Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 46 (2015) 1643 – 1652
Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

Single Image Super Resolution from Compressive Samples using Two Level Sparsity based Reconstruction

Aneesh G. Nath[a], Madhu S. Nair[b,]*, Jeny Rajan[c]

[a]*Department of Computer Science & Engg.,T K M College of Engg., Karicode, Kollam-691005, Kerala, India.*
[b]*Department of Computer Science, University of Kerala, Kariavattom, Thiruvananthapuram-695581, Kerala, India.*
[c]*Department of Computer Science & Engg., National Institute of Technology Karnataka, Surathkal, Mangalore-575025, Karnataka, India.*

Abstract

Super Resolution based on Compressed Sensing (CS) considers low resolution (LR) image patch as the compressive samples of its high resolution (HR) patch. Compressed sensing based image acquisition systems acquire less number of random linear measurements without first collecting all the pixel values. But using these compressive measurements directly to reconstruct the image causes quality issues. In this paper an image super-resolution method with two level sparsity based reconstruction via patch based image interpolation and dictionary learning is proposed. The first level reconstruction generates a low resolution image from random samples and the interpolation scheme used in this algorithm reduces the HR-LR patch coherency due to neighborhood issue which is a major drawback of single image super resolution algorithms. The dictionary based reconstruction phase generates the high resolution image from the low resolution output of the first level reconstruction phase. The experimental results proved that the proposed two level reconstruction scheme recovers more details of the image and yields improved results from very few samples (around 35-45%) than the state-of-the-art algorithms which uses low resolution image itself as input. The results are compared by considering both PSNR values and visual perception.

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: Super-resolution; Compressed Sensing; Sparsity; Dictionary learning; BM3D.

* Corresponding author. Tel.: +91-9447364158.
E-mail address: madhu_s_nair2001@yahoo.com

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.02.100

1. Introduction

The process of creating high resolution (HR) image from one or more low resolution (LR) images is known as Super-resolution (SR). In traditional methods of SR the low resolution images captured from the same scene are used to solve the lost high-frequency information. But it is more challenging if we have only one low-resolution image, and this information has to be recovered from it. In most of the applications only single low-resolution image is available and the Single image super-resolution (SISR) problem is particularly important in those situations. Compressed sensing[4] (CS) is an emerging data acquisition technique which overcomes the Shannon's sampling theorem's limitations. The motivating fact behind CS is the sparsity property of natural signals in basis like fourier, wavelet etc. In many emerging applications, the abundance of data generated by the sensing systems due to high sampling rate demands data compression before storage or transmission. Compressed Sensing combines the sampling and the compression into a single process. CS data acquisition technique enables the reduction in the number of measurements required for recovery of sparse signals or compressible signals which is sparse on some suitable basis. Reconstruction of the signals from CS measurement is done using greedy or relaxation based algorithms. In a CS based image acquisition system it acquires less number of random linear measurements (pixels at a subset of sampling lattice) without first collecting all the pixel values. Proper preprocessing technique will enable the reconstruction of image from this incomplete data. The resulting image is an LR image which is suitable for reconstruction of its original HR image. In single image super resolution using compressive sensing, a low resolution input image plays the role of the samples from compressive sensing data acquisition device and a proper dictionary which represents the high resolution image sufficiently sparse will accurately recover the high resolution image. We can summarize the problem of SISR as generating HR image X, given the degraded Low-Resolution (LR) version Y of X, which is represented as

$$Y = HBX + v \tag{1}$$

where H is the downsampling operator, B is the blurring operator and v represents the additive noise.

Based on this recovery problem, we are interested in creating X from the random samples and avoiding the coherency issues[8] in HR-LR patch pairs by proper preprocessing technique integrated to the overall super resolution scheme. The proposed system acquires only fewer numbers of samples as input instead of creating input LR image from original HR image by blurring and down sampling operations. The main motive behind the development of such a system is to provide a better quality output from the samples provided by a compressed sensing image acquisition system.

