possess more patch recurrence than LR images. They assumed the initial kernel is a delta function used to down-sample the input image. A few NNs of each small patch were found in the down-sampled version of input image. Each NN corresponds to a large patch in the original scale image, and these patch pairs construct a set of linear equations which could be solved by using weighted least squares. The root mean squared error (RMSE) between cross-scale patches was employed as the iteration criterion. Figure 1 shows the main process of cross BKE of Ref. [25]. We follow the same idea with more careful observation: the effect of the convolution on smooth patches is obviously smaller than that on structured patches (refer to Fig. 2). This phenomenon can be explained easily according to the definition of convolution. Moreover, the mean gradient magnitude  $|\text{grad}|$  is more expressive than the variance of a patch on the basis of the conclusions in Ref. [18].



**Fig. 1** Description of cross-scale patch redundancy. For each small patch  $p_i$  in  $\mathbf{Y}$ , finding its NNs  $q_{ij}^s$  in  $\mathbf{Y}^s$  which corresponds to a large patch  $q_{ij}$  in  $\mathbf{Y}$ ,  $q_{ij}^s$  and  $q_{ij}$  constitute a patch pair, and all patch pairs of NNs construct a set of linear equations which is solved using weighted least squares to obtain an updated kernel.



**Fig. 2** Blurring effect on non-smooth and smooth areas. Black boxes indicate structured areas, and red boxes indicate smooth areas. (a) Clean patches. It can be clearly seen that the structure of non-smooth patch is distinct. (b) Blurred patches corresponding to (a). The detail of non-smooth patch is obviously blurry.

## 2.3 ILS-DLA and ANR

ILS-DLA is a typical dictionary learning method. It adopts an overall optimization strategy based on least square (LS) to update the dictionary when the weight matrix is fixed so that ILS-DLA [26] is usually faster than K-SVD [17, 27]. Besides, ANR just adjusts the objective function slightly and the SR reconstruction process is theoretically based on least square method.

Supposing we have two coupled feature sets  $\mathbf{F}_L$  and  $\mathbf{F}_H$  with size  $n_L \times L$  and  $n_H \times L$ , and the number of the atoms in LR dictionary  $\mathbf{D}_L$  and HR dictionary  $\mathbf{D}_H$  is  $K$ . The training process for  $\mathbf{D}_L$  can be described as

$$\{\hat{\mathbf{D}}_L, \hat{\mathbf{W}}\} = \underset{\mathbf{D}_L, \mathbf{W}}{\operatorname{argmin}} \sum_{i=1}^L \|\mathbf{w}_i\|_p + \lambda \|\mathbf{F}_L - \mathbf{D}_L \mathbf{W}\|_2^2, \\ \text{s.t. } \|\mathbf{d}_L^i\|_2^2 = 1 \quad (1)$$

where  $\mathbf{d}_L^i$  is an atom in  $\mathbf{D}_L$ ,  $\mathbf{w}_i$  is a column vector in  $\mathbf{W}$ .  $p \in \{0, 1\}$  is the constrain of the coefficient vector  $\mathbf{w}_i$ , and  $\lambda$  is a tradeoff parameter. Equation (1) is usually resolved by optimizing one variable while keeping the other one fixed. In ILS-DLA case, least square method is used to update  $\mathbf{D}_L$  while  $\mathbf{W}$  is fixed. Once  $\mathbf{D}_L$  and  $\mathbf{W}$  were obtained then we could

compute the  $\mathbf{D}_H$  according to the same LS rule:

$$\mathbf{D}_H = \mathbf{F}_H \mathbf{W}^T (\mathbf{W} \mathbf{W}^T)^{-1} \quad (2)$$

According to the philosophy of ANR, a mapping matrix can be calculated through the weight matrix and the both dictionaries. Then, it is used to project LR feature patches to HR feature patches directly. Thus,  $L_0/L_1$  norm constrained optimization problem degenerates to an issue of matrix multiplication.

### 3 Proposed approach

#### 3.1 Improved blur kernel estimation

Referring to Fig. 1, we use  $\mathbf{Y}$  to represent the input LR image, and  $\mathbf{X}$  to be latent HR image. Michaeli and Irani [25] estimated the blur kernel through maximizing the cross-scale NLSS directly, while we minimized the dissimilarity between cross-scale patches. Despite these two ideas look like the same with each other intuitively, they are slightly different and lead to severely different performance [16]. While Ref. [16] has introduced this part of content in detail, we need to present the key component of the improved blur kernel estimation for integrated elaboration. The following objective function has reflected the idea of minimizing the dissimilarity between cross-scale patches:

$$\underset{\mathbf{k}}{\operatorname{argmin}} \sum_{i=1}^N \left\| \mathbf{p}_i - \sum_{j=1}^{M_i} z_{ij} \mathbf{R}_{ij} \mathbf{k} \right\|_2^2 + \eta \|\mathbf{Ck}\|_2^2 \quad (3)$$

where  $N$  is the number of query patches in  $\mathbf{Y}$ . Matrix  $\mathbf{R}_{ij}$  corresponds to the operation of convolving with  $\mathbf{q}_{ij}$  and down-sampling by  $s$ .  $\mathbf{C}$  is a matrix used as the penalty of non-smooth kernel. The second term of Eq. (3) is kernel prior and  $\eta$  is the balance parameter as the tradeoff between the error term and kernel prior. For the calculation of the weight  $z_{ij}$ , we can find  $M_i$  NNs in down-sampled version  $\mathbf{Y}^s$  for each small patch  $\mathbf{p}_i$  ( $i = 1, 2, \dots, N$ ) in the input image  $\mathbf{Y}$ . The “parent” patches  $\mathbf{q}_{ij}$  right above  $\mathbf{q}_{ij}^s$  are viewed as the candidate parent patches of  $\mathbf{p}_i$ . Then the weight  $z_{ij}$  can be calculated as follow:

$$z_{ij} = \frac{\exp(-\|\mathbf{p}_i - \mathbf{q}_{ij}^s\|^2 / \sigma^2)}{\sum_{j=1}^{M_i} \exp(-\|\mathbf{p}_i - \mathbf{q}_{ij}^s\|^2 / \sigma^2)} \quad (4)$$

where  $M_i$  is the number of NNs in  $\mathbf{Y}^s$  of each small patch  $\mathbf{p}_i$  in  $\mathbf{Y}$ , and  $\sigma$  is the standard deviation of noise added on  $\mathbf{p}_i$ .  $s$  is the scale factor (see

Fig. 1). Note that we apply the same symbol to express column vector corresponding to the patch here. Setting the gradient of the objective function in Eq. (3) to zero can get the update formula of  $\mathbf{k}$ :

$$\hat{\mathbf{k}} = \left( \sum_{i=1}^N \sum_{j=1}^{M_i} z_{ij} \mathbf{R}_{ij}^T \mathbf{R}_{ij} + \eta \mathbf{C}^T \mathbf{C} \right)^{-1} \sum_{i=1}^N \sum_{j=1}^{M_i} z_{ij} \mathbf{R}_{ij}^T \mathbf{p}_i \quad (5)$$

