Multi-scale Dictionary for Single Image Super-resolution

Kaibing Zhang, Xinbo Gao [1]
[1] School of E.E., Xidian University
Xi'an 710071, P.R. China

Dacheng Tao [2]
[2] QCIS and FEIT, University of Technology, Sydney
NSW 2007, Australia

Xuelong Li [3]
[3] Xi'an Institute of Optics and Precision Mechanics of CAS
Xi'an 710119, P.R. China

kbzhang0505@gmail.com,xbgao@mail.xidian.edu.cn,dacheng.tao@uts.edu.au,xuelong_li@opt.ac.cn

Abstract

Reconstruction- and example-based super-resolution (SR) methods are promising for restoring a high-resolution (HR) image from low-resolution (LR) image(s). Under large magnification, reconstruction-based methods usually fail to hallucinate visual details while example-based methods sometimes introduce unexpected details. Given a generic LR image, to reconstruct a photo-realistic SR image and to suppress artifacts in the reconstructed SR image, we introduce a multi-scale dictionary to a novel SR method that simultaneously integrates local and non-local priors. The local prior suppresses artifacts by using steering kernel regression to predict the target pixel from a small local area. The non-local prior enriches visual details by taking a weighted average of a large neighborhood as an estimate of the target pixel. Essentially, these two priors are complementary to each other. Experimental results demonstrate that the proposed method can produce high quality SR recovery both quantitatively and perceptually.

1. Introduction

Image super-resolution (SR) technique has been recognized as an effective and efficient method to produce high-resolution (HR) images that conventional digital cameras cannot capture from a real scene. This technique has potential applications in many fields such as computer vision, medical and remote sensing imaging, video surveillance, and entertainment. Therefore, it has attracted intensive attention in video and image processing communities over the past decades. Existing SR approaches can be divided into three categories: interpolation-based methods, reconstruction-based methods, and example-based methods.

- Interpolation-based SR approaches (e.g., [20]) are simple and fast but tend to produce blurring edges and unclear details.

- Reconstruction-based methods (e.g., [11, 13]) estimate an HR image by incorporating a certain prior knowledge (e.g., edge prior [14], redundancy of similarity prior [19], and gradient profile prior [13]) to make SR estimate well-posed. These approaches perform well on reconstructing sharp edges with less jaggy artifacts.

- Example-based SR approaches (e.g., [7, 3, 17, 10]) assume that the high-frequency details lost in the LR image can be effectively predicted from a set of LR and HR training image pairs. With the help of a training

(a) ASDS [4] (b) Patch redundancy [8]

(c) Sparse regression [10] (d) Our results

Figure 1. Comparison ($\times 3$) of SR results on monarch image by different methods. (a) result using ASDS SR [4]; (b) result using patch redundancy SR [8]; (c) result using sparse regression and natural image prior [10]; (d) result from our SR method. Note that our result is compelling both in the major edge and textural regions. Moreover, our method produces noiseless result.

database, these methods can generate plausible details that do not appear in the LR input. However, the fundamental problem is that the quality of resultant images heavily depends on the supporting training images. Moreover, example-based methods are difficult to preserve sharper edge and suppress unwanted artifacts.

In recent years, a large number of SR methods (e.g., [8, 14, 4]) that combine the reconstruction- and example-based methods into a unified SR framework have been acknowledged as an efficient way to produce more compelling SR results. To reconstruct a photo-realistic HR image with sharp edges and visual details, this paper introduces a multi-scale dictionary to a novel SR method that simultaneously integrates the local and non-local priors. The proposed SR method takes the following unique features:

1. To address the aforementioned problems in example-based methods, this paper directly estimates the high-frequency details from an LR input based upon the redundancies of similar patterns across different scales in natural images. In particular, we jointly learn a multi-scale dictionary by using image patches at different scales from the LR input and recover the missing details by sparsely representing the expected HR image from the learnt multi-scale dictionary, which is possible to capture the redundancies of similar patches at different scales. We further develop an example-based hallucination term and incorporate it into the reconstruction-based SR method, which aims to obtain more faithful details.

2. To achieve a reliable SR estimate, a local regularization term is introduced by reformulating the steering kernel regression (SKR) [15] to capture the local information of an image. To obtain a robust SR result, a non-local prior regularization term is added by using the non-local means (NLM) [2] filter to find similar redundancies within the same scale. We incorporate these two complementary regularization terms into the reconstruction-based SR framework to maintain sharp edges and suppress artifacts.

In this paper, a unified energy minimization model is developed to obtain a local optimal solution, where the HR reconstruction term, local and non-local regularization terms, and example-based hallucination term are integrated together. Figure 1 shows an example on the monarch image, where the ASDS method [4], patch redundancy method [8], and sparse regression [10] methods are used as baselines. This example suggests that not only noticeable ringing artifacts along major edges are suppressed, but visual details can perfectly be recovered by our method. Moreover, our result appears to be noiseless, leading to photo-realistic results. The rest of the paper is organized as follows. Related works are briefly reviewed in Section 2. Section 3 explains how to utilize the redundancy to design a multi-scale dictionary from a given LR input. In Section 4, we present the proposed SR framework. Section 5 demonstrates experimental results in comparison to the-state-of-arts. Section 6 concludes this paper.

