

## Research Article

# Multiscale and Multitopic Sparse Representation for Multisensor Infrared Image Superresolution

Xiaomin Yang,<sup>1</sup> Kai Liu,<sup>2</sup> Zhongliang Gan,<sup>3</sup> and Binyu Yan<sup>1</sup>

<sup>1</sup>College of Electronics and Information Engineering, University of Sichuan, Chengdu, Sichuan 610064, China

<sup>2</sup>College of Electrical and Engineering Information, University of Sichuan, Chengdu, Sichuan 610064, China

<sup>3</sup>College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Correspondence should be addressed to Kai Liu; [kailiu@scu.edu.cn](mailto:kailiu@scu.edu.cn)

Received 17 March 2015; Revised 4 June 2015; Accepted 9 June 2015

Academic Editor: Marco Anisetti

Copyright © 2016 Xiaomin Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Methods based on sparse coding have been successfully used in single-image superresolution (SR) reconstruction. However, the traditional sparse representation-based SR image reconstruction for infrared (IR) images usually suffers from three problems. First, IR images always lack detailed information. Second, a traditional sparse dictionary is learned from patches with a fixed size, which may not capture the exact information of the images and may ignore the fact that images naturally come at different scales in many cases. Finally, traditional sparse dictionary learning methods aim at learning a universal and overcomplete dictionary. However, many different local structural patterns exist. One dictionary is inadequate in capturing all of the different structures. We propose a novel IR image SR method to overcome these problems. First, we combine the information from multisensors to improve the resolution of the IR image. Then, we use multiscale patches to represent the image in a more efficient manner. Finally, we partition the natural images into documents and group such documents to determine the inherent topics and to learn the sparse dictionary of each topic. Extensive experiments validate that using the proposed method yields better results in terms of quantitation and visual perception than many state-of-the-art algorithms.

## 1. Introduction

High-resolution (HR) infrared (IR) images are desired in various electronic imaging applications, such as medical diagnosis, criminal investigation, surveillance, remote sensing, and aerospace. However, given the inherent limitation of relevant imaging devices or other factors, obtaining images at a desired resolution is difficult. Therefore, many efforts have been devoted to improving the spatial resolution of the IR image. Superresolution (SR) is one of the most promising methods in the research community.

At present, a large number of SR methods have been developed successfully. The existing methods for image SR can be divided into three general categories: interpolation-based methods [1, 2], reconstruction-based methods [3–6], and learning-based methods [7–11].

The interpolation-based [1, 2] scheme applies the correlation of neighboring image pixels to approximate the fundamental HR pixels. These types of methods can be easily implemented at a high speed. However, these methods may lead to the loss of detailed information.

Reconstruction-based approaches utilize additional information from low-resolution (LR) images to synthesize an HR image. These approaches are ill-posed estimation problems and require a priori information on images to regularize the solution. Therefore, various regularization methods have been proposed to improve the performance of SR reconstruction, such as the projection on convex sets [3], maximum a posteriori (MAP) [4, 5], and regularization-based method [6]. Compared with interpolation-based schemes, the reconstruction-based methods deliver better performance with a small desired magnification factor. However,

the most common defect of multiframe SR reconstruction is that, with an increase in the magnification factor, the LR inputs cannot provide sufficient information to maintain a high-quality SR reconstruction result.

Learning-based methods presume that the high-frequency details lost in the LR image can be predicted by learning the cooccurrence relationship between LR training patches and their corresponding HR patches. Freeman et al. [7] first introduced the learning idea for SR reconstruction, which uses a Markov random field model to learn the relationship between local regions of images and their underlying scenes. Various effective tools have been proposed to learn prior information, such as neighbor embedding- (NE-) based methods [8, 12], regression-based methods [9, 10], and sparse coding- (SC-) based methods [11, 13–15]. The NE-based methods estimate each desired HR image patch by linearly combining its neighbor training HR image patches. Chang et al. [12] introduced locally linear embedding from manifold learning to process the image SR task. Zhang et al. [8] proposed a partially supervised NE method. However, given the lack of prior textures and details, NE-based methods are weak in visualizing textures and details. The regression-based methods directly estimate the desired HR pixels using some complicated statistical models. Wang and Tang [9] proposed a principal component analysis-based SR reconstruction method to estimate the desired HR image. Wu et al. [10] used the kernel partial least squares regression model to handle the one-to-many mapping problem. Wu's method requires searching the neighbors in the entire training database and using the same number of principal components to synthesize the desired HR feature patches, which result in high computational costs. The SC-based SR method can better retain the most relevant reconstruction neighbors and can restore more image information than the two learning-based methods discussed above. Yang et al. [16] proposed an approach based on sparse representation, with the assumption that the HR and LR images share the same set of sparse coefficients. Therefore, the HR image can be reconstructed by combining the trained HR dictionary and the sparse coefficients of the corresponding LR image.

The abovementioned SC-based SR methods always suffer from three problems. First, due to the inherent limitation of relevant imaging devices or other factors, IR images always lack detailed information, which leads to unsatisfied IR image reconstruction results. Multiple images acquired by different sensors provide complementary information on the same scene. As such, a reasonable method of improving the resolution of the IR image is the combination of inherently complementary information from the images obtained from different multisensors. Second, a traditional sparse dictionary is learned from patches with a fixed size, which cannot capture the exact information of the images. However, the local structures of an image tend to repeat themselves many times with some similar neighbors across the natural images, not only within the same scale but also across different scales. Details missing in a local structure at a smaller scale can be estimated from similar patches at a larger scale. Different images prefer different patch sizes for optimal representation. Therefore, jointly representing an image at different scales is

important. Considering the above cues, we propose a model of obtaining multiscale patches to learn dictionaries. We use a simple model that generates pyramid images and divides such images into multiscale patches. Finally, given that dictionary learning is a key issue of the sparse representation model, considerable effort in learning dictionaries from example image patches has been exerted, leading to state-of-the-art results in image reconstruction. Many dictionary learning methods aim at learning a universal and overcomplete dictionary that represents various image structures. However, for natural images, a large number of different local structural patterns exist. The contents can vary significantly across different images or different patches in a single image. One dictionary is inadequate in capturing all of the different structures. Multiple dictionaries [15, 17] are more effective in representing various contents in an image and provide better reconstruction results than one universal dictionary [15]. Based on these observations, training patches are categorized into multiple groups based on visual characteristics in our algorithm. A subdictionary is then learned in the respective data groups. Unsuitable training sample groups used in dictionary learning lead to artifacts in example learning-based methods [18]. In this study, we group the patches into several categories. Each category corresponds to a topic. We apply the probabilistic latent semantic analysis (pLSA) model [19] to group the patches and to determine the inherent topics. That means we group the patches into several categories. Each category corresponds to a topic. We then learn the sparse dictionary for each topic. Our framework treats each group individually, thereby leading to more accurate distribution dictionaries. We conduct semantic analysis on a given patch to categorize it to a topic. The given patch can be better represented by the selected topic subdictionary. Thus, the entire image can be more accurately reconstructed using this method than using a universal dictionary, as validated by our experiments.

