Multifocus Image Fusion and Restoration With Sparse Representation

Bin Yang and Shutao Li, *Member, IEEE*

***Abstract*—To obtain an image with every object in focus, we always need to fuse images taken from the same view point with different focal settings. Multiresolution transforms, such as pyramid decomposition and wavelet, are usually used to solve this problem. In this paper, a sparse representation-based multifocus image fusion method is proposed. In the method, first, the source image is represented with sparse coefficients using an overcomplete dictionary. Second, the coefficients are combined with the choose-max fusion rule. Finally, the fused image is reconstructed from the combined sparse coefficients and the dictionary. Furthermore, the proposed fusion scheme can simultaneously resolve the image restoration and fusion problem by changing the approximate criterion in the sparse representation algorithm. The proposed method is compared with spatial gradient (SG)-, morphological wavelet transform (MWT)-, discrete wavelet transform (DWT)-, stationary wavelet transform (SWT)-, curvelet transform (CVT)-, and nonsubsampling contourlet transform (NSCT)-based methods on several pairs of multifocus images. The experimental results demonstrate that the proposed approach performs better in both subjective and objective qualities.**

***Index Terms*—Image fusion, image restoration, sparse representation.**

I. Introduction

NOWADAYS, image fusion has become an important subarea of image processing. For one object or scene, multiple images can be taken from one or multiple sensors. These images usually contain complementary information. Image fusion is the process of detecting salient features in the source images and fusing these details to a synthetic image. Through image fusion, extended or enhanced information content can be obtained in the composite image, which has many application fields, such as digital imaging, medical imaging, remote sensing, and machine vision [1]–[5].

As an example of fusion that is relevant to this paper, optica imaging cameras suffer from the problem of finite depth of field, which cannot make objects at various distances (from the sensor) all in focus. Therefore, if one object in the scene is in focus, then the other objects at different distances from the camera will be out of focus and, thus, blurred. The solution to get all the objects focused in one image is multifocus image fusion technique. In this technique, several images of a scene are captured with focus on different parts. Then, these images are fused with the hope that all the objects will be in focus in the resulting image [5]–[9].

There are various methods available to implement image fusion. Basically, these methods can be categorized into two categories. The first category is the spatial domain-based methods, which directly fuse the source images into the intensity values [10]–[13]. The other category is the transformed domain-based methods, which fuse images with certain frequency or time–frequency transforms [1], [4], [5].

Assuming that $F(\cdot)$ represents the "fusion operator," the fusion methods in the spatial domain can be summarized as

$$I_F = F(I_1, I_2, \ldots, I_K). \tag{1}$$

The simplest fusion method in spatial domain just takes the pixel-by-pixel average of the source images. However, this method often leads to undesirable side effects, such as reduced contrast [1]. If the source images are not completely registered, then a single pixel-based method, such as spatial gradient (SG)-based method [10], always results in artifacts in the fused image. Therefore, some more reasonable methods were proposed to fuse source images with divided blocks or segmented regions instead of single pixels [11]–[13]. However, the block-based fusion methods usually suffer from blockness in the fused image [11]. For the region-based method, the source images are first segmented, and the obtained regions are then fused using their properties, such as spatial frequency or SG. The segmentation algorithms, usually complicated and time consuming, are of vital importance to the fusion quality [13].

A more popular method that has been explored in recent years is by using multiscale transforms. The usually used multiscale transforms include various pyramids [14]–[17], discrete wavelet transform (DWT) [1], [5], [18], complex wavelet [19], [20], ridgelet [21], curvelet transform (CVT) [22], and contourlet [23]. The transformed domain-based methods can be summarized as

$$I_F = T^{-1}\left(F\left(T(I_1), T(I_2), \ldots, T(I_K)\right)\right) \tag{2}$$

where $T(\cdot)$ represents a multiscale transform, and $F(\cdot)$ means the applied fusion operator.

Pyramid decomposition is the earliest multiscale transform used for image fusion [14]–[17]. In this method, each source image is first decomposed into a sequence of images (pyramid) in different resolutions. Then, at each position in the transformed image, the value in the pyramid with the highest

Manuscript received January 17, 2009; revised May 22, 2009. First published October 30, 2009; current version published March 20, 2010. This work was supported in part by the National Natural Science Foundation of China under Grants 60871096 and 60835004, by the Ph.D. Programs Foundation of the Ministry of Education of China under Grant 200805320006, and by the Key Project of the Chinese Ministry of Education under Grant 2009-120. The Associate Editor coordinating the review process for this paper was Dr. Cesare Alippi.

The authors are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: yangbin01420@163.com; shutao_li@yahoo.com.cn).

Digital Object Identifier 10.1109/TIM.2009.2026612

saliency is selected. Finally, the fused image is constructed using the inverse transform of the composite images. The wavelet transform-based fusion methods employ a similar scheme to the pyramid transform-based methods. However, the performance of multiresolution transform-based methods is limited owing to that most of the multiresolution decompositions are not shift invariant, which is brought by the underlying down-sampling process [1], [5], [19]. The shift-invariant extension of the DWT can yield an overcomplete signal representation, which is suitable for image fusion [24]–[26]. Further, many advanced geometric multiscale transforms, such as CVT, ridgelet, and contourlet, have been explored in recent years and shown improved results [21]–[23], [27]–[29]. However, because the fused image obtained by transform domain-based algorithms is globally created, a little change in a single coefficient of the fused image in the transformed domain may cause all the pixel values to change in spatial domain. As a result, undesirable artifacts may be produced in the fusion process using the multiresolution transform-based methods in some cases [3].

