

# Single-Image Super-Resolution via Adaptive Joint Kernel Regression

Chen Huang  
yach23@gmail.com  
Xiaoqing Ding  
dingxq@tsinghua.edu.cn  
Chi Fang  
fangchi@tsinghua.edu.cn

State Key Laboratory of Intelligent  
Technology and Systems,  
Tsinghua National Laboratory for  
Information Science and Technology,  
Department of Electronic Engineering,  
Tsinghua University, Beijing 100084,  
China

---

## Abstract

This paper proposes an adaptive joint kernel regression framework for single-image super-resolution (SR). The basic idea is to regularize the ill-posed reconstruction problem using a regression-based prior that exploits both local structural regularity and nonlocal self-similarity of natural images. To this end, we first generalize the nonlocal means method in the local kernel regression framework, and then extend such generalized regressors to the nonlocal range. Combining them into one single regularization term leads to a joint kernel regression scheme that simultaneously exploits both image statistics in a natural manner. We further propose a measure called regional redundancy to determine the confidence of these regression groups and thus control their relative effects of regularization adaptively. Adaptive dictionary learning and dictionary-based sparsity prior are also introduced to interact with the regression prior for robustness. Quantitative and qualitative results on SR show that our method outperforms other state-of-the-art methods, and can also be applied to other inverse problems such as image deblurring.

## 1 Introduction

Single image super-resolution (SR) refers to the task of estimating a high resolution (HR) image from a *single* low resolution (LR) image. SR methods can be broadly categorized into three classes: interpolation-based methods, reconstruction-based methods, and example-based methods. Interpolation techniques (e.g. [12]) are simple and fast but tend to blur the fine details. The reconstruction-based methods (e.g. [8, 16, 17, 23]) often incorporate prior knowledge and additional information with the reconstruction constraint to obtain effective solutions because SR is an ill-posed problem. Then how to design a good *image prior* is an essential issue, and image priors are usually formulated as constraints or regularization terms in the optimization function. The example-based methods (e.g. [7, 8, 9, 10, 21, 22]) hallucinate detailed textures from a training dictionary of LR/HR image or patch pairs. However, such methods strongly rely on the chosen dictionary for satisfactory results.

In this paper, we focus on learning good image priors and robust dictionaries for SR reconstruction. Recently, natural image priors have been extensively studied in the literature.

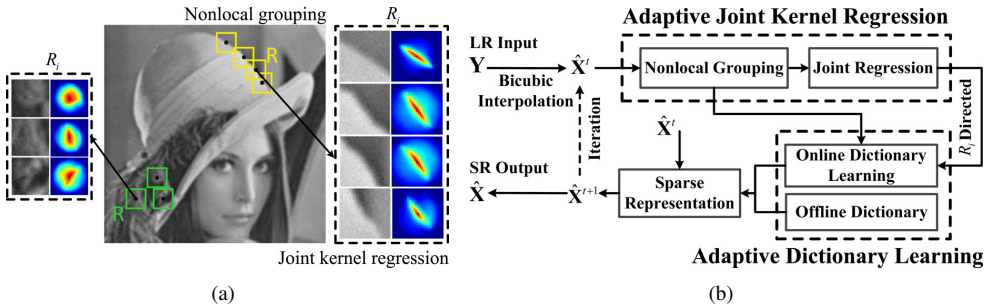


Figure 1: (a) Graphical illustration of the adaptive joint kernel regression, where the dashed boxes of different sizes indicate the regression groups with different confidence given by our *regional redundancy* measure  $R_i$  (bigger box has larger  $R_i$ ). The LR image is corrupted with a Gaussian blur ( $\sigma_b = 2$ ) and Gaussian noise ( $\sigma_n = 5$ ), and the reference patch in each group is marked as “R”; (b) Block diagram of our SR algorithm based on joint kernel regression.

The gradient profile prior is developed in [17] to preserve sharp edges, but is limited in modeling the visual complexity of real images. Yang *et al.* [21] explored the sparse representation prior of image patches with respect to a properly chosen dictionary, which usually suffers from inconsistency between neighboring patches. Later, priors of image *self-similarities* and *local/nonlocal regularities* have been exploited for more robust estimation. In [8], the nonlocal self-similarity properties both within and across spatial scales are fully exploited, but the local regularities are neglected. Zhang *et al.* [23] improved by assembling the Steering Kernel Regression [18] (SKR)-based local prior and Nonlocal Means [1] (NLM)-based nonlocal prior, whose connection, however, remains loose.

