## *Image fusion via non-local sparse K-SVD dictionary learning*

Li, Ying; Li, Fangyi; Bai, Bendu; Shen, Qiang

*Published in:*
Applied Optics

*DOI:*
10.1364/AO.55.001814

*Publication date:*
2016

*Citation for published version (APA):*
Li, Y., Li, F., Bai, B., & Shen, Q. (2016). Image fusion via non-local sparse K-SVD dictionary learning. *Applied Optics*, *55*(7), 1814-1823. https://doi.org/10.1364/AO.55.001814

# Image fusion via non-local sparse K-SVD dictionary learning

**YING LI [1,*], FANGYI LI [1,3], BENDU BAI[2], QIANG SHEN[3]**

[1] *Shaanxi Provincial Key Lab of Speech and Image Information Processing, School of Computer Science, Northwestern Polytechnical University, Xi'an 710129, China*

[2] *School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China*

[3] *Department of Computer Science, Institute of Mathematics, Physics and Computer Science, Aberystwyth University, SY23 3DB Aberystwyth, U.K*

**Corresponding author: lybyp@nwpuedu.cn*

Received 01 November 2015; revised 17 January, 2016; accepted 19 January 2016; posted 21 19 January 2016 (Doc. ID 253148); published XX Month XXXX

**Image fusion aims to merge two or more images captured via various sensors of the same scene to construct a more informative image by integrating their details. Generally, such integration is achieved through the manipulation of the representations of the images concerned. Sparse representation plays an important role in the effective description of images, offering a great potential in a variety of image processing tasks, including image fusion. Supported by sparse representation, in this paper, an approach for image fusion by the use of a novel dictionary learning scheme is proposed. The non-local self-similarity property of the images is exploited, not only at the stage of learning the underlying description dictionary but during the process of image fusion. In particular, the property of non-local self-similarity is combined with traditional sparse dictionary [1], resulting in an improved learned dictionary, that is hereafter referred to as the non-local sparse K-SVD (NL_SK_SVD) dictionary. The performance of the NL_SK_SVD dictionary is applied for image fusion using simultaneous orthogonal matching pursuit. The proposed approach is evaluated with different types of image, and compared with a number of alternative image fusion techniques. The resultant superior fused images using the present approach demonstrates the efficacy of the NL_SK_SVD dictionary in sparse image representation.**

***OCIS codes:*** *: (100.0100) Image processing; (100.2000) Digital image processing; (350.2660) Fusion*

http://dx.doi.org/10.1364/AO.99.099999

## 1. INTRODUCTION

The development of photography capture devices has resulted in various types of image that focuses on different scene properties. Brightness, color, temperature, distance, and other information may be represented in individual images. Image fusion is the process of integrating information contained within two or more images regarding a common scene into a single composite image that is more informative and suitable for human visual perception or subsequent computer processing [2]. For instance, visible images detail the spectral and spatial information in a particular scene. However, if the color and brightness of an object concerned are different from the background only slightly, it can be difficult to recognize the object visually, whereas infrared radiation (IR) images captured over the object may help to provide a precise representation of the target. Hence, the fusion of visible and IR images can present more information for both human inspection and computer-based image analysis with one integrated image [3]. Indeed, image fusion has recently been utilized as effective tools in object recognition [4], remote sensing [5], target tracking [6], surveillance [7], and defense applications that require the use of multiple images of a scene.

Image fusion algorithms can be categorized with respect to different levels of integration, including low, mid and high levels, which are respectively referred to as pixel, feature and symbolic levels [2]. A pixel-wise image fusion algorithm straightforward operates on individual pixels, and a great majority of image fusion schemes developed to date fall into this category. Higher levels of the fusion process are guided by intelligent analysis of the source images, performed in terms of fusion of objects and/or features instead of individual pixels.

Different approaches have been proposed to address the problem of pixel-wise image fusion. Existing work can be summarized into three groups: 1) spatial domain-based, 2) transform domain-based and 3) multi-resolution analysis-based. These approaches are briefly described below. Spatial domain-based methods, which are also known as component substitution-based [8], carry out fusion of image pixels in terms of particular decomposition and representation of image components. Methods developed on the basis of intensity-hue-saturation (IHS) transform [9] or Principal Component Analysis (PCA)

[10] are the commonly used representatives of such techniques, especially for panchromatic (PAN) image and multispectral image fusion. The important advantage of these methods is that they can focus on image areas of interest, but they may be less capable in dealing with the change of such areas. Transform domain-based algorithms [11] create a fused image globally through pixel-level fusion locally. In particular, to change a single coefficient in such a fused image, the whole neighborhood of an image pixel in the spatial domain may change, which may therefore, unfortunately lead to undesirable side-effects [2].

Multi-resolution analysis is a more popular strategy in pixel-wise image fusion. In such schemes, wavelet transforms can be used to fuse images. Typical representatives include wavelet transform [12] and dual-tree complex wavelet transform (DT-CWT) [13]. These techniques have been used in many successful image fusion applications [12]. However, since different wavelets normally represent different characteristics of a given image (e.g., lines and textures), use of a single wavelet transform may be restrictive.

In addition to aforementioned approaches, visual attention-based or saliency-based techniques are also applied in image fusion in light of human psychological observations [14-18]. These approaches attempt to reflect the significance of human visual attentions in performing image fusion, assuming that the fusion is to pursue an integration of visual information in the first place. They are able to preserve the full content value and retain the visual meaningful features more accurately. A particular successful mechanism is to employ saliency measures of wavelet transform coefficients, improving the performance of image fusion significantly [16]. However, this method would require wavelet decomposition to be carried out first. In addition, a number of existing fusion schemes following the visual attention-based or saliency-based approach may work well only when properties of monotonicity and heterogeneity are satisfied, thereby restricting their otherwise wide range of utility.

Different from these approaches, sparse representation (SR) techniques offer a new solution for image fusion. The general framework following this approach [19] [20] can be outlined as follows: Firstly, the two or more source images are divided into patches using a certain 'sliding window' in respect of the corresponding position of each source image, with the resulting patches represented as vectors. Then, the vectors are encoded with coefficients by decomposition in terms of the atomic vectors in a given dictionary. Next, fused coefficients are computed using the sparse coefficients of each source image with regard to a certain integration rule. Finally, integrated vectors are computed by multiplying the fused coefficients by the dictionary, and then the fusion image is constructed in relation to the composite patch, which is derived from such vectors.

Most existing work in this area concentrates on the dictionary used for sparse representation and the fusion rule for integrating images. Representative dictionaries are of two types: analytic dictionary and trained dictionary, in support of the process of sparse decomposition [1]. Rather than using analytic dictionary (e.g., the Discrete Cosine Transform (DCT) dictionary), a number of techniques [8] [21-23] focus on the dictionary learning in order to improve the ability of sparse representation, typically for applications such as medical image fusion and remote sensing image fusion. The sparse coefficients of source images are combined to form the fused coefficient, where the popular choose-absolute-maximum rule (max-abs) [12] is often adopted. There have been alternative fusion rules [23-25] proposed to combine the coefficients in an attempt to exploit and integrate more information from source images, but this is beyond the scope of the present work.