Obviously, effectively and completely extracting the underlying information of the original images would make the fused image more accurate. Different from multiscale transformations, the sparse representation using an overcomplete dictionary that contains prototype signal atoms describes signals by sparse linear combinations of these atoms [30]–[34]. Two main characteristics of sparse representation are its overcompleteness and sparsity [32]. Overcompleteness means that the number of basis atoms in the dictionary exceeds the number of image pixels or signal dimensions. The overcomplete dictionary that contains rich transform bases allows for more stable and meaningful representation of signals. Sparsity means that the coefficients corresponding to a signal are sparse, that is to say, only "a few descriptions" can describe or capture the significant structure information about the object of interest. Benefiting from its sparsity and overcompleteness, sparse representation theory has successfully been applied in many practical applications, including compression, denoising, feature extraction, classification, and so on [32]–[34]. Recent studies have shown that common image features can also be accurately described by only a few coefficients or "a few descriptions" [32]. Using the few coefficients as the salient features of images, we design a sparse representation (SR)-based image fusion scheme. In general, sparse representation is a global operation, in the sense that it is based on the gray-level content of an entire image. However, the image fusion quality depends on the accurate representation of the local salient features of source images. Therefore, a "sliding window" technique is adopted to achieve better performance in capturing local salient features and keeping shift invariance.

In the proposed method, the source images are first divided into patches, which lead to "a small size" of overcomplete dictionary to every patch. Second, the patches are decomposed by prototype signal atoms into their corresponding coefficients. The larger the coefficient is, the more salient features it contains. Third, the "choose-max" fusion rule is used to combine the corresponding coefficients of the source images. Finally, the result image is reconstructed using the combined coefficients previously obtained. In sparse representation, the dictionary (atoms) is often created by a prespecified set of functions, such as discrete cosine transforms (DCT), short-time Fourier transforms, wavelet, CVT, and contourlet. More complex dictionary can be obtained by learning from a given set of signal examples [32].

Note that most of the image fusion methods are based on the assumption that the source images are noise free. Therefore, these fusion algorithms can produce high-quality fused images if the assumption is satisfied. However, practically, the images are often corrupted by noise during acquisition or transmission processes. For multiresolution-based methods, they usually denoise the source images, first, by setting all the coefficients below a certain threshold to zero and keeping the remaining coefficients unchanged. Then, the filtered images are fused. One advantage of our proposed method is that it can simultaneously carry out denoising and fusion of noisy source images.

The rest of this paper is organized into five sections. In Section II, the basic theory of sparse representation is presented. In Section III, we propose the fusion scheme with sparse representation theory and discuss how to simultaneously carry out image restoration and fusion. Numerical experiments and discussions are detailed in Section IV. Both advantages and disadvantages of the proposed schemes, together with some suggestions about the future work, are given in Section V.

II. Basic Theory of Signal Sparse Representation

Sparse representation is based on the assumption that natural signals can be represented or approximately represented as a linear combination of a "few" atoms from dictionary [30]. For signals $\Gamma \subset \mathbf{R}^n$, sparse representation theory suggests the existence of a dictionary $\mathbf{D} \in \mathbf{R}^{n\times T}$, which contains T prototype signals that are referred to as atoms. For any signal $\boldsymbol{x} \in \Gamma$, there exists a linear combination of atoms from $\mathbf{D}$ that approximates it well. More formally, it is $\forall \boldsymbol{x} \in \Gamma$, $\exists \boldsymbol{s} \in R^T$ such that $\boldsymbol{x} \approx \mathbf{D}\boldsymbol{s}$. The vector $\boldsymbol{s}$ contains "coefficients" of $\boldsymbol{x}$ in $\mathbf{D}$. It usually assumes that $T > n$, implying that the dictionary $\mathbf{D}$ is redundant. The solution is generally not unique. Finding the smallest possible number of nonzero components of $\boldsymbol{s}$ involves solving the following optimization problem:

$$\min_{s} \|\boldsymbol{s}\|_0 \text{ subject to } \|\mathbf{D}\boldsymbol{s} - \boldsymbol{x}\| < \varepsilon \tag{3}$$

where $\|\boldsymbol{s}\|_0$ denotes the number of nonzero components in $\boldsymbol{s}$.

The above optimization is an NP-hard problem and can only be solved by systematically testing all the potential combinations of columns [31]. Thus, approximate solutions are considered instead. In the past decade, several pursuit algorithms have been proposed to solve this problem. The simplest algorithms are the matching pursuit (MP) [30] and the orthogonal MP (OMP) algorithms [32]. They are greedy algorithms that sequentially select the dictionary atoms. These methods involve the computation of the inner product between the signal and dictionary columns.

The MP algorithm aims to learn $\boldsymbol{x} \approx \mathbf{D}\boldsymbol{s}$ as follows: Let $\langle \cdot, \cdot \rangle$ denote the inner product. Initialize the residual function $\boldsymbol{r}^0$ as

$$\boldsymbol{r}^0 = \boldsymbol{x}. \tag{4}$$

Then, loop over all prototype signal atoms and select the index t_l^* of the best atom function in $\mathbf{D}$ as

$$t_l^* = \arg\max_t \left\langle \boldsymbol{r}^l, \boldsymbol{d}_t \right\rangle \tag{5}$$

with a coefficient along that dimension as

$$\boldsymbol{s}\left(t_l^*\right) = \left\langle \boldsymbol{r}^l, \boldsymbol{d}_{t_l^*} \right\rangle. \tag{6}$$

Then, update the residual function

$$\boldsymbol{r}^{l+1} = \boldsymbol{r}^l - \boldsymbol{s}\left(t_l^*\right)\boldsymbol{d}_{t_l^*} \tag{7}$$

and repeat until

$$\|\boldsymbol{r}^l\|^2 = \left\| \boldsymbol{x} - \sum_{i=1}^{l} \boldsymbol{s}\left(t_i^*\right)\boldsymbol{d}_{t_i^*} \right\|^2 < \varepsilon \tag{8}$$

or the iteration reaches the priori set number.