Equation (5) is very similar to the result of Ref. [25], which can be interpreted as maximum a posteriori (MAP) estimation on  $\mathbf{k}$ . However, there are at least three essential differentials with respect to Ref. [25]. Firstly, the motivation is different so that Ref. [25] tends to maximize the cross-scale similarity according to NLSS [12] while we minimize the dissimilarity directly according to Ref. [18]. This may not be easy to understand. However, the former leads Michaeli and Irani [25] to form their kernel update formula from physical analysis and interpretation of “optimal kernel”. The latter leads us to obtain kernel update formula from quantitating cross-scale patch dissimilarity and directly minimizing it according to ridge regression [16]. Secondly, selective patch processing measured by the average gradient amplitude  $|\text{grad}|$  was adopted to improve the result of blind BKE. Finally, the number of NNs of each small patch  $\mathbf{p}_i$  is not fixed which provides more flexibility during solving least square problem. Accordingly, the terminal criterion cannot be the totality of NNs. We use the average patch dissimilarity (APD) as terminal condition of iteration:

$$\text{APD} = \sum_{i=1}^N \sum_{j=1}^{M_i} \left\| \mathbf{p}_i - \mathbf{q}_{ij}^s \right\|_2^2 \cdot \left( \sum_{i=1}^N M_i \right)^{-1} \quad (6)$$

It is worth to note that selective patch processing is used to eliminate the effect on BKE caused by smooth patches; we selectively employ structured patches to calculate blur kernel. Specifically, if the average gradient magnitude  $|\text{grad}|$  of each query patch is smaller than a threshold, then we abandon it. Otherwise, we use it to estimate blur kernel according to Eq. (5). We typically perform search in the entire image according to Ref. [18] but this could not consume too much time because of a lot of smooth patches being filtered out.

#### 3.2 Feature extraction strategy

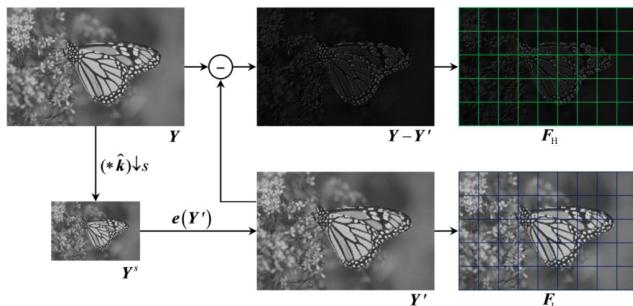
There is a data preparation stage before dictionary learning when using sparse representation to do SR

task, it is necessary to extract training features from the given input data because different feature extraction strategies will cause very different SR results. The mainstream feature extraction strategies include raw data of an image patch, the gradient of an image patch in  $x$  and  $y$  directions, and mean-removed patch etc. We adopt the back-projection residuals (BPR) model presented in Ref. [28] for feature extraction (see Fig. 3).

Firstly, we convolve  $\mathbf{Y}$  with estimated kernel  $\mathbf{k}$ , and down-sample it with  $s$ . From the view of minimizing the cross-scale patch dissimilarity, the estimated blur kernel gives us a more accurate down-sampling version of  $\mathbf{Y}$ . In order to make the feature extraction more accurate, we consider the enhanced interpolation of  $\mathbf{Y}'$ , which forms the source of LR feature space  $\mathbf{F}_L$ . The enhanced interpolation is the result of an iterative back-projection (IBP) operation [29, 30]:

$$\hat{\mathbf{Y}}'_{t+1} = \hat{\mathbf{Y}}'_t + [(\mathbf{Y}^s - \hat{\mathbf{Y}}_t^s) \uparrow s] * \mathbf{k}' \quad (7)$$

where  $\hat{\mathbf{Y}}_t^s = (\hat{\mathbf{Y}}'_t * \mathbf{k}) \downarrow s$ ,  $\mathbf{k}'$  is a back-projection filter that spreads the differential error locally. It is usually replaced by a Gaussian function. The IBP starts with the bicubic interpolation, and down-sampling operation is performed by convolving with estimated blur kernel  $\mathbf{k}$ . After a certain number of iterations, the error between  $\mathbf{Y}^s$  and  $\hat{\mathbf{Y}}_t^s$  will be sufficiently small so that the enhanced version of  $\mathbf{Y}'$  is adequately consistent with  $\mathbf{Y}^s$ . The HR feature space  $\mathbf{F}_H$  is obtained by extracting raw patches from residuals image  $\mathbf{Y} - \mathbf{Y}'$ . In fact, the back-projection residual image represents the high-pass filtered version of  $\mathbf{Y}$ . It contains essential high frequencies



**Fig. 3** Feature extraction strategy.  $e(\cdot)$  represents the enhanced interpolation operation. The down-sampled version  $\mathbf{Y}^s$  is obtained by convolving with estimated kernel  $\hat{\mathbf{k}}$  and down-sampling with  $s$ . LR feature set consists of normalized gradient feature patch extracted from  $\mathbf{Y}'$ , HR feature set is made up with the raw patches extracted from BPR image  $\mathbf{Y} - \mathbf{Y}'$ .

of  $\mathbf{Y}$ , and this is substantially helpful for dictionary training. Besides, the directly down-sampled version  $\mathbf{Y}^s$  with the estimated kernel usually is inconsistent with  $\mathbf{Y}$ . The enhanced interpolation process gives an effective adjustment via reducing local projection error. When both LR and HR feature patches get prepared, we use ILS-DLA algorithm to train our coupled dictionaries presented in Ref. [26] for fast training and unified optimization rule, see Section 2.3.

### 3.3 SR recovery via ASM

Yang et al. [13] accelerated the SR process from two directions: reducing the number of patches and finding a fast solver for  $L_1$  norm minimization problem. We adopt a similar manner for the first optimization direction, i.e., a selective patch process (SPP) strategy. However, in order to be consistent with BKE, the criterion of selecting patches is the gradient magnitude  $|\text{grad}|$  instead of the variance. The second direction Yang et al. headed to is learning a feed-forward neural network model to find an approximated solution for  $L_1$  norm sparse encoding. We employ ASM to accelerate the algorithm similar with Ref. [17]. It requires us to reformulate the  $L_1$  norm minimization problem as a least square regression regularized by  $L_2$  norm of sparse representation coefficients, and adopt the ridge regression (RR) to relieve the computationally demanding problem of  $L_1$  norm optimization. The problem then comes to be

$$\underset{\mathbf{w}}{\text{argmin}} \|\mathbf{y} - \mathbf{D}_L \mathbf{w}\|_2^2 + \mu \|\mathbf{w}\|_2 \quad (8)$$

where the parameter  $\mu$  allows alleviating the ill-posed problem and stabilizes the solution.  $\mathbf{y}$  corresponds to a testing patch extracted from enhanced interpolation version of input image.  $\mathbf{D}_L$  is the LR dictionary trained by ILS-DLA. The algebraic solution of Eq. (8) is given by setting the gradient of objective function to zero, which gives:

$$\mathbf{w} = (\mathbf{D}_L^T \mathbf{D}_L + \mu \mathbf{I})^{-1} \mathbf{D}_L^T \mathbf{y} \quad (9)$$

where  $\mathbf{I}$  is a identity matrix. Then, the same coefficients are used on the HR feature space to compute the latent HR patches, i.e.,  $\mathbf{x} = \mathbf{D}_H \mathbf{w}$ . Combined with Eq. (9):

$$\mathbf{x} = \mathbf{D}_H (\mathbf{D}_L^T \mathbf{D}_L + \mu \mathbf{I})^{-1} \mathbf{D}_L^T \mathbf{y} = \mathbf{P}_M \mathbf{y} \quad (10)$$

where mapping matrix  $\mathbf{P}_M$  can be computed offline and  $\mathbf{D}_H$  is computed by Eq. (2). Equation (10) means HR feature patch can be obtained by LR

patch multiplying with a projection matrix directly, which reduces the time consumption tremendously in practice. Moreover, the feature patches needed to be mapped to HR features via  $\mathbf{P}_M$  will be further reduced due to SPP. Though the optimization problem constrained by  $L_2$  norm usually leads to a more relaxative solution, it still yields very accurate SR results because of cross-scale BKE.

