Published in IET Image Processing
Received on 9th April 2014
Revised on 2nd September 2014
Accepted on 13th September 2014
doi: 10.1049/iet-ipr.2014.0311

ISSN 1751-9659

Simultaneous image fusion and denoising with adaptive sparse representation

Yu Liu[1], *Zengfu Wang*[1,2]

[1]*Department of Automation, University of Science and Technology of China, Hefei 230026, People's Republic of China*
[2]*Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, People's Republic of China*
E-mail: liuyu1@mail.ustc.edu.cn

Abstract: In this study, a novel adaptive sparse representation (ASR) model is presented for simultaneous image fusion and denoising. As a powerful signal modelling technique, sparse representation (SR) has been successfully employed in many image processing applications such as denoising and fusion. In traditional SR-based applications, a highly redundant dictionary is always needed to satisfy signal reconstruction requirement since the structures vary significantly across different image patches. However, it may result in potential visual artefacts as well as high computational cost. In the proposed ASR model, instead of learning a single redundant dictionary, a set of more compact sub-dictionaries are learned from numerous high-quality image patches which have been pre-classified into several corresponding categories based on their gradient information. At the fusion and denoising processes, one of the sub-dictionaries is adaptively selected for a given set of source image patches. Experimental results on multi-focus and multi-modal image sets demonstrate that the ASR-based fusion method can outperform the conventional SR-based method in terms of both visual quality and objective assessment.

1 Introduction

Image fusion technique aims at generating a composite image by integrating the complementary information from multiple source images of the same scene [1]. The source images can be obtained from either an imaging sensor whose optical parameters can be adjusted or different kinds of sensors. The composite/fused image will be more suitable for human perception or machine processing than any individual source image. Recently, image fusion technique has been widely employed in machine vision, remote sensing, medical imaging etc.

During the past two decades, various image fusion algorithms have been developed. Generally, these methods can be grouped into two categories: spatial domain fusion (SDF) and transform domain fusion (TDF) [2]. The SDF methods usually deal with image blocks [3, 4] or segmented regions [5]. The basic principle is to select image blocks or regions from source images with a certain activity level measurement [6]. However, the fused results of block-based methods usually suffer from blocking effect even though the block size can be optimised [4]. It is also difficult for the region-based methods to stably obtain high-quality result since image segmentation is really a tough task. The TDF methods share a 'decomposition-fusion-reconstruction' framework [7], namely, decompose the source images into a multi-scale domain, fuse the transformed coefficients by a given rule and reconstruct the fused image with the merged coefficients. Traditional multi-scale transforms used in TDF methods include morphological pyramid [8], gradient pyramid [9], discrete wavelet transform (DWT) [10], dual tree complex wavelet transform (DTCWT) [11], curvelet transform (CVT) [12], non-subsampled contourlet transform (NSCT) [13], non-subsampled shearlet transform [14] etc. The core idea behind these methods is that the underlying salient information of the source images can be extracted from the transformed coefficients. Most recently, some new TDF methods based on multi-scale filtering have been introduced, such as the multi-scale edge-preserving decomposition-based method [15] and the neighbour distance (ND)-based method [16]. Sparse representation (SR) has been successfully employed in many image processing applications including denoising [17], super-resolution [18, 19] and fusion [20–25]. The SR-based fusion methods can also be classified into the TDF methods since the activity level of source images is measured in the sparse domain.

In image fusion literature, compared with the great concentration on fusing clear source images, less attention has been paid to develop effective fusion algorithms for noisy ones. Since images are often corrupted by noise during acquisition or transmission, it has practical significance to study the subject of noisy image fusion. This paper mainly focuses on this point. Particularly, the main contribution of this paper is that a novel adaptive SR (ASR) model is presented for simultaneous image fusion and denoising. In traditional SR-based applications like denoising and fusion, a highly redundant dictionary is usually needed to satisfy signal reconstruction requirement because the structures of different image patches vary significantly. However, the study in [26] shows that a

highly redundant dictionary may lead to potential visual artefacts in the reconstruction result, especially when the input signal is corrupted by noise. A simple explanation is that when a dictionary has too many atoms, the possibility that one atom and a noisy input signal own similar structure is relative high. Thus, based on the SR theory, the input signal may be just replaced by the corresponding atom, rather than being denoised. Another problem caused by a redundant dictionary is the computational efficiency. For the SR-based image applications, the computational cost will increase a lot when the dictionary size (the number of atoms) becomes larger. To overcome the above two shortcomings of conventional SR-based fusion methods, we propose a novel ASR model for simultaneous image fusion and denoising in this paper. In the ASR model, instead of learning a single redundant dictionary, a set of more compact sub-dictionaries is learned from numerous high-quality image patches which have been pre-classified into several corresponding categories based on their gradient information. At the fusion and denoising processes, one of the sub-dictionaries is adaptively selected for a given set of source image patches.

A preliminary version of this work appeared in [25], where we only tested the algorithm on multi-focus images and the experiments were not very sufficient. In this paper, we conduct much more experiments on both multi-focus and multi-modal image sets to further confirm the effectiveness of the proposed method. Moreover, the descriptions of both the ASR model and the fusion method are more scientific and accurate here. This paper is structured as follows. Section 2 introduces some related work about SR and its applications on image fusion and denoising. The ASR model is introduced in Section 3. Section 4 presents the ASR-based fusion scheme in detail. The experimental results are given in Section 5. Finally, Section 6 concludes this paper.

2 Related work

On the basis of the physiological characteristics of human visual system, SR is presented to address the inherent sparsity of natural signals [27]. The basic assumption behind SR is that a signal $\boldsymbol{x} \in \boldsymbol{R}^n$ can be approximately represented by a linear combination of a 'few' atoms from a redundant dictionary $\boldsymbol{D} \in \boldsymbol{R}^{n \times m} (n < m)$, where n is the signal dimension and m is the dictionary size. That is, the signal $\boldsymbol{x}$ can be expressed as $\boldsymbol{x} \simeq \boldsymbol{D}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \in \boldsymbol{R}^m$ is the unknown sparse coefficient vector. As the dictionary is over-complete, there are numerous feasible solutions for this underdetermined system of equations. The target of SR is to calculate the sparsest $\boldsymbol{\alpha}$ which contains the fewest non-zero entries among all feasible solutions. Mathematically, the sparest $\boldsymbol{\alpha}$ can be obtained with the following sparse model

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{\alpha}||_0 \text{ s.t.} ||\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\alpha}||_2 < \varepsilon \tag{1}$$

where $\varepsilon > 0$ is an error tolerance and $\| \bullet \|_0$ denotes the l_0-norm which counts the number of non-zero entries. A popular method to solve this non-deterministic polynomial-hard problem is the orthogonal matching pursuit (OMP) [28], which iteratively updates the sparse vector by selecting the most relevant atom to the given signal.

There are two main categories of offline approaches to obtain a dictionary. The first one is using the analytical models such as discrete cosine transform and DWT, which own the advantages of simplicity and fast implementation. However, this category of dictionary is often restricted to signals of a certain type and cannot be used for an arbitrary family of signals. The second category is based on machine learning technique. Suppose that M image patches of size $\sqrt{n} \times \sqrt{n}$ are sampled from natural images and rearranged to column vectors in the $\boldsymbol{R}^n$ space, thereby the training database $\{\boldsymbol{y}_i\}_{i=1}^M$ is constructed with each $\boldsymbol{y}_i \in \boldsymbol{R}^n$. The dictionary learning model can be presented as

$$\min_{\boldsymbol{D}, \{\boldsymbol{\alpha}_i\}_{i=1}^M} \sum_{i=1}^{M} \|\boldsymbol{\alpha}_i\|_0 \text{ s.t. } \|\boldsymbol{y}_i - \boldsymbol{D}\boldsymbol{\alpha}_i\|_2 < \varepsilon, \quad i = 1, 2, \ldots, M \tag{2}$$

where $\varepsilon > 0$ is an error tolerance, $\{\boldsymbol{\alpha}_i\}_{i=1}^M$ is the unknown sparse vector corresponding to $\{\boldsymbol{y}_i\}_{i=1}^M$ and $\boldsymbol{D} \in \boldsymbol{R}^{n \times m}$ is the unknown dictionary to be learned. The K-singular value decomposition (K-SVD) algorithm [29] is one of the most commonly used methods to solve this problem.