2. Related works

The reconstruction-based methods impose some priors during the reconstruction process to obtain a reliable estimate. Especially, edge-directed prior methods [11, 13] are popular. For example, Fattal [6] enforced an edge statistics prior on the reconstructed HR image to generate sharper edges and suppress noticeable artifacts. Thus, these edge-directed approaches perform well on preserving sharper edges and suppressing noticeable artifacts. However, they fail to recover visual details and may result in unnatural results. Example-based SR approaches exceed the reconstruction-based methods in recovering plausible details by learning the correspondences between LR image patches (or pixels) and HR image patches (or pixels) from a training database consisting of LR and HR patch pairs. Representative approaches include k-nearest neighbor (k-NN) learning (e.g., [7]), manifold learning (e.g., [3, 5]), sparse coding (e.g., [17]), and regression-based (e.g., [10]) methods. Although they can generate high-frequency details, they typically tend to blur edges. In this paper, to reconstruct a photo-realistic HR image and to suppress artifacts in the reconstructed image, we propose a novel SR method that learns a multi-scale dictionary from LR input to hallucinate details while integrates the local and non-local priors to produce sharp edges and suppress unexpected artifacts.

3. Multi-scale dictionary-based hallucination

Our multi-scale dictionary-based hallucination is based upon the following two observations. First, the local structures in a natural image usually tend to repeat themselves many times, both within the same scale and across different scales. Therefore, details missing in a local structure at a smaller scale can be estimated from its similar patches at a larger scale. Secondly, different images prefer different patch sizes for optimal representation. For instance, the major edges prefer a larger scale while the sophisticated details tend to a smaller one. Therefore, it is important to jointly represent an image at different scales. Considering the above cues, we introduce a multi-scale dictionary representation [12] (originally used in image/video denoising) to example-based hallucination. Then we reformulate it as a hallucination regularization term for the reconstruction-based SR framework to maintain visual details.

3.1. Generating training images

To learn a multi-scale dictionary for representing similar redundancies of local patterns within the same scale and across different scales, we extract training image patches from the pyramid images down-sampled from the LR input.

Let Y be the LR input. The down-sampling version I_p at th-p level is given by convoluting Y with a Gaussian kernel

$$I_p = (Y * B_p) \downarrow_{s_p} \tag{1}$$

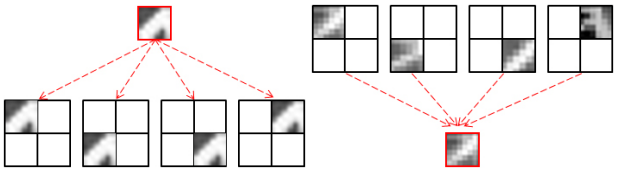
where $B_p(p = 1 \dots 4)$ stands for the Gaussian kernel with a standard deviation $\sigma^2 log(p)/log(5)$, $\downarrow_{s_p}$ denotes the down-sampling operator with factor $s_p = (0.8)_p$ at the $p-$th level.

3.2. Learning multi-scale dictionary

Consider a set of root patches $\{z_i\}_{i=1}^N$ of size $\sqrt{n} \times \sqrt{n}$ extracted from a sequence of pyramid images. Each patch can be sparsely reconstructed by using the multi-scale dictionary $D \in \mathbb{R}^{n \times k}$, containing k atoms and S different scales. The root patch is divided along the tree into the sub-patches of size $n_s = \frac{n}{4^s}$, where $s = 0, ..., S-1$ is the depth in the tree. A multi-scale dictionary $D \in \mathbb{R}^{n \times k} (k = \sum_{s=0}^{S-1} 4^s k_s)$ contains S different dictionaries $D_s \in \mathbb{R}^{n_s \times k_s}$, each of which has k_s atoms of size n_s. We employ a multi-scale K-SVD algorithm proposed in [12] to learn such a dictionary from $\{z_i\}_{i=1}^N$ such that each sample can be approximately represented by a sparsely linear combination of the atoms from D, i.e.,

$$D, \{\alpha_i\} = \arg \min_{D, \{\alpha_i\}} \sum_i \|z_i - D\alpha_i\|_2^2, s.t. \|\alpha_i\|_0 \leq L, \forall i \tag{2}$$

where α_i is the sparse representation of z_i , $\|.\|_0$ stands for the count for the nonzero entries in α_i and L is the maximal number of nonzero entries. In this paper, we introduce two special operations to extend the traditional K-SVD algorithm [1] for learning of a multi-scale dictionary: *stacking operation* and *retrieving operation*. In particular, the *stacking operation* is used to stack each atom in $D_s \in \mathbb{R}^{n_s \times k_s}$ into its possible positions and construct a multi-scale dictionary $D \in \mathbb{R}^{n \times k}$ for sparse coding, in which all atoms at different scales as well as different positions are equally treated. Correspondingly, the *retrieving operation* is used to revisit each updated atom in $D \in \mathbb{R}^{n \times k}$ at its possible positions and combine them into a single one. This operation is necessary because each atom at the scale s has 4^s possible positions. All the atoms are equally treated and updated independently. Thus, the atom $d_{sl} (1 \leq l \leq k_s)$ is updated 4^s times at its 4^s possible positions during each iteration. We revisit each atom at its possible positions in D and average the corresponding regions as the final estimate of this atom.



