A New Single-Image Super-Resolution Using Efficient Feature Fusion and Patch Similarity in Non-Euclidean Space

Rajashree Nayak[1] · Bunil Kumar Balabantaray[2] · Dipti Patra[3]

Abstract

Efficient trade-off between the reconstruction qualities and the processing time of any single-image super-resolution reconstruction (SISRR) approach is critically influenced by two major aspects. These aspects are (i) appropriate representation of image patch in feature space and (ii) effective searching of candidate patches from the pool of training patches or learned dictionary. This paper proposes a neighbor embedding-based SISRR method. Novelties of our work include integration of (i) efficient feature mapping scheme which fuses multiple correlated features naturally, (ii) faster searching of candidate patches by measuring the patch correlation in non-Euclidean space and (iii) adaptive selection of neighborhood size using patch characteristic. Correlation among features is modeled via global covariance matrix, and the fusion process enables to preserve sufficient structural, spatial correlation among patches. Distance functions based on notion of generalized eigenvalue are used for measuring patch similarity which support faster searching of candidate patches. Performance analysis of the suggested method is compared with some of the competent state-of-the-art methodologies. From the simulated result analysis, proposed work is found to be outperforming in terms of sharpened image details with diminished effect of artifacts at a reasonable computational burden.

Keywords Super-resolution · Non-Euclidean space · Covariance matrix · Neighbor embedding · Neighborhood · Similarity

1 Introduction

Machine vision systems (MVSs) have a wide range of applications in various industry, academia, etc. Usually, MVSs are assembled around the right camera resolution and utilize sensors with well-specialized optics to capture high-quality images. These images are further processed, analyzed and measured by any other hardware or software techniques for decision making. It can effortlessly inspect the image details that are hard to be seen by human eye. However, current imaging system generates images with limited resolution and needs to be enhanced for obtaining high resolution (HR) images. Resolution of a digital image is defined as the detail information contained in that image. HR images contain high pixel density, encompass essential and critical image details and hence offer enhanced visual perception. Thus, HR images become essential in various imaging applications such as forensic image analysis, video surveillance [1], medical imaging, image security analysis [2] and many more. However, the resolution of an image is exclusively influenced by the properties and specification of the capturing device. Possible reasons for the limited resolution may be due to different physical constraints, insufficient amount of photodetectors, inferior spatial sampling rate and inappropriate image capturing process. Moreover, images captured in poor environmental condition get contaminated with different degradations such as fog, haze and smog. Consequently, quality of images gets degraded and results in low resolution (LR) images. Various dehazing and defogging techniques [3,4] have been suggested in the literature to reinstate the visibility and to diminish the artifacts. These methods works

✉ Rajashree Nayak
    rajashreenayak17@gmail.com

    Bunil Kumar Balabantaray
    bunil@nitm.ac.in

    Dipti Patra
    dpatra@nitrkl.ac.in

1   Department of Electrical Engineering, National Institute of Technology Meghalaya, Shillong, India

2   Department of Computer Science Engineering, National Institute of Technology Meghalaya, Shillong, India

3   Department of Electrical Engineering, National Institute of Technology Rourkela, Rourkela, India

efficiently as the preprocessing step to the reconstruction process but hardly improve the perceived resolution. Possible approaches to boost up the spatial resolution of the captured LR images are either by redesigning the hardware (by integrating adequate number of photodetector or by increasing pixel density per unit area) or by increasing the chip size. However, resolution enhancement by redesigning the hardware is critically challenging, costly and time-consuming in nature. As an alternative, image reconstruction is a mathematical process which generates images from the acquired projections captured at different angles or at different times. It aims at retrieving information that has been lost or obscured in the imaging process without sacrificing the quality of the image. Basically, reconstruction process is intended to model the degradation phenomenon and utilizes the reverse process to retrieve the original images from the degraded, noisy, blurred and aliased images. Over the few decades, Super-resolution reconstruction (SRR) has been recognized as one of the evolving and cost effective off-line technology to enrich the resolution of the LR images. This process aims at estimating HR image by fusing the non-redundant information in the single or multiple LR images of the same scene. SRR methods overcome the limitations associated with the optics and sensor technology and efficiently enhance the resolution.

SRR methodologies have been widely used in various imaging applications such as remote sensing [5], medical imaging [6] and machine vision. In the age of advanced space technology, commercial remote sensors capture images of better spatial resolution. However, quality of images degrades due to various atmospheric effects. Consequently, results in nonaligned LR images. In that scenario, SRR techniques behave as adaptive and low-risk solutions which will effectively enhance the resolution of the archive of remotely sensed LR images. In medical imaging, resolution of the captured multimodal images of a particular patient using the same imaging device at different times may not be consistent. This is due to several physical constraints, environmental conditions and also due to the respiratory behavior of the patient. Consequently, captured images are of limited resolution. SRR methodologies help to increase the resolution of captured images for better diagnostic analysis and proper treatment planning. Another important application of SRR is in the field of video surveillance [7]. Coding of surveillance video for various application needs large volume of memory and storage. As an alternative, we can compress the original video sequences into its lower version to store them effectively. Their resolution can be further improved via SRR techniques at the user end.

SRR of a HR image by using multiple number of LR images is referred as multi-image SRR (MISRR), whereas reconstruction of a HR image with single LR image is referred as single-image SRR (SISRR). Performance of the MISRR methods relies on the accuracy of motion estimation among the LR observations which is again an open research problem. Moreover, error in the registration process affects a lot to the reconstruction process. Furthermore, availability of a series of LR observations of a particular scene encompassing sub-pixel shift among each other is not an easy task. All these limitations necessitate the use of SISRR in reconstructing a HR image. SISRR aims at estimating a pleasant HR solution with enhanced image detail from its LR image. Learning-based approaches have been drawing widespread attention in SISRR as these types of reconstruction process provide better reconstruction capability even at higher magnification factors. These approaches exploit some image priors extracted from external training image datasets to recover the missing image details in the HR image.

Figure 1 depicts the basic flowchart of learning-based single-image super-resolution reconstruction (LBSISRR) methods. Estimation of the output HR image is performed by utilizing two basic steps such as training stage followed by reconstruction stage. Training stage aims at learning the co-occurrence priors or correlation among the training image patches via some learning models. This stage involves two subsequent steps such as feature extraction step and learning step. Training stage starts with the input LR image whose resolution is to be enhanced and a set of training images. Training images contain various HR images and their corresponding LR images. Prior to feature extraction step, the training/testing HR and LR images are divided into overlapped patches. In the feature extraction step, each training/testing HR and LR image patches are represented in feature space, i.e., any predefined feature extraction operator is used to represent each image patch in terms of feature vectors. Next, the learning step involves different learning models to preserve the co-occurrence prior among images patches. This phase aims at providing potential candidate HR patches corresponding to each input LR image patch for the reconstruction process. Candidate patches are selected by measuring the patch similarity in feature space. After the successful implementation of training stage, the reconstruction stage integrates the learned knowledge as a priori term with the input LR image to preserve the missing details in the estimation process. Finally, the estimated HR patches are stitched to generate the super-resolved output image. Basic steps followed in the reconstruction process are summarized as follows:

– Generation of training image dataset containing HR images and their corresponding LR images
– Division of training images into overlapped patches
– Extraction of features for training image patches
– Division of input LR image into several patches
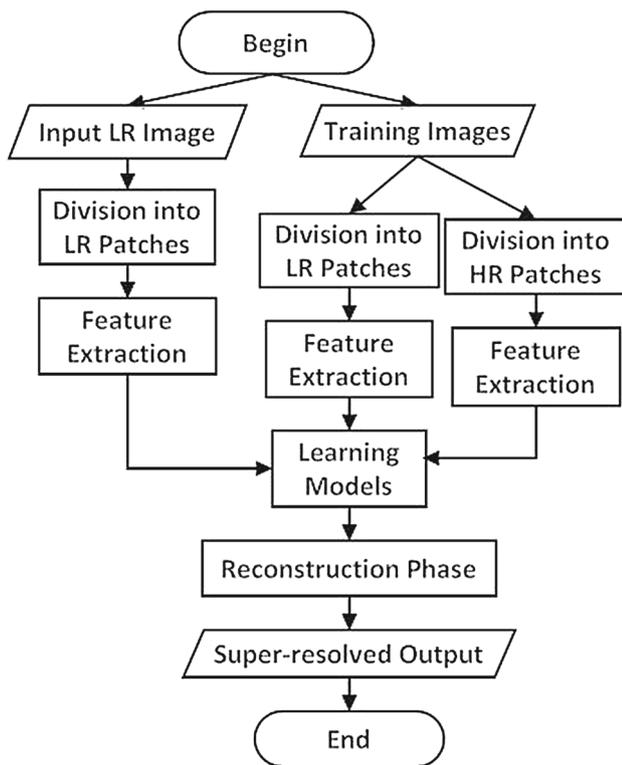– Extraction of feature for the input LR image patches

Fig. 1 Flowchart of LBSISRR method

– Establishment of learning model to learn the co-occurrence priors among patches
– Utilization of appropriate searching method to select candidate patches for each LR patch
– Estimation of HR image by utilizing the priors obtained from the learning models.

Based on the variety of learning models available in the literature, LBSISRR methods are categorized as example-based single-image SRR (EX-SISRR) [8–15] methods, neighbor-embedding-based single-image SRR (NE-SISRR) methods [16–24], and sparse coding-based single-image super-resolution (SC-SISRR) [25–35] methods. Now-a-days, deep learning-based SRR methods have been successfully implemented for providing fast and high-quality super-resolution (SR) results [36–40].

EX-SISRR methods [8–14] depend on a numerous amount of example training patches for the reconstruction process. Prediction of HR patch is achieved by learning the Markov random field (MRF) model solved by belief propagation (BP) algorithm. Spatial relationship between the LR patches and their corresponding HR patches is learned from the example images using the Markov network. Despite the recent advances, these methods suffer from heavy computational overload and have a weak generality property in preserving the boundary between two highly contrasted patches [14,15].

All these bottlenecks make these methods often too slow for practical use [29]. A detailed analysis about some of the significant EX-SISRR methods is provided in Sect. 2.1. Several NE-SISRR and SC-SISRR methods have been proposed to overcome these issues.

NE-SISRR methodologies are based on the idea of local linear embedding (LLE) from the manifold learning and estimate the HR image by linearly combining the neighboring candidate HR patches. Existing state-of-the-art NE-SISRR methodologies utilize two independent steps: (1) searching of K numbers of candidate patches for each input patch in its feature space and (2) obtaining the reconstruction weights of patches for estimating the HR image. Popularly, size of K is chosen randomly or by means of any distance function [18,23], whereas the weight of each patch is selected by least-squares minimization of reconstruction error. Reconstruction quality of the estimated HR image solely depends on the quality and spatial compatibility among the K number of candidate patches. Computational overload of the estimation process is contributed mainly due to searching of candidate patches using Euclidean distance (Eud) measure. As compared to EX-SISRR methods, NE-SISRR methods exhibit much stronger generalization capability for a range of images [21]. These methods can estimate a high-quality HR solution by utilizing relatively limited training image patches. Some important aspects which influence the performance of NE-SISRR methods are

– To obtain efficient feature mapping scheme in a low dimensional feature space which will help to preserve the missing details in the reconstruction process by selecting relevant candidate patches
– To choose the optimal size of K which will enable faithful reconstruction. As reconstruction of a HR image with fixed number of neighboring patches causes over or under-fitting of data, thus yields blurred output.
– To use a faster searching scheme to get the neighboring patches which will reduce the computational overload of the whole reconstruction process

So far, many improved NE-SISRR methods have been proposed in the literature. The detail description about these methods is provided in Sect. 2.2.

SC-SISRR methods exploit the sparsity prior to solve the reconstruction problem. These methods are based on the assumption that the LR image patch and its corresponding HR patch shares the same sparse representations. As the initial step, these methods aim at learning dictionaries from the external image datasets. Next step aims at searching the dictionary for each image patch to obtain the optimal match. Performance of these methods solely relies on dictionary learning in its feature space and main time-consuming step is the searching mechanism for obtaining the similar candidate

patches from the dictionary. In the literature, a number of advanced SC-SISRR methods have been proposed by learning compact dictionaries and/or by utilizing faster searching schemes. A brief description about some of these methods is provided in Sect. 2.3.

1.1 Gaps in the Literature and Solutions to Break Them

Varieties of SISRR approaches have been suggested in the literature (refer Tables 1, 2, 3) to enable an efficient trade-off between reconstruct quality and processing time. An appropriate balance between the quality of HR solutions and the computational speed highly depends upon the selection of training image dataset, feature extraction operation to represent each patch, searching of candidate patches and the learning model to learn the co-occurrence prior among patches. Selection of an efficient training image dataset or an over-complete dictionary from a set of example image patches plays an important role in the successful accomplishment of the reconstruction task. Quality of reconstruction gets enhanced when the training image patches encompass structural as well as statistical similarity with the input image to be super-resolved [15]. However, learning of training dataset or dictionary is performed only once for any reconstruction problem, hence may not contribute more toward the computational overload of the method. Irrespective of the type of learning model, trade-off process is critically influenced by two major aspects such as (1) selection of appropriate feature type to represent image patch and (2) searching of candidate patches from the pool of training image dataset or learned dictionary. A brief discussion about these two aspects is discussed below.

