Image super-resolution using sparse coding over redundant dictionary based on effective image representations

Muhammad Sajjad, Irfan Mehmood, Sung Wook Baik *

College of Electronics and Information Engineering, Sejong University, Seoul, Republic of Korea

Abstract

Recent years have shown a growing research interest in the sparse-representation of signals. Signals are described through sparse linear combinations of signal-atoms over a redundant-dictionary. Therefore, we propose a novel super-resolution framework using an overcomplete-dictionary based on effective image-representations such as edges, contours and high-order structures. This scheme recovers the vector of common sparse-representations between low-resolution and corresponding high-resolution image-patches by solving the $\ell_1$-regularized least-squared problem; subsequently, it reconstructs the HR output by multiplying it with the learned dictionary. The dictionary used in the proposed-technique contains more effective image-representations than those in previous approaches because it contains feature-descriptors such as edges, contours and motion-selective features. Therefore, the proposed-technique is more robust to various types of distortion. A saliency-map quickens this technique by confining the optimization-process to visually salient regions. Experimental analyses confirm the effectiveness of the proposed-scheme, and its quantitative and qualitative performance as compared with other state-of-the-art super-resolution algorithms.

1. Introduction

Image super resolution (SR) is an enthusiastic area of research and desirable for many applications. Image SR increases the pixel density of a low-resolution (LR) image to obtain a high-resolution (HR) image. Image SR promises to overcome some of the inherent resolution limitations of low-cost imaging sensors, e.g. surveillance cameras and cell phones. LR images obtained from LR digital devices are reconstructed to an acceptable resolution for efficiently utilizing the growing capabilities of HR devices, such as high-definition LCDs. In medical imaging, analysis and diagnosis are considerably difficult to achieve using low-quality images. Therefore, HR images are obtained from their corresponding low-quality images using various image SR techniques. HR technology is also used effectively in Blu-ray movies and high-quality video conferencing. Furthermore, Web videos require an enhancement in resolution for their frame sequences because they are stored at low-quality to deal with limited server storage and bandwidth. The SR problem also provides an appropriate environment through which image processing ideas and techniques can be assessed.

Various SR schemes have been developed to enhance the resolution of LR images. Interpolation-based techniques are used to construct an HR image from its corresponding LR image. However, simple interpolation methods cannot recover missing high-frequency components, and often blur the discontinuities of a zoomed image. Some classic simple interpolation methods such as nearest neighbor, bilinear, and bicubic methods, are more frequently used than non-adaptive interpolation schemes [1]. Although these methods suffer from problems of aliasing, edge halos, and blurring effects, they are very efficient in terms of time complexity, and have been used in many applications to date. The authors in [2,3] developed adaptive interpolation methods to minimize unwanted artifacts incurred from non-adaptive interpolation-based schemes. These methods determine the re-sampling points (pixels) according to the edge information; however, these methods also suffer from distortions. These methods cannot eliminate noise from the signals and hence generate noisy results. Jurio et al. [4] proposed a nonlinear image magnification scheme which associates each pixel with an interval obtained by weighted aggregation of the pixels in its neighborhood. A magnified image is obtained from the interval and using a linear $K_a$ operator. More detail regarding $K_a$ operator can be found in [4]. Jurio et al. also ignores some important features such as texture invariance and geometrical invariance information, which cause aliasing and
blurring artifacts in a reconstructed image. Further, to improve the quality of a magnified image, the SR process is cast as an inverse problem of recovering an HR image by merging LR images, based on prior information of the observation model used to map the LR image to HR image [5,6]. However, such SR schemes perform considerably poorly owing to the existence of unknown blurring operators, ill-conditioned registration problems and few LR images.

Various machine-learning based algorithms construct HR images by capturing the co-occurrence prior between LR and HR image patches. The authors in [7] proposed a learning-based method for low-level vision problems by organizing a Markov network of image patches and an underlying scene. Factorization approximation is used to learn the parameters of a Markov network from synthetic examples of image/scene pairs, and applies the parameters to image SR. In [8], an HR image patch as a linear combination of $K$ neighbors is recovered from its corresponding LR input image patch using a set of training examples based on manifold learning methods. This SR technique forms two manifolds with a similar local geometry in two distinct features spaces, i.e. LR and HR. It often fails to keep the reconstruction process free from blurring effects owing to the fixed numbers of $K$ neighbors used. In [9], the authors assumed that an LR image and its corresponding HR version have a similar local geometry. The assumptions made in [9] have been extended by many researchers using the concept of sparse coding. The authors in [10] improved the results by choosing the most relevant reconstruction neighbors based upon sparse coding. They worked directly with image patch pairs taken from LR and HR images, and determined that the use of sparse coefficients is time consuming owing to the sparse coding over a large-sample image-patch database. To further improve the algorithm in [10], the method in [11] is used to learn sparse coefficients for each patch of an LR input image, and then uses this sparse representation to recover an HR image patch. The authors trained dictionaries for both LR and HR image patches jointly by enforcing the similarity of a sparse representation, i.e., the dictionaries for LR and HR contain the same sparse representations. In [11], the algorithm from [12] is used to train dictionaries for the LR and HR features spaces. To make the method in [11] more efficient, the authors of [13] introduced some modifications by incorporating numerical shortcuts and a dimensionality reduction using a principal component analysis (PCA). In [13], the authors used the same approach to enforce a sparse coefficient similarity. For dictionary learning, they used the K-SVD dictionary procedure [14]. Furthermore, in [15], work of Yang et al. was revisited in view of the learning theory, and was improved using more flexible regression methods than a linear regression method. Unlike, the methods in [16,13], the authors in [17] proposed a framework that does not require a collection of trained LR and HR image data in advance, and do not assume the self-similarity of image patches across different feature spaces. With the theoretical support of Bayes decision theory, the authors verified that their SR framework learns and selects the optimal support vector regression (SVR) model while producing an SR image, which results in the minimum SR reconstruction error. In [18], the authors developed analysis tools for a sparse-representation based inverse problem using a redundant dictionary. They addressed some basic issues on how to regularize the solutions to underdetermined problems such as (1) the practical implications of noncompliance with theoretical compress sensing (CS) hypothesis, (2) the properties and role of dictionaries, and (3) whether the sole use of sparsity prior is sufficient. The goal of their work was to establish the relationship between compress sensing and SR, and provide a better understanding of the role of sparsity priors. In [19], the authors proposed an adaptive compressive image recovery (CIR) technique using a local piecewise autoregressive model (PAR). A set of PAR models are learned from clustered training images, and the PAR is then adaptively selected according to the local image structures for the sparse regularization of the CIR process. For SR problems, obtaining an HR output close to the original image is generally attempted. SR problems try to reduce unwanted artifacts (e.g., edge blurring, noise and blocking effects or stair cases). To recover missing information from an underlying input image, existing techniques use image descriptors, which describe the edges and lines. However, these techniques do not incorporate high-order structures such as transformational invariants for SR reconstruction. Cadieu and Olshausen [20], however, proved experimentally that such type of image representation describing high-order structures is more effective for image denoising.

To improve the quality of the HR output and reduce the vulnerability to various types of distortions, we propose a novel SR framework using a BP based sparse-coding algorithm with an overcomplete dictionary based on effective image contents. The optimization process of the proposed technique recovers the required representation coefficients, where the optimization includes the $\ell_1$-regularized least squared problem in its constraint. The $\ell_1$-regularized least squared problem is solved using an efficient solver found in [21,22]. An HR patch of the underlying LR patch can be generated as a linear combination of a few atoms from an overcomplete dictionary using the found representation coefficients. The prototype signal atoms possessed by our dictionary consist of more effective feature descriptors such as edges, lines, and/or bars, as well as contours and high-order structures. The dictionaries used by the previous approaches lack such type of feature descriptors owing to the use of dictionary procedures of K-SVD [14,12]. Hence, the proposed technique is more robust to various types of noises. Moreover, the use of a saliency map makes the proposed technique more efficient by confining the optimization process to visually salient regions of the input image. The proposed technique uses a single two-layered dictionary instead of a separate dictionary for each LR and HR feature space. The experimental results prove the superiority of the proposed technique, not only in terms of the visual results but also in a quantitative analysis, compared with other state-of-the-art schemes.

The main contributions of this paper are as follows:

- A novel and efficient framework for SR reconstruction is proposed, where dictionary learning is an offline process and can be replaced by any efficient dictionary-learning technique. In addition, the salient computational model can be easily added or removed and the minimization problem can be solved using any efficient solver.
- The common sparse coefficients can be determined using a single overcomplete dictionary, whereas in existing techniques, two separate dictionaries are used for the LR and HR feature space.
- The dictionary used in the proposed technique contains more effective image representations than those in previous approaches because it contains feature descriptors such as edges, lines, and/or bars, as well as contours and motion-selective features.
- A cost-effective saliency model is structured to speed up the reconstruction process by confining the expensive optimization process to only the salient parts of the input image.
- For sparse recovery, $\ell_1$-minimization is solved using an efficient solver found in [22] during SR reconstruction, instead of OMP and $\ell_1$-regularization linear regression, which is also known as LASSO.

The rest of this paper is organized as follows: In Section 2, the related concepts are covered for an easier understanding of the overall topic. In Section 3, the proposed method is covered in detail. In Section 4, various experiments comparing the proposed
technique with other state-of-the-art SR schemes are described. Finally, in Section 5, some concluding remarks are provided.

2. Background

In this section, the basic idea of reconstruction, sparse-representation based image SR, and the role of a dictionary in SR reconstruction are described. From this basic idea, one can easily see the core concept of SR reconstruction and the role of various constraints involved. Using sparse-representation based image SR, we describe various state-of-the-art schemes closely related to the proposed scheme. Because a dictionary plays a vital role in the reconstruction, the gradual advancement of a dictionary is presented to explore its role in SR reconstruction.