(a) Stacking operation (b) Retrieving operation

Figure 2. Overview of jointly training multi-scale dictionary. (a) Stacking each multi-scale atom to different positions before updating; (b) Retrieving a multi-scale atoms from different positions by average after updating.

Figure 2 illustrates the stacking and revisiting operations on an atom at the scale s=1, where four possible locations are considered during performing dictionary learning.

By using the two operations, we can easily obtain the multi-scale dictionary D through the traditional **K-SVD** algorithm [1]. Note that we randomly select k_0 samples as the root atoms in D_0 . For D_s, $s = 1 \dots S-1$, we directly divide the root atoms into a set of subpatches and randomly choose k_s samples of size n_s to construct its initial dictionary.

3.3. Reconstruction with sparse representation

Let X be an HR image to be reconstructed and the operator R_{ij} be a binary matrix that extracts a square patch of size $\sqrt{n} \times \sqrt{n}$ from the location (i,j) in X and represent the patch as a column vector by lexicographic ordering, i.e., $x_{ij} = R_{ij}X$. The sparse representation of x_{ij} over the multi-scale dictionary $D \in \mathbb{R}^{n \times k}$ is given as

$$\hat{x}_{ij} \approx D\alpha_{ij}, s.t. \|\alpha_{ij}\|_0 \leq L \tag{3}$$

where α_{ij} is the sparse representation of patch $R_{ij}X$ in $\hat{X}$. An HR image Z can be reconstructed by merging all the constructed patches and averaging the overlapping regions between the adjacent patches, i.e.,

$$\hat{Z} = \left(\sum_{ij} R_{ij}^T R_{ij} \right)^{-1} \sum_{ij} R_{ij}^T D\alpha_{ij}. \tag{4}$$

4. Single image SR model

Figure 3 presents the diagram of the proposed unified SR framework that includes the reconstruction term, a local prior as well as a non-local prior regularization terms, and sparse hallucination regularization term. Mathematically, it is defined as

$$X^* = \min_X \left\{ \begin{array}{l} E(X|Y) + \alpha_1 E_{local}(X) \\ +\alpha_2 E_{non\text{-}local}(X) + \alpha_3 E_{sparse}(X) \end{array} \right\} \tag{5}$$

where $E(X|Y) = \|(X * B) \downarrow -Y\|_2^2$ is the reconstruction term to ensure that the reconstructed HR image is consistent with the LR input via back-projection. The second

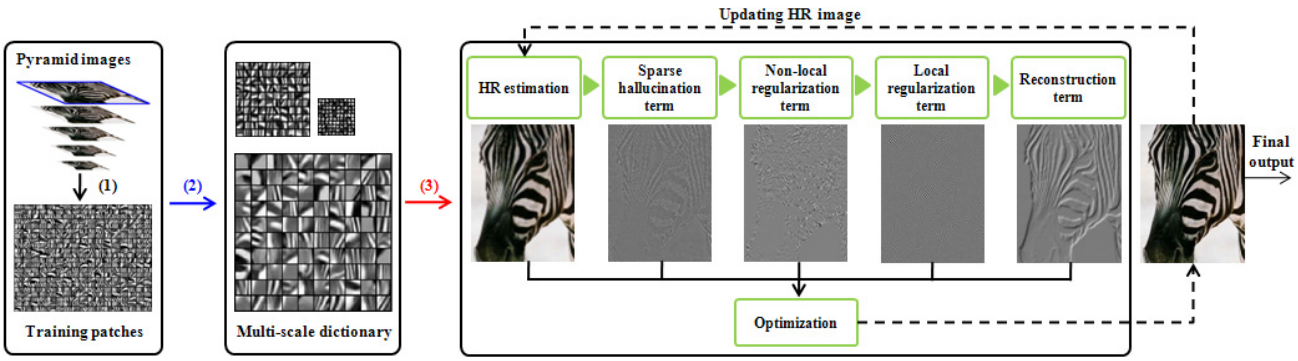


Figure 3. Overview of proposed SR framework. There are three fundamental stages. (1) Generating training samples from the pyramid images of the LR input; (2) Learning multi-scale dictionary; (3) Optimizing a unified energy function, in which the HR reconstruction term, the local as well as non-local regularization terms, and sparse hallucination regularization term are combined together and the gradient decent method is used to find a local optimal solution to the reconstructed HR image.

term $E_{local}(X)$ is the local prior regularization, which indicates that each HR pixel should be perfectly estimated from a small local area around it. The third term $E_{non\text{-}local}(X)$ is the non-local prior that assumes that each HR pixel can be predicted by weighting average of a large neighborhood. The last sparse hallucination regularization term requires that the estimated HR image has a sparse representation over a multi-scale dictionary learnt from the LR input itself.

