Sparse Representation Based Super Resolution Using Saliency and Edge Information

Saboya Yang, Jiaying Liu*, Wenhan Yang, Zongming Guo
Institute of Computer Science and Technology, Peking University, Beijing, P.R.China, 100871

***Abstract*—Sparse representation provides effective prior information for single-frame super resolution reconstruction. The diversified training samples of the general dictionary lead to the difficulty of recovering fine grained details due to the negligence of redundant structural characteristics. Thus, the dictionary which is adaptive to local structures is needed. Considering the highly structured information of saliency and edge regions, we present a novel sparse representation based super resolution approach. Salient regions are segmented to train the saliency dictionary. The same is true for edge regions. Thus, more adaptive dictionaries are acquired. When reconstructing the input image, dictionaries are chosen adaptively and then more clear details are achieved. Objective quality evaluation shows that our proposed algorithm achieves highest PSNR results comparing with the state-of-the-art methods. And subjective results demonstrate the proposed method reduces artifacts and preserves more details.**

I. Introduction

Super resolution (SR) aims to reconstruct a high resolution (HR) image from one or a series of low resolution (LR) images. It is an ill-posed inverse problem because a large amount of information may be lost in the degradation process. Thus, one of the important solutions is to provide external information for SR.

In the past decades, researchers made progress in SR. They have found from the early stage of mammal's visual processing system that the pixel set of natural images has less variability than a completely random set [1]. And in natural images, instead of the same patches, two similar patches can always be found. Taking this phenomenon into account, a relatively large database is built to preserve correspondences between HR and LR patches. It provides external information for SR. Freeman *et al.* [2] utilized Markov network to build the correspondences. This algorithm is quite time-consuming. Meanwhile, Chang *et al.* [3] adopted neighbor embedding into SR and reduced the number of training sets immensely. Since this method needs to decide the fixed number of neighbors artificially, it tends to result in under- or over-fitting.

To solve the above problem, Yang *et al.* [4] introduced sparse representation into SR. Image sparse representation refers that a patch can be approximately expressed as a linear combination of few prespecified atom patches. And the sparsity constraint is that few of linear coefficients are nonzeros. While each atom patch refers to a dictionary base, all atom patches combine to be an over-complete dictionary. In this method ScSR[4], Yang exploited linear programming to solve the sparse representation of LR patches. It is easy to understand and operate, which attracts much attention. However, ScSR extracts samples from arbitrary regions to train the dictionary. When reconstructing one specific object, the dictionary which represents individual structural characteristics is needed. Nowadays, many researchers focus on training a dictionary with general representativeness and containing different kinds of image structural properties [5]. Nonetheless, the sparse decomposition based on the over-complete dictionary is unstable for complex natural images and may produce some artifacts. This means when exploiting general dictionary, it is hard to reconstruct local features adaptively. Therefore, there is still much work to be done to acquire more adaptive dictionaries.

From the view of biological vision, when looking at one image, people usually focus on salient and edge regions. And from another view of the scientific analysis, human eyes are especially sensitive to structural information. Generally, salient regions and edge regions tend to be highly structured. Taking this property into consideration, salient and edge regions can be extracted to train specific dictionaries. And if these dictionaries are used to reconstruct corresponding regions, more details can be captured. Guided by this intuition, saliency and edge information of images is drawn into the framework to enhance the adaptivity of dictionary.

In a word, the contributions of this paper are

- obtaining adaptive dictionaries using saliency and edge information;
- reconstructing the input image by corresponding dictionaries adaptively;
- testing the idea of category-oriented dictionary and obtaining good results.

The rest of the paper is organized as follows. In Sec.II, sparse representation based SR algorithm is reviewed. Sec.III focuses on the sparse representation based SR using saliency and edge information algorithm. Then in Sec.IV the experimental results can be seen. And in the end, a brief conclusion is presented in Sec.V.

II. Overview of Sparse Representation Based SR

The traditional sparse representation SR method [4] reconstructs the image by HR and LR dictionaries. It is based on the assumption that there exists a certain relationship between LR and HR sparse coefficients. The original image is segmented

*Corresponding author
This work was supported by National High-tech Technology R&D Program (863 Program) of China under Grant 2014AA015205, Beijing Natural Science Foundation under contract No.4142021 and National Natural Science Foundation of China under contract No.61201442.