2.1. Basic idea

Image \( I \) can be modeled as a linear superposition of basis functions from dictionary \( \Phi \), combined together with a set of representation coefficient vectors \( X \):

\[
I = \Phi X + e_1.
\] (1)

The choice of \( \Phi \) and \( X = (\chi^1, \ldots , \chi^0) \) determine the image code. The vector of sparse coefficients \( \chi \) is a dynamic variable that changes according to the sparse representation of the image [12,20,23–25]. The value of \( X \) is then calculated for each image to satisfy the above equality to constitute the output of the code. The error term \( e_1 \) is included because of an improper capture of structures not well described by \( \Phi \), i.e., reconstruction errors. Suppose \( I^0_p \) and \( I^0 \) are the patches extracted at the same location from an HR image, \( I^0 \), and LR image, \( I^p \), respectively, where \( I^0 \) is the corresponding HR version of \( I^p \). The process of downsampling a particular HR patch can be presented as

\[
I^0_p = I^0_{H} * f_{DS} + e_2.
\] (2)

where \( e_2 \) is the downsampling error and \( f_{DS} \) is the interpolation operator for downsampling. The goal is to recover the high-quality HR image patch, \( I^0 \), from the LR image patch, \( I^p \), and the reconstructed \( I^0_p \) patch should be as close as possible to the original \( I^0 \), i.e., \( I^0_p \approx I^0 \). Let \( \Phi \) be an overcomplete dictionary, \( \Phi \in \mathbb{R}^{d \times j} \), where \( j \) is the number of basis functions \( (d < j) \) and \( \chi \in \mathbb{R}^d \) contains the representation coefficients of the underlying signal. Suppose a patch \( I^p \) ordered lexicographically as a column vector can be represented sparsely over dictionary \( \Phi \), i.e., \( I^p = \Phi \chi \). This can be formulated in the following form:

\[
\Phi f_{DS} = I^0_p - e \Rightarrow \| I^0_p - (\Phi \chi f_{DS}) \|^2_2 \leq e.
\] (3)

The LR patch \( I^0_p \) can be properly described as a sparse linear combination of atoms chosen appropriately from an over-complete dictionary \( \Phi \) according to the key observation derived from Eqs. (1)–(3). Thus, the patch \( I^p \) can be constructed using the same vector of sparse representation coefficient \( \chi \) over the dictionary \( \Phi \) [10,16,20–28]. The problem posed in Eq. (3) dictates that the selection of a good dictionary and minimization technique can improve the output of the image SR reconstruction.

2.2. Sparse representation based image SR

The task of image SR is to reconstruct an HR image \( I^0 \) from the observed LR image \( I^p \) as accurately as possible. This task is extremely ill posed, because for a given LR input, \( I^p \), many HR images, \( I^0 \), satisfy the previous reconstruction constraint posed in Eqs. (1)–(3). Suppose, \( I^0_p \) is the target signal to be reconstructed and \( \Phi \) is a given dictionary. The sparse coding of \( I^0_p \) over \( \Phi \) is to seek a sparser vector \( \chi \) such that \( I^0_p = \Phi \chi \). Using the sparsity prior, the representation coefficients of \( I^0_p \) over \( \Phi \) can be estimated from its observation \( I^0_p \) by solving the following \( \ell_0 \)-minimization problem

\[
\min_{\chi} \| I^0_p - \Phi \chi \|^2_2 : \| \chi \|_0 \leq \epsilon,
\] (4)

where \( \epsilon \) controls the sparsity. Alternatively, the vector of sparse coefficients \( \chi \) can also be determined as

\[
\hat{\chi} = \arg \min_{\chi} \left\{ \| I^0_p - \Phi \chi \|^2_2 + \beta \| \chi \|_0 \right\},
\] (5)

which is indeed very sparse i.e. \( \| \chi \|_0 \ll d \). The notation \( \| \cdot \|_0 \) stands for the count of non-zero entries in \( \chi \) and \( \beta \) is a constant. Because the \( \ell_0 \)-minimization is a non-convex and NP-hard problem, involving a combinatorial search that makes it computationally intractable [29], approximate solutions are considered. In the past decade, several efficient pursuit algorithms have been proposed to solve the \( \ell_0 \)-minimization. Matching pursuit (MP) and orthogonal matching pursuit (OMP) are the simplest algorithms proposed [30,31]. These iterative greedy algorithms seek sparse representation of an image patch through a sequence of mono-atomic approximations. These schemes are very simple, comprising the computation of inner-products between the image patch and dictionary atoms, and possibly deploying some least squares solvers. \( \ell_1 \)-minimization has been widely used as an alternative approach to solving the problem posed in Eq. (5) because it is the closest convex function to \( \ell_0 \)-minimization. Recent research has proved that iteratively reweighting the \( \ell_0 \)-norm sparsity regularization term can lead to better sparsity and image SR reconstruction [32]. Yang et al. [11] incorporated linear regression regularized with the \( \ell_0 \)-norm on coefficients to solve the sparse coding problem posed in Eq. (5), which is known in the statistic as a Lasso [33]. Zeyde et al. [13] used OMP to recover a vector of sparse coefficients during SR reconstruction for an underlying patch. In [34], the authors recovered the desired sparse coefficients by minimizing the \( \ell_1 \)-norm. However, they used OMP during the dictionary learning procedure for a sparse recovery. Tang et al. [15], revisited the algorithm of Yang et al. in view of learning theory, and designed a single-image SR algorithm based on the framework of \( \ell_2 \)-boosting i.e., greedy regression. Shi et al. [19] proposed an adaptive compressive image recovery technique using PAR. The authors computed the sparse PAR by solving the problem in Eq. (5) by incorporating an \( \ell_1 \)-norm based on a non-local structural sparsity regularizer. In addition, recent studies have shown that using an appropriate method to solve an \( \ell_1 \)-norm can make the recovery of a sparse representation more effective and robust.

2.3. Dictionary

Meaningful data representation is playing an energetic role in the areas of signal processing and image representation. The use of a sparse representation of signals as an impelling force for image SR and signal denoising has drawn significant research attention in the last ten years. Initially, the representation coefficients of a 1-D wavelet have been considered, but it was determined that the incorporation of a unitary wavelet is not sufficient for handling images. Hence, several redundant sparse representation coefficients were developed, e.g., a curvelet, contourlet, wedgelet, bandlet, and steerable wavelet [35]. For compression, few representation coefficients are desirable for the essential content of a signal. Therefore, the success of the JPEG2000 coding standard can be attributed to the sparsity of a wavelet [14,36]. Wavelet techniques and shift-invariant variations are among the most effective methods known for denoising [35,37]. In parallel, the image denoising problem was properly addressed by formulating a matching pursuit [38] and basis pursuit [37] as a direct sparse
decomposition technique over a redundant dictionary. Thus, an overcomplete basis set has been explored because it offers the flexibility to represent a significantly wider range of signals with more fundamental basis atoms than a single dimension [14,36,38]. Sparse and redundant data modeling is used to search for a signal representation as a linear combination of a small number of prototype signal atoms from a pre-defined overcomplete basis set. In recent years, rapid growth in interest regarding the training of overcomplete basis sets has been seen (that is, using machine-learning algorithms to learn an overcomplete dictionary directly from data, allowing the most relevant features of the signals to be efficiently replicated). This concept can be easily described using an analogy of the relationship among the vocabulary, words and dictionaries in English literature. Suppose we have to construct a particular sentence from a group of words chosen from a dictionary. Hence, there must be an algorithm that can search for a vocabulary such that only a small number of words are required to describe a given sentence, even though the set (i.e., dictionary) from which these words are taken can be significantly larger. Now, the concept can be easily understood in terms of image representation and a dictionary, e.g., the algorithm determines the coefficients of the basis functions such that only a small number of atoms are typically needed to describe a given image, even though the set from which these atoms are drawn might be significantly larger. Most importantly, the set of feature descriptors (basis functions) as a whole has to be capable of reconstructing any given image in the training set, or the image should resemble the features descriptors of the trained dictionary [20,23,39]. Taking this concept as the bedrock, various dictionaries consisting of image representations have been developed. Aharon et al. [14] developed a simple and efficient K-SVD dictionary-learning scheme. K-SVD generalizes the idea of K-means, uses a sequential updating for the dictionary, and learns image representations such as edges, lines, and/or bars. Elad et al. [35] successfully deployed a K-SVD dictionary-learning scheme for image denoising, and further, Zeyde et al. [13] incorporated the same dictionary-learning procedure for image SR reconstruction. However Zeyde et al. augmented their K-SVD dictionary by imposing a joint-learning scheme based on Yang et al.’s feature mapping. In image SR, Yang et al. train dictionaries for both the LR and HR image patches jointly by enforcing the similarity of sparse representation, that is, the dictionaries for LR and HR contain the same sparse representations. The authors in [16] used the algorithm from [12] to train their dictionaries for LR and HR features spaces. The algorithm in [12] only replicates an image representation such as line, bars, and/or edges, and cannot mimic the feature descriptors that are selective to the contours, motion, and direction of a moving object. The authors in [34] also used a KSVD dictionary-learning algorithm to simultaneously determine geometric dictionaries and sparse coefficients. Tang et al. [15] incorporated Yang et al.’s concept of dictionary learning. Moreover, the authors in [20,40,41] proved that existing dictionary-learning procedures cannot learn image representations such as contours, motion, and directions of moving objects, i.e., high-order structures. The authors proposed a hierarchical visual-information processing model that learns high-order structures, a distinction that motivated us to incorporate the work of [20,40] in the dictionary learning used in our proposed technique. The terms, set of basis functions, set of feature descriptors, and dictionary each have the same meaning and are used interchangeably throughout this paper.

3. Proposed methodology

The task of producing an HR output in the proposed framework consists of two subtasks: (1) dictionary learning and (2) image super-resolution. In dictionary learning, the procedure for learning a set of features descriptors is described that leads to the formulation of an overcomplete dictionary. This is an offline process, and is generally conducted before starting image SR. In image SR, an HR output is produced using a trained overcomplete dictionary.