4.1. Local prior regularization term

The local prior term $E_{local}(X)$ assumes that a given HR pixel can be predicted from a small neighborhood area, i.e.,

$$\hat{X}_i = \arg\min \sum_{j \in \mathcal{N}(\mathbf{x}_i)} (X_j - X_i)^2 K_{\mathbf{x}_i}(\mathbf{x}_j - \mathbf{x}_i), \quad (6)$$

where $\mathbf{x}_i$ represents the i-th 2D location in the estimated HR image X, $\mathcal{N}(\mathbf{x}_i)$ stands for the neighbors of $\mathbf{x}_i$, X_j stands for the pixel value at the location $\mathbf{x}_j$, and $K_{\mathbf{x}_i}(\mathbf{x}_j - \mathbf{x}_i)$ is a spatial kernel at location $\mathbf{x}_i$ which assigns larger weights to nearby similar pixel while smaller ones to farther non-similar pixels. We employ the steering kernel proposed in [15] to calculate the weights as follows:

$$w_{ij}^K = \frac{\sqrt{det(C_i)}}{2\pi h_k^2} exp\left(- \frac{(\mathbf{x}_j - \mathbf{x}_i)^T C_i (\mathbf{x}_j - \mathbf{x}_i)}{2h_k^2} \right), \quad (7)$$

where w_{ij}^K is the weight of the pixel x_i with respect to the pixel x_j, the matrix C_i is the symmetric gradient covariance at $\mathbf{x}_i$ in the vertical and horizontal directions, and h_k is a smoothing parameter to control the supporting range of the steering kernel. We denote the steering kernel $\{w_{ij}^K\}_{j=1}^{|\mathcal{N}(\mathbf{x}_i)|}$ as a row vector w_i^K by lexicographic ordering. With Eq. (7), the local kernel can estimate the intensity for the pixel x_i as a weighted least-squares problem:

$$\begin{aligned} \hat{\beta}_i &= \arg\min (R_{\mathbf{x}_i} X - \Phi\beta_i)^T W_i^K (R_{\mathbf{x}_i} X - \Phi\beta_i) \\ &= \arg\min \|R_{\mathbf{x}_i} X - \Phi\beta_i\|^2_{W_i^K}, \end{aligned} \quad (8)$$

where $R_{\mathbf{x}_i}$ is an operator to extract the local patch centered at the i-th 2D location from $\mathbf{x}_i$ and represent it as a vector by lexicographic ordering; $W_i^K = diag(w_i^K)$ is the weight matrix defined by the steering kernel $\{w_{ij}^K\}_{j=1}^{|\mathcal{N}(\mathbf{x}_i)|}$ and Φ is a polynomial bases. The solution of Eq. (8) is

$$\hat{\beta}_i = (\Phi^T W_i^K \Phi)^{-1} \Phi^T W_i^K R_{\mathbf{x}_i} X. \quad (9)$$

Hence, the estimate at position $\mathbf{x}_i$ is

$$\hat{X}_i = e_1^T \hat{\beta}_i, \quad (10)$$

where e_1 is a column vector with the first element equal to one, and the remaining equal to zero. To incorporate the above regression form into our SR framework, we encapsulate the regression weights related to each pixel, i.e.,

$$\hat{X} = \arg\min \{ \sum_{i \in X} \|X_i - c_i^K \cdot \mathcal{L}(\mathbf{x}_i)\|_2^2 \}, \quad (11)$$

where $c_i^K = e_1^T (\Phi^T W_i^K \Phi)^{-1} \Phi^T W_i^K$, $\mathcal{L}(\mathbf{x}_i)$ represents a column vector consisting of neighborhood pixels centered at $\mathbf{x}_i$. We reformulate Eq. (11) as the matrix form by

$$\hat{X} = \arg\min \|(I - Q)X\|_2^2, \quad (12)$$

where I is the identity matrix and

$$Q(i,j) = \begin{cases} c_{ij}^K & , j \in \mathcal{L}(\mathbf{x}_i) \\ 0 & , otherwise \end{cases} \quad (13)$$