In summary, this study makes the following three main contributions: (1) IR images always lack detailed information. Meanwhile, VI images contain abundant object edges and details, providing a more perceptual description of a scene for human eyes. This study combines the inherently complementary information from images obtained from different multisensors to improve the resolution of the IR image. (2) To learn the sparse dictionary for representing similar redundancies of local patterns within the same scale and across different scales, this study builds pyramid images downsampled from the images. Then it divides the pyramid images into multiscale patches, thereby representing the image in a more efficient manner and providing a more global look of the image. (3) The pLSA model is applied to group the patches by determining the inherent topics and to group the training patches with similar patterns. Each dictionary is learned from some type of example patches with the same topic, and multiple dictionaries are learned simultaneously. Extensive experimental results show that our proposed method achieves competitive performance compared to state-of-the-art methods.

The remainder of this paper is organized as follows: Section 2 presents the details of the proposed approach. Section 3 reports the experimental results. Section 4 discusses the conclusion.

## 2. The Proposed SR Scheme

This study proposes a novel sparse representation algorithm, which aims to combine the information of visible images, provide a more global look of the IR image, and simultaneously utilize the inherent topics of IR images in a unified framework. The proposed method can be divided into three steps: (a) combining the information of images from multisensors, (b) obtaining multiscale patches, and (c) learning multitopic sparse dictionaries. In combining the information of visible images, our framework improves the resolution of the IR image when learning the LR sparse dictionary. In obtaining multiscale patches, we build pyramid images and extract multiscale patches from such images, which can provide a more global look of the images. In presenting different structural patterns more accurately, we partition the natural images into documents and group them to determine the inherent topics using the pLSA. A compact subdictionary can then be learned for each topic.

*2.1. Combining the Information of Multisensors.* Given an observed LR IR image  $Y$ , which is a downsampled and blurred version of the HR image  $X$  of the same scene, we derive the following equation:

$$Y = HBX, \quad (1)$$

where  $H$  denotes a downsampling operator and  $B$  is a blurring filter. The goal of a single-image SR is to reconstruct the HR image  $\widehat{X}$  from the LR image  $Y$  as accurately as possible.

With the LR image  $Y$ ,  $y^i$  is the set of patch features extracted from  $Y$ :

$$y^i = \text{Extract}(Y), \quad (2)$$

where  $\text{Extract}(\cdot)$  is an operator that extracts the feature of patch  $i$  from image  $Y$ .

Recent works [16, 20] indicate that derivative features can represent the patch more efficiently than the actual intensities. The derivative features are obtained using four 1D filters:

$$\begin{aligned} f_1 &= [-1, 0, 1], \\ f_2 &= f_1^T, \\ f_3 &= [1, 0, -2, 0, 1], \\ f_4 &= f_3^T. \end{aligned} \quad (3)$$

Images acquired by multisensors provide complementary information on the same scene. IR images always lack detailed information. Meanwhile, VI images contain abundant object edges and details, providing a more perceptual description of a scene for human eyes [21]. As such, combining the detailed information in visible images to improve the resolution of the IR image is reasonable; that is, the information of an LR IR image and the information of the corresponding HR visible image are used for reconstructing an HR IR image.

Applying these four filters, we obtain four description feature vectors for each patch of the LR IR image and its corresponding HR visible image, which are concatenated as one vector in the final gradient representation of the LR patch. The information of the LR IR image and the information of its corresponding HR visible image are combined together to learn the LR sparse dictionary.

With the sparse generative model, each patch feature  $y^i$  ( $y^i \in \mathfrak{R}^n$ ) can be projected over the LR dictionary  $D_l \in \mathfrak{R}^{n \times k}$ , which characterizes the LR patches. This projection produces a sparse representation of  $y^i$  via  $\alpha_l$ , expressed as follows:

$$y^i = D_l \alpha_l \quad y^i \in \mathfrak{R}^n, \quad \|\alpha_l\|_0 \ll k, \quad (4)$$

where  $\alpha_l$  denotes sparse representation atoms. For the HR IR image, high-frequency information is obtained to present the HR patch. The corresponding HR patch feature  $x^i$  has  $x^i \in \mathfrak{R}^m$  sets of patch features extracted from the HR image  $X$  obtained as follows:

$$x^i = \text{Extract}(X). \quad (5)$$

Reapplying the sparse generative model, we have

$$x^i = D_h \alpha_h, \quad (6)$$

where  $D_h \in \mathfrak{R}^{m \times k}$  is the HR dictionary that characterizes the HR patches and is coupled with  $D_l$  through the relation  $D_l = HBD_h$ . This relation indicates that each atom in  $D_h$  has its corresponding LR version in  $D_l$  and vice versa. We assume that the sparse representation of an LR patch in terms of  $D_l$  can be directly used to recover the corresponding HR patch from  $D_h$ ; namely,  $\alpha_l = \alpha_h$ . The process of Sparse representation-based SR by combining the information of visible images is described in Figure 1.

As such, the reconstructed HR image  $\widehat{X}$  can be built by applying the sparse representation in each  $y^i$  and then using the estimated  $\alpha_l$  with  $D_h$  to obtain each  $\widehat{x}^i$ , which together form the image  $\widehat{X}$ .

The SC is clearly a bridge between the LR and HR patches. The dictionaries  $D_l$  and  $D_h$  have a key role in generating such SC. The dictionaries  $D_l$  and  $D_h$  can be easily generated from a set of samples using algorithms, such as K-SVD [11] and efficient SC [13, 14, 17, 22].

*2.2. Obtaining Multiscale Patches.* It is observed that different images prefer different patch sizes for optimal performance [13]. Reference [13] even observed the oversmoothing of artifacts when using unsuitable patches. An explanation for this phenomenon is that dictionary learning from patches with a fixed size cannot capture the exact information of the images. One size of the sample patches corresponds to one scale. However, selecting the exact patch size of the image is difficult. As such, having a multiscale dictionary avoids selecting the patch size in advance. A multiscale treatment can help represent the image in a more efficient manner. In our proposed multiscale framework, we focus on simultaneously obtaining the multiscale patches. First, pyramid

