Single Image Super-Resolution Using Joint Dictionary

Mr. Kiran Jadhav, Dr. Ramesh Kulkarni, Mrs. Ashwini S.Sawant
(M.E Student) (Professor) (Assistant Professor)
Electronics and Telecommunication Department
V.E.S. Institute of Technology, Mumbai.
kjkiranjadhav7@gmail.com, ramesh.kulkarni@ves.ac.in, ashwini.sawant@ves.ac.in

Abstract— In this paper, we utilize the application of sparse representation on single image super-resolution problems. It is found from the image statistics that image patches can be well-represented as a sparse linear combination of elements from an appropriately selection over-complete trained dictionary. In this algorithm, a sparse representation for each patch of the low-resolution image, and then the coefficients in the low-resolution domain used for reconstruction of the high-resolution counterpart. By jointly training the low-resolution and high-resolution dictionaries and selecting the best match for the local patch, a super-resolution image is generated. By increasing the similarity index between the low-resolution and high-resolution local patches, it is possible to get best results. Experimental results on single image super-resolution demonstrate the superiority of the proposed method.

Index Terms— Bicubic interpolation, Joint dictionary, Sparse coding, Sparse representation, Super-resolution.

1 INTRODUCTION

Super-Resolution is termed as a set of methods whose aim is to improve the spatial resolution of images. It is a technique to increase the resolution of an imaging system. It also refers as the problem of constructing high-resolution image from one or multiple low-resolution images. In effect, super-resolution extrapolates the high frequency components, and minimizes the aliasing and blurring during the image capturing process. Therefore, super-resolution can be distinguished in two main families: single image super-resolution method and multi-frame super-resolution method. Super-resolution is a very intersecting area in research. It is considered to be one of the most promising techniques that can help to overcome the limitations due to optics and sensor resolution. Based on the multiple images super-resolution, many methods have been proposed (e.g., [1], [2], [3]).

The goal of this paper is that constructing high-resolution image form single observed low-resolution image, with the help of the joint training dictionary. Therefore, it is referred as "single image super-resolution method". Super-resolution techniques have been proved useful in case where greater clarity in image is required. Super-resolution has a wide range of applications such as satellite imaging, medical imaging (magnetic resonance imaging (MRI) and computer tomography (CT)), and object recognition. Super-resolution can be very useful in case where multiple images of the same scene are easily obtained and help for getting more details from them. For surveillance and forensic purposes, it is often required to get more information of region of interest (ROI).

The super-resolution task is typical example of an inverse problem of recovering the original high-resolution image from one or more the low-resolution images. Linear formulation of the imaging model enables to formulate the super-resolution problem. This problem is ill-posed because of the lack of availability of low-resolution images and unknown blurring operators. Since there is no direct solution to the ill-posed problem, regularization procedures are necessary to stabilize the inversion of this ill-posed problem, such as [1], [4], [5].

However, when number of available low-resolution images is small or large magnification factor, the functionality of the reconstruction-based super-resolution algorithm degrades rapidly which may be produced smooth output but losing important high frequency details [2]. Simple resolution enhancement methods have been commonly used in image processing for noise removal based on smoothing and interpolation techniques. Another category of SR algorithm is based on interpolation [6], [7], [8]. Bilinear and Bicubic interpolation are the simple methods that generate smooth images with ringing and jagged artifacts. By exploiting the natural image priors, interpolation will give more favorable results.

Machine learning technique is third category of super-resolution algorithm. In this technique, it tries to capture the coexistent prior between low-resolution and high-resolution image patches. An example based learning method [9] learned that how to predict low-resolution to high-resolution via a Markov Random Field (MRF) solved by belief propagation. This method extends Primal Sketch priors [10] is developed to enhance blurred edges, ridges and corners. In order to overcome this problem, Locally Linear Embedding (LLE) [11] proposed via multiple repetitive learning. Due to over- or under-fitting problem, blurred results are obtained by using a fixed number K neighbors. Super-resolution method [12] is presented new approach based on sparse representation to thwart super-resolution problem. For training a dictionary, it requires a multiple training images which consumes lots of time.

