

# Learning Parametric Distributions for Image Super-Resolution: Where Patch Matching Meets Sparse Coding

Yongbo Li, Weisheng Dong✉, Guangming Shi, Xuemei Xie

School of Electronic Engineering, Xidian University, Xi'an, China, 710071

yongboli@stu.xidian.edu.cn, {wsdong, xmxie}@mail.xidian.edu.cn, gmshi@xidian.edu.cn

## Abstract

*Existing approaches toward Image super-resolution (SR) is often either data-driven (e.g., based on internet-scale matching and web image retrieval) or model-based (e.g., formulated as an Maximizing a Posterior (MAP) estimation problem). The former is conceptually simple yet heuristic; while the latter is constrained by the fundamental limit of frequency aliasing. In this paper, we propose to develop a hybrid approach toward SR by combining those two lines of ideas. More specifically, the parameters underlying sparse distributions of desirable HR image patches are learned from a pair of LR image and retrieved HR images. Our hybrid approach can be interpreted as the first attempt of reconciling the difference between parametric and non-parametric models for low-level vision tasks. Experimental results show that the proposed hybrid SR method performs much better than existing state-of-the-art methods in terms of both subjective and objective image qualities.*

## 1. Introduction

Image super resolution (SR) aiming to reconstructed a high-resolution (HR) image from a single low-resolution (LR) image is a classic ill-posed inverse problem. Prior knowledge of the underlying HR image is required to compensate the loss of the information. To incorporate the prior knowledge, the image SR is commonly formulated as a *Maximum a Posterior* estimation problem, where a posterior probability is maximized to reconstruct a visually pleasant HR image. To faithfully reconstruct the HR image, the selection of the prior distribution of the underlying HR image is of great importance.

The sparsity prior of natural images has been extensively exploited in the literature. The gradient sparse prior is often employed to suppress the noise and other visual artifacts during the reconstruction [12, 8, 11]. However, fine image details are also often removed. Recently, coupled with dictionary learning, patch-based sparse coding approaches

have shown the popularity and effectiveness for image SR [6, 5, 18, 17, 19]. The sparse codes of the desired HR image patches are first inferred from the LR image patches, and then used to reconstruct the HR image patches. Due to the ill-posed nature of SR problem, the selection of an appropriate sparse regularizer is critical for the success of these methods. Typically, parametric sparse models (e.g., the Laplacian, Generalized Gaussian) are adopted. The model parameters can be either manually selected or estimated from the input LR images [18, 6, 19]. Advanced sparse methods considering the structural correlations between the patches have also been proposed [19, 7]. However, due to the lack of sufficient information, it is rather challenging to accurately estimate the sparse models from the LR image, especially for large scaling factors.

The learning-based methods have also attracted many attentions. In [10, 14, 18], a large number of LR/HR patch pairs are constructed and the high-frequency details of the input LR patches are inferred from the training LR/HR pairs. Recent studies show that the learning-based SR approaches can become more effective when the training images are more correlated with the given LR image [20, 15]. In [20, 15], similar images are first retrieved from a large image database, and then the high-frequency details are learned from the retrieved images. Despite its effective for SR the HR images, visual artifacts can often be observed due to the inconsistencies between the learned high-frequency details and the observed LR image.

To address the limitations of the sparsity-based and the learning-based methods, we propose a novel image SR approach that combines the sparse coding and the image retrieval based learning into a unified SR framework. For a given LR image, a set of HR images that are highly correlated with the desired HR image are retrieved from a database. Instead of learning the high-frequency details, we propose to learn the parametric sparse models for the desired HR patches from both the retrieved HR image and the LR image. Specifically, the Laplacian models with nonzero means are learned. With the learned sparse models, which are

much accurate than those learned from a single LR image, the sparse coding accuracy can be much improved, leading to significant improvements of the SR performance. To the best of our knowledge, this is the first time that the parametric nonzero-mean sparse models are learned from the retrieved correlated images. Experimental results on natural images show that the proposed method outperforms the current state-of-the-art image SR methods in terms of both PSNR and visual quality.

## 2. Related works

We briefly review the image SR methods, i.e., the sparse representation based methods and the learning-based methods, which are relevant and related to the proposed method.

**Sparse representation** based methods exploit the fact that the desired HR images admit sparse expansions with respect to a basis or dictionary. Given the observed the LR image  $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$ , where  $\mathbf{H} \in \mathbb{R}^{M \times N}$  is the blurring and downsampling matrix and  $\mathbf{n} \in \mathbb{R}^N$  denotes the additive Gaussian noise, the sparsity-based SR problem can be formulated as [6, 7]

$$(\mathbf{x}, \boldsymbol{\alpha}_i) = \underset{\mathbf{x}, \boldsymbol{\alpha}_i}{\operatorname{argmin}} \sum_i \|\mathbf{R}_i \mathbf{x} - \mathbf{D} \boldsymbol{\alpha}_i\|_2^2 + \lambda \sum_i \psi(\boldsymbol{\alpha}_i) \quad (1)$$

s.t.  $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 \leq \sigma_n^2$

where  $\mathbf{R}_i$  denotes the matrix extracting the patch at position  $i$  and  $\mathbf{D}$  denotes the dictionary,  $\psi(\cdot)$  denotes the sparse regularization term,  $\lambda$  are the regularization parameters. A common selection of  $\psi(\cdot)$  is the  $\ell_1$  norm. Recent studies show that structured sparse regularizer can lead to better results. In [7], an effective nonlocally centralized sparse representation (NCSR) model has been proposed, where the nonlocal structured sparsity is exploited. The NCSR method is very effective for image SR for small magnification factors (smaller than 3). However, for large magnification factors, it is rather challenging to obtain an accurate nonlocal estimation of the sparse codes and the NCSR becomes much less effective, as will be shown in the experimental results.