This paper is organized as follows. The related works reported in the literature is given in section2. Section 3 describes about the proposed scheme of super-resolution which includes the dictionary training and reconstruction phases with the preprocessing step. Experimental results are shown in Section 4 and conclusions are drawn in Section 5.

2. Literature Survey

The work of Tsai et al.[1] in 1984 made SR reconstruction as one of the most active research areas ever since. The main technique proposed in this area varies from frequency domain approach of Borman et al.[2] to spatial domain approach of Sung et al.[3]. Conventional approach in super resolution is to generate a SR image from a number of low-resolution images using some constraints (e.g. Bilateral Total Variation[13] and Huber Markov Random Fields[12] (MRF)), with the help of maximum-a-posteriori (MAP) like regularization technique.

Single Image SR (SISR) can be classified as follows:

1) Reconstruction-based super-resolution[11, 14, 16] without training process but with the help of well defined constraints for the target high-resolution image.

2) Learning-based super-resolution algorithms[18-21] makes use of a dictionary which is trained and tested using training set of images.

Interpolation based algorithms reconstruct the image details by interpolating the LR input image and enhance the edges by making it sharper. The standard methods in this category are bilinear and bicubic interpolation. Another well-known algorithm is Back Projection[17]. This method gradually sharpens the edges while it iterates through the image. The Dai et al.'s approach[16] is also an example of above category which enforces the continuity by extracting and blending them with the results of interpolation. Fattal's[14] work is based on high frequency information reconstruction from edge statistics. A training set of HR image patches and its LR image patches are collected into a database and using this database to recover the lost image information is called example-based super-resolution method[19]. But the main drawback of example-based algorithms is that the training set and test set similarity is the main criterion of performance. The sparse representation method based on CS[4] theory exploits the high-dimension signals' linear relationship and uses it to recover HR information from their low-dimension projections. Yang et al.[5, 6] proposed the idea of sparsity which works with sparse representation of patch pairs sampled from HR and LR images. In the method[5] trained over complete dictionaries are used for representing them as sparse vectors to get the co-occurrence in order to increase the speed and accuracy. Yang et al.[6] proposed a method to make the representations of these patch pairs more compact. This method achieved significant performance improvement. Later, Zeyde et al.[7] modified this approach to make it more efficient and faster. Zhang et al.'s[15] method of training two dictionaries called main dictionary learning and the residual dictionary learning with sparse representation, recovers main high-frequency (MHF) and residual high-frequency (RHF) information and thereby achieves better results.

3. Proposed two level reconstruction scheme

As any other learning based super resolution methods, our approach also consists of dictionary learning phase and image synthesis phase. Though many state-of-the-art dictionary learning methods are available in the literature, we adopt the method of Zeyde et al.[7], which is proved to be faster and efficient.

3.1. Dictionary learning phase

This stage starts with collecting high resolution image set H_{ORG} from the training data. Since we assume our algorithm to work with a set of random samples from a CS based acquisition system, we are interested in taking only few samples from these images. For that purpose the image is multiplied by a randomly generated binary sampling pattern with most of zero values (55-75% of zeros in ideal case). This leads to the formation of the samples which is same as one that provided by the CS acquisition device, and is denoted by H_{SAMP}. These samples or CS observations are given as the input to the first level reconstruction step called non-local patch based pre-processing.