In this paper, we proposed most flexible method that is based on sparse representation using sparse coding which helps to escape over- or under-fitting problem of Locally Linear Embedding (LLE) [11] and therefore, it gives better quality results. To avoid the above difficulties, sparse representation is efficient and recent method used to improve quality of an image. However, sparse coding is taken too much time when it is applied over a large sampled image patch database directly.

This paper focuses on the problem of generating the super-resolution version from input low-resolution image. In previously mentioned approach [13], we will depend on patches from the input image. However, instead of that we learned complex representation by directly working on the image patch pairs sampled from high- and low-resolution images [13], which helps to improve speed and complexity of the algorithm. Accurate recovery is very difficult because of the ill-posed nature of super-resolution problem. But sparse representation of the image patch provides both effectiveness and robustness which helps in regularizing the inverse problem.

Now sparse representation is used in many fields related to inverse problems. Sparse representation has been successfully applied in many area of image processing such as denoising [14] and restoration [15]. To learn an over-complete dictionary from natural image patches, the K-SVD algorithm [16] is used and successfully applied to the image denoising problem. In our setting, we will work with two joint dictionaries, D_h for high-resolution patches, and D_l for low-resolution patches. The sparse representation of a low-resolution patch in terms of D_l will be directly used to recover the high-resolution patch from D_h. Directly, we do not compute the sparse representation of the high-resolution patch. Here we obtain a uniform solution by allowing patches to overlap and demanding that the reconstructed high-resolution patches agree on the overlapped areas. In this paper, we try to learn two over-complete dictionaries in a probabilistic model similar to [17]. To learn effectiveness of the compact dictionaries that is applied to the generic images for super-resolution, image patch pairs must have same sparse representation with respect to D_h and D_l, and combining them simultaneously with proper normalization.

In case of online recovery, the sparse representation uses the only low-resolution dictionary D_l and high-resolution dictionary D_h is used to estimate the output high-resolution image. The sparse representation is robust to noise as suggested in [14]. Our algorithm is more robust to noise in the test image and it also performs denoising and super-resolution simultaneously while most other methods cannot perform.

a) Organization of the Paper: This paper is organized as follows: in Section 2, we discuss review of the super-resolution via sparse representation. In Section 3, we discuss how to learn joint dictionaries for the high- and low-resolution image patches respectively. In Section 4, we demonstrate experimental results for regularizing the image super-resolution problem.

b) Notations: High-resolution images are denoted by X and Low-resolution images are denoted by Y, and x and y denote the high-resolution and low-resolution image patches respectively. D denotes the dictionary for sparse coding. High-resolution image patches dictionary and low-resolution image patches dictionary are denoted by D_h and D_l respectively.

2 SUPER-RESOLUTION VIA SPARSE REPRESENTATION [12]

"Single image super-resolution" is termed as the generating high-resolution image X from single low-resolution image Y of the same scene. In order to solve this ill-posed problem, two constraints are assumed: (1) A reconstruction constraint: which requires that recovered image X should be compatible with the input image Y as per the principle of the image observation model; and (2) A sparsity prior: which assumes that the sparse representation of image X can be reconstructed from low-resolution version and it can be sparsely represented with respect to over complete dictionary.

