Image Super-Resolution using Single Image Semi Coupled Dictionary Learning

Hemant S Goklani hsgoklani@gmail.com
ECED, SVNIT
Surat, 395007, India

Shravya S shravya2490@gmail.com
ECED, SVNIT
Surat, 395007, India

Jignesh N Sarvaiya jns@eced.svnit.ac.in
ECED, SVNIT
Surat, 395007, India

Abstract

Obtaining a high resolution image from a low resolution image plays an important role in many image processing applications. In Single Image Super Resolution (SISR), the desired high resolution output image is synthesized from a single low resolution input image. In this paper, Single Image Semi Coupled Dictionary Learning (SI-SCDL) method is proposed, where the dictionaries to represent the high and low resolution images are trained from the input image itself. In the proposed method, the online training stage is employed, where the dictionaries are learnt online and it does not require any external training database. Simulation results show that the proposed SI-SCDL method performs better when compared to other mentioned methods.

Keywords: Super-resolution, Single Image Super Resolution (SISR), Single Image Semi Coupled Dictionary Learning (SI-SCDL).

1. INTRODUCTION

Image super-resolution deals with increasing the spatial resolution of any low resolution image. The applications of image super-resolution are seen in the fields of satellite imaging, medical imaging, forensic science, etc. Among the various methods available in literature, image interpolation methods, i.e., the bicubic interpolator [1] and the edge guided interpolators [2], interpolate the missing high resolution pixel as the weighted average of its local neighbors. Image super-resolution problem can also be viewed as the inverse problem, where series of linear inverse estimators of high resolution image are computed based on different priors on the image regularity [4]. This paper concentrates on SISR, wherein the high resolution image is obtained from the single low resolution input image. In Multi Image Super-Resolution (MISR), multiple low resolution images of the same scene are required to reconstruct the high resolution image. In various practical applications, multiple images of the same scene may not be available to employ the MISR method. Yet another method of image super-resolution is an example or learning-based methods offers better results, but these models require a training database and quality of results depend on the size and type of database [5], [6]. To overcome this limitation, recent approaches assumed the re-occurrence of similar patches within and across image scales from which the high resolution image patches are predicted [7]. Utilizing the fact that image patches can be well represented using sparse representation of vectors, Yang et.al, proposed image super-resolution as a sparse representation problem where the image patches are well represented using the trained dictionary [8], [9]. Yang et.al employed Joint Dictionary Learning (JDL), where the dictionaries are trained to together assume the sparse representations of both the low and high resolution images are the same. This was further extended to obtain Semi-Coupled Dictionary

Learning (SCDL), which assumes a mapping between sparse representations [14]. Though the SCDL method outperforms JDL method, it requires an external training database to train the dictionaries. The dictionaries also have to be trained for the required scaling factor. Accuracy of multiple image based SCDL algorithm is dependent on the training data. The quality of the results obtained depend on the size and type of training data utilized. If a large amount of similar class training data was available only then satisfactory results will be obtained. In real world scenario, for an unknown input image, multiple images of the similar class may not be available as the training database. In this situation, the proposed SISR based SCDL can be applied to reconstruct high resolution images from a single input image only. In the proposed method online dictionary training is employed where the dictionary is trained specifically to every input low resolution image. The training data are obtained from the input low resolution image itself. The proposed method does not require any additional dictionary training time and memory space for storage of learned dictionaries.

2. PROPOSED METHOD

The proposed Single Image SCDL method can be represented by two stages, (i) Online training stage (ii) Execution stage as shown in Fig. 1.