Another trend in SR is to combine the reconstruction- and example-based methods into a unified framework to produce more compelling results. For example, the dictionary-induced priors learned from example images are introduced into the regularized reconstruction in [5]. In fact, SR can be viewed as a *regression* problem aiming to map LR images to target HR images. Then in this sense, dictionary-based methods do regression using bases learned from an external database or the input image itself, while regression models directly estimate HR pixels (kernel learning) or regularize the estimator. As for the regression models, examples include SKR [18], Gaussian Process Regression (GPR) [9], Kernel Ridge Regression (KRR) [10] and Non-Local Kernel Regression (NLKR) [22], and they can all be effectively exploited as a prior for SR reconstruction. Among them, NLKR is promising to overcome the drawbacks of literature [23] by unifying the local and nonlocal priors into a single model in a complimentary way, but it discards the further potential enabled by the higher-order statistics. Besides, it needs a separate deblurring process which is ill-posed by itself.

When it comes to the dictionary-based regression, preparing an appropriate dictionary is nontrivial. Traditional choices are the analytically designed wavelet bases, but they are lacking in flexibility to a given image. Other methods instead learn a dictionary (usually over-complete) from an image database using techniques like K-SVD [13] and Principle Components Analysis (PCA) [6]. These methods often enforce sparsity under the dictionary representation. However, the flexibility is still limited since the dictionary is only learned to perform well on average. Online dictionary learning from the given image offers a promising alternative to exploit rich information contained in the input [2, 5]. One drawback is that the learning process easily runs into the risk of building dictionaries with many artifacts under

image corruptions (e.g. noise).

In this paper, we propose an Adaptive Joint Kernel Regression (AJKR)-based SR algorithm that combines the local and nonlocal image priors in a coherent framework. Our approach is different from others in several ways: 1) we combine a set of NLM-generalized kernel regressors, which are more consistent with our nonlocal collaborative framework; 2) the proposed *regional redundancy* measure introduces higher-order statistics at the region level for each regression group, making the overall framework more coherent and adaptive (see Fig. 1(a)); and 3) an adaptive PCA-based dictionary learning scheme is adopted to bridge the gap of dictionaries learned online and offline by mixing them, and more importantly such a scheme together with its induced sparsity prior can adapt to the AJKR process in response to the regional redundancy measure (see the block diagram in Fig. 1(b)).

The rest of this paper is organized as follows. Section 2 reviews related works. In Section 3 we describe our AJKR framework and its advantages over other models, and the adaptive dictionary learning scheme is also discussed. Experimental results and comparisons with state-of-the-art methods are provided in Section 4. We conclude the paper in Section 5.

## 2 Related Works

Single image SR aims to estimate an HR image  $\mathbf{X} \in \mathbb{R}^n$  from a single LR image  $\mathbf{Y} \in \mathbb{R}^m$  (lexicographically ordered vector and  $m < n$ ). The imaging model is usually expressed as

$$\mathbf{Y} = \mathbf{D}\mathbf{H}\mathbf{X} + \mathbf{V}, \quad (1)$$

where  $\mathbf{D} \in \mathbb{R}^{m \times n}$  and  $\mathbf{H} \in \mathbb{R}^{n \times n}$  are the downsampling matrix and blurring matrix respectively, and  $\mathbf{V} \in \mathbb{R}^m$  is assumed to be an additive Gaussian white noise vector. Since the inverse problem is highly ill-posed, regularization techniques are required to introduce prior knowledge to restrict the solution space. Typically, the task of SR reconstruction is formulated as a regularized least-square optimization problem as follows

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{H}\mathbf{X}\|_2^2 + \lambda \mathbf{C}(\mathbf{X}), \quad (2)$$

where  $\lambda$  is the parameter balancing the effects of the fidelity term and the regularization term  $\mathbf{C}(\mathbf{X})$ . Most of the past works focus on designing different formulations for  $\mathbf{C}(\cdot)$ .

### 2.1 Local Kernel Regression

Natural images often present relatively stable local structures, hence a query pixel can be estimated from its neighboring pixels within a small area. This local structural regularity offers an alternative for designing the regularization term  $\mathbf{C}(\cdot)$ , and can usually be expressed as a kernel regression function  $\hat{z}(\mathbf{x}_i)$

$$\hat{z}(\mathbf{x}_i) = \arg \min_z \sum_{j \in \mathcal{N}(\mathbf{x}_i)} (Y_j - z)^2 w_{ij}^K, \quad (3)$$

where  $\mathcal{N}(\mathbf{x}_i)$  denotes the local neighbors of location  $\mathbf{x}_i$ ,  $Y_j$  denotes the pixel observation at  $\mathbf{x}_j$ , and  $w_{ij}^K$  is the spatial kernel that describes the similarity between pixels at  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . In SKR [18], the steering kernel is adopted to account for the gradient which is defined as

$$w_{ij}^K = \frac{\sqrt{\det(\mathbf{Q}_i)}}{2\pi h_k^2} \exp\left(-\frac{(\mathbf{x}_j - \mathbf{x}_i)^T \mathbf{Q}_i (\mathbf{x}_j - \mathbf{x}_i)}{2h_k^2}\right), \quad (4)$$

where  $\mathbf{Q}_i$  is the symmetric gradient covariance matrix, and  $h_k$  is a smoothing parameter.