Considering a signal $x \in R^n$ can be referred as sparse linear combination over the over-complete dictionary D of K atoms, where $\boldsymbol{x}$ be a high-resolution image (patch).To be more exact, let $D \in R^{nxK}$, where D is an over-complete dictionary of K atoms ($K > n$) which is large matrix learned using high-resolution image patches. In practice, relation between high-resolution patch $\boldsymbol{x}$ and low-resolution patch y which is downsampled and blurred version of x can be represented as:

$$y = SBx = Lx, \tag{1}$$

where $\boldsymbol{y}$ is a downsampled and blurred version of $\boldsymbol{x}$, S is a downsampling operator, B is a unknown blurring operator and $L \in R^{kxn}$ which denote their combined effect with $k < n$. Signal $\boldsymbol{x}$ can be written as $x = D\alpha_0$, where $\alpha_0 \in R^K$. Eq.(1) written as :

$$y = LD\alpha_0, \tag{2}$$

Eq. (2) indicates is that low-resolution patch y will have the same sparsely representation coefficients α_0.The equation $x = D\alpha_0$ is not progressed if the dictionary D is over-completed for the unknown coefficients α. After getting the sparse representation, high-resolution patch x can be reconstructed by:

$$x^{\wedge} \approx D\alpha_0 \tag{3}$$

The joint dictionary training process will be discussed in Section 3. The generic images are used for training.

3 JOINT DICTIONARY LEARNING AND PROPOSED ALGORITHM

In this section, we discuss to process of the joint dictionary training. Initially we train two dictionary which having same sparse representation. Our goal is to learn high- and low-resolution dictionaries. After training these two dictionaries, both dictionaries are combining and then learned to ensure same sparse representation.

Fig. 1. shows the flow of joint dictionary training diagram using multiple generic images. The input for the training is the collection of high-resolution images $X^h = \{x_1, x_2, ..., x_n\}$ and corresponding low-resolution images $Y^l = \{y_1, y_2, ..., y_n\}$. Let $P = \{ X_h, Y_l \}$ represents the sampled trained image patch pair. Because of ill-posed problem, it is impossible to achieve the goal. To find the sparse representation over the over-complete dictionary, specially sparse coding is used to speed up the computation. The problem is solved in two steps:

(1) Separate sparse coding for high- and low-resolution patches are

$$D_h = \arg\min_{\{D_h,Z\}} \left\| X^h - D_h Z \right\|_2^2 + \lambda \left\| Z \right\|_1, \quad (4)$$

and

$$D_l = \arg\min_{\{D_l,Z\}} \left\| Y^l - D_l Z \right\|_2^2 + \lambda \left\| Z \right\|_1, \quad (5)$$

respectively, and

(2) By combined eqs. (4) and (5), a common sparse coding is generated which is represented mathematically represented as follows:

$$\{D_h, D_l\} = \min_{\{D_h,D_l,Z\}} \left(\frac{1}{I}\right) \left\| X^h - D_h Z \right\|_2^2 + \left(\frac{1}{J}\right) \left\| Y^l - D_l Z \right\|_2^2 + \lambda \left(\frac{1}{I} + \frac{1}{J}\right) \left\| Z \right\|_1, \quad (6)$$

Here, I and J are the dimensions of high-resolution and their corresponding low-resolution image patches in the form of vector L_1 and L_2 norm is used to make sure that solution is sparsest and to avoid uncertainty due to scaling respectively.

ALGORITHM

1: Apply a low-resolution image Y. Call both high-resolution dictionary D_h and low-resolution dictionary D_l.

2: Divide the input low-resolution image Y into 5 x 5 patches. For each 5 × 5 patch y of Y, overlap 4 pixels starting from the top-left corner in each direction, and calculate the mean pixel value 'm'.

3: Optimization problem is solved by using following eq :

$$\min_{\alpha} \left\| \tilde{D}\alpha - \tilde{y} \right\|_2^2 + \lambda \left\| Z \right\|_1, \quad (7)$$

where λ is sparsity regularization parameter.

4: From that, high-resolution patch $x = D_h \alpha^*$ is obtained and then keep the patch $x + m$ into a high-resolution image X_0.