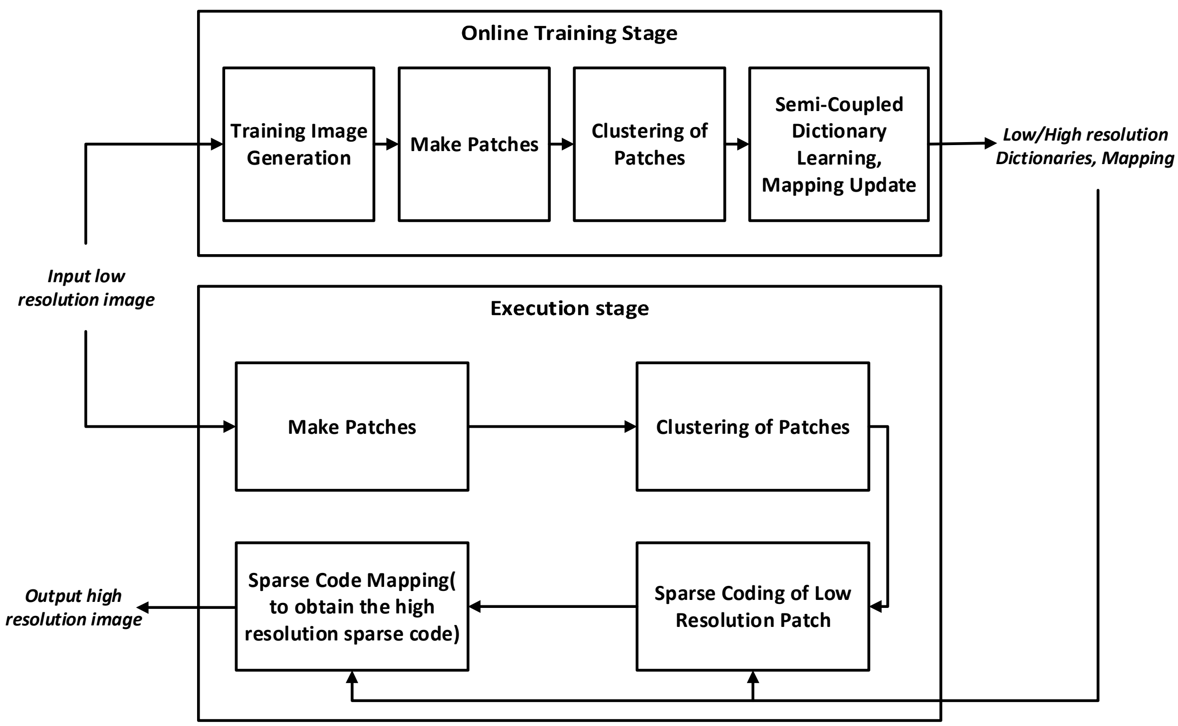


FIGURE 1: Online training stage and execution stage of SI-SCDL

The low resolution image is given as the input to both the training stage and the execution stage. The first step in the online training stage is training data (low-high resolution image pairs) formation. Low and high resolution image patches are extracted from the training data. The patches are clustered, and for every cluster low, high resolution dictionaries and the mapping are learnt. After the completion of training stage, execution stage is processed. The input image patches are extracted followed by clustering. For every clustered low resolution image patch, using the corresponding dictionary from the online training stage, the sparse representation is obtained. The high resolution patch sparse representation is obtained using the corresponding map- ping function. Using the high resolution dictionary obtained from the training stage, the final output high resolution image is obtained.

2.1 Training Data Formation

In the proposed method online dictionary training is employed, wherein the low and high resolution dictionaries are trained separately for every input image. This requires the training data to be obtained from the input image. The training data are comprised of low-high resolution image pairs. From the input low resolution image, image pyramid is constructed as $\{LR_i\}_{i=1}^{-n}$ *i.e.* $\{LR_{-1}, LR_{-2}, ...\}$ by down sampling the input image by the required scaling factor. Every down sampled image is up sampled again using bicubic interpolation method by the same factor to obtain its higher resolution version $\{HR_i\}_{i=1}^{-n}$. The image pyramid thus formed consists of a set of low resolution image which form the LR space and high resolution image set form the HR space. The low-high resolution image pairs form the training data for online dictionary learning as shown in Fig 2.

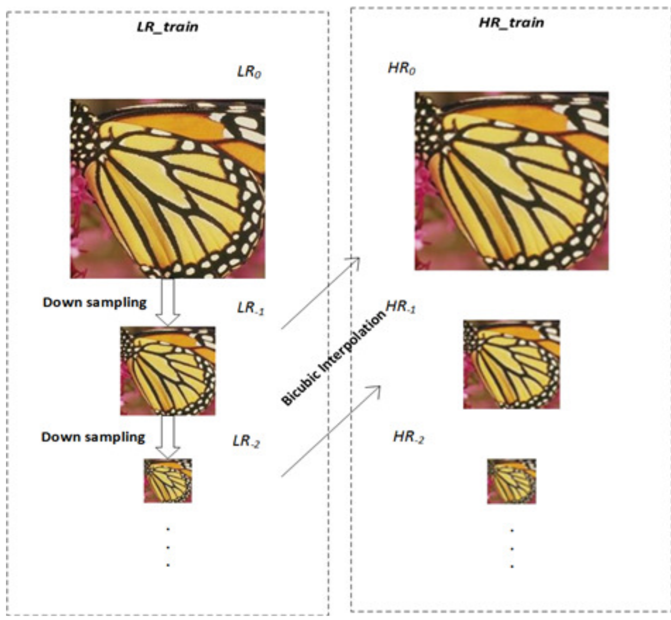


FIGURE 2: Training data formation in SI-SCDL.