Because the computation of the MP algorithm is intensive, in this paper, the OMP is used. The OMP method updates the residual as

$$\boldsymbol{r}^{l+1} = \boldsymbol{x} - \boldsymbol{P}_{\mathrm{span}}\left\{\boldsymbol{d}_{t_1^*}, \boldsymbol{d}_{t_2^*}, \ldots, \boldsymbol{d}_{t_l^*}\right\}\boldsymbol{x} \tag{9}$$

where P denotes the orthogonal projection onto a subspace. The OMP algorithm, in parallel, applies the Gram–Schmidt orthogonalization upon the chosen atoms for the efficient computation of projections. Ideally, the number of iterations is equal to the number of nonzeros in $\boldsymbol{s}$. Other well-known pursuit approaches include the basis pursuit (BP) [35] and the focal underdetermined system solver (FOCUSS) [33]. The BP replaces the ℓ^0-norm with an ℓ^1-norm in (3), whereas the FOCUSS uses the ℓ^p-norm with as a replacement for the ℓ^0-norm.

III. Image Fusion Method

A. Sparse Representation for Image Fusion

Since the sparse representation globally handles an image, it cannot directly be used with image fusion, which depends on the local information of source images. In our method, we divide the source images into small patches and use the fixed dictionary $\mathbf{D}$ with small size to solve this problem. In addition, a sliding window technique is adopted to make the sparse representation shift invariant, which is of great importance to image fusion.

We assume that source image I is divided into many image patches. As shown in Fig. 1, to facilitate the analysis, the jth patch with size $n \times n$ is lexicographically ordered as a vector $\boldsymbol{v}^j$. Then, $\boldsymbol{v}^j$ can be expressed as

$$\boldsymbol{v}^j = \sum_{t=1}^{T} \boldsymbol{s}^j(t)\boldsymbol{d}_t \tag{10}$$

where $\boldsymbol{d}_t$ is an atom from a given overcomplete dictionary, and $\mathbf{D} = [\boldsymbol{d}_1 \ \cdots \ \boldsymbol{d}_t \ \cdots \ \boldsymbol{d}_T]$, which contains T atoms. $\boldsymbol{s}^j = [s^j(1) \ \cdots \ s^j(t) \ \cdots \ s^j(T)]^T$ is the sparse representation obtained in (3).

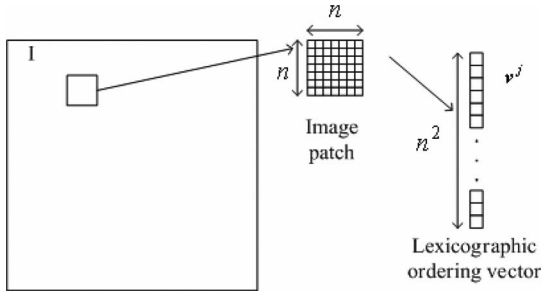

Fig. 1. Selected image patch and its lexicographic ordering vector.

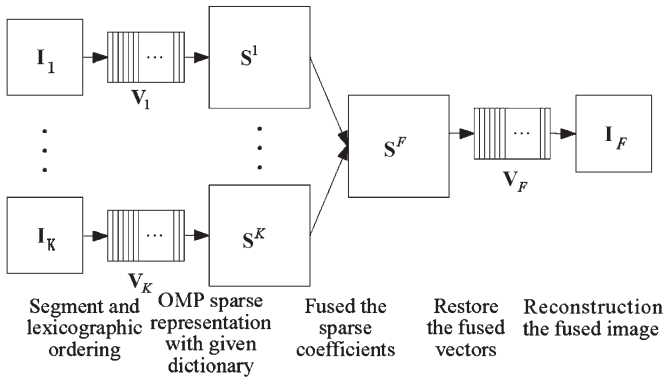

Fig. 2. Schematic diagram of the proposed SR-based fusion method.

Assume that the vectors responding to all the patches in image I are constituted into one matrix $\mathbf{V}$. Then, $\mathbf{V}$ can be expressed as

$$V = [\boldsymbol{d}_1 \ \boldsymbol{d}_2 \ \cdots \boldsymbol{d}_T] \begin{bmatrix} s^1(1) & s^2(1) & \cdots & s^J(1) \\ s^1(2) & s^2(2) & \cdots & s^J(2) \\ \vdots & \vdots & \ddots & \vdots \\ s^1(T) & s^2(T) & \cdots & s^J(T) \end{bmatrix} \tag{11}$$

where J is the number of image patches. Let $\mathbf{S} = [\boldsymbol{s}^1, \boldsymbol{s}^2, \cdots, \boldsymbol{s}^J]$. Then, (11) can be expressed as

$$\mathbf{V} = \mathbf{DS} \tag{12}$$

where $\mathbf{S}$ is a sparse matrix.