In this paper, an image fusion scheme in sparse representation is proposed using a novel learning dictionary, through the exploitation of the non-local self-similarity property of images. This property has been utilized in image denoising and restoration previously [26] [27]. However, no prior work has been reported that make explicit use of image self-similarities in dictionary training and image fusion. The performance of the proposed approach is evaluated in comparison with a number of state-of-the-art techniques and over a range of images, demonstrating that its potential in achieving improved image fusion results.

The remainder of this paper is organized as follows. Section 2 gives an overview of related work in sparse representation (SR) for image fusion, including a brief description of SR and dictionary learning, and of the property of non-local self-similarity. Section 3 presents the proposed image fusion scheme with a novel learned dictionary. Section 4 details the experimental results and performance analysis. Section 5 concludes the paper, including a discussion of further research.

## 2. RELATED WORK

This section presents relevant background work, including basic concepts of sparse representation, dictionary and dictionary learning, and the property of non-local self-similarity in images.

### A. Sparse representation

Sparse representation of signals has resulted in significant development in signal processing techniques. It has also found success in applications to image processing since images are simply two-dimensional signals [28]. Following this approach, without losing generality, *n*×*n*-sized image blocks are utilized when sparse representation is employed, rather than the whole image, where n is a much smaller number than the size of a given image.

Let $x \in \mathbb{R}^{n^2}$ be a column vector transformed from a certain $n \times n$ image block, by ordering the pixels lexicographically. Suppose that a dictionary $D \in \mathbb{R}^{n^2 \times N}$ contains $N$ atomic vectors named atoms, each of which is an $n^2$-sized column vector. In the general sparse representation theory the vector $x$ can be represented as a linear combination of the atoms and a random perturbation (noise) vector with respect to the dictionary $D$, such that $x = D\alpha + e$, where $\alpha \in \mathbb{R}^N$ whose elements are termed the sparse coefficients, and $e \in \mathbb{R}^{n^2}$ represents the random noise with bounded energy, $\|e\|_2 \le \varepsilon$. The sparsity of the coefficients indicates that there are only a small number of nonzero entries in the vector $\alpha$. That is, as few atoms as possible are used to represent the original image. Often the size of the dictionary $D$ is required to be N>n , implying that the dictionary is of redundant or over-complete information. The target of sparse decomposition is to best represent the transformed vector using the over-complete dictionary $D$ and the sparse coefficient $\alpha$. In particular, the ideal, sparsest $\alpha$ is sought, which may be obtained through the following optimization:

$$\min_{\alpha} \|\alpha\|_0^0 \text{ subject to } \left\|x - D\alpha\right\|_2 \le \delta \qquad \textbf{(1)}$$

where $\|\cdot\|_0$ is the $L_0$ norm counting the number of nonzero entries in the vector and $\delta \ge 0$ is a preset error tolerance.

There are two significant parts that play an important role in sparse decomposition. One is the method for solving the above optimization problem. In general, it is a non-deterministic polynomial-time hard (NP-hard) problem, thus, two groups of suboptimal algorithms that approximate the optimization solution have been developed: 1) matching pursuit, which is a type of greedy algorithm, represented by methods such as Matching Pursuit (MP) [29] and Orthogonal Matching Pursuit (OMP) [30]; and 2) relaxation, represented by methods such as

Basic Pursuit (BP) [31] and FOcal Underdetermined System Solver (FOCUSS) [32].

The construction of dictionary is another key to improving the performance of sparse representation. As mentioned previously, the representative dictionaries are analytic dictionary and trained dictionary in the process of sparse decomposition [1]. An analytic dictionary is built upon mathematical modeling of data, e.g., via curvelet transform or contourlet transform [1]. Dictionary training is a much more recent approach to dictionary design. It has been strongly influenced by the latest advances in sparse representation theory itself. Two typical training algorithms for learning dictionary are: Method of Optimal Directions (MOD) [33] and K-SVD [34]. Compared to analytic dictionaries, trained dictionaries have proven to be able to produce state-of-the-art results in many practical image processing applications [1].

### B. Dictionary learning

Both routes for dictionary construction in the process of sparse representation have their respective strengths and weaknesses [35]. In particular, an analytic dictionary presents a formulated mathematical model of the given data, which is highly structured and entails fast numerical implementation, but unfortunately lacks adaptability. Opposite to this, a trained dictionary from a set of examples is adaptable, although this is achieved at the expense of generating an unstructured dictionary, which can be computationally more costly to apply.

The sparse dictionary [1] proposed in [35] may be seen as another learning based dictionary in general. However, as one of the parametric training methods, which combines analytic dictionary and trained dictionary, it is a hybrid technique that helps bridge the gap between these two approaches. As such, it gains the benefits of both. Specific strengths include low complexity, compact representation, and stability under noise and reduced overfitting, leading to a simple and flexible dictionary representation which is both adaptive and efficient [35].

More concretely, the goal of learning sparse dictionary is to decompose a trained dictionary on a fixed analytic dictionary, with the latter named a base dictionary hereafter. The following optimization problem needs to be solved:

$$\min_{A,\Gamma} \left\| X - \Phi A \Gamma \right\|_F^2 \quad \text{subject to} \quad \begin{cases} \forall i \left\| \Gamma_i \right\|_0^0 \le t \\ \forall j \left\| a_j \right\|_0^0 \le p, \left\| \Phi a_j \right\|_2 = 1 \end{cases} \tag{2}$$

where $X$ is the set of training examples, $\Phi$ is the base dictionary, $A$ is the sparse dictionary representation, and $\Gamma$ is the sparse representation of training samples such that $X \approx \Phi A \Gamma$. In the constraining conditions, $p, t$ are the target atom sparsity and target training samples sparsity, respectively.

A seemingly straightforward approach is to sparse-code the atoms of the learned dictionary $D_0$ in the base dictionary $\Phi$ to obtain $D = \Phi A \approx D_0$. However, this naive approach is suboptimal; only when the base dictionary $\Phi$ is sufficiently compatible with $D_0$, the representation in $A$ may be sparse.

Having recognized this, a K-SVD-like learning scheme is proposed [35] to train the sparse dictionary from examples, which is hereafter termed Sparse K-SVD algorithm (SK-SVD). In the process of training a sparse dictionary, the basic framework of K-SVD algorithm is utilized. Briefly, there are the two key steps: 1) sparse-coding for each samples in the training set, and 2) dictionary-updating for each representation in $A$ and $\Gamma$. These steps alternate for a predefined number of iterations. A detailed description of the SK-SVD algorithm can be found in [35] and hence, is omitted here.

In addition to compact expression and fast implementation, the parameterization of a sparse dictionary helps improve generalization and reduce the training set size. As a result, the training method can be applied to learn larger dictionaries than the MOD or K-SVD, thereby beneficial for dealing with large image blocks [1].

### C. Non-local self-similarity in images

In sparse representation of images, a source image is partitioned into patches using a 'sliding window', which partially overlap with one another. These patches are utilized to learn sparse dictionary and image decomposition. This is in order to reduce the computational complexity whilst relaxing the requirement of data storage. However, there may exist a great amount of similarity and hence redundancy between patches that are extracted from one single image. Also, similarity may exist between not only local patches in a certain neighborhood but non-local patches within the entire image, as illustrated in Fig. 1.

