

Image fusion based on multi-scale transform and sparse representation: an image energy approach

Fatemeh Fakhari¹, Mohammad.R. Mosavi¹✉, Mehdi.M. Lajvardi²

¹Department of Electrical Engineering, Iran University of Science and Technology, Narmak, Tehran 16846-13114, Iran

²Khatam Al-Anbia University, Tehran, Iran

✉ E-mail: m_mosavi@iust.ac.ir

ISSN 1751-9659

Received on 13th February 2017

Revised 6th June 2017

Accepted on 16th July 2017

E-First on 25th September 2017

doi: 10.1049/iet-ipr.2017.0104

www.ietdl.org

Abstract: Image fusion is a process to enhance the human perception of different images from the same scene. Nowadays, two popular methods in the signal/image fusion, namely, multi-scale transform (MST) and sparse representation (SR) are being used. This study uses an image energy approach to enhance a fusion rule based on the combination of MST and SR methods. Each source image is first decomposed to its sub-bands using the selected MST method. Then, SR is applied to the low-pass band and maximum absolute (max-abs) rule merges the high-pass bands. The activity level of the sparse coefficients is measured based on the energy differences of the source images. When the gap energy is high enough, a coefficient with maximum L_2 -norm is selected; otherwise, maximum L_1 -norm is considered. Finally, by applying inverse MST to the attained bands, the fused image is reconstructed. The popular MSTs, such as discrete wavelet transform, dual-tree complex wavelet transform and non-sub-sampled contourlet are used. The experiments are carried out on several standard and real-life images. The measurement results confirm that the proposed method has enhanced the contrast, clarity and visual information of the fused results.

1 Introduction

Image fusion is a process in which information from several images is extracted and integrated in a single image in order to achieve better perception [1–3]. Fused images are more informative and can be more compatible with the human eye [4–6]. Therefore, they are employed in many applications such as medical imaging, surveillance, military and remote sensing to enhance their performances. Some fusing scenarios are multi-focus, multi-view and multi-modal. In the multi-modal field, source images are being prepared by various sensors [e.g. CT-MRI images in medical imaging, visible-millimetre wave (MMW) images and visible-infrared (IR) images]. Multi-scale transform (MST) can be named among the most popular image fusion methods [7]. Different kinds of MSTs are: discrete wavelet transform (DWT) [8], dual tree complex wavelet transform (DTCWT) [9], curvelet transform (CT) [10], steerable pyramid (SP) [11], non-sub-sampled contourlet transform (NSCT) [12] and non-sub-sampled Shearlet transform [13].

In most MST-based fusion methods, source images are decomposed to their sub-bands and then, according to the selective fusion rules, these sub-bands are merged. Finally, a fused image is reconstructed by applying the inverse MST on the merged sub-bands. Every sub-band contains salient features of the images. In general, low-and high-pass bands are approximation and details of a source image, respectively. Conventionally, low-pass bands are fused using value averaging while high-pass bands fused by maximum absolute values (max-abs). Based on the max-abs rule, the activity level is measured by the absolute value and sharp edges and fine details lead to coefficients with high-absolute values.

Low-pass band contains most of image energy. Hence, the averaging rule is not suitable and some energy in the source images may be lost. This problem is known in multi-modal fusion in which source images are captured from different sensors and each sensor receives different physical characteristics from a same scene. Thus, a given area in the source image may represent different levels of brightness. The problem can be solved by performing sparse representation (SR) on the low-pass bands, in which most of the energy in source images is preserved [14].

SR is one of the most robust methods in image/signal expansion and has attracted many attentions in image processing in recent years. Generally, signal expansion is a representation of a signal as a linear combinations of some basis [15]. Similarly, SR of a signal, $x \in R^n$, is a linear combination of a few atoms of a frame or dictionary, $D \in R^{n \times m}$. This dictionary is over-complete ($n < m$) and row full-rank. The SR-based method aims to minimise error ($\|x - Da\|_2$) while the condition of $\|\alpha\|_0$ is also minimised ($\alpha \in R^m$ is sparse coefficient).

In 1996, Olshausen and Field [15] claimed that neural cells in mammalian visual cortex use sparse coding for their representation. The SR method was first used in image processing by Aharon and Elad for denoising, in 2006. Then, they introduced K-SVD algorithm in order to learn a dictionary [16]. Yang *et al.* primarily used patch-based sparsity fusion in multi-focus images in 2010 [17]. Since then, various patch-based sparsity fusion techniques such as Joint Sparsity Model (JSM) were developed [18–20]. The joint sparse coefficients are innovative and common and the fused coefficients are weighted average of these sparse codes. The main idea is selection of the proper weights. Moreover, a combined MST and SR method has been reported [14].

Studies on image fusion domain can be divided into three categories. Some researchers work on pre-processing operation such as segmentation. In fact, they focus on fusion at destination level, as the fusion of IR and visible images which has been used in reference [21]. They try to extract the targets of IR image and fuse them with its pair visible image. Some others interests are fusion methods (new spatial or frequency transforms, such as MST, SR and so on). The last groups try to extend or suggest new fusion rules. However, ordinary rules, such as averaging and max-abs, were proposed earlier and new rules have been rarely reported in recent years.

In this paper, first fusion methods then, fusion rules will be discussed. Here, we decompose source images to their MST sub-bands and extract sparse coefficients from the low-pass band. Usually, activity levels of coefficient vectors are compared and more informative vectors are selected. This level measurement is known as fusion rule. Here, the activity levels are indicated by absolute value, L_1 -norm or L_2 -norm depends on each specific