B. Proposed Fusion Scheme

Assume that there are K registered source images $I_1, \ldots, I_K$ with size of $M \times N$. Then, the proposed fusion scheme based on image sparse representation, shown in Fig. 2, takes the following steps.

Use the sliding window technique to divide each source image I_k, from left-top to right-bottom, into patches of size $n \times n$, i.e., the size of the atom in the dictionary. Then, all the patches are transformed into vectors via lexicographic ordering, and all the vectors constitute one matrix $\mathbf{V}_k$, in which each column corresponds to one patch in the source image I_k. The size of $\mathbf{V}_k$ is $(n \cdot n) \times ((M - n + 1) \cdot (N - n + 1))$.

For the jth column vector $\boldsymbol{v}_{k_j}$ in $\mathbf{V}_k$, its sparse representation is calculated using the OMP method. The OMP iterations will stop when the representation error drops below the specified tolerance. Then, we get a very sparse representation vector $\boldsymbol{s}_{k_j}$ for $\boldsymbol{v}_{k_j}$.

Then, the activity level of $\boldsymbol{s}_{k_j}$ responding to the kth source image I_k is obtained as

$$A_{k_j} = \|\boldsymbol{s}_{k_j}\|_1. \tag{13}$$

Fuse the corresponding columns of all sparse representation matrix $\mathbf{S}_1, \ldots, \mathbf{S}_k, \ldots, \mathbf{S}_K$ of the source images to generate $\mathbf{S}_F$ according to their activity levels. The jth column of $\mathbf{S}_F$ is obtained as

$$\boldsymbol{s}_{F_j} = \boldsymbol{s}_{k_j^*}, \qquad k_j^* = \arg\max_{k_j}(A_{k_j}). \tag{14}$$

The vector representation of the fused image $\mathbf{V}_F$ can be calculated by

$$\mathbf{V}_F = \mathbf{D}\mathbf{S}_F. \tag{15}$$

Finally, the fused image I_F is reconstructed using $\mathbf{V}_F$. Reshape each vector $\mathbf{v}_{F_j}$ in $\mathbf{V}_F$ into a block with size $n \times n$ and then add the block to I_F at its responding position. This can be seen as the inverse process of Fig. 1. Thus, for each pixel position, the pixel value is the sum of several block values. Then, the pixel value is divided by the adding times at its position to obtain the reconstructed result.

C. Restoration and Fusion

In practice, the source images for fusion are often corrupted by noise during the acquisition or transmission process. In traditional methods, image restoration and image fusion are separately treated. Little effort has been made to combine them [36]. The other benefit of the proposed scheme is to simultaneously conduct image restoration and fusion by using the advantage of sparse representation in image restoration.

Assume that image I is contaminated by additive zero-mean white Gaussian noise with standard deviation σ. For the jth patch's vector of image I $\boldsymbol{v}^j$, its denoising solution using the maximum *a posteriori* estimator is

$$\hat{\boldsymbol{s}}^j = \min_{\boldsymbol{s}^j} \left\|\boldsymbol{s}^j\right\|_0 \text{ subject to } \left\|\boldsymbol{v}^j - \mathbf{D}\boldsymbol{s}^j\right\|_2 < C\sigma \tag{16}$$

where C is a constant [34]. Then, all the patches' sparse representation vectors constitute the sparse coefficient matrix $\hat{\mathbf{S}}$.

Then, for source images $I_1, \ldots, I_K$, the corresponding overlapped patches and vectoring matrix are $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_K$. Their sparse representations $\hat{\mathbf{S}}_1, \hat{\mathbf{S}}_2, \ldots, \hat{\mathbf{S}}_K$ are obtained according to (16). The restored image matrices are calculated by $\hat{\mathbf{V}}_1 = \mathbf{D}\hat{\mathbf{S}}_1$, $\hat{\mathbf{V}}_2 = \mathbf{D}\hat{\mathbf{S}}_2, \ldots, \hat{\mathbf{V}}_K = \mathbf{D}\hat{\mathbf{S}}_K$. The restored and fused image vector representation $\hat{\mathbf{V}}_F$ can be reconstructed from $\hat{\mathbf{V}}_1$, $\hat{\mathbf{V}}_2, \ldots, \hat{\mathbf{V}}_K$. Letting $F(\cdot)$ be the fusion operator, then

$$\hat{\mathbf{S}}_F = F(\hat{\mathbf{S}}_1, \hat{\mathbf{S}}_2, \ldots, \hat{\mathbf{S}}_K). \tag{17}$$

The fused image vector representation can be obtained by

$$\hat{\mathbf{V}}_F = \mathbf{D}\hat{\mathbf{S}}_F. \tag{18}$$

Finally, the fused image from the restored source images is reconstructed from $\hat{\mathbf{V}}_F$.

IV. Experiments

The proposed method is compared with existing special multifocus image fusion methods (e.g., the SG [10] and the morphological wavelet transform (MWT) [8]) and general multifocus image fusion methods (the DWT [1], [5], [6], the stationary wavelet transform (SWT) [7], [24], the CVT [29], and the nonsubsampling contourlet transform (NSCT) [27], [28] on several pairs of multifocus images).