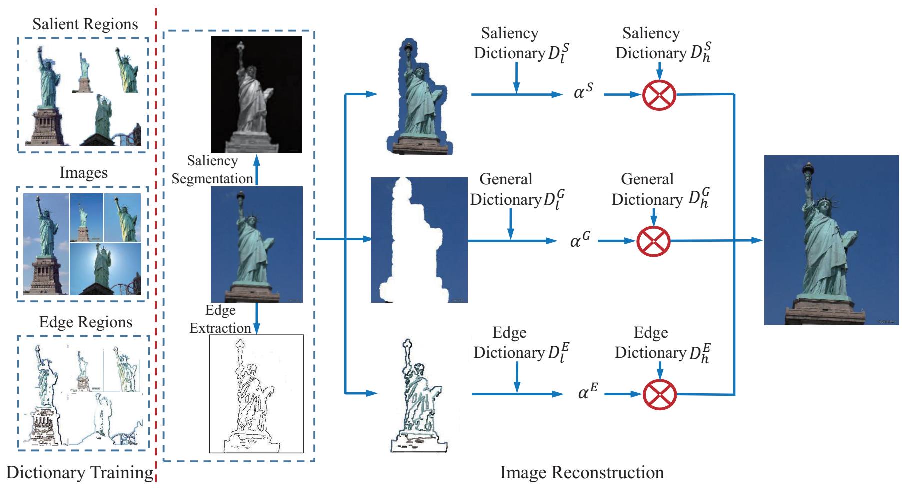


Fig. 1. The Framework of the Sparse Representation Based SR Using Saliency and Edge Information Algorithm.

into small patches first and then expressed as a linear combination of local primitives under the sparsity constraint. Next, the method is split into two steps: dictionary learning and image reconstruction. Dictionary learning is actually the process of searching for the optimal basis under sparse representation. And the optimal basis needs to satisfy the constraint of the uniqueness of sparse representation. To formulate the problem, let D be a over-complete dictionary, and D_i is the dictionary base of D. $\|\cdot\|_0$ equals the number of nonzero elements in one vector, which is actually the sparsity constraint. And it is replaced by $\|\cdot\|_1$ to solve the non-convex problem. Meanwhile, y_i corresponds to each training sample and sparse coefficient α represents the sparse representation of y_i in dictionary D. λ is the regularization parameter.

$$\begin{aligned} &\arg\min_{D,\alpha} \sum_i \|y_i - D\alpha\|_2^2 + \lambda\|\alpha\|_1, \\ &s.t. \|D_i\|_2^2 \leq 1, i = 1, 2, ..., n. \end{aligned} \tag{1}$$

In (1), D and α are terms to be solved. When these two terms are both unknown, the problem becomes non-convex, which is hard to solve. But if one term is fixed to solve another, the problem is a convex optimization problem. Therefore, when learning the dictionary D, as (2), the sparse coefficient α is initialized at first to update the dictionary D. Then in reverse, α is updated based on D. After iterating in this way, the final dictionary D is obtained.

$$\begin{aligned} &D = \arg\min_{D,\alpha} \sum_i \|y_i - D\alpha\|_2^2 + \lambda\|\alpha\|_1, \\ &s.t. \|D_i\|_2^2 \leq 1, i = 1, 2, ..., n. \end{aligned} \tag{2}$$

The same is true when solving the sparse coefficient α. The dictionary D is initialized and used to update the sparse coefficient α. Then D is updated based on α. After iterations, the final sparse coefficient α is acquired.

$$\begin{aligned} &\alpha = \arg\min_{D,\alpha} \sum_i \|y_i - D\alpha\|_2^2 + \lambda\|\alpha\|_1, \\ &s.t. \|D_i\|_2^2 \leq 1, i = 1, 2, ..., n. \end{aligned} \tag{3}$$

But this method did not consider redundant structural characteristics, which affected the adaptivity of the dictionary. Thus, we propose a sparse representation based SR algorithm using saliency and edge information.

III. Sparse Representation Based SR Using Saliency and Edge Information

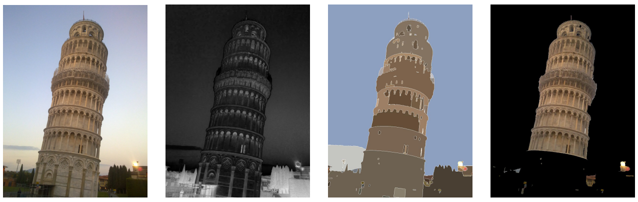
Taking the highly structured property of salient regions and edges into consideration, we propose a SR algorithm using saliency and edge information based on sparse representation. The framework of the proposed algorithm is illustrated in Fig.1. It is separated into two stages: dictionary training and image reconstruction.