Non-local Patch based pre-processing:

In order to create an image suitable for training dictionaries from incomplete data, this framework proposes an algorithm incorporating patch-based nonlocal prior into image interpolation. The likelihood term is given by observation constraint and the prior term given by sparsity constraint. In this method the sparsity constraint is given non-locally by Block matching 3D[9] (BM3D) based method which in turn helps to avoid neighborhood issues. Here x denotes unobservable image and y denotes observation data. We take $y(m,n) = x(m,n)\ S(m,n)$, where S is a binary sampling pattern and (m, n) are the spatial coordinates. Since sparsity is a prior term for this ill-posed problem's regularization, only a few number of transform coefficients are taken as significant. The main challenge in this method is to construct a good transform T which sparsify the image maximum. The thresholding operator δ gives the sparsity constraint. The transform coefficients are represented by $X = Tx$. The similar patches are packed to use sparsity constraint. In this experiment, on a block-by-block basis all local variances are calculated and its average is set to be the threshold value. To maintain quality constraint the maximum error value between current image and previous image, ε, is set to 0.001. From the quantized samples using the location information of observation data and intensity parameter λ, the images are reconstructed. The following algorithm does this pre-processing. The algorithm gives a low resolution image suitable for training the dictionary.

Algorithm 1: Nonlocal Patch-based Image Reconstruction
Input: sampling pattern S and observation data y
Output: reconstructed image $\hat{x}$

1. Let $\hat{x}^{(0)}$ be the locally interpolated image.
2. Calculate average (δ) of all local variance block by block basis from $\hat{x}^{(0)}$.
   Set $k = 0$.
3. Extract blocks of size $N_1 \times N_1$ from $\hat{x}^{(k)}$ and find similar blocks.
4. Stack the similar blocks together to form a 3D array.
5. Apply 3D decorrelating unitary transforms to these arrays to find sparse representation.
6. Filter the transform coefficients by threshold operator δ_k.
7. Reconstruct the patches from filtered coefficients by inverse transform.
8. Construct the image $\hat{x}^{(k)}$ from the patches.
9. Project $\hat{x}^{(k)}$ onto set of observation constraint:
   For all (m,n) where $S(m,n)=1$
   Set $\hat{x}^{(k)}(m,n) = y(m,n)$.
10. If $\left\|\hat{x}^{(k)} - \hat{x}^{(k-1)}\right\| < \varepsilon$ then stop and return $\hat{x}$.
    else $k=k+1$ and set $\delta_k = \delta - (k-1)\Delta$ and repeat the process from step 3.

Dictionary Training:

The set of images from the preprocessing step serves the input to the dictionary training step. To incorporate with the steps followed by Zeyde et al.[7], we consider the upscaled LR image denoted by L_{ORG} as the degraded version of the high resolution image H_{ORG}. The degraded image L_{ORG} is supposed to be created by down sampling and blurring operations. Then high frequency features are extracted by subtracting L_{ORG} from H_{ORG}. As a next step, local patches are extracted from H_{ORG} and L_{ORG}, considering only locations $k \in \Omega$ to generate the data-set $P = \{p_l^k, p_h^k\}_k$ for dictionary training. LR image L_{ORG} is filtered by using R number of high pass filters such as Laplacian to extract local features of their high-frequency content. In order to save the computations in training, the dimensionality of the input LR patches is reduced by Principal Component Analysis (PCA) algorithm as the last step before dictionary learning stage. Next, the K-SVD[10] dictionary training algorithm is used, giving the set of patches as input.

Low Resolution Dictionary Training:

The dictionary training stage starts with the low-resolution patches $\{\hat{p}_l^k\}_k$ extracted from low resolution image. As the result of applying K-SVD dictionary learning algorithms to these patches, the dictionary $L_D \in R^{n\times m}$ is created.

$$L_D, \{q^k\} = \arg\min_{L_D, \{q^k\}} \left\| p_l^k - L_D \{q^k\} \right\|^2 \quad \text{s.t} \quad \left\| q^k \right\|_0 \le L \quad \forall k. \tag{2}$$