As a powerful signal modelling technique, SR has achieved a great success in image processing community in recent years. Elad and Aharon [17] presented an SR-based image denoising method with learned dictionary. In [17], the image is assumed to be corrupted by additive white Gaussian noise with zero-mean and standard deviation σ. Note that the SR-based denoising process is performed locally on small image patches with overlaps, namely, the sliding-window technique is employed. For each image patch $\boldsymbol{x}$, the error tolerance in (1) is adaptively set as $\varepsilon = \sqrt{n} C \sigma$ [30] to calculate the sparse vector $\boldsymbol{\alpha}$

$$\min_{\boldsymbol{\alpha}} ||\boldsymbol{\alpha}||_0 \text{ s.t. } ||\boldsymbol{x} - \boldsymbol{D}\boldsymbol{\alpha}||_2 < \sqrt{n} C \sigma \tag{3}$$

where $C > 0$ is a constant. Then, with the learned dictionary, the corresponding denoised patch $\boldsymbol{y}$ is reconstructed by

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{\alpha} \tag{4}$$

Finally, all the processed patches are put back into the image coordinate space to form the denoised image. Since each pixel may be simultaneously involved by several patches, its final value should be averaged over the number of related patches.

SR was first introduced into image fusion by Yang and Li in [20]. The sliding-window technique is also adopted in their method to make the fusion process more robust to noise and mis-registration. The most important contribution of [20] is the sparse coefficient vector used to measure the activity level of source images. Particularly, among all the source sparse vectors, the one owning the maximal l_1-norm is selected as the fused sparse vector ('max-$L1$' fusion rule). The fused image is reconstructed with all the fused sparse vectors and the dictionary. Many improved SR-based fusion methods [21–25] have been proposed since this fundamental work appeared. These publications show that the SR-based methods, which own clear advantages over traditional multi-scale transform (e.g. DTCWT and NSCT) based methods, are capable of leading to state-of-the-art results.

3 ASR model

Recently, Dong *et al.* [19] presented an SR-based image super-resolution method by dividing the sparse domain with several compact sub-dictionaries. However, there are two

doi: 10.1049/iet-ipr.2014.0311

concerns about the sub-dictionary learning approach proposed in [19]. First, the natural image patches used for sub-dictionary learning are clustered using the K-means algorithm, so the learning process will be very time-consuming if the number of training patches is very large (e.g. 100 000). Second, and more importantly, the reasonability of clustering result is not easy to ensure since the K-means algorithm depends much on the initial centre of each cluster. In this paper, instead of applying the unsupervised learning technique like the K-means clustering, we use a supervised classification approach to group each patch into a category based on its gradient information. The detailed sub-dictionary learning scheme is described as follows.

To construct the training set for dictionary learning, numerous image patches of size $\sqrt{n} \times \sqrt{n}$ are randomly sampled from a set of high-quality natural images. The mean value of each sampled patch is subtracted to zero before training to guarantee that the mean value of each atom in the learned dictionary is also zero. Then, to ensure that only the patches with enough edge structures are included in the training set, a patch whose intensity variance is less than a given threshold will be excluded [18, 19]. Let $\boldsymbol{P} = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_M\}$ denote the training set which is formed by M preserved meaningful image patches.

As mentioned above, the patches in set $\boldsymbol{P}$ should be classified into several categories to learn a series of compact sub-dictionaries. To this end, the gradient dominant direction of a patch is employed for classification. Like the orientation assignment approach in the scale-invariant feature transform descriptor [31], we obtain a patch's dominant direction using a gradient orientation histogram which is formed from the gradients of all the inside pixels. For each $\boldsymbol{p}_i \in \boldsymbol{P}$, its horizontal gradient $\boldsymbol{G}_x(x, y)$ and vertical gradient $\boldsymbol{G}_y(x, y)$ are obtained with the 'Sobel' operator

$$\boldsymbol{G}_x(x, y) = \boldsymbol{p}_i(x, y) * \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \tag{5}$$

$$\boldsymbol{G}_y(x, y) = \boldsymbol{p}_i(x, y) * \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \tag{6}$$

where * denotes the two-dimensional convolution operation. The gradient magnitude $\boldsymbol{G}(x, y)$ and orientation $\boldsymbol{\Theta}(x, y)$ are calculated by

$$\boldsymbol{G}(x, y) = \sqrt{\boldsymbol{G}_x(x, y)^2 + \boldsymbol{G}_y(x, y)^2} \tag{7}$$

$$\boldsymbol{\Theta}(x, y) = \tan^{-1}(\boldsymbol{G}_y(x, y)/\boldsymbol{G}_x(x, y)) \tag{8}$$

Suppose that an orientation histogram has K bins equally dividing the 360° range of orientations. Fig. 1*a* shows the example of $K = 6$. To form the histogram for patch $\boldsymbol{p}_i$, the gradient orientation $\boldsymbol{\Theta}(x, y)$ at each pixel (x, y) in $\boldsymbol{p}_i$ is first quantised into one of the K bins, and then the gradient magnitude $\boldsymbol{G}(x, y)$ is added to the corresponding bin. Let $\{\theta_1, \theta_2, \ldots, \theta_K\}$ denote the obtained orientation histogram. Naturally, the bin having the highest peak in the histogram is corresponding to the dominant direction of $\boldsymbol{p}_i$. However, some irregular patches may not have clear dominant directions, that is, the differences among the K bins in their histograms are not very distinctive. In this situation, since a compact sub-dictionary is learned from the patches which have relative similar structures, it is more reasonable to use a dictionary learned from the patches having various structures. Thus, there are totally $K + 1$ sub-dictionaries $\{\boldsymbol{D}_0, \boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_K\}$, in which $\boldsymbol{D}_0$ is learned from all the patches in $\boldsymbol{P}$, whereas $\{\boldsymbol{D}_k | k = 1, \ldots, K\}$ is learned from the patches in a corresponding subset $\{\boldsymbol{P}_k | k = 1, \ldots, K\}$ of $\boldsymbol{P}$. Let k_i denote the index of $\boldsymbol{P}_k$ that the patch $\boldsymbol{p}_i$ should be grouped into. On the basis of the above considerations, k_i is obtained with the following classification rule

$$k_i = \begin{cases} 0, & \text{if } \dfrac{\theta_{\max}}{\sum_{k=1}^{K} \theta_k} < \dfrac{2}{K} \\ k^*, & \text{otherwise} \end{cases} \tag{9}$$

where $\theta_{\max} = \max\{\theta_1, \theta_2, \ldots, \theta_K\}$ is the highest peak in the histogram and $k^* = \arg\max_k \{\theta_k | k = 1, 2, \ldots, K\}$ is the index of $\theta_{\max}$. $k_i = 0$ means that the p_i is an irregular patch. This rule indicates that a patch will be viewed as an irregular patch if the ratio $\theta_{\max} / \sum_{k=1}^{K} \theta_k$ is smaller than $2/K$. Otherwise, the patch will be classified into the subset $\boldsymbol{P}_{k^*}$.

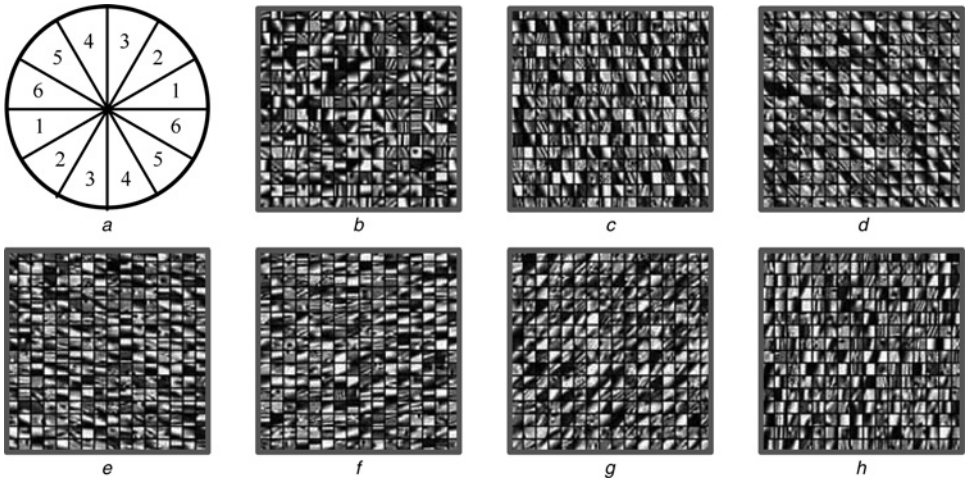

Fig. 1 *Learning sub-dictionaries in the ASR model*

a Illustration of dividing the 360° range of orientations
b–h Learned sub-dictionaries

By applying the rule (9) to all training patches in $\boldsymbol{P}$, $\{\boldsymbol{P}_k|k=1, \ldots, K\}$ can be constructed. Finally, we learn the $K+1$ sub-dictionaries $\{\boldsymbol{D}_0, \boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_K\}$ with the K-SVD algorithm from $\{\boldsymbol{P}_k|k=1, \ldots, K\}$. An example of $K=6$ is given in Figs. 1*b–h*. In this example, we randomly sampled 100 000 image patches of size 8×8 from 50 natural images. The size of each sub-dictionary is set to 256. After using $\text{var}(p)<10$ to remove the smooth patches, the obtained $\boldsymbol{P}$ consists of 82 576 patches. Then, via rule (9), each set of $\{\boldsymbol{P}_k|k=1, \ldots, K\}$ obtains about 10 000 patches. Fig. 1*b* shows the sub-dictionary $\boldsymbol{D}_0$ learned from all the 82 576 patches. The other six sub-dictionaries are shown in Figs. 1*c–h*, from which we can see that the atoms in each sub-dictionary generally have consistent gradient directions.