In order to approximate the local structure better, higher-order estimation can be used. Then  $\hat{z}(\mathbf{x}_i)$ , being regarded as a regression function, can be modeled to be polynomial where the image is assumed to be locally smooth to some order

$$\hat{\mathbf{a}}_i = \arg \min_{\mathbf{a}} \|\mathbf{Y}_i - \Phi \mathbf{a}\|_{\mathbf{W}_i^K}^2, \quad (5)$$

where  $\mathbf{Y}_i$  is a patch of the neighboring pixels around  $\mathbf{x}_i$  represented as a vector,  $\mathbf{W}_i^K = \text{diag}[w_{i1}^K, w_{i2}^K, \dots, w_{iL}^K]$  with  $L = |\mathcal{N}(\mathbf{x}_i)|$ , and  $\mathbf{a}$  represents the regression coefficients of the polynomial bases from Taylor expansion defined by  $\Phi$  (say second-order)

$$\Phi = \begin{bmatrix} 1 & (\mathbf{x}_1 - \mathbf{x}_i)^T & \text{vech}^T \left\{ (\mathbf{x}_1 - \mathbf{x}_i) (\mathbf{x}_1 - \mathbf{x}_i)^T \right\} \\ \vdots & \vdots & \vdots \end{bmatrix}, \quad (6)$$

where the  $\text{vech}(\cdot)$  operator stacks the lower triangular part of a matrix into a column vector.

The first element of the regression coefficient vector  $\hat{\mathbf{a}}_i$  is the desired pixel value estimation at  $\mathbf{x}_i$ , therefore

$$\hat{z}(\mathbf{x}_i) = \hat{a}_{i0} = \mathbf{e}_1^T (\Phi^T \mathbf{W}_i^K \Phi)^{-1} \Phi^T \mathbf{W}_i^K \mathbf{Y}_i, \quad (7)$$

where  $\mathbf{e}_1$  is a vector with the first element equal to one and the rest zero.

## 2.2 NLM-based Estimation

The NLM method [1], with the patch repetition assumption, restores a pixel by a nonlocal averaging of pixels with similar neighborhoods in the whole image. It is the basis of designing the nonlocal self-similarity constraint in SR reconstruction.

Mathematically, NLM algorithm is formulated as a least-square optimization problem

$$\hat{z}(\mathbf{x}_i) = \arg \min_z \sum_{j \in \mathcal{P}(\mathbf{x}_i)} (Y_j - z)^2 w_{ij}^N, \quad (8)$$

where  $\mathcal{P}(\mathbf{x}_i)$  denotes the index set for similar pixels of  $\mathbf{x}_i$  found in a nonlocal range. The weight  $w_{ij}^N$  depends on the weighted Euclidean distance between the two involved patches  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$  centered at  $\mathbf{x}_i$  and  $\mathbf{x}_j$  to describe their similarity, and is given by

$$w_{ij}^N = \exp \left( - \frac{\|\mathbf{Y}_i - \mathbf{Y}_j\|_{\mathbf{W}_G}^2}{h_n^2} \right), \quad (9)$$

where  $\mathbf{W}_G$  is the weight matrix of a Gaussian kernel, and  $h_n$  is the decay parameter.

The least-square problem in Eq. (8) has a closed form solution as

$$\hat{z}(\mathbf{x}_i) = \frac{\sum_{j \in \mathcal{P}(\mathbf{x}_i)} w_{ij}^N Y_j}{\sum_{j \in \mathcal{P}(\mathbf{x}_i)} w_{ij}^N}, \quad (10)$$

which can be regarded as a zero-order regression compared with Eq. (7) as claimed in [3].