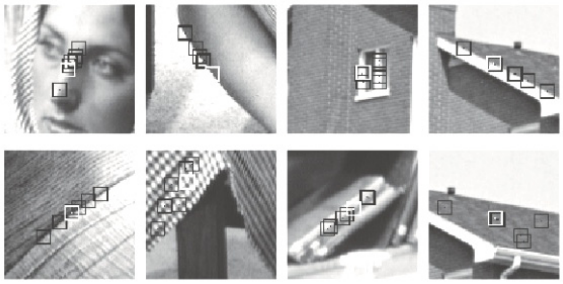


**Fig. 1** Redundancy of image existing in local and non-local patches.

Interestingly, the self-similarity contained in non-local patches can be extracted and such information has been utilized in a number of different image processing problems in recent years. The original application was for image denoising [26], where the denoised value for a pixel $i$ is a weighted average of all points whose Gaussian neighborhood looks like the neighborhood of $i$. Note that the weight of pixel $i$ in connection with another pixel depends on the similarity between them, which is related to the similarity of the intensity gray levels between the two neighboring patches. This means that the non-local means not only compares the grey level in a single point but also the geometrical configuration in a whole neighborhood.

The combination of image self-similarity and sparse representation has been put forward as a novel strategy for image processing tasks. In particular, the BM3D procedure [36] exploits both self-similarities and sparsity for image denoising, but it is based on classical, fixed orthogonal dictionaries. The learned dictionary is first used with the corresponding models of image self-similarities for image reconstruction [27], which makes it possible to effectively restore raw images from digital cameras at a reasonable speed and memory cost. Through the use of non-local means self-similarities in natural images are utilized to average out the noise among similar patches. A great deal of further effort has also been made to improve the noise-removal performance by the use of non-local self-similarity [37] [38].

Inspired by the above observation, in this paper, an image fusion approach using sparse representation techniques is proposed, with a focus on learning dictionary for SR. In particular, a sparse K-SVD dictionary trained with the property of image non-local self-similarity, named non-local sparse K-SVD dictionary, is addressed below.

## 3. IMAGE FUSION WITH NON-LOCAL SPARSE K-SVD DICTIONARY

In this work, the property of non-local self-similarity is applied both in the process of learning dictionary, and at the stage of image fusion. The proposed architecture is shown in Fig. 2.

### A. Training sample selection

In order to make full use of the non-local self-similarity property, it is necessary to identify similar patches of each training sample image. For a given patch, similar patches can be computed using Euclidean distance metric. Training samples for a learning dictionary are then reconstructed with the original one and its similar patches, where the training samples are built as outlined in Fig. 3.