1.1.1 Selection of Feature Type

After doing a critical review from Tables 1, 2 and 3, it observed that most of the SISRR methods [14,16,18,20, 23,25,26] commonly use first- and second-order gradient operator to represent each patch. These operators individually preserve the local edge information of patches, but their combination may not preserve the local neighborhood compatibility. Moreover, second-order gradients are sensitive to noises. As compared to them, feature vectors using norm luminance (NL) [18,24] efficiently preserve the low frequency information but fail to capture the high frequency information of the patch. Zernike moment (ZM)-based feature vectors [24] are insensitive to noise, efficiently preserve the global compatibility, but have limited reconstruction capability. Field-of-experts (FoE)-based feature vectors in [22] and histograms of oriented gradients (HoG)-based feature vectors in [19,21] help to preserve more image details for images rich in edge information but provide limited perfor-

mance to images rich in structural and geometrical contents. Methods in [15,28,32,34,35] utilize pixel intensity to represent image patch. However, intensity values of the patches are much sensitive to noise and contrast variation. Hence, a small disturbance in intensity value may result a significant change in further processing. Methods in [30,31] use texture of the image as the feature vector to represent image patch. However, this feature vector cannot preserve the structural and statistical correlation among patches, consequently provide limited performance for images rich in geometrical and structural content. Disadvantages of these aforementioned methods necessitate selecting a state-of-the-art feature extraction operator which will fit to variety of image types in preserving high frequency information and spatial correlation in both LR and HR space.

1.1.2 Searching of Candidate Patches

About the searching of candidate patches, almost all the EX-SISRR and NE-SISRR methods unexceptionally utilize KNN-based searching scheme where EuD or its variants are used for measuring the patch similarity [21,35]. EuD makes the searching process computationally intensive, does not preserve sufficient spatial relationship among patches and is sensitive to small amount of disturbances, hence may not select potential candidate patches. Consequently, estimated HR image suffers from various artifacts. Similarly, almost all SC-SISRR methods popularly use orthogonal matching pursuit (OMP) or linear programming (LP) to search the dictionary. This searching process utilizes l_2 norm for minimizing the error which is again computationally prohibitive. However, methods in [21,26,29,32] utilizes some faster searching techniques which results in reducing the computational complexity of the reconstruction process with an compromise with the reconstruction quality. All these aspects motivated us to propose a SISRR method which will provide an efficient trade-off between reconstruction quality and processing speed.

This paper proposes an enhanced yet computationally efficient NE-SISRR method which will enable faster searching of potential candidate patches and will facilitate better preservation of image details with diminished effect of artifacts. Major contributions and novelty of the work include

– *Proposition of a low-dimensional feature mapping scheme to represent image patch in its feature space.* Each patch is represented as a combination of several raw feature attributes such as intensity profile, texture, edge, statistical and spatial information. Fusion of these feature attributes enable to preserve sufficient structural, statistical and spatial homogeneity among patches in the reconstruction process.

– *(2) Development of a faster searching scheme to get potential candidate patches by measuring the similarity in non-Euclidean space.* Patch similarity is measured by distance function based on eigenvalues rather than Eud which speeds up the searching process

– *(3) Implication of an automotive scheme to choose the neighborhood size of K based on the input patch characteristic.* Neighborhood size of each input patch is adaptive in nature and is decided by the patch characteristics.

Afterward, the reconstruction weights are obtained by minimizing the reconstruction error. Finally, the output HR image is estimated by the linear combination of neighboring candidate HR patches. Performance analysis of the suggested method is compared in some of the significant state-of-the-art learning-based methodologies and found to be outperforming in terms of image quality as well as faster searching speed.

Rest of the paper is organized as follows: Sect. 2 provides a brief discussion about some of the significant LBSISRR methods in the literature. Detailed description of the proposed reconstruction method is provided in Sect. 3. Performance evaluation and analysis of the proposed method compared with some of the state-of-the-art LBSISRR methods along with some faster SISRR methods are performed in Sect. 4. Section 5 concludes the paper.

2 Related LBSISRR Methods

2.1 Example-Based Single-Image SRR (EX-SISRR) Methodologies

Freeman et al. [8] pioneered the concept of example-based SRR of single image. In this method, authors select numerous amount of training patches for the reconstruction process. Band-pass filtered feature map is used to represent each patch. Nearest neighbor (NN) search method is used to find candidate patches for each input patch and Eud is used to measure the patch similarity. MRF network is used to learn the correlation among related HR-LR image patches. BP algorithm is used to train the MRF network. Maximum a posteriori (MAP) approach is used to estimate the HR patch by utilizing the learned prior knowledge. However, this method fails to preserve sufficient spatial correlation among patches due to integration of Gaussian smoothness functions in the computation of compatibility functions. Moreover, heavy computational overload, slow rate of convergence make this method prohibitive in practical applications [14]. Number of methods have been proposed in the literature to address these limitations. Authors in [9,10] proposed example-based methods which enable faster convergence rate, whereas methods in [12–15] aim at preserving more image details in the reconstructed output image. In method [9], each image patch is

represented by Gaussian derivative. Position constraint operation is used to compute compatibility function and squared difference is used as the patch similarity measure, whereas primal sketch priors are used to represent patch in its feature space. In [11], searching of candidate patches is done via K-means algorithm to improve the rate of convergence of the reconstruction process. In [12], authors employ Contourlet coefficients to represent raw image patches and weighted Euclidean distance is used to measure the patch similarity. Method in [13] utilizes structural contents to represent each patch and patch similarity is measured via modified Chi-square distance. In [14,15], authors utilize image Euclidean distance (IMED) to measure patch similarity. Method in [15] uses a probability way of searching candidate patches and utilizes edge preserving compatibility functions to preserve more image details. Table 1 provides a brief comparative analysis of the popularly used EX-SISRR methods. From the critical review of the literature, it is observed that EX-SISRR methodologies provide sharper output images by selecting appropriate compatibility functions. But computational burden of these methods remain high due to utilization of numerous amount of example images for the reconstruction process and EuD to measure patch similarity. Hence, these methods remain unsuitable for real-time applications.

2.2 NE-Based Single-Image SRR (NE-SISRR) Methodologies

NE-SISRR methodologies are based on the idea of LLE from the manifold learning. These methods presume that HR and LR image patches share identical local geometry and estimate the HR image as the linear combination of K nearest neighboring patches. In [16], authors introduce the concept of LLE based SRR of image. This method utilizes first- and second-order gradient to represent each patch and uses the fixed size of K for the reconstruction process. The reconstructed output cannot well preserve the high frequency details in the images and sometimes produces too much blurry output images. Afterward, a variety of NE-SISRR methodologies [17–24] are proposed in the literature. In [17], histogram-based matching is used to select efficient training images and image residuals generated by image pyramid are used as feature vectors. Method in [18] emphasizes the importance of edge and neighborhood size in the reconstruction process. This method utilizes edge detection and feature selection to generate discriminate training patches. Combination of first-order gradient and NL feature vectors is used to represent each patch. Method in [19], proposes a partially-supervised neighbor embedding (NE)-based SRR (NE-SRR) process and uses the notion of multi-manifold assumption. In this method, each image patch is represented by means HoG feature vectors. K-mean algorithm via unsupervised Gaussian mixture model is employed to assign class labels in the

Table 1 Comparative analysis of EX-SISRR methods

Methods (Year)	Generation of training dataset	Feature extraction operation measure	Searching process and patch similarity	Learning model	Observations
[8] (2002)	Random	Bandpass filtered feature map	NN using Euclidean distance	MRF + BP	Produces blurry and unpleasing output
[9] (2003)	Randomly selected natural images	Gaussian derivatives	NN by minimizing squared difference sketch priors	MRF and BP with primal	Yields better output for generic images
[10] (2007)	Randomly selected face Images	Horizontal face features	NN via patch position constraint	MRF with compatibility checking algorithm	Increases convergence rate but produces artifacts in output
[11] (2009)	Random face images with sufficient details	Cluster Mean	K-means algorithm	MRF	Faster searching speed with blurred output
[12] (2011)	Random selection of face images	Contourlet coefficient	NN using weighted Euclidean distance	MRF via BP	Produces sharper results than the above methods
[13] (2014)	From the input image itself	Structural similarity	Non-local means algorithm using modified Chi-square distance	MRF with NLM	Produces good results for document images but quality degrades for generic images
[14] (2018)	Mutual information-based method	First- and second-order gradients	NN using image Euclidean distance	MRF with modified compatibility functions	Ringing artifacts dominant for images rich in geometrical structures
[15] (2018)	Spatiogram-based method	Intensity profile	Epitomic matching via probability-based scheme compatibility functions	MRF + edge preserving high computational overload	Produces sharper output image at a

searching of neighboring patches. In [20], discrete-cosine-transform (DCT) coefficients are used as feature vectors. Neighborhood size is selected based on K-means algorithm. In [21], each image patch is represented by HoG features. The initial training dataset is divided into a set of training subsets via K-means clustering. L_2 norm is used as the distance measure in the clustering process. With the HoG, training image pairs in a particular subset carry similar geometry. In this method, searching of candidate neighboring patches and reconstruction weights are performed simultaneously via variants of Robust-SL0 algorithm. This algorithm diminishes the limitation of the KNN searching with EuD and enhances the speed of reconstruction process. In [22], a new feature extraction operator, i.e., FoE is utilized to represent each image patch. Non-local similarity constraint is considered to find the optimal weight in the reconstruction process. K-means clustering using GMM is performed to divide the initial training dataset into 32 number of sub clusters. KNN searching process via EuD is utilized for measuring patch similarity. In [23], a NE-SRR method is proposed which is robust to the outliers present in the input image. This method employs histogram-based matching criterion to select training image dataset and global neighborhood size is selected based on Euclidean and geodesic distance measure. Each patch is represented by eight feature vectors based on gradients along different orientations. In [24], a new feature vectors based on combination of ZM and NL is employed to preserve global structure in the reconstruction process. Hierarchical method using residual variance is used to select the global size of K in the searching process. Table 2 describes a brief analysis of some of the state-of-art NE-SISRR methods.

2.3 Sparse Coding-Based Single-Image Super-Resolution (SC-SISRR) Methodologies

SC-SISRR methods have been widely used in recent years. These methods assume that the LR, HR patches have the similar sparse representation. Any LR patch can be reconstructed as a sparse linear combination of different atoms from an appropriately chosen dictionary. They do not work directly on the raw training image patches rather different compact dictionaries are learned from the set of training patches to capture the co-occurrence priors. In the training step, different learning techniques are utilized to learn the dictionaries followed by the sparse coding (SC) algorithm to obtain the optimal sparse representation (α^*) of each LR patch with respect to the learned over complete dictionary. In the reconstruction step, learned α^* of each LR serves as sparse priors for the reconstruction of its HR patch. In the final step, the estimated HR patches are merged by averaging the overlapping area to obtain the output HR image. Later, various regularization constrains are further used to provide sharpened output by minimizing the effect of reconstruction error.

Yang et al. in [25] applied the SC algorithm for both generic and face image SR. In this method, probabilistic approach is used to learn the over-complete dictionary pairs. Linear programming algorithm is used to find the sparse coefficients. This method founds to be outperforming in providing high-quality SR images but computational time of this method is very high. As an advancement, Zeyde et al. in [26] proposes a faster sparse representation based method. Instead of simultaneous learning of dictionary pairs, authors utilize K-SVD to learn the LR dictionary and pseudo-inverse to learn the HR dictionary. Principal component analysis (PCA) is used for dimension reduction of patch features and orthogonal matching pursuit (OMP) is utilized as the SC algorithm. This SC algorithm enables faster rate of convergence than the SC approach in [25]. Afterward, a variety of methods have been proposed in the literature based on different dictionary learning (DL) strategy as well as SC algorithms to maintain an efficient balance between the computational time and the image quality. In [28], authors utilize semi-coupled dictionary learning (SCDL) approach for cross-style image SRR. In [28], the proposed learning approach enables simultaneous learning of dictionary pairs as well as mapping functions to preserve better compatibility between two style domains. Moreover, K-means clustering as well as image nonlocal-redundancy is also integrated in the SCDL model to enhance its stability. The proposed method serves as an effective alternative for photo-sketch synthesis problems at a cost of computational overload. Authors in [29] propose a faster yet efficient SRR method using anchored neighborhood regression approach by integrating the notion of NE with SC. Dictionary is learned by sparse model and the dictionary atoms are grouped into neighborhoods based on the correlation among the atoms. The SC model is modified as a least-square regression problem. Estimation of HR patch is achieved by performing mapping of projection matrix of its nearest neighborhood atoms in the dictionary. Eud is used to measure the distance between anchored atoms and trained samples. In [30], an improved ODL approach is used as the dictionary learning purpose. Particle swarm optimization (PSO) is used in the dictionary training stage to provide the information about atom updating and direction of optimization process. This helps to speed up the reconstruction rate. In [31], authors utilize Gabor filtered feature vectors to represent each image patch. In [32], an evolutionary SC based method is proposed. This method utilizes the adaptive Genetic algorithm as the SC model to obtain α^* of each LR patch. This evolutionary SC model speeds up the reconstruction process than by using the conventional l_0- or l_1-based SC models. In [33], a super-pixel based sparse representation model is proposed for hyper-spectral (HS) image SRR. This method constructs two dictionaries such as a spectral dictionary from HS image and a transformed dictionary from multispectral image. The generalized simultaneous OMP (G-

Table 2 Comparative analysis of NE-SISRR methods

Methods (Year)	Generation of training dataset	Feature extraction operation	Searching process and size of K	Learning model	Observations
[16] (2004)	Images from Kodak web-site	First + Second-order gradient	KNN, $K = 5$	LLE	High frequency details are not well preserved
[17] (2006)	Histogram-based matching for images in FERET database	Image residuals	KNN, $K = 5$	LLE	Preserves sharp contrast, edges. Block effects predominant in homogeneous region of image
[18] (2009)	Discriminate training patches using edge information	First-order gradient + NL	KNN, ($K = 2$ and 5)	LLE	Preserves more image details. Adaptive selection of size K produces less blocky effect
[19] (2011)	Same images used in method [18]	HoG	K-means clustering	LLE	Enhanced output at additional computational complexity
[20] (2012)	50 images from MIT database	DCT	K-means clustering	LLE	Enhanced image quality and better processing speed
[21] (2012)	Subsets of training images using HoG-based clustering	HoG + First + Second-order gradient	variants of Robust-SL0 algorithm	LLE + Sparse	Accelerates speed of SRR. Does not perform well for images rich in textures
[22] (2014)	Clustering based using FoE features	FoE features	KNN	LLE + Non-local self similarity	Well preserves the edges and textures. Computationally heavy in measuring EuD among FoE features
[23] (2016)	Histogram based	Gradients along eight orientations	KNN Global size of K	LLE+ Robust PCA	Produces enhanced output with minimized outliers at a cost of computational overload
[24] (2017)	Randomly selected standard images	ZM + NL	KNN, K =1-15	LLE	Produces blurry output for images rich in texture and geometry

SOMP+) is used to solve the sparse representation. Fractional abundance coefficient matrix is used for the reconstruction of final HS HR image. In [34], authors propose multiple dictionary pairs based on different structural contents of image via improved fuzzy C-means (IFCM) clustering method. A weighted SC model is used to obtain α^* of each LR patch. Method proposed in [35] enables a faster searching of candidate patches by utilizing KNN-based dictionary pair for each patch rather than the unified dictionary pair. Binary encoding scheme is used to accelerate the KNN retrieval process. Moreover, the proposed method uses large sized patches which encompass more information and hence preserves more image details in the reconstruction process. The proposed method enhances the processing speed as well as

reconstruction quality. A brief discussion about some of the significant state-of-the-art SC-SISRR methods is provided in Table 3. Some of the parameter and abbreviations used in this table are given in.[1]

2.4 Deep Learning-Based SRR

In recent years, deep learning-based SRR of image [36–39] has made excellent breakthroughs in providing high-quality reconstruction. The reconstruction process preforms end to

---

[1] α^*, optimal sparse coefficient; LASSO, least absolute shrinkage and selection operator, NCC, normalized cross-correlation, G-SOMP+, generalized simultaneous OMP, HSI, hyper-spectral image, IFCM, improved fuzzy C-means.