**Learning-based** methods propose to learn the high-frequencies from large training datasets. Generally, a large number of LR/HR patch pairs are built. Many efforts have been made to learn the mapping functions from the LR patches to the high-frequencies details of the HR patches [10, 18, 16, 3]. More recently, Yue *et al.* [20] propose to learn the fine image details from correlated HR images. A set of similar HR images are first retrieved from a image database. Then, a more correlated training set containing LR/HR patch pairs are constructed, from which the high-frequencies image details can be better learned. Despite its effectiveness for hallucinating the fine image details of the HR image, visual artifacts can also be often observed in the SR results by the method of [20]. The key idea of

the proposed method is to learn more accurate sparse prior distributions from both the correlated HR patches of the retrieved image and the nonlocal similar patches of the LR image. With the learned sparse prior, much better SR results have been achieved.

## 3. Sparse representation with learned sparse distributions

We first present the proposed sparse representation with learned nonzero-mean sparse models and then introduce the image retrieve and patch matching method for searching for similar HR image patches from a image database.

### 3.1. Sparse representation with nonzero-mean sparse distributions

Let  $\mathbf{x}_i \in \mathbb{R}^n$  denote a patch of size  $\sqrt{n} \times \sqrt{n}$  extracted from the image  $\mathbf{x}$  at position  $i$ . The sparse representation of  $\mathbf{x}_i$  under dictionary  $\mathbf{D} \in \mathbb{R}^{n \times K}$ ,  $K \geq n$  amounts to solving the following optimization problem

$$\boldsymbol{\alpha}_i = \underset{\boldsymbol{\alpha}_i}{\operatorname{argmin}} \|\mathbf{x}_i - \mathbf{D} \boldsymbol{\alpha}_i\| + \lambda \psi(\boldsymbol{\alpha}_i). \quad (2)$$

where  $\psi(\cdot)$  is a sparse regularizer. The sparse model of Eq. (2) can be interpreted as a MAP estimator, i.e.,

$$\boldsymbol{\alpha}_i = \underset{\boldsymbol{\alpha}_i}{\operatorname{argmax}} \{\log P(\mathbf{x}_i | \boldsymbol{\alpha}_i) + \log P(\boldsymbol{\alpha}_i)\}, \quad (3)$$

where the likelihood term  $\log P(\mathbf{x}_i | \boldsymbol{\alpha}_i)$  is characterized by the Gaussian distribution with variance  $\sigma_n^2$

$$P(\mathbf{x}_i | \boldsymbol{\alpha}_i) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{1}{2\sigma_n^2} \|\mathbf{x}_i - \mathbf{D} \boldsymbol{\alpha}_i\|_2^2\right). \quad (4)$$

It is easy to verify that when  $P(\boldsymbol{\alpha}_i)$  is chosen to be an IID zero-mean Laplacian distribution,  $\psi(\cdot)$  equals to the common  $\ell_1$  norm. Instead of assuming the zero-mean, we use the nonzero-mean IID Laplacian distribution, as

$$P(\boldsymbol{\alpha}_i) = \prod_j P(\alpha_{i,j}) \propto \prod_j \exp\left(-\frac{\sqrt{2} |\alpha_{i,j} - \beta_{i,j}|}{\theta_{i,j}}\right), \quad (5)$$

where  $\beta_{i,j}$  and  $\theta_{i,j}$  denotes the mean and standard deviation, respectively. By substituting Eqs. (4) and (5) into Eq. (3), we obtain

$$\boldsymbol{\alpha}_i = \underset{\boldsymbol{\alpha}_i}{\operatorname{argmin}} \|\mathbf{x}_i - \mathbf{D} \boldsymbol{\alpha}_i\|_2^2 + 2\sqrt{2}\sigma_n^2 \sum_j \frac{1}{\theta_{i,j}} |\alpha_{i,j} - \beta_{i,j}|. \quad (6)$$

The objective function of Eq.(6) is similar to the NCSR model proposed in [7]. However, our sparse model is developed from the MAP estimator with nonzero-mean sparse distributions.

Considering the rich repetitive structures in natural images, we extend the sparse model of Eq.(6) to simultaneous sparse coding. For each exemplar patch  $\mathbf{x}_i$ , we collect a set of patches similar to  $\mathbf{x}_i$  via patch matching in a large window centered at  $i$ . Let  $\mathbf{X}_i = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,m}] \in \mathbb{R}^{n \times m}$  denote the grouped patch set, whose each column corresponds to a patch (including  $\mathbf{x}_i$  itself). As each patch contain similar structures, we assume that the sparse codes of each patch can be characterized by the same parametric distribution  $P(\alpha_i)$ . This leads to the following simultaneous sparse coding

$$\mathbf{A}_i = \underset{\mathbf{A}_i}{\operatorname{argmin}} \|\mathbf{X}_i - \mathbf{D}\mathbf{A}_i\|_F^2 + 2\sqrt{2}\sigma_n^2 \sum_{k=1}^m \|\Lambda(\alpha_{i,k} - \beta_i)\|_1 \quad (7)$$

where  $\Lambda = \operatorname{diag}(\frac{1}{\beta_j})$  and  $\mathbf{A}_i = [\alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,m}] \in \mathbb{R}^{K \times m}$ . Similar to [7], the dictionary  $\mathbf{D}$  is chosen to be the local PCAs learned from each patch clusters generated by  $k$ -means clustering. Now the task is how to accurately estimate the distribution parameters  $\beta_i$  and  $\theta_i$ . A natural way is to iteratively estimate them from the set of similar patches  $\mathbf{X}_i$ . However, due to the lack of sufficient information, it is difficult to accurately estimate them from the LR image patches. In the next subsection, we will present the proposed approach to estimate  $\beta_i$  and  $\theta_i$  from both the input LR patches and the retrieved HR images.