2.2 Clustering in SI-SCDL

Patches are extracted from the training data obtained, and is clustered using K-means algorithm to group the similar patches together. Separate dictionaries are learnt for each cluster, which aids for better image representation.

2.3 Dictionary training and mapping update for SI-SCDL

Let $\{LR_{train_i}\}_{i=1}^{n}$ and $\{HR_{train_i}\}_{i=1}^{n}$ denote the training dataset. From these training images the patches are extracted as $\{LR_{patch_i}\}_{i=1}^{pl}$ and $\{HR_{patch_i}\}_{i=1}^{ph}$. The patches are divided into CN clusters as $\{LR_{cl_c}\}_{c=1}^{CN}$ and $\{HR_{cl_c}\}_{c=1}^{CN}$. For every cluster, dictionary pair $\{D_{LR_c}\}_{c=1}^{CN}, \{D_{HR_c}\}_{c=1}^{CN}$ and the mapping function $\{M_c\}_{c=1}^{CN}$ are learned. The semi coupled dictionary and the mapping between the sparse representation of individual cluster can be found by minimizing the below equation as,

$$\begin{aligned} \min_{D_{LRc}, D_{HRc}, M_c(.)} & E_{rr}(D_{LRc}, LR\_cl_c) + E_{rr}(D_{HRc}, HR\_cl_c) \\ & + \gamma E_{map}(M_c(\alpha_{LRc}), \alpha_{HRc}) + \lambda E_{reg}(\alpha_{LRc}, \alpha_{HRc}, M_c(.), D_{LRc}, D_{HRc}), \quad \forall c = \{1, 2..CN\} \end{aligned} \tag{1}$$

The first two terms in Equ.(1) represent the error in the representation of low resolution patch and high resolution patch using their respective dictionaries D_{LR_c} , D_{HR_c} and their sparse representation α_{LR_c} and α_{HR_c} respectively. E_{map} term represents the error in mapping between the sparse coefficients. E_{reg} is added to regularize the sparse coefficients and the mapping. The sparse representation of LR_{cl_c} and HR_{cl_c} over D_{LRc} and is related via the mapping $M_c(.)$. Equ. (1) jointly optimizes the dictionary pair and the mapping function. Assuming the mapping function to be linear and bidirectional, then the Equ. (1) can be rewritten as.

$$\min_{D_{LRc},D_{HRc},M_c} \| LR\_cl_c - D_{LRc}\alpha_{LRc} \|_F^2 + \| HR\_cl_c - D_{HRc}\alpha_{HRc} \|_F^2 + \gamma \| \alpha_{HRc} - M_c\alpha_{LRc} \|_F^2$$
$$+ \lambda_L \| \alpha_{LRc} \|_1 + \lambda_H \| \alpha_{HRc} \|_1 + \lambda_M \| M_c \|_F^2 \quad \text{s.t } \| D_{LRci} \|_{l_2} \le 1,\ \| D_{HRci} \|_{l_2} \le 1 \ \ \forall i,c \tag{2}$$

where γ, λ_L, λ_H, λ_M are the regularization parameter to balance the importance of the terms in the equation. D_{LRc} and D_{HRc} are the dictionaries for the low and high resolution image clusters LR_{cl_c} and HR_{cl_c} respectively. M_c is the mapping function between the sparse coefficients. This equation is convex with respect to each of them individually keeping the others fixed. M_c is assumed be identity matrix then the sparse coefficients becomes equal. Equ. (2) can be solved by the below three steps:

i. Sparse Coding

The dictionaries D_{LRc} and D_{HRc} , and the mapping function M_c are assumed to be fixed. M_c is assumed to be linear bidirectional, thus α_{LR_c} can be transformed to α_{HR_c} and vice versa. M_c is initialized as the Identity matrix. Based on above, the sparse coefficients are calculated as,

$$\min_{\alpha_{LRc}} \| LR\_cl_c - D_{LRc}\alpha_{LRc} \|_F^2 + \gamma \| \alpha_{HRc} - M_{LRc}\alpha_{LRc} \|_F^2 + \lambda_L \| \alpha_{LRc} \|_1 , \quad \forall c$$
$$\min_{\alpha_{HRc}} \| HR\_cl_c - D_{HRc}\alpha_{HRc} \|_F^2 + \gamma \| \alpha_{LRc} - M_{HRc}\alpha_{HRc} \|_F^2 + \lambda_H \| \alpha_{HRc} \|_1 , \ \forall c \tag{3}$$

where λ_L and λ_H are the regularization parameters used to balance the fidelity of the representation with sparsity. Equ. (3) represent the multi task LASSO problem, can be solved using the LASSO algorithm [10].