A. Experiment Setups

There are three key issues in multiresolution image fusion methods: 1) activity level measurement; 2) coefficient combination; and 3) consistency verification. In our experiment, the most popular settings are selected. 1) Activity level measurement: using the absolute value of the coefficients at each coefficient location. 2) Coefficient combination: selecting the coefficient with the largest activity level at each pixel location (choose-max). 3) Consistency verification: using a small majority filter (window-based verification), and the widow size is 3×3 [5]. For all four multiresolution methods, the source images are decomposed with four levels. The basis for DWT and SWT is "db1." For NSCT, 2, 8, 8, and 16 directions are used in the high-frequency scales. For the multiresolution-based denoising, we use K-sigma hard thresholding solution [27]. The threshold is

$$TH = K\sigma \tag{19}$$

for each scale of the coefficients, where K is a constant, and σ is the noise standard deviation.

For the proposed method, an overcomplete separable version of the DCT dictionary obtained by sampling the cosine wave in different frequencies is adopted because of fast implementation and effective performance. The moving step of the sliding window technique is one pixel for the following experiments. All the experiments are implemented in Matlab 6.5 and on a Pentium(R) 1.7-GHz PC with 512-MB RAM. In addition to the subjective evaluation, two objective evaluation criteria, i.e., 1) root mean square error (RMSE) [1] and 2) $\mathrm{Q}^{\mathrm{AB/F}}$ metric [37], are used to compare the different fusion methods. The smaller the value of the RMSE, the better the fusion result. The $\mathrm{Q}^{\mathrm{AB/F}}$ measure should be as close to 1 as possible.

B. Fusion Results of Clear Images

In the first experiment, four pairs of multifocus source images without noise, as shown in Fig. 3, are used to test the proposed method. Fig. 3(a) is near focused, where the clock is in "focus" and clear in vision, whereas the student is out of "focus" and blurred. Fig. 3(b) is far focused, and the situations for the clock and the student are contrary. The other images in Fig. 3 have a similar situation, with the left column images focusing on the near objects and the right column images on the far objects.

For the proposed method, two parameters, i.e., 1) the block size and 2) the global error ε in (3), should be investigated. First, we fix the global error ε to 0.1 and investigate the influence of block size. Then, for exploring the effect of the global error

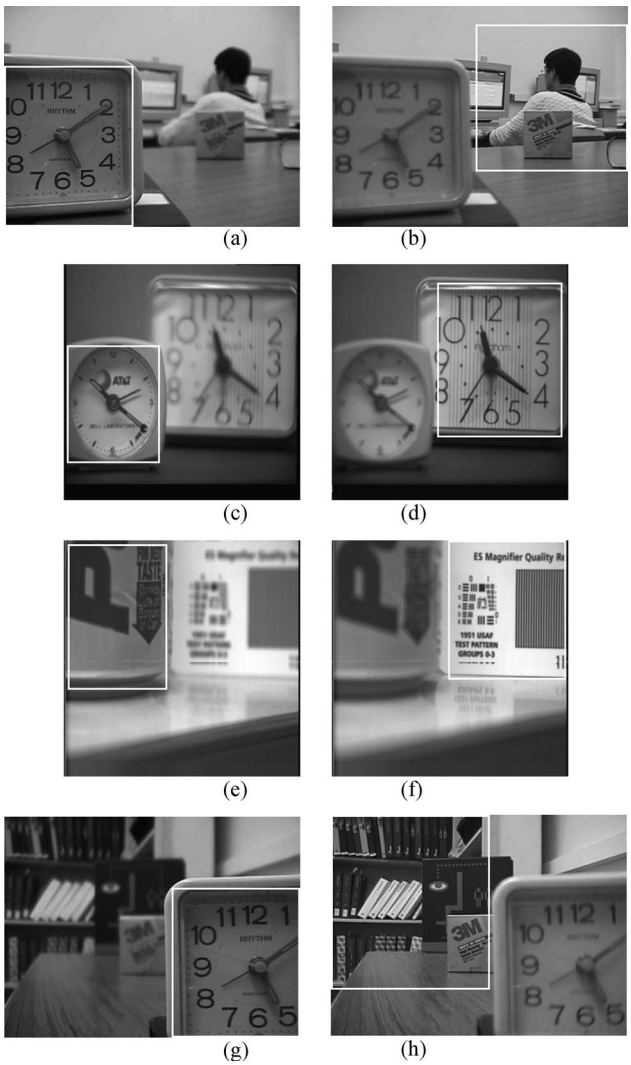

Fig. 3. Multifocus source images.

ε, the block size is fixed to 8×8. For RMSE, the ground truth regions are labeled as white rectangles in Fig. 3. The effects of different block size and global error on RMSE and $Q^{AB/F}$ are shown in Fig. 4(a)–(d). From Fig. 4(a)–(d), we can obviously conclude that the performance of the proposed method improves as the block size increases and the global error decreases. However, in this situation, the elapsed time is also raised, as shown in Fig. 4(e) and (f). We believe that it will not be a problem because of the hardware's high-speed development. In addition, we can see from Fig. 4(b) and (d) that when the global error is smaller than 2, the fused results are all with good quality.

Two fused results of Fig. 3(a) and (b) with global error, i.e., 5 and 10, are shown in Fig. 5. When the global error is 5, the fused image shown in Fig. 5(a) loses some details, such as the second hand of the clock and the details of the table. However, most of the important details of the two source images are reserved. Even if the global error is 10, the result can still be acceptable, and the proposed method does not result in artificial distortion in the fused image. From the analysis, we can conclude that it is reasonable to set the global error ε in (3) to 0.1 and the block size to 8×8.