## 3 Proposed Algorithm

### 3.1 Adaptive Joint Kernel Regression

The well-established methods of SKR [18] and NLM [1] allow us to make use of cues from local structural regularity and nonlocal self-similarity, respectively. However, they are not yet effectively utilized to attain the full power. We here propose an Adaptive Joint Kernel Regression (AJKR) framework to simultaneously exploit both image priors in a higher-order collaborative manner. To start with, the NLM is generalized in the local kernel regression framework following [3], and is given by

$$\hat{\mathbf{a}}_i = \arg \min_{\mathbf{a}} \|\mathbf{Y}_i - \Phi \mathbf{a}\|_{\mathbf{W}_i^N}^2, \quad (11)$$

where  $\mathbf{W}_i^N = \text{diag} [w_{i1}^N, w_{i2}^N, \dots, w_{iL}^N]^1$  again with  $L = |\mathcal{N}(\mathbf{x}_i)|$ .

The difference with respect to Eq. (5) lies in the choice of kernel weights as the patch similarity weight matrix  $\mathbf{W}_i^N$  is used instead of the spatial kernel matrix  $\mathbf{W}_i^K$ . In the zero-order case, the matrix  $\Phi$  simplifies to a column vector of all ones and Eq. (11) boils down to the NLM algorithm Eq. (8), but in the local range.

The role of this generalization is two-fold. For the kernel regression, the patch-based similarity can take into account the small-scale structures (see examples in Fig. 1(a)), and thus avoids the too much attention paid by spatial kernels in SKR to the dominant structural orientation and contour that are sometimes detail destroying. From the viewpoint of NLM, the higher-order version is more descriptive of the underlying image structure by structural regression, and allows for negative weights during averaging [3]. More radically, it paves the way for our framework “harmonization” in a complete nonlocal sense.

Since the NLM-generalized regressors can be regarded as localized versions of NLM, which somewhat violate its original concept, we extend them to the nonlocal range  $\mathcal{P}(\mathbf{x}_i)$  defined over a group of similar local patches to genuinely use the power of nonlocal redundancy. Here we perform grouping simply via patch matching. Once a group is built, the combination of its member regressors weighted by their mutual similarities  $w_{ij}^N$  leads to a joint regression scheme that exploits local and nonlocal priors simultaneously and collaboratively

$$\hat{\mathbf{a}}_i = \arg \min_{\mathbf{a}} \sum_{j \in \mathcal{P}(\mathbf{x}_i)} w_{ij}^N \|\mathbf{Y}_j - \Phi \mathbf{a}\|_{\mathbf{W}_j^N}^2. \quad (12)$$

From this optimization function, we can see the local kernel regression regularizes the observations found by nonlocal search via structural regression, while the nonlocal self-similarity enhances the robustness of local estimation by providing redundancies. Besides, the kernel regressors generalized from NLM are more consistent with this nonlocal fusion. The pixel estimation at  $\mathbf{x}_i$  is again the first element of the vector solution of Eq. (12)

$$\hat{z}(\mathbf{x}_i) = \mathbf{e}_1^T \hat{\mathbf{a}}_i = \mathbf{e}_1^T \left[ \Phi^T \left( \sum_{j \in \mathcal{P}(\mathbf{x}_i)} w_{ij}^N \mathbf{W}_j^N \right) \Phi \right]^{-1} \Phi^T \sum_{j \in \mathcal{P}(\mathbf{x}_i)} w_{ij}^N \mathbf{W}_j^N \mathbf{Y}_j. \quad (13)$$

We then define the row vectors  $\mathbf{k}_{ij}^T = \mathbf{e}_1^T \left[ \Phi^T \left( \sum_{j \in \mathcal{P}(\mathbf{x}_i)} w_{ij}^N \mathbf{W}_j^N \right) \Phi \right]^{-1} \Phi^T \cdot w_{ij}^N \mathbf{W}_j^N$  to be the *equivalent kernels* with which we perform the regression for  $\mathbf{x}_i$ . Since this single form of

<sup>1</sup>Note the weight  $w_{ij}^N$  here is calculated in the same way as the patch similarity weight in Eq. (9), so we use the same symbols for consistency concerns. However, the weights here are calculated from patch pairs in the local neighborhood  $\mathcal{N}(\mathbf{x}_i)$  instead of the nonlocal  $\mathcal{P}(\mathbf{x}_i)$  in Eq. (9), which is worth noting to avoid confusion.