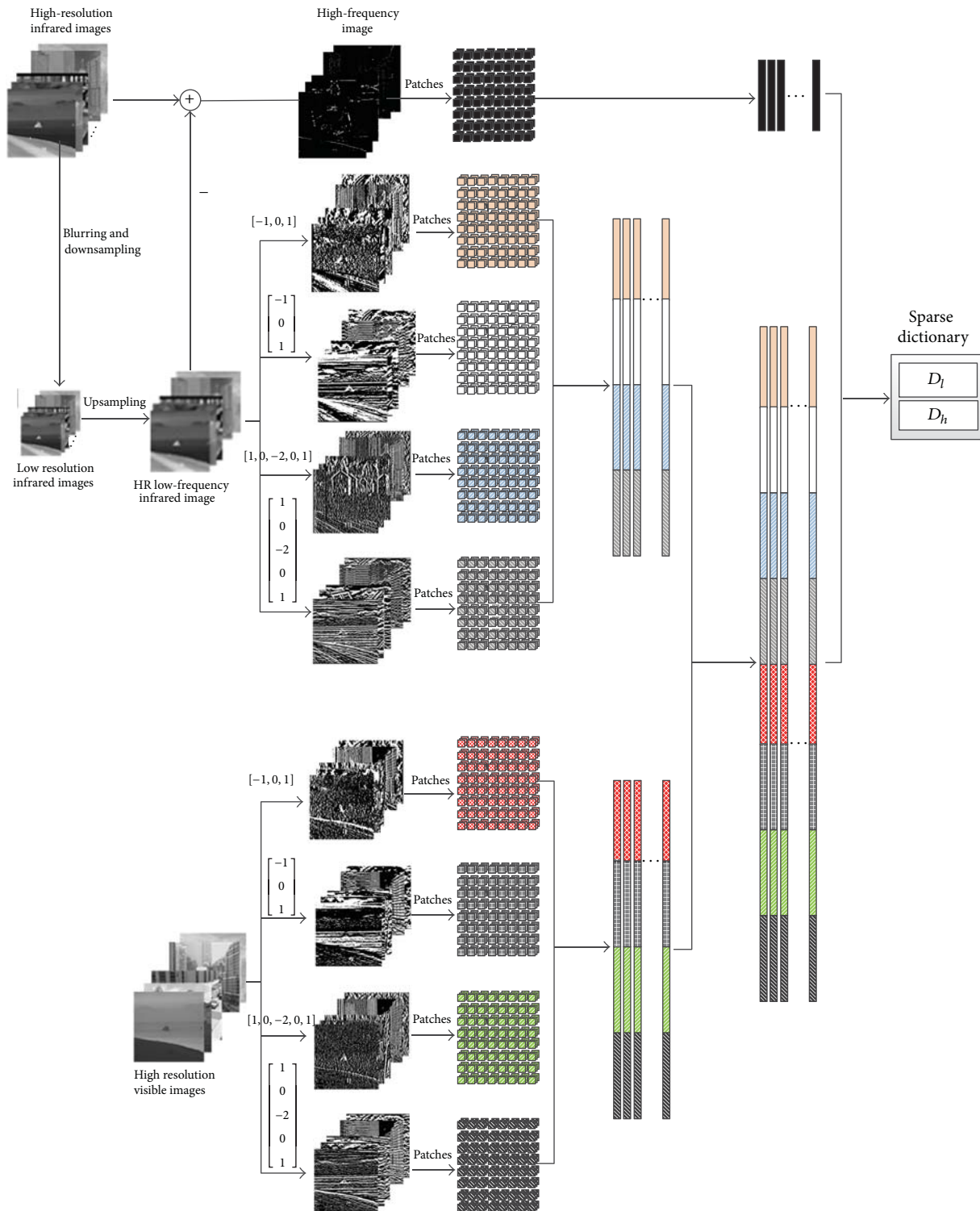


FIGURE 1: Sparse representation-based SR by combining the information of visible images.

images downsampled from the images are built to learn the sparse dictionary for representing similar redundancies of local patterns within the same scale and across different scales. Second, multiscale patches from the pyramid images are then extracted.

Pyramid transform is an effective multiresolution analysis approach. During pyramid transform, each pixel in the low spatial pyramid is obtained by downsampling from its adjacent low-pass filtered HR image. Sequential pyramid images are constructed, as shown in Figure 2. Pyramid images



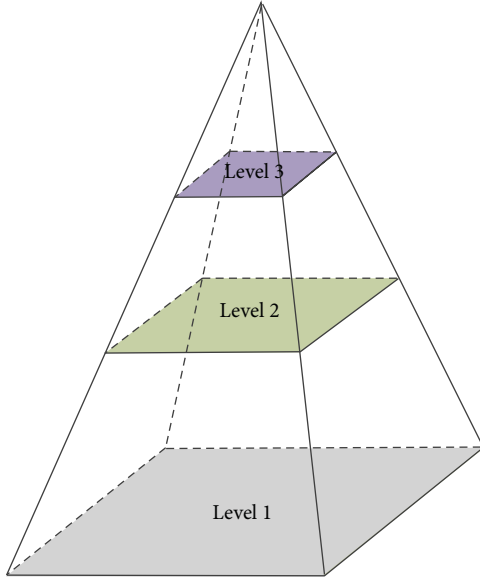


FIGURE 2: A three-level spatial pyramid.

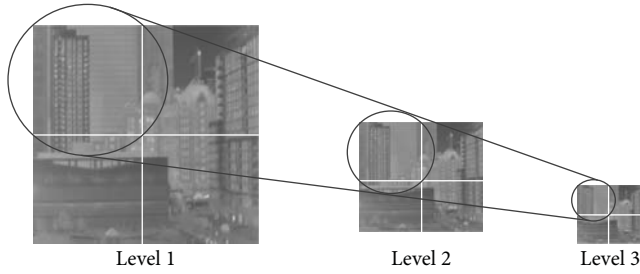


FIGURE 3: Gaussian pyramid images. The original image (level 1) and its Gaussian pyramids are shown from left to right.

can be generated by Gaussian smooth filtering, as shown in Figure 3.

Let  $I(x, y)$  denote the original image. The downsampled version  $I_p(x, y)$  at the  $p$ th level is obtained by convoluting  $I(x, y)$  with a Gaussian kernel  $G(x, y, p\sigma)$ , as follows:

$$G(x, y, p\sigma) = \frac{1}{2\pi p^2 \sigma^2} \exp\left(-\frac{x^2 + y^2}{2p^2 \sigma^2}\right), \quad (7)$$

$$I_p(x, y) = I(x, y) \times G(x, y, p\sigma) \downarrow s_p,$$

where  $\downarrow s_p$  denotes the downsampling operator, with the factor  $s_p = (0.5)^p$  at the  $p$ th level.

After generating the pyramid images, we use the quadtree model [15] to extract multiscale patches from the pyramid images, as shown in Figure 4. We consider a set of large root patches of size  $\sqrt{N} \times \sqrt{N}$  extracted from the sequential pyramid images. The root patch is then divided into subpatches of size  $N_s = N/4^s$  ( $s = 0, \dots, S-1$ ) along the tree, where  $S$  is the depth of the tree. After obtaining multiscale patches, we can learn dictionaries from the patches of different scales. Figure 5 illustrates the process of extracting multiscale patches from the pyramid images.

**2.3. Learning the Multitopic Dictionary.** We partition the natural images into documents and group them to determine the inherent topics using pLSA and to present the different structural patterns more accurately. Each dictionary is learned from some type of example patches with the same topic, and multiple dictionaries are learned simultaneously. The example image patches are classified into many topics by the pLSA model. Given that each topic consists of many patches with similar patterns, a compact subdictionary can be learned for each topic. For an image patch to be coded, the best subdictionary that is most relevant to the given patch is selected. Considering that the given patch can be better represented by the selected subdictionary, the entire image can be more accurately reconstructed than when a universal dictionary is used, as validated by our experiments. The use of multitopic dictionary learning has two main advantages: (1) the training patches are divided into some topics, which ensure that the subdictionary represents the statistical model of the example patches more accurately and (2) the training patches enhance the speed of dictionary learning on each topic and the final reconstruction accuracy through the transfer of knowledge between topics.

**2.3.1. Standard pLSA.** The pLSA [19], which is an extension of LSA [23], provides a probabilistic formulation to model documents in a text collection. The pLSA assumes that the words are generated from a mixture of latent aspects, which can be decomposed from a document. The pLSA model has been used successfully in image classification, image retrieval, and image annotation. The pLSA model ignores the orders of words in a document and instead uses the counts of words occurring in a document. We briefly outline the principle of the pLSA in this subsection. More details can be found in [19].