## 4 Experimental results

All the following experiments are performed on the same platform, i.e., a Philips 64 bit PC with 8.0 GB memory and running a single core of Intel Xeon 2.53 GHz CPU. The core differences between the proposed method and Ref. [16] are the feature extraction and SR recovery. The former is mainly aiming at reducing local projection error and improving the quality of the training samples further. The latter is primarily used to accelerate the reconstruction of the latent HR image.

### 4.1 Experiment settings

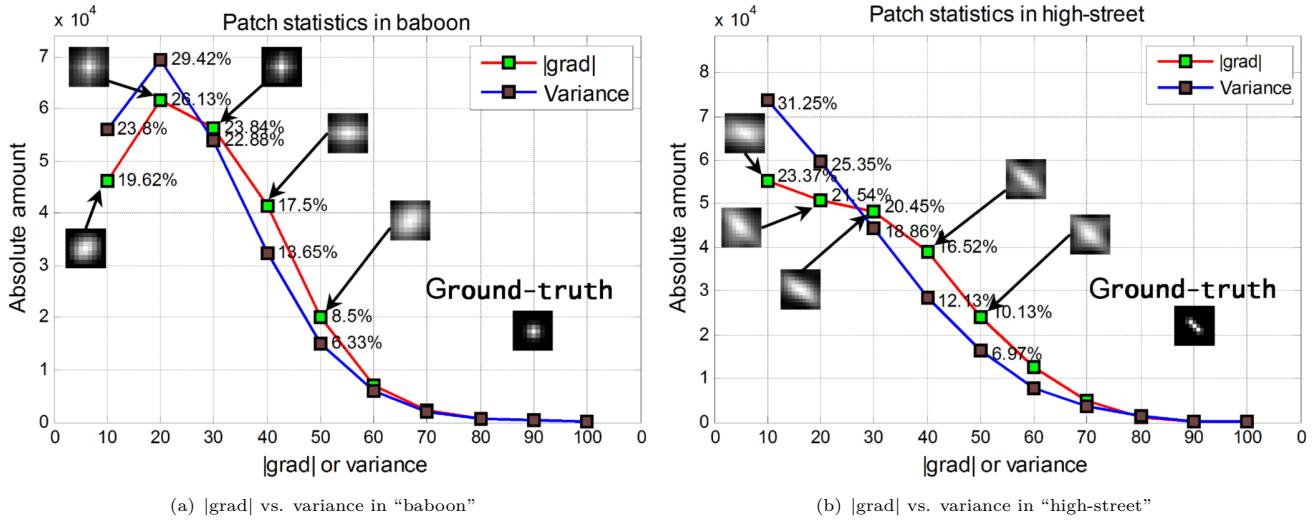
We quintessentially perform  $\times 2$  and  $\times 3$  SR in our experiments on blind BKE. The parameter settings in BKE stage are partially the same with Refs. [16] and [25], i.e., when scale factor  $s = 2$ , the size of small query patches  $\mathbf{p}_i$  and candidate patches  $\mathbf{q}_{ij}^s$  of NNs are typically set to  $5 \times 5$ , while the sizes of “parent” patches  $\mathbf{q}_{ij}$  are set to  $9 \times 9$  and  $11 \times 11$ ; when performing  $\times 3$  SR, query patches and candidate patches do not change size but “parent” patches are set to be  $13 \times 13$  patches. Noise standard deviation  $\sigma$  is assumed to be 5. Parameter  $\eta$  in Eq. (3) is set to be 0.25, and matrix  $\mathbf{C}$  is chosen to be the derivative matrix corresponding to  $x$  and  $y$  direction of “parent” patches. The threshold of gradient magnitude  $|\text{grad}|$  for selecting query patches varies in 10–30 according to the different images. In the processing of feature extraction, the enhanced interpolation starts with the bicubic interpolation and down-sampling operation is performed by convolving with estimated blur kernel  $\mathbf{k}$ , and the back-projection filter  $\mathbf{k}'$  is set to be a Gaussian kernel with the same size of  $\mathbf{k}$ . The tradeoff parameter  $\mu$  in Eq. (8) is set to be 0.01 and the number of iteration for dictionary training is 20.

### 4.2 Analysis of the metric in blind BKE

Comparisons for blind BKE usually include the accuracy of the estimated kernels and the efficiency, and these two elaborations have been presented in our previous work [16] in detail. We intend to analyze the impact on blind BKE through discriminating the query patches instead of simply comparing the final results with some related works. The repeated conclusions will be ignored here. We collected patches from three natural image sets (Set2, Set5, and Set14) and found that the values of  $|\text{grad}|$  and variance mostly fall into the range of [0, 100]. So the entire statistical range is set to be [0, 100] and the statistical interval for  $|\text{grad}|$  and variance is typically set to be 10.

We sampled the  $500 \times 400$  “baboon” image and got 236,096 query patches, and 235,928 patches from  $540 \times 300$  “high-street” (dense sampling). It is distinctly observed that the statistical characteristics of  $|\text{grad}|$  and variance are very similar with each other in Fig. 4. The query patches with threshold  $\leq 30$  account for the most proportion for both  $|\text{grad}|$  and variance, and we could get similar conclusion from other images. However, the relative relation between them reverses around 30 (value may be different from images but the reverse determinately exists). This is an intuitive presentation that why we adopt the  $|\text{grad}|$  instead of variance as the metric of selecting patches based on the philosophy of dropping the useless smooth patches as many as possible and keeping the structured patches as far as possible. More systematically theoretical explanations could be found in Ref. [18].

Moreover, the performance of blind BKE is obviously affected by the threshold on  $|\text{grad}|$ . The optimal kernel was pinned beside the threshold in Fig. 4. We can see that the estimated kernel by our method is not close to the ground-truth one infinitely as the threshold increasing because the useful structured patches reduce as well. Usually, the estimated kernel at the threshold of “turning point” is closest to the ground-truth. When the threshold is set to be 0, it actually degenerates to the algorithm of Ref. [25], which does not give the best result in most instances. In general case, the quality of recovery declines with the increase of  $|\text{grad}|$  like the second illustration in Fig. 4. But there indeed exist special cases like the first illustration in Fig. 4 for the PSNRs



**Fig. 4** Comparisons between statistical characteristics and threshold effect on estimated kernels. The testing images are “baboon” from Set14 and “high-street” from Set2, and the blur kernels are a Gaussian with  $hsize = 5$  and  $sigma = 1.0$  ( $9 \times 9$ ) and a motion kernel with  $length = 5$ ,  $theta = 135$  ( $11 \times 11$ ) respectively. We only display estimated kernels when threshold on  $|\text{grad}| \leq 50$ .

of recovered images rise firstly and then fell with the threshold increasing.