4 Simultaneous fusion and denoising method

The schematic diagram of our ASR-based fusion algorithm is shown in Fig. 2. Suppose that there are J pre-registered source images $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots \boldsymbol{I}_J\}$ of size $H\times W$. All the source images are assumed to be corrupted by additive white Gaussian noise with zero-mean and standard deviation σ ($\sigma=0$ means the images are clear). With the learned sub-dictionaries $\{\boldsymbol{D}_0, \boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_K\}$, the detailed fusion scheme consists of the following six steps.

Step 1: For each source image $\boldsymbol{I}_j$, apply the sliding-window technique to extract all possible patches of size $\sqrt{n}\times\sqrt{n}$ with a step length of one pixel from upper left to lower right. Let $\{\boldsymbol{s}_1^i, \boldsymbol{s}_2^i, \ldots, \boldsymbol{s}_J^i\}_{i=1}^N$ denote a set of patches in $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots \boldsymbol{I}_J\}$ with the same position i, where $N=(H-\sqrt{n}+1)\times(W-\sqrt{n}+1)$ is the number of extracted patches in each source image.

Step 2: For each position i, rearrange $\{\boldsymbol{s}_1^i, \boldsymbol{s}_2^i, \ldots, \boldsymbol{s}_J^i\}$ into column vectors $\{\boldsymbol{v}_1^i, \boldsymbol{v}_2^i, \ldots, \boldsymbol{v}_J^i\}$, and then normalise the mean value of each $\boldsymbol{v}_j^i$ in $\{\boldsymbol{v}_1^i, \boldsymbol{v}_2^i, \ldots, \boldsymbol{v}_J^i\}$ to zero to obtain $\hat{\boldsymbol{v}}_j^i$ by

$$\hat{\boldsymbol{v}}_j^i = \boldsymbol{v}_j^i - \bar{v}_j^i \cdot \mathbf{1} \tag{10}$$

where $\bar{v}_j^i$ is the mean value over all the elements of $\boldsymbol{v}_j^i$ and $\mathbf{1}$ denotes an all-one valued $n\times 1$ vector.

Step 3: Among $\{\hat{\boldsymbol{v}}_1^i, \hat{\boldsymbol{v}}_2^i, \ldots, \hat{\boldsymbol{v}}_J^i\}$, pick out the $\hat{\boldsymbol{v}}_m^i$ which has the largest variance. Then, construct the gradient orientation histogram for $\hat{\boldsymbol{v}}_m^i$ based on the approach presented in Section 3, and select one sub-dictionary from $\{\boldsymbol{D}_0, \boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_K\}$ using the rule in (9). Let $\boldsymbol{D}_{k_i}$ denotes the adaptively selected sub-dictionary.

Step 4: Calculate the sparse coefficient vector $\boldsymbol{\alpha}_j^i$ of each $\hat{\boldsymbol{v}}_j^i$ in $\{\hat{\boldsymbol{v}}_1^i, \hat{\boldsymbol{v}}_2^i, \ldots, \hat{\boldsymbol{v}}_J^i\}$ with $\boldsymbol{D}_{k_i}$ by

$$\min_{\boldsymbol{\alpha}_j^i} ||\boldsymbol{\alpha}_j^i|| \text{ s.t. } ||\hat{\boldsymbol{v}}_j^i - \boldsymbol{D}_{k_i}\boldsymbol{\alpha}_j^i||_2 < \sqrt{n}C\sigma + \varepsilon \tag{11}$$

where $C>0$ is a constant and $\varepsilon>0$ is an error tolerance.

Step 5: Merge the J obtained sparse vectors $\{\boldsymbol{\alpha}_1^i, \boldsymbol{\alpha}_2^i, \ldots, \boldsymbol{\alpha}_J^i\}$ with the 'max-$L1$' fusion rule [20, 21]. The detailed reason that the 'max-$L1$' rule is appropriate for image fusion can be found in [21]. Specifically, the l_1-norm of the sparse vector is employed as the activity level measurement of the corresponding source image patch, and the fused sparse vector $\boldsymbol{\alpha}_{\text{F}}^i$ of $\{\boldsymbol{\alpha}_1^i, \boldsymbol{\alpha}_2^i, \ldots, \boldsymbol{\alpha}_J^i\}$ is obtained with the following rule

$$\boldsymbol{\alpha}_{\text{F}}^i = \boldsymbol{\alpha}_{j^*}^i, \quad j^* = \arg\max_j \{||\boldsymbol{\alpha}_j^i||_1|, \quad j=1, 2, \ldots, J\} \tag{12}$$

The merged mean value $\bar{v}_{\text{F}}^i$ of $\{\bar{v}_1^i, \bar{v}_2^i, \ldots, \bar{v}_J^i\}$ is accordingly set as

$$\bar{v}_{\text{F}}^i = \bar{v}_{j^*}^i \tag{13}$$

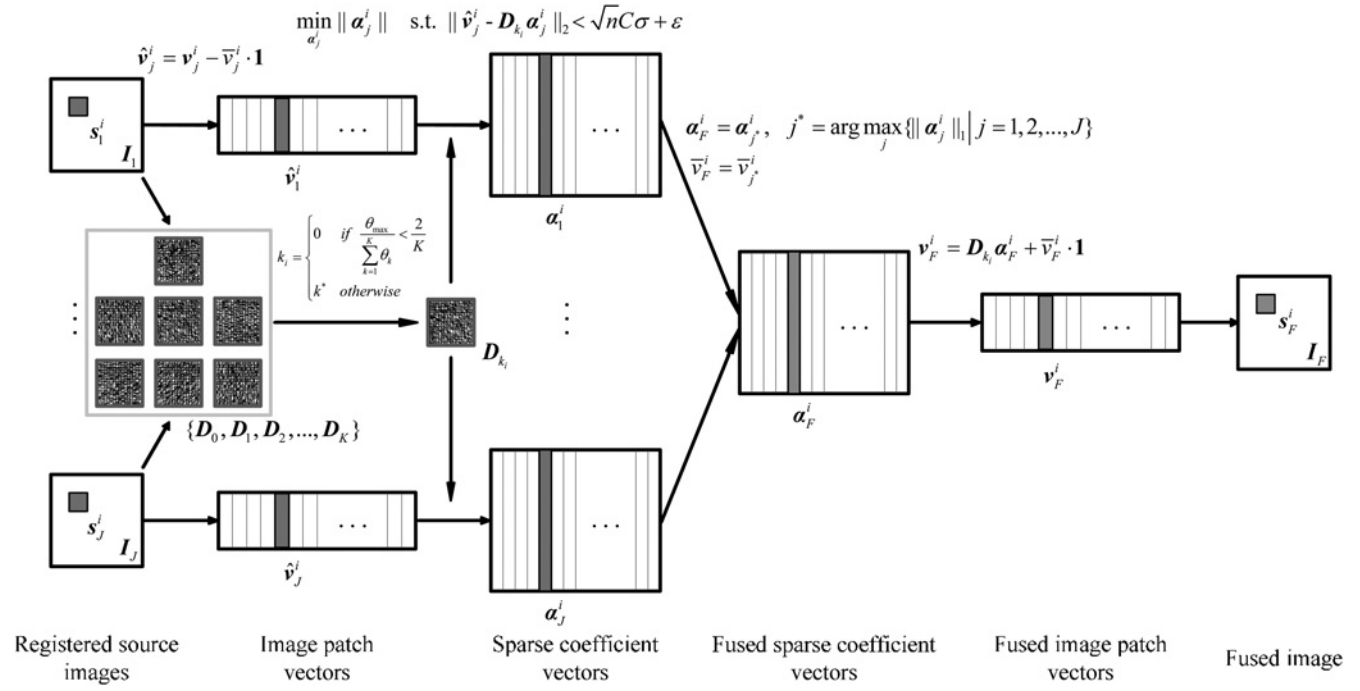

Fig. 2 *Schematic diagram of the ASR-based fusion algorithm*

doi: 10.1049/iet-ipr.2014.0311

Finally, the fused result of $\{\boldsymbol{v}_1^j, \boldsymbol{v}_2^j, \ldots, \boldsymbol{v}_J^j\}$ is calculated by

$$\boldsymbol{v}_{\mathrm{F}}^j = \boldsymbol{D}_{k_j}\boldsymbol{\alpha}_{\mathrm{F}}^j + \bar{v}_{\mathrm{F}}^j \cdot \mathbf{1} \tag{14}$$