A corpus that contains  $M$  documents is denoted by  $D = \{d_1, \dots, d_M\}$ , and each document  $d_i$  is represented with the count of its words from a vocabulary  $W = \{w_1, \dots, w_N\}$ . The entire corpus is summarized by the  $N \times M$  cooccurrence matrix  $N$ , where each entry  $n(d_i, w_j)$  indicates the count of the word  $w_j$  in the document  $d_i$ . In the framework of the pLSA, the observed word  $W$  is conditionally independent of the document  $D$  given a latent variable  $Z = \{z_1, \dots, z_K\}$ , which is referred to as the “latent aspect.” The graphical model shown in Figure 6(a) illustrates the form of the joint probability of  $P(w_j, d_i, z_k) = P(d_i)P(z_k | d_i)P(w_j | z_k)$  in the pLSA model. The joint probability of the observed variables is obtained by marginalizing over the latent aspect  $z_k$ :

$$P(d_i, w_j) = P(d_i) \sum_{k=1}^K P(z_k | d_i) P(w_j | z_k). \quad (8)$$

Equation (6) expresses each document as a convex combination of  $K$  aspect vectors, which results in matrix decomposition, as shown in Figure 6(b). Each document is essentially modeled as a mixture of aspects, the histogram for a particular document being composed of a mixture of the histograms corresponding to each aspect.

The model parameters of pLSA are the two conditional distributions  $P(z_k | d_i)$  and  $P(w_j | z_k)$ , which are estimated

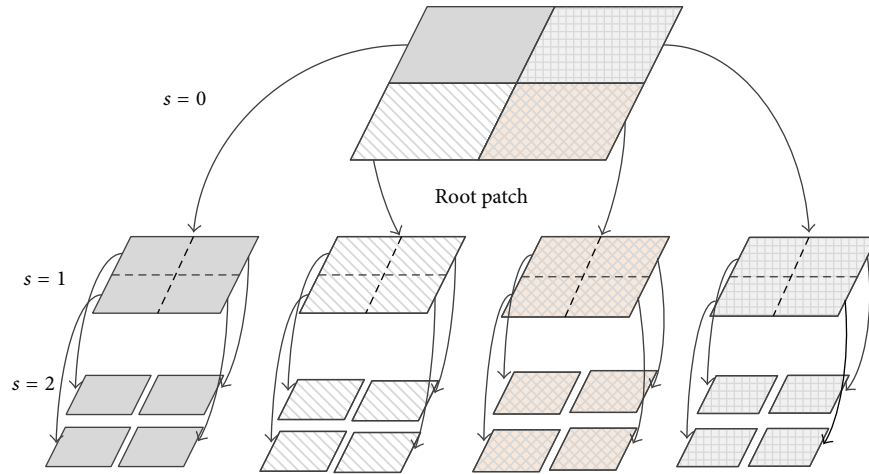


FIGURE 4: Multiscale patches based on the quadtree model.

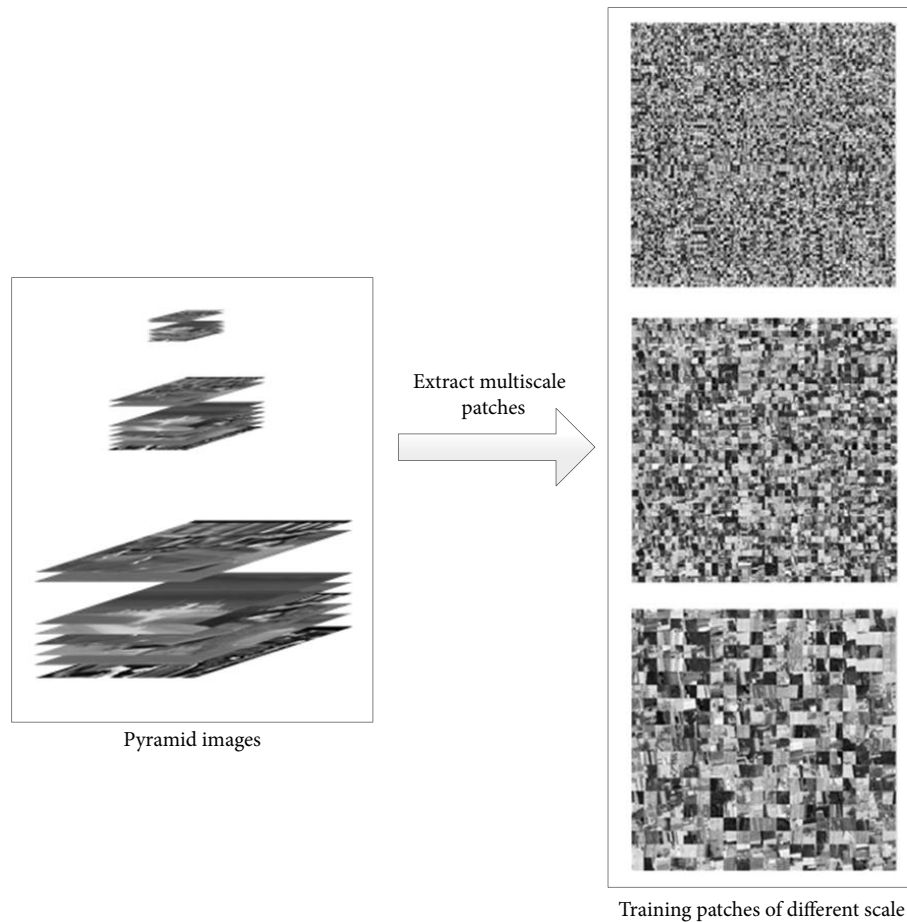


FIGURE 5: Extraction of multiscale patches.

using the expectation-maximization (EM) algorithm on a set of training documents.  $P(w_j | z_k)$  characterizes each aspect and remains valid for documents out of the training set. By contrast,  $P(z_k | d_i)$  is relative only to the specific documents and cannot carry any prior information to an unseen document.

The EM algorithm is used to compute the parameters  $P(w_j | z_k)$  and  $P(z_k | d_i)$  by maximizing the log-likelihood of the observed data:

$$L = \sum_{i=1}^N \sum_{j=1}^M n(d_i, w_j) \log P(d_i, w_j). \quad (9)$$

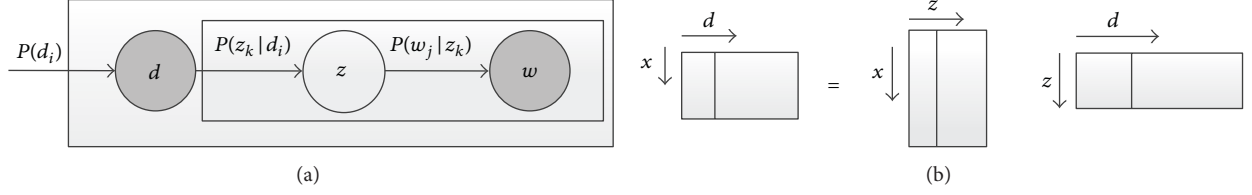


FIGURE 6: (a) Graphical model representation of pLSA; (b) matrix decomposition of conditional distribution.