3.1. Dictionary learning

The statistical properties of the natural images can be used to comprehend the variations in features of the receptive field found in the simple cells of the visual cortex [42]. The receptive fields of the primary visual cortex (V1) are localized with respect to time and space, and have band-pass features [43]. The receptive fields of visual area middle temporal (MT) also known as visual area V5 are orientated based on their sensitivity to the velocity (direction and motion) stimulus [44]. The receptive fields of simple cells in V1 can be considered as edge, bar, and line detectors because they provide a high response to the oriented edge, bar, and line. Similar to simple cells, complex cells in V1 are selective to a spatio-temporal orientation, and their responses are highly sensitive to local phase information [45]. The joint localization of the spatial and frequency domains has led to the idea that this joint localization mimics a Gabor filter by minimizing the uncertainty in both domains [42,46]. The role of V1 and MT in visual-information processing and their hierarchical relationship was replicated in the form of a three-layer hierarchical model [20,40,41] that learns a highly overcomplete basis (features) from time-varying natural images. In this hierarchical relationship, the first layer is the input layer, which represents the time-varying natural pixel intensities. The training image sequences for the input layer are obtained from open access datasets [47,48]. The time-varying image sequences consist of animal footage in grasslands along rivers and streams. They contain a variety of motions, e.g., camera motions, background motion, motion borders, from the occlusions (introduced through tracking) and motions of the animals in the scene. The time-varying natural-image sequences obtained for training purposes were spatially low-pass filtered and whitened, as described in [24], because such data in their raw form pose potential problems owing to data corruption and artifacts at the highest spatial frequency of the image, and because of huge inequalities in variance, in addition to the different directions of the input space. A technique introduced by [49] to improve these defects is to sphere the data by making the variance equal in all directions. Sphering can be achieved using a whitening filter (circularly symmetric) with frequency response, $W(f) = f$, because the amplitude of the spectrum falls at rate of $1/f$ in all directions in the 2D frequency plane; this implies that the high-intensity values are boosted and the lower frequencies are weakened. To maintain a tradeoff between high and low frequencies, it is appropriate to filter the training data set using a circular symmetric low-pass filter as

$$L(f) = e^{-|f|/f_0^n},$$

where $f_0$ is the cut-off frequency of 200 cycles/picture, and $n$ is the steepness parameter, i.e., $n = 4$. Whitening and low-pass filters can be combined to preprocess the training image sequences with a frequency response:

$$R(f) = W(f) \cdot L(f) = f^{-|f|/s^n}.$$  

The second layer (hidden layer) of the hierarchical model replicates the Gabor-like properties (lines, bars, and edges, i.e., amplitude information) of the simple cells, and responds to the local phase information of the receptive fields of complex cells in V1. The amplitude indicates the degree of presence of the corresponding feature, while the phase represents its exact position. It was
shown experimentally that many of the observed response characteristics of neurons in V1 may be imitated in terms of a sparse coding model of images, as explained in the complex basis function model for replicating the receptive fields of both simple and complex cells of V1:

\[ l_t = \Re \{ X_t^{\Phi V_1} \} + e_t, \]

where \( l_t \) is the image intensity as a function of space \((x^2)\) and time \( t \) (at the subscript). The set of basis functions \( \Phi_{V_1} \) now has both real and imaginary parts instead of only a real part, as in Eq. (1). In addition, \( \Re \) indicates the real part of the argument, and * denotes a complex conjugate. The set of representation coefficient vectors, \( X_t \), is also complex:

\[ X_t = a_t e^{i \theta_t}, \]

where \( a_t \) is the amplitude controlling the real and imaginary parts of basis function \( \Phi_{V_1} \) (forming a two-dimensional subspace), and phase \( \theta_t \) determines the position within each subspace. Amplitude \( a_t \) and phase \( \theta_t \) represent the simple and complex cells of V1, respectively. The learning process for the second layer, which replicates the receptive fields of V1, is given by minimizing the following energy function for the dictionary \( \Phi_{V_1} \):

\[
\min_{a_t, \theta_t} \sum_t \left| l_t - X_t^{\Phi V_1} \right|^2 + \sum_t \left( \lambda_t a_t + \lambda_t |a_t|^2 \right),
\]

where the term \( \lambda_t a_t \) and \( \lambda_t |a_t|^2 \) impose sparseness and temporal stability on the time rate of the change in amplitude, respectively. The amplitude and phase are computed by minimizing Eq. (10) with respect to \( a_t \) and \( \theta_t \). Because there is no penalty on phase \( \theta_t \) variables, these variables will therefore rotate as needed in order to efficiently match the sparsity and reduce the variation in amplitude in the time-varying images. The learning rule for the second layer basis functions achieved by minimizing the problem posed in Eq. (10) with respect to \( \Phi_{V1} \) using coefficients \( a_t \) and \( \theta_t \) inferred in (10), is as follows:

\[
\Delta \Phi_{V1} \propto \frac{1}{\sigma^2} \sum_t \left( l_t - \Re \{ X_t^{\Phi V_1} \} \right) X_t.
\]

The dictionary \( \Phi_{V1} \) emulates the receptive fields of neurons in V1. The third layer of the hierarchical model mimics the receptive fields of neurons in MT, which are sensitive to the motion and direction of objects. The decomposition of time-varying images into amplitude \( a_t \) and phase \( \theta_t \) variables produces a non-linear representation of the image content, enabling the hierarchical model to learn a higher-order structure in a simple form. The process of decomposition may be exploited by the third layer of processing of the hierarchical model. The dynamics of objects moving in continuous trajectories over short epochs will be encoded in the population activity of the phase \( \theta_t \) variable. As in Eqs. (1) and (8), the phase variable can also be modeled as follows:

\[
\theta_t = \Phi_{MT} \delta_t + \nu_t,
\]

The phase \( \delta_t \) is represented in terms of a set of basis functions, \( \Phi_{MT} \), and the presence or absence of these basis functions is denoted by the coefficient \( \lambda_t \). The additive noise term \( \nu_t \) indicates uncertainty or noise in the approximation of the time-rate variation. The energy function can be minimized as

\[
\min_{\delta_t} \sum_t \sum_{k \in \{0, 1\}} - \cos \left( \theta_t^i - \Phi_{MT} \delta_t \right) + \lambda_t \delta_t,
\]

where \( \lambda_t \delta_t \) imposes a sparse, independent distribution on coefficient \( \delta_t \). Now, the coefficients are computed by minimizing the problem posed in Eq. (13) with respect to \( \delta_t \). The learning rule for the minimization problem posed in Eq. (13) with respect to \( \Phi_{MT} \) using \( \delta_t \), as previously inferred, is

\[
\Delta \Phi_{MT} = \sum_{i \in \{0, 1\}} \sin(\theta_t - \Phi_{MT} \delta_t) \delta_t.
\]

The two hidden layers are not independent from each other. The forward and backward propagations affect the inference of the coefficients \( a_t \), \( \theta_t \), and \( \delta_t \). Because the inference of \( \delta_t \) depends on \( \theta_t \), the inference of \( \theta_t \) in turn depends on both \( \delta_t \) and \( a_t \). This forward and backward propagation resembles the ventral and dorsal streams of visual processing in a primate’s visual cortex [50]. A higher level of abstraction adapted to the statistics of a visual word forms additional constraints on the lower levels. This feedback from higher-level representation of MT can in turn alter the inference of V1. The second and third layers are converged to a set of localized, oriented, bandpass, and transformation feature descriptors using a Matlab package developed by [20,40,41]. The model was trained on a natural image sequence obtained from public repositories [47,48]. The second and third layers were trained on different sized image patches to learn a set of features descriptors (basis functions). The learned set of basis functions demonstrates feature descriptors that are sensitive to lines, edges, and bars, as well as a high-order structure such as the velocity of a moving pattern. To understand the replication process of a hierarchical model in detail, the readers are referred to [20,40,41].

3.2. Image super-resolution

Users usually focus on the salient region of an image or video. Therefore, expensive image processing operations are only applied to the salient regions to speed up the overall SR reconstruction. To make the proposed SR scheme cost effective, a saliency map of the input image is first computed, following which SR reconstruction for HR output is applied. Only salient regions are magnified using the proposed technique, and the remaining regions are magnified through a low-cost bicubic interpolation.

3.2.1. Saliency model

Saliency algorithm in the proposed SR scheme detects most prominent regions of the source input image \( I^p \), and only these salient features are reconstructed by the proposed technique. The saliency map of the input image \( I^p \) for the proposed technique consists of three short steps. In the first step, multi-scale contrast map [51] is calculated, and in the second step, Laplacian of Gaussian map [52] is computed. In the third step, both maps are fused to obtain the salient map of the input image \( I^p \). Because contrast is used for local feature detection in various applications of computer vision and image processing [53,54], a multi-scale contrast feature image, \( f_c \), is calculated in a Gaussian image pyramid as a linear combination of different contrasts:

\[
f_c = \sum_{l=1}^{N} \sum_{p \in \omega(p)} \left( f_l^0(p) - f_l^0(p) \right)^2,
\]

where \( \omega(p) \) is a \( 9 \times 9 \) window. \( f_l^0 \) in \( I^p \) represents the \( l^{th} \) level image in the Gaussian pyramid, and the total number of levels is six, i.e., \( N = 6 \). The resultant feature map \( f_c \) is normalized in the range \( [0, 1] \). In \( f_c \), the high-contrast boundaries, i.e., the salient regions, are highlighted by assigning low scores to the homogenous regions inside the most important object. An example of a multi-scale contrast is shown in Fig. 1(a)-(c).

In the second step, the theory of edge detection presented by Harr and Hildreth is used to expose the salient features of the input image [52,55]. Their research analyzed the orientation and direction selectivity of human visual perception. For further details, refer to [52]. The input image is blurred by applying a Gaussian operator to reduce the intensity of the structures (including noise) at a smaller scale of less than standard deviation, \( \sigma \). The Gaussian
function is smooth in both the spatial and frequency domains. The second-order derivative, $\nabla^2$, which has a significant benefit of being isotropic (invariant to rotation), is applied to the blurred image. Not only does the isotropic characteristic correspond to human visual perception, it allows the proposed technique to react equally to variations in intensity in any mask direction. The Marr–Hildreth algorithm consists of convolving the Laplacian of the Gaussian mask with an input image $I_O$: \[
abla^2 G(x, y) = \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} e^{-\frac{x^2 + y^2}{2\sigma^2}},
\] (17) and the Laplacian of the Gaussian $\nabla^2 G(x, y)$ is \[
abla^2 G(x, y) = \left[\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4}\right] e^{-\frac{x^2 + y^2}{2\sigma^2}},
\] (18) where $\sigma = 3.5$ and $n = 6\sigma$ (is the size of the Gaussian filter is $n \times n$), and Laplacian $\nabla^2$ of $G(x, y)$ obtained using a threshold approximately equal to 4% of the maximum value of Eq. (18). The Laplacian filter uses a $3 \times 3$ neighborhood approach. To obtain a complete saliency map $f_s$ for the proposed technique, a multi-scale contrast map $f_c$ and feature representation map of high contrast information $f_e$ are combined: \[
f_s(x, y) = f_c(x, y) + f_e(x, y).
\] (19)

The saliency map $f_s$ of input image $I^0$ obtained from Eq. (19) is given in Fig. 1(d). The color-bar on the right side of Fig. 1(d) indicates that the pixels are more salient, with a value of greater than 0.5. These salient pixels possess high contrast information. This method of saliency has been tested on many images obtained from public databases [56,57].