ii. Dictionary Learning: With the sparse coefficients α_{LR_c} and α_{HR_c} calculated from the Equ. (1) kept fixed and the mapping function M_c fixed as in the above step, the dictionaries D_{LRc} and D_{HRc} are updated as,

$$\min_{D_{LRc},D_{HRc}} \| LR\_cl_c - D_{LRc}\alpha_{LRc} \|_F^2 + \| HR\_cl_c - D_{HRc}\alpha_{HRc} \|_F^2$$
$$\text{s.t} \| D_{LRci} \|_{l_2} \le 1, \| D_{HRci} \|_{l_2} \le 1, \ \forall c \tag{4}$$

Where the term definition remains the same as in Equ.(1) and (2). Equ (4) is QCQP problem and is solved as stated in [11].

iii. Mapping Update: After sparse coding and dictionary learning, it is kept fixed for mapping updated as,

$$\min_{M_c} \| \alpha_{HRc} - M_c \alpha_{LRc} \|_F^2 + (\lambda_M / \gamma) \| M_c \|_F^2, \quad \forall c \tag{5}$$

where M_c is the mapping and α_{HR_c} , α_{LR_c} are the sparse coefficients of the high and low resolution patch belonging to the c^{th} cluster. λ_M , γ are the regularization parameters. Equ. (5) is a ridge regression problem and is solved to obtain the mapping function as,

$$M_c = \alpha_{HRc} \alpha_{LRc}^T (\alpha_{LRc} \alpha_{LRc}^T + (\lambda_M / \gamma) I)^{-1}, \quad \forall c \tag{6}$$

where I represents the identity matrix. After the above mentioned three steps the dictionary pair $\{D_{HR_c}\}_{c=1}^{CN}$, $\{D_{LR_c}\}_{c=1}^{CN}$ and the stable bilinear $\{\mathrm{M}_c\}_{c=1}^{CN}$ is learnt is learnt for all the clusters. After successful completion of online system training, multiple cluster specific low and high resolution dictionaries and feature space mapping is learnt.

2.4 Execution stage of the proposed SI-SCDL method

After the training stage, given a low resolution LR image, it is first converted into $\{LR_i\}_{i=1}^{p}$ patches. Initial estimate of the high resolution version is obtained by applying the simple method of bicubic interpolation. High and low resolution patches are clustered using K-means algorithm as $\{LR_c\}_{c=1}^{CN}$ and $\{HR_c\}_{c=1}^{CN}$. For every patch in $\{LR_i\}_{i=1}^{p}$, its corresponding high resolution patch in $\{HR_i\}_{i=1}^{p}$ is obtained by solving the below optimization equation to obtain the sparse representation α_{LR_c} and α_{HR_c} as,

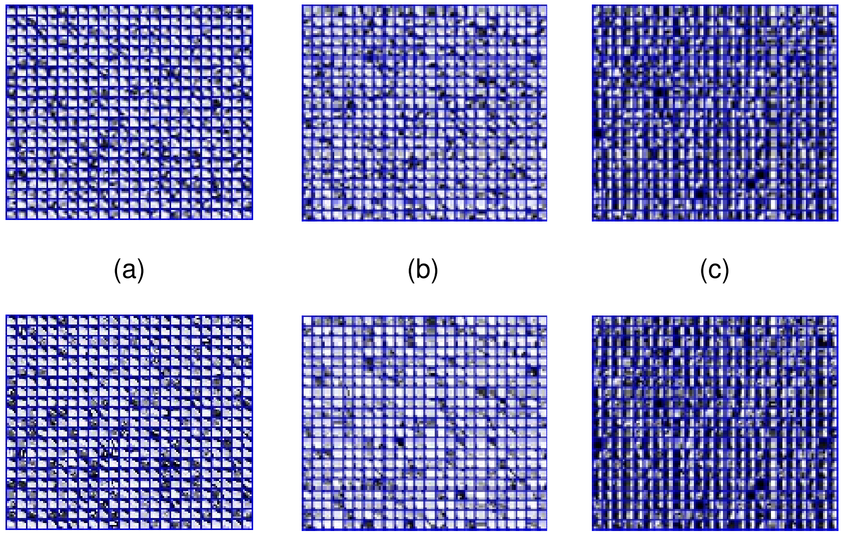
$$\begin{aligned} &\min_{\alpha_{LRc}, \alpha_{HRc}} \| LR_c - D_{LRc} \alpha_{LRc} \|_F^2 + \| HR_c - D_{HRc} \alpha_{HRc} \|_F^2 \\ &+ \gamma \| \alpha_{HRc} - \mathrm{M_c}\, \alpha_{LRc} \|_F^2 + \lambda_L \| \alpha_{LRc} \|_1 + \lambda_H \| \alpha_{HRc} \|_1 \end{aligned} \tag{7}$$