### 3.2. Parametrical sparse distribution learning from reference images

Given a LR image  $\mathbf{y}$ , we retrieve a set of HR images from a large image database. The retrieved HR images containing similar contents to the LR image can be used to learn the sparse distributions of the desired HR patches. Specifically, we propose to learn the parametrical Laplacian distributions. This reduces to the learning of the means  $\beta_i$  and variances  $\theta_i^2$  for  $P(\alpha_i)$ . As the retrieved images may have different scales, illuminations and viewpoints. The retrieved images are aligned to the LR image with global registrations. After alignment, we search for similar HR patches from the aligned HR images, denoted as  $\{\mathbf{z}\}_{r=1}^R$ . The exemplar patches, denoted as  $\hat{\mathbf{x}}_i$  are extracted from the initially interpolated HR image. For the expression conveniences, we leave the details of the image retrieve, registrations and patch matching in the next subsection.

Let  $\{\tilde{\mathbf{z}}_{i,l}\}_{l=1}^L$  denote the set of similar patches that are within the  $L$ -th closest patches to  $\hat{\mathbf{x}}_i$  grouped from the retrieved image  $\{\mathbf{z}\}_{r=1}^R$ . Then, the expectation of  $\hat{\mathbf{x}}_i$  can be estimated as

$$\boldsymbol{\mu}_{i,1} = \sum_{l=1}^L w_{i,l} \tilde{\mathbf{z}}_{i,l} \quad (8)$$

where  $w_{i,l}$  is computed as  $w_{i,l} = (1/c_1) \exp(-\|\tilde{\mathbf{z}}_{i,l} - \hat{\mathbf{x}}_i\|/h_1)$ , wherein  $c_1$  is the normalization constant and  $h_1$  is

a predefined constant. In addition to the external reference patches, we can also obtain an estimate of the expectation from the nonlocal similar patches, as

$$\boldsymbol{\mu}_{i,2} = \sum_{k=1}^m w_{i,k} \hat{\mathbf{x}}_{i,k} \quad (9)$$

where  $w_{i,k} = (1/c_2) \exp(-\|\hat{\mathbf{x}}_{i,k} - \hat{\mathbf{x}}_i\|/h_2)$ , wherein  $c_2$  is the normalization constant and  $h_2$  is a predefined constant. Then, a more accurate estimate of the expectation of  $\beta_i$  can be obtained by the combination of  $\beta_{i,1}$  and  $\beta_{i,2}$ , as

$$\beta_i = \Delta \beta_{i,1} + (\mathbf{I} - \Delta) \beta_{i,2}, \quad (10)$$

where  $\beta_{i,1} = \mathbf{D}^\top \boldsymbol{\mu}_{i,1}$  and  $\beta_{i,2} = \mathbf{D}^\top \boldsymbol{\mu}_{i,2}$  denotes the representation coefficient vectors of  $\boldsymbol{\mu}_{i,1}$  and  $\boldsymbol{\mu}_{i,2}$  with respect to  $\mathbf{D}$ , respectively, and  $\Delta = \gamma \operatorname{diag}(\delta_j) \in \mathbb{R}^{K \times K}$ , where  $\delta_j$  are computed according to the energy ratio of  $\beta_{i,1}(j)$  and  $\beta_{i,2}(j)$ , i.e.,

$$\delta_j = \frac{r_j^2}{r_j^2 + 1/(r_j^2)}, \quad r_j = \beta_{i,1}(j)/\beta_{i,2}(j), \quad (11)$$

where the sparse vector with more energy is encouraged to take large weights.  $\gamma$  is defined as

$$\gamma = \begin{cases} 1, & d_{min} \leq T \\ \epsilon, & otherwise \end{cases} \quad (12)$$

where  $d_{min}$  denotes the minimum distance between the query patch and the candidate patches,  $T$  is a predefined threshold, and  $\epsilon$  is a small constant. With  $\gamma$  we can reduce the contributions from the dissimilar patches.

With the two sets of similar patches, the standard derivations  $\theta_i$  for each coefficients  $\alpha_i$  can also be estimated, as

$$\theta_i = \Delta \theta_{i,1} + (\mathbf{I} - \Delta) \theta_{i,2}, \quad (13)$$

where  $\theta_{i,1}$  and  $\theta_{i,2}$  are the maximum likelihood estimates of the standard derivations, computed as

$$\theta_{i,1} = \frac{1}{L} \sum_{l=1}^L (\tilde{\alpha}_{i,l} - \beta_{i,1})^2, \quad \theta_{i,2} = \frac{1}{L} \sum_{k=1}^m (\hat{\alpha}_{i,k} - \beta_{i,2})^2, \quad (14)$$

where  $\tilde{\alpha}_{i,l} = \mathbf{D}^\top \tilde{\mathbf{z}}_{i,l}$  and  $\hat{\alpha}_{i,k} = \mathbf{D}^\top \hat{\mathbf{x}}_{i,k}$ . In Fig. 1, we show the coefficients distributions learned by the proposed method and the NCSR, and the real distribution of the original image. It can be clearly seen that the learned distribution by the proposed method is more close to read distribution than that of NCSR, verifying the effectiveness of the proposed method.

### 3.3. Image retrieval and patch matching

The proposed framework of similar patches matching from the database is shown in Fig.2, which consists of three

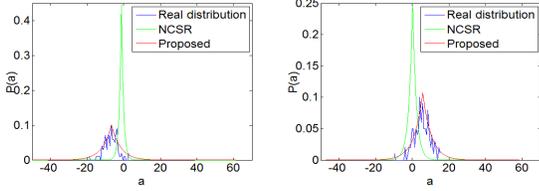


Figure 1: Coefficients distributions comparison on similar patches collected for exemplar patches of test image “a”. Here, the coefficients distributions associated with the 4<sup>th</sup> atom of the DCT dictionary is plotted. Similar observation can be obtained for other coefficients of other atoms.