Table 3 Comparative analysis of SC-SISRR methods

Methods (Year)	Type of dictionary	Dictionary learning algorithm	Feature extraction operation	Searching of dictionary for α^*	Sparse models	Observations
[25] (2010)	Joint dictionary	Probabilistic model	First + second-order derivative	LASSO	SC model + K-mean clustering	Provides enhanced output than the method in [16] at a computational overload
[26] (2011)	Dual dictionary	K-SVD + Pseudo-inverse	HPF + PCA	OMP	Conventional SC using l_2 norm	Sharper result. Lesser artifacts, much faster than the method in [25]
[27] (2012)	Coupled dictionary	Supervised descent method	Gradient features	LASSO	Conventional SC using conjugate gradient descent	Outperforms joint dictionary training methods
[28] (2012)	Semi-coupled dictionary	Minimization of energy function	Patch intensity	LASSO	Conventional SC + Clustering + non-local similarity	Effective method in cross-style image synthesis
[29] (2014)	Neighborhood of dictionary	K-SVD	First + second order	NN	NE+Sparse model using	Faster searching of dictionary
[30] (2015)	Over-complete dictionary	ODL using PSO	Texture feature	OMP	Conventional SC using l_2 norm	Provides improved result for remote sensing images
[31] (2015)	Dual dictionary	K-SVD features	Gabor	NCC	Conventional SC using l_1 norm	Preserves more image details than the method in [25]
[32] (2016)	Over-complete dictionary	L_1-regularized least square	Pixel intensity	Adaptive GA	Evolutionary SC model	Reduces dictionary searching complexity
[33] (2017)	Spectral dictionary	ODL	Spectral signature	G-SOMP+	Conventional SC model	Superior result for HSI SRR
[34] (2018)	Multiple dictionary pairs	IFCM clustering methods	Pixel intensity	l_0 minimization	Weighted SC model	Provides improved results for infrared images at a higher computational cost
[35] (2019)	KNN-based dictionary	KNN	Gray values of patch	Binary encoding method	Conventional SC model	Provides better results in flat areas. Dictionary training process is time-consuming

end mapping of a LR image into a HR image either via convolutional neural network (CNN) model [36,37] or by Generative-Adversarial-Nets (GAN) model [38] or by deep shallow convolutional network [39]. Generally, the state-of-the-art deep learning-based methodologies diverge from each other in terms of: varieties of network architectures, variants of loss functions and most importantly different types of learning strategies. Wang et al. [40] provide an excel-

lent survey in the context of SRR via deep leaning models. Irrespective of the great accomplishment of these methods in terms of reconstruction accuracy with reduced computational burden, they suffer from some critical bottlenecks which should be solved carefully. To design an optimal network which will efficiently combine local as well as global along with low as well as high-level information at a reduced spatial and temporal cost is still an open problem [40]. The next difficult aspect of implementing these methods is selecting an appropriate loss function which will guide to optimize the model parameters. In [38], Ledig et al. have utilized cross-entropy based adversarial loss in the reconstruction process to provide an enhanced output. However, training the GAN architecture is quite difficult and sometimes unstable in nature [40].

3 Proposed Method

In this section, we will discuss about the proposed NE-SISRR method known as efficient neighbor embedding-based SRR (ENESR) method. The proposed work is based on the idea of the work in method [16] with several important modifications. Both of these methods are based on the concept of LLE which aims at estimating the HR image by linearly combining the neighboring candidate HR patches. Method in [16] utilizes first- and second-order gradient to represent each patch and employs fixed neighborhood size K for the reconstruction process. Consequently, the estimated output cannot well preserve the high frequency details in the reconstruction process. Sometimes, produces too much blurry output solutions. Moreover, this method exploits distance function based Eud to measure patch similarity. Hence, searching of candidate patches from the pool of training image patches consumes more time. On the contrary, proposed work fuses a number of feature attributes naturally to represent each image patch. The fusion process helps to preserve sufficient structural, spatial and statistical correlation among image patches. An automotive scheme is proposed to choose adaptive neighborhood size K for each image patch. Aiding to that, distance functions based on the notion of generalized eigenvalue which lies in computationally cheap non-Euclidean space is exploited to measure patch similarity. Detailed description on the contributions and novelty of the proposed work are discussed as follows:

- *Proposition of a low-dimensional feature mapping scheme to represent image patch in its feature space.*

  1. To the best of our knowledge, till now the LBSISRR methods either use edge information or combination of edge and smoothness information to represent the image patch. Consequently, reconstruction solutions well preserve the high frequency information in the reconstruction process, but provides limited performance for images rich in structural and statistical content. This necessitates selecting an efficient feature extraction operator which will preserve not only the high frequency information but also the spatial and structural correlation in both LR and HR space.

  2. To accomplish this, a low-dimensional feature mapping scheme is proposed which enables compact yet informative way of representing a patch in its feature space. Each patch is represented by a set of discriminative feature attributes. Correlation among these feature attributes is modeled via global covariance matrix. The matrix enables to fuse multiple correlated features naturally. This fusion process helps to preserve sufficient structural, spatial and statistical correlation among each other. Hence, allows faithful reconstruction of HR patch.

  3. Moreover, unlike individual raw feature the unified feature attributes are insensitive to noise; invariant to scale and illumination variations, consequently provides robust performance.

- *Development of a faster searching scheme to get potential candidate patches by measuring the similarity in non-Euclidean space*

  1. In the searching of candidate neighbors, unexceptionally, almost all methods use the Eud or its variants to measure the patch similarity. Eud-based measure utilizes intensity profile of the patch for its computation hence there is no provision to preserve the spatial correlation among pixels. Nevertheless, this measure is much sensitive to noise and suffers from curse of dimensionality.

  2. In this work, the patch similarity is measured in non-Euclidean space. Each patch is modeled as a global co-variance matrix which lie in a Riemannian manifold [41]. This matrix is symmetric and positive definite in nature. Most importantly, the diagonal and off-diagonal elements of this matrix possess variance and correlation among feature attributes.

  3. Inter-variance gap and inter-correlation map among feature vectors are measured by distance functions based on generalized eigenvalues which surely do not lie on Euclidean space, thus enables a faster searching process.

- *Implication of an automotive scheme to choose the neighborhood size of K based on the input patch characteristic*

  1. In the literature, various strategies such as K-mean clustering [19–21] or by distance measure [23] have been suggested to select the global size of K in the NE-SISRR methods. However, all these strategies aid

additional computational load to the reconstruction process. However, our proposed work uses a cheap strategy to do that.

2. Similarity values of the input patch with its candidate HR patches are sorted in descending order. Sorted similarity values are plotted with respect to the patch index. Next, gradient of this similarity curve is plotted. The global minima point of the gradient curve is considered as the threshold value ς. Training image patch index whose gradient values are lesser than equal to the threshold value ς are considered as the K number of neighboring patches

Processing time of any LBSISRR approach is critically influenced by effective searching of candidate patches from the pool of training patches. Selection of K number of candidate patches is performed by matching the similarity among patches. Each image patch is represented in its feature space. State-of-the-art LBSISRR methods utilize several raw image statistics such as gradient, pixel intensity, filter responses, DCT, HoG and FoE features to represent each image patch. Let us consider the popularly used feature extraction operator such as first- and second-order gradients along vertical and horizontal direction of an image patch of size $(z \times z)$. Now, each patch is represented by $(4z^2 \times 1)$ length of feature vector. Patch similarity is determined by measuring similarity among two $(4z^2 \times 1)$ feature vectors using Eud-based functions. Dimensionality of the feature representation of an image patch depends on the patch size selected. Consequently, complexity of the searching process increases with increase in patch size. On the other hand, proposed low-dimensional feature mapping scheme enables to fuse d dimension of raw feature attributes to a covariance matrix of size $(d \times d)$. The fusion process preserves sufficient image details in the reconstruction process. Instead of working on the high-dimensional feature vectors of image patches, proposed work employs distance functions using eigenvalues of $(d \times d)$ matrix to measure the patch similarity. Distance functions based on generalized eigenvalues lie in non-Euclidean space and hence converge faster that the Eud-based functions. Hence, through the proposed low-dimensional feature mapping scheme, we can effectively reduce the computational time of the searching process while enhancing the SR quality of the reconstructed solution.

Y_L is the input LR image to be super-resolved to obtain the final HR image $\hat{X}_H$. X_H is the ground truth HR image. Y_L image is up-sampled by P-spline interpolation scheme [42] by a factor of u to obtain Y_U so as to make its size equal to the size of final HR image $\hat{X}_H$. $\left\{ y_U^j \right\}_{j=1}^{N}$ are the N number of overlapped patches of Y_U image and $\left\{ \hat{x}_H^j \right\}_{j=1}^{N}$ are the corresponding N number of HR patches to be estimated.

$TR_{optimal}$ is the optimal training image dataset consisting of HR images H_{TRbest} and LR images L_{TRbest}. $\left\{ l_{TRbest}^p \right\}_{p=1}^{M}$ and $\left\{ h_{TRbest}^p \right\}_{p=1}^{M}$ are M number of overlapped LR and HR training image patches, respectively. $\left\{ NE(y_L^j) \right\}_{j=1}^{K}$ is K number of neighboring LR training patches of the input LR patch (y_L^j) and $\left\{ NE(x_H^j) \right\}_{j=1}^{K}$ are their corresponding HR patches, respectively. Equation (1) represents the mathematical formulation of the imaging model which is used to generate the LR counterpart image Y_L of the HR image X_H.

$$Y_L = De\,Br\,Wa\,X_H + \eta \qquad (1)$$

where De is the decimation matrix, Br is the blurring operation, Wa performs the warping operation to generate the LR image from its corresponding HR image. η is the amount of AWGN which corrupts the imaging model. Terminologies and notations utilized in this paper are described in Table 4 and pseudo-code of the proposed ENESR method is depicted in Algorithm 1.

Algorithm 1 Pseudo-code of the proposed ENESR method

Preprocessing: $Y_L, u, TR_{optimal}, H_{TRbest}, L_{TRbest}$
   $TR_{optimal}$ is selected as given in section 3.1

Preprocessing: **(i)** Interpolate Y_L by factor u to yield Y_U
   (ii) Divide Y_U to N number of overlapped patches $\left\{ y_U^j \right\}_{j=1}^{N}$ of size $z \times z$ each
   (iii) Divide H_{TRbest} images to M number of overlapped HR patches $\left\{ h_{TRbest}^p \right\}_{p=1}^{M}$ of size $z \times z$ each
   (iv) Represent each overlapped patch in low dimensional feature space as discussed in section 3.2

Output: $\hat{X}_H$

   Begin algorithm

1: **for** each test image patch y_U^j in $\left\{ y_U^j \right\}_{j=1}^{N}$ **do**

2:    Find K nearest neighbors NE_K from $\left\{ h_{TRbest}^p \right\}_{p=1}^{M}$ by measuring the patch similarity in the non-Euclidean space as described in section 3.3 and utilize an adaptive scheme to select the size of K as given in section 3.4

3:    Compute the reconstruction weights W_K of the neighboring patches as described in section 3.5

4:    Embed the reconstruction weights contributed by the NE_K HR patches to estimate the HR patch $\hat{x}_H^j$ as given in Eq. (9)

5: **end for**

6: $\hat{X}_H$ is estimated by stitching the estimated HR patches $\left\{ \hat{x}_H^j \right\}_{j=1}^{N}$ by preserving the local compatibility and smoothness priors among patches. The overlapped region of patches are averaged in the stitching process.