Figure 2: Visualization of the redundancy measure  $R_i$ . From left to right: the original image and its  $R_i$  map (brighter intensity means larger value), the corrupted image and its  $R_i$  map.

equivalent kernels can supply complementary views towards the regularity in natural images, we plug them into the SR optimization function in Eq. (2) to act as the regularization  $\mathbf{C}(\cdot)$

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{H}\mathbf{X}\|_2^2 + \lambda \sum_{i=1}^n \left\| X_i - \sum_{j \in \mathcal{P}(\mathbf{x}_i)} \mathbf{k}_{ij}^T \mathbf{X}_j \right\|_2^2, \quad (14)$$

where  $X_i$  denotes the pixel to be estimated at location  $\mathbf{x}_i$ , and  $\mathbf{X}_j$  is its similar patch centered at  $\mathbf{x}_j$  found in the nonlocal range. To derive the matrix form, we should note it is highly possible that the similar patches  $\{\mathbf{X}_j\}$  overlap. This means when they are returned to their original locations, the multiple kernel weights in the overlapping part should be added up. To this end, an operator  $F(\cdot)$  is defined to realize this process, where  $F([\mathbf{k}_{i1}^T, \mathbf{k}_{i2}^T, \dots, \mathbf{k}_{iS}^T])$ ,  $S = |\mathcal{P}(\mathbf{x}_i)|$  packs the equivalent kernel set of  $\mathbf{x}_i$  into the  $i^{\text{th}}$  row of the *equivalent kernel matrix*  $\mathbf{K}$ . As a result, we can rewrite Eq. (14) as

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{H}\mathbf{X}\|_2^2 + \lambda \|\mathbf{I} - \mathbf{K}\mathbf{X}\|_2^2, \quad (15)$$

where  $\mathbf{I}$  is the identity matrix.

So far, we have obtained a regularized model that captures both the local structures and their dependencies in groups gained from the nonlocal redundancy. However, an important fact has been neglected, *i.e.* the patches in a formed group may not be mutually similar enough to enable reasonable joint regression. Actually, the degree of patch redundancy varies significantly across different regions within an image, which largely affects the correctness of our joint regression-based regularization. Therefore, we propose an explicit measure of regional redundancy to determine the confidence of each regression group that gives

$$R_i = \sum_{j \in \mathcal{P}(\mathbf{x}_i)} (w_{ij}^N)^2, \quad w_{ij}^N = \exp\left(-\frac{\|\mathbf{X}_i - \mathbf{X}_j\|_{\mathbf{W}_G}^2}{h_n^2}\right), \quad (16)$$

which can also be regarded as penalizing the patch distances due to the way  $w_{ij}^N$  is calculated. Obviously, the smaller the distances are (*i.e.* the larger  $R_i$ ), the more similar the grouped patches are and the more patch redundancy there is in the nonlocal region. Hence we give the name of regional redundancy to  $R_i$ . Figure 2 shows example maps of  $R_i$ , with large values at smooth areas and edges while small values at textures. This measure is also shown to be reasonably robust to corruptions with similar distributions, because it introduces such higher-order statistics by lifting to the region level. Then we use it to adaptively control the

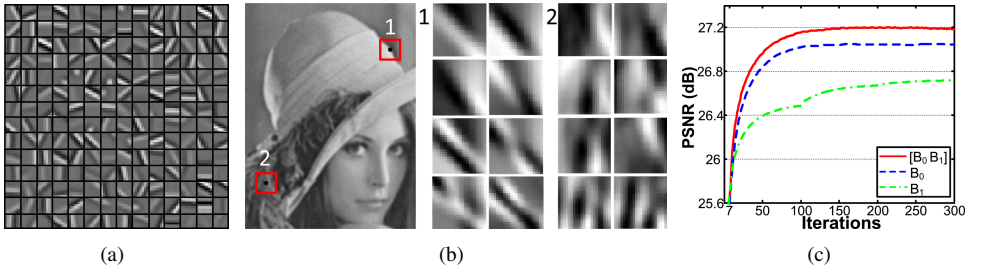


Figure 3: (a) Example of the offline PCA dictionary  $\mathbf{B}_0$  (centroids); (b) Examples of the online PCA dictionaries  $\mathbf{B}_1$ , where the 8 first atoms are shown; (c) PSNR curves ( $\times 3$ ) versus iterations for different dictionary learning schemes with the corrupted *lena* image.

relative effects of regularization among all the regression groups (see Fig. 1(a)) and have

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{H}\mathbf{X}\|_2^2 + \lambda \|(\mathbf{I} - \mathbf{K})\mathbf{X}\|_{\mathbf{R}}^2, \quad (17)$$

where the diagonal matrix  $\mathbf{R} = \text{diag}[R_1, R_2, \dots, R_n]$ . This makes our idea of joint regression more adaptive and complete by building a global *image-region vision* at higher level.