### 4.3 Comparisons for SR recovery

Compared with several currently proposed methods (such as A+ ANR [27] and SRCNN [31] etc.), the reconstruction efficiency of our method sometimes is slightly low but almost in the same magnitude. Due to the anchored space mapping, the proposed method was accelerated substantially with regard to some typical sparse representation algorithms like Refs. [11] and [32]. Table 1 and Table 2 present the quantitative comparisons using PSNRs and SR recovery time to compare objective index and SR efficiency. Four recent proposed methods including Ref. [32], A+ ANR (adjusted anchored neighborhood regression) [27], SRCNN (super-resolution convolutional neural network) [31], and JOR (jointly optimized regressors) [33] are picked out as the representative of non-blind methods (presented in Table 1), and three blind methods, NBS (nonparametric blind super-resolution) [25], SAR-NBS (simple, accurate, and robust nonparametric blind super-resolution) [34], and Ref. [16] are concurrently listed in Table 2 with the proposed method. It’s worth noting that PSNR needs reference images as base line. Because the input images are blurred by different blurring kernels so that the observation data is seriously degenerated and non-blind methods usually give very bad results in this case, we referred the recovered

images to the blurred input images. The average PSNRs and running time are collected ( $\times 2$  and  $\times 3$ ) over four image sets (Set2, Set5, Set14, and B100). Besides, we set the threshold on  $|\text{grad}|$  around to the “turning point” adaptively for the best BKE estimation instead of pinning it to a fixed number (e.g., 10 in Ref. [16]). The methods listed in Table 1 and Table 2 are identical to the methods presented in Figs. 5–8.

As shown in Table 1 and Table 2, the proposed algorithm obtained higher objective evaluation than other blind or non-blind algorithms in both  $s = 2$  and/or  $s = 3$  case. For fairness, it excludes the time of preparing data and training dictionaries for all of these methods. Firstly, four non-blind methods in Table 1 fail to recover real images when the inputs are degraded seriously though some of them provided very high speed. And, both the accuracy and efficiency of estimating kernels via Michaeli et al. [25] are not high enough which has been illustrated in Ref. [16], and SR recovery performed by Ref. [11] is very time-consuming. While the same process of BKE with us was executed in Ref. [16] (fixed threshold on  $|\text{grad}|$ ), the SR reconstruction with SPP is essentially still low-efficiency. The proposed method adopted adaptive  $|\text{grad}|$  threshold to improve the quality of BKE and the enhanced interpolation on input images reduced the mapping errors brought by estimated kernels further. On the other hand, ASM increased the speed of the algorithm in essence. This is mainly

**Table 1** Performance of several non-blind methods (without estimating blur kernel)

Image set	Scale	Zeyde et al. ([32])		A+ ANR ([27])		SRCNN ([31])		JOR ([33])	
		PSNR (dB)	Time (s)	PSNR (dB)	Time (s)	PSNR (dB)	Time (s)	PSNR (dB)	Time (s)
Set2	$\times 2$	30.4314	9.0178	30.6207	2.2749	30.5952	2.0517	30.6334	8.3200
Set5		30.4607	7.1648	31.1095	1.3894	31.1021	1.2044	31.1252	7.8399
Set14		30.8179	11.4795	31.1023	2.3714	31.0683	2.1742	31.1214	9.2874
B100		30.5784	8.4876	31.1007	1.5497	31.0271	1.4263	31.1098	8.1431
Set2	$\times 3$	28.0296	6.1883	28.2544	1.7831	28.2527	1.6397	28.2605	8.4166
Set5		28.1548	4.1546	28.3327	1.1173	28.3314	0.8564	28.3397	7.4879
Set14		28.4706	8.6470	28.6147	1.9406	28.6107	1.7867	28.6244	8.8371
B100		28.2381	6.0365	28.5836	1.5733	28.5694	1.5049	28.6042	8.2165

**Table 2** Performance of several blind methods (adaptive threshold on  $|\text{grad}|$ )

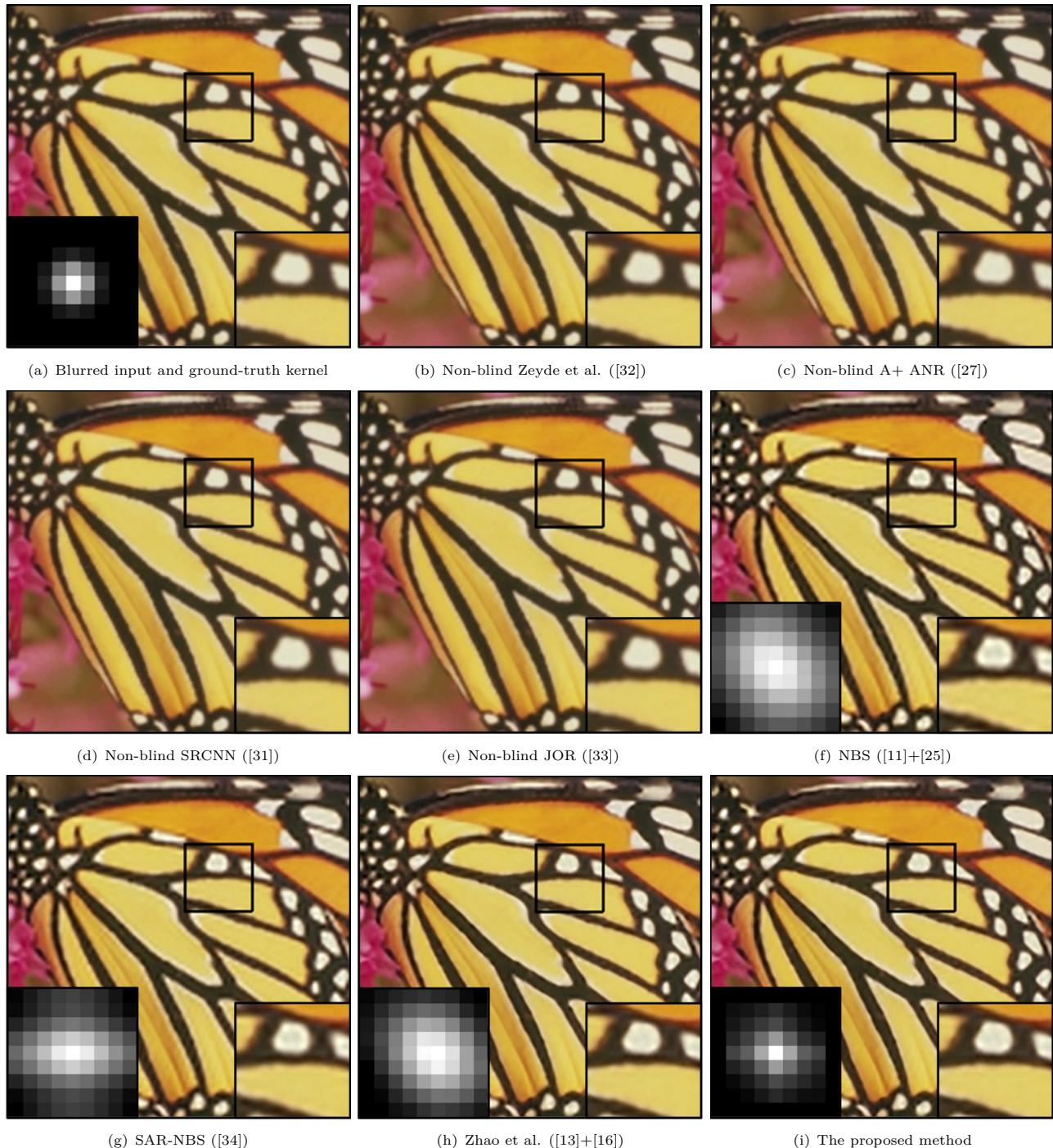
Image set	Scale	NBS ([11] + [25])		SAR-NBS ([34])		Zhao et al. ([13] + [16])		Ours	
		PSNR (dB)	Time (s)	PSNR (dB)	Time (s)	PSNR (dB)	Time (s)	PSNR (dB)	Time (s)
Set2	$\times 2$	31.1417	126.7434	31.4371	—	31.8745	25.8316	32.1795	2.0476
Set5		31.4734	97.4641	31.7217	—	32.0146	19.7842	32.3914	1.2078
Set14		31.3043	108.7918	31.5841	—	31.9243	26.7849	31.9167	2.1607
B100		31.4971	102.4352	31.7638	—	31.9673	22.4876	32.2671	1.3844
Set2	$\times 3$	29.2549	98.4870	29.8749	—	30.3498	21.4876	30.6648	1.5743
Set5		29.3631	86.9237	29.8379	—	30.4716	15.9829	30.5176	1.0748
Set14		29.2461	95.4881	29.6477	—	29.9831	24.9573	30.2472	1.8476
B100		28.9771	91.4877	29.6942	—	30.2752	19.7481	30.3317	1.4977