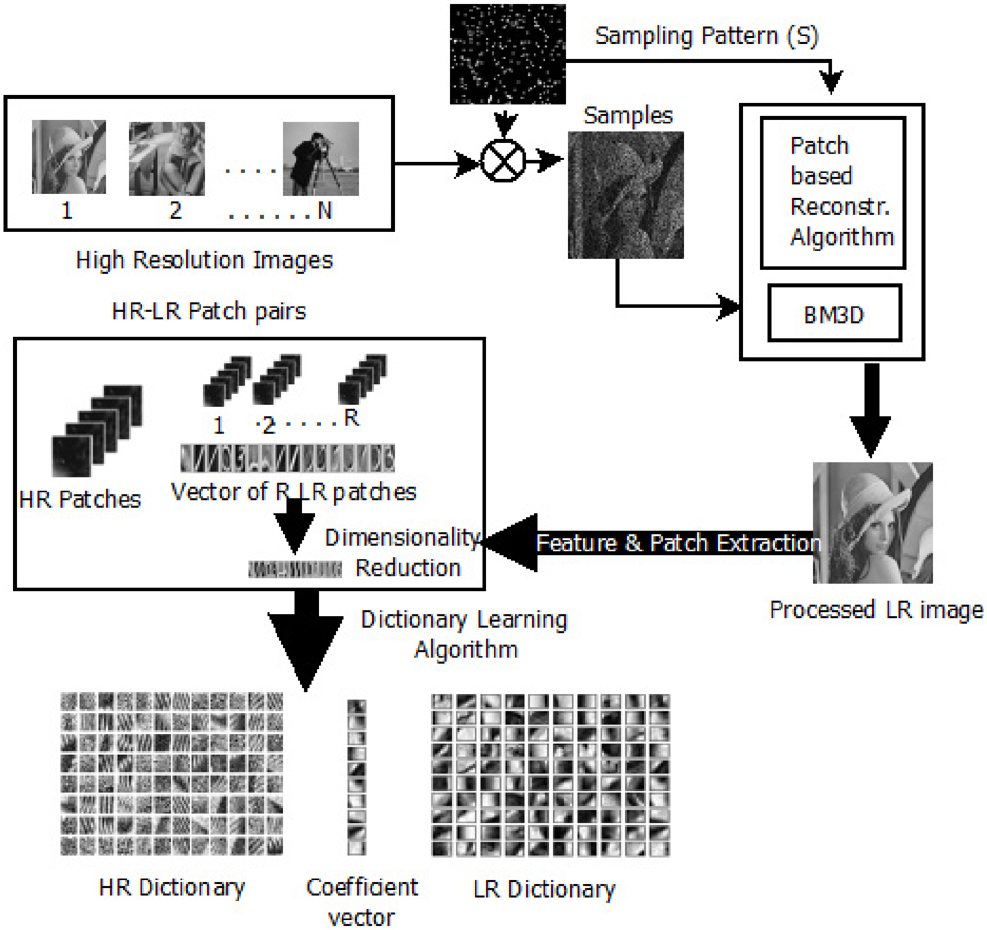

Fig. 1. Illustration of dictionary learning phase.

This training process also generates the sparse representation coefficient vectors q^k corresponding to the training patches $\{\hat{p}_l^k\}_k$ as a side product. Here $\|\cdot\|_0$ is the l_0 norm which gives the count of nonzero values in the vector.

High Resolution Dictionary Learning:

By approximating $P_h^k \approx H_D q^k$, the HR patch P_h^k can be recovered. For this the already generated sparse representation vector of LR patch q^k is multiplied with the high-resolution dictionary H_D. The high resolution dictionary H_D can be found to get the correct approximation. So, H_D is the dictionary matrix which minimizes the approximation error.

$$H_D = \arg\min_{H_D} \sum_k \left\| p_h^k - H_D \left\{ q^k \right\} \right\|_2^2 \qquad \text{i.e.} \quad H_D = \arg\min_{H_D} \left\| \mathrm{P}_h - H_D \mathrm{Q} \right\|_F^2 \tag{3}$$

The following Pseudo-Inverse expression is used for solving this problem.