Table 4 Notations and terminologies used

Symbol	Interpretation	Symbol	Interpretation
De	Decimation matrix	H_{TRbest}	HR images $\in TR_{optimal}$
Br	Blurring matrix	L_{TRbest}	LR images $\in TR_{optimal}$
Wa	Warping matrix	$\{h_{TRbest}^p\}_{p=1}^M$	M number of overlapped patches of H_{TRbest} images
η	Gaussian noise	$\{l_{TRbest}^p\}_{p=1}^M$	M number of overlapped patches of L_{TRbest} images
Y_L	Input LR image	$\{NE(y_L^j)\}_{j=1}^K$	K number of neighborhood LR patches for y_L^j
X_H	Ground truth HR image	$\{NE(x_H^j)\}_{j=1}^K$	K number of neighborhood HR patches for y_L^j
Y_U	Up-sampled LR image	u	Up-sampling factor
$\hat{X}_H$	Estimated HR image	$\rho_1(.)$ and $\rho_2(.)$	Distance functions
N, M	Number of overlapped patches	No	Normalization factor
$\{\hat{x}_H^j\}_{j=1}^N$	N number of estimated HR patches	Sim	Similarity index between two patches
$\{y_L^j\}_{j=1}^N$	N number of overlapped patches of Y_L	ς	Threshold value
$TR_{optimal}$	Optimal training image dataset	A_i	Feature attribute vector
λ_i	Eigenvalues of matrix	$C_p(x, y)$	Global covariance matrix

3.1 Selection of $TR_{optimal}$

Proposed work utilizes the methodology used in [15] for the selection of optimal training image dataset $TR_{optimal}$. Selection of $TR_{optimal}$ via spatiogram-based matching criterion preserves sufficient statistical and structural similarity with the input test image.

3.2 Selection of Features

This subsection describes the representation of each image patch as a compact set of feature attributes. Y_U and H_{TRbest} $\in TR_{optimal}$ are divided into patches of size $(z \times z)$ with 2 pixel overlapping. Each patch is represented by a set of preferably compact and discriminative feature attributes. The raw feature attributes exhibit some correlation among each other. This correlation among feature vectors is modeled by a global covariance matrix. The covariance matrix enables to fuse multiple correlated features naturally [41]. Nevertheless, it is robust to noise and invariant to variations of scale, rotation and illumination level [41]. Correlation between two patches is based on measuring the similarity between their corresponding global covariance matrices. The global covariance matrices lie on non-Euclidean space [41]. Hence, the notion of a generalized eigenvalue is considered for measuring the similarity which enables a faster matching process.

Let a patch p of size $(z \times z)$ is represented by a d dimensional attribute vector $\{A_i\}_{i=1,2...n}$. Equation (2) represents the formulation of its corresponding $(d \times d)$ covariance

matrix $C_p(x, y)$. $(x, y)^{th}$ component of the C_p matrix is described in Eq. (3) [41].

$$C_p = \frac{1}{n-1} \sum_{I=1}^n (A_i - \mu)(A_i - \mu)^T \tag{2}$$

$$
C_p(x, y) = \frac{1}{n-1} \left[\sum_{i=1}^n A_i(x) A_i(y) - \frac{1}{n} \sum_{i=1}^n A_i(x) \sum_{i=1}^n A_i(y) \right]
\tag{3}
$$

where μ is the mean of the feature attribute vector A_i.

In this work, we integrate 9 different feature vectors in the formation of vector A_i and is represented in Eq. (4).

$$\{A\} = [x, y, |I_p|, G_I, S_j]^T \quad I = SC \times O; \; j = 1, 2 \tag{4}$$

where (x, y) represents the pixel location, $|I_p|$ is the absolute intensity profile of the patch p, G_I are the multi-scale and multi-orientation Gabor features and S_j are the first-order and second-order statistical information of the patch p. SC and O are the total number of scales and orientations of the Gabor filter, respectively.

The pixel locations (x, y) in the vector A_i are correlated to other features which contribute toward the non-diagonal entries of the covariance matrix. Gabor features are used because they capture geometrical as well as textural information effectively. The low-frequency Gabor filters often

behave as a local image smoother and represent image patch at coarse resolution. The high-frequency Gabor filters often act as efficient edge detectors irrespective of the nature of underlying intensity distributions [43]. Unlike conventional edge detectors, the multi-oriented Gabor features easily differentiate edges in horizontal as well as vertical directions [43]. However, computational cost for the computation of Gabor features is high. To reduce the computational cost, the approximate method provided in [44] are adopted in this work. The first- and second-order statistical information such as mean and co-variance measure of the patch are embedded to make the vector A_i adequate to characterize image patch efficiently. With these 9 raw feature vectors, each patch p is represented by a (9×9) covariance matrix $C_p(x, y)$.

Some basic properties of the covariance matrix $C_p(x, y)$ are presented as follows:

- it is symmetric and positive definite which lie in Riemannian manifold
- unlike individual raw features, it is robust to noise, insensitive to change in scale, rotation and illumination
- it enables to integrate multiple correlated features
- diagonal elements exhibit the variance of the raw feature vectors
- off-diagonal elements of the matrix hold the correlation among feature vectors

3.3 Selection of Candidate Patches

In this section, candidate HR patches for each input LR patch are searched over the image patches of $H_{TRbest} \in TR_{optimal}$. Given two patches p, q with their corresponding global covariance matrices $C_p(x, y)$, $C_q(x, y)$, measurement of similarity between two patches is done w.r.t their covariance matrices. Similarity between two patches will be maximum when inter-variance gap of feature attributes will be minimum and the inter-correlation of feature attributes will be maximum. Measurement of both inter-variance gap and inter-correlation map is guided by specified distance function and minimization of distance measure signifies the maximization of the similarity between two patches. As the covariance matrices do not lie in Euclidean space, distance function is based on generalized eigenvalues which follow the Lie group structure of positive definite matrices [41]. Equation (5) defines the formulation of similarity between patches p and q using their covariance matrices $C_p(x, y)$ and $C_q(x, y)$, respectively.

$$
\begin{aligned}
Sim &\left\{ C_p(x, y), C_q(x, y) \right\} \\
&= \rho_1 \left\{ C_{pD}(x, y), C_{qD}(x, y) \right\} \\
&\quad + \rho_2 \left\{ C_{pND}(x, y), C_{qND}(x, y) \right\}
\end{aligned}
\tag{5}
$$

where $\rho_1(\cdot)$, $\rho_2(\cdot)$ are the distance functions to measure the inter-variance gap and inter-correlation map, respectively. Equations (6) and (7) provide the formulation of $\rho_1(\cdot)$, $\rho_2(\cdot)$, respectively.

$$
\begin{aligned}
\rho_1 &\left\{ C_{pD}(x, y), C_{qD}(x, y) \right\} \\
&= sqrt \left[\sum_{i=1}^{n} \ln^2 \lambda_i \left(C_{pD}, C_{qD} \right) \right]
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
\rho_2 &\left\{ C_{pND}(x, y), C_{qND}(x, y) \right\} \\
&= \left\| \log \left(C_{pND}^{-\frac{1}{2}} C_{qND} C_{pND}^{-\frac{1}{2}} \right) \right\|_F
\end{aligned}
\tag{7}
$$

where $C_{pD}(x, y)$, $C_{qD}(x, y)$ are the diagonal matrices of $C_p(x, y)$, $C_q(x, y)$, respectively. $C_{pND}(x, y)$, $C_{qND}(x, y)$ are same as the $C_p(x, y)$, $C_q(x, y)$ matrices by keeping all the off-diagonal entries unchanged and replacing all the diagonal entries by one. $\lambda_i \left(C_{pD}, C_{qD} \right) = \lambda_1, \lambda_2 \ldots \lambda_n$ are joint eigenvalues of the matrices of C_{pD}, C_{qD} by solving the equation $\det(\lambda C_{pD} - C_{qD}) = 0$. $\| \cdot \|_F$ is the Frobenius norm. Frobenius norm of a $m \times n$ matrix A is defined as the square root of the sum of the absolute squares of its elements. $\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$.

3.4 Selection of Size K

Size of K is adaptively chosen according to the patch similarity values. After measuring the patch similarity (Sim) of the input patch y_L^j with respect to the training image patches $\{h_{TRbest}^p\}_{p=1}^M$ by using the formulation given in Eq. (5), the Sim values of patches are sorted in descending order. Sorted similarity values are plotted with respect to the patch index. Next, gradient of this similarity curve is plotted. The global minima point of the gradient curve is considered as the threshold value ς. Training image patch index whose gradient values are lesser than equal to the threshold value ς is considered as the K number of neighboring patches for y_L^j patch. Figure 2 shows the graphical representation of the selection of size K of an arbitrary patch y_L^j.

First row shows the plot of sorted Sim values against the training image patch index. Second row shows the gradient of the above curve. Global minima of the curve is shown in circle and ς value is marked by the arrow mark. Here, the ς is lying at the 10th number training image patch index and hence the size of K is 10. Top most 10 training image patches are serving as the neighborhood candidate patches for the input patch y_L^j. Likewise, size of K is different for each LR patch and it depends on the patch similarity of that particular patch with the training image patches $\{h_{TRbest}^p\}_{p=1}^M$. Consequently, neighborhood size K for the reconstruction process is adaptive in nature and diminishes the effect of over and under fitting of data. The overall steps for selecting the opti-
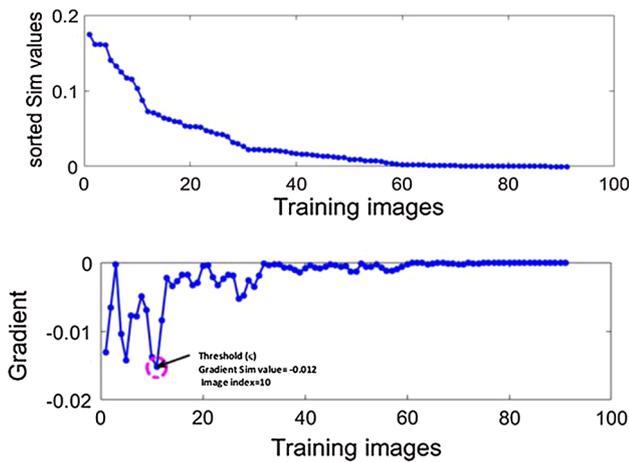
Fig. 2 Selection of size K

mal candidate patches for each input LR patch are described in Algorithm 2.

Algorithm 2 discusses about the steps to select K number of candidate HR patches for each input patch. Input to this searching process is N number of input patches $\left\{y_U^j\right\}_{j=1}^N$ and M number of training HR patches $\left\{h_{TRbest}^p\right\}_{p=1}^M$. Searching process aims at selecting K number of HR training patches $\left\{h_{TRbest}^i\right\}_{i=1}^K$ from M number of training HR patches in $\left\{h_{TRbest}^p\right\}_{p=1}^M$ for each input patch y_U^j in $\left\{y_U^j\right\}_{j=1}^N$. y_U^j is the jth patch of up-sampled image Y_U. Y_U is obtained by up-sampling the input LR image Y_L by a factor of u using P-spline interpolation [42]. Image Y_U is divided into N number of patches $\left\{y_U^j\right\}_{j=1}^N$ with the same size as that of the each h_{TRbest}^p patch. Now, feature vector A_i and its corresponding co-variance matrix for each input patch y_{UR}^j and each training HR image patches $\left\{h_{TRbest}^p\right\}_{p=1}^M$ are computed using Eqs. (3) and (4), respectively. Correlation among the co-variance matrix of input patch y_U^j and M number of co-variance matrix of training patches $\left\{h_{TRbest}^p\right\}_{p=1}^M$ are measured using Eq. (5). Now we are available with M number of correlated values which are termed as Sim values. By using the procedure as discussed in Sect. 3.4, K number of candidate training patches $\left\{h_{TRbest}^i\right\}_{i=1}^K$ for the input patch y_U^j are selected. This process is continued till each patch of $\left\{y_U^j\right\}_{j=1}^N$ selects its corresponding K number of candidate HR patches. Afterward, reconstruction weights are computed for the linear embedding process to estimate the final HR image.

Algorithm 2 : Selection of neighborhood patches

Input: Y_L, magnification factor u and $\left\{h_{TRbest}^p\right\}_{p=1}^M$

Output: $\left\{NE(Y_U^j)\right\}_{j=1}^K$

   Begin algorithm

1: Interpolate Y_L by factor u to yield Y_U

2: Divide Y_U into N number of patches $\left\{y_U^j\right\}_{j=1}^N$ with the same size as that of the each h_{TRbest}^p patch

3: Compute the vector A_i for j^{th} input patch y_U^j using Eq. (4).

4: Compute the covariance matrix $C_{\left(y_U^j\right)}$ using Eq. (3)

5: **for** each training HR image patches $h_{TRbest}^p \in \left\{h_{TRbest}^p\right\}_{p=1}^M$ **do**

6:    Compute the vector A_i

7:    Compute its corresponding covariance matrix $C_{\left(h_{TRbest}^p\right)}$

8: **end for**

9: Store M number of covariance matrices $\left\{C_{\left(h_{TRbest}^p\right)}^i\right\}_{i=1}^M$ for the training HR image patches

10: Compute the similarity values Sim among $\left\{C_{\left(y_U^j\right)}\right\}$ and $\left\{C_{\left(h_{TRbest}^p\right)}^i\right\}_{i=1}^M$ using Eq. (5)

11: Store and sort the values of Sim in descending order

12: Plot the values of sorted Sim values with respect to the training patch index

13: Global minimum of the gradient curve plotted is treated as the ς value.

14: Select K number of patches $\left\{h_{TRbest}^i\right\}_{i=1}^K \in \left\{h_{TRbest}^p\right\}_{p=1}^M$ having the gradient of Sim value $\leq \varsigma$ as candidate neighboring patches for the jth input patch y_U^j.