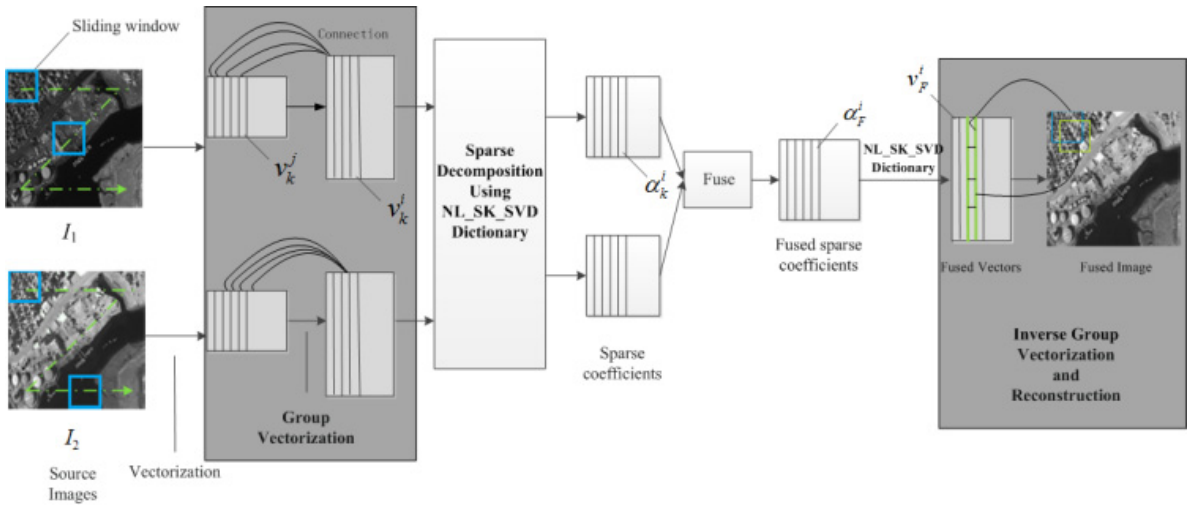



**Fig. 2.** Image Fusion via non-local sparse K-SVD dictionary

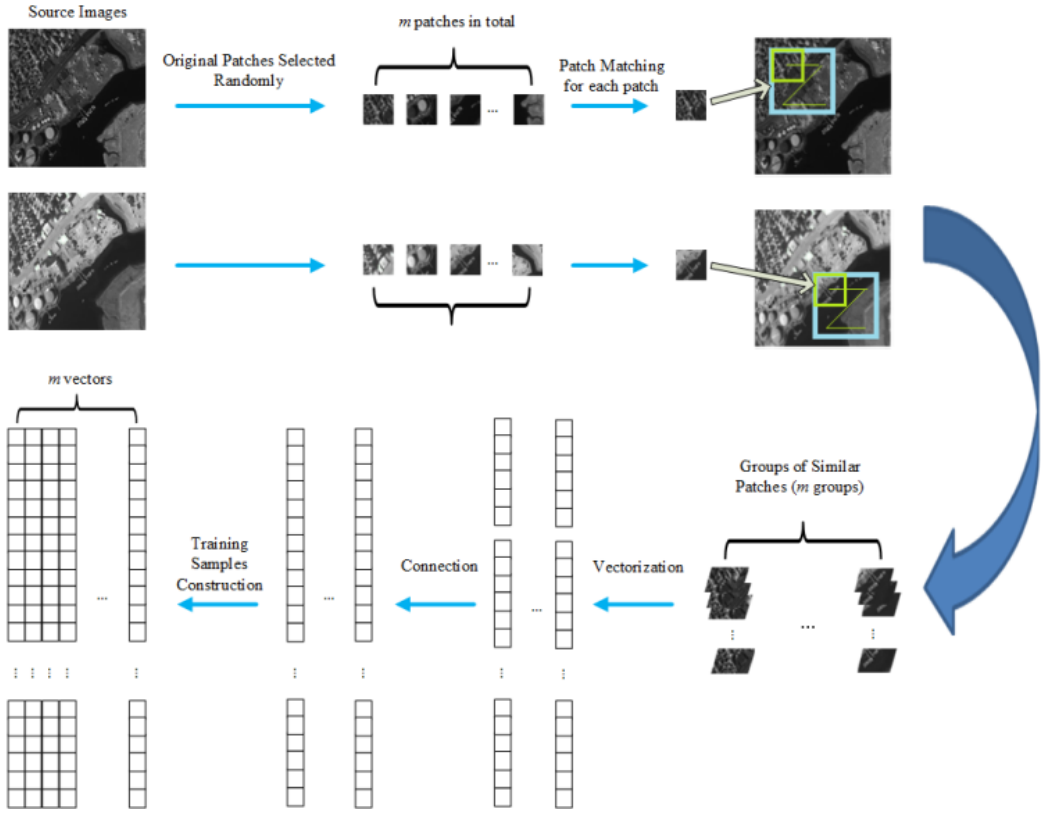



**Fig. 3.** Training samples building based on non-local self-similarity

To obtain the non-local self-similarity training samples, a certain number of, say $m$, rectangular patches are first selected randomly, each of which is of a preset size $n \times n$. Then for each of the resulting patches, termed the original ones hereafter, $r$ most similar patches are taken, each of which best matches this original using Euclidean distance metric within a $p \times q$ neighboring area. Next, as reflected in Fig. 3, the original and its similar patches are transformed into column vectors via lexicographic ordering, and a new overall vector is constructed collapsing these vectors head to tail. This process is repeated for m times, resulting in a matrix $X \in \mathbb{R}^{(r+1)\cdot n^2 \times m}$, containing the required training samples, with each column being a training example.

Given the training matrix, the learning dictionary can be obtained by directly applying the sparse K-SVD algorithm [35], resulting in a non-local sparse K-SVD dictionary.

### B. Non-local sparse K-SVD dictionary based image fusion

The non-local self-similarity property of images is herein further applied to perform image fusion. This is inspired by the work of [20]. In particular, the fusion process is implemented with the simultaneous orthogonal matching pursuit (SOMP) procedure [39], guaranteeing that different source images are sparsely decomposed into the same subset of dictionary bases. Again, the entire image fusion flowchart is shown in Fig. 2.

The learning dictionary is built on the basis of samples constructed from non-local self-similarity patches. Similarly, patches selected for fusion (by a sliding window) are required to be extended to becoming high dimensional vectors in groups, with respect to the non-local self-similarity property also. There are two reasons for such extension: 1) According to the sparse representation theory, the dimension of vectors for fusion is supposed to correspond to the dimension of atoms within the sparse dictionary (the NL_SK_SVD dictionary in this work). 2) Since the atoms in the NL_SK_SVD dictionary are constructed with patches of non-local self-similarity, image blocks (i.e., vectors) for fusion are required to be extended such that their corresponding sparse representation will be of sufficient sparsity. Thus, group vectorization needs to be carried out before the SOMP stage, and an inverse group vectorization process is performed before fused image reconstruction (as highlighted in Fig. 2).

Without loss of generality, suppose that $K$ source images which are of size $M \times N$ are geometrically registered. Then, following the framework of Fig. 2, the proposed image fusion scheme works as follows:

- **Source image partition**: Divide each source image $I_k$, $k$=1,2,..., $K$, from left-top to right-bottom, into every possible image patch of size $n \times n$, which is the same as the size of the atoms in the dictionary. Then all the patches are transformed into vectors via lexicographic ordering, resulting in a matrix $\{v_k^j\}_{j=1}^{(M-n+1)\cdot(N-n+1)} \in \mathbb{R}^{n^2 \times ((M-n+1)\cdot(N-n+1))}$, where $v$ denotes the label of vector and $n^2$ denotes the dimensionality of each vector.
- **Group vectorization**: For each vector in the first matrix $\{v_k^j\}, k = 1$, find $r$ most similar vectors in addition to itself throughout the matrix, using the conventional Euclidean distance metric as the criterion in measuring the similarity, the smaller the distance the more similar the vectors. Then, collect these $r+1$ vectors and transform them into a high dimensional vector by connecting them head to tail. This process results in an extended matrix $\{V_k^j\}_{j=1}^{(M-n+1)\cdot(N-n+1)} \in \mathbb{R}^{n^2\cdot(r+1)\times((M-n+1)\cdot(N-n+1))}$. Repeat this process for the remaining matrices $\{v_k^j\}, k = 2$ to $K$, extending each vector of the same index. Hence, $K$ extended matrices are obtained.
- **Vector decomposition**: Decompose the vectors at each position, $j$, in extended matrices $\{V_k^j\}_{j=1}^K$ that are extracted from different source images into their corresponding sparse representations, $\alpha_1^j, \alpha_2^j, \ldots, \alpha_K^j$, using the SOMP algorithm with the non-local sparse K-SVD dictionary.
- **Sparse coefficient integration**: Combine the sparse coefficient vectors using the max-abs rule at each position $j$:

$$\alpha_F^j(t) = \alpha_k^j(t), k = \arg\max_{k=1,2,\ldots,K}(|\alpha_k^j(t)|) \quad \textbf{(3)}$$

  where $\alpha_k^j(t)$ is the $t$ th value of the vector $\alpha_k^j(t)$, $t = 1,2,\ldots,T$. The use of this absolute-maximum rule is to combine coefficients here. This is because the fused image is expected to being represented from the most important information, instead of a certain form of averaging which may result in smoother edge or line. A fused vector is then obtained by $v_F^j = D\alpha_F^j$, where $D$ is the non-local sparse K-SVD dictionary.
- **Inverse group vectorization and reconstruction**: For each fused vector $v_F^j$, divide it equally into $r+1$ sub-vectors. Each resulting sub-vector is reshaped into an image block of size $n \times n$ and is then added onto the emerging reconstructed fusion image $I_F^j$ at the corresponding position. The final reconstructed fusion image $I_F$ is obtained by dividing each individual pixel value by the adding times at its position.

Note that in the final fused image, any patch is in general, reconstructed by averaging with multiple patches, rather than just one patch in the previous SOMP approach [20]. Hence, more information will be integrated from those non-local patches which are similar to a certain extent, resulting in a better fusion performance.

## 4. EXPERIMENTAL RESULTS AND ANALYSIS

The performance of the proposed approach for image fusion using non-local sparse K-SVD dictionary is evaluated in this section. Experimental investigation conducted involves different types of source image and is compared with three dictionary learning based image fusion schemes.

### A. Experimental setup

Six pairs of source images are fused, each pair of which are captured in the same scene by different sensors. These source images are shown in Fig. 4. In particular, Figs. 4 (1a) and (1b) are multi-focus images and the fused image is supposed to illustrate both clocks clearly. Figs. 4 (2a) and (2b) show a computed tomography (CT) and a magnetic resonance imaging (MRI) image, the structures of bone and areas of soft tissue are captured respectively. Obviously, such information merging can play a significant role in successful medical diagnosis. Infrared images and visual images are often integrated to fuse information in order to enhance image details. Infrared images are shown in Figs. 4 (3a) and (4a), while optical images are shown in Figs. 4 (3b) and (4b), respectively for two different scenes. The information contained in an infrared image is complementary to that in the corresponding visual image. Also, it is a common practice to fuse remote sensing images, integrating information acquired by the aerial sensor technologies regarding an interesting object or phenomenon without making physical contact with the object. Such images can be very different, ranging from those of different spatial resolutions to different spectral resolutions. Figs. 4 (5a) and (5b) show two images obtained using different bands of a multi-spectral scanner, and Figs. 4

(6a) and (6b) show a high spatial resolution image and a hyperspectral image, respectively.