In this way, we can write the local prior regularization term as:

$$E_{local}(X) = \|(I - Q)X\|_2^2. \quad (14)$$

4.2. Non-local regularization term

We reformulate the NLM filter [2] to form the non-local regularization term. Mathematically, the NLM filter calculates a pixel by averaging of similar pixels within a large neighborhood by matching not only their own pixel values but also their local neighboring pixels, i.e.,

$$\hat{X}_i = \frac{\sum\limits_{j \in \mathcal{P}(\mathbf{x}_i)} w_{ij}^N X_j}{\sum\limits_{j \in \mathcal{P}(\mathbf{x}_i)} w_{ij}^N} \tag{15}$$

where $\mathcal{P}(\mathbf{x}_i)$ denotes the index set that consists of the coordination of pixels similar to the pixel X_i. The weight w_{ij}^N stands for the similarity between the local patches $R_{\mathbf{x}_i}X$ (centered around the center pixel X_i) and $R_{\mathbf{x}_j}X$ (centered around the center pixel X_j). The similarity weight is estimated by

$$w_{ij}^N = exp\left(- \frac{\|R_{\mathbf{x}_i}X - R_{\mathbf{x}_j}X\|_G^2}{h_n^2} \right), \tag{16}$$

where h_n is a global filter parameter that controls the decay of the exponential expression in the weighting computation. The parameter G is a kernel matrix (e.g., Gaussian kernel) that assigns a large weight to the pixels nearby the center pixel of the image patch. For the non-local similarity regularization term, we rewrite it in the following form

$$\hat{X} = \arg\min\{\sum_{i \in X} \|X_i - \mathbf{c}_i^N \,.\, \mathcal{S}(\mathbf{x}_i)\|_2^2\}, \tag{17}$$

where $\mathcal{S}(\mathbf{x}_i)$ represents a column vector by stacking similar pixels centered at $\mathbf{x}_i$ in lexicographical ordering and $\mathbf{c}_i^N$ is the row vector consisting of the corresponding NLM weights. Then the Eq. (17) can be further rewritten as

$$\hat{X} = \arg\min \|(I - P)X\|_2^2 \tag{18}$$

where I is the identity matrix and

$$P(i,j) = \begin{cases} c_{ij}^N & , j \in \mathcal{S}(\mathbf{x}_i) \\ 0 & , otherwise \end{cases} . \tag{19}$$

In such a way, we form the non-local prior regularization term as

$$E_{non\text{-}local}(X) = \|(I - P)X\|_2^2. \tag{20}$$

4.3. Sparse hallucination regularization term

The sparse hallucination regularization term enforces another constraint that the reconstructed HR image X should perfectly be consistent with by a multi-scale dictionary learnt from the LR input itself, i.e.,

$$\|X - Z\|_2^2 \leq \varepsilon^2. \tag{21}$$

We formulate the sparse hallucination regularization term as:

$$E_{sparse}(X) = \|X - Z\|_2^2. \tag{22}$$

4.4. Optimization

To find a local optimal solution of

$$\begin{aligned} E(X) = \|(X * B) \downarrow - Y\|_2^2 + \alpha_1 \|(I - Q)X\|_2^2 \\ + \alpha_2 \|(I - P)X\|_2^2 + \alpha_3 \|X - Z\|_2^2, \end{aligned} \tag{23}$$

we employ the gradient descent method to optimize it:

$$X^{(t+1)} = X^{(t)} - \tau \nabla E(X), \tag{24}$$

where t is iteration times and τ is the step size. The gradient of the energy function is written as

$$\begin{aligned} \nabla E(X) = \big((X * B) \downarrow - Y\big) \uparrow * B + \alpha_1 (I - Q)^T (I - Q)X + \\ \alpha_2 (I - P)^T (I - P)X + \alpha_3 (X - Z). \end{aligned} \tag{25}$$

When implementing, we initialize the HR image $X^{(0)}$ by bicubic interpolation and reconstruct Z by sparse representation over a learnt multi-scale dictionary from the input LR image Y .

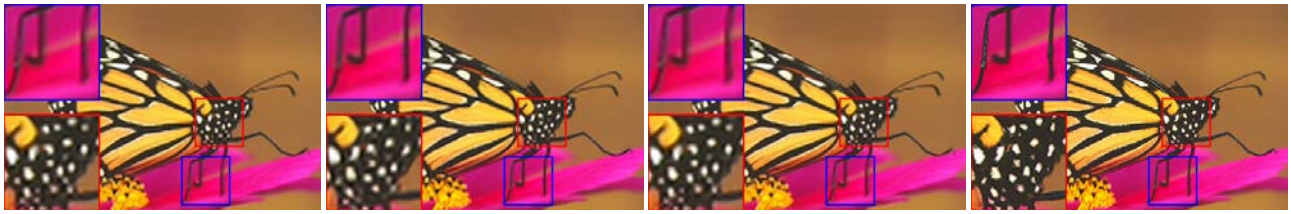
5. Experimental results

Experimental configuration. The simulated LR images are generated from the original HR versions by a 7×7 Gaussian kernel (with standard deviation 1.1 for down-sampling factor of 3 in our experiments) and then decimated by a specified factor 3. In our tests, the local analysis window for computing the SKR weights is the size of 7×7. The structure sensitive parameter (cf. [15]) and the smoothing parameter h_k are set to 0.3 and 1.5, respectively. The patch size applied in calculating the similarity weights is set to 7×7 and the filter parameter h_n is set to 5. The search radius of 13×13 pixels is manually specified for finding the related neighbors and the top 10 neighbors are selected to calculate the NLM weight matrix. In our experiments, 20 times iterations are performed to learn the multi-scale dictionary, and the number of atoms for each scale is set to 256. The size of root patch $\sqrt{n} \times \sqrt{n}$ is set to $\sqrt{12} \times \sqrt{12}$. In addition, the step size τ is set to 4, the parameters α_1, α_2 , and α_3 are set to 0.25, 0.1, and 0.02 respectively through experimental adjustment.