3.2.2. Reconstruction of HR image

Suppose $P$ is an input LR image and $H^i$ is its corresponding HR image. There are two feature spaces, an LR feature space, $L = \mathbb{R}^d$, and HR feature space, $H = \mathbb{R}^d$. In other words, $L$ is the observation space, and $H$ is the latent space, where the signals have a sparse representation in terms of the set of basis functions. There exists a mapping function $F$ that relates patch $I_p^O$ in $L$ to its
matching patch \( I_p^\ell \) in \( L \), i.e., \( F : H \rightarrow L \), using an overcomplete dictionary \( \Phi \in \mathbb{R}^{d \times j} \) that contains \( j \) prototype signal atoms for columns \( \{d\}_{j=1} \). The problem is to identify the vector of sparse representation, \( \chi \), using trained dictionary \( \Phi \) for space \( L \) and its corresponding space \( H \), such that any given signal \( b^\ell \in \mathbb{P}^L \) can be represented as a sparse linear combination of prototype signal atoms. In this section, patch-wise SR reconstruction is performed over a set of basis functions \( \Phi \) [10,14,36]. The proposed algorithm will infer the HR image patch for each corresponding patch. The LR image \( \hat{I}^\ell \) is divided into a set of overlapping patches \( \{ I_p^\ell \}_{k=1} \), each of size \( n \times n \). The LR input image \( \hat{I}^\ell \) is scaled up by a factor of \( \gamma \) using a bicubic interpolation technique resulting in \( I^\ell \), which is the blurred version, i.e., the interpolated image \( I^\ell \) loses some high-frequency information. \( I^\ell \) is also divided into a set of patches \( \{ I_p^\ell \}_{k=1} \), the same as \( \hat{I}^\ell \). The trained redundant dictionary \( \Phi \) consists of two layers, \( \Phi_{\text{V}} \) and \( \Phi_{\text{MT}} \), as mentioned in Section 3.1, where \( \Phi_{\text{V}} \) is responsible for recovering the representation coefficients for the orientation and direction selectivity features, and \( \Phi_{\text{MT}} \) is used for the motion and direction selectivity features. To easily understand the role of each parameter in the proposed scheme, a summary of the input and output parameters is described in Table 1.

Before recovering the vector of sparse coefficients \( \chi \), a pre-processing step is applied to both \( \hat{I}^\ell \) and \( I^\ell \) similar to the approaches in [7,10,58]. In [7], a high-pass filter is used to extract the edge information from the input LR image patches. In [58], features are extracted using the first- and second-order gradients of the input LR image. In the proposed technique, the desired pre-processing is employed on a full image instead of on each patch individually to avoid the boundary problems occurring from a small patch size. The features are extracted using the Laplacian of Gaussian \( \nabla^2 \)G (Eq. (18));

\[
\begin{align*}
\hat{f}^\ell &= \hat{I}^\ell - \nabla^2 \hat{G}, \\
\hat{f}_p^\ell &= \hat{I}^\ell - \nabla^2 \hat{G}.
\end{align*}
\]

The features obtained in Eqs. (20) and (21) are used to ensure that the recovered coefficients of representation fit the most relevant part of the LR patch, and are more suitable for precisely estimating the HR output. These features are also important for predicting the lost high-frequency contents in \( I^\ell \) for the target HR image \( \hat{I}^\ell \). In addition, patches \( \hat{I}_p^\ell \) and \( \hat{I}_p^\ell \) are extracted from \( \hat{f}^\ell \) and \( \hat{f}_p^\ell \), respectively. The problem is how to recover the vector of sparse coefficients \( \chi \) for each input LR patch \( I_p^\ell \) using the trained dictionary \( \Phi \), where \( \Phi = \begin{bmatrix} \Phi_{\text{V}} \\ \Phi_{\text{MT}} \end{bmatrix} \). The found vector of sparse representation \( \chi \) is then used to recover the HR patch \( \hat{I}_p^\ell \) by multiplying it with \( \Phi \). The vector of sparse coefficients \( \chi \) can be determined for the given patch by solving the following optimization problem

\[
\min_{\chi} \| \chi \|_0 : \| f_p^\ell \Phi_{\text{V}} - f_p^\ell \Phi \chi \|_2^2 + \| f_p^\ell \Phi_{\text{MT}} - f_p^\ell \Phi \chi \|_2^2 < \epsilon.
\]

where \( \epsilon \) can be chosen according to the standard deviation \( \sigma \) of noise [35]. Eq. (22) can be used to predict the presence or absence of common sparse representations between two spaces, \( \phi' \) and \( \phi'' \) (that is, the degree to which these feature descriptors are present), in terms of the representation coefficient vector \( \chi \). Solving Eq. (22) using the \( \ell_0 \)-norm is an NP-hard problem, leading to an approximate solution being acquired using a greedy algorithm such as forward selection [60]. Several pursuit algorithms such as MP, OMP, and optimized OMP have been proposed to solve the \( \ell_0 \) minimization problem [38]. With \( \ell_1 \)-norm, the problem is convex and can be solved efficiently as long as the desired representation coefficients \( \chi \) are sufficiently sparse. In addition, \( \ell_1 \)-minimization does better in terms of true sparsity and signal recovery [22,61]. Hence, the minimization problem in Eq. (22) can be reformulated using the \( \ell_1 \)-norm

\[
\min_{\chi} \| f_p^\ell \Phi_{\text{V}} - f_p^\ell \Phi \chi \|_2^2 + \| f_p^\ell \Phi_{\text{MT}} - f_p^\ell \Phi \chi \|_2^2 + \| \chi \|_1.
\]

Theoretically, Eq. (23) is an \( \ell_1 \)-regularized least squared (LS) problem [22,29,33], where \( \beta > 0 \) balances the sparsity of the solution and fidelity of the approximation. The sparsity term \( \| \chi \|_1 \) is relaxed to \( \| \chi \|_2 \), leading to the well-known basis pursuit problem [33,37]. Some basic properties of the problem posed in Eq. (23) are as follows. (1) Limiting behavior: The limiting point has the minimum \( \ell_1 \)-norm among all points that satisfy the first two terms of Eq. (23), which leads to a high fidelity as \( \beta \rightarrow 0 \). (2) Finite convergence to zero as \( \beta \rightarrow \infty \): In Eq. (23), more attention is needed on the sparsity of \( \chi \) because \( \| \chi \|_1 \) is the dominant term. The convergence of the problem in Eq. (23) occurs for a finite value of \( \beta > \beta_{\text{max}} = [2\Phi \hat{I}^\ell \|_\infty \] \) where \( \beta \| \chi \|_1 = \max \| \chi \|_1 \) denotes the \( \ell_\infty \)-norm of vector \( u \), and for \( \beta > \beta_{\text{max}} \) the optimal solution of an \( \ell_1 \)-regularized LS is zero. Moreover, the regularization path of the problem in Eq. (23), i.e., the family of solutions as \( \beta \) varies (0, \( \infty \)), has a piecewise-linear solution path property [62]. The \( \ell_1 \)-regularized LS problem in Eq. (23) is solved using the efficient solver in [22].

### Table 1

Summary of the input and output parameters in the proposed SR reconstruction scheme.

<table>
<thead>
<tr>
<th>Description of model parameters</th>
<th>( f )</th>
<th>( f )</th>
<th>( \Phi )</th>
<th>( \Phi_{\text{V}} )</th>
<th>( \Phi_{\text{MT}} )</th>
<th>( \gamma )</th>
<th>( R )</th>
<th>( \chi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f ) Input image</td>
<td>Multi-scale contrast feature map of the input image ( \hat{I}^\ell )</td>
<td>Feature map containing edge information extracted using Laplacian of Gaussian ( \nabla^2 )G</td>
<td>Final saliency map of the input image</td>
<td>Contains features such as edges, lines, and bars, as well as contours</td>
<td>Contains high-order structures such as motion and direction selectivity features</td>
<td>Zooming factor</td>
<td>The matrix for extracting the regions of overlap between the target and previously reconstructed patches</td>
<td>Vector of sparse coefficients (optimal value)</td>
</tr>
</tbody>
</table>
closely constrain the representation coefficients $\chi$ in terms of $\Phi$ with previously computed HR patches, the optimization problem in Eq. (23) can be modified as
\[
\min_{\chi} \| \chi \|_1 \quad s.t. \quad \| f_{p} - f_{p}^o \Phi X^0 \|_2^2 + \| f_{p} + \tilde{f}_{p} - \tilde{f}_{p}^o \Phi X^0 \|_2^2 < \epsilon_1 \\
\| \Phi^o - \phi^o \|_2^2 + \| \Phi^o + \Phi^o - \phi^o \|_2^2 < \epsilon_2.
\]
(24)
where $\Phi$ is the matrix that extracts the regions of overlap between the target and previously reconstructed patches, and $\phi^o$ and $\phi^o$ contain the values on the overlap of the previous LR and reconstructed HR patches, and enforce the representations common in $f_{p}^o$ and $f_{p}^o$. The constrained $\ell_1$-regularized LS problem in Eq. (24) can be simplified by reformulating as
\[
\min_{\chi} \| Q_{p}^0 f_{p}^o - \tilde{Q}_{p}^0 \Phi X^0 \|_2^2 + \| Q_{p}^0 f_{p}^o - \tilde{Q}_{p}^0 \Phi X^0 \|_2^2 + \| \Phi \|_1 \\
\]
where $Q_{p}^o = \left[ \begin{array}{c} f_{p}^o \\ \phi^o \end{array} \right]$, $\tilde{Q}_{p}^0 = \left[ \begin{array}{c} f_{p}^o \\ \phi^o \end{array} \right]$, $Q_{p}^0 = \left[ \begin{array}{c} f_{p}^o \\ \phi^o \end{array} \right]$, $\tilde{Q}_{p}^0 = \left[ \begin{array}{c} f_{p}^o \\ \phi^o \end{array} \right]$. (25)

The optimal solution $\chi^*$ can be obtained by solving Eq. (25), and this optimal vector of sparse representation coefficients $\chi$ is then used to determine the corresponding HR patch $f_{p}^o$ by multiplying it with $\Phi$
\[
f_{p}^o = \Phi \chi^*.
\]
(26)
The reconstructed patch $f_{p}^o$ is placed in its corresponding position in the output image. The entire reconstruction process of an HR image is summarized in Algorithm 1, and a pictorial representation is provided in Fig. 2.