Step 6: Iterate steps 2–5 for all the N sets of source image patches in $\{\boldsymbol{s}_1^i, \boldsymbol{s}_2^i, \ldots, \boldsymbol{s}_J^i\}_{i=1}^N$ to obtain the fused results $\{\boldsymbol{v}_{\mathrm{F}}^i\}_{i=1}^N$. Let $\boldsymbol{I}_{\mathrm{F}}$ denote the fused image. For each $\boldsymbol{v}_{\mathrm{F}}^i$, reshape it into a $\sqrt{n} \times \sqrt{n}$ patch $\boldsymbol{s}_{\mathrm{F}}^i$ and then plug $\boldsymbol{s}_{\mathrm{F}}^i$ into its original position in $\boldsymbol{I}_{\mathrm{F}}$. As the patches are overlapped, $\boldsymbol{I}_{\mathrm{F}}$ is finally obtained with each pixel's value being averaged over its accumulation times.

5 Experiments

In this section, we first present the detailed experimental setups including test image sets and various methods for comparison. Then, the fusion metrics used for quantitative assessment are briefly introduced. Finally, the experimental results are exhibited and discussed.

5.1 Experimental setups

In our experiments, the proposed method is tested on 12 pairs of multi-focus images and 8 pairs of multi-modal images, as shown in Fig. 3. To verify the effectiveness of the proposed method on fusing noisy images, zero-mean Gaussian noises with standard deviations of 5, 10, 15 and 20 are artificially added to these clear source images.

One advantage of most TDF methods is they can naturally integrate the denoising process into the fusion framework. For the fusion methods based on multi-scale transforms such as CVT and NSCT, the transform coefficients of noisy source images are first denoised. Specifically, the coefficients with absolute values less than some given thresholds are set to zero while others remain unchanged. Then, the denoised source coefficients are merged. For the SR-based fusion method, the combination of fusion and denoising is even more straightforward, which have been introduced before. However, it is often difficult for most SDF methods to combine the denoising and fusion together. Therefore, to evaluate the proposed method objectively, only TDF methods are compared in this paper.

The ASR-based method is compared with five popular or latest TDF methods based on DWT [10], CVT [12], NSCT [13], ND [16] and SR [20]. The parameters of these methods are set as follows. For the DWT, CVT, NSCT and ND methods, the decomposition levels are all set to 4. The low-frequency bands are merged with the 'averaging' rule, whereas the 'choosing maximum absolute' rule with a 3×3 window-based consistency verification scheme [10] is adopted to fuse high-frequency bands. For the DWT method, the wavelet basis 'db4' is applied. For the NSCT method, we use the 'pyrexc' filter as the pyramid filter and the 'vk' filter as the directional filter. Moreover, the direction numbers of the four decomposition levels are selected as 4, 8, 8 and 16. All the above parameters are set according to the results in a survey paper [32], which fully investigates the optimal parameter settings for various multi-scale transform-based fusion methods. For the SR and ASR methods, the patch size is set to 8×8 and the error tolerance ε is set to 0.1 according to the study in [20]. Furthermore, either the SR or ASR method has two versions with different dictionary sizes of 128 and 256. Let SR-128, SR-256, ASR-128 and ASR-256 denote the four

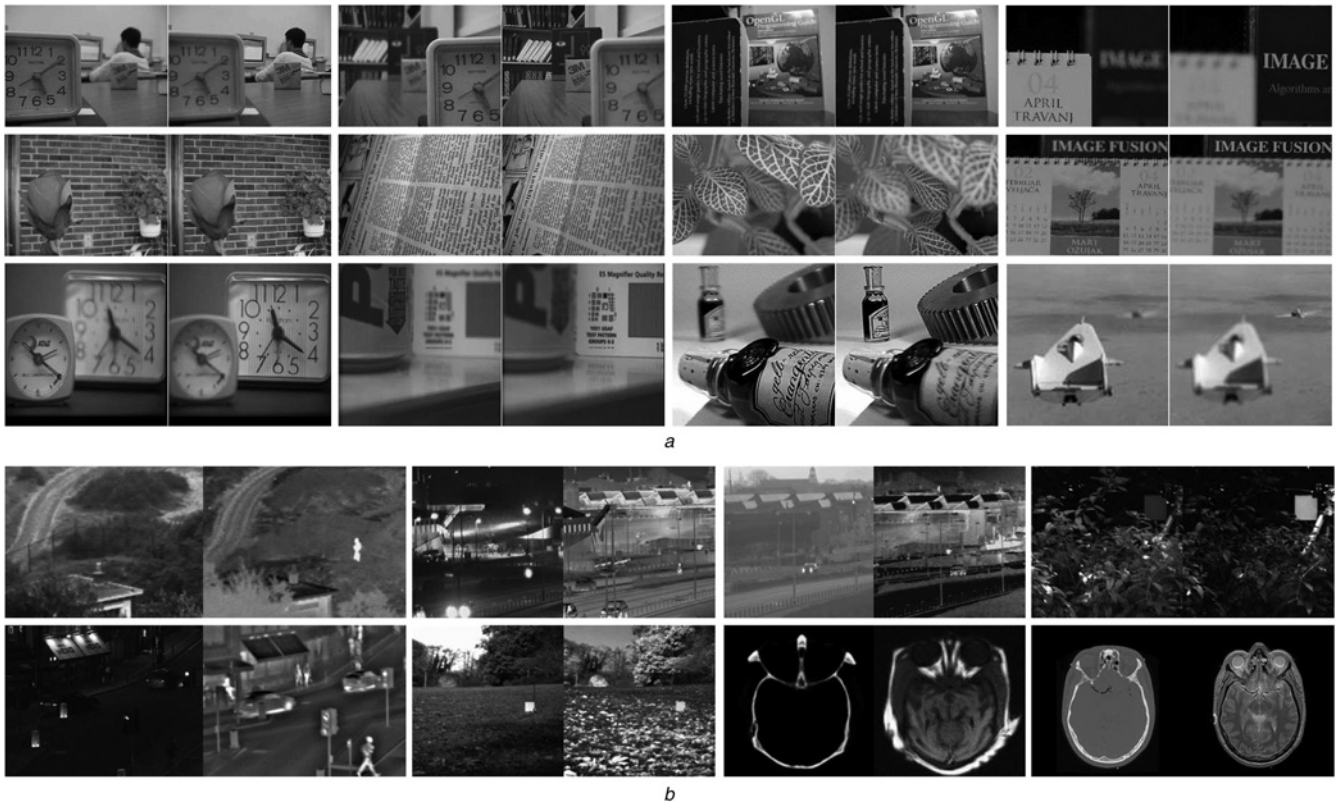

Fig. 3 *Source images used in our experiments*

a 12 pairs of multi-focus images
b 8 pairs of multi-modal images

corresponding methods. All the dictionaries are learned with the K-SVD method using default parameters given in [29]. For either ASR-128 or ASR-256, the number of sub-dictionaries is fixed to 7, that is, $K = 6$ in (9).