Equ.(7) can be solved alternatively for α_{LR_c} and α_{HR_c} . After obtaining α_{HR_c} each patch can be obtained as,

$$HR_c = D_{HRc} \alpha_{HRc} \tag{8}$$

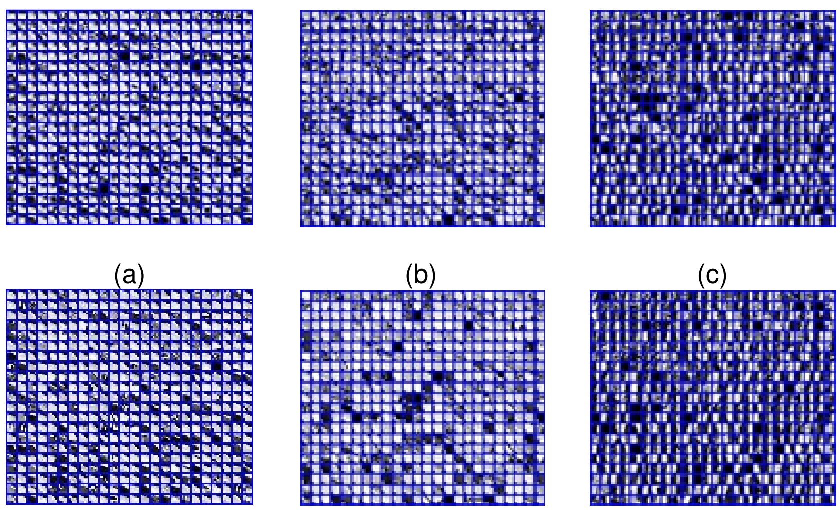
3. SIMULATION RESULTS

This section presents simulation results for evaluating the performance of the proposed SI-SCDL method. The algorithm was run on Telsa K10 GPU accelerator using Matlab 2014a. The proposed algorithm is compared with various state-of-the-art methods such as Bicubic interpolation [1], New Edge Directed Interpolation (NEDI) [3], ScSR [9] and NE [15]. Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are the parameters which are considered for quantitative comparison. The zoomed in part of the image is used for qualitative analysis.



(d) (e) (f)

FIGURE 3: Initial low resolution dictionaries (a) Cluster 2 (b) Cluster 4 (c) Cluster 6, Initial high resolution dictionaries (d) Cluster 2 (e) Cluster 4 (f) Cluster 6.



(d) (e) (f)

FIGURE 4: Learned low resolution dictionaries (a) Cluster 2 (b) Cluster 4 (c) Cluster 6, Learned high resolution dictionaries (d) Cluster 2 (e) Cluster 4 (f) Cluster 6.

The images considered for super-resolution are taken from the USC-SIPI [12] and Berkeley image segmentation databases [13]. Single image super-resolution with scaling factor 2 is considered. The ground truth high resolution image considered is 256 × 256, which is down-sampled to obtain the low resolution image of size 128 × 128. In the proposed method for obtaining the training images using bicubic interpolation 3 levels of low, high image resolution images are considered. Thus the lowest resolution of the image considered for training is 32 × 32. Overlapping image patches of size 5 × 5 is considered. The numbers of dictionary atoms for every cluster for both D_{LRc} and D_{HRc} is 512 each and no of clusters are set to 64 empirically. We empirically set the regularization parameters as $\lambda = \gamma = 0.01$ and $\lambda_R = 0.001$. Initial low resolution dictionary and high resolution dictionary for cluster 2, 4, 6 are shown in Fig. 3. After the dictionary learning stage, the final low and high resolution dictionaries are shown in Fig. 4. It can be seen that the cluster specific dictionaries differ from one another. They are learnt to better represent image patches belonging to specific cluster.

Table 1 and Table 2 compares the results of different image super-resolution methods in terms of PSNR and SSIM. It can be seen that proposed SI-SCDL method achieves the highest PSNR and SSIM values for most of the images considered and outperforms many super-resolution approaches, such as NEDI [3] and ScSr [9].