**Algorithm 1. Reconstruction of HR image**

1. Input: $f^0$, $\Phi$ = $\begin{bmatrix} \Phi_{v1} \\ \Phi_{Mr} \end{bmatrix}$
2. Compute the saliency map $f_{s}$ of $f^0$
   a. Compute multi-scale contrast map $f_{s}$, using Eq. (15)
   b. Compute the Laplacian of Gaussian map $f_{s}$, using Eq. (16)
   c. Compute $f_{s}$ by fusing $f_{s}$ and $f_{s}$, using Eq. (19)
3. Set HR $f^o = 0$, scale up $f^0$ to $f^o$ by factor $\gamma$ (the same size as $f^0$) using classic bicubic interpolation
4. Build feature maps $f^0$ and $\tilde{f}^o$ by filtering $f^0$ and $\tilde{f}^o$ using Eqs. (20) and (21), respectively
5. Extract corresponding patches $f_{p}^0$, $\tilde{f}_{p}^0$, $f_{p}^o$, $\tilde{f}_{p}^o$, and $f_{p}^o$, $\tilde{f}_{p}^o$, respectively
6. Repeat for each patch of $f_{p}^0$ and $f_{p}^o$ (corresponding patches)
   1. If the extracted patch $f_{p}^0$ is salient according to the saliency map $f_{s}$
   1.1. Solve the optimization problem using Eq. (25) for the optimal solution $\chi^*$
   1.2. Recover HR patch $f_{p}^o$ using Eq. (26)
   1.3. Add HR patch $f_{p}^o$ to the corresponding position in $f^o$
   2. Else
   2.1. $f_{p}^o = \tilde{f}_{p}^o$
   2.2. Add HR patch $f_{p}^o$ to the corresponding position in $f^o$
   3. End If
7. Until (the last patch)
8. Output: HR image $f^o$

As mentioned earlier, $\Phi_{Mr}$ contains high-order structures that are more selective for the speed and direction of a moving pattern. Cadieu and Olshausen [20] proved experimentally that such type of image representations describing high-order structures are more effective in removing noises that occur from camera or object motions. Eq. (25) can be extended to a video sequence by exploiting the temporal dimension. Let $f_{p}^o$ and $f_{p}^o$ represent the input LR and interpolated image sequence, respectively, with index $t$ for the sequence, and $f_{p}^o$ and $f_{p}^o$ are their corresponding extracted patches. In addition, $(i,j,t)$ in the index indicates the location in sequence $t$
\[
\min_{\chi_{ij,t}} \sum_{t} \sum_{ij} \left\| Q_{p}^{0,i,j,t} f_{p}^{0,i,j,t} - \tilde{Q}_{p}^{0,i,j,t} \Phi X_{ij,t} \right\|_2^2 \\
+ \sum_{t} \sum_{ij} \left\| Q_{p}^{1,i,j,t} f_{p}^{1,i,j,t} - \tilde{Q}_{p}^{1,i,j,t} \Phi X_{ij,t} \right\|_2^2 \\
+ \sum_{t} \sum_{ij} \rho_{ij,t} \left\| X_{ij,t} \right\|_1.
\]
(27)

Given the optimal solution $X_{ij,t}$, the HR patch can be reconstructed as
\[
f_{p}^{1,i,j,t} = \Phi X_{ij,t}.
\]
(28)
The reconstructed patch is placed in its corresponding position in the video sequence. The HR reconstruction of video sequence is briefly given in Algorithm 2. Only salient patches are SR reconstructed by solving the minimization problem posed in Eq. (25) in the case of a single image, and the minimization problem posed in Eq. (27) in the case of a video sequence. The remaining patches, which are not salient, are simply interpolated using a classic bicubic interpolation algorithm. To zoom a color image, the YCbCr color space is used. The benefit of a YCbCr color space is based on the features of human visual perception, where the Y component denotes the luma, i.e., brightness; $Cb$ is the blue difference (B − Y); and $Cr$ is the red difference (R − Y). The human eye is more perceptive to the luma as compared to chrominance information. The proposed technique gives less importance to the color information because such signals are smoother than a luma’s signals. Thus, the $Cb$ and $Cr$ components of the source image are magnified using a bicubic interpolation method, and the Y component is magnified through Algorithm 1.

**Algorithm 2. Reconstruction of HR video**

1. Input: $f_{p}^0$, $\Phi$ = $\begin{bmatrix} \Phi_{v1} \\ \Phi_{Mr} \end{bmatrix}$
2. Repeat for each frame in $f_{p}^0$
   2.1. Run Algorithm 1 from step 2 to step 8
3. Until (the last frame)
4. Output: HR video $f^1$

**4. Experimental results**

In this section, the proposed technique is evaluated both qualitatively and quantitatively. For this purpose, various image quality assessment metrics were incorporated with no compromise to any standard schemes. The dictionary $\Phi$ was trained from $10^5$ patches of size $9 \times 9$ sampled independently from a training dataset. The training datasets were downloaded from open access databases [47,48]. Unless explicitly mentioned, the size of the dictionary was maintained as $2^{10}$ in all experiments. Each atom of the dictionary consists of 81 elements. The regularization parameter $\beta$ was taken as $\beta = 0.01 \rho_{\max}$, where the value of $\rho_{\max}$ was computed using
Fig. 3. PSNR of the distorted and reconstructed videos (shown in Table 2) by the Yang, Zeyde and proposed techniques. For the proposed technique, the videos are reconstructed using $\Phi_{V1}$ and $\Phi (\Phi_{V1} + \Phi_{NT}$ combined) separately. The numbers on the radius in each sub-figure are the PSNR, and those on the circumference indicate the frame number.
the formula $\beta > \beta_{\text{max}} = \|2\Phi F^T_2\|_\infty$. To evaluate the proposed technique properly, various types of experiments were conducted. Moreover, the validity of the learned responses possessed by the dictionary used in the proposed SR technique was confirmed through various experiments described in [20,40,41]. Hence, we avoided repeating previously conducted experiments and focus instead on an evaluation of SR. Because our main goal is to analyze the role of a dictionary and sparse coding in the reconstruction of an HR image from its corresponding LR image in the proposed framework, we concentrate solely on a quantitative analysis of the proposed and other state-of-the-art SR schemes. The following sections describe the experimental results and analysis in detail.

4.1. Dataset

This sub-section describes the source of the datasets used for the purpose of evaluation. The images and video sequence were taken from CIPR [63] and Live database [64]. The images downloaded from CIPR were originally Kodak images in Raster file format with a resolution of 768 × 512. The size and format of these images could be adjusted according to the needs of the experiment. Eight video segments were downloaded from [64,65]. The contents of these video sequences consist of natural scenes containing different types of motions. Each video sequence has both an uncompressed (reference video) and corresponding compressed version. The frame size is the same in both versions. The compressed videos were distorted using the MPEG-2 compression. MPEG-2 compression produces a fairly uniform video distortion and/or quality, both spatially and temporally. The bit rate required to compress the recovery strength of each reconstruction SR technique, the compressed videos were distorted using the MPEG-2 compression. MPEG-2 compression produces a fairly uniform video distortion and/or quality, both spatially and temporally. The bit rate required to compress the video sequences for a specified visual quality varied dramatically depending on the content. The compression rates varied from 1 to 2 Mbps, depending on the reference sequence. Both the reference images and their corresponding distorted version were downloaded from the Live database. The images were distorted using JPEG2000, Gaussian blur, and white noise. The set of basis functions ($\Phi_{VI}$ and $\Phi_{MT}$) used in the experiments was trained on natural image sequences obtained from public repositories [47,48].

4.2. Quantitative evaluation

The methods used for image and video quality assessment, and the mechanism of the quantitative analysis, are described in this sub-section. The proposed technique was compared with four others, i.e., a bicubic technique, and methods by Jurio et al. [4] (Jurio), Yang et al. [16] (Yang) and Zeyde et al. [13] (Zeyde). The software packages of Jurio, Yang and Zeyde were obtained from the corresponding authors. For this purpose, various image/video quality assessment metrics were used. The metrics include the peak-signal-to-noise ratio (PSNR), cross-correlation coefficient (CCC), mean absolute error (MAE), and structural similarity index metric (SSIM) [66]. The software for the SSIM can be downloaded from the Live image and video quality database [64].

4.2.1. Robustness to distortion

The proposed technique was compared quantitatively with two other techniques, Yang and Zeyde. Both the Yang and Zeyde techniques can handle image HR and denoising simultaneously. The bicubic and Jurio techniques can only magnify an input image and do not have a denoising property. For these types of algorithms, the input image must be free of noise; otherwise, the algorithms will magnify the noisy data. The proposed technique also has the ability to handle image HR and denoising simultaneously. Therefore, to explore the role of $\Phi_{VI}$ and $\Phi_{MT}$ in SR reconstruction, a series of experiments were performed. In this experiment, we had two versions of each video sequence (a reference version and a distorted version), as shown in Table 2. The bit rate of the distorted video depends on the contents of the video. First, each frame of the distorted video is downsampled by a factor of three. The Yang, Zeyde, and the proposed technique reconstruct the downsampled version of the video at the actual size, and the PSNR is then computed for each SR-reconstructed video. To determine the recovery strength of each reconstruction SR technique, the PSNR is also calculated for the original distorted video with respect to its reference video. This process of downsampling and reconstruction was repeated for all video sequences shown in Table 2. To analyze the contribution of $\Phi_{MT}$ during the proposed SR reconstruction, the video was reconstructed separately using $\Phi_{VI}$ and $\Phi_{MT}$. Fig. 3 shows the PSNR score for all eight videos shown in Table 2. To simplify Fig. 3, the average PSNR of five consecutive frames is given for all video sequences except Mobile & Calendar and Park Run, where the average PSNR is given for every ten successive frames.