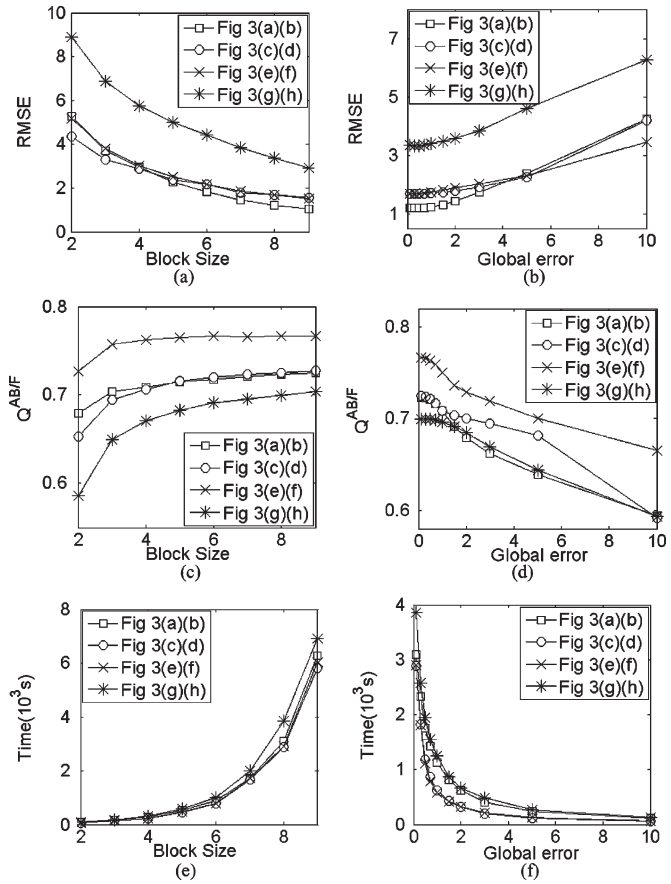

Fig. 4. Fusion performance of the proposed method with different block size and global error.

Fig. 5. Fused results of the proposed method with different global error. (a) and (b) Fused images with global error of 5 and 10 respectively.

We also investigated the effect of the moving step of the sliding window technique on fusion performance. When no overlapping, i.e., "distinct window" technique, is adopted, some fused results have obvious blocking artifacts, which is because the "distinct window" technique is intrinsically shift variant. The moving step from 1 to 8 corresponds to the changing from "sliding window" to "distinct window." Additional experiments also demonstrated that the proposed method performs best with the smallest moving step.

Now, the proposed method with the above settings is compared with different fusion methods based on SG, MWT, DWT, SWT, CVT, and NSCT. Only the fused results of the proposed method are given, as shown in Fig. 6, because of the lack of space. In fact, the student's head in Fig. 3(a) and (b) has slight motion. For easy observation, we magnify the head regions of

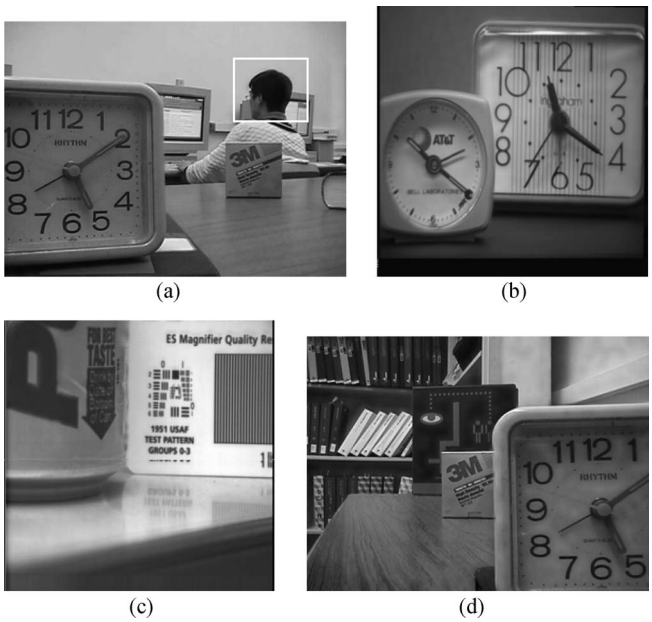

Fig. 6. Fused results of the different fusion methods. (a)–(g) Fused images from SG-, MWT-, DWT-, SWT-, CVT-, NSCT-, and SR-based methods.

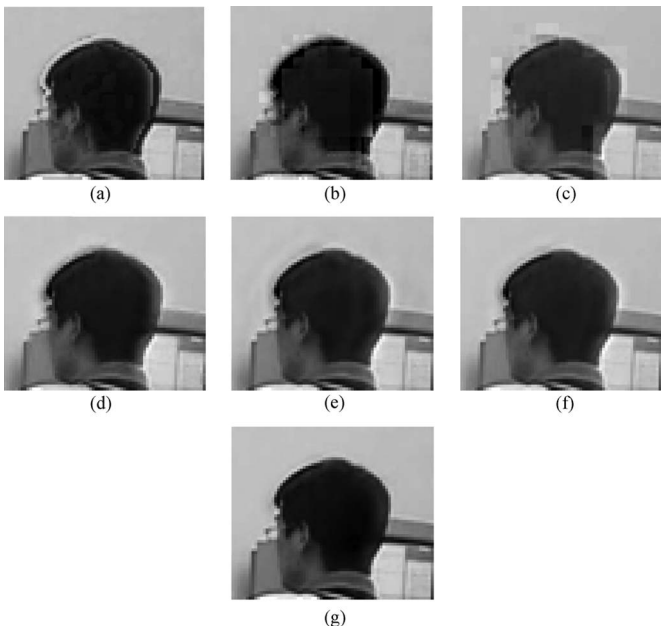

Fig. 7. Magnified head regions of the fused results with different methods. (a)–(g) Results from SG-, MWT-, DWT-, SWT-, CVT-, and NSCT-based methods.

the fused results with different methods in Fig. 7. There are serious reconstruction artifacts in the SG-, MWT-, and DWT-fused images and circle blurring effect in the CVT-fused image because of their down-sampling process. Although the fused results obtained by SWT and NSCT have better appearances than DWT and CVT, the contrast of the fused images is reduced to some extent. The proposed method provides the result with the best visual appearance.