Image	Bicubic[1]	NEDI[3]	ScSR[9]	NE[15]	Proposed SI-SCDL
Cameraman	26.15	27.87	27.73	27.76	**27.89**
Fence	24.83	25.43	26.33	26.52	**26.86**
Foreman	35.63	38.89	38.65	38.85	**39.33**
Girl	36.39	36.94	36.35	36.55	**37.07**
Starfish	30.00	32.51	33.48	33.48	**33.53**

TABLE 1: PSNR (dB) of different methods (Scaling factor = 2).

Image	Bicubic[1]	NEDI[3]	ScSR[9]	NE[15]	Proposed SI-SCDL
Cameraman	0.86	0.90	0.90	0.90	**0.91**
Fence	0.74	0.80	0.80	0.81	**0.82**
Foreman	0.94	0.97	0.97	0.97	**0.98**
Girl	0.90	0.91	0.89	0.90	**0.91**
Starfish	0.89	0.94	**0.96**	0.95	**0.96**

TABLE 2: SSIM of different methods (Scaling factor = 2).

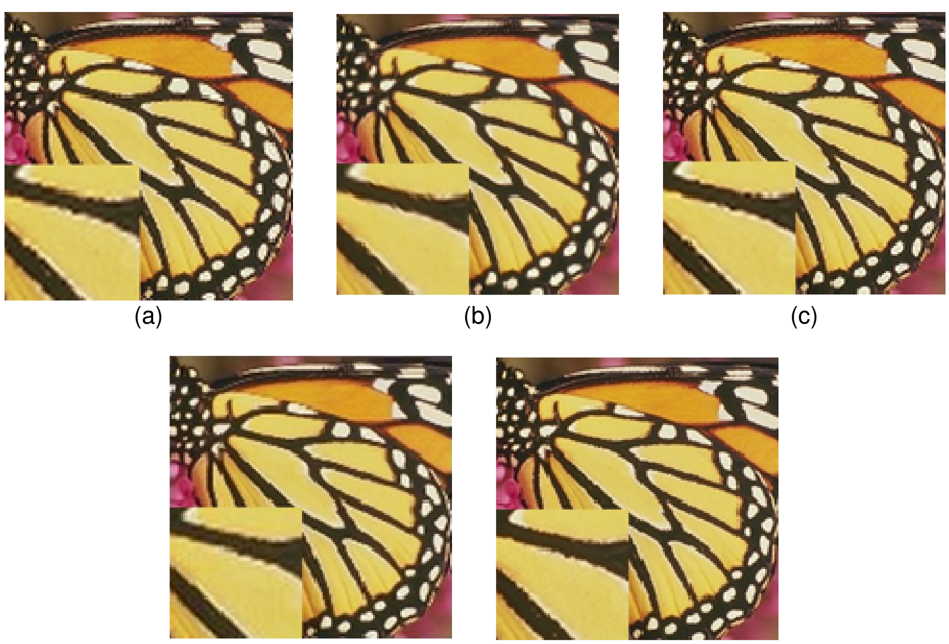
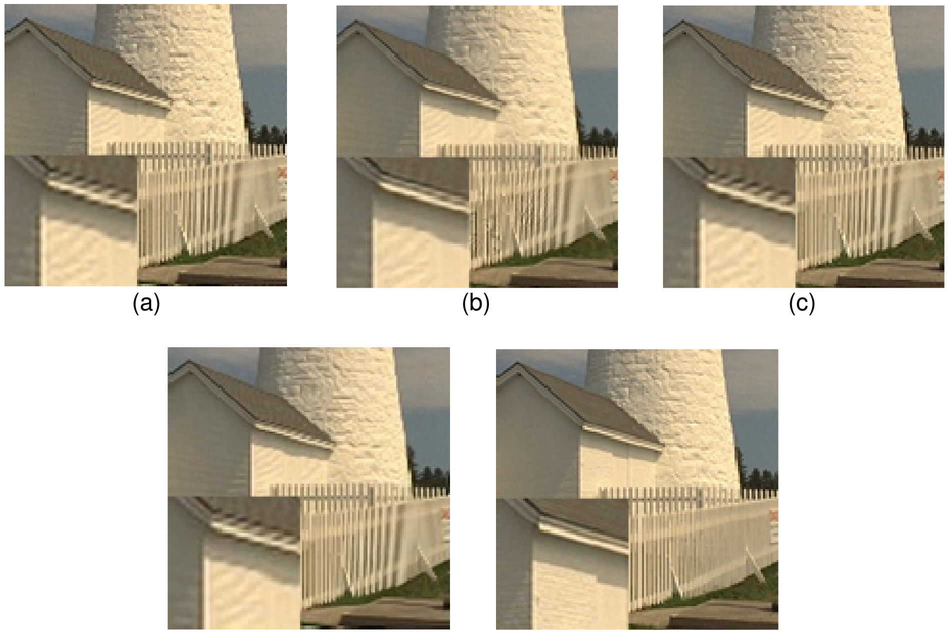