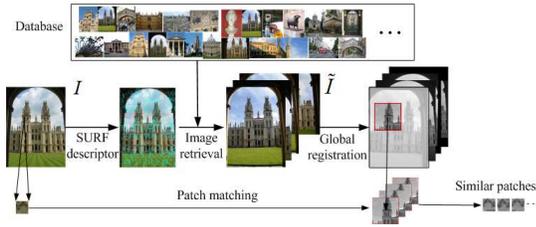


Figure 2: Framework of patch matching from database.



Figure 3: Results of image retrieval. (a) The query image; (b) The retrieved images.

steps, i.e., image retrieval, global registration and patch matching.

For image retrieve, the LR image is first interpolated to obtain an initial estimate of the original HR image, and then an image retrieval method is applied. Specifically, the SURF [2, 1] feature descriptors are used for image retrieval, and the method of [13] is employed to retrieve the correlated images. However, any existing image retrieval methods can be used in this work. Please refer to [13] for the details of the retrieval method. An example of image retrieval results is shown in Fig.3.

As the retrieved correlated images may be with different scales, illumination and viewpoints, they cannot be used for patch matching directly. For better patch matching accuracy, the correlated images should be aligned to the LR image. To register the retrieved image to the initially estimated HR image, the correspondences between the feature points of the initially recovered HR image and the correlated images are first established. Similar to [20], the SIFT feature descriptors are used to establish the correspondences. Then,

the registration homographic matrixes for each correlated image are estimated using the Random Sample Consensus (RANSAC) algorithm [9]. Finally, the retrieved HR images are then aligned to the LR image with the estimated homographic matrixes.

The transformed images, denoted as  $z_r$  are then used for patch matching. For an exemplar patch  $\hat{x}_i$  extracted from the initially interpolated HR image, we search for similar patches in  $z_r$ . As the exemplar patches  $\hat{x}_i$  are generally very small (e.g.,  $7 \times 7$ ) and thus contain little structural information of the original HR patches, the patch matching with  $\hat{x}_i$  is not accurate. To ensure the patch containing sufficient structural information, large patches are used for patch matching. Let  $\hat{p}_i \in \mathbb{R}^r$  denote the patch of size  $\sqrt{q} \times \sqrt{q}$  ( $q > n$ ) extracted from  $\hat{x}$  at position  $i$ . Considering the inconsistencies between the interpolated HR image and the retrieved high-quality HR images, we simulate the interpolated HR images for each retrieved HR image by first blurring and downsampling followed by interpolation. Let  $\hat{z}_r$  denote the interpolated version of  $z_r$ . Then, for a given query patch  $\hat{p}_i$ , we search for similar patches in each  $\hat{z}_r$  within a large searching window centered at position  $i$ . For robustness, the patch distance proposed in [20] is adopted, i.e.,

$$d(\hat{p}_i, \hat{q}_{i,l}) = \|\hat{p}_i - \hat{q}_{i,l}\|_2^2 + \rho \|\nabla \hat{p}_i - \nabla \hat{q}_{i,l}\|_2^2 \quad (15)$$

where  $\hat{q}_{i,l}$  denote the patches of size  $\sqrt{q} \times \sqrt{q}$  extracted within the searching windows of  $\hat{z}_r$ ,  $\rho$  is a constant ( $\rho = 10$  in our implementation) and  $\nabla$  denotes the operator extracting the gradients. Note that before computing the patch distance the DC components of the patches are removed. Generally, the value of the minimum distance (denoted as  $d_{min}$ ) between the query patch and the candidate patches indicates the quality of the patch matching. To better adapt to the local image structures, we adaptively adjust the size of the query patch  $\hat{p}_i$  according to  $d_{min}$  as in [20]

$$\sqrt{q} = \begin{cases} 21, & d_{min} \leq 1000 \\ 17, & 1000 < d_{min} \leq 1500 \\ 13, & 1500 < d_{min} \leq 2500 \\ 9, & d_{min} > 2500 \end{cases} \quad (16)$$

After patch matching, the patches that are within the first  $L$  closest to  $\hat{p}_i$  are selected, denoted as  $\hat{q}_{i,l}$ ,  $l = 1, 2, \dots, L$ . Then, smaller patches of size  $\sqrt{n} \times \sqrt{n}$  are extracted from  $z_r$  at the corresponding positions of  $\hat{q}_{i,l}$ . Finally, the matched HR patches are obtained as  $\tilde{z}_{i,l} = z_{i,l} - \hat{z}_{i,l} + \hat{x}_i$ . The grouped similar patches are then used to learn the sparse distributions, as described in the previous section. A much better estimate of the original HR image can then be obtained with learned sparse distributions, as will be described in the next section.

When a better estimate of the HR image is obtained, we directly use the patch  $\hat{x}_i$  to search for similar patches in the

retrieved images  $\mathbf{z}_r$ , using the Euclidean distance. After patch matching, the DC components of the grouped similar patches are set as that of  $\hat{\mathbf{x}}_i$ .

#### 4. Image super-resolution with learned sparse distribution

The image SR problem with the learned sparse model can be formulated as

$$(\mathbf{x}, \mathbf{A}_i) = \operatorname{argmin}_{\mathbf{x}, \mathbf{A}_i} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \eta \sum_i \{ \|\tilde{\mathbf{R}}_i \mathbf{x} - \mathbf{D}\mathbf{A}_i\|_F^2 + 2\sqrt{2}\sigma_n^2 \sum_{k=1}^m \|\Lambda_i(\alpha_{i,k} - \beta_i)\|_1 \}, \quad (17)$$

where  $\tilde{\mathbf{R}}_i \mathbf{x} \doteq [\mathbf{R}_{i,1}\mathbf{x}, \mathbf{R}_{i,2}\mathbf{x}, \dots, \mathbf{R}_{i,m}\mathbf{x}] \in \mathbb{R}^{n \times m}$  denotes the data matrix formed by the similar patches. A standard approach for solving Eq.(17) is to adopt alternative optimization.