15: $\left\{NE(Y_U^j)\right\}_{j=1}^K = \left\{h_{TRbest}^i\right\}_{i=1}^K$

16: Return $\left\{NE(Y_U^j)\right\}_{j=1}^K$

   End algorithm

3.5 Computation of Reconstruction Weight and Estimation of HR Image $\left\{\hat{x}_H^j\right\}_{j=1}^N$

Optimal reconstruction weights W_K of the neighboring patches are computed by minimizing the local reconstruction error as performed in [16,18]. For each y_U^j, K nearest neighbor patches $\left\{NE(Y_U^j)\right\}_{j=1}^K$ are searched. W_K is achieved by minimizing the error for y_U^j as given in Eq. (8).
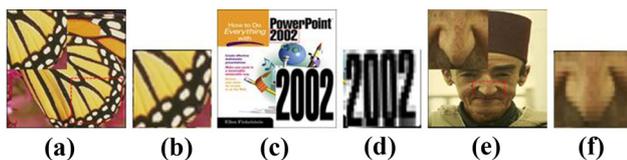
(a)   **(b)**   **(c)**   **(d)**   **(e)**   **(f)**

Fig. 3 **a**, **b** HR and LR images of 'butterfly.bmp,' **c**, **d** HR and LR images of 'ppt3.bmp,' **e**, **f** HR and LR images of '189080.jpg'

(a)   **(b)**   **(c)**   **(d)**
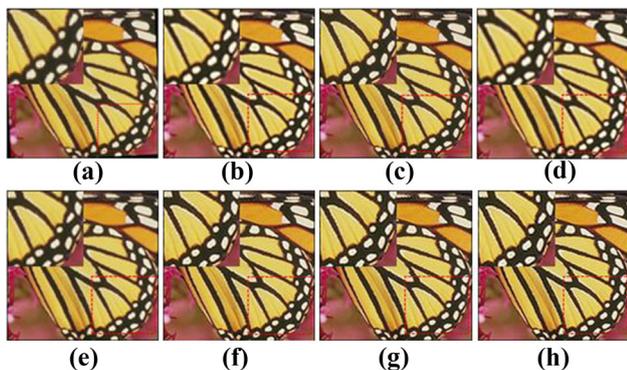
(e)   **(f)**   **(g)**   **(h)**

Fig. 4 Reconstructed HR solutions of 'butterfly.bmp' image using different methods at a magnification factor of $3\times$. **a** Gao et al. [21], **b** Zhu et al. [22], **c** Mishra et al. [23], **d** Mishra et al. [24], **e** Nayak et al. [15], **f** Liu et al. [35], **g** Ledig et al. [38], **h** ENESR (proposed)

$$\varepsilon r^{(j)} = \left\| Y_U^j - \sum_{Y_U^P \in NE_j} W_{jp} Y_U^P \right\|^2 \qquad (8)$$

subject to constraint $\sum_{Y_U^P \in NE_j} W_{jp} = 1$ and $W_{jp} = 0$ for any patch $Y_U^P \notin NE_j$. NE_j is the nearest neighbor patches of y_U^j. Estimation of $\hat{x}_H^j$ is done by following the formulation as described in Eq. (9).

$$\hat{x}_H^j = \sum_{Y_U^P \in NE_j} W_{jp} \hat{x}_U^P. \qquad (9)$$

$\hat{X}_H$ is estimated by stitching the estimated HR patches $\left\{\hat{x}_H^j\right\}_{j=1}^{N}$ by preserving the local compatibility and smoothness priors among patches. The overlapped region of patches are averaged in the stitching process to preserve better inter-patch compatibility.

4 Simulations and Result Analysis

Effectiveness of the proposed method in providing an efficient trade-off between reconstruction quality and the processing speed is validated through simulation data experiments for both synthetic and real-time images. Performance evaluation of the proposed work is compared with some of the
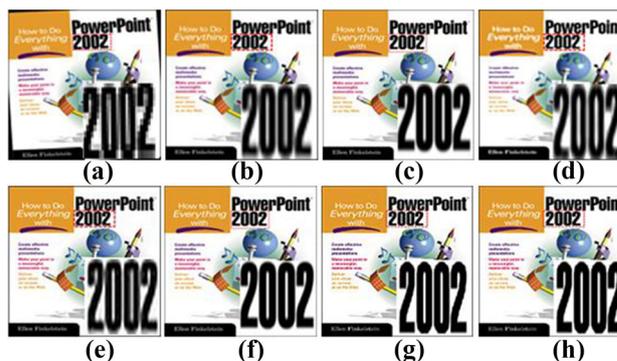

(a)   **(b)**   **(c)**   **(d)**

(e)   **(f)**   **(g)**   **(h)**

Fig. 5 Reconstructed HR images for 'ppt3.bmp' image using different reconstruction methods at $3\times$ zooming factor. **a** Gao et al. [21], **b** Zhu et al. [22], **c** Mishra et al. [23], **d** Mishra et al. [24], **e** Nayak et al. [15], **f** Liu et al. [35], **g** Ledig et al. [38], **h** ENESR (proposed)

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

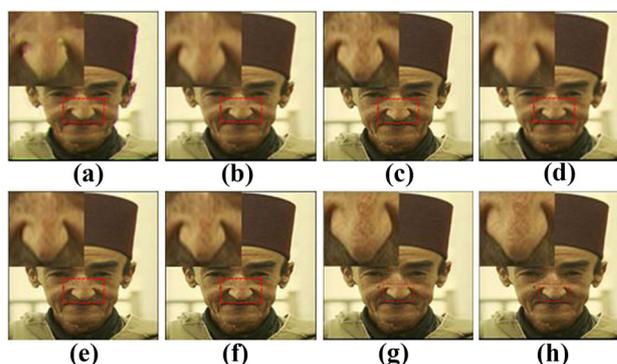**(e)**   **(f)**   **(g)**   **(h)**

Fig. 6 Reconstructed HR solutions of '189080.jpg' image using different methods at a zooming factor of $3\times$. **a** Gao et al. [21], **b** Zhu et al. [22], **c** Mishra et al. [23], **d** Mishra et al. [24], **e** Nayak et al. [15], **f** Liu et al. [35], **g** Ledig et al. [38], **h** ENESR (proposed)

existing state-of-the-art LBSISRR methodologies [15,21–24,35] along with the deep learning-based method proposed in [38]. Along with visual perception, several subjective image quality measures such as PSNR, SSIM, FSIM, ringing measure (RM) and blur measure (BM) [45] are used to quantify the reconstruction quality for synthetic data experiments. To have a fair comparison, several objective quality measures such as Spearman rank order correlation coefficient (SRCC), Pearson linear correlation coefficient (PLCC) and root mean square error (RMSE) are also computed to quantify the reconstruction quality of the estimated HR images. We have used the methodology used in [46] for the computation of these parameters. Reconstruction process is carried out for $2\times$, $3\times$ and $4\times$ magnification factors. For synthetic-data experiments, images used in [15,25] are used for training purpose and images from 'Set5' [47], 'Set14' [26] and 'BSD200' [48] are utilized for testing purpose. For real-time experiments, images from MDSP [49] dataset are used. Each of the images in the MDSP dataset contains a sequence of image frames with some global and local motion with each other.

For the simulation purpose, one of the image sequence is used as the testing image while rest of the frames are utilized for training purpose. The blind image quality index (BIQI) [50] and natural image quality evaluator (NIQE) [51] indices are used to quantify the reconstruction accuracy for real-time data experiments. Higher value signifies better reconstruction quality.

4.1 Experimental Setting

In the reconstruction process, $TR_{optimal}$ is generated as done in [15]. Size of K is chosen adaptively as described in Sect. 3.4. The degradation process as defined in Eq. (1) utilizes the blurring matrix Br as (5×5) Gaussian LPF with smoothing parameter 0.1. Warping matrix Wa executes $[-1$ to $1]$ translation along both x and y directions along with 2 degrees of rotation. Down-sampling operation is performed by an amount of 2 to 4. Noise η is added to make the SNR as 20 dB. In the real-time experiments, the degradation process ignores the warping operation as the image sequences comprise of local as well as global shift among each other. Size of LR patch $z \times z$ is taken as (5×5) with two pixels overlapping. For extracting the Gabor features of patches, scale $SC = 1$ and orientation $O = 4$, i.e., $O = \left\{ \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi \right\}$ are used. Parameters of the compared LBSISRR methods are chosen in the similar fashion as provided in their original work.

4.2 Performance Analysis of Synthetic Data Experiments

In this section, we have discussed the performance analysis of the proposed work with some of the existing state-of-the-art methods by performing experimentation on synthetic test images from the 'Set5,' 'Set14' and 'BSD200' datasets at different magnification factors. For objective quality analysis, reconstructed HR solutions of some test images from these datasets are also provided.

Figure 3 depicts the original HR and their corresponding LR versions of 'butterfly.bmp,' 'ppt3.bmp' and '189080.jpg' images from the 'Set5,' 'Set14' and 'BSD200' datasets, respectively.

Reconstructed HR solutions of these there test images at $3\times$ magnification factor are presented in Figs. 4, 5 and 6. Corresponding subjective quality measures of the reconstructed images are given in Table 5.

From the visual analysis as perceived from images in Figs. 4, 5 and 6g and tabular data analysis, it is observed that the deep learning-based SRR method in [38] provides superior performance. However, here our soul aim is to propose a faster and efficient LBSISRR method as compared to the existing ones. Hence, we will analyze the effectiveness of our proposed work with respect to the compared LBSISRR methods. The reconstructed HR images using the method in [21] well preserves the image details by utilizing the HOG feature. However, reconstruction quality of this method degrades for images rich in textural contents as observed in Fig. 6a for '189080.jpg' image. However, method in [22] utilizes FoE features and non-local based constraints, consequently produces enhanced output even for images rich in textural content. But, measuring patch similarity by means of EuD among feature vector makes the method time-consuming in nature. Likewise, method in [24] is also providing some pleasant output by utilizing ZM based features. This method well preserves the global structure of image and robust to noise but have limited reconstruction capability for large-scale reconstruction. Consequently, some unwanted artifacts dominant in the reconstructed solutions as observed in Fig. 5d for 'ppt3.bmp' image. As compared to the above discussed methods, method in [23] is quite robust to noise and outliers present. This method also suggests an automated scheme for selecting the global size K for the reconstruction process. The estimated solutions as shown in Figs. 4, 5 and 6c show the efficacy of this method in producing the enhanced output. However, utilization of geodesic and EuD makes the reconstruction process computationally expensive. Method in [15] proposes edge preserving compatibility functions to well preserve the compatibility among neighboring patches and hence produces better HR results than the methods in [21,22,24]. This can be observed from the images in Figs. 4, 5 and 6e. However, utilization of intensity profile to represent patch in its feature space makes this method sensitive to noise and outliers. Nevertheless, computational complexity of this method is too high due to epitomic-based neighborhood searching process. Reconstructed solutions as shown in Figs. 4, 5 and 6f via the method in [35] produce sharper HR images by utilizing SC-based learning and KNN-based dictionary instead of unified dictionaries. Consequently, this method provides enhanced image quality at a lesser computational overload. From visual perception of reconstructed HR images as shown in Figs. 4, 5 and 6h, it is clearly seen that the proposed method produces sharp image details with diminished effect of artifacts than the other methods. Improvement of reconstruction quality of the proposed work is due to the integration of several raw image features which in turn provides an unified framework to preserve the complete set of high frequency components in the reconstruction process. To provide a fair comparative analysis, average values of different subjective quality measures of the reconstructed solutions for three image datasets at different magnification factors are provided in Tables 6, 7 and 8. Quality measure values using the method in [38] are shown in boldfaces and the values by using the proposed ENESR method are shown in italic and boldfaces.

Table 5 Subjective quality measures of the reconstructed HR solutions for 'butterfly.bmp,' 'ppt3.bmp' and '189080.jpg' test images at zooming factor of 3×

Quality measures	Bi-cubic	[21]	[22]	[23]	[24]	[15]	[35]	[38]	ENESR (proposed)
'butterfly.bmp'									
PSNR (dB)	26.53	28.90	29.62	33.14	30.71	31.64	33.67	**39.30**	*34.52*
SSIM	.823	.844	.850	.870	.852	.867	.874	**.912**	*.880*
FSIM	.831	.865	.878	.913	.900	.908	.926	**.950**	*.933*
RM	8.77	6.42	5.43	2.01	4.80	3.22	2.14	**0.681**	*1.77*
BM	10.55	7.73	6.66	3.41	5.58	6.40	2.93	**1.07**	*2.26*
'ppt3.bmp'									
PSNR (dB)	26.44	28.34	29.23	32.24	30.61	31.85	33.48	**37.10**	*34.48*
SSIM	.665	.673	.701	.748	.712	.726	.758	**.837**	*.797*
FSIM	.676	.714	.734	.826	.784	.804	.836	**.894**	*.840*
RM	11.2	8.11	7.20	4.28	5.14	6.89	3.83	**1.54**	*2.34*
BM	12.6	9.14	8.34	5.24	6.70	6.10	4.64	**2.01**	*3.18*
'189080.jpg'									
PSNR (dB)	27.44	28.64	29.44	31.24	30.17	30.90	32.16	**36.10**	*33.14*
SSIM	.789	.801	.813	.836	.820	.826	.841	**.900**	*.853*
FSIM	.804	.823	.835	.874	.851	.868	.880	**.934**	*.896*
RM	15.6	11.1	10.3	6.95	7.22	7.14	4.01	**1.27**	*3.20*
BM	19.8	12.0	11.4	7.41	8.84	8.32	5.32	**1.93**	*4.04*

Table 6 Average values of subjective quality measures of the reconstructed HR images for 'Set5' dataset [47] at 2×, 3× and 4× zooming factors

Quality measures	Bi-cubic	[21]	[22]	[23]	[24]	[15]	[35]	[38]	ENESR (proposed)
'2 ×'									
PSNR (dB)	33.34	34.26	35.03	37.13	35.18	36.05	38.23	**40.40**	*39.11*
SSIM/FSIM	.830/.841	.873/.887	.886/.895	.913/.947	.892/.903	.902/.917	.957/.964	**.989/.990**	*.968/.972*
RM/BM	10.3/11.8	8.28/9.81	7.10/8.88	3.20/4.42	6.13/7.71	5.05/6.89	2.34/3.66	**1.03/1.68**	*1.46/2.02*
'3 ×'									
PSNR (dB)	30.36	31.18	32.42	33.07	32.63	32.81	33.80	**35.64**	*34.55*
SSIM/FSIM	.808/.826	.854/.867	.861/.870	.882/.903	.876/.890	.879/.892	.897/.921	**.921/.951**	*.907/.928*
RM/BM	14.7/16.5	9.36/10.5	8.72/9.94	4.47/5.35	7.75/9.04	5.66/8.80	2.89/4.65	**1.96/2.05**	*2.21/2.77*
'4 ×'									
PSNR (dB)	28.40	29.04	29.62	30.23	29.72	29.80	30.55	**32.06**	*31.11*
SSIM/FSIM	.779/.813	.791/.827	.801/.845	.823/.850	.813/.844	.816/.848	.839/.857	**.881/.902**	*.846/.860*
RM/BM	17.3/20.7	10.7/13.4	9.60/10.6	6.78/8.68	7.53/10.2	7.17/7.76	3.54/6.50	**2.78/4.06**	*3.07/5.18*

Tables 9 and 10 show the average values of the objective quality assessment parameters of the reconstructed HR solutions at magnification factors of 3× and 4×, respectively, for different dataset images. By analyzing the data present in Tables 6, 7, 8, 9 and 10, it is observed that, both the subjective and objective quality measures are consistent with the visual perception and our proposed ENESR method outperforms the compared methods. However, methods in [23,35] are providing comparable results with the ENESR method.