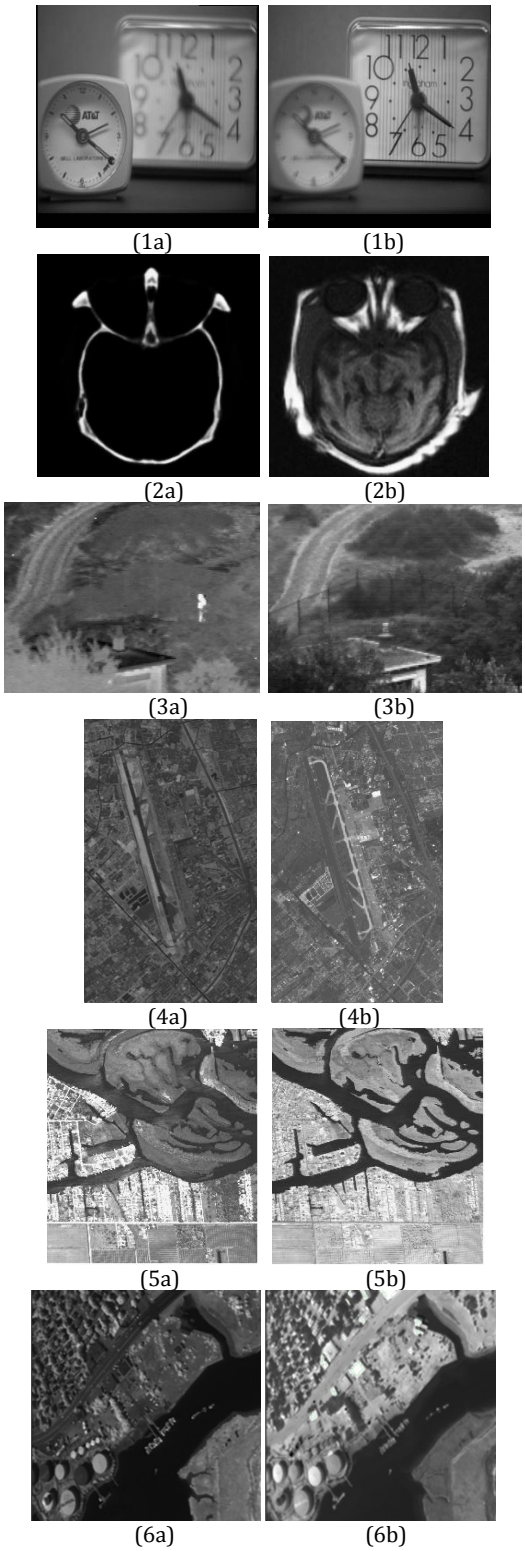


**Fig. 4** Six pairs of source images.

As mentioned previously, source images are divided into patches by a sliding window, with a size of *n*×*n*. In this experimental study, the size of each patch is empirically set to 8x8 to facilitate comparison, which is commonly used in the literature. In general, it may be difficult to capture information on structures and textural features if the patch areas are too small, but computing with larger blocks will cost a long time for sparse representation and dictionary learning. However, the learning samples are built with non-local patches, involving high dimensional dictionary atoms. Summarizing these reasons, sliding windows of 8×8 are taken in performing the experimental evaluation below.

The performance of the proposed approach is compared with three other dictionary learning based image fusion schemes. These are: 1) SOMP that uses DCT dictionary (DCT) [20], 2) SOMP that uses K-SVD learning dictionary (K-SVD) [34], and 3) SOMP that uses Sparse K-SVD learning dictionary (SK-SVD) [35]. As mentioned previously, the dictionary for sparse representation can be divided into two categories: analytic and learning based respectively. For comparison, DCT is employed to form a mathematical model as for the analytic dictionary, and the trained dictionaries are constructed from data with the K-SVD and Sparse K-SVD algorithms.

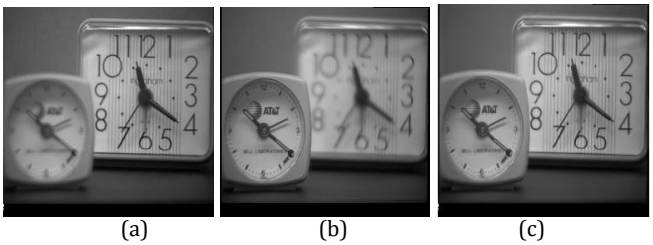
Other parameters used in the experimentation are set as follows. The number of training samples ($m$) for non-local sparse K-SVD dictionary is 2000, which are selected from the source images randomly; the neighbor window ($p \times q$) for similar patch selection is set to 10x10; the number of most similar patches ($r$) is 3; the size of DCT analytic dictionary is 64 x 192; the size of both K-SVD and SK-SVD dictionary is 64 x 200; the training parameters for NL_SK_SVD dictionary are: the size of the initial DCT dictionary is 256 x 200, the sparsity of target training sample representation and target atom representation are set to 20 and 10, respectively; and the total number of iterations is 10.

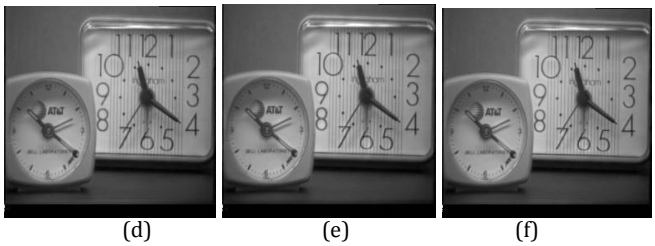
To ensure comprehensive and fair comparison over the fusion of grey value images, five quantitative object assessment criteria are used, including: Mutual Information ($MI$) [40], $Q^{AB/F}$ [41], $Q^0$ [42], Standard Deviation ($SD$) [24] and Average Grade ($AG$). Since there is not an ideal or ground truth fusion outcome that may be utilized as reference in performing such image fusion tasks, all results are adjudged relatively in terms of their respective performance measurement values. The higher such a value, the better the corresponding fusion result.

### B. Results and Discussion

To demonstrate the performance of the different image fusion schemes, fused images of each pair of source images shown in Fig. 4 are presented in Fig. 5–Fig. 10. Quantitative evaluation measurements are compared in Table 1, where the best results are indicated in bold.

Figures 5(a) and (b) show the source images (1a) and (1b), and Figs. 5(c)-(f) illustrate the fusion results of applying SOMP, based on three trained dictionaries mentioned previously and the proposed dictionary learning approach, respectively. All four methods seem to be able to lead to an accreditable fusion result, but the quantitative assessments shown in Table 1 (regarding Fig. 5) indicate that the proposed scheme obtains an overall slightly better fused image.