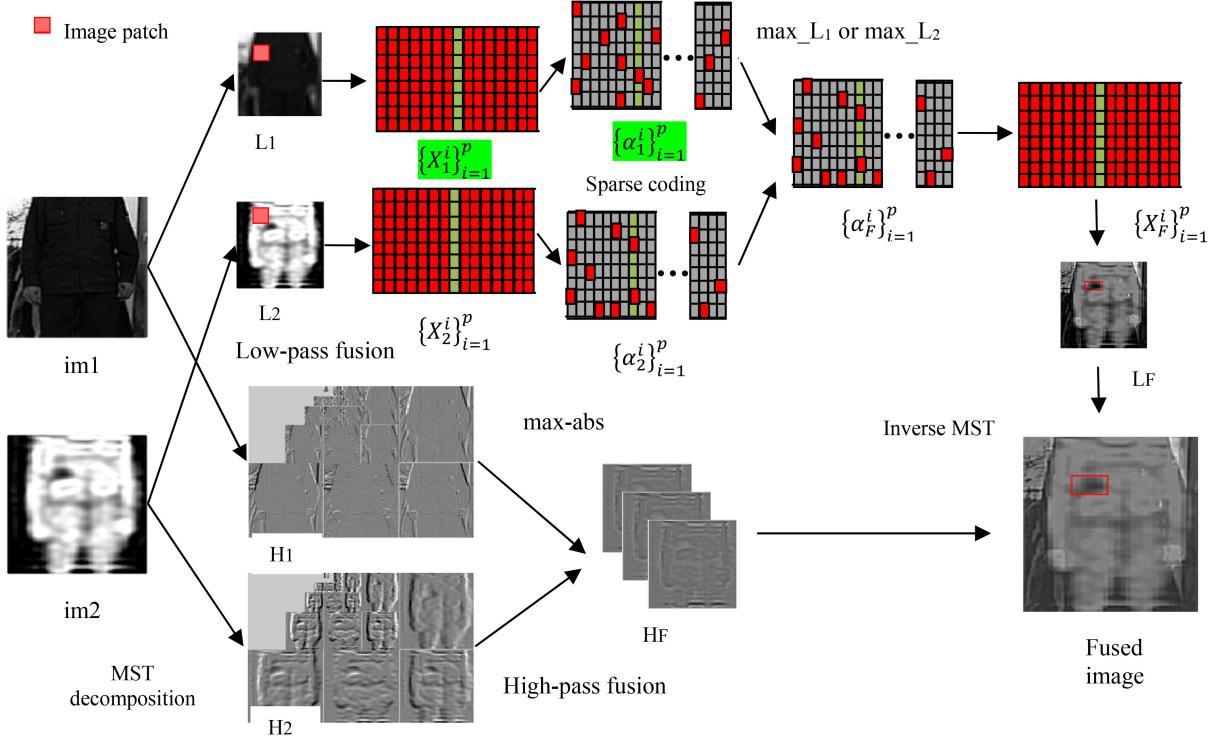


Fig. 1 Schematic diagram of the proposed method (the number of patches is p)

situation. Energy of images are calculated. If their energy differences are high enough (e.g. in multi-modal images), the ‘max- L_2 -norm’ rule will be suitable; otherwise, ‘max- L_1 -norm’ is used. Also, the ordinary max-abs rule is performed on the MST high bands. Briefly, we focus on the rule of fusion and performance of three MSTs; DWT, DTCWT and NSCT which are compared for further validity. Five fields of standard, real-life image databases are used including: multi-focus, medical multi-modal, multi-modal IR-visual and multi-modal MMW-visual images.

The paper was composed with the following structure. First, our method is proposed and explained in Section 2. In Section 3, experimental results are reported. A summary of the significant results are given in Section 4.

2 Proposed fusion method

The aim of image fusion is to make a more informative image of a scene in a single image. Different spatial and frequency transforms provide ways to extract image information. The multi-scale transforms decompose an image to its frequency sub-bands. Every sub-band contains specific information. That is, high-pass bands present details, such as edges and low-pass bands contain outlines like texture and energy.

SR and MST are similar. However, the SR-based method handles images in spatial domain. Since, the SR-based method utilises a universal dictionary (over-complete), it results in better quality than the MST-based method. SR is a very simple model as signals are represented only by few linear combinations of the dictionary atoms. Olshausen and Field could show signal representation of visual cortex. Interestingly, it has three salient features: localised, oriented and band-pass [15]. MST does not contain these features; however, SR dose because SR includes minimisation problem $\|x - Da\|_2^2$, and sparsity problem, $\|\alpha\|_0$ (more explanation has reported in [15]).

In this section, some details on the proposed patch-based sparsity fusion method are mentioned and schematically shown in Fig. 1.

2.1 Patch-based sparsity fusion

If an image matrix is rearranged completely to a single vector, it will cause high-computation cost. For example, an image of size of 256×256 creates a vector of length 65,536 and needs a larger over-

complete dictionary. Therefore, the possibility of choosing a smaller over-complete dictionary paves the way for reducing computation cost. However, on the other hand, this method causes the defect of image blocking. To remove this problem, patches selection occurs with overlapping which itself leads to smoothing the details and edges in the overlapping areas. Hence, it is preferred to segment the image into smaller patches, e.g. patches of size 8×8 . These patches are extracted by a sliding window technique in which the window slides the image from top left-hand side corner to bottom right-hand side corner. The technique benefits from being a shift invariant and spatial-time method. The advantages of the combined MST- and SR-based method over the MST-based or SR-based methods have been comprehensively elaborated in [14].

2.2 Proposed method

The proposed method includes four following steps. Here, just two images have been considered for simplicity.

Step 1: Running MST

Every source image $\{im1, im2\}$ is decomposed by a specific MST into their low-pass $\{low1, low2\}$ and high-pass bands $\{high1, high2\}$.

Step 2: Fusion of the low-pass bands

Every low-pass band $\{low1, low2\}$ is divided into image patches of size $\sqrt{n} \times \sqrt{n}$ by running the sliding window technique from top left-hand side to bottom right-hand side. The number of image patches with image size of $M \times N$ can be given by $(M - \sqrt{n} + 1) \times (N - \sqrt{n} + 1)$ if the window's step length is set to one. Each patch is rearranged to a column vector and subtracted from its mean value. Having a zero mean value of the vector is similar to DC offset removing and known as data centring in signal processing [22].

We denote two corresponding vectors at the same position i , as $x_{i,1}$ and $x_{i,2}$. Sparse coefficients $\{\alpha_{i,1}, \alpha_{i,2}\}$ of the vectors $\{x_{i,1}, x_{i,2}\}$ are computed by orthogonal matching pursuit (OMP) algorithm [23]:

$$\alpha_{i,1} = \arg \min_{\alpha} \|\alpha\|_0 \quad \text{subject to} \quad \|x_{i,1} - D\alpha\|_2 < \varepsilon \quad (1)$$

$$\alpha_{i,2} = \arg \min_{\alpha} \|\alpha\|_0 \quad \text{subject to} \quad \|x_{i,2} - D\alpha\|_2 < \varepsilon \quad (2)$$

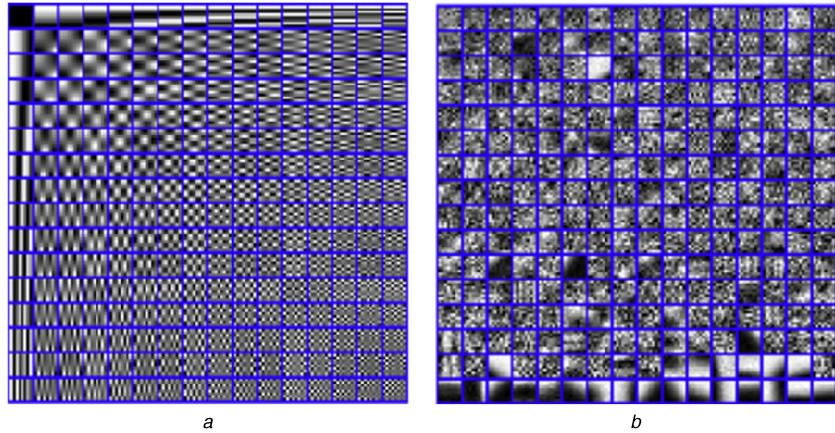


Fig. 2 Dictionary basis are obtained through offline and learning methods
(a) Offline method is Gabor filter, (b) Dictionary is trained by KSVD [21]

where the over-complete dictionary, \mathbf{D} , is obtained through two methods; in the first one, which is offline, the basis is general and independent of signal such as discrete cosine transform. The second method is dictionary learning with consideration of a series of samples (random image patches). The examples of the dictionaries are shown in Fig. 2.