The RMSE and the $Q^{AB/F}$ between source and fused images of the seven methods are listed in Table I, where the values in bold indicate the highest quality measure obtained over all

TABLE I
OBJECTIVE PERFORMANCE OF DIFFERENT FUSION METHODS

Measures	Methods	Fig. 3 (a)(b)	Fig. 3 (c)(d)	Fig. 3 (e)(f)	Fig. 3 (g)(h)
RMSE	SG	3.163	2.905	3.455	5.822
	MWT	6.868	6.020	7.717	10.349
	DWT	4.109	4.504	3.432	8.576
	SWT	4.154	3.820	3.082	8.159
	CVT	3.865	3.870	2.235	7.694
	NSCT	3.592	3.730	2.142	7.742
	SR	**1.215**	**1.679**	**1.695**	**3.365**
$Q^{AB/F}$	SG	0.710	0.703	0.754	0.661
	MWT	0.642	0.650	0.698	0.630
	DWT	0.685	0.685	0.737	0.668
	SWT	0.705	0.715	0.761	0.699
	CVT	0.692	0.696	0.745	0.684
	NSCT	0.712	0.716	0.759	**0.702**
	SR	**0.723**	**0.725**	**0.766**	0.700

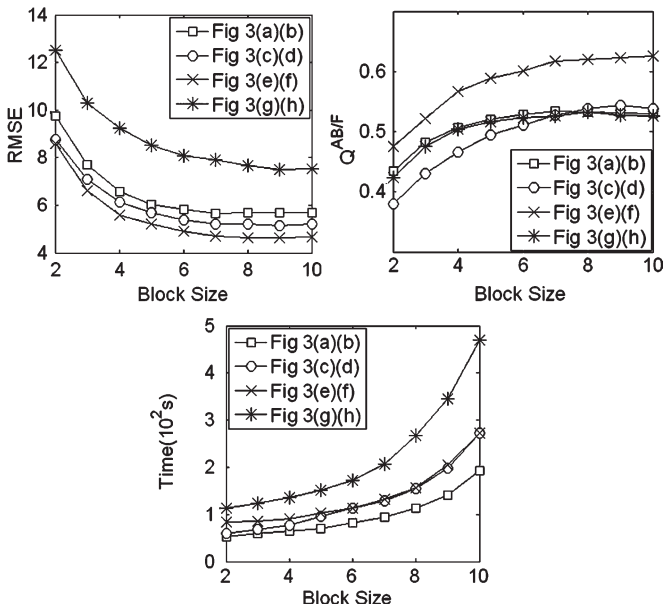

Fig. 8. Performance of the proposed method on noised images with difference block size.

fusion methods. From Table I, we can see that the results of the proposed method are obviously better than those of the other six methods on RMSE criterion. As to the $Q^{AB/F}$ criterion, our method only loses to the NSCT method in Fig. 3(g) and (h).

C. Fusion Results of Noised Images

The proposed restoration and fusion method is compared with the CVT- or NSCT-based approaches, which have been proved very effective in image denoising application. For the CVT- and NSCT-based method, the K in (19) is set to 4 for the finest scale and set to 3 for the remaining scales according to successful cases [27]. The multifocus source images in Fig. 3 are corrupted by adding zero-mean additive Gaussian white noise with $\sigma = 15$. The constant C in (16) is set to 1.15 [34]. We first test the effect of block size on the proposed method. From

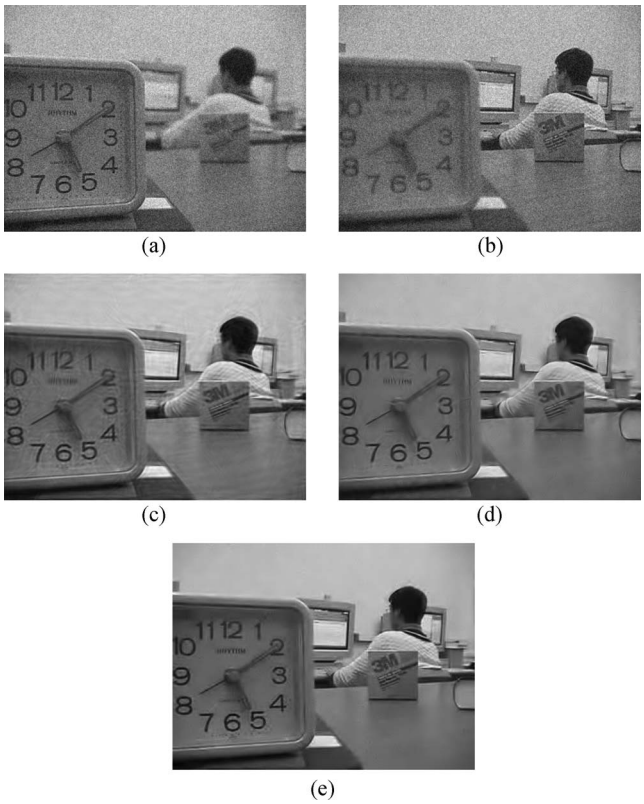

Fig. 9. Restored and fused results. (a) and (b) Two noised images. (c)–(e) Fused and restored images using CVT, NSCT, and SR.