When the source images are corrupted by noise, we only compare the ASR method with CVT, NSCT and SR methods for the following two reasons. First, in image denoising applications, the advantages of the CVT [33], NSCT [34] and SR [17] based methods over traditional multi-resolution (e.g. DWT) methods have been widely verified. Second, the latest ND method [16] is only employed to fuse clear images. For the CVT and NSCT methods, the hard thresholding approaches presented in [33, 34] are used for source image denoising, respectively. For SR-128, SR-256, ASR-128 and ASR-256, the parameter C in (11) is set to 1.15 [17].

5.2 Objective evaluation metrics

It is not an easy task to quantitatively evaluate the quality of a fused image because the reference image (ground truth) does not exist in practice. Although many fusion metrics have been proposed, none of them is universally believed to be more reasonable than others. Thus, it is necessary to use several metrics for evaluation. According to a recent survey by Liu *et al.* [35], the fusion metrics can be classified into four categories: information theory-based metrics, image feature-based metrics, image structural similarity-based metrics and human perception-based metrics. In this paper, we choose one metric from each of the above four categories. The four selected metrics are briefly introduced as follows. Uniformly, let $\boldsymbol{A}$ and $\boldsymbol{B}$ denote two source images of size $H \times W$, whereas $\boldsymbol{F}$ as the fused image.

(1) Normalised mutual information (MI) Q_{MI} [36]. Q_{MI} is an information theory-based fusion metric which can overcome the instability of traditional MI metric [37]. The definition of Q_{MI} is

$$Q_{\mathrm{MI}} = 2\left[\frac{\mathrm{MI}(\boldsymbol{A}, \boldsymbol{F})}{H(\boldsymbol{A}) + H(\boldsymbol{F})} + \frac{\mathrm{MI}(\boldsymbol{B}, \boldsymbol{F})}{H(\boldsymbol{B}) + H(\boldsymbol{F})}\right] \tag{15}$$

where $H(\boldsymbol{X})$ is the entropy of image $\boldsymbol{X}$ and $\mathrm{MI}(\boldsymbol{X}, \boldsymbol{Y})$ is the MI between images $\boldsymbol{X}$ and $\boldsymbol{Y}$. Q_{MI} calculates the amount of information in $\boldsymbol{F}$ obtained from $\boldsymbol{A}$ and $\boldsymbol{B}$.

(2) Gradient-based fusion metric Q_{G} [38]. Q_{G} is an image feature-based metric. It is a popular fusion metric which evaluates the extent of gradient information extracted from the source images. Q_{G} is calculated by

$$Q_G = \frac{\sum_{x=1}^{\mathrm{H}} \sum_{y=1}^{\mathrm{W}} (\boldsymbol{Q}^{\mathrm{AF}}(x, y)\boldsymbol{w}^A(x, y) + \boldsymbol{Q}^{\mathrm{BF}}(x, y)\boldsymbol{w}^B(x, y))}{\sum_{x=1}^{\mathrm{H}} \sum_{y=1}^{\mathrm{W}} (\boldsymbol{w}^A(x, y) + \boldsymbol{w}^B(x, y))} \tag{16}$$

where $\boldsymbol{Q}^{\mathrm{AF}}(x, y) = \boldsymbol{Q}_g^{\mathrm{AF}}(x, y)\boldsymbol{Q}_\alpha^{\mathrm{AF}}(x, y)$, $\boldsymbol{Q}_g^{\mathrm{AF}}(x, y)$ and $\boldsymbol{Q}_\alpha^{\mathrm{AF}}(x, y)$ denote the edge strength and orientation preservation values at pixel (x, y). The definition of $\boldsymbol{Q}^{\mathrm{BF}}(x, y)$ is same with that of $\boldsymbol{Q}^{\mathrm{AF}}(x, y)$. The weighting factors $\boldsymbol{w}^A(x, y)$ and $\boldsymbol{w}^B(x, y)$ indicate the significances of $\boldsymbol{Q}^{\mathrm{AF}}(x, y)$ and $\boldsymbol{Q}^{\mathrm{BF}}(x, y)$, respectively.

(3) Yang's fusion metric Q_Y [39]. Q_Y is a structural similarity-based fusion metric which measures the level of structural information of source images preserved in the fused image. Q_Y is defined as (see (17))

where SSIM is structural similarity [40] between two images, w is a local window and $\lambda(w)$ is the local weight.

(4) Chen–Blum metric Q_{CB} [41]. Q_{CB} is a human perception-based fusion metric which addresses the major features in human visual system. The calculation of Q_{CB} is relative complex and more details can be found in [41].

For each of the four metrics, a larger value indicates a better fused result. To guarantee the objectivity of evaluation results, all the four metrics are calculated using an evaluation toolbox implemented by *Z*. Liu (the first author of [35]).

5.3 Experimental results and discussion

Figs. 4*a* and *b* show two clear multi-focus source images, between which a slight motion of the student's head exists. The fused results of different methods are shown in Figs. 4*c–j*. The fused image of the DWT method suffers serious artefacts since DWT is not shift-invariant. The fused results of the CVT, NSCT and ND methods are much better, but the artefacts around the student's head still exists mainly because the low-frequency bands are simply merged with the 'averaging' rule. The fused images of SR-128, SR-256, ASR-128 and ASR-256 are in high visual quality, but the differences among Figs. 4*g–j* are not very obvious in terms of visual perception.

A fusion example of clear visible and infrared images is shown in Fig. 5. The visible image shown in Fig. 5*a* captures more spatial details such as the brushwood and fence, whereas thermal objects such as the person can be easily extracted from the infrared image shown in Fig. 5*b*. Figs. 5*c–j* show the different fused results. Similarly, the artefacts in the fused image of the DWT method are the most serious. The CVT, NSCT and ND methods can extract enough spatial details from the visible image, but the person in their fused results looks darker than that in other results. The fused results of SR-128, SR-256, ASR-128 and ASR-256 are in high quality for both spatial details and thermal objects are well preserved. With more careful observation, we can see that the spatial information is more accurate in the fused image obtained by ASR-256 than those of SR-128, SR-256 and ASR-128 (see the edges of house in the scene).

Fig. 6 shows a fusion example for noisy multi-focus images. The two noisy images shown in Figs. 6*c* and *d* are obtained by injecting zero-mean white Gaussian noise with $\sigma = 10$ into the two source images shown in Figs. 6*a* and *b*, respectively. The different fused results obtained from the noising inputs are shown in Figs. 6*e–j*. The results of the CVT and NSCT methods suffer from both low visibility and serious artefacts. The fused images of SR-128 and SR-256 have much better visual quality, but some undesirable artefacts still exist (see the top border of the far clock in the scene). The fused images obtained by ASR-128 and ASR-256 own high clarity without obvious artefacts.

$$Q_Y = \begin{cases} \lambda(w)\mathrm{SSIM}(\boldsymbol{A}, \boldsymbol{F}|w) + (1 - \lambda(w))\mathrm{SSIM}(\boldsymbol{B}, \boldsymbol{F}|w), & \mathrm{SSIM}(\boldsymbol{A}, \boldsymbol{B}|w) \geq 0.75 \\ \max\{\mathrm{SSIM}(\boldsymbol{A}, \boldsymbol{F}|w), \mathrm{SSIM}(\boldsymbol{B}, \boldsymbol{F}|w)\}, & \mathrm{SSIM}(\boldsymbol{A}, \boldsymbol{B}|w) < 0.75 \end{cases} \tag{17}$$

doi: 10.1049/iet-ipr.2014.0311

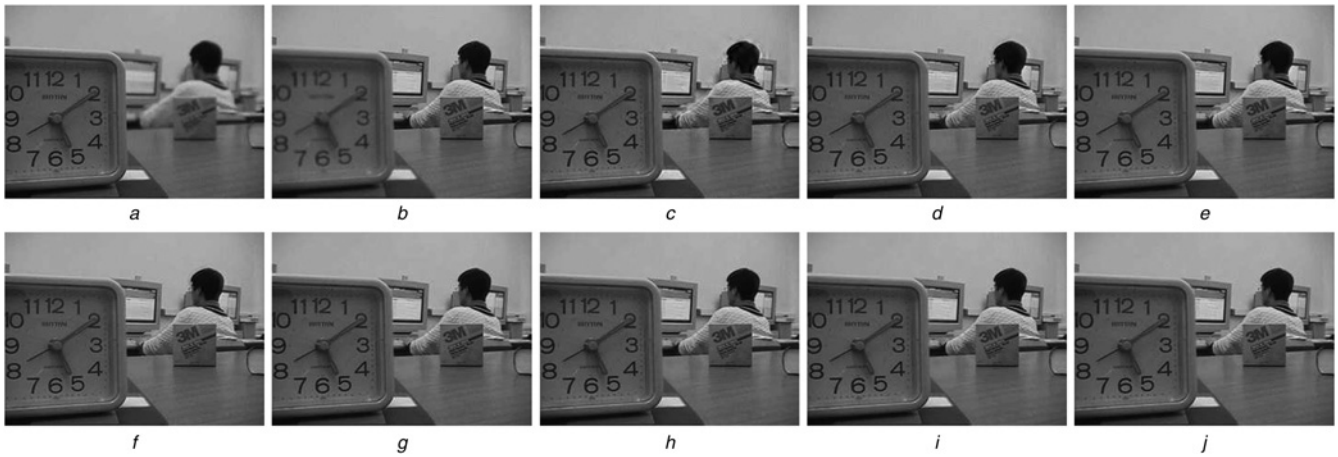

Fig. 4 *Example of clear multi-focus image fusion*

a Foreground-focused source image
b Background-focused source image
c Fused image by DWT
d Fused image by CVT
e Fused image by NSCT
f Fused image by ND
g Fused image by SR-128
h Fused image by SR-256
i Fused image by ASR-128
j Fused image by ASR-256