The calculated sparse coefficients should be merged according to a specific fusion rule. The energy in images must be measured to select a proper rule. If their energy differences are significant, the L_2 -norms of the corresponding sparse coefficients $\{\alpha_{i,1}, \alpha_{i,2}\}$ is compared; otherwise, L_1 -norm is used. The ‘max- L_1 -norm’ and ‘max- L_2 -norm’ rules exploit the activity level measurement principles. The image energy is obtained from the sum of squared grey-levels [24] (according to (3)). Representing the source image energy by E_{im} , a suitable fusion rule contains the following steps:

$$E_{im} = \sum_{x=1, y=1}^{x=M, y=N} f^2(x, y), \quad f(x, y) \text{ is image's grey - level} \quad (3)$$

$$\text{if } E_{im1} \cong E_{im2} \rightarrow \alpha_{i,F} = \begin{cases} \alpha_{i,1} & \text{if } \|\alpha_{i,1}\|_1 > \|\alpha_{i,2}\|_1 \\ \alpha_{i,2} & \text{o.w} \end{cases} \quad (4)$$

$$\text{if } E_{im1} \ll E_{im2} \text{ or } E_{im1} \gg E_{im2} \rightarrow \alpha_{i,F} = \begin{cases} \alpha_{i,1} & \text{if } \|\alpha_{i,1}\|_2 > \|\alpha_{i,2}\|_2 \\ \alpha_{i,2} & \text{o.w} \end{cases} \quad (5)$$

where, $\alpha_{i,F}$ refer to fused sparse coefficients and $v_{i,F}$ is defined as fused vector and is calculated by:

$$x_{i,F} = \mathbf{D}\alpha_{i,F} + m_{i,F} \quad (6)$$

Moreover, the fused mean value, $m_{i,F}$, is obtained by (assuming, $m_{i,1}$ and $m_{i,2}$ are mean values of the patches):

$$m_{i,F} = \begin{cases} m_{i,1} & \text{if } \alpha_{i,F} = \alpha_{i,1} \\ m_{i,2} & \text{o.w} \end{cases} \quad (7)$$

The above process is repeated until all the fused vectors of the whole image patches, $x_{i,F}$, are given. These vectors are converted to image patches and finally the patches are put together to construct the low-pass sub-band, ‘low_F’, of the fused image. The block diagram of the low-pass band fusion rule is illustrated in Fig. 3.

Step 3: Fusion of the high-pass bands

The high-pass bands of an image contains the image details, such as edges. More informative coefficients have higher absolute value. Thus, the max-abs is a suitable fusion rule for these bands.

Step 4: Inverse MST

The low-pass and high-pass bands of the ultimate fused image are obtained from the Step 2 and Step 3, respectively. Now, it is the time to perform the inverse MST to attain the fused image from these bands.

2.3 Advantages of the proposed method over individual MST or SR

The MST-based fusion method has several drawbacks including the low-band averaging and losing most of the images energy [14] as mentioned in the first section. For more clarification, an example of multi-modal images is represented in Fig. 4 consist of a pair of visible and MMW images. The source images in the multi-modal fusion are taken from different sensors and each sensor receives different physical characteristics from the same scene. As can be seen from Fig. 4, the visible image shows edges and textures, while the concealed metallic object can be detected as a bright region in the millimetre image [25]. The sum of squared grey-levels is an indication of the image energy. Larger grey-levels mean higher energy lead to the brighter areas. Whereas, the averaging rule is not accurate in high differences values. The proposed method can solve the problem by applying SR on the low-pass bands. In fact, subtraction of the mean value from the patches preserves most of the source images energy and transfers it to the ultimate fused image in the SR method.

Another obstacle with MST-based fusion method is determination of its decomposition levels. To extract enough details, a large number of the decomposition levels must be picked. However, it can arise low-band coefficients which affect many pixels. Therefore, a tiny noise or error appears in the source image (due to miss-registration) and a distorted fused image is produced. Conversely, in the proposed method, spatial details of the low band are extracted using the SR method and four decomposition level is sufficient [14, 26].

On the other hand, the SR-based image fusion causes some bottlenecks. As the sliding windows are overlapped, some fine details, such as the edges, will be smoothed (called the blocking effect). Moreover, the SR method typically results in a high-computational cost [27]. For example, the SR must be run on 62,001 patches for a source image of size 256×256 when the patches size is 8×8 and the sliding window step is 1 [28]. Thus, the SR-based method image fusion may takes several minutes. Our proposed method solves the mentioned problems as the following:

- i. We perform the SR method on the low band of an image. Thus, sliding widow can choose large step since there are no important details in the low-band and they are actually in the high bands (see ref. [14]). As a result, the ‘blocking effect’ is diminished.
- ii. Most MST’s low band of an image is very smaller than the source image. Therefore, their SR computational cost is significantly reduced and the second problem is also obviated.

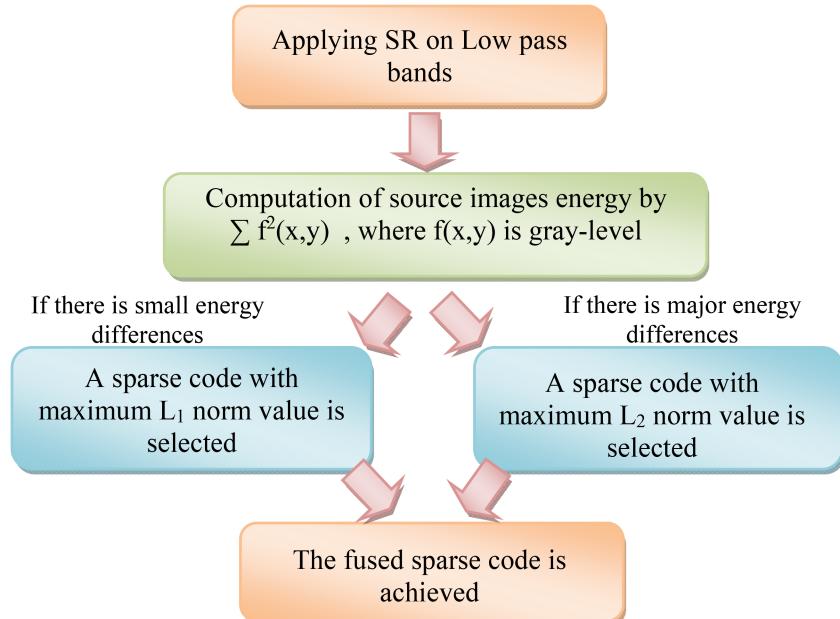


Fig. 3 Block diagram of the low-pass band fusion rule

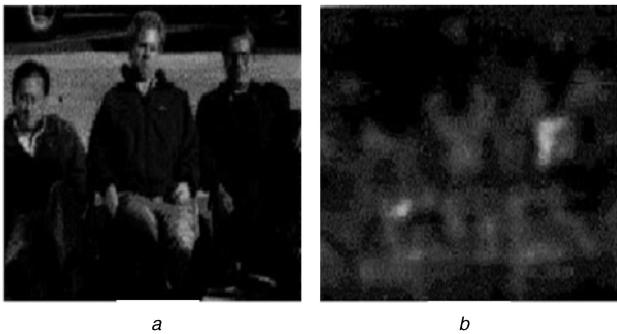


Fig. 4 Multi-modal source images reflect specific attribute of the same scene
(a) Visible image, (b) Millimetre wave image