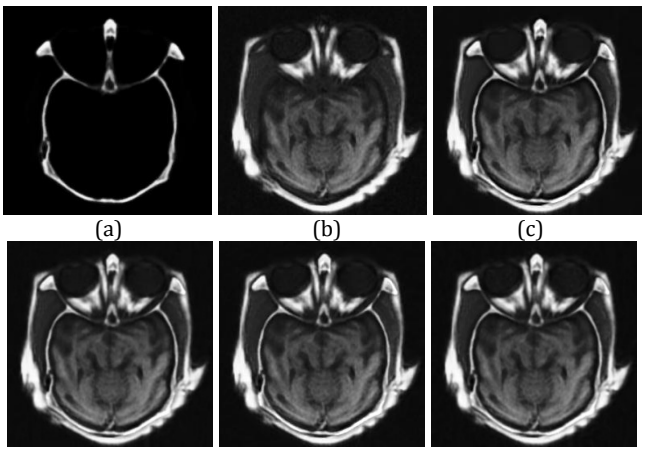




(d) (e) (f)

**Fig. 5** Fusion results of source images (1a) and (1b) in Fig. 4 based on different trained dictionaries: (a) source image (1a), (b) source image (1b), (c) DCT dictionary, (d) K-SVD dictionary, (e) SK-SVD dictionary, (f) proposed trained dictionary.

The results on medical image fusion are shown in Figs. 6 (c)-(f). Both the bone structure and the areas of soft tissue can be seen clearly in all fused images. However, the textures in the top are more prominent in (d), (e) and (f), obtained using trained dictionaries. In particular, the quantitative measurements given in Table 1 (regarding Fig. 6) indicate that the proposed approach achieves better results.



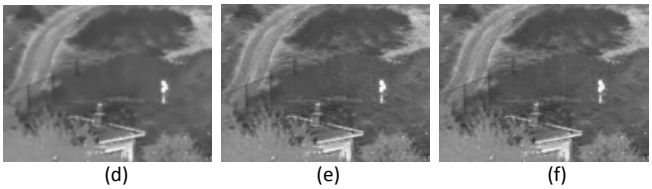
(a) (b) (c)

(d) (e) (f)

**Fig. 6** Fusion results of source images (2a) and (2b) in Fig. 4 based on different trained dictionaries: (a) source image (2a), (b) source image (2b), (c) DCT dictionary, (d) K-SVD dictionary, (e) SK-SVD dictionary, (f) proposed trained dictionary.

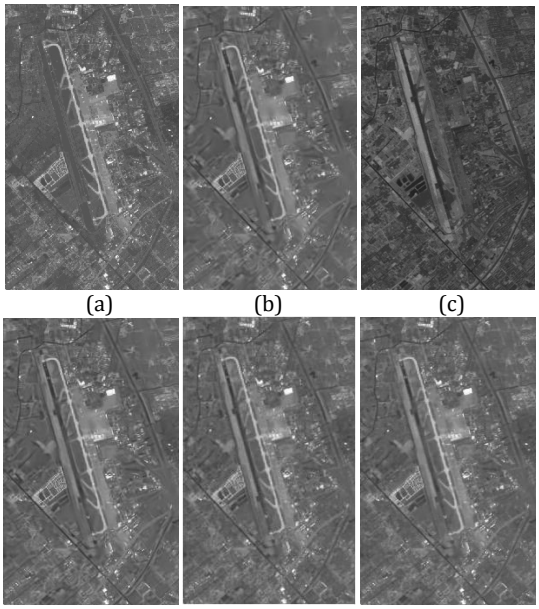
Fig. 7, Fig. 8 and Table 1 present the qualitative and quantitative results of fusing infrared images (3a) (4a) and visual images (3b) (4b) in Fig. 4, respectively. Although the objects, e.g., person, airfield runway, can be seen clearly among all of the fused images, details are more plentiful in Fig. 7 (f) and Fig. 8 (f) obtained by proposed approach. Comparing results in Table 1 (regarding Fig. 7 and Fig. 8), it can also be seen that there is much more structural similarity between fused image (f) and source images. Furthermore, fused image (f) contains more information, with distribution of grey scale being more spread out so that the resolution is much higher.

(a) (b) (c)



(d) (e) (f)

**Fig. 7** Fusion results of source images (3a) and (3b) in Fig. 4 based on different trained dictionaries: (a) source image (3a), (b) source image (3b), (c) DCT dictionary, (d) K-SVD dictionary, (e) SK-SVD dictionary, (f) proposed trained dictionary.

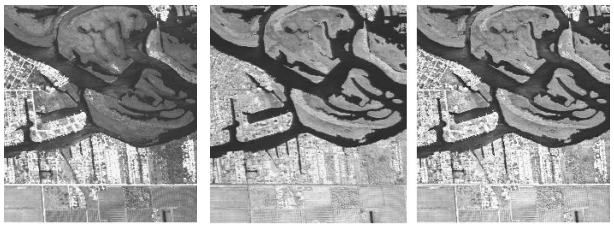


(a) (b) (c)

(d) (e) (f)

**Fig. 8** Fusion results of source images (4a) and (4b) in Fig. 4 based on different trained dictionaries: (a) source image (4a), (b) source image (4b), (c) DCT dictionary, (d) K-SVD dictionary, (e) SK-SVD dictionary, (f) proposed trained dictionary.

Fig. 9 and Table 1 (regarding Fig. 9) present the fusion results of multi-spectral images shown in Figs. 4 (5a) and (5b). Qualitatively, the performance of the proposed scheme is better than that of the rest, e.g., the distribution of grey scale is more spread out and the details are clearer. Quantitative comparisons shown in Table 1 (regarding Fig. 9) reinforce this observation as the work leads to fused image being more structurally similar to the source images and having higher resolution.

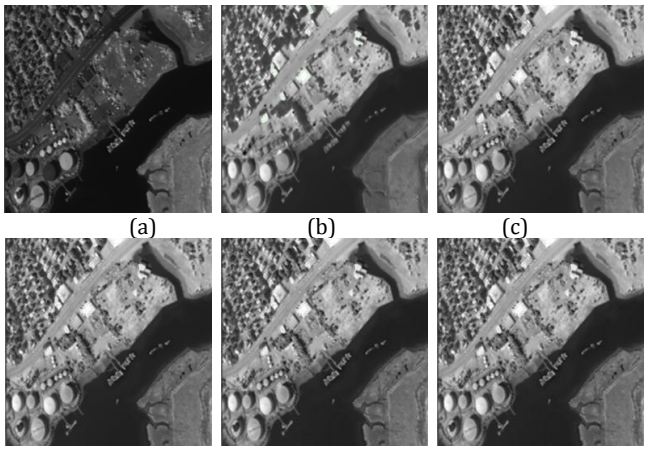


(a) (b) (c)

(d) (e) (f)

**Fig. 9** Fusion results of source images (5a) and (5b) in Fig. 4 based on different trained dictionaries: (a) source image (5a), (b) source image (5b), (c) DCT dictionary, (d) K-SVD dictionary, (e) SK-SVD dictionary, (f) proposed trained dictionary.

The fusion results of high-spatial-resolution image (Fig. 4 (6a)) and hyperspectral image (Fig. 4 (6b)) are shown in Fig. 10. Seen from the fused image qualitatively, detail information is more abundant and edges are much clearer in image (f) that is returned by the use of the proposed scheme. Table 1 (regarding Fig. 10) shows the comparative quantitative evaluation outcomes. There is much more structural similarity between fused image (f) and source images. Also, more information is integrated in fused image (f), and the distribution of grey scale is more spread out so that the resolution is much higher, showing a better fusion image as a whole.