due to the adjustment of the objective function and the constraint conversion from  $L_0/L_1$  norm to  $L_2$  norm. Actually, the improvement of our method is not only reflected in SR recovery stage, but also reflected in BKE (through SPP) and dictionary training (through ILS-DLA) which are not usually mainly concerned by most of researchers. But these preprocessing procedures are still very different when big data need to be processed.

Figures 5–8 show the visual comparisons between several typical SR algorithms and our method. For layout purpose, all images are diminished when inserted in the paper. Still note that the input images of all algorithms are obtained through reference images blurred with different blur kernel. Namely, input image data is set to be of low quality in our experiments for the sake of simulating many actual application scenarios. Though it is well known that non-blind SR algorithms presented in the illustrations are efficient for many SR tasks, they fail to offset the blurring effect in testing images without a more precise blurring kernel. There is also significant difference about the estimated kernels and reconstruction results among blind algorithms. The BKE process of Ref. [25] is actually close to our method when the threshold on  $|\text{grad}|$  is 0 and the criterion for iteration is

MSE. Shao et al. [34] solved the blur kernel and HR image simultaneously by minimizing a bi- $L_0-L_2$ -norm regularized optimization problem with respect to both an intermediate super-resolved image and a blur kernel, which is extraordinarily time-demanding (so not shown in Table 1). More important is that the fitting problem caused by useless smooth patches still exists in these methods. Although the idea of our method is simple, it could avoid the fitting problem by dropping the smooth query patches reasonably according to the internal statistics of a single natural image.

Figures 9 and 10 present SR recovery results of other two real images (“fence” and “building”), which were captured by our cell-phone with slightly joggle and motion. It is easily noticed that all blind methods produce a better result than non-blind methods which even can not offset the motion deviation basically. Comparing Fig. 9(f)–Fig. 9(i) and Fig. 10(f)–Fig. 10(i), we can find the visible difference in estimated kernels and recovered results produced by different blind methods. Particularly, SAR-NBS tends to overly sharpen the high frequency region and gives obvious distortion in final images. The results of Zhao et al. [16] look more realistic but the reconstruction accuracy is not high enough compared with our approach.



**Fig. 5** Visual comparisons of SR recovery with low-quality “butterfly” image from Set5 ( $\times 2$ ). The ground-truth kernel is a  $9 \times 9$  Gaussian kernel with  $hsize = 5$  and  $sigma = 1.25$ , and threshold on  $|\text{grad}|$  is 18.

## 5 Conclusions

We proposed a novel single image SR processing framework aiming at improving the SR effect and reducing SR time consumption in this paper. The proposed algorithm mainly consists of blind blur

kernel estimation and SR recovery. The former is based on the idea of minimizing dissimilarity of cross-scale image patches, which leads us to obtain kernel update formula by quantitating cross-scale patch dissimilarity and directly minimizing it according to least square method. The reduction of SR time mainly relies on an ASM process



**Fig. 6** Visual comparisons of SR recovery with low-quality “high-street” image from Set2 ( $\times 3$ ). The ground-truth kernel is a  $13 \times 13$  motion kernel with  $len = 5$  and  $theta = 45$ , and threshold on  $|grad|$  is 24.

with LR/HR dictionaries trained by ILS-DLA algorithms a selective patch processing strategy measured by  $|grad|$ . Therefore, the SR effect is mainly guaranteed by improving the quality of training samples and the efficiency of SR recovery is mainly guaranteed by anchored space mapping and selective patch processing. They ensure the improvement of time performance via reducing the number of query patches and translating  $L_1$  norm constrained optimization problem into  $L_2$  norm constrained anchor mapping process. Under the equal conditions, all above-mentioned processes make our SR algorithm achieve better results than several outstanding blind and non-blind SR approaches proposed previously with a much higher speed.

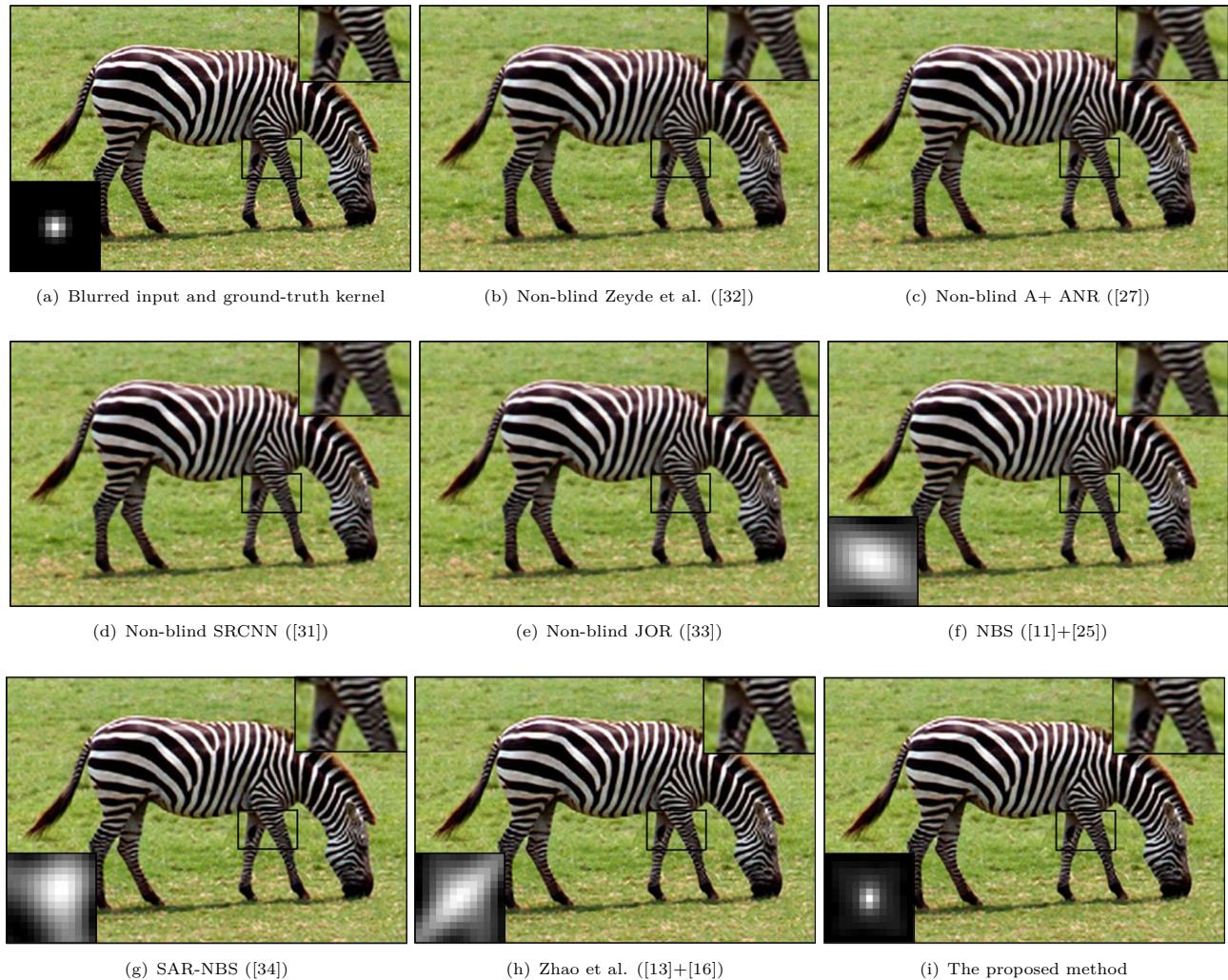
## Acknowledgements

We would like to thank the authors of Ref. [34], Mr. Michael Elad and Mr. Wen-Ze Shao, for their