The last example shown in Fig. 7 is given for noisy medical image fusion. Fig. 7*a* shows a computed tomography (CT) image which mainly illuminates the bone structure, whereas Fig. 7*b* shows a magnetic resonance imaging (MRI) image which addresses the soft tissue structures. In clinical medicine, to help doctors make more accurate diagnoses, it is very worthwhile to compose a single image which simultaneously contains the information of CT and MRI images. The arrangement of Fig. 7 is same as that of Fig. 6 and the standard deviation of injected Gaussian noise is also 10. The fused results of the last four methods have higher contrast in soft tissue structures than those of the CVT and NSCT methods. Some important bond structural information is lost by SR-128 and SR-256, whereas ASR-128 and ASR-256 can obtain more useful information from two source images.

Tables 1 and 2 list the average quantitative assessments of different methods on multi-focus images and multi-modal images, respectively. The maximum in each line is shown in bold, which implies the best performance over all

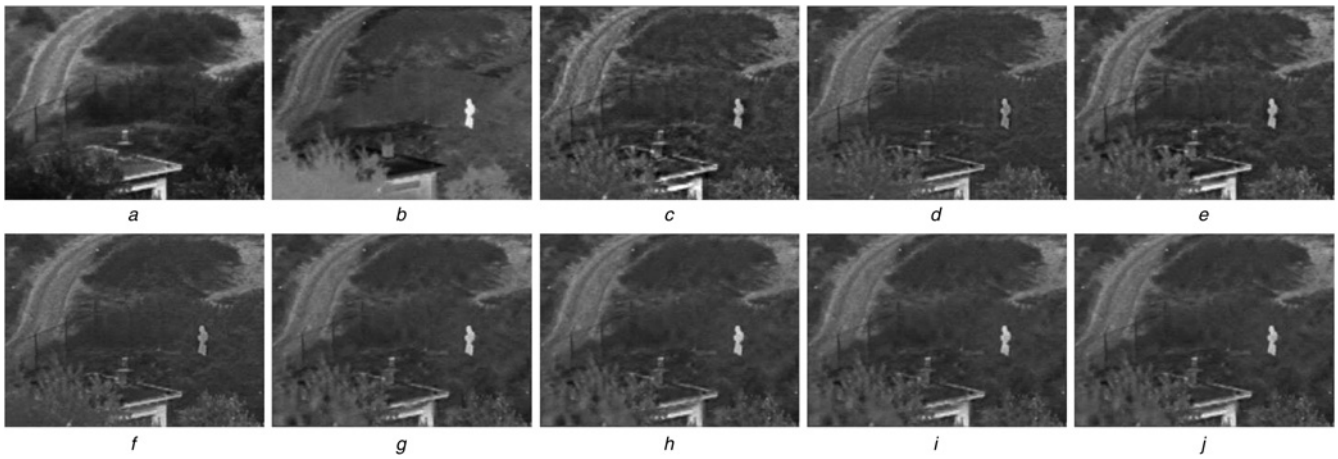

Fig. 5 *Example of clear multi-modal image fusion*

a Visible image
b Infrared image
c Fused image by DWT
d Fused image by CVT
e Fused image by NSCT
f Fused image by ND
g Fused image by SR-128
h Fused image by SR-256
i Fused image by ASR-128
j Fused image by ASR-256

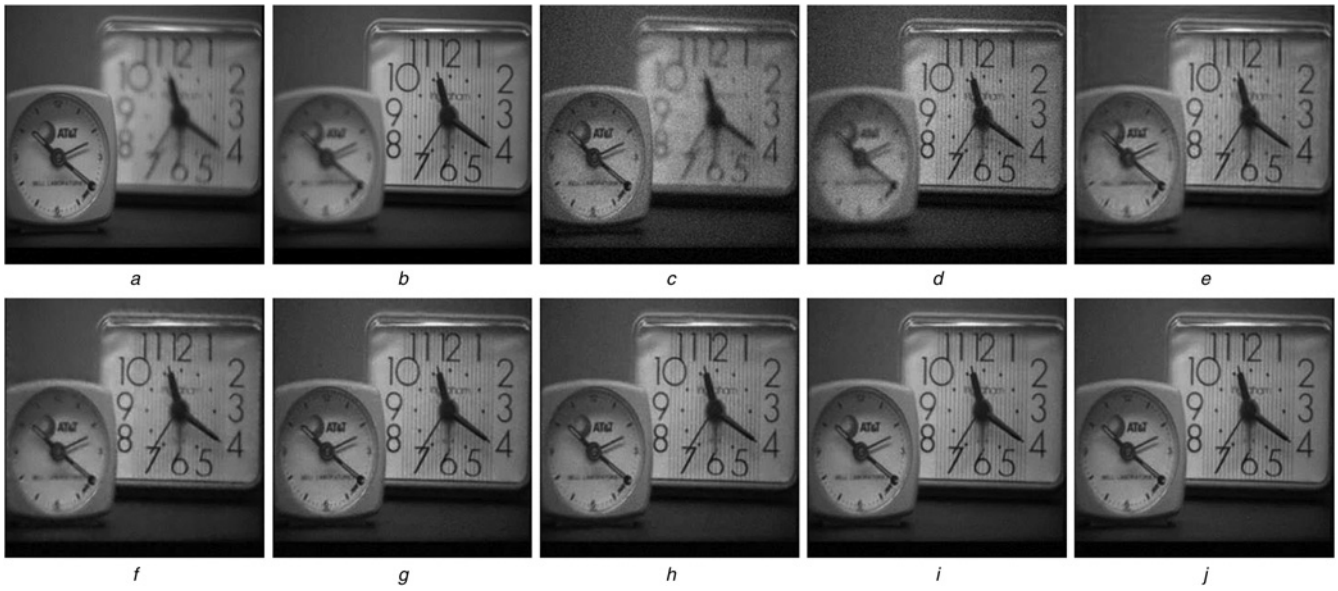

Fig. 6 *Example of noisy multi-focus image fusion*

a Foreground-focused clear source image
b Background-focused clear source image
c Foreground-focused noisy source image
d Background-focused noisy source image
e Fused image by CVT
f Fused image by NSCT
g Fused image by SR-128
h Fused image by SR-256
i Fused image by ASR-128
j Fused image by ASR-256