(a) (b) (c)

(d) (e) (f)

**Fig. 10** Fusion results of source images (6a) and (6b) in Fig. 4 based on different trained dictionaries: (a) source image (6a), (b) source image (6b), (c) DCT dictionary, (d) K-SVD dictionary, (e) SK-SVD dictionary, (f) proposed trained dictionary.

Finally, it is interesting to investigate the time cost for fusing individual images given the fixed block size of 8x8. The time consumed for fusing the images of Figs. 5-10 is presented in Table 1. It is not surprising that the computational cost with the methods involving trained dictionaries (e.g., K-SVD, SK-SVD and NL_SK_SVD) is heavier than the use of analytic dictionary (DCT). This is because the trained dictionaries need to generate an unstructured dictionary learned from training samples before it is applied in image fusion. The proposed method costs a little more than its originals since similar patches are grouped in terms of the non-local self-similarity, both in the process of learning dictionary and at the stage of image fusion. However, significantly better performance is resulted, as reflected both from the object perception of the fused images and from the quantitative assessments. The slight extra computational is therefore, worthwhile spending.

## 5. CONCLUSION

An image fusion approach using a novel dictionary learning scheme has been proposed in this paper. The non-local self-similarity property of images is combined with sparse dictionary, resulting in an innovative trained dictionary strategy, non-local sparse K-SVD dictionary. The approach has been applied to support the implementation of image fusion. Comparative experimental studies, carried out over different types of source image, have demonstrated that better performance can be obtained than that obtainable by the state-of-the-art dictionary-based image fusion techniques.

Whilst the approach shows stronger performance than existing methods, there are a number of areas in which further development can help further strengthen its ability. In particular, computational complexity is fairly significant, especially in the processes of dictionary learning, block group vectorization and fused image reconstruction. Thus, computationally more efficient means for the formulation of non-local self-similarity of images would be desirable to support scale-up applications. Also, in the present work, once the non-local sparse K-SVD dictionary is constructed from samples, the fusion scheme is performed via patches in the source images, sequentially and independently. Hence, an investigation into accelerative strategies would be helpful, e.g., through the use of GPU's powerful floating-point arithmetic ability to speed up CPU computation.

**Funding Inforamtion.** Doctoral Program of Higher Education of China (20126102110041); Research Fund for the Key Project of Technology Research Plan of Ministry of Public Security, China (2014JSYJA018); Astronautical Supporting Technology Foundation of China (2014-HT-XGD).

**Acknowledgment**. We are grateful to the anonymous reviewers for their constructive comments which have helped improve this work.

**Table 1**. Quantitative assessment of dictionary-based fusion schemes

| Images | Dictionary-based fusion scheme | Quantitative assessment | | | | | |
|---|---|---|---|---|---|---|---|
| | | MI | $Q^{AB/F}$ | $Q_0$ | SD | AG | Time(s) |
| Fig. 5 | DCT | 6.6857 | 0.6873 | 0.9826 | 51.9643 | 6.1203 | **1147.93** |
| | K-SVD | 6.7456 | 0.6867 | **0.9836** | 51.7522 | 6.0605 | 1504.15 |
| | SK-SVD | 6.7587 | 0.6813 | 0.9816 | **52.4784** | 6.2995 | 1309.44 |
| | NL_SK_SVD | **6.9215** | **0.6964** | 0.9828 | 52.3050 | **6.7614** | 1638.21 |
| Fig. 6 | DCT | 3.8108 | 0.7492 | 0.5128 | 60.6513 | **7.0382** | **1229.97** |
| | K-SVD | 3.9427 | 0.7486 | 0.5128 | 60.8059 | 6.7494 | 1556.67 |
| | SK-SVD | 3.9713 | 0.7499 | 0.5109 | 60.8364 | 6.7452 | 1344.39 |
| | NL_SK_SVD | **4.0363** | **0.7547** | **0.5134** | **60.9799** | 6.8155 | 1667.24 |
| Fig. 7 | DCT | 2.4602 | 0.3113 | 0.5752 | 29.0536 | 2.8899 | **1811.75** |
| | K-SVD | 2.4855 | 0.2951 | 0.5760 | 28.9278 | 2.4950 | 2126.08 |
| | SK-SVD | **2.6044** | 0.3945 | 0.5813 | 29.3203 | 3.5290 | 1984.93 |
| | NL_SK_SVD | 2.4904 | **0.4265** | **0.5839** | **29.9317** | **4.0029** | 2501.89 |
| Fig. 8 | DCT | 1.2135 | 0.4122 | 0.6071 | 20.1436 | 7.0774 | **1243.42** |
| | K-SVD | 1.3193 | 0.4697 | 0.6241 | **21.0037** | **8.3126** | 1556.87 |
| | SK-SVD | 1.3405 | 0.4766 | 0.6285 | 20.5817 | 7.5891 | 1325.25 |
| | NL_SK_SVD | **1.3853** | **0.4895** | **0.6310** | 20.7641 | 7.8854 | 1701.74 |
| Fig. 9 | DCT | 4.1085 | 0.5972 | **0.9268** | 63.8325 | 14.5884 | **1162.67** |
| | K-SVD | 4.1507 | 0.5880 | 0.9262 | 62.7391 | 14.2204 | 1531.53 |
| | SK-SVD | **4.1623** | 0.5884 | 0.9265 | 63.9187 | 15.0886 | 1250.21 |
| | NL_SK_SVD | 4.1110 | **0.6052** | 0.9246 | **65.8581** | **21.6851** | 1772.49 |
| Fig. 10 | DCT | 4.3077 | 0.5410 | **0.6968** | 54.4286 | 9.9177 | **1142.87** |
| | K-SVD | 4.3519 | 0.5412 | 0.6967 | 54.5196 | 10.1919 | 1482.39 |
| | SK-SVD | **4.4086** | 0.5423 | 0.6967 | 54.5327 | 10.2603 | 1341.91 |
| | NL_SK_SVD | 4.4059 | **0.5424** | 0.6967 | **54.5337** | **10.2736** | 1526.03 |