In the dictionary training stage, besides training a general dictionary, we respectively extract patches from salient regions and edge regions to train customized saliency dictionary and edge dictionary. More details can be viewed in Sec.III-A and Sec.III-B. The salient regions are segmented based on the intuition that they are more visually notable comparing with the neighborhood. And the edges are extracted by Canny detector. In Sec.III-C, the image is reconstructed by improved SR based on sparse representation. As shown in the right side of Fig.1, the image is reconstructed after segmentation. Different regions are restored utilizing the corresponding dictionaries.

A. Saliency Dictionary Learning

The distinction of training samples distinguishes the saliency dictionary and the general dictionary. While training the saliency dictionary, patches are sampled from salient regions instead of arbitrary regions. To achieve this, the salient regions have to be detected first. As a matter of fact, it is easy to find the patches extracted from the salient regions highly relevant, considering the highly structured feature of salient regions. And there remain few unrelated patches in the training samples. Therefore, the obtained dictionary is highly adapted with the salient regions in the input image, which contributes to improve the performance of the algorithm.

Inspired by [6], salient regions are segmented using the contrast filter. To filtrate and obtain regions with larger contrast, the distance between the average feature vectors of the pixels in the subregion and in the neighborhood is measured in CIELab color space. The saliency map shown in Fig.2(b) is



(a) (b) (c) (d)

Fig. 2. An example of salient region segmentation. (a) Original image, (b) Saliency map, (c) Saliency segmentation, (d) Salient region.

acquired after calculating under different scales. It illustrates what human eyes feel. And every pixel value in the saliency map equals the saliency value of this pixel. Afterwards, the map is over-segmented by K-means, where the K seeds for K-means are determined utilizing the hill-climbing algorithm [7]. Then many pixels are gathered together as indicated in Fig.2(c). After calculating average saliency value for each region, we define the regions with average saliency value higher than the threshold as salient regions. Finally, the segmentation result in Fig.2(d) is generated.

With the segmentation result, only patches in the salient regions are extracted as samples y_i^S to train the saliency dictionary D^S as follows:

$$\begin{aligned} &D^S = \arg\min_{D^S,\alpha^S} \sum_i \|y_i^S - D^S\alpha^S\|_2^2 + \lambda\|\alpha^S\|_1, \\ &s.t.\|D_i^S\|_2^2 \le 1, i = 1, 2, ..., n. \end{aligned} \tag{4}$$

B. Edge Dictionary Learning

Similar to the saliency dictionary, samples to train edge dictionary are extracted only from edge regions. Therefore, edges need to be detected in advance. And the edge dictionary can represent characteristics of edge regions because edge information has the capacity to independently display inherent properties of objects' surface shape.

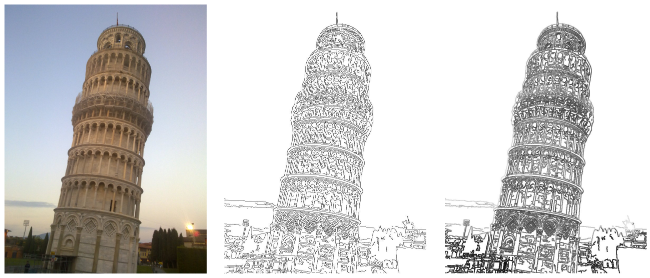
Edges are actually discontinuous regions in which luminance changes rapidly. Ideally, applying edge detection algorithms to the input image results in a group of continuous curves. And these curves clearly show the boundaries of objects, the curved boundaries of the discontinuous surfaces and so on. Hence the edge detection algorithms help get rid of irrelevant information. In this way, the computation cost is reduced and the structural properties of the objects are reserved. As a result, we can extract the edges effectively. In this paper, Canny detector is chosen to extract edges as Fig.3(b). And in Fig.3(c), the edge regions are defined as regions around the edges with the specified width.

After extracting edges, only patches in the edge regions are extracted as samples y_i^E to train the edge dictionary D^E as follows:

$$\begin{aligned} &D^E = \arg\min_{D^E,\alpha^E} \sum_i \|y_i^E - D^E\alpha^E\|_2^2 + \lambda\|\alpha^E\|_1, \\ &s.t.\|D_i^E\|_2^2 \le 1, i = 1, 2, ..., n. \end{aligned} \tag{5}$$

C. Image Reconstruction Using Saliency and Edge Information

According to the traditional sparse representation SR algorithm, the sparse coefficient α is calculated as (3). Assuming that LR dictionary D_l and HR dictionary D_h share the same sparse coefficient, the HR image can be calculated by multiplying sparse coefficient α and HR dictionary D_h. And $x \in R^m$ is the vector expression of the patch.