##### 4.1. Solving for $\mathbf{A}_i$

For a fixed  $\mathbf{x}$ , Eq.(17) reduces to a sequences of simultaneous sparse coding problems, as

$$\mathbf{A}_i = \operatorname{argmin}_{\mathbf{A}_i} \|\tilde{\mathbf{R}}_i \mathbf{x} - \mathbf{D}\mathbf{A}_i\|_F^2 + 2\sqrt{2}\frac{\sigma_n^2}{\eta} \sum_{k=1}^m \|\Lambda_i(\alpha_{i,k} - \beta_i)\|_1, \quad (18)$$

which is equivalent to solving a sequences of sparse coding problem, i.e.,

$$\alpha_{i,k} = \operatorname{argmin}_{\alpha_{i,k}} \|\mathbf{R}_{i,k}\mathbf{x} - \mathbf{D}\alpha_{i,k}\|_2^2 + 2\sqrt{2}\frac{\sigma_n^2}{\eta} \|\Lambda_i(\alpha_{i,k} - \beta_i)\|_1. \quad (19)$$

With fixed  $\beta_i$  and  $\Lambda_i$ , the sparse coding problem of Eq.(19) is convex and can be iteratively solved by the surrogate based algorithm [4].

$$\alpha_{i,k}^{(t+1)} = \mathcal{S}_\tau(\mathbf{v}_{i,k}^{(t)} - \beta_i) + \beta_i, \quad (20)$$

where  $\mathbf{v}_{i,k}^{(t)} = \mathbf{D}^\top(\mathbf{R}_{i,k}\mathbf{x} - \mathbf{D}\alpha_{i,k}^{(t)})/c + \alpha_{i,k}^{(t)}$ ,  $\mathcal{S}_\tau(\cdot)$  denotes the soft-thresholding operator with threshold  $\tau = (2\sqrt{2}\sigma_n^2)/(c\eta\theta_i)$ , and  $c$  is a constant to guarantee the convexity of the surrogate function. Specifically, the sparse coding problem of Eq. (19) can be solved in just one step when the dictionary is orthogonal. In this work, the orthogonal local PCA dictionaries are used. Please refer to [6] for the details of local PCA dictionaries.

##### 4.2. Solving for the whole image $\mathbf{x}$

With fixed  $\mathbf{A}_i$ , the whole image can be reconstructed by solving

$$\mathbf{x} = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \eta \sum_i \|\tilde{\mathbf{R}}_i \mathbf{x} - \mathbf{D}\mathbf{A}_i\|_F^2 \quad (21)$$

---

#### Algorithm 1 Image SR with Learned Distributions

---

##### Initialization:

- Initialize  $\hat{\mathbf{x}}$  by bicubic interpolation and set  $\eta$  and  $\sigma_n$ ;
- Compute the local PCA dictionaries via  $k$ -means and PCA;
- Retrieve correlated images  $\{\mathbf{z}_r\}$  and align them to  $\hat{\mathbf{x}}$ ;

##### Outer loop:

 Iteration on  $t = 1, 2, \dots, T_{max}$ 

- Group a set of similar patches  $\mathbf{X}_i$  for each  $\hat{\mathbf{x}}_i$  from the current estimate of  $\mathbf{x}$ ;
- Search for similar patches  $\{\tilde{\mathbf{z}}_{i,l}\}$  to  $\hat{\mathbf{x}}$  from the retrieved images as described in sec.3.3
- Learn the nonzero expectations  $\beta_i$  and standard derivations  $\theta_i$  from both  $\mathbf{X}_i$  and  $\{\tilde{\mathbf{z}}_{i,l}\}$ ;
- Inner loop** (solve Eq.(17)): iteration on  $J = 1, 2, \dots, J$ ;  
  - Update sparse codes  $\mathbf{A}_i^{(j+1)}$  by solving Eq.(18);
  - Update the whole image  $\hat{\mathbf{x}}^{(j+1)}$  by computing Eq.(22);
  - Set  $\mathbf{x}^{(t+1)} = \mathbf{x}^{(j+1)}$  if  $j = J_{max}$  **End for**
- If  $\operatorname{mod}(t, t_0) = 0$ , update the PCA dictionaries  $\mathbf{D}$  using  $\mathbf{x}^{(t+1)}$ ;  
**End for**

##### Output: $\mathbf{x}^{(t+1)}$ .

---

which admits a closed-form solution as

$$\mathbf{x} = (\mathbf{H}^\top \mathbf{H} + \eta \sum_i \tilde{\mathbf{R}}_i^\top \tilde{\mathbf{R}}_i)^{-1} (\mathbf{H}^\top \mathbf{y} + \eta \sum_i \tilde{\mathbf{R}}_i^\top \mathbf{D}\mathbf{A}_i), \quad (22)$$

and  $\tilde{\mathbf{R}}_l^\top \tilde{\mathbf{R}}_l = \sum_{k=1}^m \mathbf{R}_k^\top \mathbf{R}_k$ ,  $\tilde{\mathbf{R}}_i^\top \mathbf{D}\mathbf{A}_i = \sum_{k=1}^m \mathbf{R}_k^\top \mathbf{D}\alpha_{i,k}$ . The matrix to be inverted in Eq.(22) is very large, we use the conjugate gradient algorithm to solve Eq.(22). The proposed SR algorithm is summarized in **Algorithm 1**. The local PCA dictionaries are updated in every  $t_0$  iterations to save computational complexity.