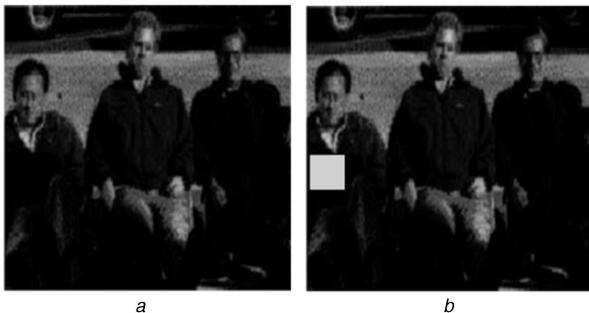


Fig. 5 Applying a change on a small region to evaluate which norm is more sensitive to energy variations
(a) Visible image without changing, (b) Visible image with a small region that is made brighter

2.4 Why L_2 -norm is performed better than L_1 -norm in some cases?

The low-pass band holds most energy of an image. Thus, its fusing process needs more attention. Liu *et al.* merged these bands with the SR-based scheme [14]. Based on their method, the mean value of each patch is subtracted to zero and preserved in SR. Hence, most of the image energy is transferred to the ultimate fused image. They use ‘max- L_1 -norm’ to compare the activity level of coefficients according to (4). However, we prove here that ‘max- L_1 -norm’ is not suitable in all cases, especially when there is significant energy differences in source images. A change in the visible image of Fig. 4 will enable us to investigate this subject. As

shown in Fig. 5, a small region of visible image has got brighter and as a result, the whole image energy is increased. The MST method is performed on each image (the converted and unconverted image) and SR is run on the low-pass band. Then, L_1 - and L_2 -norms of each sparse coefficient are calculated. Fig. 6 shows these norm values of each sparse coefficient for the pre- and post-conversion. In Fig. 6, the norm values before the indicating arrows (placed between 20th to 30th patches) are related to the patches placed in the changed region. In order to define which norm is more affected, the standard deviation (SD) has been measured and listed in Table 1. Notably, the change increases the SD of L_2 -norm, while it is decreased for L_1 -norm. Thus, L_2 is more sensitive to energy changes in the comparison of the two norm. When energy difference of two source images is significant, the activity level is measured by ‘max- L_2 -norm’ criterion rather than ‘max- L_1 -norm’. This is the main idea of the proposed rule.

3 Experimental results

Three popular types of MSTs including DWT, DTCWT and NSCT are performed on the five pairs of different real-life, standard image groups, such as multi-focus, passive millimetre wave (PMMW)- and MMW-visible, IR-visible, and medical (according to Fig. 7) to assess the validity of the proposed method. The real-life PMMW and visible images are employed in this work. In PMMW image, a concealed metallic objects are shown as dark regions. The size of images is 256×256 and assumed they are all registered precisely. The patches size and window step were chosen 8×8 and 1 for the SR, respectively. The OMP error was also set to 0.1. Elad and Aharon [16] dictionary trained by K-SVD algorithm after 180 iterations on 100,000 image patches is used. The dictionary size is also set as 64×256 to guarantee the over-completeness. Quality measurement criteria, such as entropy, SD, mutual information (MI) and $Q^{AB/F}$ are used to evaluate the proposed method performance. In addition, another measurement criterion namely ‘the fused image energy’ was also introduced in the current study in which image energy is calculated by (3).

3.1 Discrete wavelet transform-SR

In this section, each of four methods as DWT, SR, DWT + SR with L_1 -norm and DWT + SR with L_2 -norm are performed on databases shown in Fig. 7 and the fused images are presented in Figs. 8 and 9. Measurement criteria and the results are summarised in Table 2. The ‘db1’ basis and four decomposition levels are chosen for the wavelet. The image size is 256×256 and after four levels of DWT,

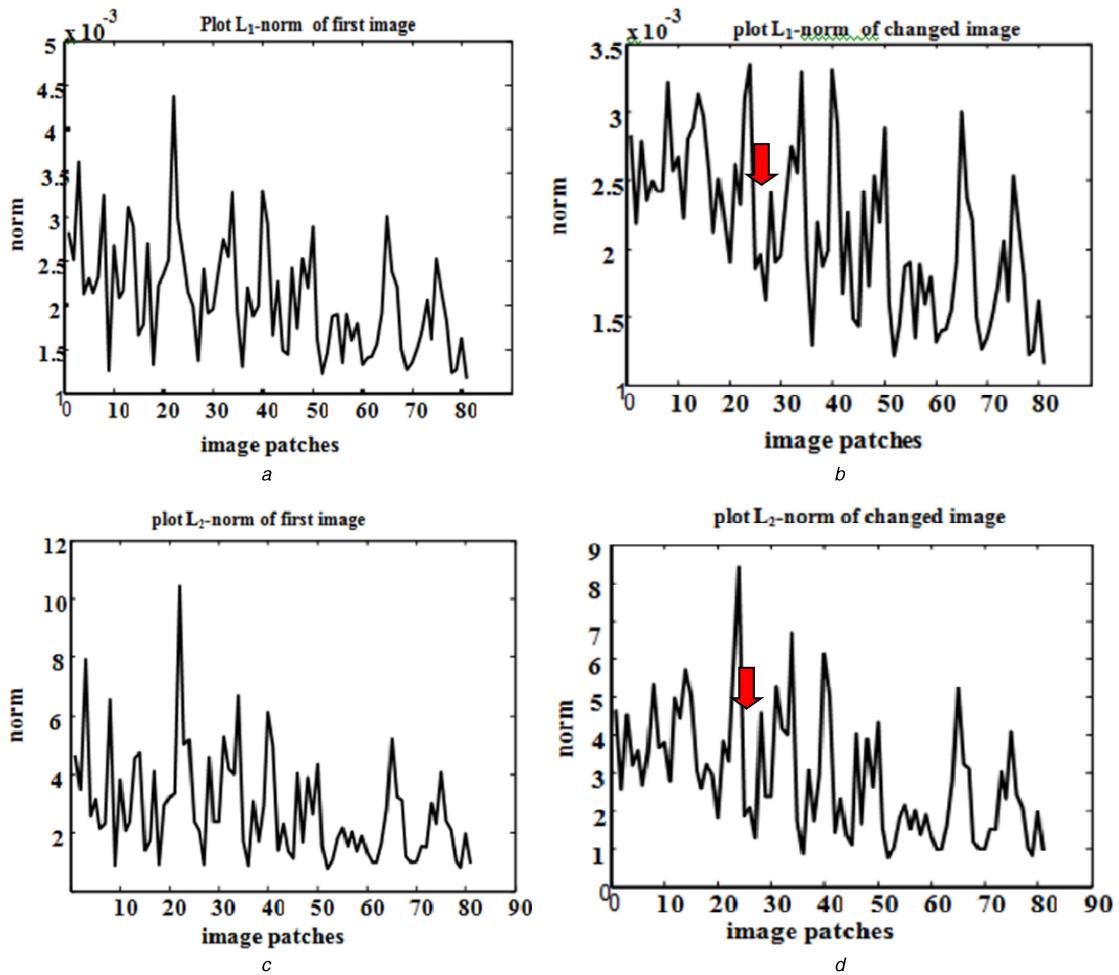


Fig. 6 Plot of L_1 and L_2 norms, before and after increasing the some grey-level of pixels