In Fig. 3, $\Phi_{VI}$ and $\Phi_{VI} + \Phi_{MT}$ (in the legend) represent the proposed technique incorporating $\Phi_{VI}$ and $\Phi$ ($\Phi_{VI}$ and $\Phi_{MT}$) for SR reconstruction, respectively. To make the analysis results easier to read, the difference in the averaged PSNR of the proposed technique, shown in Table 3, was compared with that from distorted (D), Yang (Y), and Zeyde (Z) sequences. From Fig. 3 and Table 3, it can be observed that the net PSNR scores of the proposed technique (in both cases using $\Phi_{VI}$ and $\Phi$) are better than those of the Yang and Zeyde techniques. This is because the latter two methods use dictionary features such as edges, lines, and bars [12,14], whereas the dictionary used by the proposed scheme contains more effective image representations such as edges, lines, and bars, as well as contours and high-order structures. The hierarchical relationship between $\Phi_{VI}$ and $\Phi_{MT}$ not only improves the reconstruction process of an HR image but also simultaneously conducts denoising of the input image. It was observed that when there is little or no motion in the video samples, the Yang and Zeyde technique perform well; however, when large motion are present these methods reduce the power in high-spatial frequencies.

For further analysis of the SR reconstruction process for a single image, ten reference images and their corresponding compressed (JPEG2000) and blurred (Gaussian) counterparts were downloaded from [36]. The distorted images were downsampled by a factor of

<table>
<thead>
<tr>
<th>Video no.</th>
<th>Video name</th>
<th>Video description</th>
<th>fps</th>
<th># of Frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Blue sky</td>
<td>Circular camera motion showing a blue sky and some trees</td>
<td>25</td>
<td>217</td>
</tr>
<tr>
<td>2</td>
<td>Mobile and calendar</td>
<td>Camera pan, toy train moving horizontally with a calendar moving vertically in the background</td>
<td>50</td>
<td>500</td>
</tr>
<tr>
<td>3</td>
<td>Park run</td>
<td>Camera pan, a person running across a park</td>
<td>50</td>
<td>500</td>
</tr>
<tr>
<td>4</td>
<td>Pedestrian area</td>
<td>Still camera, shows some people walking about in a street intersection</td>
<td>25</td>
<td>250</td>
</tr>
<tr>
<td>5</td>
<td>River bed</td>
<td>Still camera, shows river bed containing some pebbles and water</td>
<td>25</td>
<td>250</td>
</tr>
<tr>
<td>6</td>
<td>Rush hour</td>
<td>Still camera, shows rush hour traffic on a street</td>
<td>25</td>
<td>250</td>
</tr>
<tr>
<td>7</td>
<td>Sunflower</td>
<td>Still camera, shows a bee moving over a sunflower in close-up</td>
<td>25</td>
<td>250</td>
</tr>
<tr>
<td>8</td>
<td>Tractor</td>
<td>Camera pan, shows a tractor moving across some fields</td>
<td>25</td>
<td>250</td>
</tr>
</tbody>
</table>
three and then reconstructed to the original size using the proposed, Yang, and Zeyde algorithms. Their PSNR and SSIM are computed with respect to the corresponding reference image.

Table 3
The difference in the average PSNR of the proposed technique from that of the distorted (D), Yang (Y), and Zeyde (Z) sequences in both cases, i.e., using only $\phi_{1}$ or $\phi$.

<table>
<thead>
<tr>
<th>Video name</th>
<th>$\phi_{1} - D$</th>
<th>$\phi_{1} + \phi_{2} - D$</th>
<th>$\phi_{2} - Y$</th>
<th>$\phi_{2} + \phi_{3} - Y$</th>
<th>$\phi_{3} - Z$</th>
<th>$\phi_{3} + \phi_{4} - Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile and calendar</td>
<td>9.50948</td>
<td>10.9171</td>
<td>3.61048</td>
<td>5.02097</td>
<td>4.65108</td>
<td>6.06157</td>
</tr>
<tr>
<td>Park run</td>
<td>9.14748</td>
<td>10.5079</td>
<td>3.56871</td>
<td>5.62181</td>
<td>5.01190</td>
<td>6.34886</td>
</tr>
<tr>
<td>Pedestrian area</td>
<td>11.0792</td>
<td>12.4887</td>
<td>2.90947</td>
<td>4.71713</td>
<td>3.45426</td>
<td>5.37498</td>
</tr>
<tr>
<td>Tractor</td>
<td>7.51169</td>
<td>9.22217</td>
<td>2.91032</td>
<td>4.52106</td>
<td>3.63438</td>
<td>4.94485</td>
</tr>
</tbody>
</table>

In Table 4, the PSNR and SSIM scores demonstrate that the proposed technique is superior to the Yang and Zeyde techniques. They also prove that the proposed technique is more robust to noise.

Table 4
PSNR and SSIM scores for the Yang, Zeyde, and proposed technique on images compressed using JPEG2000.

<table>
<thead>
<tr>
<th>Image</th>
<th>Bit-rate</th>
<th>PSNR</th>
<th>SSIM</th>
<th>PSNR</th>
<th>SSIM</th>
<th>PSNR</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building-1</td>
<td>0.40141</td>
<td>24.17876</td>
<td>0.780724</td>
<td>23.47425</td>
<td>0.76523</td>
<td>27.40601</td>
<td>0.880724</td>
</tr>
<tr>
<td>Building-2</td>
<td>0.37358</td>
<td>24.01148</td>
<td>0.767836</td>
<td>23.3398</td>
<td>0.75333</td>
<td>27.21285</td>
<td>0.867836</td>
</tr>
<tr>
<td>Bike-1</td>
<td>0.6557</td>
<td>27.83947</td>
<td>0.842067</td>
<td>26.38577</td>
<td>0.82886</td>
<td>31.19042</td>
<td>0.919067</td>
</tr>
<tr>
<td>Bike-2</td>
<td>0.33249</td>
<td>25.85407</td>
<td>0.764125</td>
<td>25.17301</td>
<td>0.75572</td>
<td>29.05392</td>
<td>0.884125</td>
</tr>
<tr>
<td>Cap-1</td>
<td>0.60354</td>
<td>33.92342</td>
<td>0.832471</td>
<td>32.8112</td>
<td>0.82995</td>
<td>36.85884</td>
<td>0.97832</td>
</tr>
<tr>
<td>Cap-2</td>
<td>0.40137</td>
<td>30.75253</td>
<td>0.87332</td>
<td>29.19566</td>
<td>0.85108</td>
<td>34.36709</td>
<td>0.91847</td>
</tr>
<tr>
<td>LightHouse-1</td>
<td>0.65051</td>
<td>29.12919</td>
<td>0.872668</td>
<td>28.02496</td>
<td>0.86649</td>
<td>32.52831</td>
<td>0.959668</td>
</tr>
<tr>
<td>LightHouse-2</td>
<td>0.24231</td>
<td>27.16303</td>
<td>0.809581</td>
<td>26.388</td>
<td>0.80160</td>
<td>30.45121</td>
<td>0.899581</td>
</tr>
<tr>
<td>Monarch-1</td>
<td>0.4277</td>
<td>33.5162</td>
<td>0.863242</td>
<td>31.917</td>
<td>0.86641</td>
<td>36.92586</td>
<td>0.948242</td>
</tr>
<tr>
<td>Monarch-2</td>
<td>0.20004</td>
<td>30.64123</td>
<td>0.825856</td>
<td>29.6221</td>
<td>0.82954</td>
<td>33.97663</td>
<td>0.909856</td>
</tr>
<tr>
<td>CoinInFountain-1</td>
<td>0.70372</td>
<td>29.13472</td>
<td>0.842656</td>
<td>28.09905</td>
<td>0.83807</td>
<td>32.49539</td>
<td>0.918565</td>
</tr>
<tr>
<td>CoinInFountain-2</td>
<td>0.18741</td>
<td>24.8941</td>
<td>0.714426</td>
<td>23.67831</td>
<td>0.70378</td>
<td>28.02332</td>
<td>0.838426</td>
</tr>
<tr>
<td>Parrots-1</td>
<td>0.68851</td>
<td>33.67944</td>
<td>0.878253</td>
<td>32.6849</td>
<td>0.87048</td>
<td>37.90917</td>
<td>0.953251</td>
</tr>
<tr>
<td>Parrots-2</td>
<td>0.38193</td>
<td>32.93861</td>
<td>0.848398</td>
<td>31.27949</td>
<td>0.83134</td>
<td>35.87835</td>
<td>0.941398</td>
</tr>
<tr>
<td>Statue-1</td>
<td>2.4112</td>
<td>36.15753</td>
<td>0.904121</td>
<td>34.60228</td>
<td>0.89092</td>
<td>38.54257</td>
<td>0.994121</td>
</tr>
<tr>
<td>Statue-2</td>
<td>0.37665</td>
<td>29.61748</td>
<td>0.850371</td>
<td>28.50945</td>
<td>0.84250</td>
<td>33.02288</td>
<td>0.957371</td>
</tr>
<tr>
<td>Sailing-1</td>
<td>0.22253</td>
<td>28.15178</td>
<td>0.81749</td>
<td>27.5328</td>
<td>0.80675</td>
<td>32.52603</td>
<td>0.930493</td>
</tr>
<tr>
<td>Sailing-2</td>
<td>0.23477</td>
<td>30.67541</td>
<td>0.837474</td>
<td>29.88482</td>
<td>0.83731</td>
<td>34.03032</td>
<td>0.919474</td>
</tr>
<tr>
<td>WomanHat-1</td>
<td>0.96524</td>
<td>32.00304</td>
<td>0.906204</td>
<td>30.69551</td>
<td>0.89302</td>
<td>35.85227</td>
<td>0.988204</td>
</tr>
<tr>
<td>WomanHat-2</td>
<td>0.23561</td>
<td>30.59589</td>
<td>0.814401</td>
<td>30.07272</td>
<td>0.80021</td>
<td>34.92353</td>
<td>0.929401</td>
</tr>
</tbody>
</table>
that the later techniques. The same is true for Table 5, where two images of each corresponding reference image filtered using a circular-symmetric 2-D Gaussian kernel of standard deviation are shown in the Sigma (σ) column.