$$H_D = \mathrm{P}_h \mathrm{Q}^+ = \mathrm{P}_h \mathrm{Q}^{\mathrm{T}} (\mathrm{Q}\mathrm{Q}^{\mathrm{T}})^{-1} \tag{4}$$

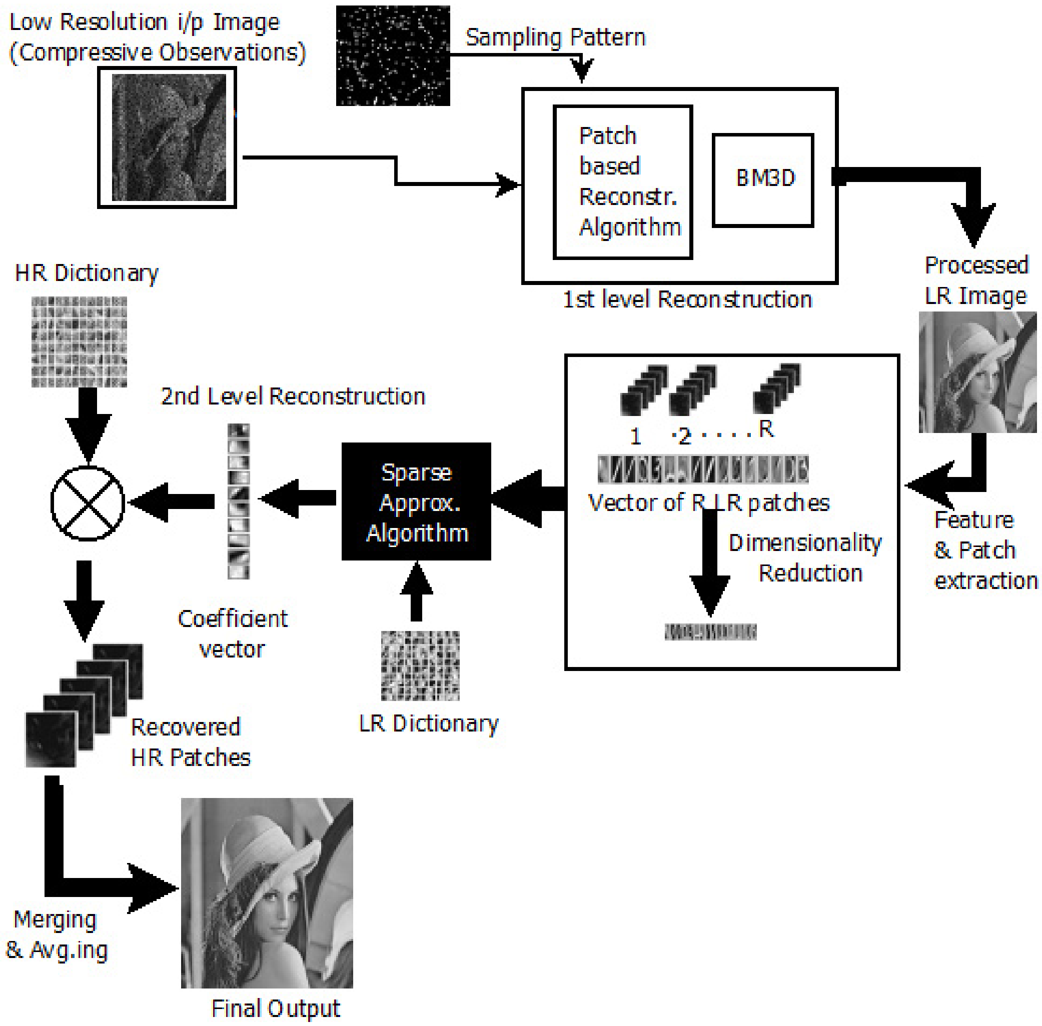

Fig. 2. Illustration of image synthesis phase.