Tables 11 and 12 provide the improvement of different subjective and objective quality measures of the proposed

method compared to the two comparable methods in [23,35] at different magnification factors for the three test image datasets. From the above tabular data analysis, it is observed that, improvement in different quality measures of the proposed method decreases with increase in magnification factors. However, our proposed method provides remarkable improvement in image quality even at higher magnification factor which signifies the efficacy of the method in providing high-quality image details.

Table 13 provides the average processing time for dataset images, whereas Table 14 provides the average processing

Table 7 Average values of the subjective quality measures of reconstructed HR images for 'Set14' dataset [26] at zooming factors of 2×, 3× and 4×

Quality measures	Bi-cubic	[21]	[22]	[23]	[24]	[15]	[35]	[38]	ENESR (proposed)
'2 ×'									
PSNR (dB)	30.20	30.65	31.08	32.60	31.68	32.10	32.94	**35.55**	*33.74*
SSIM/FSIM	.868/.871	.887/.890	.891/.892	.935/.937	.902/.903	.911/.921	.940/.941	**.981/.982**	*.951/.952*
RM/BM	10.3/12.1	9.83/11.7	8.21/10.0	5.05/7.44	6.18/8.20	5.41/8.30	4.64/6.60	**2.49/4.41**	*3.34/5.16*
'3 ×'									
PSNR (dB)	27.53	28.07	28.51	30.73	28.83	29.14	31.13	**33.62**	*31.73*
SSIM/FSIM	.773/.776	.813/.814	.825/.827	.839/.845	.827/.828	.829/.830	.843/.844	**.861/.864**	*.852/.855*
RM/BM	13.3/16.9	9.60/10.3	8.86/9.81	6.15/7.23	7.32/8.42	6.54/7.60	5.20/6.31	**3.46/4.41**	*4.35/5.32*
'4 ×'									
PSNR (dB)	25.94	26.12	26.74	27.76	27.01	27.25	28.07	**29.15**	*28.64*
SSIM/FSIM	.701/.70	.714/.715	.720/.721	.758/.760	.730/.732	.736/.737	.761/.763	**.790/.802**	*.767/.768*
RM/BM	13.5/19.3	10.0/10.6	9.32/10.0	6.40/7.40	7.74/8.36	6.83/7.93	5.41/6.41	**4.42/4.67**	*5.02/5.49*

Table 8 Average values of subjective quality measures of reconstructed HR images for 'BSD200' dataset [48] at zooming factors of 2×, 3× and 4×, respectively

Quality measures	Bi-cubic	[21]	[22]	[23]	[24]	[15]	[35]	[38]	ENESR (proposed)
'2 ×'									
PSNR (dB)	27.36	28.14	28.56	29.63	28.65	29.06	30.14	**34.43**	*30.84*
SSIM/FSIM	.852/.863	.870/.886	.872/.887	.893/.907	.886/.895	.890/.900	.898/.910	**.922/.943**	*.914/.924*
RM/BM	14.4/18.6	8.84/13.4	8.32/12.2	6.24/8.03	7.56/8.91	6.65/9.44	5.78/7.05	**3.46/4.31**	*5.11/6.63*
'3 ×'									
PSNR (dB)	25.92	26.03	26.72	27.42	26.80	26.89	27.83	**30.62**	*28.46*
SSIM/FSIM	.746/.761	.769/.785	.772/.786	.788/.807	.776/.790	.784/.792	.793/.812	**.821/.845**	*.802/.821*
RM/BM	16.3/20.8	13.7/17.1	10.5/14.1	7.08/8.21	9.12/10.4	8.55/9.48	6.14/7.34	**3.80/5.84**	*5.84/7.00*
'4 ×'									
PSNR (dB)	24.63	25.01	25.32	26.04	25.66	25.83	26.34	**29.32**	*26.88*
SSIM/FSIM	.672/.721	.693/.743	.699/.746	.707/.760	.700/.751	.703/.756	.712/.764	**.762/.781**	*.718/.770*
RM/BM	20.6/23.6	17.2/20.2	14.8/17.1	9.06/10.0	12.1/14.2	11.3/12.1	8.12/9.20	**5.11/6.13**	*7.89/8.94*

Table 9 Average values of objective quality measures of the reconstructed HR images at a zooming factor of 3×

Quality measures	Bi-cubic	[21]	[22]	[23]	[24]	[15]	[35]	[38]	ENESR (proposed)
'Set5'									
SRCC	.3016	.4424	.4517	.5215	.4853	.5016	.6632	.8271	.7314
PLCC	.3624	.4943	.5023	.5516	.5226	.5314	.7010	.8642	.7619
RMSE	3.492	2.726	2.418	2.210	2.341	2.398	1.432	1.068	1.215
'Set14'									
SRCC	.2062	.3801	.3912	.4516	.4113	.4311	.5014	.6588	.5913
PLCC	.3491	.4817	.5241	.6218	.5373	.6013	.7228	.8323	.7896
RMSE	4.682	2.808	2.562	2.314	2.446	2.416	1.606	1.07	1.321
'BSD200'									
SRCC	.4432	.5064	.5234	.6019	.5617	.5777	.6324	.8218	.7011
PLCC	.4713	.5214	.5619	.6414	.5828	.5770	.6811	.8414	.7316
RMSE	5.412	3.441	3.212	3.110	3.281	3.216	2.321	1.108	2.01

Table 10 Average values of objective quality measures of the reconstructed HR images at a zooming factor of 4×

Quality measures	Bi-cubic	[21]	[22]	[23]	[24]	[15]	[35]	[38]	ENESR (proposed)
'Set5'									
SRCC	.2110	.4213	.4310	.5168	.4646	.4837	.6446	.8113	.7012
PLCC	.2624	.4717	.4823	.5347	.4918	.5026	.6865	.8411	.7213
RMSE	6.442	4.726	4.818	2.793	4.660	4.765	2.213	1.792	2.113
'Set14'									
SRCC	.1189	.3601	.3722	.4406	.4023	.4110	.4818	.6470	.5632
PLCC	.2007	.4606	.5003	.6117	.5115	.5613	.7003	.8117	.7414
RMSE	8.832	4.469	4.314	2.762	3.317	3.116	2.201	1.87	2.061
'BSD200'									
SRCC	.2175	.4416	.4537	.5891	.4817	.4945	.6210	.8116	.6800
PLCC	.2220	.4572	.4680	.6295	.5441	.5321	.6771	.8330	.7114
RMSE	8.882	5.642	6.013	3.600	4.413	3.967	3.01	1.63	2.85

Table 11 Average improvement of various subjective quality measures of proposed method over the methods in [23] and [35] at different magnification factors

Datasets	ΔPSNR (dB) [23]/[35]	ΔSSIM (%) [23]/[35]	ΔFSIM (%) [23]/[35]	ΔRM (%) [23]/[35]	ΔBM (%) [23]/[35]
2×					
'Set 5'	1.98/0.88	5.68/1.13	2.57/0.82	54.3/37.6	54.2/44.8
'Set 14'	1.14/0.8	1.68/1.16	1.157/1.5	33.86/28.01	30.65/21.81
'BSD200'	1.23/0.7	2.29/1.75	1.83/1.51	18.10/11.6	17.43/5.96
3×					
'Set 5'	1.48/0.75	2.75/1.10	2.69/0.76	50.5/23.52	48.2/40.40
'Set 14'	1.00/0.60	1.52/1.06	1.16/1.05	29.26/16.34	26.41/15.68
'BSD200'	1.04/0.63	1.74/1.12	1.70/1.09	17.51/4.89	14.73/4.63
4×					
'Set 5'	0.88/0.56	2.71/0.82	1.16/0.35	54.71/13.27	40.32/20.31
'Set 14'	0.88/0.57	1.17/0.78	1.04/0.65	21.56/7.20	25.81/14.35
'BSD200'	0.84/0.54	1.53/0.83	1.29/0.78	12.91/2.83	10.68.81/2.82

Table 12 Average improvement of various objective quality measures of proposed method over the methods in [23] and [35] at magnification factors 3× and 4×

Datasets	ΔSRCC(%) [23]/[35]	ΔPLCC (%) [23]/[35]	ΔRMSE [23]/[35]
3×			
'Set 5'	28.69/9.32	27.60/7.99	0.995/0.217
'Set 14'	23.62/15.20	21.25/8.45	0.993/0.288
'BSD200'	14.14/9.79	12.32/6.90	1.10/0.311
4×			
'Set 5'	26.29/8.07	25.86/4.82	0.68/0.11
'Set 14'	21.76/14.45	17.49/5.54	0.70/0.14
'BSD200'	13.38/8.67	11.51/4.82	0.75/0.16

time in reconstructing the HR images in three 'Set5,' 'Set14' and 'BSD100' datasets.

4.2.1 Comparison with the State-of-the-Art Faster LBSISRR Methods

In this subsection, performance of our proposed method is compared with some of the existing state-of-the-art faster LBSISRR methods [21,26,29,32,35] to show its superiority in providing a high-quality solution at a faster processing speed. Figure 7 shows the HR solutions of the 'butterfly.bmp,' 'ppt3.bmp' and '189080.jpg' images by using the above mentioned reconstruction methods.

Table 15 describes both subjective and objective image quality assessment parameters of these reconstructed solutions at a magnification factor 3×. Table 16 describes the average values of both subjective and objective image quality assessment parameters along with average processing speed (in seconds) of the reconstructed solutions at different magnification factors. Averaging process takes the quality measure values and processing time for reconstructing the

Table 13 Average processing time (seconds) at different magnification factors for 'Set5' dataset images

Zooming factor	[21]	[22]	[23]	[24]	[15]	[35]	[38]	ENESR (proposed)
'2×'	32.01	50.68	35.38	46.65	106	5.03	**3.14**	*12.84*
'3×'	42.14	61.32	45.32	56.18	132	7.00	**4.11**	*19.3*
'4×'	53.32	72.65	56.60	63.14	168	13.19	**6.12**	*22.12*

Table 14 Average processing time (seconds) at different magnification factors for images in three datasets

Zooming factor	[21]	[22]	[23]	[24]	[15]	[35]	[38]	ENESR (proposed)
'2×'	28.01	46.28	31.33	42.35	103	4.13	**3.11**	*8.34*
'3×'	34.10	55.30	41.12	52.16	129	6.72	**4.10**	*10.3*
'4×'	49.32	68.34	52.10	58.21	163	11.49	**6.10**	*15.12*

| (a) | (b) | (c) | (d) | (e) | (f) |
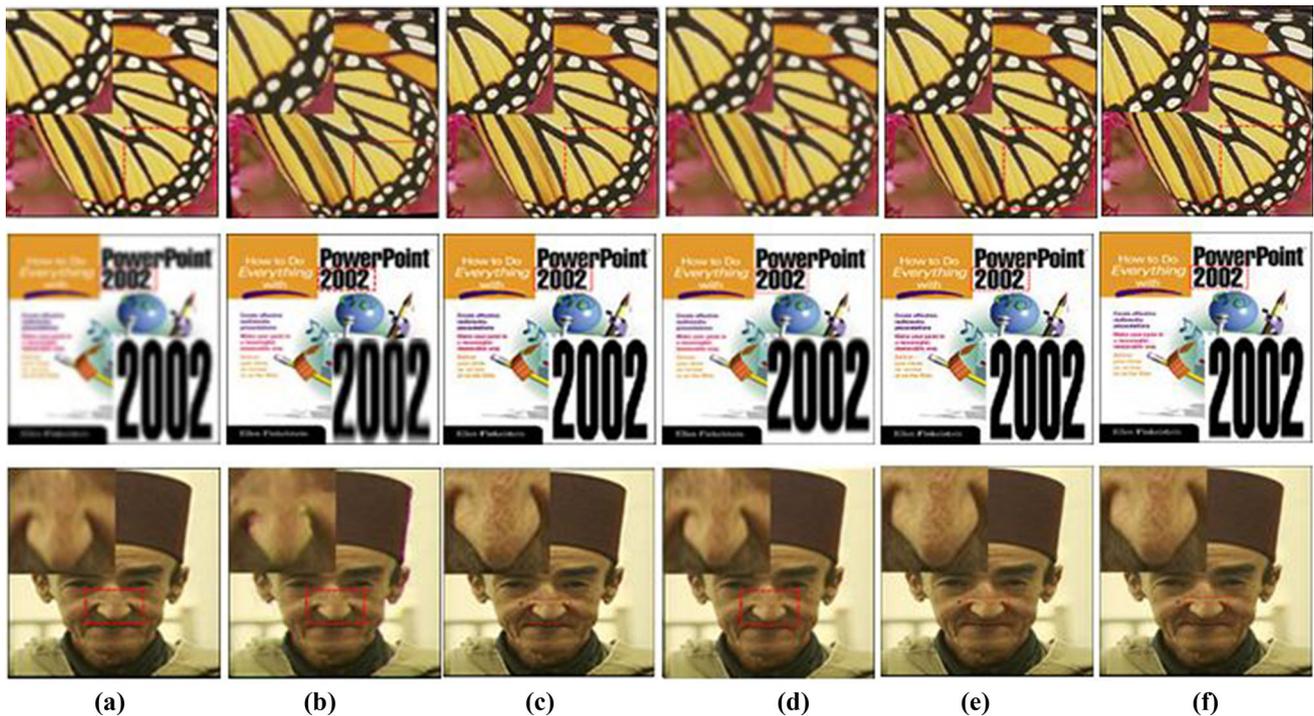
Fig. 7 Reconstructed HR solutions of 'butterfly.bmp,' 'ppt3.bmp' and '189080.jpg' images using different faster LBSISRR methods at a zooming factor of 3×. **a** Zeyde et al. [26], **b** Gao et al. [21], **c** Timofte et al. [29], **d** Ahmadi et al. [32], **e** Liu et al. [35], **f** ENESR (proposed)