## References


1. Rubinstein, Ron, Alfred M. Bruckstein, and Michael Elad. "Dictionaries for sparse representation modeling." Proceedings of the IEEE **98**, 1045-1057 (2010).
2. Goshtasby, A. Ardeshir, and StavriNikolov, "Image fusion: advances in the state of the art." Inf. Fusion **8**, 114-118 (2007).
3. Yan, Xiang, Hanlin Qin, Jia Li, Huixin Zhou, Jing-guo Zong, and Qingjie Zeng, "Infrared and visible image fusion using multiscale directional nonlocal means filter." Appl. Opt. **54**, 4299-4308 (2015).
4. Bai, Xiangzhi. "Image fusion through feature extraction by using sequentially combined toggle and top-hat based contrast operator." Appl. Opt. **51**, 7566-7575 (2012).
5. Simone, Giovanni, Alfonso Farina, Francesco Carlo Morabito, Sebastiano B. Serpico, and Lorenzo Bruzzone. "Image fusion techniques for remote sensing applications." Inf. Fusion **3,** 3-15 (2002).
6. Sworder, David D., John E. Boyd, and G. A. Clapp. "Image fusion for tracking manoeuvring targets." International Journal of Systems Science **28**, 1-14 (1997).
7. Karali, A. Onur, Serdar Cakir, and Tayfun Aytaç. "Multiscale contrast direction adaptive image fusion technique for MWIR-LWIR image pairs and LWIR multifocus infrared images." Appl. Opt. **54,** 4172-4179 (2015).
8. Li, Shutao, Haitao Yin, and Leyuan Fang. "Remote sensing image fusion via sparse representations over learned dictionaries." IEEE Trans. Geosci. Remote Sens. **51,** 4779-4789 (2013).
9. Choi, Myungjin. "A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter." IEEE Trans. Geosci. Remote Sens. **44**, 1672-1682 (2006).
10. Chavez, Pats, Stuart C. Sides, and Jeffrey A. Anderson. "Comparison of three different methods to merge multiresolution and multispectral data- Landsat TM and SPOT panchromatic." Photogrammetric Engineering and Remote Sensing **57**, 295-303 (1991).
11. Nikolov, Stavri, Paul Hill, David Bull, and Nishan Canagarajah. "Wavelets for image fusion." Wavelets in signal and image analysis. Springer Netherlands, 2001.213-241.
12. Pajares, Gonzalo, and Jesus Manuel De La Cruz. "A wavelet-based image fusion tutorial." Pattern Recognition **37,** 1855-1872 (2004).
13. Lewis, J. J., R. J. O'callaghan, S. G. Nikolov, D. R. Bull, and C. N. Canagarajah. "Region-based image fusion using complex wavelets."

Seventh International Conference on Information Fusion (FUSION). Vol. 1. 2004.
14. Atrey, Pradeep K., M. Anwar Hossain, Abdulmotaleb El Saddik, and Mohan S. Kankanhalli. "Multimodal fusion for multimedia analysis: a survey." Multimedia Systems **16**, 345-379 (2010).
15. Hua, Xian-Sheng, and Hong-Jiang Zhang. "An attention-based decision fusion scheme for multimedia information retrieval." In Advances in Multimedia Information Processing-PCM 2004, pp. 1001-1010.
16. Jian, Muwei, Junyu Dong, and Yang Zhang. "Image fusion based on wavelet transform." In Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on, vol. 1, (IEEE, 2007) pp. 713-718.
17. Luo, X. Y., J. Zhang, and Q. H. Dai. "Saliency-based geometry measurement for image fusion performance." IEEE Transactions on Instrumentation and Measurement, **61**, 1130-1132 (2012).
18. Chen, Yanfei, and Nong Sang. "Attention-based hierarchical fusion of visible and infrared images." Optik-International Journal for Light and Electron Optics **126** 4243-4248 (2015).
19. Yang, Bin, and Shutao Li. "Multifocus image fusion and restoration with sparse representation." IEEE Trans. Instrumentation and Measurement **59**, 884-892 (2010).
20. Yang, Bin, and Shutao Li. "Pixel-level image fusion with simultaneous orthogonal matching pursuit." Inf. Fusion **13**, 10-19 (2012).
21. Li, Shutao, Haitao Yin, and Leyuan Fang. "Group-sparse representation with dictionary learning for medical image denoising and fusion." IEEE Trans. Biomedical Engineering **59**, 3450-3459 (2012).
22. Yin, Haitao. "Sparse representation with learned multiscale dictionary for image fusion." Neurocomputing **148,** 600-610 (2015).
23. Zhang, Qiheng, Yuli Fu, Haifeng Li, and Jian Zou. "Dictionary learning method for joint sparse representation-based image fusion." Optical Engineering **52**, 057006-057006 (2013).
24. Ting, Liu, and Cheng Jian. "A New Algorithm for Image Fusion via Sparse Representation." International Conference on Automatic Control and Artificial Intelligence (ACAI) (2012), pp. 151-154.
25. Liu, Yu, and Zengfu Wang. "Multi-focus image fusion based on sparse representation with adaptive sparse domain selection." IEEE 2013 Seventh International Conference on Image and Graphics (ICIG) (IEEE, 2013), pp. 591-596.
26. Buades, Antoni, BartomeuColl, and J-M. Morel."A non-local algorithm for image denoising." IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE, 2005), Vol. 2.
27. Mairal, Julien, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. "Non-local sparse models for image restoration." IEEE 12th International Conference on Computer Vision (ICCV) (IEEE, 2009), pp.2272-2279.
28. Elad, Michael, Mario AT Figueiredo, and Yi Ma. "On the role of sparse and redundant representations in image processing." Proceedings of the IEEE **98,** 972-982 (2010).
29. Mallat, Stéphane G., and Zhifeng Zhang. "Matching pursuits with time-frequency dictionaries." IEEE Trans. Signal Processing **41**, 3397-3415 (1993).
30. Tropp, Joel A. "Greed is good: Algorithmic results for sparse approximation." IEEE Trans. Information Theory **50**, 2231-2242 (2004).
31. Chen, Scott Shaobing, David L. Donoho, and Michael A. Saunders. "Atomic decomposition by basis pursuit." SIAM journal on scientific computing **20**, 33-61 (1998).
32. Gorodnitsky, Irina F., and Bhaskar D. Rao. "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm." IEEE Trans. Signal Processing **45**, 600-616 (1997).
33. Engan, Kjersti, Sven Ole Aase, and J. HakonHusoy. "Method of optimal directions for frame design." in IEEE Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (IEEE, 1999), pp. 2443-2446.
34. Aharon, Michal, Michael Elad, and Alfred Bruckstein. "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation." IEEE Trans. on Signal Processing **54,** 4311-4322 (2006).
35. Rubinstein, Ron, Michael Zibulevsky, and Michael Elad. "Double sparsity: Learning sparse dictionaries for sparse signal approximation." IEEE Trans. on Signal Processing **58**, 1553-1564 (2010).
36. Dabov, Kostadin, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. "Image denoising by sparse 3-D transform-domain collaborative filtering." IEEE Trans. Image Processing **16**, 2080-2095 (2007).
37. Wang, Jin, Yanwen Guo, Yiting Ying, Yanli Liu, and Qunsheng Peng. "Fast non-local algorithm for image denoising." in IEEE International Conference on Image Processing (IEEE, 2006), pp.1429-1432.
38. Kervrann, Charles, Jérôme Boulanger, and Pierrick Coupé. "Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal." Scale Space and Variational Methods in Computer Vision. Springer Berlin Heidelberg, 520-532(2007).
39. Tropp, Joel A., Anna C. Gilbert, and Martin J. Strauss. "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit." Signal Processing **86**, 572-588 (2006).
40. Qu, Guihong, Dali Zhang, and Pingfan Yan. "Information measure for performance of image fusion." Electronics letters **38**,313-315 (2002).
41. Xydeas, C. S., and V. Petrović. "Objective image fusion performance measure." Electronics Letters **36**, 308-309 (2000).
42. Piella, Gemma, and Henk Heijmans. "A new quality metric for image fusion." in IEEE Proceedings of the International Conference on Image Processing (ICIP) (IEEE 2003), pp.III173-III176.