(a & c) are the L_1 -norm and L_2 -norm of pre-conversion, respectively (Fig. 5a), (b & d) are the L_1 -norm and L_2 -norm of post-conversion, respectively (Fig. 5b)

Table 1 Comparison of L_1 and L_2 norms by application of a change on the image which is depicted in Fig. 5

| | SD of L_1 -norm | SD of L_2 -norm |
|----------------------------------|--------------------|-------------------|
| Image without changing (Fig. 5a) | 6.43×10^3 | 1.5×10^3 |
| Image with changing (Fig. 5b) | 6.38×10^3 | 2×10^7 |

the size of the low-pass sub-bands will be 32×32 . Using the 8×8 sliding window and setting step length to 1, the number of image-patches will be 81. In Table 2, the bold values are the best result of specific criterion among the four methods. As listed in the Table 2, for the source images with large energy gap (such as IR-visible and medical images), the superiority of the proposed method (DWT + SR- L_2 -norm) over DWT + SR- L_1 -norm is obvious. SR has yielded better in some cases; however, it takes long performing time (typically, several minutes). Therefore, it is not suitable for a fast processing.

Among the proposed method criteria, the SD is specifically better than the others. In fact, it is a criterion of contrast and clarity which is related to the image energy. The fused image energy criterion (3) is shown in Table 3 besides the energy of the source images. The proposed method increases the fused image energy.

3.2 Dual-tree complex wavelet transform-SR

Table 4 lists the measurement criterion for DTCWT, SR, DTCWT + SR with L_1 -norm and DTCWT + SR with L_2 -norm methods. Filters named LeGall 5-3 and Qshift-06 are used in DTCWT [29]. The ultimate fused images obtained by performing each method, are also shown in Fig. 10.

3.3 Non-sub-sampled contourlet transform-SR

The NSCT with characteristics of non-subsampled pyramid and non-subsampled directional filter bank has the best performance among the MST methods [30]. On the other hand, the NSCT are shift-invariant. The results of the four methods namely NSCT, SR, NSCT + SR- L_1 -norm and NSCT + SR- L_2 -norm methods are presented in Table 5 and the fused images are shown in Fig. 11. In this paper, ‘pyrexc’ as a pyramid and ‘vk’ as a directional filters are used. The direction number of the four decomposition levels are selected from fine to coarse as 16, 8, 8 and 4, respectively.

4 Conclusion

In the paper, we proposed an energy approach for image fusion. We suggested that the source images energy must be considered prior to image fusion and a suitable fusion rule was declared with regard to the energy differences. A fusion method was expressed based on MST along with SR. First, the source images including five groups of multi-focus, medical, multi-modal visible-IR and visible-MMW and -PMMW were decomposed to their sub-bands using MST. Then, SR was performed on their low-pass bands to produce the sparse codes. The sparse coefficients were merged considering which one had higher activity level. Finally, the high bands were fused with the conventional ‘max-abs’ rule.

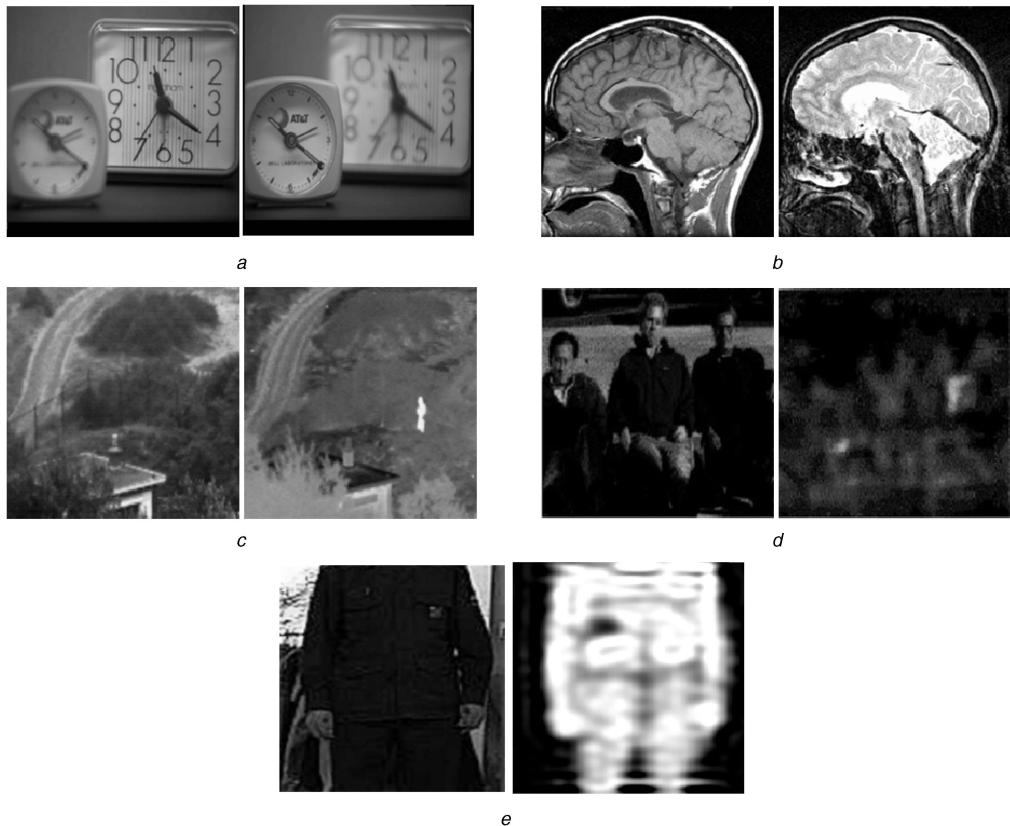


Fig. 7 Databases are in five categories as (from left- to right-hand side)

(a) multi-focus, (b) medical (MRI-CT), (c) multi-modal (visible-infrared), (d) multi-modal (visible-millimetre wave), (e) real-life data (visible-PMMW)