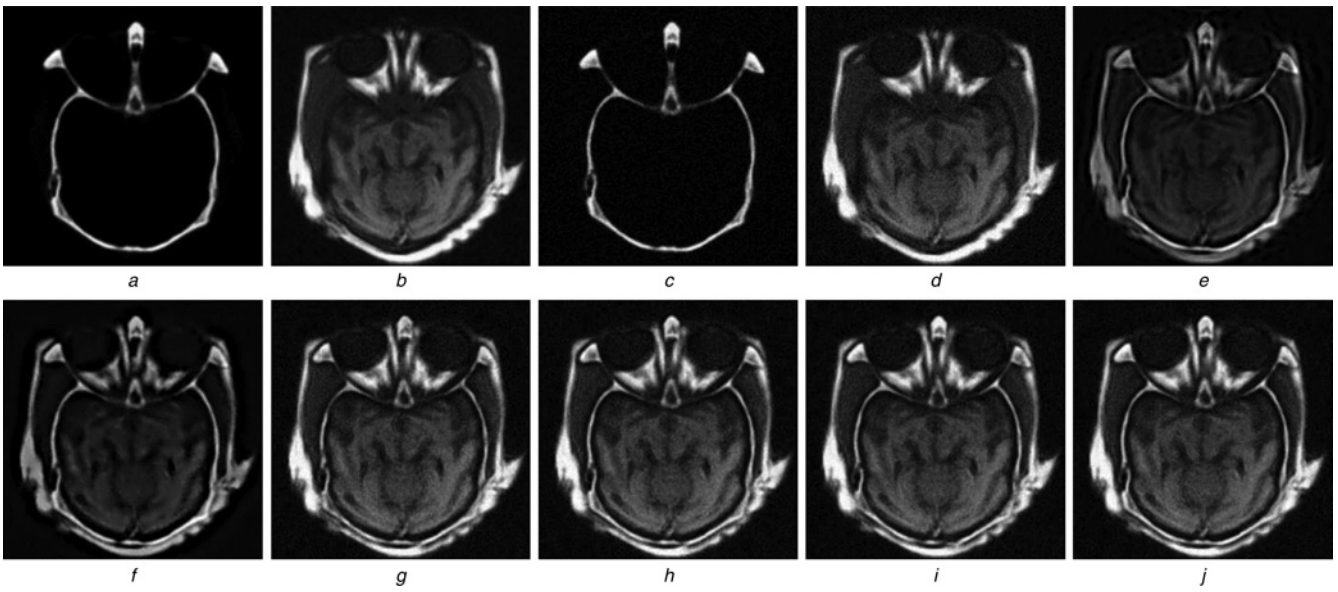

Fig. 7 *Example of noisy multi-modal image fusion*

a Clear CT image
b Clear MRI image
c Noisy CT image
d Noisy MRI image
e Fused image by CVT
f Fused image by NSCT
g Fused image by SR-128
h Fused image by SR-256
i Fused image by ASR-128
j Fused image by ASR-256

doi: 10.1049/iet-ipr.2014.0311

Table 1 Average quantitative assessments of different methods on 12 pairs of multi-focus images

Fusion metrics	Standard deviations	Methods							
		DWT	CVT	NSCT	ND	SR-128	SR-256	ASR-128	ASR-256
Q_{MI}	0	0.7906	0.8473	0.9024	0.9082	1.0871	**1.0925**	1.0886	1.0914
	5	—	0.7336	0.7975	—	0.8726	**0.8794**	0.8765	0.8783
	10	—	0.6760	0.7436	—	0.7890	0.7942	0.7946	**0.7970**
	15	—	0.6398	0.7072	—	0.7336	0.7369	0.7392	**0.7398**
	20	—	0.6122	0.6703	—	0.6981	0.6977	**0.7008**	0.6996
Q_G	0	0.6579	0.6934	0.7157	0.7168	0.7291	0.7294	0.7282	**0.7296**
	5	—	0.6123	0.6328	—	0.6550	0.6537	**0.6580**	0.6575
	10	—	0.5498	0.5677	—	0.5929	0.5944	0.5970	**0.5981**
	15	—	0.4965	0.5078	—	0.5437	0.5434	**0.5495**	0.5478
	20	—	0.4528	0.4525	—	0.4987	0.4994	**0.5058**	0.5055
Q_Y	0	0.8541	0.8819	0.9252	0.9358	0.9442	**0.9457**	0.9440	0.9453
	5	—	0.7227	0.7561	—	0.7846	0.7861	0.7875	**0.7902**
	10	—	0.6439	0.6804	—	0.6953	0.6966	0.6983	**0.6998**
	15	—	0.5912	0.6278	—	0.6422	0.6426	**0.6484**	0.6479
	20	—	0.5491	0.5832	—	0.6053	0.6040	**0.6119**	0.6107
Q_{CB}	0	0.6885	0.6921	0.7382	0.7245	0.7551	0.7562	0.7555	**0.7564**
	5	—	0.6587	0.6739	—	0.6930	0.6935	0.6951	**0.6957**
	10	—	0.6227	0.6289	—	0.6522	0.6517	**0.6575**	0.6569
	15	—	0.5948	0.5919	—	0.6242	0.6251	**0.6303**	0.6289
	20	—	0.5705	0.5611	—	0.5985	0.6003	0.6099	**0.6106**

methods. The underlying characteristics behind Tables 1 and 2 can be mainly summarised into the following three points.

(1) The SR- and ASR-based methods clearly outperform traditional multi-scale transform-based methods in terms of both clear and noisy situations. Thus, we only concentrate on the four SR- and ASR-based methods next.
(2) When the source images are clear, the performance order is generally ASR-256 > SR-256 > ASR-128 > SR-128, where the symbol '>' indicates 'better than'. However, the differences among these four methods are very slight.
(3) When the source images are corrupted by noise, the order is generally {ASR-128, ASR-256} > {SR-128, SR-256} with considerable differences, which indicates that the proposed ASR model is more effective than the conventional SR model in this situation. Furthermore, ASR-256 > ASR-128 is not always valid especially when the standard deviation is larger than 10, and so is SR-256 > SR-128. In other words, when the corrupted noise is serious, a larger dictionary size may not lead to a better result. The main reason for this is that a highly redundant dictionary may cause potential artefacts in the reconstruction result, which has been explained in Section 1.

At last, we compare the computational efficiency of the four SR- and ASR-based methods. For clear comparison, we normalise the averaging running time of ASR-128 to a basic time unit (about 120 s for fusing two 256×256 source images in MATLAB on a 3 GHz computer). The results are listed in Table 3, from which we can see that the computational cost of the ASR model only increase by about 15% relative to the SR model when the dictionary size is fixed. Moreover, ASR-128 is more efficient than SR-256. Considering that ASR-128 can broadly compete with SR-256 in clear situation and outperform it in noisy

Table 2 Average quantitative assessments of different methods on 8 pairs of multi-modal images

Fusion metrics	Standard deviations	Methods							
		DWT	CVT	NSCT	ND	SR-128	SR-256	ASR-128	ASR-256
Q_{MI}	0	0.3623	0.3866	0.4185	0.4582	0.5033	0.5076	0.5069	**0.5118**
	5	—	0.3733	0.4056	—	0.4564	0.4592	0.4581	**0.4606**
	10	—	0.3628	0.3966	—	0.4235	0.4243	0.4276	**0.4280**
	15	—	0.3553	0.3879	—	0.4059	0.4066	0.4107	**0.4108**
	20	—	0.3476	0.3815	—	0.3952	0.3945	**0.4003**	0.3991
Q_G	0	0.5488	0.5569	0.6578	0.6454	0.6631	0.6661	0.6654	**0.6685**
	5	—	0.5021	0.5804	—	0.6089	0.6104	0.6131	**0.6176**
	10	—	0.4469	0.5089	—	0.5496	0.5529	0.5583	**0.5601**
	15	—	0.4035	0.4485	—	0.4793	0.4839	**0.4895**	0.4888
	20	—	0.3676	0.3983	—	0.4299	0.4312	**0.4370**	0.4357
Q_Y	0	0.7407	0.7424	0.8257	0.8259	0.8484	0.8492	0.8487	**0.8554**
	5	—	0.6207	0.6639	—	0.6992	0.7029	0.7030	**0.7049**
	10	—	0.5652	0.6024	—	0.6395	0.6411	**0.6443**	0.6439
	15	—	0.5260	0.5529	—	0.5775	0.5796	0.5835	**0.5847**
	20	—	0.4926	0.5111	—	0.5477	0.5462	**0.5511**	0.5508
Q_{CB}	0	0.5171	0.5226	0.5545	0.5402	0.5735	0.5743	0.5742	**0.5753**
	5	—	0.5071	0.5181	—	0.5319	0.5330	0.5358	**0.5369**
	10	—	0.4956	0.4980	—	0.5017	0.5032	**0.5082**	0.5073
	15	—	0.4678	0.4720	—	0.4796	0.4805	**0.4875**	0.4867
	20	—	0.4494	0.4521	—	0.4614	0.4608	**0.4703**	0.4682

Table 3 Computational costs of different methods

Methods	SR-128	SR-256	ASR-128	ASR-256
computational costs	1	1.59	1.15	1.82

situation, the ASR-based method also has an advantage over the SR-based method in terms of efficiency.