As mentioned earlier, some previous works such as [22, 23] are based on similar considerations about the unified local/nonlocal priors. Compared with NLKR [22], our framework is more coherent with the NLM-generalized kernel regressors embedded in, and is also more adaptive with the regional redundancy measure that introduces higher-order statistics to account for the inter-group variance rather than in a blindly “group-wise” way as in NLKR. Besides, we incorporate such model into the SR reconstruction framework as a regularization prior, thus estimating the HR image as a whole rather than pixel by pixel in NLKR. Compared with Zhang *et al.* [23], we enjoy the advantage of better capturing both image priors in a collaborative framework instead of crudely imposing two penalty terms.

### 3.2 Adaptive Dictionary Learning

The above AJKR framework can further benefit from dictionary-based methods that do local regression using the learned bases. We here propose a patch-based adaptive dictionary learning scheme by combining the dictionaries learned online ( $\mathbf{B}_1$  from input image) and offline ( $\mathbf{B}_0$  from external database). To learn  $\mathbf{B}_0$ , we adopt the adaptive PCA strategy in [6] due to its simplicity and use the same training data (see Fig. 3(a)). For the  $\mathbf{B}_1$ , since we have already formed many groups of similar patches in the AJKR process, we learn for each group an adaptive dictionary using PCA to adequately code the patches in it (see Fig. 3(b)), which is a popular choice in the literature [2, 5]. Figure 3(c) evaluates different dictionary learning schemes on *lena*, showing superiority of the combined dictionary  $\mathbf{B} = [\mathbf{B}_0 \mathbf{B}_1]$ .

Once the dictionary  $\mathbf{B} \in \mathbb{R}^{L \times d}$  is built, we can represent an image patch  $\mathbf{X}_i \in \mathbb{R}^L$  as a linear combination of the atoms in  $\mathbf{B}$  such that  $\mathbf{X}_i = \mathbf{B}\alpha_i$ ,  $\alpha_i \in \mathbb{R}^d$ . Let the whole image be the average of all the overlapping patch estimates,  $\alpha$  the concatenation of all  $\alpha_i$ , and  $\circ$  the representation operator, we then reach to our final optimization function for SR as

$$\hat{\alpha} = \arg \min_{\alpha} \|\mathbf{Y} - \mathbf{D}\mathbf{H}\mathbf{B} \circ \alpha\|_2^2 + \lambda \|(\mathbf{I} - \mathbf{K})\mathbf{B} \circ \alpha\|_{\mathbf{R}}^2 + \beta \|\alpha\|_1, \quad (18)$$

where  $\beta$  is the regularization parameter of the sparsity term. Notice that the constraint provided by AJKR may not be sufficient to regularize the optimization problem in Eq. (18). For

example, an extreme case exists when a patch does not look like any other in the image (*e.g.* with unique structures), then the joint kernel regression in a group of dissimilar patches can be severely erroneous and hence suppressed to a large extent due to small  $R_i$ . Therefore, we further enforce the sparsity of representation coefficients  $\alpha$  by penalizing its  $\ell_1$  norm.

Indeed, some recent works (*e.g.* [14, 15]) also consider mixing the online and offline dictionaries. Compared with them, our combined dictionary together with its induced sparsity prior distinguishes itself by offering the ability to adapt to the other prior (regression-based) in a unified framework. The superiority of the combined dictionary  $\mathbf{B}$  in Fig. 3(c) is a result of such adaptation. Specifically, if no/few similar patches are found for a given one, its near-zero redundancy measure  $R_i$  actually cancels out the erroneous regression prior as mentioned above, thus reducing Eq. (18) to sparse coding only at that patch, which is performed mainly over the bases in  $\mathbf{B}_0$  because those in  $\mathbf{B}_1$  learned online from mutually dissimilar patches deviate too far from the true signal space. While in the case of high patch redundancy, the large  $R_i$  imposes a strong effect of the joint regression prior, and on the other hand, the online dictionary  $\mathbf{B}_1$  dominates for enforcing the remaining sparse representation prior. Figure 1(b) illustrates the block diagram of our overall algorithm, and the optimization problem (Eq. (18)) at each iteration can be efficiently solved by the iterative shrinkage algorithm [4].

## 4 Experiments

In this section we compare our method AJKR with several representative related as well as state-of-the-art SR algorithms. The quantitative results are carried out only on the illuminance channel for color images, in terms of the metrics of PSNR and Structural SIMilarity index (SSIM) [19]. Since in real-world SR tasks the observed LR images are often contaminated by noise, the robustness of SR methods with respect to noise is also evaluated.