Here $\alpha^* = \arg\max P(\alpha).P\left(\tilde{y} \middle| \alpha, \tilde{D}\right)$ (8)

where

$$P(\alpha) = \frac{1}{2b} \exp\left(- \frac{\left\| \alpha \right\|_1}{b} \right)$$

$$P\left(\tilde{y} \middle| \alpha, \tilde{D}\right) = \frac{1}{2\sigma^2} \exp\left(- \frac{1}{2\sigma^2} \left\| \tilde{D}\alpha - \tilde{y} \right\|_2^2 \right) \quad (9)$$

where b is the variance of the laplacian prior on α, and σ^2 is the variance of the noise assumed on the data $\tilde{y}$.

5: But X_0 is not satisfied reconstruction constraint. So we use gradient descent. Then we search nearest image to X_0 which satisfies the reconstruction constraint:

$$X^* = \arg\min_x \left\| SHX - Y \right\|_2^2 + c \left\| X - X_0 \right\|_2^2 \quad (10)$$

where X^* is estimation of high-resolution image *i.e* super-resolution image.

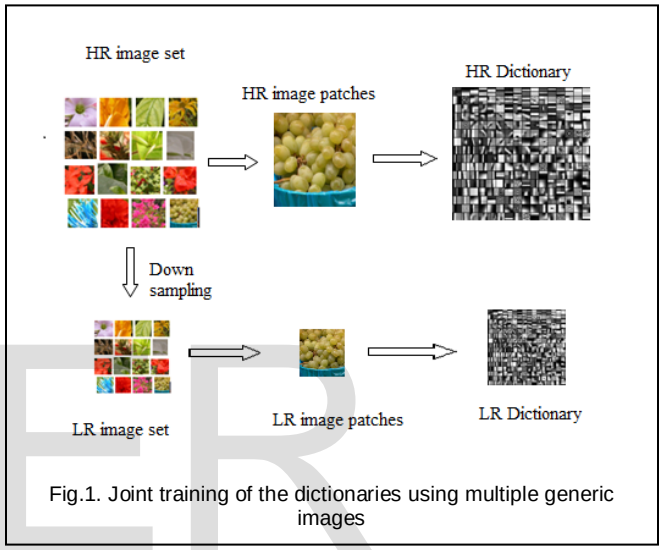


Fig.1. Joint training of the dictionaries using multiple generic images

4 EXPERIMENTAL RESULTS

In this paper, the input-low resolution image is upscale by a factor of 2 for generic images. After recursive experiment, it is observed that 5x5 is an optimum patch size. In this paper, we use patches of size 5x5 low-resolution. However, between adjacent patches, it can be overlap by pixel of 4, and corresponding for high-resolution patches of size 10×10 can be overlap by pixel of 8. Here human eyes are easier to detect changes in color images. Therefore, we work only on illuminance channel only.

Firstly, we trained two dictionaries for high-resolution and low-resolution image patches randomly sampled at the rate of 100,000 high-resolution / low-resolution patch pairs from each set of training images. Dictionary size 1024 is fixed to balance between image quality and computation. The λ = 0.15 is used throughout the experiment which produce superior results.

Performances of the different methods are assessed both visually and qualitatively by Peak signal to noise ratio (PSNR), Structural similarity (SSIM) and image quality index (IQI).

System configuration: Intel(R) Core(TM) i5-3470 CPU@3.20GHz machine with 4GB RAM is used for simulation.

TABLE 1
PSNR PERFORMANCE FOR DIFFERENT METHODS

Upscale	PSNR(dB)		
	Zoom Image	Bicubic Interpolation	Sparse Recovery
2	30.98	32.79	35.04
4	30.49	32.92	33.43
6	30.06	32.35	32.49

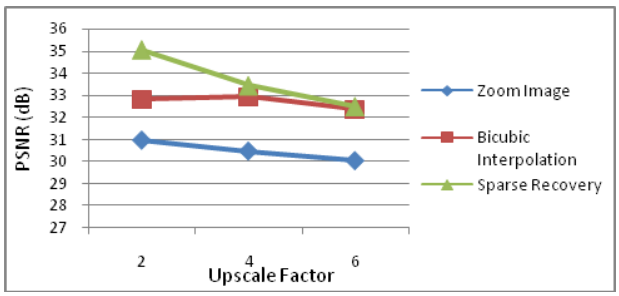


Fig.2. Zoom image, Bicubic interpolation, Sparse recovery for Lena image.