Comparison with the state of the art algorithms. We conduct the SR experiments on a variety of natural images with textures and edges. Since the human visual system (HVS) is more sensitive to the luminance channel than the chrominance channels, we transform *RGB* values into *YCbCr* space and carry out the SR process on the luminance channel in *YCbCr*. The chrominance channels (*Cb* and *Cr*)

(a) PSNR: 21.16, SSIM: 0.785 (b) PSNR: 22.05, SSIM: 0.825 (c) PSNR: 23.02; SSIM: 0.853 (d) PSNR: 23.49, SSIM: 0.864

Figure 4. Comparison of SR results ($\times 3$) on zebra image. (a) shows the result of back-projection; (b) shows the results of combing local prior regularization term; (c) shows the results of combining local and non-local regularization terms. The model incorporating the reconstruction term, local and non-local terms, and sparse reconstruction terms produces sharper edges and visual details as shown in (d).



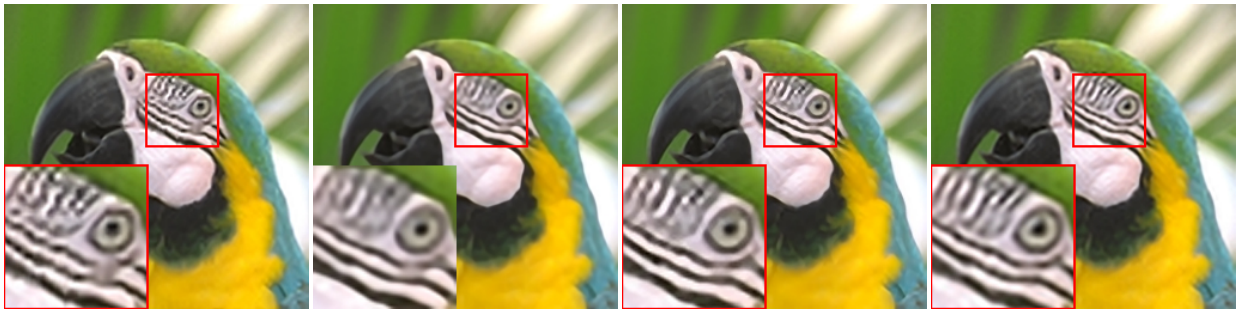
(a) PSNR: 24.17, SSIM: 0.871 (b) PSNR: 24.73, SSIM: 0.886 (c) PSNR: 24.76, SSIM: 0.888 (d) Original

Figure 5. Comparison of SR results ($\times 3$) on butterfly image obtained by different scale dictionary. (a) shows the result of $S = 1$ (single scale); (b) shows the results of $S = 2$ (two scales); (c) shows the results of $S = 3$ (three scales). The model incorporating sparse hallucination term with multi-scale dictionary produces sharper edges and visual details than with single one as shown as (b) and (c).

are directly magnify to the desired size through the bicubic interpolation algorithm. To demonstrate the effectiveness of the proposed SR method, Figure 4 compares the SR results on zebra image by incorporating different regularization terms in different ways. Figure 4 (a) shows the results from back-projection algorithm, where both serious jaggy artifacts along edges and annoying details are produced. Figure 4 (b) shows results of incorporating the construction term and the local prior term. Although the result is significantly improved, there are still a lot of unpleasing artifacts. Figure 4 (c) shows the results of incorporating local and non-local regularization terms and the reconstruction term. In this result, the artifacts along edges are effectively suppressed, leading to more compelling results than Figure 4 (a) and (b). However, the fine details cannot perfectly be recovered. Figure 4 (d) shows the results of incorporating the reconstruction term, local and non-local regularization terms, and sparse hallucination regularization term. Compared with the results in Figure 4 (a)-(c), these results in Figure 4 (d) show top level quality both along edges and in textural regions.

In Figure 5 , we test our method on butterfly image and compare SR results obtained from the dictionaries at different scales. We can see that the reconstructed HR images obtained from the multi-scale dictionary shown in Figure 5 (b) and (c) are better in terms of quantitative and visual quality than that obtained from the single-scale dictionary as shown in Figure 5 (a).