3.2. Image synthesis phase

The Image Synthesis Stage generates high resolution image from the random samples provided by the CS acquisition system. The input provided to this phase are the image samples which are the same as the one that provided by the CS acquisition device denoted by H_{SAMP}. These samples or CS observations are given as the input to the first level reconstruction step which generates the low resolution image from the CS observations. We use the same nonlocal image interpolation algorithm and upscaling process to create the upscaled version of the image denoted by L_{ORG}. This kind of preprocessing helps to avoid the neighborhood issue[8] and create an image with same size of the required high resolution image. This pre-processed image is the input for the next level reconstruction. The second level reconstruction uses the trained dictionary to enhance the low-resolution image L_{ORG}. We assume that this image is generated by same degradation operations (blur and scale-down) from a high-resolution image

H_{ORG} as done in the training phase. The same R numbers of high-pass filters that we have applied in the training phase are used for filtering the image for extracting the features. Patches are extracted from these R images from locations $k \in \Omega$. After the PCA based dimensionality reduction operation, the OMP algorithm is applied to $\{\hat{p}_l^k\}_k$ patches with the number of atoms for sparse representation as L. The output of OMP algorithm is sparse representation vectors $\{q^k\}_k$. These vectors are multiplied by the high-resolution dictionary H_D to generate the approximated HR patches, $\{H_D q^k\}_k = \{\hat{p}_h^k\}_k$. The solution to the minimization problem given below gives the final image $\hat{H}_{ORG}$ from $\hat{p}_h^k$.

$$\hat{y}_h = \arg\min_{\hat{y}_h} \sum_k \left\| R_k(\hat{y}_h - y_l) - \hat{p}_h^k \right\|_2^2 \tag{5}$$

where R_k is the patch extraction operator for extracting $n \times n$ sized patches at location k from the image. This problem attempts to make the patches extracted from difference image $\hat{y}_h - y_l$ as close as possible to the approximated patches $\{\hat{p}_h^k\}_k$. The Least-Squares solution to this problem is given by

$$\hat{H}_{ORG} = \hat{y}_h = y_l + \left[\sum_k R_k^T R_k\right]^{-1} \sum_k R_k^T \hat{p}_h^k \tag{6}$$

4. Experimental results

The proposed algorithm is implemented in MATLAB R2012a using K-SVD for learning dictionary and OMP algorithm for sparse approximation. The computer system used for simulation is Intel Core i5 – 2410M CPU at 2.30GHz with 4GB of RAM. In this framework, only very few percentage of the samples are collected from the image. The image is downscaled with a scaling factor 2 before taking the samples, in order to show the upscaling performance. The parameters in patch based non-local image interpolation are fixed in all experiments as $\Delta = 0.02, \varepsilon = 0.001$. The initial interpolation is done by Delaunay triangulation based interpolation. Adding more and more images for training would lead to improved results. The feature extraction from the low resolution image is performed with four filters: $f_1 = [1,-1] = f_2^T$ and $f_3 = [1,-2, 1] = f_4^T$. The size of each patch is set to $n = 81$ (9 × 9), and after applying PCA, the dimensionality has been reduced from 324 (4×81) dimensions to 30 dimensions. In dictionary training, number of iterations in K-SVD algorithm is set to 40, with number of atoms in dictionary as m = 1000. The value of L which defines the sparsity is given as 3.

For evaluating the performance of the proposed system with varying amount of samples, we took samples from 1% to 100% at an interval of 5 for standard 4 images of varying sizes. Analyzing the PSNR comparison of the results which is shown as graph in figure 3, it is clear that around 35% to 45% of samples are sufficient enough to produce a reasonable output for all the images. It proves that this method is able to produce high quality image from fewer amount of random samples while most of the existing super resolution algorithms fail to handle the random samples directly.