(a) (b) (c)

Fig. 3. An example of edge region extraction. (a) Original image, (b) Edge extraction, (c) Edge region.

$$x = D_h\alpha. \tag{6}$$

After obtaining the saliency dictionary and the edge dictionary, general dictionary is trained using patches from arbitrary regions. In the proposed algorithm, different regions are reconstructed by different dictionaries as shown in (7). Edge regions utilize the LR edge dictionary D_l^E to do sparse decomposition and obtain the corresponding edge sparse coefficients α^E. And when multiplying the edge sparse coefficients α^E and the HR edge dictionary D_h^E, the HR edge regions are obtained. Same as edge regions, salient regions use the LR saliency dictionary D_l^S and obtain the corresponding saliency sparse coefficients α^S. Then the HR salient regions are obtained after multiplying the HR saliency dictionary D_h^S. In the meantime, other regions are reconstructed using the corresponding LR general dictionary D_l^G and HR general dictionary D_h^G. Finally, the super-resolved image is obtained.

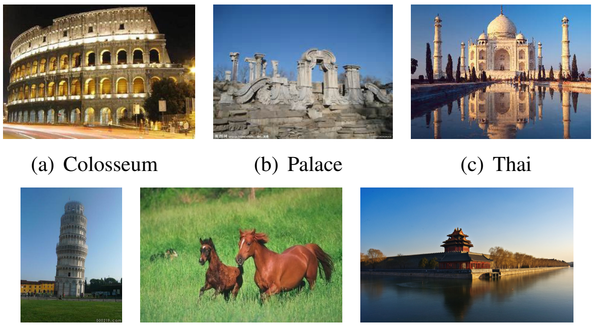
$$x = \begin{cases} D_h^E\alpha^E, x \in Edge \\ D_h^S\alpha^S, x \in Saliency \\ D_h^G\alpha^G, x \in General \end{cases}. \tag{7}$$

IV. Experimental Results

To evaluate the effectiveness of the proposed method, we conduct experiments of $2\times$ super resolution on several test sets. The LR input images are generated from the original HR images by bicubic downsampling. We collect six image sets from the Internet to test the algorithm. As indicated in Fig.4, they are *Colosseum*, *Palace*, *Thai*, *Tower*, *Horse* and *Turret*. And take image set *Tower* as an example, all images in set *Tower* are related to the same object taken from different views at different time in Fig.5. These datasets have been released on our website*.

In experiments, the LR patch size is 7×7, and the overlap between LR patches is $[6, 6]$. The salient regions and the edge regions are extracted to train the corresponding saliency dictionary and edge dictionary. We compare the proposed algorithm with the baseline Bicubic method and ScSR [4]. The algorithm is evaluated in both objective and subjective ways.

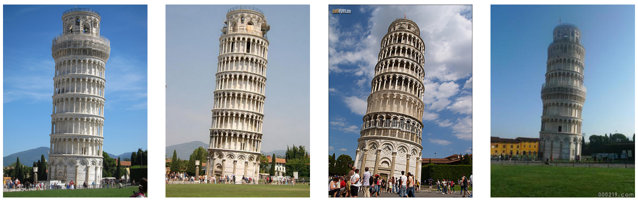
*http://www.icst.pku.edu.cn/course/icb/Projects/SRSEI.html



(a) Colosseum (b) Palace (c) Thai

(d) Tower (e) Horse (f) Turret

Fig. 4. Test Image Set



(a) (b) (c) (d)

Fig. 5. *Tower* Image Set

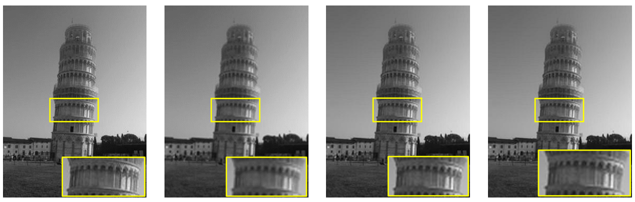
The objective results are measured by Peak Signal to Noise Ratio (PSNR) in Table.I. The average gain of our proposed algorithm is 0.05 dB over ScSR. And the subjective results on *Tower* are illustrated in Fig.6.