The steps of the EM algorithm are described as follows:

(1) E-step: the conditional distribution  $P(z_k | d_i, w_j)$  is computed from the previous estimate of the parameters:

$$P(z_k | d_i, w_j) = \frac{P(z_k | d_i) P(w_j | z_k)}{\sum_{l=1}^K P(z_l | d_i) P(w_j | z_l)}. \quad (10)$$

(2) M-step: the parameters  $P(w_j | z_k)$  and  $P(z_k | d_i)$  are updated with the new expected value  $P(z_k | d_i, w_j)$ :

$$P(w_j | z_k) = \frac{\sum_{i=1}^N n(d_i, w_j) P(z_k | d_i, w_j)}{\sum_{m=1}^M \sum_{i=1}^N n(d_i, w_j) P(z_k | d_i, w_m)}, \quad (11)$$

$$P(z_k | d_i) = \frac{\sum_{j=1}^M n(d_i, w_j) P(z_k | d_i, w_j)}{\sum_{j=1}^M n(d_i, w_j)}.$$

**2.3.2. Our Method.** Given a collection of IR images, we intend to determine the inherent topics of the images. We use general terms [24], such as topics, documents, and words, which are mostly used in the text of the literature. In our application, we define the atoms of the sparse dictionary as the “words” of the vocabulary and the sliding window of the sparse dictionary as the “document.” The sliding window consists of patches. Figure 7 shows the  $L \times L$  sliding window (large blue square) and one patch (small red square) in it. All of the documents are grouped by “topic” based on the cooccurrences of different words within and across the documents. Our method has the following five steps: (1) vocabulary formulation, (2) document representation, (3) topic learning, (4) subdictionary construction, and (5) superresolution image reconstruction (SRIR). Our method is illustrated in Figure 8.

**Vocabulary Formulation.** We need to represent each document by a collection of words from a vocabulary. A general sparse dictionary  $D_G$  with  $M$  atoms  $\alpha_m$  ( $m = 1, 2, \dots, M$ ) is learned over all of the patches to construct the vocabulary. Each atom in  $D_G$  is defined as a word of the vocabulary. All of the atoms of  $D_G$  produce the vocabulary for the pLSA model.

**Document Representation.** We assume that document  $d_i$  has  $L$  patches  $p_l$  ( $l = 1, 2, \dots, L$ ). We represent each patch in the document using a linear combination of atom  $\alpha_m$  ( $m = 1, 2, \dots, M$ ) from the general dictionary. We denote the atoms representing patch  $p_l$  as  $\alpha_{l,m}$ . We denote the count of

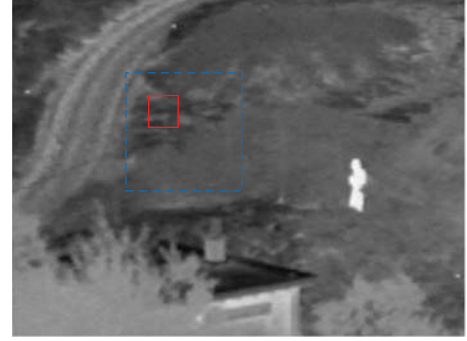


FIGURE 7: Sliding window and patches.

vocabulary  $\alpha_j$  in document  $d_i$  as  $n(w_j, d_i)$ , where  $n(w_j, d_i) = \sum_{l=1}^L \alpha_{l,j}$ . We then use the pLSA model to learn the latent topic of the documents.

**Topic Learning.** All of the documents can be summarized by the  $M \times N$  cooccurrence matrix, where each entry  $n(d_i, w_j)$  indicates the count of the word  $\alpha_j$  in document  $d_i$  and  $N$  is the total number of documents. The EM algorithm is used to compute the parameters  $P(w_j | z_k)$  and  $P(z_k | d_i)$  by maximizing the log-likelihood of the observed data. After learning,  $P(z_k | d_i)$  represents the mixture proportions of each document. The maximum value for each of the document can be assumed as the document topic assignment.

**Subdictionary Construction.** We assume  $K$  determined topics. All of the documents are then classified into  $K$  group Documents = [documents<sub>1</sub>, documents<sub>2</sub>, ..., documents<sub>K</sub>]. For one document group documents<sub>i</sub>, we collect all of the patches that belong to these documents and denote these patches as patches<sub>i</sub>. As such, we can obtain  $K$  group Patches = [patches<sub>1</sub>, patches<sub>2</sub>, ..., patches<sub>K</sub>]. We aim to learn  $K$  compact subdictionaries  $D_k$  ( $k = 1, 2, \dots, K$ ) from Patches. Each of the patches<sub>k</sub> ( $k = 1, 2, \dots, K$ ) is apparently expected to have the same distinctive patterns. We use the SRIR for each group's patches<sub>k</sub> to learn the subdictionary for each topic, such that the most suitable subdictionary for each given local image patch can be selected using the pLSA model.

**SRIR.** We divide the LR image into overlapping documents and the documents into overlapping patches. Then, we represent each document in the same manner as that conducted



FIGURE 8: Illustration of our method.

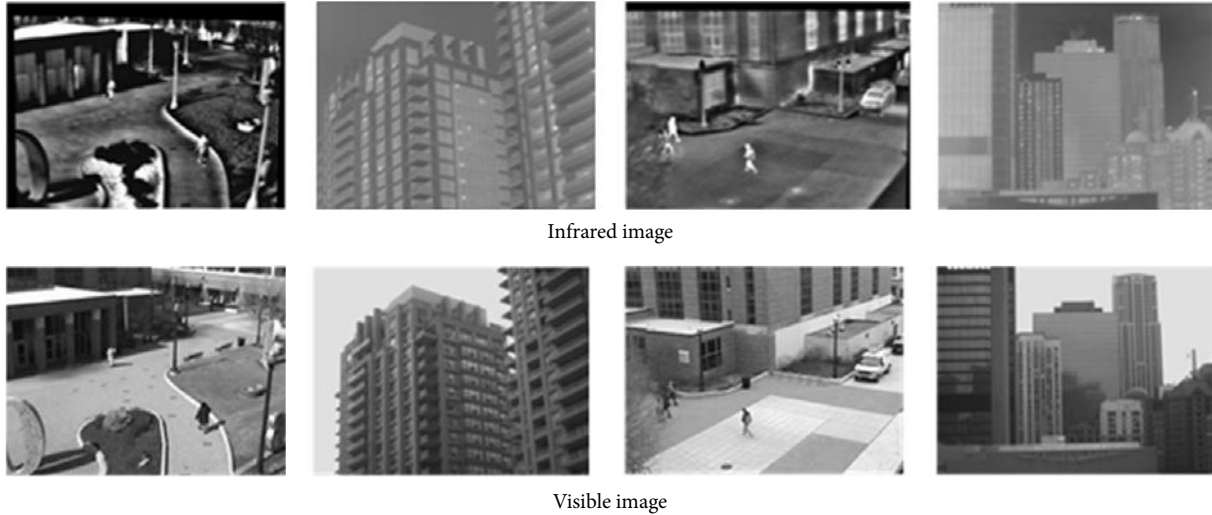


FIGURE 9: Some examples of the infrared images and visible images used in our experiments.

during topic discovery. Each document is analyzed by using the EM algorithm to determine its topic assignment. Each patch of a document is reconstructed by using the topic corresponding to the subdictionary. We do this for all of the documents in the test image and then take the average of all overlapping portions to obtain the reconstructed HR image.