In order to further test the effectiveness of our proposed SR model in recovering textural details, Figure 6 compares our method with three state-of-the-arts, including ASDS [4], patch redundancy [8], and sparse regression [10] methods. As shown in Figure 6, the ASDS method performs well in preserving edge and suppressing noise. However, the resultant images tend to smooth high frequency details as shown in Figure 6 (a) in the red insets. The patch redundancy method exploits the redundancies in the LR image to hallucinate details, which is demonstrated to be better at generating sharpening edges than synthesizing faithful details as shown in Figure 6 (b). The results of sparse regression method shown in Figure 6 (c) are sharper along edges. However, the method tends to blur the sophisticated textual details, leading to unnatural results. By contrast, our method can produce sharper edges and more faithful details with minimal artifacts as shown in Figure 6 (d). In Figure 7, we further compare our results with example-based methods, including neighbor embedding (NE)-based method [3], sparse coding (SC)-based method [17], and the method in [18]. The NE-based method uses the LLE algorithm to estimate the optimal weights to synthesize the HR images, which can introduce many high-frequency details. However, the resultant images tend to blur high-frequency details. The results of SC-based method can efficiently produce many plausible details from the LR-HR dictionary pairs. However, unwanted artifacts also are introduced. Moreover, note that there are noticeable ringing artifacts generated along salient edges. The method in [18] uses the Sparse-Land model similar to [17] to scale-up an LR image.



(a) ASDS [4] (b) Patch redundancy [8] (c) Sparse regression [10] (d) Our result

Figure 6. Comparison ($\times 3$) of SR results on parrot image. The results of [4] have richer details. The method of patch redundancy in [8] produces sharper edges, and method in [10] produces sharpening edges and visual details. However, the feathers shown in the red insets are not faithful to the LR input. Our result shows more reasonable textural details in the reconstructed HR image.

The results from it produce unfaithful details and jaggy artifacts along edges as shown in Figure 7 (c). Our method produces both sharper edges and more faithful details with less artifacts along major edges and in textural regions through combining the local, non-local and hallucination regularization priors. We also report the objective quality by the PSNR and SSIM indices[16] with the original HR images. We can see that our results achieve the top level results. Figure 8 demonstrate more challenging results ($\times 4$ magnification) on four real LR images with rich textures and salient edges in comparison to Gaussian process regression (GPR)-based SR approach in [9]. The GPR-based method predicts each pixel by its neighbors through GPR model. The presented results of GPR-based method are produced via carrying out two times magnification of a factor of 2. Our results are directly magnified by a factor of 4, which is more challenging than a step by step scheme used in [9] under larger magnification. All of these results demonstrate that our SR method performs better both in preserving edges and recovering visual details. Tables 1 summarizes the PSNR and SSIM scores [16] for compared methods on four representative images. Our method consistently outperforms other methods in all the test images.

6. Conclusion

In this paper, we proposed a novel single image SR approach by integrating example- and reconstruction-based SR methods. The example-based method estimates the high frequency details by jointly learning a multi-scale dictionary from a given LR image and produces expected details by sparse representation over the learnt multi-scale dictionary. The reconstruction-based method produces sharper edges and suppresses unwanted artifacts by taking local and non-local priors as regularization terms. Then, we presented a unified SR framework that incorporates the reconstruction constraint term, local and non-local regularization terms, and sparse hallucination regularization term. It is experimentally shown that the proposed methods can produce sharper edges and more faithful details in comparison to the other state-of-the-art SR approaches. Under the proposed SR framework, we can further improve the local and non-local regularization terms by incorporating other effective regression models,

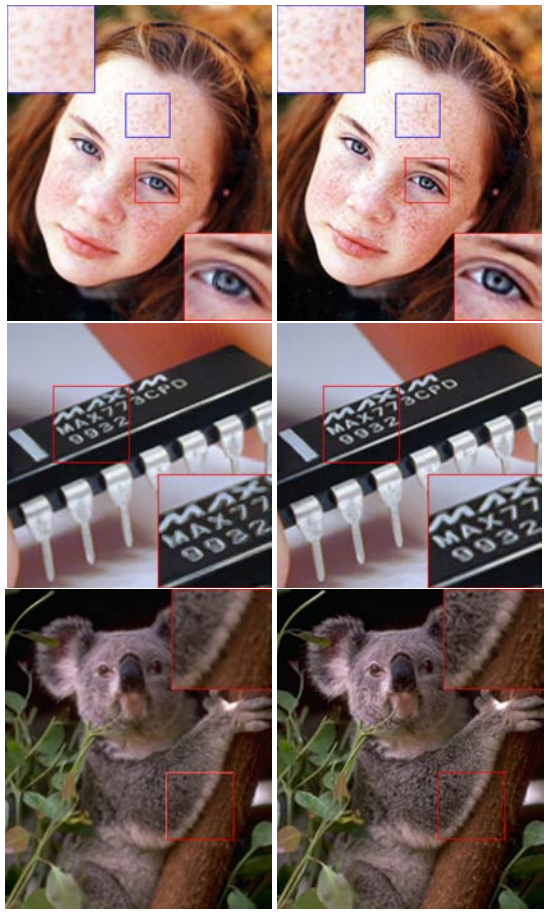
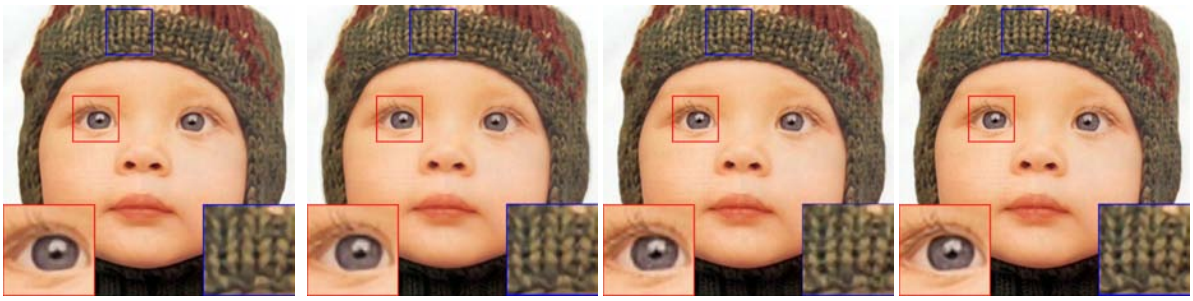