The results shown in figure 4 are the visual comparison of the proposed method with Bicubic interpolation and Yang et al.[6] methods. Table 1 shows the PSNR comparison of the same. The values given along with the PSNR of proposed method are the percentage of samples with which the proposed method produce better output than the other two methods. It is to be noticed that bicubic and Yang et. al methods work on full samples of low resolution images, mean time the proposed method takes only very less random samples and still produce comparable results. In the case of Boat image, only with 30% of samples, the proposed method produces PSNR value of 25.4122 which is greater than the PSNR value obtained by using Bicubic and Yang et al. methods. Similarly, only 35% of samples used in the reconstruction process of the proposed method are enough to outperform the other two methods in the case of Barbara image. As stated in previous results, the PSNR values of results produced from 35% to 45% of samples by our method are greater than that of both these methods. So in terms of visual perception and PSNR values the proposed algorithm performs much better than the state-of-the-art methods.

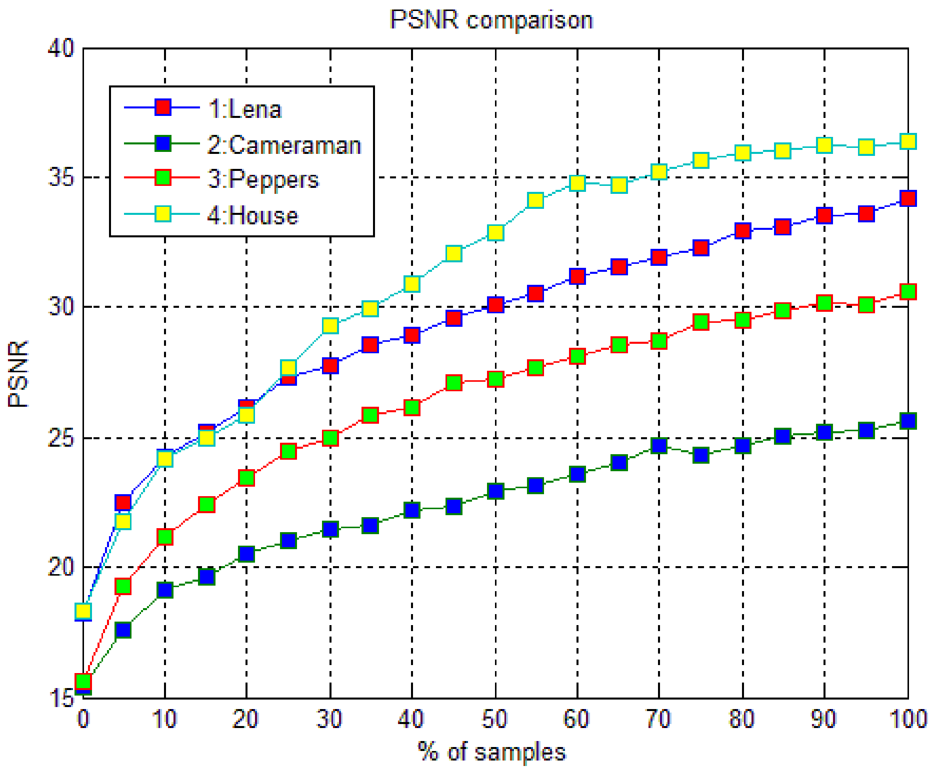

Fig. 3. PSNR comparison with different sample size.

Table 1. PSNR (in dB) comparison

Image	Bicubic	Yang et al.'s method	Proposed method	% of samples taken in the proposed method
Boat	24.7095	25.2444	25.4122	30%
House	30.9195	32.7651	32.9941	45%
Lena	27.9286	28.3229	28.5850	35%
Peppers	25.1243	26.9227	27.0790	45%
Man	25.8304	26.4995	26.6092	40%
Barbara	21.8747	22.2540	22.9364	35%
Fingerprint	22.0286	23.3575	23.7090	40%
Couple	24.5410	25.1162	25.1211	35%
Cameraman	21.4369	21.9221	22.2173	40%

Fig.4. Visual comparison of the results; Column (a) Low resolution downscaled input image; (b) Results of Bicubic method; (c) Results of Yang et al.'s method; (d) Compressive samples given as input to the proposed method; (e) Results of the proposed method.