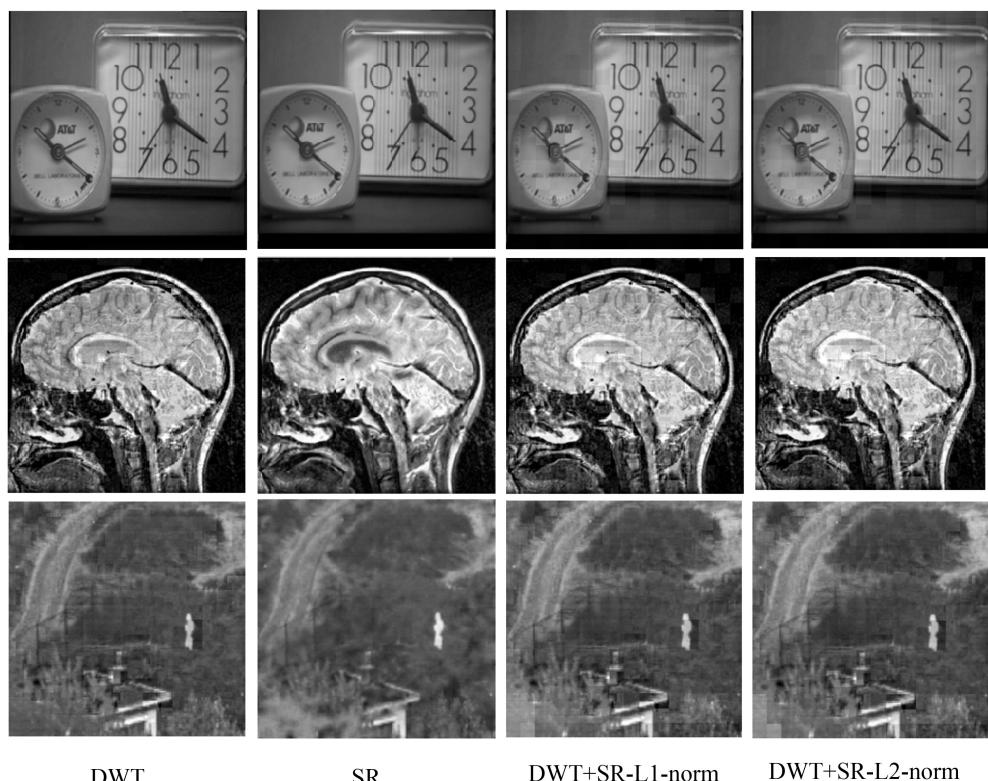


Fig. 8 Results of using DWT, SR, DWT + SR with L_1 -norm and DWT + SR with L_2 -norm of proposed methods (from left- to right-hand side) on following databases

The activity level in SR was measured by two rules: ‘max- L_1 -norm’ or ‘max- L_2 -norm’. For determination of the suitable fusion rule, the energy of the source images was considered. When there was a large energy gap (such as the case for multi-modal images obtained from different sensors), ‘max- L_2 -norm’ was selected; if

not, the max- L_1 -norm was used. The determined rule cause more energy transfer to the fused image and enhanced the contrast and clarity of the ultimate fused image. These improvements were confirmed by numerical criteria. Finally, we compared our proposed method with the latest work. In multi-focus images, the

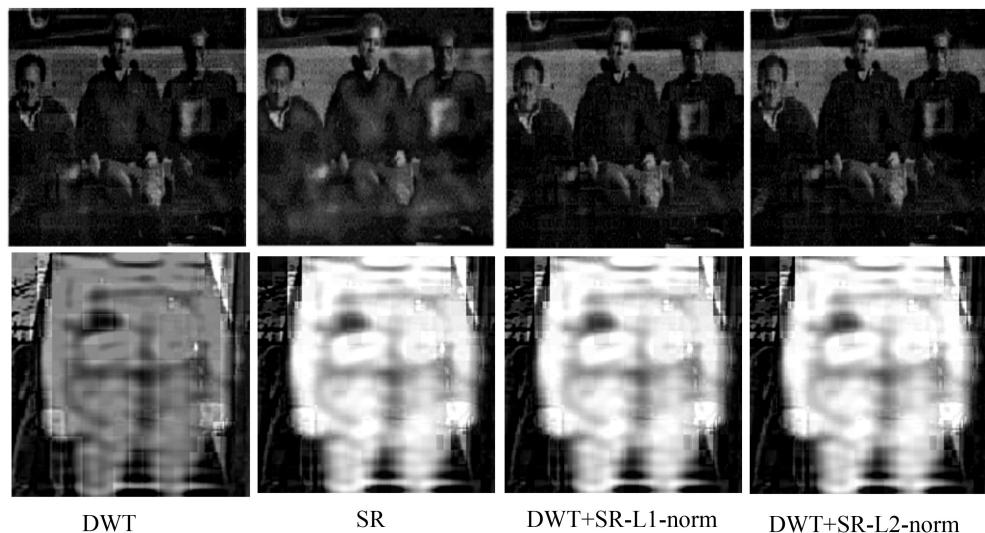


Fig. 9 Results of using DWT, SR, DWT + SR with L_1 -norm and DWT + SR with L_2 -norm of proposed methods (from left- to right-hand side) on following databases

Table 2 Measurement criteria of using the DWT, SR, DWT + SR_L1-norm and DWT + SR_L2-norm (proposed method)

| Images | Metrics | DWT | SR | DWT + SR_L1-norm | DWT + SR_L2norm** |
|---------------|------------------------|-------|-------------|------------------|-------------------|
| Multi-focus | Entropy | 7.32 | 7.31 | 7.4 | 7.4 |
| | SD | 51.24 | 50.76 | 51.27 | 50.96 |
| | MI | 6.88 | 8.16 | 5.92 | 5.93 |
| | Q ^{AB/F} [26] | 0.68 | 0.72 | 0.65 | 0.65 |
| Medical | Entropy | 7.51 | 7.82 | 7.53 | 7.54 |
| | SD | 77.01 | 78.7 | 81.2 | 84 |
| | MI | 3.12 | 4.03 | 3.06 | 3.06 |
| | Q ^{AB/F} | 0.45 | 0.57 | 0.45 | 0.45 |
| Visible-IR | Entropy | 6.64 | 6.54 | 6.8 | 6.92 |
| | SD | 28.05 | 26.83 | 32.48 | 33.5 |
| | MI | 1.28 | 1.36 | 1.4 | 1.5 |
| | Q ^{AB/F} | 0.44 | 0.38 | 0.44 | 0.44 |
| Visible-MMW | Entropy | 6.02 | 6.78 | 5.9 | 5.9 |
| | SD | 34.75 | 36.62 | 40.2 | 40.63 |
| | MI | 1.87 | 4.3 | 1.94 | 2 |
| | Q ^{AB/F} | 0.59 | 0.71 | 0.58 | 0.58 |
| Visible-PMMW* | Entropy | 7.18 | 7.8 | 7.2 | 7.1 |
| | SD | 54.64 | 70.5 | 98.62 | 100.18 |
| | MI | 2.63 | 3.1 | 3.45 | 3.54 |
| | Q ^{AB/F} | 0.54 | 0.62 | 0.5 | 0.49 |

* Real-life data.