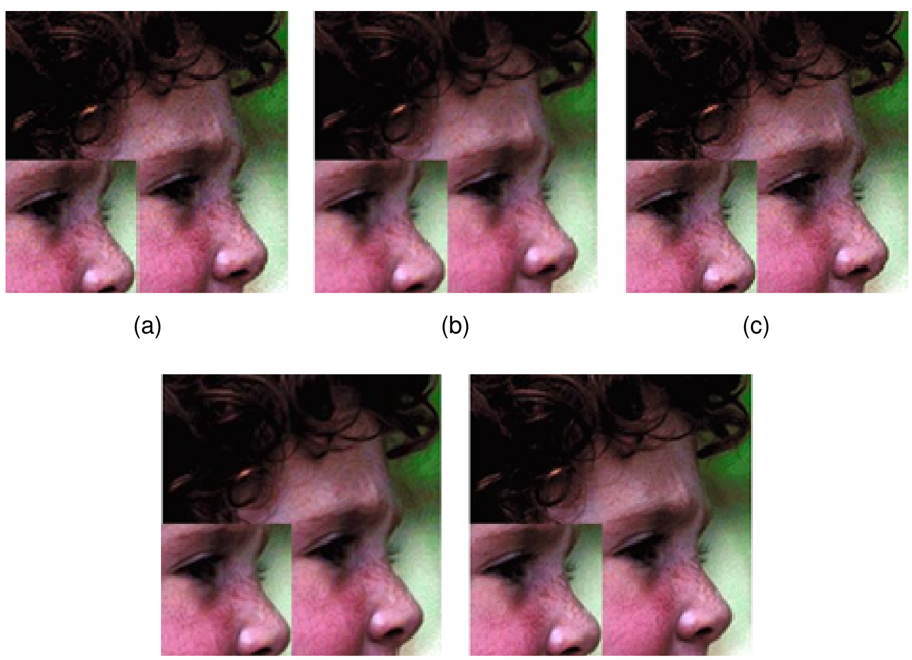
FIGURE 5: The comparison of the proposed algorithm and other methods on Butterfly image and the local magnification also shown in the lower left corner (a) Bicubic [1] (b) NEDI [3] (c) ScSR [9] (d) Proposed SI-SCDL (e) Original image.



(a) (b) (c)

(d) (f)

Figure 6: The comparison of the proposed algorithm and other methods on Fence image and the local magnification also shown in the lower left corner (a) Bicubic [1] (b) NEDI [3] (c) ScSR [9] (d) Proposed SI-SCDL (e) Original image.



(a) (b) (c)

(d) (f)

Figure 7: The comparison of the proposed algorithm and other methods on Girl image and the local magnification also shown in the lower left corner (a) Bicubic [1] (b)NEDI [3] (c) ScSR [9] (d) Proposed SI-SCDL (e) Original image.