From the zoomed comparison of the highlighted part in the original image by different methods in Fig.6, the proposed method outperforms ScSR. Compared with ScSR, our method reduces blocking effects and artifacts along the edges in Fig.6(d) on account of the specific dictionaries. It proves that the saliency and edge information contributes to obtain adaptive dictionaries and recover more structural information.

TABLE I

PSNR COMPARISON OF SR RESULTS ON TEST IMAGES

Sequence	Bicubic	ScSR	Proposed
Colosseum	24.50	25.69	**25.75**
Palace	29.43	30.95	**31.03**
Thai	24.21	24.94	**25.00**
Tower	32.03	33.21	**33.26**
Horse	29.87	30.90	**30.94**
Turret	32.50	34.04	**34.05**



(a) Original (b) Bicubic (c) ScSR (d) Proposed

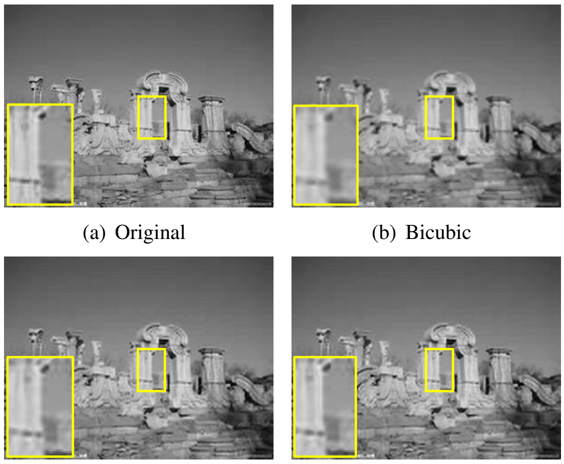
Fig. 6. Subjective comparison of different algorithms on *Tower*. (a) Original image, (b) Bicubic, (c) ScSR, (d) Proposed.

Meanwhile, we extend the algorithm from object-oriented to category-oriented. In one image set, there may exist images about the same category of objects, which means the limitation is not only the same object. In this way, we combine four image sets *Colosseum*, *Palace*, *Thai* and *Tower* into one *Architecture* set. After re-training the dictionaries, we obtain the comparison results presented in Table.II and Fig.7. According to the results, we realize that the proposed method still improves the performances. This leads to the conclusion that our method has relatively comprehensive adaptability, and it is not limited to the same object.

TABLE II

PSNR COMPARISON OF SR RESULTS ON TEST IMAGES

Sequence	Bicubic	ScSR	Proposed
Colosseum	24.50	25.69	**25.75**
Palace	29.43	30.95	**31.07**
Thai	24.21	24.94	**25.04**
Tower	32.03	33.21	**33.34**



(a) Original (b) Bicubic

(c) ScSR (d) Proposed

Fig. 7. Subjective comparison of different algorithms on *Palace*. (a) Original image, (b) Bicubic, (c) ScSR, (d) Proposed.

V. Conclusions

In this paper, according to sparse representation SR framework, we focus on how to acquire a more adaptive dictionary. Salient regions and edge regions are extracted to train saliency and edge dictionaries. Different regions of the input image are restored using corresponding dictionaries to acquire the final HR image. Experimental results indicate the proposed method outperforms other methods in both objective and subjective quality.

References

[1] D. J. Field. What is the goal of sensory coding. *Neural Computation*, July 1994, 6(4): 559-601.

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

[3] H. Chang, D. Y. Yeung, Y. Xiong. Super-resolution through neighbor embedding. *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, 275-282.

[4] J. Yang, J. Wright, T. Huang, Y. Ma. Image super resolution via sparse representation. *IEEE Transactions on Image Processing*, Nov 2010, 19(11): 2861-2873.

[5] M. Protter, I. Yavneh, M. Elad. Closed-form mmse estimation for signal denoising under sparse representation modeling over a unitary dictionary. *IEEE Transactions on Signal Processing*, July 2010, 58(7): 3471-3484.

[6] R. Achanta, F. Estrada, P. Wils, S. Susstrunk. Salient region detection and segmentation. *International Conference on Computer Vision Systems*, 2008, 66-75.

[7] T. Ohashi, Z. Aghbari, A. Makinouchi. Hill-climbing algorithm for efficient color-based image segmentation. *IASTED International Conference On Signal Processing, Pattern Recognition, and Applications*, 2003.