In all experiments we use HR patches of  $7 \times 7$  pixels ( $L = 49$ ) with a 4-pixel overlap for both local kernel regression and nonlocal patch matching. We set the support of the nonlocal searching to be  $S = 10$  nearest neighbors in a window of size  $21 \times 21$ , and use the parameters  $\lambda = 50, \beta = 0.27, h_n = 10$ . We first performed SR experiments on synthetic LR images. The LR images were generated from 5 standard test images by a truncated  $7 \times 7$  Gaussian kernel ( $\sigma_b = 1.6$ ) and down-sampled by a factor of 3. A Gaussian noise ( $\sigma_n = 5$ ) was also added.

Table 1 shows the quantitative performance of our method, for both noiseless and noisy cases, in comparison to three regression-based methods, GPR [9], KRR [10] and the method of Zhang *et al.* [23], and to two dictionary-based methods, Centralized Sparse Representation (CSR) [5] and Sparse Coding (SC) method [21]. The GPR and KRR are two recent regression methods that capture from input image the mapping between LR and HR patches via Gaussian process regression and sparse kernel regression, respectively. For all the five compared methods, we used their default parameter setup. As shown in the table, our method constantly outperforms the others across all metrics for both cases, with the largest PSNR improvements (over GPR) at 4.05 dB (noiseless) and 2.78 dB (noisy) on average which are quite significant. The demonstrated robustness to noise of our method can be attributed to the redundancy measure-guided prior adaptation and the responsive dictionary scheme.

Figure 4 offers an example for visual comparison. We can see our AJKR method can preserve much sharper edges and more details than the other three regression methods. Our result is also free of the jaggart artifacts with the other dictionary-based methods.

We further compare with other state-of-the-art algorithms on real images in Fig. 5. Compared with NLKR [22] and the methods of Shan *et al.* [16], Glasner *et al.* [8] and Freedman



Table 1: Comparison of SR results ( $\times 3$ , PSNR/SSIM) for both noiseless and noisy cases.

Images	GPR	SC	KRR	Zhang <i>et al.</i>	CSR	AJKR
Bike	21.9 / 0.635	23.3 / 0.739	23.2 / 0.717	24.1 / 0.786	24.7 / 0.802	<b>25.4 / 0.833</b>
Butterfly	23.0 / 0.796	24.6 / 0.821	24.8 / 0.855	26.9 / 0.896	28.2 / 0.921	<b>28.6 / 0.937</b>
Girl	31.7 / 0.765	30.9 / 0.804	32.5 / 0.788	33.0 / 0.810	33.7 / 0.826	<b>33.9 / 0.840</b>
Parrot	26.6 / 0.854	28.4 / 0.883	28.4 / 0.883	29.6 / 0.900	30.7 / 0.918	<b>30.8 / 0.935</b>
Plants	29.9 / 0.832	31.3 / 0.879	31.5 / 0.872	32.9 / 0.897	34.0 / 0.921	<b>34.6 / 0.936</b>
Noisy Bike	21.8 / 0.620	23.0 / 0.698	22.9 / 0.685	23.2 / 0.711	23.8 / 0.736	<b>24.3 / 0.762</b>
Noisy Butterfly	22.6 / 0.775	24.3 / 0.787	24.4 / 0.816	25.7 / 0.850	26.8 / 0.888	<b>27.0 / 0.894</b>
Noisy Girl	31.0 / 0.741	30.4 / 0.734	31.3 / 0.736	31.5 / 0.749	32.0 / 0.764	<b>32.2 / 0.782</b>
Noisy Parrot	26.4 / 0.829	27.7 / 0.800	27.8 / 0.816	28.5 / 0.856	29.5 / 0.878	<b>29.6 / 0.892</b>
Noisy Plants	29.4 / 0.807	30.0 / 0.795	30.3 / 0.805	30.8 / 0.825	31.7 / 0.860	<b>32.0 / 0.868</b>

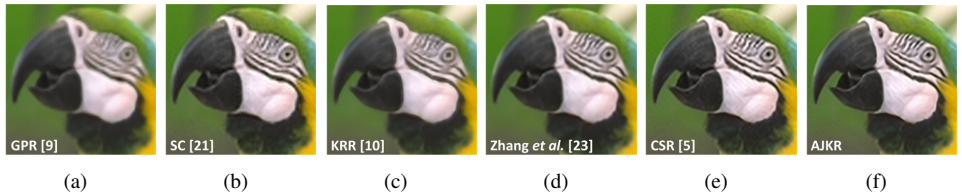
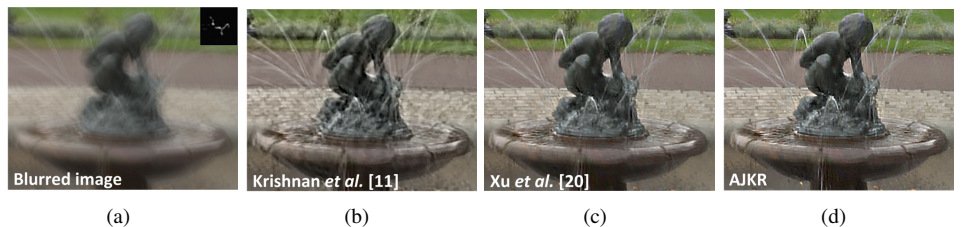
Figure 4: Visual comparison of SR results on *Parrot* image ( $\times 3$ , noiseless case).