6 Conclusions

In this paper, we first present a novel ASR model by learning several compact sub-dictionaries to represent the sparse domain. Then, a simultaneous image fusion and denoising method based on the ASR model is proposed in detail. Particularly, the gradient information of a local patch is employed for training patch classification in the dictionary learning stage as well as adaptive sub-dictionary selection in the fusion process. Experimental results on multi-focus and multi-modal image sets demonstrate that the ASR-based fusion method can outperform the SR-based method in terms of both visual quality and quantitative assessment, especially in the practical situation that the source images are corrupted by noise. Furthermore, with more compact dictionaries applied in the ASR model, the algorithm can be more efficient than the SR-based method using a single redundant dictionary.

To fairly verify the effectiveness of the ASR model used in image fusion, we believe it is reasonable to use the classic SR model for comparison, although some fusion methods based on improved SR model have been introduced, such as the simultaneous OMP (SOMP) in [21] and joint SR (JSR) in [22]. Actually, it is worthwhile to note that these improved SR models can be naturally combined into the proposed ASR-based fusion framework because the ASR model proposed in this paper and the SOMP or JSR model are in different hierarchies (the two 'model' have different meanings). For example, with an adaptively selected sub-dictionary in the ASR-based fusion framework, the SOMP or JSR can be used just as with a redundant dictionary in the SR-based fusion framework. Also because of this, there is still large room for improving the ASR-based fusion method in the future.

7 Acknowledgments

The authors would like to thank the editor and anonymous reviewers for their valuable comments and constructive suggestions. The authors also thank Professor Zheng Liu for providing the code for objective assessment. This work was supported by the National Science and Technology Projects (no. 2012GB102007).

8 References

1 Goshtasby, A.A., Nikolov, S.: 'Image fusion: advances in the state of the art', *Inf. Fusion*, 2007, **8**, (2), pp. 114–118
2 Stathaki, T.: 'Image fusion: algorithms and applications' (Academic Press, London, UK, 2008)
3 Li, S., Kwok, J., Wang, Y.: 'Combination of images with diverse focuses using the spatial frequency', *Inf. Fusion*, 2001, **2**, (3), pp. 169–176
4 Aslantas, V., Kurban, R.: 'Fusion of multi-focus images using differential evolution algorithm', *Expert Syst. Appl.*, 2010, **37**, (12), pp. 8861–8870
5 Li, S., Yang, B.: 'Multifocus image fusion using region segmentation and spatial frequency', *Image vis. Comput.*, 2008, **26**, (7), pp. 971–979
6 Huang, W., Jing, Z.: 'Evaluation of focus measures in multi-focus image fusion', *Pattern Recognit. Lett.*, 2007, **28**, (4), pp. 493–500
7 Piella, G.: 'A general framework for multiresolution image fusion: from pixels to regions', *Inf. Fusion*, 2003, **4**, (4), pp. 259–280
8 Toet, A.: 'A morphological pyramidal image decomposition', *Pattern Recognit. Lett.*, 1989, **9**, (4), pp. 255–261
9 Petrović, V.S., Xydeas, C.S.: 'Gradient-based multiresolution image fusion', *IEEE Trans. Image Process.*, 2004, **13**, (2), pp. 228–237
10 Li, H., Manjunath, B.S., Mitra, S.K.: 'Multisensor image fusion using the wavelet transform', *Graph. Models Image Process.*, 1995, **57**, (3), pp. 235–245
11 Lewis, J.J., O'Callaghan, R.J., Nikolov, S.G., *et al.*: 'Pixel- and region-based image fusion with complex wavelets', *Inf. Fusion*, 2007, **8**, (2), pp. 119–130
12 Nencini, F., Garzelli, A., Baronti, S., *et al.*: 'Remote sensing image fusion using the curvelet transform', *Inf. Fusion*, 2007, **8**, (2), pp. 143–156
13 Zhang, Q., Guo, B.: 'Multifocus image fusion using the nonsubsampled contourlet transform', *Signal Process.*, 2009, **89**, (7), pp. 1334–1346
14 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
15 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
16 Zhao, H., Shang, Z., Tang, Y., *et al.*: 'Multi-focus image fusion based on the neighbor distance', *Pattern Recognit.*, 2013, **46**, (3), pp. 1002–1011
17 Elad, M., Aharon, M.: 'Image denoising via sparse and redundant representations over learned dictionaries', *IEEE Trans. Image Process.*, 2006, **15**, (2), pp. 3736–3745
18 Yang, J., Wright, J., Huang, T., *et al.*: 'Image super-resolution via sparse representation', *IEEE Trans. Image Process.*, 2010, **19**, (11), pp. 2861–2873
19 Dong, W., Zhang, L., Shi, G., *et al.*: 'Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization', *IEEE Trans. Image Process.*, 2011, **20**, (7), pp. 1838–1857
20 Yang, B., Li, S.: 'Multifocus image fusion and restoration with sparse representation', *IEEE Trans. Instrum. Meas.*, 2010, **59**, (4), pp. 884–892
21 Yang, B., Li, S.: 'Pixel-level image fusion with simultaneous orthogonal matching pursuit', *Inf. Fusion*, 2012, **13**, (1), pp. 10–19
22 Yu, N., Qiu, T., Bi, F., *et al.*: 'Image features extraction and fusion based on joint sparse representation', *IEEE J. Sel. Top. Signal Process.*, 2011, **5**, (5), pp. 1074–1082
23 Yin, H., Li, S., Fang, L.: 'Simultaneous image fusion and super-resolution using sparse representation', *Inf. Fusion*, 2013, **14**, (3), pp. 229–240
24 Iqbal, M., Chen, J.: 'Unification of image fusion and super-resolution using jointly trained dictionaries and local information contents', *IET Image Process.*, 2012, **6**, (9), pp. 1299–1310
25 Liu, Y., Wang, Z.: 'Multi-focus image fusion based on sparse representation with adaptive sparse domain selection'. Proc. Int. Conf. Image Graphics, Qingdao, China, July 2013, pp. 591–596
26 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
27 Olshausen, B., Field, J.: 'Emergence of simple-cell receptive field properties by learning a sparse code for natural images', *Nature*, 1996, **381**, (6583), pp. 607–609
28 Mallat, S., Zhang, Z.: 'Matching pursuits with time-frequency dictionaries', *IEEE Trans. Signal Process.*, 1993, **41**, (12), pp. 3397–3415
29 Aharon, M., Elad, M., Bruckstein, A.: 'K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation', *IEEE Trans. Signal Process.*, 2006, **54**, (11), pp. 4311–4322
30 Mairal, J., Elad, M., Sapiro, G.: 'Sparse representation for color image restoration', *IEEE Trans. Image Process.*, 2008, **17**, (1), pp. 53–69
31 Lowe, D.: 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vis.*, 2004, **60**, (2), pp. 91–110
32 Li, S., Yang, B., Hu, J.: 'Performance comparison of different multi-resolution transforms for image fusion', *Inf. Fusion*, 2011, **12**, (2), pp. 74–84
33 Starck, J., Candès, E., Donoho, D.: 'The curvelet transform for image denoising', *IEEE Trans. Image Process.*, 2002, **11**, (6), pp. 670–684
34 da Cunha, A.L., Zhou, J., Do, M.N.: 'The nonsubsampled contourlet transform: theory, design, and applications', *IEEE Trans. Image Process.*, 2006, **15**, (10), pp. 3089–3101

doi: 10.1049/iet-ipr.2014.0311

35 Liu, Z., Blasch, E., Xue, Z., *et al*.: 'Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (1), pp. 94–109
36 Hossny, M., Nahavandi, S., Creighton, D.: 'Comments on 'information measure for performance of image fusion', *Electron. Lett.*, 2008, **44**, (18), pp. 1066–1067
37 Qu, G., Zhang, D., Yan, P.: 'Information measure for performance of image fusion', *Electron. Lett.*, 2002, **38**, (7), pp. 313–315
38 Xydeas, C., Petrović, V.: 'Objective image fusion performance measure', *Electron. Lett.*, 2000, **36**, (4), pp. 308–309
39 Yang, C., Zhang, J., Wang, X., *et al*.: 'A novel similarity based quality metric for image fusion', *Inf. Fusion*, 2008, **9**, (2), pp. 156–160
40 Wang, Z., Bovik, A., Sheikh, H., *et al*.: 'Image quality assessment: from error visibility to structural similarity', *IEEE Trans. Image Process.*, 2004, **13**, (4), pp. 600–612
41 Chen, Y., Blum, R.: 'A new automated quality assessment algorithm for image fusion', *Image Vis. Comput.*, 2009, **27**, (10), pp. 1421–1432