### 3. Experimental Results

**3.1. Samples and Settings.** In our experiments, the IR images and corresponding visible images were obtained from [25] <http://www.dgp.toronto.edu/~nmorris/data/IRData/>. Samples of the training images are shown in Figure 9. The LR images used in all the experiments were downsampled from the HR images. In our experiments, the LR images were generated by shrinking the corresponding HR images with the scale factor of 3.

We employed the peak signal-to-noise ratio (PSNR) and the structural similarity measurement (SSIM) to evaluate the superresolved image and assess the performance of the proposed method. The mean values of the PSNR and SSIM of all of the test images were used as the quality index. The PSNR evaluates the reconstruction quality based on the pixel intensity. The SSIM measures the similarity between two images based on their structural information. The SSIM metric needs a “perfect” reference image for comparison and provides a normalized value between  $[0, 1]$ , where “0” indicates that the two images are totally different, whereas “1” confirms that the two images are the same. Thus, higher values of PSNR and SSIM indicate a result with better quality.

**3.2. Reconstruction Results.** In this section, we conduct several experiments to evaluate the effectiveness of the proposed method.

*Experiment 1* (comparison with the state-of-the-art algorithms). The proposed method was tested using some IR images to validate the effectiveness of the proposed resolution enhancement method in terms of visual fidelity and objective criterion. We compare our algorithm with some well-known image SR algorithms, such as the nearest neighbor, cubic B-spline interpolation method, and Yang’s method [16], to validate the efficiency of our method. In our method, the root patch size is  $16 \times 16$ , the depth of the tree is 3, and the number of training patches in the training process is 100,000. For the multitopic dictionary, the number of atoms in the general dictionary is 1,000. The number of atoms is the same in the multitopic dictionaries. We assume  $K$  determined topics ( $K = 500$ ). For Yang’s method, the number of atoms is 1,000. We present the SR results of images (with a scale factor of 3) obtained using different methods in Figure 10. We extract the region after magnification within the red box to show the details after SR. We observe that the bicubic interpolation method blurs the sharpness of the edges and misses some fine details in the reconstructed images. Yang’s method [16] recovers a significant number of details but produces many jagged and ringing artifacts, along with edges or details. The proposed method obtains better visual quality than all of the other three competing methods.

Moreover, the PSNR and SSIM values of the SR results on LR images using various algorithms are listed in Table 1.



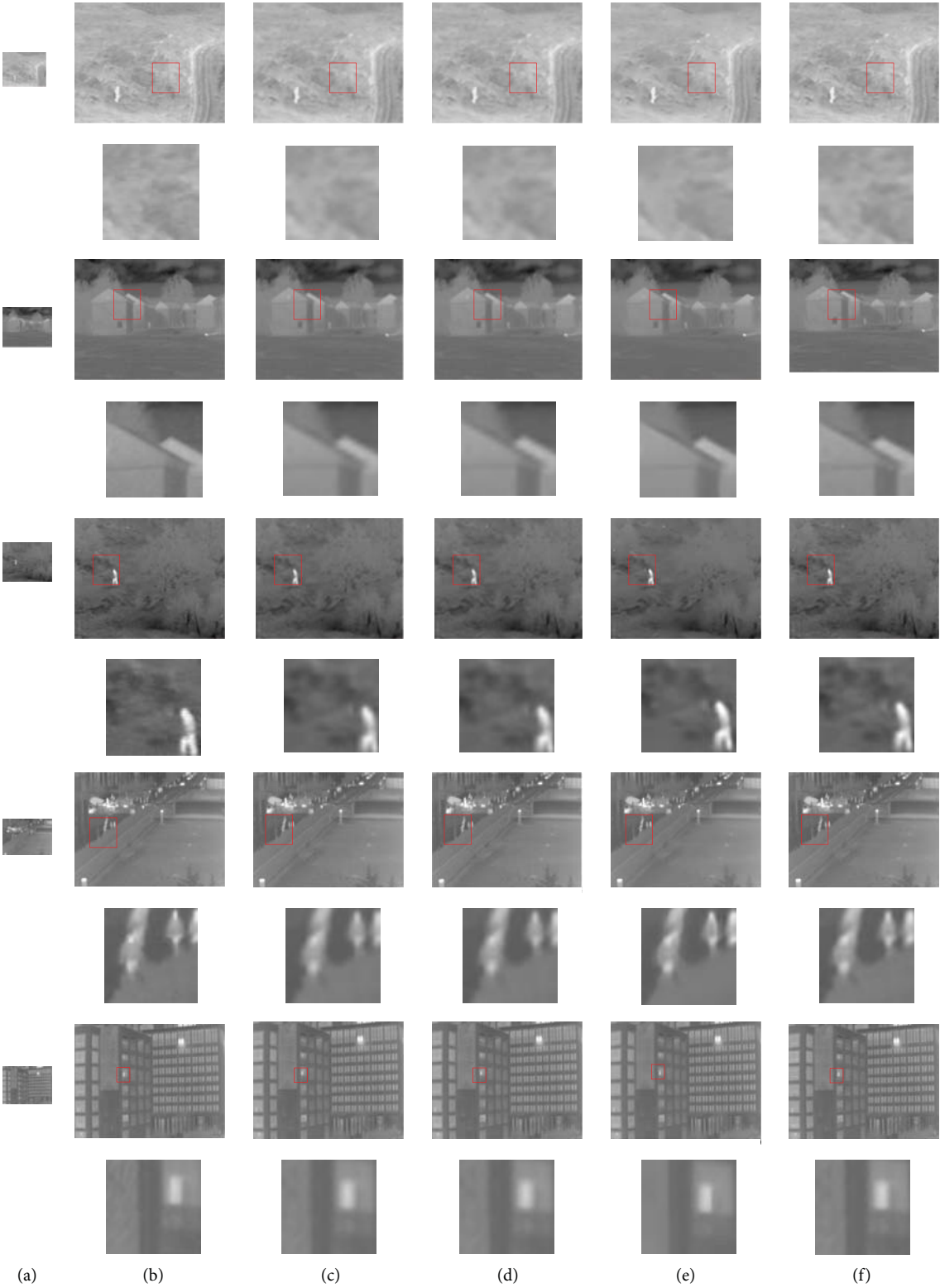
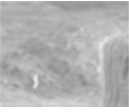






FIGURE 10: Visual comparison of four test images: (a) LR image, (b) original HR image, (c) results obtained using the nearest neighbor interpolation, (d) results obtained using the cubic B-spline interpolation, (e) results obtained using the sparse representation-based method, and (f) results obtained using the proposed method.

TABLE 1: Numerical results of Figure 10.




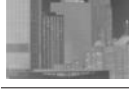
Method	Nearest neighbor		Cubic B-spline		Yang's method		Our method	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
	36.77	0.4440	36.80	0.4458	36.98	0.5394	37.32	0.5857
	38.06	0.7210	38.07	0.7279	38.17	0.7691	38.20	0.8107
	36.19	0.4652	36.19	0.4713	36.63	0.6005	36.68	0.6405
	36.49	0.4933	36.48	0.5007	36.73	0.5477	36.93	0.5964
	32.91	0.3621	32.93	0.3646	34.02	0.3702	34.10	0.3741

We observe that the average PSNR and SSIM gains of the proposed method over Yang's method [16] and the bicubic interpolation method are in dB, which show that the SR results from the proposed method have better objective quality in terms of PSNR and SSIM.