Table 1 and Fig.2. shows the comparison of different methods for different upscaling for an image Lena. From table 1, it is clear that the sparse recovery method is more superior than any other methods.

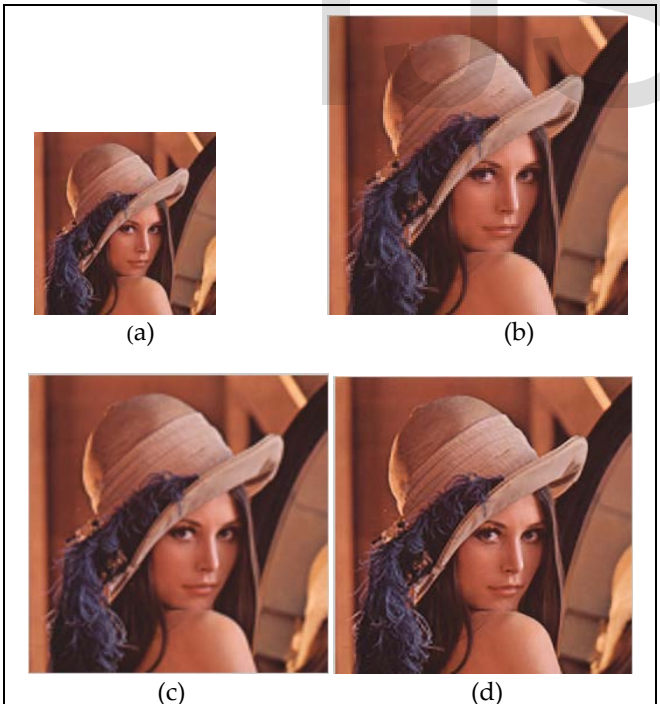


Fig. 3. Results of the Lena with magnification factor two of 5x5 patch sizes. Left to Right: (a) low-resolution input, (b) zoom image (PSNR = 30.98dB), (c) bicubic interpolation (PSNR = 32.79dB), (d) Sparse Recovery (PSNR = 35.04dB).

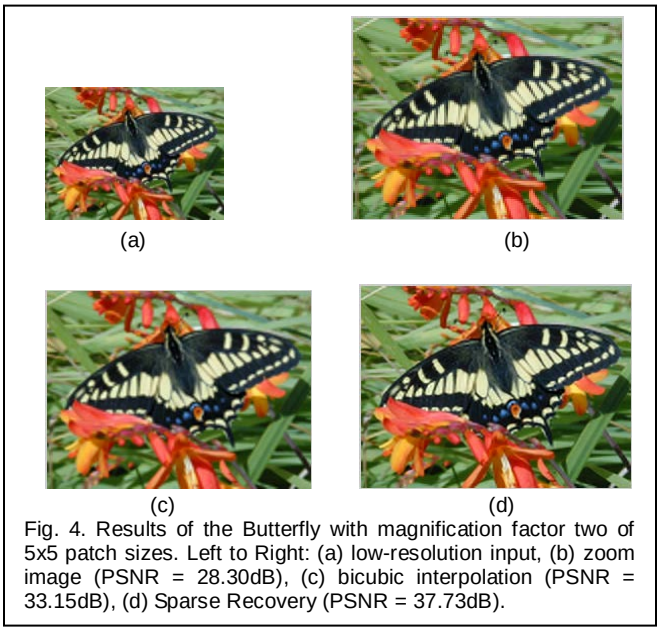


Fig. 4. Results of the Butterfly with magnification factor two of 5x5 patch sizes. Left to Right: (a) low-resolution input, (b) zoom image (PSNR = 28.30dB), (c) bicubic interpolation (PSNR = 33.15dB), (d) Sparse Recovery (PSNR = 37.73dB).