images in 'Set5,' 'Set14' and 'BSD100' datasets. From the visual perception and tabular data analysis, it is observed that, proposed ENESR method outperforms the compared faster LBSISRR methods. This is because, our proposed method efficiently fuses several raw image features to preserve sufficient amount of structural, statistical and spatial image details in the estimated solutions with reduced effect of artifacts. However, method in [29] provides comparable results with the proposed method. This method utilizes the combination of sparse coding and neighbor embedding to learn the dictionary. Unlike, conventional EuD-based measures, candidate patches are searched by considering the correlation with the

dictionary atoms and hence sufficient spatial correlation have been preserved among neighboring patches. Consequently, yields a better quality of reconstruction and considered as the next state-of-the-art faster LBSISRR method.

4.2.2 Entropy of Reconstructed HR Solutions

Along with the above mentioned subjective and objective quality measures, entropy of the LR and its corresponding reconstructed HR images using different LBSISRR methods are also provided here to show the information content in the reconstructed solutions. Entropy values of the reconstructed

Table 15 Quality measures of 'butterfly.bmp,' 'ppt3.bmp' and '189080.jpg' images using proposed and compared faster LBSISRR methods at a zooming factor of 3×

Methods	PSNR (dB)	SSIM	FSIM	RM	BM	SRCC	PLCC	RMSE
'butterfly.bmp'								
[26]	28.67	.846	.868	6.00	7.28	.4432	.5545	3.112
[21]	28.40	.844	.865	6.42	7.73	.4321	.5432	3.321
[29]	33.81	.876	.928	2.10	2.76	.6865	.7654	2.110
[32]	28.85	.848	.875	5.60	6.85	.5014	.5748	3.001
[35]	33.67	.874	.926	2.14	2.93	.6756	.7555	2.342
Proposed	*34.52*	*.880*	*.933*	*1.77*	*2.26*	*.7112*	*.8234*	*1.412*
'ppt3.bmp'								
[26]	28.45	.688	.718	7.40	8.65	.4567	.5003	2.434
[21]	28.34	.673	.714	8.11	9.14	.4432	.4956	2.675
[29]	33.54	763	.841	3.12	4.02	.6657	.7008	1.897
[32]	28.68	.695	.722	6.55	7.54	.4615	.5123	2.121
[35]	33.48	.758	.836	3.83	4.64	.6638	.7000	1.930
Proposed	*34.48*	*.797*	*.840*	*2.34*	*3.18*	*.7231*	*.7342*	*1.753*
'189080.jpg'								
[26]	28.73	.807	.826	10.34	11.65	.4578	.4973	3.332
[21]	28.64	.801	.823	11.1	12.0	.4649	.5057	3.456
[29]	32.50	.847	.887	3.97	4.45	.5367	.5743	2.123
[32]	28.85	.811	.829	5.46	7.23	.4700	.5188	3.176
[35]	32.16	.841	.880	4.01	5.32	.5321	.5770	2.459
Proposed	*33.14*	*.853*	*.896*	*3.20*	*4.04*	*.5432*	*.5842*	*1.801*

Table 16 Comparison of average values (for images in 'Set5,' 'Set14' and 'BSD100' datasets) of various image quality measures of the reconstructed HR images using proposed scheme along with compared faster LBSISRR methods at different zooming factors

| Methods	PSNR (dB)	SSIM	FSIM	RM	BM	SRCC	PLCC	RMSE	Processing time (s)
3×									
[26]	28.56	.8192	.8243	10.32	12.12	.4512	.5010	2.81	20.21
[21]	28.39	.8183	.8220	10.88	12.63	.4419	.4991	2.99	34.10
[29]	30.96	.8465	.8618	4.63	5.17	.6002	.7192	1.73	5.60
[32]	28.63	.8200	.8376	9.94	10.14	.4617	.5146	2.06	11.24
[35]	30.92	.8443	.8590	4.74	6.1	.5990	.7016	1.79	6.72
Proposed	31.58	.8536	.8760	4.13	5.03	.7314	.7610	1.515	10.34
4×									
[26]	26.60	.738	.768	11.84	14.11	.4108	.4721	4.63	22.14
[21]	26.53	.732	.761	12.63	14.73	.4076	.4631	4.94	49.32
[29]	28.40	.772	.813	5.56	7.25	.6017	.7032	2.40	8.63
[32]	26.68	.740	.773	11.07	12.11	.4231	.4856	4.11	19.18
[35]	28.32	.770	.810	5.69	7.37	.5824	.6880	2.47	11.49
Proposed	28.69	.777	.819	5.32	6.54	.6481	.7247	2.34	15.12
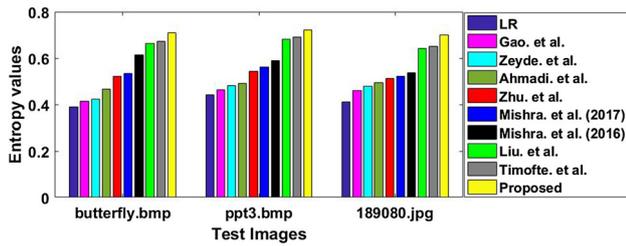
Fig. 8 Bar plot of the entropy values of different SISRR methods for estimating 'butterfly.bmp,' 'ppt3.bmp' and '189080.jpg' images at a zooming factor of 3×

'butterfly.bmp,' 'ppt3.bmp' and '189080.jpg' images at a magnification factor of 3× are plotted in Fig. 8.

Formulation for the calculation of entropy and information gain is provided in Eq. (10).

$$En = -\sum_{i=1}^{M}\sum_{j=1}^{N} p(i,j)\log_2 p(i,j)$$

$$IG = \frac{En(HR)}{En(LR)} \tag{10}$$

where M, N are size of images, $p(i,j)$ pixel values and $\log_2 p(i,j)$ probability density.

From Fig. 8, it is quite evident that, for each of the images the HR solutions contain higher information than its corresponding LR images. However, our proposed ENESR method outperforms the compared methods and methods in [29,35] provide comparable results. Information gain (IG) of a method is computed by taking the ratio of the entropy values of the reconstructed HR solution to its corresponding LR image. Higher value of IG signifies more information content. Values of IG for 'butterfly.bmp,' 'ppt3.bmp' and '189080.jpg' images of the proposed method are: 1.82, 1.63, 1.70. Values of IG for the method in [35] are: 1.70, 1.54, 1.55, whereas values for the method in [29] are: 1.72, 1.56, 1.70, respectively. Thus, average improvement of the proposed ENESR method over the methods in [29,35] is around 6.98% and 5.81 %, respectively.

4.2.3 Reconstruction Quality Versus Processing Speed

This section discusses about the trade-off between the reconstruction accuracy and the processing speed behavior of the reconstruction process. This is demonstrated by plotting the average computational speed (in seconds) against the avg. PSNR(dB) of the reconstructed HR solutions. The plot is provided for the proposed, compared state-of-the-art LBSISRR methods and faster SRR methods at different magnification factors. Figures 9 and 10 show the plot of average running time Vs average PSNR at 3× and 4×, respectively.
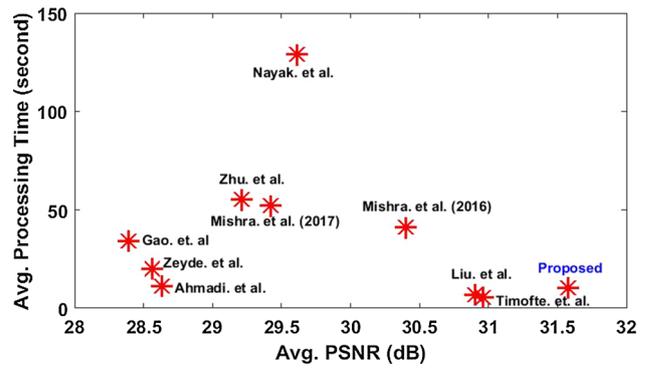

Fig. 9 Avg. processing speed (s) versus Avg. PSNR (dB) of the proposed and the state-of-the-art LBSISRR methods at a magnification factor of 3×. Blue color represents the proposed ENESR method
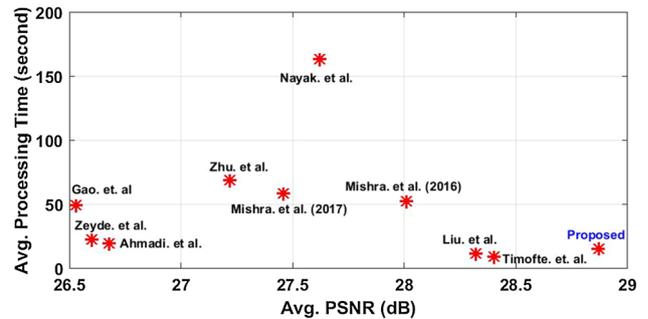

Fig. 10 Avg. processing speed (s) versus Avg. PSNR (dB) of the proposed and the state-of-the-art LBSISRR methods at a zooming factor 4×. Blue color represents the proposed ENESR method
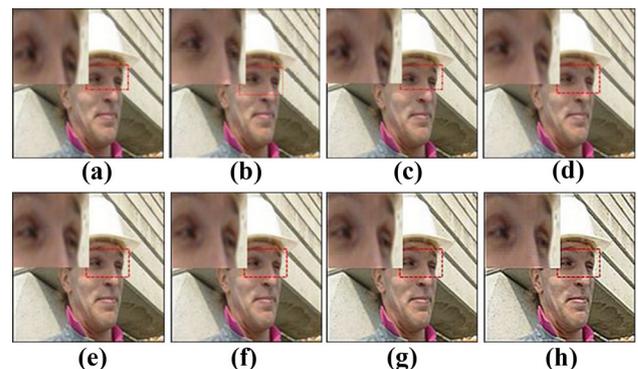

Fig. 11 Reconstructed HR solutions of 'Foreman' sequence using different methods at a magnification factor 3× along with the BIQI/NIQE values **a** Gao et al. [21] (19.46/.5612), **b** Zhu et al. [22] (20.01/.5718), **c** Mishra et al. [23] (22.43/.6013), **d** Mishra et al. [24] (20.68/.5819), **e** Nayak et al. [15] (20.86/.5867), **f** Liu et al. [35] (23.60/.6312), **g** Ledig et al. [38] (32.24/.6615), **h** ENESR (proposed) (25.01/.6519)

From the plot, it is observed that the methods in [15,22] consume more processing time to converge. This is due to the integration of EuD-based measure to compute the compatibility functions and patch similarity. As compared to other methods, method in [29,35] outperforms in terms of computational speed. Integration of SC based learning method by
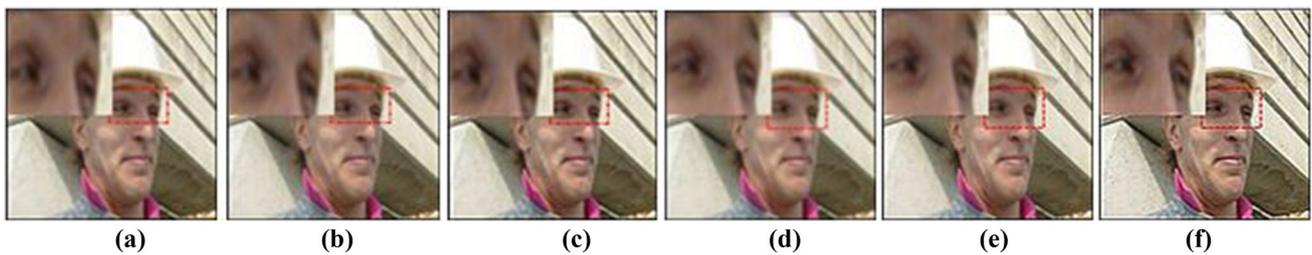
Fig. 12 Reconstructed HR solutions of 'Foreman' sequence using proposed and compared faster LBSISRR methods at a magnification factor 3× along with their BIQI/NIQE values **a** Zeyde. et al. (19.90/.5703), **b** Gao et al. [21] (19.46/.5612), **c** Timofte et al. [29] (24.11/.6410), **d** Ahmadi et al. [32] (19.97/.5711), **e** Liu et al. [35] (23.60/.6312), **f** ENESR (proposed) (25.01/.6519)