*Experiment 2* (effect of multisensor). To validate the effectiveness of multisensor by combining the information of visible images, we compared multisensor SRIR with traditional SRIR algorithm as Yang's method [16]. The number of training patches in the training process is 100,000. For the dictionary learning step, the number of atoms in the dictionary is 1,000. Figure 11 shows the SR results of the IR image. Figure 11(c) shows the results of the traditional SRIR algorithm as Yang's method [16], where severely jagged artifacts along the edges and annoying details are produced. The SR result is limited. Figure 11(d) shows the results of combining the information of visible images. We observe that the result is significantly improved qualitatively and quantitatively. The PSNR and SSIM values of the SR results on LR images using various algorithms are listed in Table 2.

*Experiment 3* (effect of multiscale patches). We compare the SR results obtained from the dictionaries using multiscale patches and one fixed-scale patch. In the multiscale patches-based method, the root patch size is  $16 \times 16$ . In the fixed-scale patch, fixed patches with three different patch sizes  $4 \times 4$ ,  $8 \times 8$ , and  $16 \times 16$  are analyzed. The number of training patches in the training process is 100,000. For the dictionary learning step, the number of atoms in the dictionary is 1,000. The reconstruction results are shown in Figure 12. We have observed that different images prefer different patch sizes for

TABLE 2: Numerical results of Figure 11.

Method	Yang's method		Multisensor	
	PSNR	SSIM	PSNR	SSIM
	34.28	0.4123	35.53	0.4237
	36.63	0.6005	36.65	0.6247
	36.73	0.5477	36.87	0.5751
	33.07	0.3731	34.57	0.3769

optimal performance. The multiscale treatment can help represent the image in a more efficient manner, thereby allowing applications to provide a more global look of the image. We observe that the reconstructed HR images obtained from the multiscale patches-based method, as shown in Figure 12(f), are better in terms of quantitation and visual perception than those obtained from the single-scale patches-based methods, as shown in Figures 12(c) to 10(e). The PSNR and SSIM values of the SR results on LR images using various algorithms are listed in Table 3.

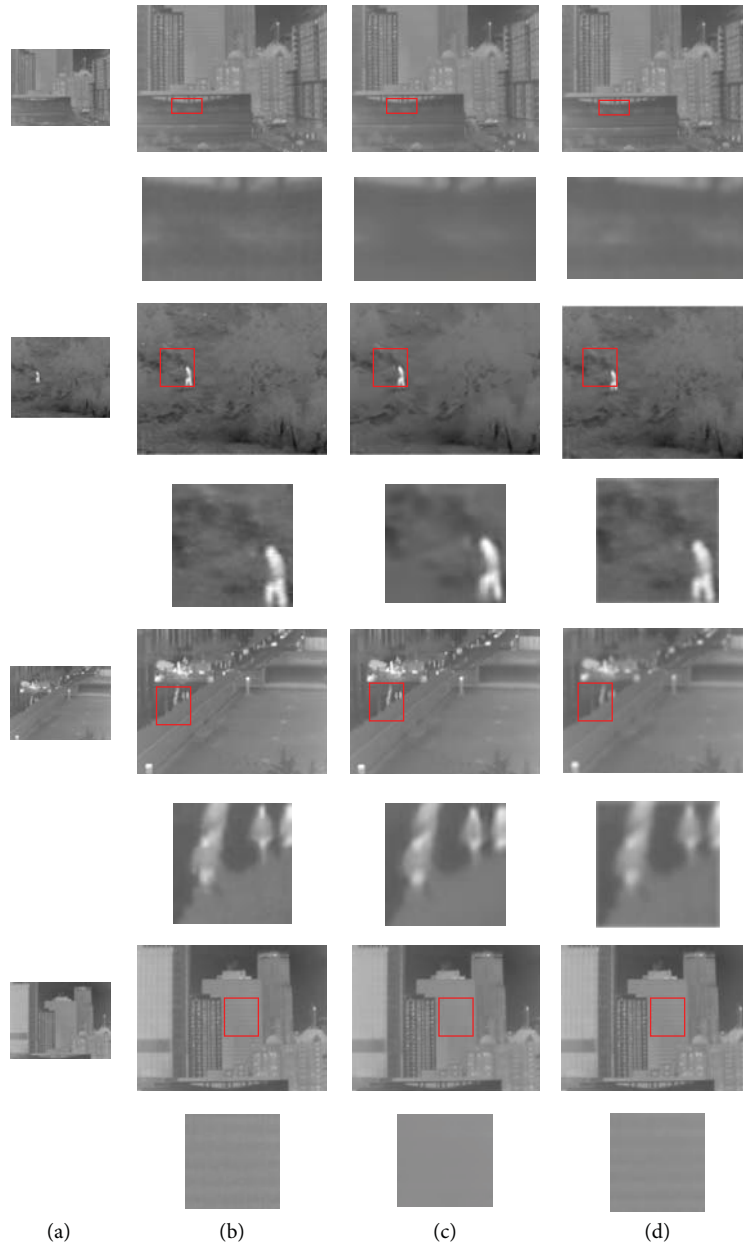


FIGURE 11: Visual comparison of four test images: (a) LR image, (b) original HR image, (c) results obtained using the sparse representation-based method, and (d) results obtained using the multisensor based method.

#### 4. Conclusion

We proposed a novel sparse representation-based image SR method. The algorithm combines detailed information in visible images to improve the resolution of the IR image. Given the complementary nature of these types of information, the proposed method can generate state-of-the-art results in SR tasks. Considering the fact that the optimal sparse domains of natural images can vary significantly across different images and different image patches in a single image, the proposed method uses a simple model that generates pyramid images and divides the pyramid images into multiscale patches to

represent the image in a more efficient manner. We also partition the natural images into documents and group the documents to determine the inherent topics using pLSA and to learn the sparse dictionary of each topic using the sparse dictionary learning technique. Extensive experimental results show that our proposed method can achieve competitive performance compared to state-of-the-art methods.

#### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

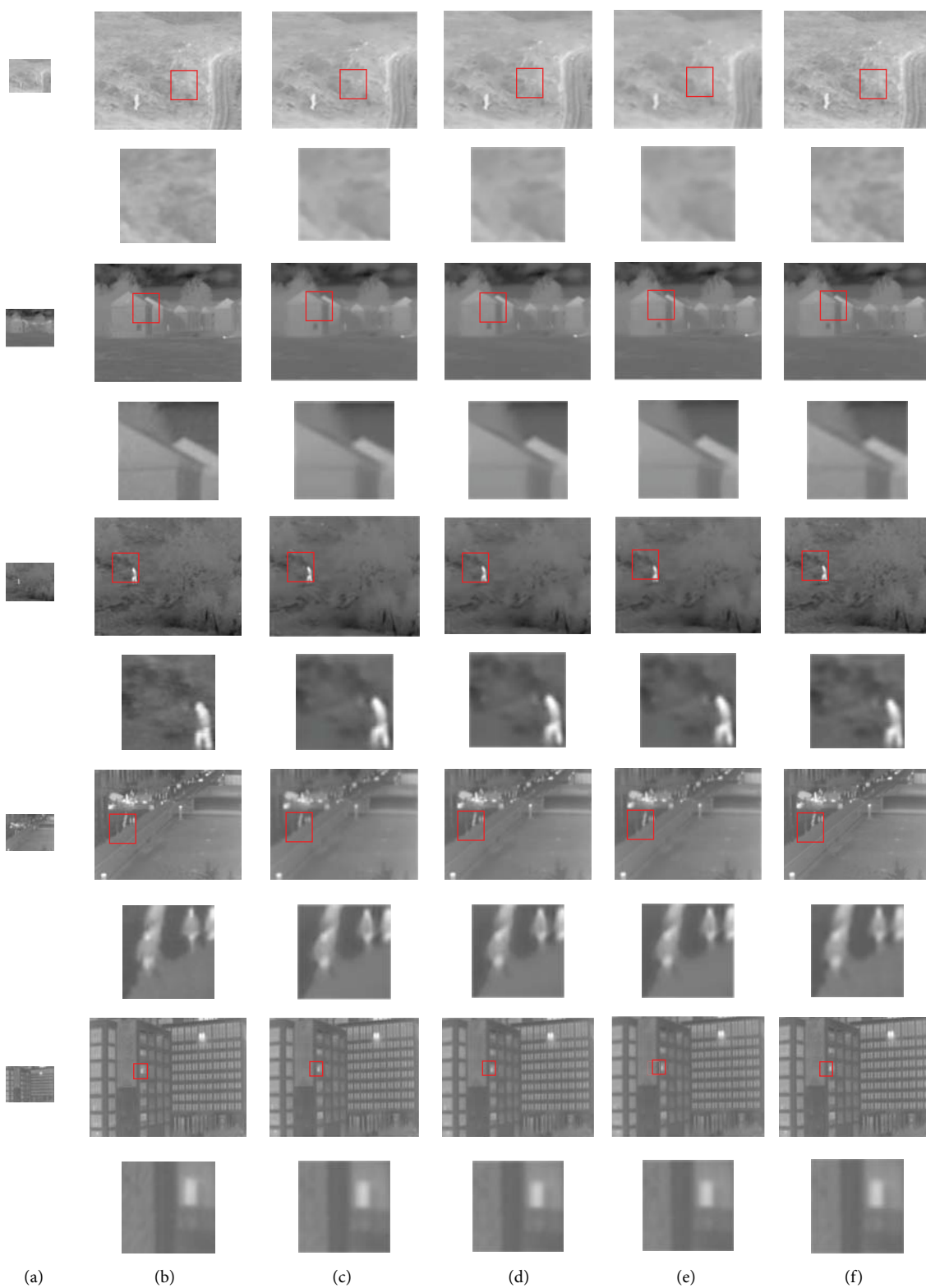
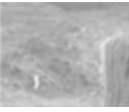






FIGURE 12: Visual comparison of four test images: (a) LR image, (b) original HR image, (c) results obtained with patch size  $4 \times 4$ , (d) results obtained with patch size  $8 \times 8$ , (e) results obtained with patch size  $16 \times 16$ , and (f) results obtained using the proposed method.



TABLE 3: Numerical results of Figure 12.

Method	Patch size $4 \times 4$		Patch size $8 \times 8$		Patch size $16 \times 16$		Multiscale patches	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
	36.96	0.5378	36.93	0.5327	36.92	0.5298	37.27	0.5679
	38.07	0.6247	38.15	0.6538	38.18	0.7721	38.19	0.7894
	36.64	0.6109	36.24	0.5371	36.17	0.4928	36.65	0.6362
	36.51	0.4959	36.72	0.5468	36.57	0.5179	36.89	0.5683
	34.04	0.3723	34.01	0.3703	33.72	0.3674	34.06	0.3726

## Acknowledgments

The research is sponsored by the National Natural Science Foundation of China (nos. 61271330 and 61411140248), the Research Fund for the Doctoral Program of Higher Education (no. 20130181120005), the National Science Foundation for Postdoctoral Scientists of China (no. 2014M552357), the Science and Technology Plan of Sichuan Province (no. 2014GZ0005), and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

## References

- [1] R. Kimmel, "Demosaiicing: image reconstruction from color CCD samples," *IEEE Transactions on Image Processing*, vol. 8, no. 9, pp. 1221–1228, 1999.
- [2] X. Li, "Demosaiicing by successive approximation," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 370–379, 2005.
- [3] T. S. Huang and R. Y. Tsai, "Multi-frame image restoration and registration," *Advances in Computer Vision and Image Processing*, vol. 1, pp. 317–339, 1984.
- [4] H. F. Shen, L. P. Zhang, B. Huang, and P. X. Li, "A MAP approach for joint motion estimation, segmentation, and super resolution," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 479–490, 2007.
- [5] L. P. Zhang, H. Y. Zhang, H. F. Shen, and P. X. Li, "A super-resolution reconstruction algorithm for surveillance images," *Signal Processing*, vol. 90, no. 3, pp. 848–859, 2010.
- [6] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.
- [7] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.
- [8] K. Zhang, X. Gao, X. Li, and D. Tao, "Partially supervised neighbor embedding for example-based image super-resolution," *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 2, pp. 230–239, 2011.
- [9] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 35, no. 3, pp. 425–434, 2005.
- [10] W. Wu, Z. Liu, and X. He, "Learning-based super resolution using kernel partial least squares," *Image and Vision Computing*, vol. 29, no. 6, pp. 394–406, 2011.
- [11] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [12] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 1, pp. 275–282, Washington, DC, USA, June–July 2004.
- [13] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *Multiscale Modeling & Simulation*, vol. 7, no. 1, pp. 214–241, 2008.
- [14] G. Monaci and P. Vanderqheynst, "Learning structured dictionaries for image representation," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '04)*, vol. 4, pp. 2351–2354, IEEE, Singapore, October 2004.
- [15] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces*, vol. 6920 of *Lecture Notes in Computer Science*, pp. 711–730, 2012.
- [16] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

- [17] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.
- [18] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with non-local means and steering kernel regression," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4544–4556, 2012.
- [19] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine Learning*, vol. 42, no. 1-2, pp. 177–196, 2001.
- [20] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.
- [21] W. Wu, X. Yang, Y. Pang, J. Peng, and G. Jeon, "A multifocus image fusion method by using hidden Markov model," *Optics Communications*, vol. 287, no. 1, pp. 63–72, 2013.
- [22] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," in *Proceedings of the 22nd Annual Conference on Neural Information Processing Systems (NIPS '08)*, pp. 1033–1040, December 2008.
- [23] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391–407, 1990.
- [24] P. Purkait and B. Chanda, "Image upscaling using multiple dictionaries of natural image patches," in *Computer Vision—ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5–9, 2012, Revised Selected Papers, Part III*, vol. 7726 of *Lecture Notes in Computer Science*, pp. 284–295, Springer, Berlin, Germany, 2013.
- [25] N. J. W. Morris, S. Avidan, W. Matusik, and H. Pfister, "Statistics of infrared images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR' 07)*, pp. 1–7, June 2007.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