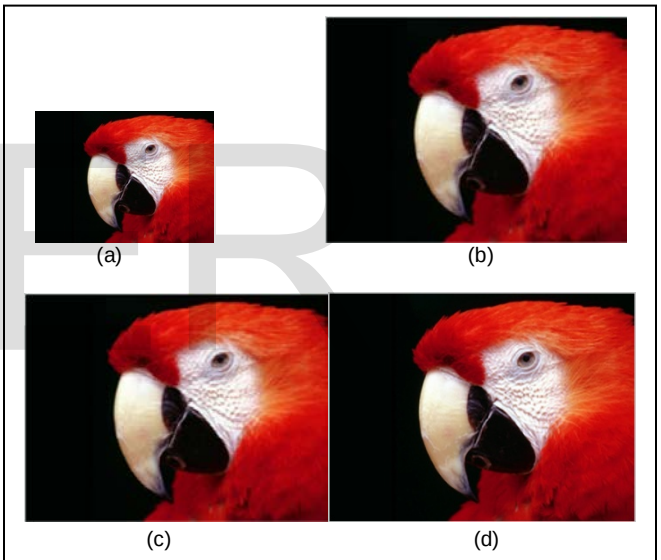


Fig. 5. Results of the Parrot with magnification factor two of 5x5 patch sizes. Left to Right: (a) low-resolution input, (b) zoom image (PSNR = 32.94dB), (c) bicubic interpolation (PSNR = 334.92dB), (d) Sparse Recovery (PSNR = 36.15dB).

Fig. 3, 4, and 5, it is clear that the proposed method gives much superior results than any other methods. When the Lena image is zoomed by a factor 2, it is get blurred which reduces the quality of the image. When the same image is reconstructed by bicubic interpolation and sparse recovery, the sparse recovery shows the superior results as shown in table 1.

Table 2, 3 and 4, shows more experimental results of PSNR, SSIM and IQI values for different patch sizes withdifferent test images.

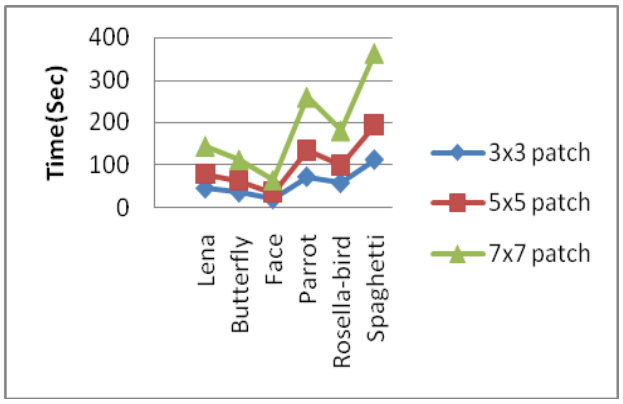


Fig.6. Time Estimation for different patche sizes

Fig.6, shows time estimation for 3x3, 5x5 and 7x7 patch sizes. From this figure, we can conclude that the time required for bigger patch size is more. But from the experimental results, it is clear that the best reconstructed image with moderate time can be achieved by 5x5 patch size. This is justified by Fig.7.

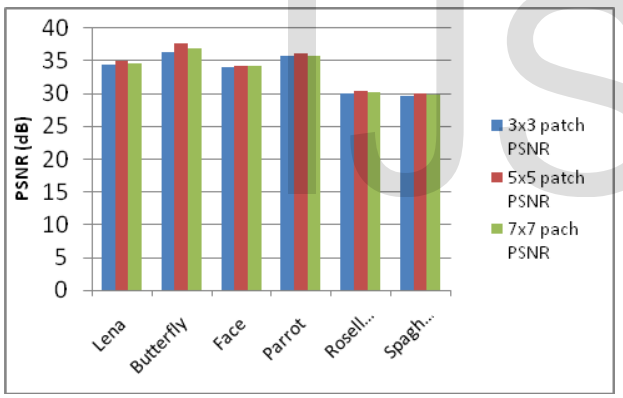


Fig.7. Comparision for PSNR for different patch sizes.