Table 17 Average values of BIQI/NIQE measures for real-time reconstructed HR images in MDSP database [49] at different zooming factors

Zooming factors	[21]	[22]	[23]	[24]	[15]	[29]	[35]	[38]	ENESR (proposed)
'2 ×'	26.3/.581	28.2/.601	32.2/.641	29.2/.612	31.3/.621	33.5/.691	33.3/.682	**41.7/.731**	*34.2/.701*
'3 ×'	22.4/.562	25.7/.583	30.7/.622	26.4/.591	28.9/.601	31.4/.670	31.1/.662	**39.2/.721**	*31.7/.680*
'4 ×'	20.1/.531	23.1/.551	28.9/.591	25.4/.560	27.0/.572	30.0/.650	29.9/.642	**37.4/.711**	30.7/.660

Table 18 Average improvement of ∆BIQI (%)/∆NIQE (%) of proposed method over the methods in [35] and [29] at different magnification factors

Zooming factor	[35]	[29]
'2 ×'	2.63/2.75	1.99/1.44
'3 ×'	2.30/2.74	1.98/1.46
'4 ×'	2.54/2.74	1.96/1.44

utilizing the KNN-based dictionary clusters enables faster convergence in method [35]. However, method in [29] utilizes an off-line anchored regression process to obtain the neighborhood by measuring the correlation with the dictionary atoms. Most importantly, this method integrates the notion of SC along with neighbor embedding approach to interpret the learned dictionary in a smaller space. Consequently, the method enables to preserve sufficient image details and converges at a faster rate than the methods using SC only [26,32,35]. However, our proposed method is consuming a little more time than the comparable methods in [29,35] but yields high-quality reconstruction accuracy. This is due to the use of covariance-based feature representation of image patches. At higher magnification factors, i.e., at 3× 4×, proposed ENESR method achieves an average gain of PSNR value by .45 dB to .64 dB at a cost of 4 to 6 seconds more computational time than the next state-of-the-art LBSISRR methods in [29,35]. Hence, methods in [29,35] can be efficiently used for applications where the computation speed is a major concern with the sacrifice in image quality. On the contrary, our proposed ENESR method maintains an efficient trade-off between computational speed and recon-

struction accuracy. Hence, it can be successfully applied in various real-time applications.

4.3 Performance Analysis for Real-Time Data Experiments

In this section, we have provided the performance analysis of real-time test images for different magnification factors. Figure 11 depicts the reconstructed HR solutions of the Foreman sequence along with their corresponding BIQE and NIQE values at a magnification factor of 3×.

Figure 12 depicts the reconstructed HR solutions of the Foreman sequence along with their corresponding BIQE and NIQE values when compared with faster reconstruction methods at zooming factor of 3×.

Table 17 provides the average values of BIQI and NIQE measures of the reconstructed HR solutions from MDSP dataset [49] at different magnification factors. From the visual perception and tabular data analysis, it is observed that, even for real-time data analysis also, our proposed method is providing superior reconstruction quality than the other methods. Methods in [29] and [35] are providing comparable results to the proposed work. But, with increase in magnification factor, the reconstruction accuracy of these methods severely degrades and artifacts dominant. On the other hand, our proposed method maintains its reconstruction accuracy even at larger magnification factors.

Table 18 provides the average improvement of BIQI and NIQE measures of the proposed method over the methods in [29] and [35] at different magnification factors for MDSP dataset.

5 Conclusion

In this paper, we have presented an efficient neighbor embedding-based super-resolution reconstruction scheme by utilizing an enhanced way of representing image patch in feature space and a faster way of searching candidate patches. Each patch is efficiently represented by fusing a number of low- and high-level feature attributes which in turn provides sufficient essential information for reconstructing high-quality image details. Moreover, this unified means of patch representation makes the reconstruction process robust to noise. Size of K is chosen adaptively and patch similarity is measured in non-Euclidean space. This overcomes the bottleneck of Euclidean distance-based similarity measures. Qualitative and quantitative analysis of reconstruction process suggests that methods in [29,35] provide comparable results with the proposed method and hence considered as the next two state-of-the-art methods. Our proposed work produces an average gain in PSNR by 0.37-0.68 dB, average improvements of SSIM, FSIM, RM, BM, SRCC and IG measures are by 0.91–1.08%, 1.09–1.94%, 6.45–12%, 11.26–17.54%, 10.13–18.10% and 6.98%, respectively, over the method in [35] at a cost of 3–5 s more processing time. Moreover, as compared to method in [29], our method achieves an average gain in PSNR by .30–.60 dB, average gain in SSIM, FSIM, RM, BM, SRCC and IG measures are by .64–.8%, .73–1.62%, 4.31–10.78%, 2.7–9.7%, 7.15–17.9% and 5.8% at a cost of 4–6 s processing time. All these analysis suggest that the proposed work produces a better trade-off between sharper image details and running time than the compared state-of-the-art methods.

References

1. Hamza, R.; Hassan, A.; Huang, T.; Ke, L.; Yan, H.: An efficient cryptosystem for video surveillance in the internet of things environment. Complexity **2019**, (2019)

2. Hamza, R.; Hewage, C.; Titouna, F.: Investigation of three-dimentional image security based on improved image randomised encryption method. NED Univ. J. Res. (2018)

3. Singh, D.; Kumar, V.; Kaur, M.: Single image dehazing using gradient channel prior. Appl. Intell. **49**(12), 4276–4293 (2019)

4. Singh, D.; Kumar, V.: Single image defogging by gain gradient image filter. Sci. China Inf. Sci. **62**(7), 79101 (2019)

5. Jiang, K.; Wang, Z.; Yi, P.; Jiang, J.: A progressively enhanced network for video satellite imagery superresolution. IEEE Signal Process. Lett. **25**(11), 1630–1634 (2018)

6. Jurek, J.; Kociński, M.; Materka, A.; Elgalal, M.; Majos, A.: Cnn-based superresolution reconstruction of 3D MR images using thick-slice scans. Biocybernet. Biomed. Eng. **40**(1), 111–125 (2020)

7. Zhang, L.; Zhang, H.; Shen, H.; Li, P.: A super-resolution reconstruction algorithm for surveillance images. Signal Process. **90**(3), 848–859 (2010)

8. Freeman, W.T.; Jones, T.R.; Pasztor, E.C.: Example-based super-resolution. IEEE Comput. Graph. Appl. **22**(2), 56–65 (2002)

9. Sun, J.; Zheng, N.-N.; Tao, H.; Shum, H.-Y.: Image hallucination with primal sketch priors. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings, vol. 2, pp. II–729 IEEE (2003)

10. Huang, D.J.; Siebert, J.P.; Cockshott, W.P.; Xiao, Y.J.: Super-resolution of face images based on adaptive markov network. In: Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, 2007, pp. 742–747. IEEE (2007)

11. Banerjee, J.; Namboodiri, A.M.; Jawahar, C.; Contextual restoration of severely degraded document images. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, vol. 2009, 517–524. IEEE (2009)

12. Wu, W.; Liu, Z.; Gueaieb, W.; He, X.: Single-image super-resolution based on markov random field and contourlet transform. J. Electron. Imaging **20**(2), 023005–023005 (2011)

13. Chen, X.; Qi, C.: Document image super-resolution using structural similarity and markov random field. IET Image Process. **8**(12), 687–698 (2014)

14. Nayak, R.; Krishna, L.S.; Patra, D.: Enhanced super-resolution image reconstruction using MRF model. In: Progress in Intelligent Computing Techniques: Theory, Practice, and Applications, pp. 207–215. Springer (2018)

15. Nayak, R.; Patra, D.: New single-image super-resolution reconstruction using MRF model. Neurocomputing **293**, 108–129 (2018)

16. Chang, H.; Yeung, D.-Y.; Xiong, Y.: Super-resolution through neighbor embedding. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004, vol. 1, pp. I–I. IEEE (2004)

17. Chan, T.-M.; Zhang, J.: Improved super-resolution through residual neighbor embedding. J. Guangxi Normal Univ. **24**(4), 1–8 (2006)

18. Chan, T.-M.; Zhang, J.; Pu, J.; Huang, H.: Neighbor embedding based super-resolution algorithm through edge detection and feature selection. Pattern Recognit. Lett. **30**(5), 494–502 (2009)

19. Zhang, K.; Gao, X.; Li, X.; Tao, D.: Partially supervised neighbor embedding for example-based image super-resolution. IEEE J. Sel. Top. Signal Process. **5**(2), 230–239 (2010)

20. Cao, M.; Gan, Z.; Zhu, X.: Super-resolution algorithm through neighbor embedding with new feature selection and example training. In: 2012 IEEE 11th International Conference on Signal Processing (ICSP), vol. 2, pp. 825–828. IEEE (2012)

21. Gao, X.; Zhang, K.; Tao, D.; Li, X.: Image super-resolution with sparse neighbor embedding. IEEE Trans. Image Process. **21**(7), 3194–3205 (2012)

22. Zhu, Q.; Sun, L.; Cai, C.: Non-local neighbor embedding for image super-resolution through FoE features. Neurocomputing **141**, 211–222 (2014)

23. Mishra, D.; Majhi, B.; Sa, P.K.; Dash, R.: Development of robust neighbor embedding based super-resolution scheme. Neurocomputing **202**, 49–66 (2016)

24. Mishra, D.; Majhi, B.; Sa, P.K.: Improved feature selection for neighbor embedding super-resolution using zernike moments. In: Proceedings of International Conference on Computer Vision and Image Processing, pp. 13–24. Springer (2017)

25. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y.: Image super-resolution via sparse representation. IEEE Trans. Image Process. **19**(11), 2861–2873 (2010)

26. Zeyde, R.; Elad, M.; Protter, M.: On single image scale-up using sparse-representations. In: International Conference on Curves and Surfaces, pp. 711–730. Springer, Berlin (2010)

27. Yang, J.; Wang, Z.; Lin, Z.; Cohen, S.; Huang, T.: Coupled dictionary training for image super-resolution. IEEE Trans. Image Process. **21**(8), 3467–3478 (2012)

28. Wang, S.; Zhang, L.; Liang, Y.; Pan, Q.: Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2216–2223. IEEE (2012)

29. Timofte, R.; De Smet, V.; Van Gool, L.: A+: Adjusted anchored neighborhood regression for fast super-resolution. In: Asian Conference on Computer Vision, pp. 111–126. Springer, Berlin (2014)

30. Wang, L.; Geng, H.; Liu, P.; Lu, K.; Kolodziej, J.; Ranjan, R.; Zomaya, A.Y.: Particle swarm optimization based dictionary learning for remote sensing big data. Knowl.-Based Syst. **79**, 43–50 (2015)

31. Nayak, R.; Patra, D.; Harshavardhan, S.: Sparse representation based image super resolution reconstruction. In: TENCON 2015–2015 IEEE Region 10 Conference, pp. 1–6. IEEE (2015)

32. Ahmadi, K.; Salari, E.: Single-image super resolution using evolutionary sparse coding technique. IET Image Process. **11**(1), 13–21 (2016)

33. Fang, L.; Zhuo, H.; Li, S.: Super-resolution of hyperspectral image via superpixel-based sparse representation. Neurocomputing **273**, 171–177 (2018)

34. Yang, X.; Wu, W.; Liu, K.; Chen, W.; Zhou, Z.: Multiple dictionary pairs learning and sparse representation-based infrared image super-resolution with improved fuzzy clustering. Soft Comput. **22**(5), 1385–1398 (2018)

35. Liu, N.; Xu, X.; Li, Y.; Zhu, A.: Sparse representation based image super-resolution on the knn based dictionaries. Optics Laser Technol. **110**, 135–144 (2019)

36. Dong, C.; Loy, C.C.; He, K.; Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2015)

37. Kim, J.; Lee, J.K.; Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)

38. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690 (2017)

39. Wang, Y.; Wang, L.; Wang, H.; Li, P.: End-to-end image super-resolution via deep and shallow convolutional networks. IEEE Access **7**, 31959–31970 (2019)

40. Wang, Z.; Chen, J.; Hoi, S.C.: Deep learning for image super-resolution: a survey. arXiv preprint arXiv:1902.06068 (2019)

41. Tuzel, O.; Porikli, F.; Meer, P.: Region covariance: A fast descriptor for detection and classification. In: Computer Vision-ECCV, vol. 2006, pp. 589–600 (2006)

42. Nayak, R.; Patra, D.: Super resolution image reconstruction using penalized-spline and phase congruency. Comput. Electr. Eng. **62**, 232–248 (2017)

43. Ou, Y.; Sotiras, A.; Paragios, N.; Davatzikos, C.: Dramms: deformable registration via attribute matching and mutual-saliency weighting. Medical Image Anal. **15**(4), 622–639 (2011)

44. Zhan, Y.; Shen, D.: Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method. IEEE Trans. Med. Imaging **25**(3), 256–272 (2006)

45. Marziliano, P.; Dufaux, F.; Winkler, S.; Ebrahimi, T.: Perceptual blur and ringing metrics: application to jpeg2000. Signal Process. Image Commun. **19**(2), 163–172 (2004)

46. Tang, L.; Sun, K.; Liu, L.; Wang, G.; Liu, Y.: A reduced-reference quality assessment metric for super-resolution reconstructed images with information gain and texture similarity. Signal Process. Image Commun. **79**, 32–39 (2019)

47. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding (2012)

48. Martin, D.; Fowlkes, C.; Tal, D.; Malik J.; et al.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Iccv Vancouver (2001)

49. Milanfar, P.; Farsiu, S., Robinson, D.: MDSP super-resolution and demosaicing datasets (2016)

50. Moorthy, A.K.; Bovik, A.C.: A two-step framework for constructing blind image quality indices. IEEE Signal Process Lett. **17**(5), 513–516 (2010)

51. Mittal, A.; Soundararajan, R.; Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal Process. Lett. **20**(3), 209–212 (2012)