kind help in running their blind SR method [34], which thus enables an effective comparison with their method. This work is partially 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 and 2015GZ0212), Key Program of Education Department of Sichuan Province (Grant Nos. 11ZA130 and 13ZA0169), and the innovation funds of Southwest University of Science and Technology (Grant No. 15ycx053).

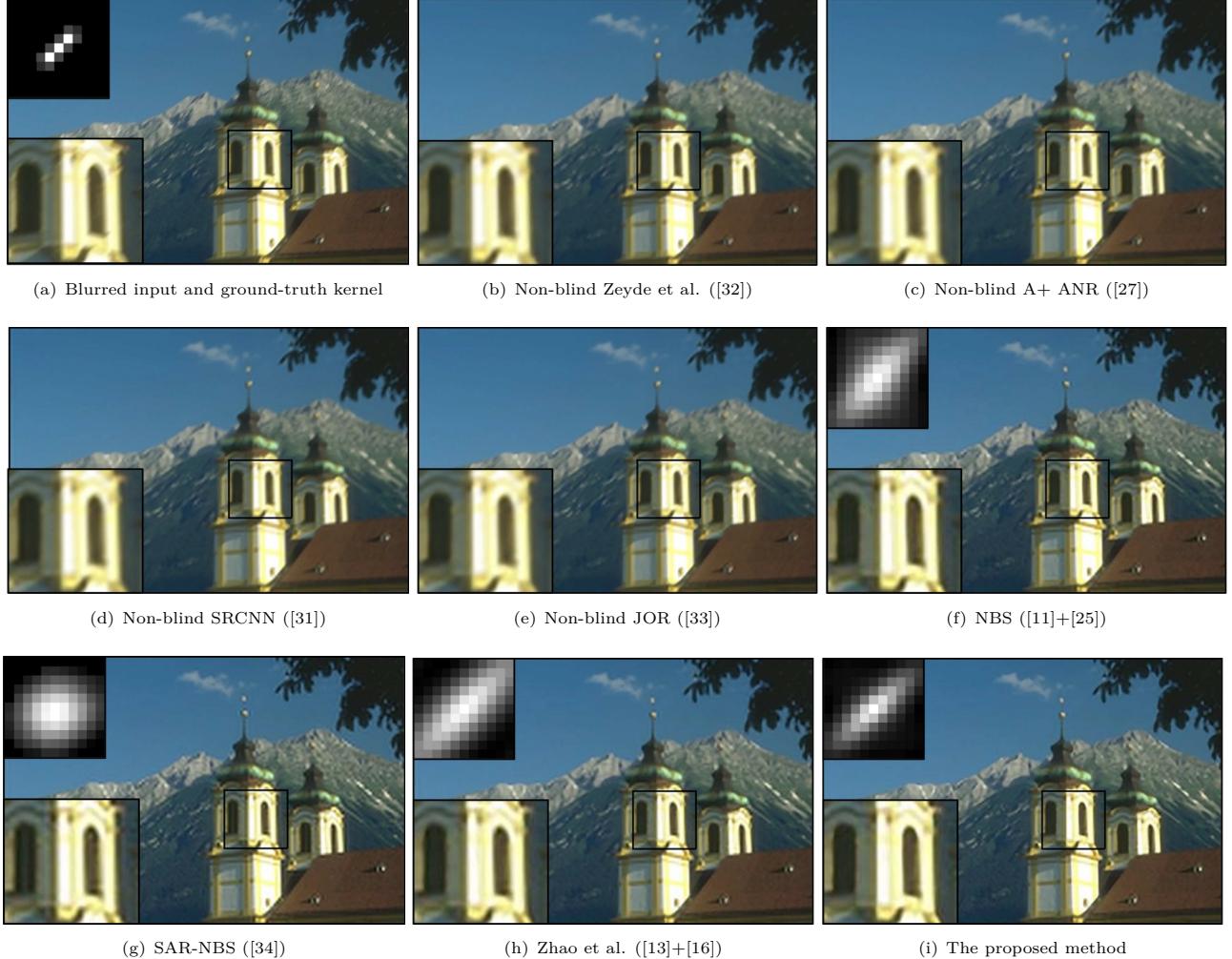
## References

- [1] Freedman, G.; Fattal, R. Image and video upscaling from local self-examples. *ACM Transactions on Graphics* Vol. 30, No. 2, Article No. 12, 2011.
- [2] Babacan, S. D.; Molina, R.; Katsaggelos, A. K. Parameter estimation in TV image restoration