5. Conclusion

This paper presents an image super-resolution approach for creating high resolution images directly from compressive samples utilizing non-local patch based interpolation combined with dictionary learning. Experimental analysis prove that the proposed method is able to overcome the restriction of other dictionary learning methods which cannot utilize the random samples from a CS based acquisition system and also dealt with the neighborhood issue of single image super resolution schemes. Quantitative and subjective visual analysis shows that the proposed method can attain good results with less number of random samples.

References

1. R. Y. Tsai and T. S. Huang, Multiple frame image restoration and registration, *Advances in Computer Vision and Image Processing*, Greenwich, CT: JAI Press Inc., pp. 317-339, 1984.
2. Sean Borman and Robert L. Stevenson, Super-resolution from image sequences - A review, *Proceedings of the 1998 Midwest Symposium on Circuits and Systems*, pp. 374-378, 1998.
3. Sung C. Park, Min K. Park, and Moon G. Kang, Super-resolution image reconstruction: a technical overview, *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21-36, 2003.
4. D. L. Donoho, Compressed sensing, *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
5. J. Yang, J. Wright, T. S. Huang, and Y. Ma, Image super-resolution as sparse representation of raw image patches, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008.
6. J. Yang, J. Wright, T. Huang, and Y. Ma, Image superresolution via sparse representation, *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
7. R. Zeyde, M. Elad, and M. Protter, On Single Image Scale-Up using Sparse-Representations, *Curves & Surfaces, LNCS 6920*, pp. 711–730, 2011.
8. Kevin (Xu) Su, Qi Tian, Qing Xue, Nicu Sebe, Jingsheng Ma, Neighborhood issue in single-frame super resolution, *Proceedings of IEEE International Conference on Multimedia and Expo*, July 2005.
9. Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian,, Image denoising with block-matching and 3D filtering, *Electronic Imaging'06,* Proc. SPIE 6064, No. 6064A-30, 2006.
10. M. Aharon, M. Elad, and A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
11. Pradeep Sen, Soheil Darabi, Compressive Image Super-resolution, *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference*, pp. 1235 – 1242, Nov. 2009.
12. 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, no. 12, pp. 1621–1633, Dec. 1997.
13. S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, Fast and robust multiframe super-resolution, *IEEE Transactions on Image Processing*, vol. 13, no.10, pp. 1327–1344, Oct. 2004.
14. R. Fattal, Image upsampling via imposed edge statistics, *ACM Transactions Graphics*, vol. 26, no. 3, article no. 95, 2007.
15. Jian Zhang, Chen Zhao, Ruiqin Xiong, Siwei Ma, Debin Zhao, Image super resolution via dual dictionary learning and sparse representation, *IEEE International Symposium of Circuits and Systems,* 2012, arXiv:1208.3723.
16. S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong, Soft edge smoothness prior for alpha channel super resolution, *Proceedings of IEEE Conference on Computer Vision and Pattern Classification*, pp. 1–8, 2007.
17. M. Irani and S. Peleg, Motion analysis for image enhancement: Resolution, occlusion, and transparency, *Journal of Visual Communication and Image Representation*, vol. 4, no. 4, pp. 324–335, Dec 1993.
18. 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.
19. Freeman W.T., Jones, T.R., Pasztor, E.C., Example-based super-resolution, *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp.56–65, 2002.
20. J. Sun, N. N. Zheng, H. Tao, and H. Shum, Image hallucination with primal sketch priors, *Proceedings of Conference on Computer Vision and Pattern Recognition,* vol. 2, pp. 729–736, 2003.
21. H. Chang, D.-Y. Yeung, and Y. Xiong, Super-resolution through neighbour embedding, *Proceedings of IEEE Conference on Computer Vision and Pattern Classification*, vol. 1, pp. 275–282, 2004.