## 5. Experimental results

Seven images from the Oxford Building dataset<sup>1</sup> shown in Fig.4 are used as the test images. The other remaining images of Oxford building database are served as training images. The LR images are generated by first blurring the HR images with a  $7 \times 7$  Gaussian kernel with standard deviation of 1.6, then down-sampling with factors of 3, 4 and 6. For color images, we only apply the proposed method to the illumination channel and use bi-cubic interpolator for the chromatic channels. The main parameters of the proposed algorithm are set as follows:  $R = 4$  correlated images are retrieved, the number of similar patches from retrieved images and the initially reconstructed HR images are  $L = 10$

<sup>1</sup><http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/>



Figure 4: The test images. From left to right, the images are named as “a” to “g”

and  $m = 10$ , respectively,  $T_{max} = 10$  and  $J = 10$ . The proposed method is compared with three recently SR methods, i.e., the dictionary learning based SR method (denoted as DL-SR) [18], the image retrieved based SR method (denoted as IR-SR) [20] and the NCSR method [7]. Note that the NCSR is the current state-of-the-art reconstruction based method. For a fair comparison, we use the retrieved images to train the LR/HR dictionaries pair that will be used in the DL-SR method [18].

The PSNR and SSIM results of the test methods are reported in Table 1. It can be seen that the proposed method performs much better than the NCSR method [7]. The PSNR gain over NCSR can be up to 2.72dB. This verifies the effectiveness of the proposed method. When compared to the IR-SR method [20], the average PSNR gain is around 1dB.

Parts of the reconstructed images by the test methods are shown in Figs.5-6. We can see that the NCSR method [7] is very effective in recovering the large scale edges. However it fails to recover fine image details, especially for large scaling factors. The IR-SR method can better recover the fine image details. However, visual artifacts can be observed in the reconstructed image. The proposed method takes both advantages of the effective nonlocal sparse models and the retrieved-based methods. It can not only well reconstruct large edges but also well recover fine image details.

## 6. Discussions

In this paper, we proposed a novel image SR method that learns the sparse distributions from correlated images. For a given LR image, a set of high-quality images that similar to the LR image are first retrieved from a large image database. Then, the parametric sparse distributions are learned from both the LR image and the retrieved images. Specifically, the expectations and the standard derivations of the sparse codes are learned. With the learned sparse distributions, more accurate sparse regularization terms can be built. Experimental results show that the proposed method outperforms the current state-of-the-art image SR methods in terms both PSNR and visual quality.

**Limitations.** Despite the effectiveness in reconstructing HR images, the proposed method has some limitations for the general images, for which the image retrieval method cannot find correlated images. In Fig. 7, we show the SR results on two general images *Lena* and *Helicopter*. We can see that both the PSNRs of the proposed method

are comparable to those of NCSR. For this case, the proposed method reduces to the conventional sparsity-based SR method, where the sparse models are learned using the LR images only.

## 7. Acknowledgements

This work was supported in part by the Major State Basic Research Development Program of China (973 Program) under Grant 2013CB329402, in part by the Natural Science Foundation (NSF) of China under Grants(No. 61471281, 61472301, 61227004 and 61390512), in part by the Program for New Scientific and Technological Star of Shaanxi Province under Grant 2014KJXX-46, in part by the Fundamental Research Funds of the Central Universities of China under Grant BDY081424, in part by the Specialized Research Fund for the Doctoral Program of Higher Education (No. 20130203130001), and by the 111 Project (No. B07048).

## References

- [1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.
- [2] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *Computer vision—ECCV 2006*, pages 404–417. Springer, 2006.
- [3] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *Proc. of the IEEE CVPR*, volume 1, pages I–I, 2004.
- [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] W. Dong, D. Zhang, and G. Shi. Centralized sparse representation for image restoration. In *Proc. of the IEEE ICCV*, pages 1259–1266, 2011.
- [6] W. Dong, D. Zhang, G. Shi, and X. Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *Image Processing, IEEE Transactions on*, 20(7):1838–1857, 2011.
- [7] W. Dong, L. Zhang, G. Shi, and X. Li. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 22(4):1620–1630, 2013.
- [8] C. Fernandez-Granda and E. J. Candes. Super-resolution via transform-invariant group-sparse regularization. In *Proc. of the IEEE ICCV*, pages 3336–3343, 2013.
- [9] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image