** Proposed method.

Table 3 Source images and fused images energise provided by DWT, DWT + SR_L1-norm and DWT + SR_L2-norm

| Image | Energy | Fused image energy | DWT + SR_L1-norm | DWT + SR_L2 norm |
|-----------------------|---------------------|---------------------|---------------------|---------------------|
| multi-focus image one | 7.745×10^8 | 7.893×10^8 | 7.883×10^8 | 7.884×10^8 |
| multi-focus image two | 7.8×10^8 | | | |
| medical CT | 1.224×10^9 | 9.79×10^8 | 1.089×10^9 | 1.23×10^9 |
| medical MRI | 7.437×10^8 | | | |
| visible | 5.603×10^8 | 5.912×10^8 | 6.608×10^8 | 6.662×10^8 |
| infrared | 6.741×10^8 | | | |
| visible | 1.509×10^8 | 1.25×10^8 | 1.607×10^8 | 1.6×10^8 |
| millimetre wave | 1.06×10^8 | | | |
| visible | 2.33×10^8 | 7.35×10^8 | 2.05×10^9 | 2.1×10^9 |
| PMMW | 2.803×10^9 | | | |

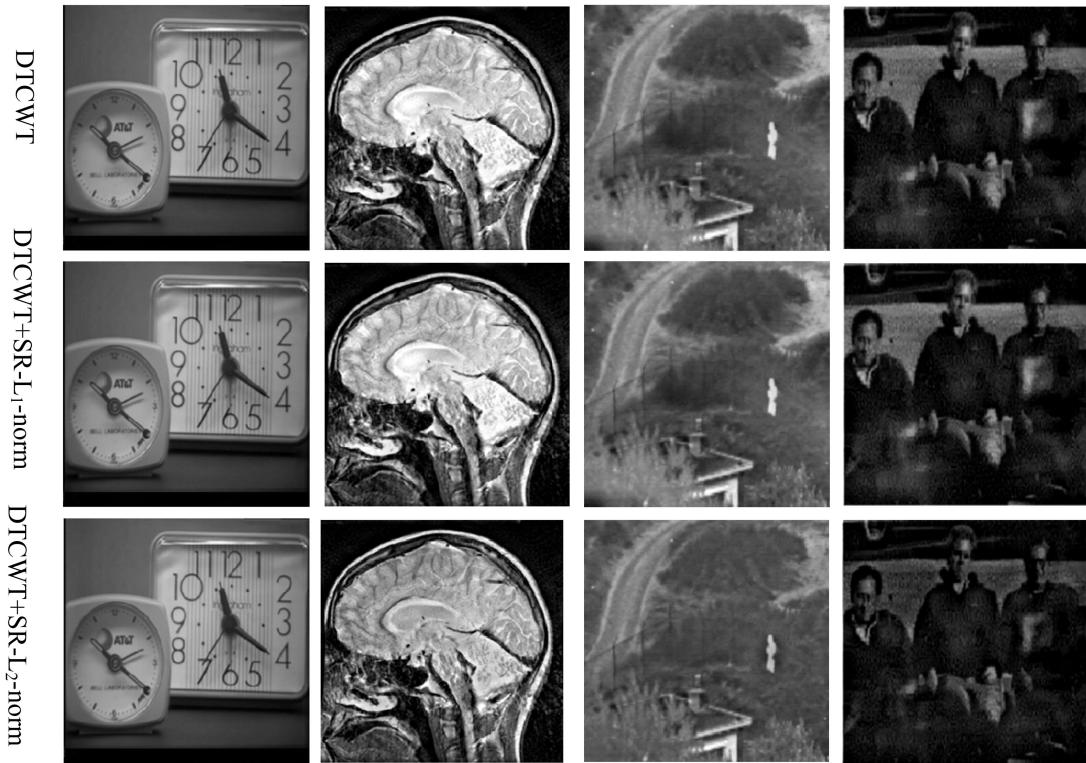


Fig. 10 Results of using DTCWT, DTCWT + SR with L_1 -norm and DTCWT + SR (proposed method) on following four image groups

Table 4 Measurement criteria of using the DTCWT, SR, DTCWT + SR_L₁-norm and DTCWT + SR_L₂-norm (proposed method)

| Images | Metrics | DTCWT | SR | DTCWT + SR_L ₁ -norm | DTCWT + SR_L ₂ -norm* |
|--------------|-------------------|-------|-------------|---------------------------------|----------------------------------|
| multi-focus | Entropy | 7.34 | 7.31 | 7.33 | 7.34 |
| | SD | 50.9 | 50.76 | 51.1 | 51.1 |
| | MI | 7.21 | 8.16 | 7.44 | 7.35 |
| | Q ^{AB/F} | 0.71 | 0.72 | 0.71 | 0.71 |
| medical | Entropy | 7.5 | 7.82 | 7.6 | 7.6 |
| | SD | 74.28 | 78.7 | 86.56 | 86.71 |
| | MI | 3.18 | 4.03 | 3.27 | 3.25 |
| | Q ^{AB/F} | 0.5 | 0.57 | 0.5 | 0.5 |
| visible-IR | Entropy | 6.47 | 6.54 | 6.9 | 7.00 |
| | SD | 26.19 | 26.83 | 33.6 | 33.9 |
| | MI | 1.31 | 1.36 | 1.53 | 1.6 |
| | Q ^{AB/F} | 0.46 | 0.38 | 0.46 | 0.46 |
| visible-MMW | Entropy | 6.08 | 6.78 | 6.5 | 6.5 |
| | SD | 31.81 | 36.62 | 40.26 | 40.34 |
| | MI | 1.8 | 4.3 | 2.3 | 2.4 |
| | Q ^{AB/F} | 0.6 | 0.71 | 0.64 | 0.64 |
| visible-PMMW | Entropy | 7.16 | 7.8 | 7.36 | 7.18 |
| | SD | 50.21 | 70.5 | 91.58 | 97.4 |
| | MI | 2.65 | 3.1 | 3.58 | 3.76 |
| | Q ^{AB/F} | 0.52 | 0.62 | 0.56 | 0.52 |



Fig. 11 Results of using NSCT, NSCT + SR with L_1 -norm and NSCT + SR with L_2 -norm from left- to right-hand side on following databases

Table 5 Measurement criteria of using the NSCT, SR, NSCT + SR_L₁-norm and NSCT + SR_L₂-norm (proposed method)

| Images | Metrics | NSCT | SR | NSCT + SR_L ₁ -norm | NSCT + SR_L ₂ -norm* |
|-------------|-------------------|------------|------------|--------------------------------|---------------------------------|
| multi-focus | Entropy | 7.33 | 7.31 | 7.33 | 7.33 |
| | SD | 51 | 50.76 | 51.05 | 51.1 |
| | MI | 7.33 | 8.16 | 8.3 | 8.28 |
| | Q ^{AB/F} | 0.72 | 0.72 | 0.79 | 0.73 |
| medical | Entropy | 7.46 | 7.82 | 7.8 | 7.85 |
| | SD | 76.6 | 78.7 | 78.08 | 78.9 |
| | MI | 3.3 | 4.03 | 4.08 | 4.1 |
| | Q ^{AB/F} | 0.56 | 0.57 | 0.57 | 0.58 |
| visible-IR | Entropy | 6.54 | 6.54 | 6.71 | 6.75 |
| | SD | 27.05 | 26.83 | 30.37 | 31.06 |
| | MI | 1.31 | 1.36 | 1.62 | 1.7 |
| | Q ^{AB/F} | 0.5 | 0.38 | 0.47 | 0.47 |
| visible-MMW | Entropy | 6.04 | 6.78 | 6.79 | 6.8 |
| | SD | 32.5 | 36.62 | 37.62 | 37.9 |
| | MI | 1.9 | 4.3 | 4.3 | 4.3 |
| | Q ^{AB/F} | 0.65 | 0.71 | 0.72 | 0.72 |

results of the proposed method were similar to those of the state-of-the-arts methods since the energy differences were not significant in these images. On the other hand, the multi-modal images, especially IR-visible images, were significantly improved after applying the method and most of the source images energy was preserved. However, it must be noted that the individual SR gave the best results in medical imaging; however, due to its high-computational cost, the proposed method was more applicable.