Figure 5: SR results on real images with a more challenging magnification factor 4.

Figure 6: Real image deblurring (compared with Krishnan *et al.* [11] and Xu *et al.* [20]).

and Fattal [7], our improvements are evident. Note that we use the reported results of NLKR since its source code is not available (so not compared previously). Fig. 6 shows an image deblurring example by omitting the downsampling operator  $\mathbf{D}$  from Eq. (18).

## 5 Conclusions

This paper introduces an adaptive joint kernel regression framework for single-image super-resolution. This framework can make better use of the local and nonlocal image priors in

a higher-order collaborative manner. An adaptive PCA-based dictionary learning scheme is also integrated to interact with the regression prior for robustness. The large variety of experiments show that the proposed algorithm achieves state-of-the-art performance and is robust with minimum artifacts even under large amounts of noise and blur. The extension to image deblurring also shows great potential to produce outstanding results as compared with the best performing deblurring methods.

## 6 Acknowledgments

This work was supported by the National Basic Research Program of China (973 program) under Grant No. 2013CB329403.

## References

- [1] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 60–65, June 2005.
- [2] P. Chatterjee and P. Milanfar. Clustering-based denoising with locally learned dictionaries. *IEEE Transactions on Image Processing*, 18(7):1438–1451, July 2009.
- [3] Priyam Chatterjee and Peyman Milanfar. A generalization of non-local means via kernel regression. In *Proc. IS&T-SPIE Computational Imaging VI*, page 68140P, 2008.
- [4] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.
- [5] Weisheng Dong, Lei Zhang, and Guangming Shi. Centralized sparse representation for image restoration. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1259–1266, 2011.
- [6] Weisheng Dong, Lei Zhang, Guangming Shi, and Xiaolin Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on Image Processing*, 20(7):1838–1857, 2011.
- [7] Gilad Freedman and Raanan Fattal. Image and video upscaling from local self-examples. *ACM Trans. Graph.*, 28(3):1–10, 2010.
- [8] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. In *IEEE International Conference on Computer Vision (ICCV)*, pages 349–356, 2009.
- [9] He He and Wan Chi Siu. Single image super-resolution using Gaussian process regression. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 449–456, June 2011.
- [10] Kwang In Kim and Younghee Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1127–1133, June 2010.

- [11] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 233–240, 2011.
- [12] Xin Li and M. T. Orchard. New edge-directed interpolation. *IEEE Transactions on Image Processing*, 10(10):1521–1527, 2001.
- [13] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, 2008.
- [14] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2272–2279, October 2009.
- [15] Daniele Perrone, Avinash Ravichandran, René Vidal, and Paolo Favaro. Image priors for image deblurring with uncertain blur. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 114.1–114.11, 2012.
- [16] Qi Shan, Zhaorong Li, Jiaya Jia, and Chi Keung Tang. Fast image/video upsampling. *ACM Trans. Graph.*, 27(5):1–7, October 2008.
- [17] Jian Sun, Zongben Xu, and Heung Shum. Image super-resolution using gradient profile prior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, June 2008.
- [18] H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. *IEEE Transactions on Image Processing*, 16(2):349–366, 2007.
- [19] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [20] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural  $l_0$  sparse representation for natural image deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.
- [21] Jianchao Yang, J. Wright, T. S. Huang, and Yi Ma. Image super-resolution as sparse representation of raw image patches. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, June 2008.
- [22] Haichao Zhang, Jianchao Yang, Yanning Zhang, and Thomas S. Huang. Non-local kernel regression for image and video restoration. In *Proc. European Conference on Computer Vision (ECCV)*, pages 566–579, 2010.
- [23] Kaibing Zhang, Xinbo Gao, Dacheng Tao, and Xuelong Li. Single image super-resolution with non-local means and steering kernel regression. *IEEE Transactions on Image Processing*, 21(11):4544–4556, 2012.