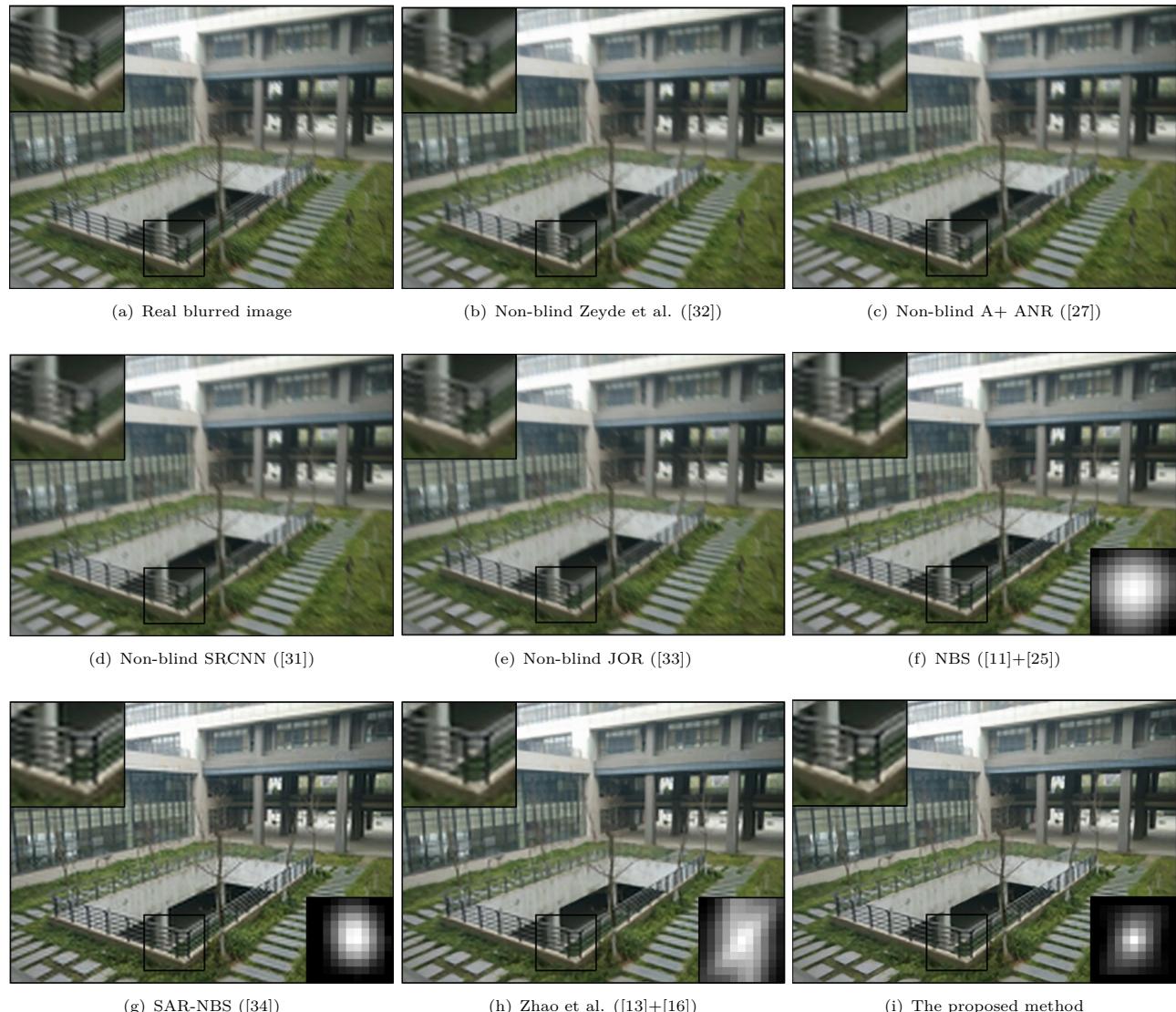
**Fig. 7** Visual comparisons of SR recovery with low-quality “zebra” image from Set14 ( $\times 3$ ). The ground-truth kernel is a  $13 \times 13$  Gaussian kernel with  $hsize = 5$  and  $theta = 1.25$ , and threshold on  $|\text{grad}|$  is 16.

- using variational distribution approximation. *IEEE Transactions on Image Processing* Vol. 17, No. 3, 326–339, 2008.
- [3] Babacan, S. D.; Molina, R.; Katsaggelos, A. K. Total variation super resolution using a variational approach. In: Proceedings of the 15th IEEE International Conference on Image Processing, 641–644, 2008.
  - [4] Babacan, S. D.; Molina, R.; Katsaggelos, A. K. Variational Bayesian super resolution. *IEEE Transactions on Image Processing* Vol. 20, No. 4, 984–999, 2011.
  - [5] Sun, J.; Xu, Z.; Shum, H.-Y. Image super-resolution using gradient profile prior. In: Proceedings of IEEE Conference on Computer Vision Pattern Recognition, 1–8, 2008.
  - [6] Freeman, W. T.; Jones, T. R.; Pasztor, E. C. Example-based super-resolution. *IEEE Computer Graphics and Applications* Vol. 22, No. 2, 56–65, 2002.
  - [7] Sun, J.; Zheng, N.-N.; Tao, H.; Shum, H.-Y. Image hallucination with primal sketch priors. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 729–736, 2003.
  - [8] Chang, H.; Yeung, D. Y.; Xiong, Y. Super-resolution through neighbor embedding. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 275–282, 2004.
  - [9] Roweis, S. T.; Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science* Vol. 290, No. 5, 2323–2326, 2000.
  - [10] Bevilacqua, M.; Roumy, A.; Guillemot, C.; Morel, M.-L. A. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: Proceedings of the 23rd British Machine Vision Conference, 135.1–135.10, 2012.
  - [11] Yang, J.; Wright, J.; Huang, T.; Ma, Y. Image super-resolution as sparse representation of raw image patches. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1–8, 2008.
  - [12] Glasner, D.; Bagon, S.; Irani, M. Super-resolution from a single image. In: Proceedings of IEEE 12th



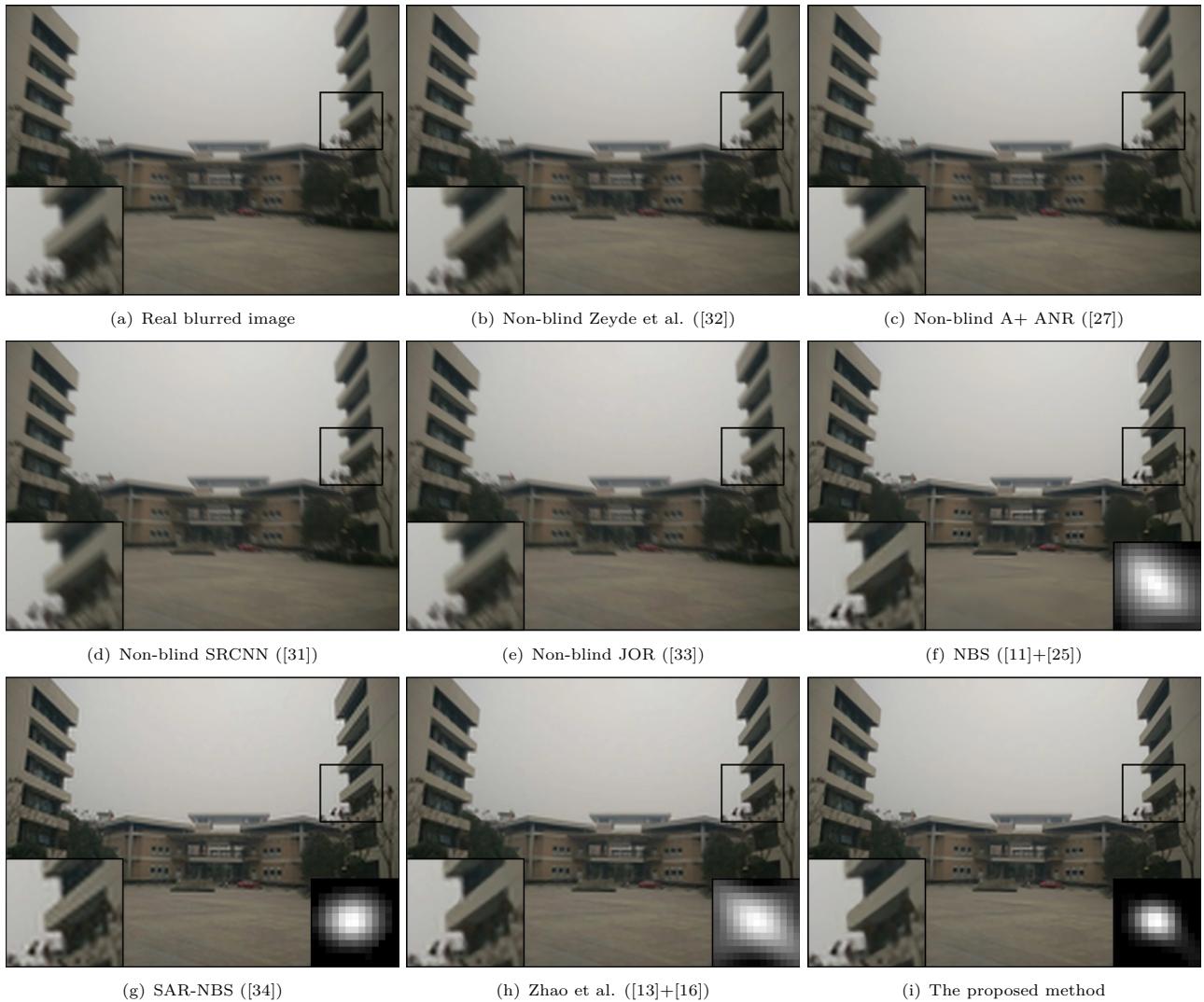
**Fig. 8** Visual comparisons of SR recovery with low-quality “tower” image from B100 ( $\times 2$ ). The ground-truth kernel is a  $11 \times 11$  motion kernel with  $len = 5$  and  $tetheta = 45$ , and threshold on  $|\text{grad}|$  is 17.