Reference

- [1] Lee, H., Yeom, S., Guschin, V.P., *et al.*: ‘Image fusion of visual and millimeter wave images for concealed object detection’, Proc. IEEE Int. Conf. on Infrared and Millimeter Wave, Busan, September 2009, pp. 1–2
- [2] Li, X., He, M., Roux, M.: ‘Multifocus image fusion based on redundant wavelet transform’, *IET Image Process.*, 2010, **4**, (4), pp. 283–293
- [3] Jiang, Y., Wang, M.: ‘Image fusion using multiscale edge-preserving decomposition based on weighted least squares filter’, *IET Image Process.*, 2014, **8**, (3), pp. 183–190
- [4] Sahu, D. K., Parsai, M. P.: ‘Different image fusion techniques’, *Int. J. Mod. Eng. Res.*, 2012, **2**, (5), pp. 4298–4301
- [5] Hossny, M., Nahavandi, S., Creighton, D., *et al.*: ‘Towards autonomous image fusion’. 11th IEEE Int. Conf. on Control Automation Robotics and Vision (ICARCV), 2010, pp. 1748–1754
- [6] Wald, L.: ‘Some terms of reference in data fusion’, *IEEE Trans. Geosci. Remote Sens.*, 1999, **37**, (3), pp. 1190–1193
- [7] Liu, Y., Wang, Z.: ‘Simultaneous image fusion and denoising with adaptive sparse representation’, *IET Image Process.*, 2014, **9**, (5), pp. 347–357
- [8] Li, H., Manjunath, B., Mitra, S.: ‘Multi-sensor image fusion using the wavelet transform’, *Graph. Models Image Process.*, 1995, **57**, (3), pp. 235–245
- [9] Lewis, J., O’Callaghan, R., Nikolov, S., *et al.*: ‘Pixel and region-based image fusion with complex wavelets’, *Inf. Fusion*, 2007, **8**, (2), pp. 119–130
- [10] Nencini, F., Garzelli, A., Baronti, S., *et al.*: ‘Remote sensing image fusion using the curvelet transform’, *Inf. Fusion*, 2007, **8**, (2), pp. 143–156
- [11] Liu, Z., Tsukada, K., Hanasaki, K., *et al.*: ‘Image fusion by using steerable pyramid’, *Pattern Recog. Lett.*, 2001, **22**, pp. 929–939
- [12] Zhang, Q., Guo, B.: ‘Multi-focus image fusion using the non-subsampled contourlet transform’, *Signal Process.*, 2009, **89**, (7), pp. 1334–1346
- [13] Gao, G., Xu, L., Feng, D.: ‘Multi-focus image fusion based on non-subsampled shearlet transform’, *IET Image Process.*, 2013, **7**, (6), pp. 633–639
- [14] Liu, Y., Liu, Sh., Wang, Z.: ‘A general framework for image fusion based on multi-scale transform and sparse representation’, *Inf. Fusion.*, 2015, **24**, pp. 147–164
- [15] Olshausen, B. A., Field, D. J.: ‘Emergence of simple-cell receptive field properties by learning a sparse code for natural images’, *Nature*, 1996, **381**, (6583), pp. 607–609
- [16] Elad, M., Aharon, M.: ‘Image denoising via sparse and redundant representations over learned dictionaries’, *IEEE Trans. Image Process.*, 2006, **15**, (12), pp. 3736–3745
- [17] Yang, B., Li, S.T.: ‘Multi-focus image fusion and restoration with sparse representation’, *IEEE Trans. Instrum. Meas.*, 2010, **59**, (4), pp. 884–892
- [18] Yin, H., Li, Sh.: ‘Multimodal image fusion with joint sparsity model’, *Opt. Eng., SPIE*, 2011, **6700**, pp. 1–11
- [19] Yao, Y., Guo, P., Xin, X., *et al.*: ‘Image fusion by hierarchical joint sparse representation’, *Cogn. Comput.*, 2014, **6**, (3), pp. 281–292
- [20] Yang, B., Li, S.: ‘Visual attention guided image fusion with sparse representation’, *Opt.-Int. J. Light Electron. Opt.*, 2014, **125**, (17), pp. 4881–4888
- [21] Xiaoqi, L., Baohua, Zh., Ying, Zh., *et al.*: ‘The infrared and visible image fusion algorithm based on target separation and sparse representation’, *Infrared Phys. Tech.*, 2014, **67**, pp. 397–407
- [22] Mairal, J., Bach, F., Ponce, J.: ‘Sparse modeling for image and vision processing’, 2014, arXiv preprint arXiv, p. 1411.3230
- [23] Cai, T., Wang, L.: ‘Orthogonal matching pursuit for sparse signal recovery with noise’, *IEEE Trans. Inf. Theory*, 2011, **57**, (7), pp. 4680–4687
- [24] Lewis, J. P.: ‘Fast normalized cross-correlation.’, *Vis. Interface*, 1995, **10**, (1), pp. 120–123
- [25] Pergande, A. N.: ‘History and challenges of passive millimeter wave imaging’, *SPIE Def. Secur. Sens Intern. Soc. Opt. Photonics*, **2013**, pp. 890005–890006
- [26] Li, S., Yang, B., Hu, J.: ‘Performance comparison of different multi-resolution transforms for image fusion’, *Inf. Fusion*, 2011, **12**, (2), pp. 74–84
- [27] Elad, M., Yavneh, I.: ‘A plurality of sparse representations is better than the sparsest one alone’, *IEEE Trans. Inf. Theory*, 2009, **55**, (10), pp. 4701–4714
- [28] Yang, B., Li, S.: ‘Pixel-level image fusion with simultaneous orthogonal matching pursuit’, *Inf. Fusion*, 2012, **13**, (1), pp. 10–19
- [29] Selesnick, I. W., Baraniuk, G. R., Kingsbury, N.: ‘The dual-tree complex wavelet transform’, *IEEE Signal Proc. Mag.*, 2005, **22**, pp. 123–151
- [30] Shi, T., Jintao, X., Jianyu, Y., *et al.*: ‘Research on image fusion for visual and PMMW images using NSCT and region detection’. In 3rd Int. Conf. on Multimedia Technology (ICMT-13), Atlantis Press, Nov. 2013