Table 1: PSNR and SSIM results of the test methods for different scaling factors

Images	DL-SR [18]			NCSR [7]			IR-SR [20]			Proposed method		
	3	4	6	3	4	6	3	4	6	3	4	6
a	30.04	28.95	27.19	32.23	30.31	27.9	32.53	30.55	28.03	<b>33.79</b>	<b>31.79</b>	<b>29.00</b>
	0.8168	0.7749	0.7234	0.8734	0.8206	0.7531	0.8908	0.8496	0.7931	<b>0.9053</b>	<b>0.8691</b>	<b>0.8148</b>
b	25.20	24.31	22.84	26.83	25.43	23.44	26.78	25.21	23.20	<b>27.80</b>	<b>26.31</b>	<b>24.17</b>
	0.6958	0.6261	0.5430	0.7799	0.7099	0.6065	0.7794	0.7102	0.6003	<b>0.8140</b>	<b>0.7536</b>	<b>0.6536</b>
c	27.52	26.44	24.6	29.22	27.53	25.19	30.38	28.32	25.83	<b>31.24</b>	<b>29.36</b>	<b>26.87</b>
	0.7509	0.6930	0.6084	0.8298	0.7633	0.6594	0.8668	0.8032	0.6996	<b>0.8769</b>	<b>0.8227</b>	<b>0.7333</b>
d	27.71	26.87	25.21	29.21	27.83	25.95	29.16	27.71	25.64	<b>29.99</b>	<b>28.54</b>	<b>26.68</b>
	0.7824	0.7418	0.6909	0.8385	0.7824	0.7342	0.8384	0.7857	0.7240	<b>0.8632</b>	<b>0.8141</b>	<b>0.7573</b>
e	24.95	24.27	22.85	25.99	24.90	23.38	26.81	25.27	23.45	<b>27.68</b>	<b>26.15</b>	<b>24.29</b>
	0.7214	0.6752	0.6101	0.7889	0.7305	0.6506	0.8143	0.7561	0.6691	<b>0.8425</b>	<b>0.7899</b>	<b>0.7073</b>
f	28.24	27.28	25.76	30.17	28.65	26.61	30.29	28.77	26.92	<b>31.25</b>	<b>29.66</b>	<b>27.62</b>
	0.8157	0.7750	0.7177	0.8704	0.8312	0.7715	0.8794	0.8368	0.7780	<b>0.8976</b>	<b>0.8622</b>	<b>0.8043</b>
g	28.62	27.61	25.90	30.13	28.49	26.68	31.95	29.17	26.98	<b>32.85</b>	<b>30.38</b>	<b>27.99</b>
	0.7802	0.7304	0.6678	0.8454	0.7842	0.7049	0.9086	0.8476	0.7603	<b>0.9129</b>	<b>0.8646</b>	<b>0.7846</b>
Avg.	27.47	26.53	24.91	29.11	27.59	25.59	29.7	27.86	25.72	<b>30.66</b>	<b>28.88</b>	<b>26.66</b>
	0.7647	0.7166	0.6516	0.8323	0.7746	0.6971	0.8540	0.7985	0.7178	<b>0.8732</b>	<b>0.8251</b>	<b>0.7507</b>
Gain	<b>3.19</b>	<b>2.35</b>	<b>1.75</b>	<b>1.55</b>	<b>1.29</b>	<b>1.07</b>	<b>0.96</b>	<b>1.02</b>	<b>0.94</b>	–	–	–
	<b>0.1085</b>	<b>0.1085</b>	<b>0.0991</b>	<b>0.0409</b>	<b>0.0505</b>	<b>0.0536</b>	<b>0.0192</b>	<b>0.0266</b>	<b>0.0329</b>			

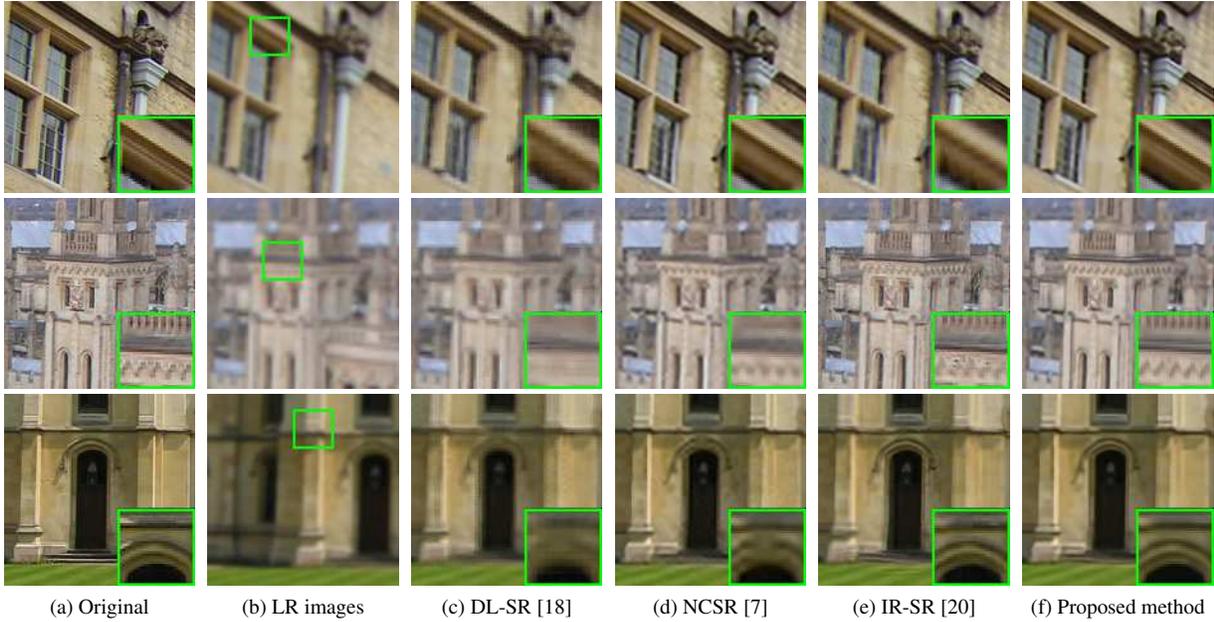


Figure 5: SR results of scaling factor 3.(a) Parts of the original images of “b” and “c” and “f”; (b) LR images; (c) DL-SR [18] (PSNR=25.20dB,27.52dB,28.24dB); (c) NCSR [7] (PSNR=26.83dB,29.22dB,30.17dB); (d) IR-SR [20] (PSNR=26.78dB,30.38dB,30.29dB); (e) **Proposed method** (PSNR=**27.80dB,31.24dB,31.25dB**)

analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

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

[11] C. Liu and D. Sun. On bayesian adaptive video super resolution. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(2):346–360, 2014.

[12] A. Marquina and S. J. Osher. Image super-resolution by tv-regularization and bregman iteration. *Journal of Scientific Computing*, 37(3):367–382, 2008.

[13] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Proc. of the IEEE CVPR*, pages 1–8, 2007.