- International Conference on Computer Vision, 349–356, 2009.
- [13] Yang, J.; Wang, Z.; Lin, Z.; Cohen, S.; Huang, T. Coupled dictionary training for image super-resolution. *IEEE Transactions on Image Processing* Vol. 21, No. 8, 3467–3478, 2012.
  - [14] Yang, J.; Wright, J.; Huang, T.; Ma, Y. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing* Vol. 19, No. 11, 2861–2873, 2010.
  - [15] He, L.; Qi, H.; Zaretzki, R. Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 345–352, 2013.
  - [16] Zhao, X.; Wu, Y.; Tian, J.; Zhang, H. Single image super-resolution via blind blurring estimation and dictionary learning. In: *Communications in Computer and Information Science*, Vol. 546. Zha, H.; Chen, X.; Wang, L.; Miao, Q. Eds. Springer Berlin Heidelberg, 22–33, 2015.
  - [17] Timofte, R.; De, V.; Van Gool, L. Anchored neighborhood regression for fast example-based super-resolution. In: Proceedings of IEEE International Conference on Computer Vision, 1920–1927, 2013.
  - [18] Zontak, M.; Irani, M. Internal statistics of a single natural image. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 977–984, 2011.
  - [19] Yang, C.-Y.; Huang, J.-B.; Yang, M.-H. Exploiting self-similarities for single frame super-resolution. In: *Lecture Notes in Computer Science*, Vol. 6594. Kimmel, R.; Klette, R.; Sugimoto, A. Eds. Springer Berlin Heidelberg, 497–510, 2010.
  - [20] Zoran, D.; Weiss, Y. From learning models of natural image patches to whole image restoration. In: Proceedings of IEEE International Conference on Computer Vision, 479–486, 2011.
  - [21] Hu, J.; Luo, Y. Single-image superresolution based on local regression and nonlocal self-similarity. *Journal of Electronic Imaging* Vol. 23, No. 3, 033014, 2014.



**Fig. 9** Visual comparisons of SR recovery with real low-quality image “fence” captured with slight joggle ( $\times 2$ ). Threshold on  $|\text{grad}|$  is 21.

- [22] Zhang, Y.; Liu, J.; Yang, S.; Guo, Z. Joint image denoising using self-similarity based low-rank approximations. In: *Proceedings of Visual Communications and Image Processing*, 1–6, 2013.
- [23] Michaeli, T.; Irani, M. Blind deblurring using internal patch recurrence. In: *Lecture Notes in Computer Science*, Vol. 8691. Fleet, D.; Pajdla, T.; Schiele, B.; Tuytelaars, T. Eds. Springer International Publishing, 783–798, 2014.
- [24] Guillemot, C.; Le Meur, O. Image inpainting: Overview and recent advances. *IEEE Signal Processing Magazine* Vol. 31, No. 1, 127–144, 2014.
- [25] Michaeli, T.; Irani, M. Nonparametric blind super-resolution. In: *Proceedings of IEEE International Conference on Computer Vision*, 945–952, 2013.
- [26] Engan, K.; Skretting, K.; Husøy, J. H. Family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation. *Digital Signal Processing* Vol. 17, No. 1, 32–49, 2007.
- [27] Timofte, R.; De Smet, V.; Van Gool, L. A+: Adjusted anchored neighborhood regression for fast super-resolution. In: *Lecture Notes in Computer Science*, Vol. 9006. Cremers, D.; Reid, I.; Saito, H.; Yang, M.-H. Eds. Springer International Publishing, 111–126, 2014.
- [28] Bevilacqua, M.; Roumy, A.; Guillemot, C.; Morel, M.-L. A. Super-resolution using neighbor embedding of back-projection residuals. In: *Proceedings of the 18th International Conference on Digital Signal Processing*, 1–8, 2013.
- [29] Irani, M.; Peleg, S. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation* Vol. 4, No. 4, 324–335, 1993.
- [30] Irani, M.; Peleg, S. Improving resolution by image registration. *CVGIP: Graphical Models and Image Processing* Vol. 53, No. 3, 231–239, 1991.

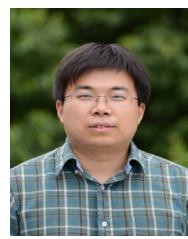


**Fig. 10** Visual comparisons of SR recovery with real low-quality image “building” captured with slight motion ( $\times 2$ ). Threshold on  $|\text{grad}|$  is 26.

- [31] Dong, C.; Chen, C. L.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In: *Lecture Notes in Computer Science*, Vol. 8692. Fleet, D.; Pajdla, T.; Schiele, B.; Tuytelaars, T. Eds. Springer International Publishing, 184–199, 2014.
- [32] Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations. In: *Lecture Notes in Computer Science*, Vol. 6920. Boissonnat, J.-D.; Chenin, P.; Cohen, A. et al. Eds. Springer Berlin Heidelberg, 711–730, 2010.
- [33] Dai, D.; Timofte, R.; Van Gool, L. Jointly optimized regressors for image super-resolution. *Computer Graphics Forum* Vol. 34, No. 2, 95–104, 2015.
- [34] Shao, W.-Z.; Elad, M. Simple, accurate, and robust nonparametric blind super-resolution. In: *Lecture Notes in Computer Science*, Vol. 9219. Zhang, Y.-J. Ed. Springer International Publishing, 333–348, 2015.



**Xiaole Zhao** received his B.S. degree in computer science from the School of Computer Science and Technology, Southwest University of Science and Technology (SWUST), China, in 2013. He is now studying in the School of Computer Science and Technology, SWUST for his master degree. His main research interests include digital image processing, machine learning, and data mining.



**Yadong Wu** now is a full professor with the School of Computer Science and Technology, Southwest University of Science and Technology (SWUST), China. He received his B.S. degree in computer science from Zhengzhou University, China, in 2000, and M.S. degree in control theory and control

engineering from SWUST in 2003. He got his Ph.D. degree in computer application from University of Electronic Science and Technology of China. His research interest includes image processing and visualization.



**Jinsha Tian** received her B.S. degree in Hebei University of Science and Technology, China, in 2013. She is now studying in the School of Computer Science and Technology, Southwest University of Science and Technology (SWUST), China for her master degree. Her main research interests include digital image processing and machine learning.



**Hongying Zhang** now is a full professor with the School of Information Engineering, Southwest University of Science and Technology (SWUST), China. She received her B.S. degree in applied mathematics from Northeast University, China, in 2000, and M.S. degree in control theory and control

engineering from SWUST in 2003. She got her Ph.D. degree in signal and information processing from University of Electronic Science and Technology of China, in 2006. Her research interest includes image processing and biometric recognition.

**Open Access** The articles published in this journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Other papers from this open access journal are available free of charge from <http://www.springer.com/journal/41095>. To submit a manuscript, please go to <https://www.editorialmanager.com/cvmj>.