Figure 8. Comparison ($\times 4$) of SR results on face, chip, and raccoon images. The results of GPR in [9] shown at the left column have jagging artifacts along edges and blurring details in textural regions. Our results at the right column have less ringing artifacts and more reasonable details in all the cases.



(a) PSRN: 33.36, SSIM: 0.895 (b) PSRN: 33.83, SSIM: 0.902 (c) PSRN: 34.48, SSIM: 0.914 (d) PSRN: 35.01, SSIM: 0.922

Figure 7. Comparison ($\times 3$) of SR results on Child image by different example-based methods. (a) NE-based SR in [3]; (b)SC-based SR in [17]; (c) Results in [18]; (d) Proposed method. As shown, our method produces both sharper results with minimal artifacts and more faithful textural details on the hat of child.

Table 1. PSNR and SSIM scores.

PSNR	Zebra	Parrots	Butterfly	Child
Proposed	**23.49**	**30.02**	**24.81**	**35.01**
ASDS	22.41	29.44	24.33	34.89
Method in [8]	22.39	27.95	24.20	33.23
Sparse regression	22.89	29.64	24.54	34.88
SSIM				
Proposed	**0.864**	**0.911**	**0.886**	**0.922**
ASDS	0.838	0.908	0.875	0.918
Method in [8]	0.851	0.891	0.881	0.901
Sparse regression	0.849	0.905	0.882	0.919

such as Gaussian process regression [9] or non-local kernel regression [19]. Furthermore, other multi-scale analysis tools such as Gaussian pyramid based modeling also can be introduced to construct a multi-scale dictionary.

7. Acknowledgement

This work is supported by the National Basic Research Program of China (973 Program) (Grant No. 2012CB316400), the National Natural Science Foundation of China (Grant Nos. 61125204, 61172146, 60832005, 61125106, 91120302, 61072093, 61101250), the Fundamental Research Funds for the Central Universities, and the Ph.D. Programs Foundation of Ministry of Education of China (Grant No. 20090203110002); the State Administration of STIND (Grant No. B1320110042); and the Australian ARC discovery project (ARC DP-120103730).

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-svd: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE TSP*, 54(11):4311–4322, November 2006.

[2] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *SIAM MMS*, (2), 2005.

[3] H. Chang, D. Y. Yeung, and Y. Xiong. Super-resolution through neigbor embedding. In CVPR, 2004.

[4] W. Dong, L. Zhang, G. Shi, and X. Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE TIP*, 20(7):1838–1857, July 2011.

[5] W. Fan and D. Y. Yeung. Image hallucination using neighbor embedding over visual primitive manifolds. In CVPR, 2007.

[6] R. Fattal. Image upsampling via imposed edge statistics. *ACM Transactions on Graphics*, 26(3):95:1–95:8, 2007.

[7] W. T. Freeman, T. R. Jones, and E. C. Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002.

[8] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In ICCV, 2009.

[9] H. He and W. C. Siu. Single image super-resolution using gaussian process regression. In CVPR, 2011.

[10] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE TPAMI*, 32(6):1127–1133, January 2010.

[11] Z. Lin and H. Shum. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE TPAMI*, 26(1):83–97, 2002.

[12] J. Mairal, G. Sapiro, and M. Elad. Learning multiscale sparse representations for image and video restoration. *SIAM Multiscale Modeling and Simulation*, 7(1):214–241, April 2008.

[13] J. Sun, J. Sun, Z. Xu, and H. Shum. Image superresolution using gradient profile prior. In CVPR, 2008.

[14] Y. W. Tai, S. Liu, M. S. Brown, and S. Lin. Super resolution using edge prior and single image detail synthesis. In CVPR, 2010.

[15] H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. *IEEE TIP*, 16(2):349–366, 2007.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Quality assessment: from error measurement to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.

[17] J. Yang, J. Wright, Y. Ma, and T. Huang. Image super-resolution as sparse representation of raw image patches. In CVPR, 2008.

[18] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. *Curves & Surfaces*, pages 24–30, 2010.

[19] H. Zhang, J. Yang, Y. Zhang, and T. Huang. Non-local kernel regression for image and video restoration. In ECCV, 2010.

[20] L. Zhang and X. Wu. An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE TIP*, 15(8):2226–2238, August 2006.