Fig.8, shows different test images used in our experiments.

5 CONCLUSION

This paper, presents a simple and novel framework for single image super-resolution approach. This approach is based on sparse representations and jointly trained dictionary. Experimental results are better and demonstrated the effectiveness of the proposed algorithm. As compared to zoom and bicubic interpolation method, the proposed approach has been improved a PSNR of 4.06 dB and 2.25 dB respectively for Lena image, 9.43 dB and 4.58 dB respectively for butterfly image and similarly 3.21 dB and 1.23dB respectively for parrot image. Ideally optimized patch size is 5x5. Table 3 show 5x5 patches is often required efficient time and give superior results in qualitatively and quantitatively. Hence, selection of the patch size is very important factor in this approach. Experimental results indicated effectiveness of proposed algorithm.

REFERENCES

[1] M. E. Tipping and C. M. Bishop, "Bayesian image super-resolution," in *Advances in Neural Information and Processing Systems 16 (NIPS)*, 2003.

[2] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, Sep. 2002.

[3] M. Elad and A. Feuer. "Super-resolution reconstruction of image sequences". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):817–834, 1999.

[4] R. C. Hardie, K. J. Barnard, and E. A. Armstrong, "Joint map registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, vol. 6, pp. 1621– 1633, 1997.

[5] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super-resolution," *IEEE Transactions on Image Processing*, vol. 13, pp. 1327–1344, 2004.

[6] H. S. Hou and H. C. Andrews, "Cubic spline for image interpolation and digital filtering," *IEEE Transactions on Signal Processing*, vol. 26, pp. 508–517, 1978.

[7] S. Dai, M. Han,W. Xu, Y.Wu, and Y. Gong, "Soft edge smoothness prior for alpha channel super resolution," in *IEEE Conference on Computer Vision and Pattern Classification (CVPR)*, 2007, pp. 1–8.

[8] J. Sun, Z. Xu, and H. Shum, "Image super-resolution using gradient profile prior," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.

[9] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low level vision," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 25–47, 2000.

[10] J. Sun, N. N. Zheng, H. Tao, and H. Shum, "Image hallucination with primal sketch priors," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2003, pp. 729–736.

[11] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *IEEE Conference on Computer Vision and Pattern Classification (CVPR)*, vol. 1, 2004, pp. 275–282.

[12] Yang, J., Wright, J., Huang, T. and Ma, Y., "Image superresolution via sparse representation," *IEEE Trans. on Image Processing*, vol. 19, no. 11, pp. 2861-2873, Nov. 2010.

[13] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.

[14] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, pp. 3736–3745, 2006.

[15] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Modeling and Simulation*, vol. 7, pp. 214–241, 2008.

[16] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transaction on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[17] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Advances in Neural Information Processing Systems (NIPS)*, 2007.