### 4.2.2. Preservation of original details

In this section, we analyze how well the proposed technique preserves the original details of an actual signal during SR reconstruction. For this purpose, all reference images provided by [56,57,64] were guaranteed to be distortion free. Further, the reference images were downsampled by a factor of three. These downsampled images were then magnified to the original size using the bicubic, Jurio, Yang, Zeyde, and proposed algorithms. This procedure was repeated for CCC, PSNR, MAE, and SSIM. CCC was computed to evaluate how close the magnified information of the input image is to the original information of the reference image. Table 6 shows the scores of CCC for ten images with respect to the corresponding SR technique. The statistics of CCC in Table 6 indicate that the computed frequencies of a zoomed image using the proposed technique have a higher correlation with the original details than the other techniques listed in the table. For most of the images, the value for CCC of the proposed technique is greater than that of the other techniques. The Jurio technique has the lowest score and the Yang and bicubic techniques have good scores when compared to the Zeyde technique. Table 7 shows the PSNR for the underlying techniques. From the table, the ratio between the strength of the maximum achievable power of the signal and the corrupting noise can be evaluated. The maximum value of the PSNR indicates the high quality of the magnified signal. The results of the proposed technique indicate a higher maximum achievable power of the signal than the other techniques. The Jurio technique scores indicate that the ratio of noise corruption is significantly higher than in the other techniques. However, the bicubic and Yang achieve higher quality signals than the Zeyde technique.

### Table 5

<table>
<thead>
<tr>
<th>Image</th>
<th>Sigma (σ)</th>
<th>Yang</th>
<th>Zeyde</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>SSIM</td>
<td></td>
<td>PSNR</td>
</tr>
<tr>
<td>Building-1</td>
<td>0.848926</td>
<td>24.2208</td>
<td>0.842107</td>
<td>23.12862</td>
</tr>
<tr>
<td>Building-2</td>
<td>1.536428</td>
<td>21.8914</td>
<td>0.758629</td>
<td>21.55698</td>
</tr>
<tr>
<td>Bike-1</td>
<td>0.906218</td>
<td>26.7385</td>
<td>0.840102</td>
<td>25.39175</td>
</tr>
<tr>
<td>Bike-2</td>
<td>1.307261</td>
<td>24.8207</td>
<td>0.787110</td>
<td>22.22783</td>
</tr>
<tr>
<td>Cap-1</td>
<td>0.677051</td>
<td>35.5150</td>
<td>0.870892</td>
<td>33.92519</td>
</tr>
<tr>
<td>Cap-2</td>
<td>1.164031</td>
<td>32.9366</td>
<td>0.839773</td>
<td>32.35698</td>
</tr>
<tr>
<td>LightHouse-1</td>
<td>0.82026</td>
<td>28.1898</td>
<td>0.873283</td>
<td>27.04732</td>
</tr>
<tr>
<td>LightHouse-2</td>
<td>1.135385</td>
<td>26.9271</td>
<td>0.861157</td>
<td>26.21606</td>
</tr>
<tr>
<td>Monarch-1</td>
<td>0.906218</td>
<td>32.1939</td>
<td>0.895363</td>
<td>30.48154</td>
</tr>
<tr>
<td>Monarch-2</td>
<td>1.708303</td>
<td>27.6096</td>
<td>0.864912</td>
<td>27.33656</td>
</tr>
<tr>
<td>CoinInFountain-1</td>
<td>0.848926</td>
<td>29.7130</td>
<td>0.76068</td>
<td>28.38463</td>
</tr>
<tr>
<td>CoinInFountain-2</td>
<td>1.45049</td>
<td>27.1567</td>
<td>0.820963</td>
<td>26.68496</td>
</tr>
<tr>
<td>Paras-1</td>
<td>0.791634</td>
<td>35.7125</td>
<td>0.888896</td>
<td>34.03466</td>
</tr>
<tr>
<td>Paras-2</td>
<td>1.249969</td>
<td>29.7634</td>
<td>0.858868</td>
<td>28.9004</td>
</tr>
<tr>
<td>Statue-1</td>
<td>0.848926</td>
<td>31.8697</td>
<td>0.886896</td>
<td>30.73821</td>
</tr>
<tr>
<td>Statue-2</td>
<td>1.307261</td>
<td>27.8487</td>
<td>0.812063</td>
<td>27.39967</td>
</tr>
<tr>
<td>Sailing-1</td>
<td>0.562467</td>
<td>33.2998</td>
<td>0.928218</td>
<td>31.68116</td>
</tr>
<tr>
<td>Sailing-2</td>
<td>1.45049</td>
<td>28.6678</td>
<td>0.856839</td>
<td>28.34424</td>
</tr>
<tr>
<td>WomanHat-1</td>
<td>0.677051</td>
<td>34.4057</td>
<td>0.890153</td>
<td>32.98634</td>
</tr>
<tr>
<td>WomanHat-2</td>
<td>1.020802</td>
<td>32.6322</td>
<td>0.884091</td>
<td>31.87632</td>
</tr>
</tbody>
</table>

### Table 6

<table>
<thead>
<tr>
<th>Image</th>
<th>Bicubic</th>
<th>Jurio</th>
<th>Yang</th>
<th>Zeyde</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building</td>
<td>3.195663</td>
<td>5.373438</td>
<td>1.991189</td>
<td>2.945663</td>
<td>1.009916</td>
</tr>
<tr>
<td>Bike</td>
<td>3.329294</td>
<td>5.464486</td>
<td>2.318773</td>
<td>3.079294</td>
<td>1.639777</td>
</tr>
<tr>
<td>Cap</td>
<td>2.412308</td>
<td>4.350063</td>
<td>1.806136</td>
<td>2.162308</td>
<td>0.959761</td>
</tr>
<tr>
<td>LightHouse</td>
<td>4.779636</td>
<td>7.073468</td>
<td>3.433398</td>
<td>4.529636</td>
<td>2.291525</td>
</tr>
<tr>
<td>Monarch</td>
<td>3.088682</td>
<td>5.108215</td>
<td>2.339282</td>
<td>2.838682</td>
<td>1.104276</td>
</tr>
<tr>
<td>Sailing</td>
<td>2.61908</td>
<td>4.598388</td>
<td>1.990448</td>
<td>2.36908</td>
<td>1.028857</td>
</tr>
<tr>
<td>WomanHat</td>
<td>5.931369</td>
<td>9.095435</td>
<td>4.690041</td>
<td>5.683169</td>
<td>2.269465</td>
</tr>
<tr>
<td>Bike</td>
<td>34.39708</td>
<td>22.38751</td>
<td>34.9495</td>
<td>34.40784</td>
<td>35.93031</td>
</tr>
<tr>
<td>Cap</td>
<td>28.33095</td>
<td>21.24695</td>
<td>30.39309</td>
<td>27.53387</td>
<td>31.12308</td>
</tr>
<tr>
<td>Monarch</td>
<td>31.9138</td>
<td>22.04602</td>
<td>33.35651</td>
<td>31.59265</td>
<td>33.39847</td>
</tr>
<tr>
<td>Statue</td>
<td>32.13976</td>
<td>23.91937</td>
<td>33.83238</td>
<td>31.05037</td>
<td>34.14534</td>
</tr>
<tr>
<td>Sailing</td>
<td>32.01882</td>
<td>21.03861</td>
<td>32.65177</td>
<td>30.77636</td>
<td>34.28225</td>
</tr>
<tr>
<td>WomanHat</td>
<td>33.98338</td>
<td>22.74622</td>
<td>34.44351</td>
<td>32.98908</td>
<td>35.18281</td>
</tr>
</tbody>
</table>

### Table 7

<table>
<thead>
<tr>
<th>Image</th>
<th>Bicubic</th>
<th>Jurio</th>
<th>Yang</th>
<th>Zeyde</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Building</td>
<td>24.26533</td>
<td>20.39248</td>
<td>25.66298</td>
<td>23.41782</td>
<td>27.60124</td>
</tr>
<tr>
<td>Bike</td>
<td>27.3073</td>
<td>23.55273</td>
<td>28.62331</td>
<td>26.12456</td>
<td>30.27111</td>
</tr>
<tr>
<td>Cap</td>
<td>34.39708</td>
<td>22.38751</td>
<td>34.9495</td>
<td>34.40784</td>
<td>35.93031</td>
</tr>
<tr>
<td>LightHouse</td>
<td>28.33095</td>
<td>21.24695</td>
<td>30.39309</td>
<td>27.53387</td>
<td>31.12308</td>
</tr>
<tr>
<td>Monarch</td>
<td>31.9138</td>
<td>22.04602</td>
<td>33.35651</td>
<td>31.59265</td>
<td>33.39847</td>
</tr>
<tr>
<td>Statue</td>
<td>32.13976</td>
<td>23.91937</td>
<td>33.83238</td>
<td>31.05037</td>
<td>34.14534</td>
</tr>
<tr>
<td>Sailing</td>
<td>32.01882</td>
<td>21.03861</td>
<td>32.65177</td>
<td>30.77636</td>
<td>34.28225</td>
</tr>
<tr>
<td>WomanHat</td>
<td>33.98338</td>
<td>22.74622</td>
<td>34.44351</td>
<td>32.98908</td>
<td>35.18281</td>
</tr>
</tbody>
</table>
technique produces a large signal error during the zooming process. The Yang technique produces fewer errors than the bicubic or Zeyde technique.

Under the assumption that human visual perception is highly adapted for extracting structural information from a scene [66], we evaluate the quality of an SR reconstructed image based on the degradation of the structural information. During the SR process, some of the structural information of the input image that could have become distorted tends to be ignored by the previous evaluation metrics. For this purpose, we compute the SSIM to measure the degree of degradation of the structural information in a reconstructed image. The SSIM statistics in Table 9 indicate that, to a great extent, the proposed technique retains the structural information of the reconstructed image distortion free. The Yang technique had the second-highest SSIM scores, and the Jurio technique again had the lowest. The Zeyde scheme had a slightly greater score than the bicubic technique. All of the quantitative analyses conducted above prove that the proposed technique preserves the original detail of the input image, and that the recovered signal is close to the original signal of the reference image.