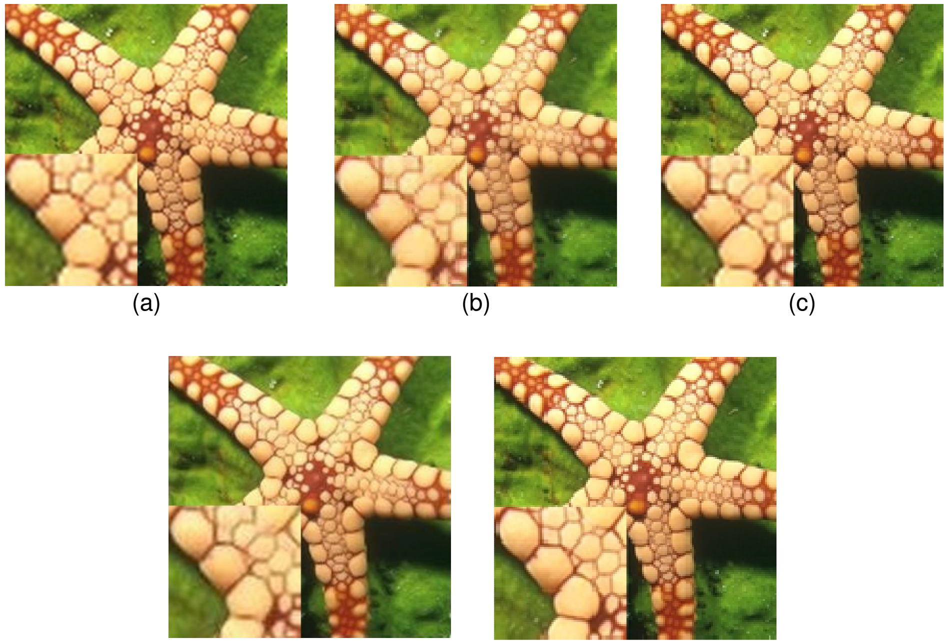


Figure 8: The comparison of the proposed algorithm and other methods on Starfish image and the local magnification also shown in the lower left corner (a) Bicubic [1] (b) NEDI [3] (c) ScSR [9] (d) Proposed SI-SCDL (e) Original image.

In case of Foreman image, maximum PSNR improvement of 3.7dB is and SSIM improvement of 0.08 for Fence image is seen. On an average over the image considered, NEDI and ScSR shows an improvement of 1.72dB and 1.9dB as compared to Bicubic respectively. However, the proposed SI-SCDL method on average provides PSNR improvement of 2.34 dB as compared to Bicubic. From the zoomed in region it can be observed that the results of proposed SI-SCDL method is relatively similar to the original image. The results for butterfly, fence, girl and starfish images for scale 2 are shown in Fig. 5, Fig. 6, Fig.7 and Fig. 8 respectively.

4. CONCLUSIONS AND FUTURE SCOPE OF PROPOSED WORK

The proposed method obtains the training dataset for dictionary learning from the input image itself and does not require any external training database. The training stage is online and dictionary is learnt specific for every individual input *i.e.,* the SI-SCDL method becomes input specific. From the simulation results it can be noted that the proposed SI-SCDL method outperforms many existing methods both in terms of visual quality and performance measures. Future research on the proposed work should include mapping function improvement through patch feature based dictionary learning and achieve single image super-resolution for higher scaling factor.

5. REFERENCES

[1] R. Keys, “Cubic convolution interpolation for digital image processing. Acoustics, Speech and Signal Processing”, IEEE Trans on, vol. 29(6), pp. 1153-1160, 1981.

[2] X. Zhang and X. Wu. Image interpolation by adaptive 2-d autoregressive modeling and soft-decision estimation. IEEE Trans on IP, vol. 17(6), pp. 887-896, 2008.

[3] X. Li and M. Orchard, “New edge directed interpolation, IEEE Trans on IP, vol.10, no. 10, pp.1521-1527, 2001.

[4] S. Mallat and G. Yu, “Super-resolution with sparse mixing estimators”, IEEE Trans on IP, vol.19(11), pp. 2889-2900, 2010.

[5] W. T. Freeman, T. Jones, and E. Pasztor “Example based super-resolution”, IEEE Computer Graphics and Applications, 2002.

[6] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution via sparse representation”, IEEE Trans. Image Process.,vol 19(11), pp.2861-2873, 2010.

[7] Daniel Glasner, Bagon Shai, Michal Irani, “Super-resolution from a single image”,12th International Conference on Computer Vision, IEEE, 2009.

[8] Jianchao Yang, John Wright, Thomas Huang, and Yi Ma, “Image super-resolution as sparse representation of raw image patches”, IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2008.

[9] Jianchao Yang, John Wright, Thomas Huang, and Yi Ma, “Image super-resolution via sparse representation”, IEEE Transactions on Image Processing, vol. 19, no. 11, pp. 2861-2873, 2010.

[10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression”, The Annals of statistics, vol. 32(2), pp.407-499, 2004.

[11] M. Yang, L. Zhang, J. Yang, and D. Zhang, “Metaface learning for sparse representation based face recognition”, In ICIP, pp. 1601-1604. IEEE, 2010.

[12] A. G.Weber, “The USC-SIPI Image Database”, tech. rep., University of Southern California, Signal and Image Processing Institute, Department of Electrical Engineering, Los Angeles, CA 90089, 2564 USA, 3740 McClintock Ave, 1997.

[13] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics”, in Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, 2, vol.2, pp. 416-423, 2001.

[14] S. Wang, L. Zhang, L. Y., and Q. Pan, Semi-coupled dictionary learning with applications in image super-resolution and photo-sketch synthesis, in International Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012.

[15] Chang, Hong, Dit-Yan Yeung, and Yimin Xiong. "Super-resolution through neighbor embedding." In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, vol. 1, pp. I-I. IEEE, 2004.