TABLE 2
PSNR (dB), SSIM AND IQI RESULTS OF SIX TEST IMAGES FOR 3X3 PATCH SIZE

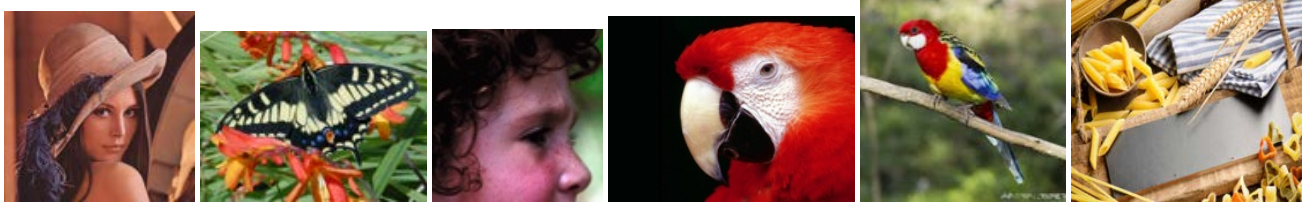
Test images		Zoom method			Bicubic Interpolation			Sparse recovery (Proposed method)			Time(sec)
		PSNR	SSIM	IQI	PSNR	SSIM	IQI	PSNR	SSIM	IQI	
Test 1	Lena	30.98	0.89	0.79	32.79	0.91	0.82	34.44	0.93	0.85	45.522004
Test 2	Butterfly	28.3	0.92	0.93	33.15	0.97	0.97	36.43	0.98	0.98	36.085286
Test 3	Face	32.22	0.84	0.75	32.96	0.85	0.76	34.09	0.88	0.8	20.305224
Test 4	Parrot	32.94	0.92	0.77	34.92	0.94	0.8	35.77	0.95	0.81	72.243137
Test 5	Rosella-bird	28.55	0.9	0.86	29.58	0.92	0.89	30.14	0.93	0.89	57.638561
Test 6	Spaghetti	26.87	0.98	0.81	28.95	0.98	0.84	29.66	0.98	0.87	112.195074

TABLE 3
PSNR (dB), SSIM AND IQI RESULTS OF SIX TEST IMAGES FOR 5X5 PATCH SIZE

Test images		Zoom method			Bicubic Interpolation			Sparse recovery (Proposed method)			Time(sec)
		PSNR	SSIM	IQI	PSNR	SSIM	IQI	PSNR	SSIM	IQI	
Test 1	Lena	30.98	0.89	0.79	32.79	0.91	0.82	35.04	0.94	0.86	77.772317
Test 2	Butterfly	28.3	0.92	0.93	33.15	0.97	0.97	37.73	0.99	0.99	62.379799
Test 3	Face	32.22	0.84	0.75	32.96	0.85	0.76	34.35	0.89	0.81	34.868287
Test 4	Parrot	32.94	0.92	0.77	34.92	0.94	0.8	36.15	0.95	0.82	133.885506
Test 5	Rosella-bird	28.55	0.9	0.86	29.58	0.92	0.89	30.35	0.93	0.9	99.385789
Test 6	Spaghetti	26.87	0.98	0.81	28.95	0.98	0.84	30.14	0.98	0.88	194.17036

TABLE 4
PSNR (dB), SSIM AND IQI RESULTS OF SIX TEST IMAGES FOR 7X7 PATCH SIZE

Test images		Zoom method			Bicubic Interpolation			Sparse recovery (Proposed method)			Time(sec)
		PSNR	SSIM	IQI	PSNR	SSIM	IQI	PSNR	SSIM	IQI	
Test 1	Lena	30.98	0.89	0.79	32.79	0.91	0.82	34.7	0.94	0.86	145.679477
Test 2	Butterfly	28.3	0.92	0.93	33.15	0.97	0.97	36.9	0.99	0.98	113.824185
Test 3	Face	32.22	0.84	0.75	32.96	0.85	0.76	34.25	0.88	0.8	66.048155
Test 4	Parrot	32.94	0.92	0.77	34.92	0.94	0.8	35.87	0.95	0.81	260.917562
Test 5	Rosella-bird	28.55	0.9	0.86	29.58	0.92	0.89	30.23	0.93	0.9	182.017596
Test 6	Spaghetti	26.87	0.98	0.81	28.95	0.98	0.84	29.78	0.98	0.87	364.127184



(a) Test 1 (b) Test 2 (c) Test 3 (d) Test 4 (e) Test 5 (f) Test 6

Fig.8. Different Test Images used for demonstration.