[14] J. Sun, Z. Xu, and H.-Y. Shum. Image super-resolution using gradient profile prior. In *Proc. of the IEEE CVPR*, pages 1–8,

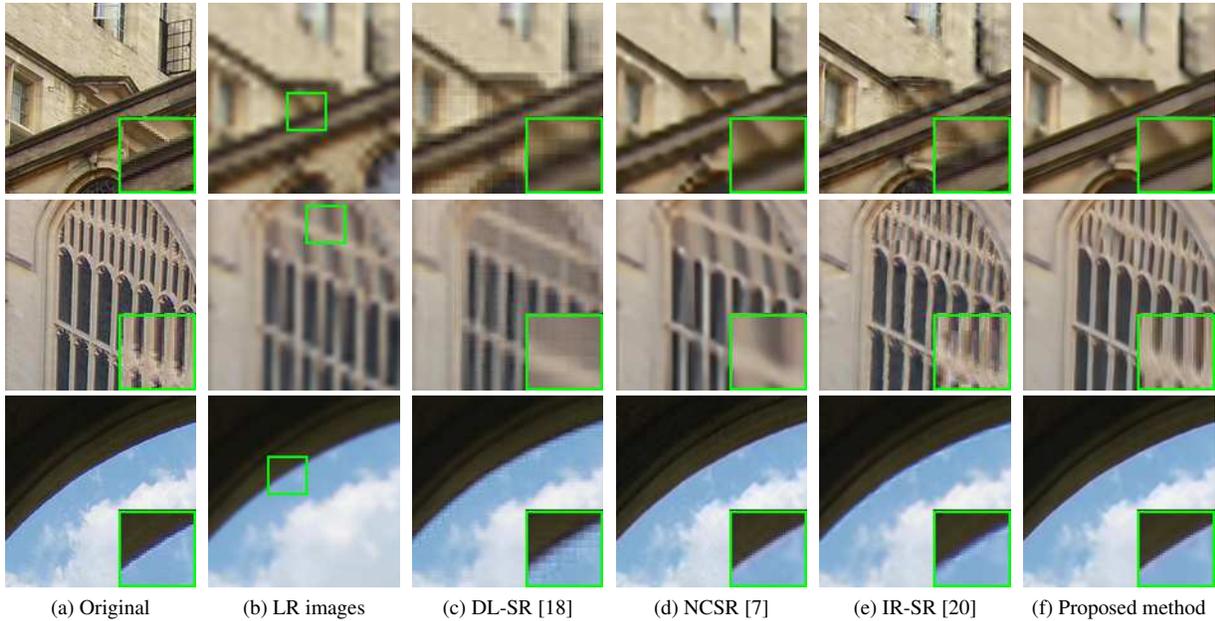


Figure 6: SR results of scaling factor of 6. (a) Parts of the original images of “b”, “c” and “f”; (b) LR images; (c) DL-SR [18] (PSNR=22.84dB,24.60dB,30.17dB); (d) NCSR [7] (PSNR=23.44dB,25.19dB,26.61dB); (e) IR-SR [20] (PSNR=23.20dB,25.83dB,26.92dB); (f) **Proposed method** (PSNR=24.17dB,26.87dB,27.62dB)

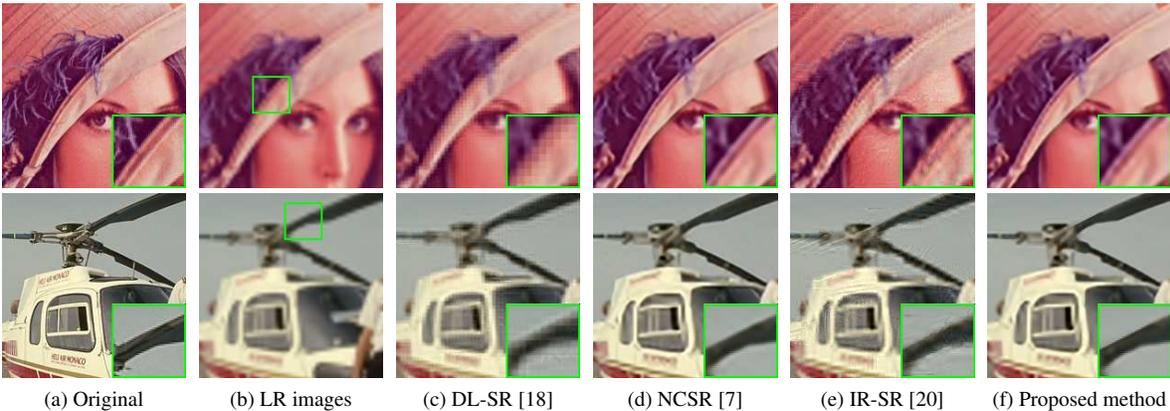


Figure 7: SR results of on general images (scaling factor of 4). (a) Parts of the original images of *Lena* and *Helicopter*; (b) LR images; (c) DL-SR [18] (PSNR=28.99dB,23.13dB); (d) NCSR [7] (PSNR=31.63dB,24.82dB); (e) IR-SR [20] (PSNR=27.73dB,23.31dB); (f) **Proposed method** (PSNR=31.43dB,24.76dB)

2008.

- [15] L. Sun and J. Hays. Super-resolution from internet-scale scene matching. In *Proc. of the IEEE ICCP*, pages 1–12, 2012.
- [16] Z. Wang, Y. Yang, Z. Wang, S. Chang, J. Yang, and T. S. Huang. Learning super-resolution jointly from external and internal examples. *arXiv preprint arXiv:1503.01138*, 2015.
- [17] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In

*Proc. of the IEEE CVPR*, pages 1–8, 2008.

- [18] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *Image Processing, IEEE Transactions on*, 19(11):2861–2873, 2010.
- [19] G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *Image Processing, IEEE Transactions on*, 21(5):2481–2499, 2012.
- [20] H. Yue, X. Sun, J. Yang, and F. Wu. Landmark image

super-resolution by retrieving web images. *Image Processing, IEEE Transactions on*, 22(12):4865–4878, 2013.