4.3. Execution time measurement

It is also important to evaluate the efficiency of the proposed technique with respect to saliency map and dictionary size. To assess the role of the saliency map, we apply thresholding to the signal is close to the original signal of the reference image.

The percentages of reduction of the PSNR and SSIM vary from 0.05% to 0.04% and 0.045% to 0.095% (less than 1%), respectively. Fig. 4 shows three images and their corresponding saliency maps. It is clear from Fig. 4 that the regions of high contrast and rich textures were highlighted through the use of the saliency map. The time was computed in seconds for all experiments describe herein. In addition, all of the images used in this experiment were distortion free.

To evaluate the tradeoff between time and quality, we tested the reconstruction process by incorporating dictionaries of different sizes. We also analyzed the impact of the dictionary size on the proposed technique by computing the PSNR and SSIM of the recovered images. The size and zooming factor of the input images were kept the same as in the previous experiment. The proposed technique was also evaluated computationally with the Yang and Zeyde techniques. The PSNR and SSIM scores are not given for the Yang and Zeyde techniques because both these techniques were already evaluated for the same purpose. The size of the dictionary is directly proportional to the execution time of the proposed technique. However, for larger dictionaries, high PSNR and SSIM score are produced, as shown in Table 11. The computational time of the proposed technique is approximately two- to three-times less than in the Zeyde and Yang techniques. The Yang technique has a longer computational time for many reasons. For example, it uses back projection to satisfy the reconstruction constraints, and does not use a saliency map to limit the reconstruction process to the salient regions. The Zeyde technique does not use back projection or a saliency map. Both the Yang and Zeyde techniques use a joint-dictionary (one dictionary each for the LR and HR feature spaces) concept to recover a latent signal \( \mathbf{f}_0 \) by providing an observed signal \( \mathbf{I}_0 \). They infer the sparse representation of \( \mathbf{f}_0 \) in the LR feature space with respect to the LR dictionary, and use this sparse representation to recover the corresponding HR patch \( \mathbf{I}_0 \) from the HR dictionary. Both dictionaries contain the same sparse representation [13,16]. The joint dictionaries used by the Yang and Zeyde techniques are not indeed customized to each individual space, and cannot capture the possible complex relationships between different signal spaces arising under various scenarios [36,56].

Moreover, in SR reconstruction, the computational complexity of the minimization solver dominates the total computational complexity of the SR reconstruction algorithm. An efficient minimization solver can make the SR technique more efficient. OMP is a greedy approach to solving the \( \ell_0 \)-minimization for a sparse recovery. Zeyde et al. used OMP for sparse recovery. Yang et al. employed \( \ell_1 \)-regularized linear regression, known in statistics as a Lasso [33], for a sparse representation recovery. In the proposed SR reconstruction, we solve the problem posed in Eq. (25), which is an \( \ell_1 \)-regularized LS using the fast solution of \( \ell_1 \)-minimization of Donoho et al. [21,22]. The latter two are considered to be in the category of BP; however, their solvers are different. It is generally believed that BP algorithms can produce more precise solutions.
than OMP, but usually require a higher time complexity \[67\]. The computational complexities in asymptotic notations (i.e., big O) of the mentioned solvers are described in detail in \[22,29,33,37,68\]. Therefore, we only concentrate on the specification-dependent execution time for the overall framework of SR reconstruction which is shown in Tables 10–12. The term specification dependent indicates that the execution time varies according to the changes brought about by the specifications of the system upon which the algorithm is being tested.

4.4. Impact of back projection

In \[10,16,26,57]\, a back projection procedure was applied to satisfy the reconstruction constraint, in which \(I^n\) is projected on to the

<table>
<thead>
<tr>
<th>Image</th>
<th>PSNR</th>
<th>SSIM</th>
<th>Execution time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peppers</td>
<td>26.871</td>
<td>0.913</td>
<td>6.72</td>
</tr>
<tr>
<td>House</td>
<td>31.035</td>
<td>0.935</td>
<td>8.83</td>
</tr>
<tr>
<td>Lena</td>
<td>27.482</td>
<td>0.938</td>
<td>7.19</td>
</tr>
<tr>
<td>Zebra</td>
<td>30.618</td>
<td>0.920</td>
<td>7.75</td>
</tr>
<tr>
<td>Athens</td>
<td>24.370</td>
<td>0.839</td>
<td>8.01</td>
</tr>
</tbody>
</table>

Table 11
Impact of dictionary sizes, and computational assessment of the proposed technique.

<table>
<thead>
<tr>
<th>Image</th>
<th>Proposed</th>
<th>Yang</th>
<th>Zeyde</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peppers</td>
<td>29.411</td>
<td>30.635</td>
<td>26.37</td>
</tr>
<tr>
<td>House</td>
<td>33.284</td>
<td>34.51</td>
<td>29.45</td>
</tr>
<tr>
<td>Lena</td>
<td>30.793</td>
<td>31.19</td>
<td>28.90</td>
</tr>
<tr>
<td>Zebra</td>
<td>33.251</td>
<td>33.98</td>
<td>29.01</td>
</tr>
<tr>
<td>Athens</td>
<td>28.262</td>
<td>28.90</td>
<td>29.83</td>
</tr>
</tbody>
</table>

Table 12
PSNR and SSIM with and without a back projection procedure.
solution space \((l^1 + f_{DS} + \eta) = l^0\) by solving the following optimization problem:

\[
\hat{l}^1 = \arg\min_{l^1} \|l^1 - f_{DS}\| \quad \text{s.t.} \quad (l^1 + f_{DS} + \eta) = l^0.
\]

The proposed algorithm was evaluated by incorporating the back projection procedure. The results of the SSIM and PSNR show that such a back projection procedure is not needed. In fact, it only increases the execution cost by removing some minor distortion, i.e., insignificant artifacts. However, it has a slight effect on the structural information of the reconstruction information. The statistics in Table 12 indicate that back projection only increases the cost of the algorithm. It also affects the structural information and introduces minute errors into the magnified signals. The downsampling and reconstruction factors were kept the same as those described in the previous section.

### 4.5. Qualitative analysis

The visual quality of the proposed technique was evaluated through a comparison with the bicubic, Jurio, Zeyde, and Yang techniques. For this purpose, the seven parameters of Wittman [1] were considered to assess the visual quality. The parameters are as follows: (1) Geometric invariance: The reconstruction technique must preserve the relative size and shape of the objects in the input image. (2) Contrast invariance: The luminance information of the object and the overall contrast of the image must be preserved. (3) Noise: The magnification technique must have the ability to keep the input image free from noise and bad artifacts such as ringing near the borders. (4) Edge preservation: The boundaries and edges must be preserved. (5) Aliasing: Edges must be free from jaggedness and staircases. (6) Texture preservation: Texture regions must not be blurred or smoothed by the reconstruction process. Finally, (7) Over-smoothing: The method should not produce undesirable piecewise-constant or block regions. A qualitative evaluation is shown in Fig. 5. The first column shows the distorted images with the cropped part enclosed by the red rectangle. The distorted cropped part is shown above the corresponding distorted image in the first column, and the original cropped part (cropped from the reference image) is shown below. They are cropped from the same location. These distorted and original cropped parts are magnified at 4×. Each column is labeled based on the corresponding technique. The monarch image compressed using JPEG2000 (at a bit rate: 0.042317) was taken from the Live database [64]. The size of the cropped parts taken from the compressed and original Monarch images is 99 × 56. The reconstructed versions of these cropped parts indicate that the bicubic and Jurio techniques completely fail to distinguish a signal from noise in the case of signal distortion, and magnify the distorted signal as it is. However, the proposed technique recovers the original signal to a greater extent, and has good visual quality compared to the Zeyde and Yang techniques, that is, the wing veins and white dots on the wings are close to those of the original magnified cropped image. When we compare the distorted versions of Monarch, it is evident that the visual quality is improved in ascending order from the bicubic technique to the proposed technique. The Bike image was blurred by motion noise. A motion filter (15 pixels, and 25 degree angle) was convolved with the Bike image in the counterclockwise direction. The 85 × 53 sized images were cropped from the distorted and reference image of Bike and magnified 4×. The spokes, cassette, dropout, and upper parts of the tire of the magnified images were analyzed. The visual results show that the proposed technique reconstructs the spokes, cassette, dropout, and ribs of the tire to a greater extent, and comparatively better than the Yang and Zeyde techniques. However, the bicubic and Jurio schemes fail in such recovery. The Jurio technique oversharpens the input image. The same procedure was adopted for the Parrots image and was convolved using Gaussian filter (7 radius, and 7 standard deviation) to obtain the distorted version. The cropped parts of the Parrots image are 63 × 47 in size. In Fig. 5, the visual quality shows that the proposed technique covers up the seven parameters of Wittman [1] to a greater extent than the other schemes.

### 5. Conclusion

This work presented a novel framework for image SR reconstruction, leading to a state-of-the-art performance, surpassing other well-known recently published algorithms. The proposed technique recovers the latent signal of the corresponding observed signal using an overcomplete basis set based on the effective image representations. The sets of basis functions are highly selective to spatial variations and velocity. The learning process of the proposed method enforces the reconstruction of an HR patch from the corresponding LR patch by combining the sparse linear combinations of the feature descriptors (signal atoms) from a single two layered dictionary according to the found representation coefficients. The algorithm also becomes more efficient by limiting the optimization process to visually salient regions of the input image using a saliency map. The superiority claimed by the proposed technique was proved in the experimental section through an extensive quantitative and qualitative analysis. In future work, the proposed technique may be made more efficient for video SR and deblurring. The dictionary used for the SR purpose in the proposed scheme can be elaborated to incorporate more types of features descriptors according to other layers of human visual perception for greater robustness to various types of distortion.

### Acknowledgments

This research is supported by MSIP (Ministry of Science, ICT & Future Planning), South Korea in the ICT R&D Program 2014 (R0112-14-1014).

### References


J. Friedman, motionGallery, 2013.


R. Center, Redwood center for theoretical neurosciences (2013).

MotionGallery, motionGallery, 2013.