TABLE II
COMPARISON OF RESTORED AND FUSED RESULTS FROM DIFFERENT METHODS

Measures	Methods	Fig. 3 (a)(b)	Fig. 3 (c)(d)	Fig. 3 (e)(f)	Fig. 3 (g)(h)
RMSE	CVT	7.902	10.850	6.940	6.435
	NSCT	6.756	10.082	6.334	5.493
	SR	**5.657**	**7.572**	**5.094**	**4.649**
$Q^{AB/F}$	CVT	0.489	0.485	0.478	0.577
	NSCT	0.527	**0.528**	0.515	0.605
	SR	**0.538**	0.522	**0.535**	**0.621**

the results shown in Fig. 8, we can see that it is appropriate to set a block size of 8×8 for the proposed method.

The noised versions in Fig. 3(a) and (b), the restored and fused results of CVT, NSCT, and our proposed method are shown in Fig. 9. The fused image of the CVT-based method reserves the details, such as texture on the table, which has been removed in Fig. 9(d) and (e). However, it produces more artificial edges. Fig. 9(d) is visually better than Fig. 9(c), but it also has more artificial distortions than Fig. 9(e). Therefore, the restored and fused image by the proposed method has much better visibility than those obtained by the other two methods. The RMSE and the $Q^{AB/F}$ between the clear images in Fig. 3 and the restored and fused images are listed in Table II. The values demonstrate that the proposed method is effective and is superior to the CVT- and NSCT-based methods.

Then, we compare the above three methods with different noise standard deviations. Gaussian white noises with standard deviation of 5, 10, 15, 20, and 25 are added to Fig. 3(a) and (b),

TABLE III
PERFORMANCE OF RESTORATION AND FUSION WITH DIFFERENT NOISE STANDARD DEVIATIONS

Measures	Methods	Noise standard deviations				
		5	10	15	20	25
RMSE	CVT	5.171	6.646	7.963	8.969	9.835
	NSCT	4.484	5.665	6.776	7.824	8.667
	SR	**2.809**	**4.364**	**5.680**	**6.815**	**8.067**
$Q^{AB/F}$	CVT	0.587	0.528	0.493	0.459	0.420
	NSCT	0.615	0.565	0.528	0.491	0.451
	SR	**0.616**	**0.571**	**0.534**	**0.501**	**0.463**

Fig. 10. Twenty ground truth images.

respectively. The corresponding objective measures are listed in Table III. From Table III, we can see that, with the increasing noise standard deviation, the fused and restored results become worse for all of them. However, the proposed method provides obviously better result than the CVT- and NSCT-based method in the slight noise situation.

D. More Results on Artificial Source Images

To confirm the effectiveness of the proposed method, 20 standard images shown in Fig. 10 are used as ground truth. From each image, two out-of-focus images are created by Gaussian blurring. Then, the blurred images with different focus points are taken as the source images. The original image is taken as the reference image. The SG-, MWT-, DWT-, SWT-, CVT-, and NSCT-based methods are used for comparison. In this experiment, the global error ε in (3) is fixed to 0.1, and the block size is fixed to 8×8. In addition, the blurred images from Fig. 10 are the added Gaussian noise with standard deviation of 15. Then, the restored and fused methods with CVT, NSCT, and proposed method are performed. The RMSE and the $Q^{AB/F}$ of the fused images using different methods on the clear images and noised versions are listed in Table IV. The reported values are the average and standard deviation of the 20 source images. For the clear source images, we can observe from Table IV that the SR-based method outperforms other methods with the $Q^{AB/F}$ and RMSE criterion. Moreover, the NSCT- and SWT-based methods follow the proposed method. The SG-based

TABLE IV
AVERAGE OBJECTIVE PERFORMANCE OF DIFFERENT FUSION METHODS ON ARTIFICIAL SOURCE IMAGES

Measures	Methods	Clear case	Noise case
RMSE	SG	3.070 ±1.439	N/A
	MWT	5.431 ±1.656	N/A
	DWT	3.449 ±0.816	N/A
	SWT	2.566 ±0.528	N/A
	CVT	3.820 ±0.853	11.112 ±1.510
	NSCT	2.182 ±0.462	9.455 ±1.172
	SR	**1.715** ±0.470	**8.857** ±0.959
$Q^{AB/F}$	SG	0.797 ±0.026	N/A
	MWT	0.741 ±0.034	N/A
	DWT	0.781 ±0.027	N/A
	SWT	0.785 ±0.025	N/A
	CVT	0.763 ±0.025	0.419 ±0.066
	NSCT	0.797 ±0.022	0.465 ±0.071
	SR	**0.801** ±0.023	**0.494** ±0.065

method also provides good fused results. The DWT-, CVT-, and MWT-based methods perform worst. For the noised image restoration and fusion, we can see from Table IV that the proposed method is obviously superior to the CVT- and NSCT-based methods.

V. CONCLUSION

In this paper, we have presented a multifocus image fusion method based on sparse representation theory. The contributions of this paper can be concluded into two aspects. First, we explored the image sparse representation theory to resolve the image fusion problem, which can simultaneously conduct image fusion and restoration. Second, the "sliding window" technique is introduced to overcome the shift variance problem, which may increases the complexity but will not be a problem owing to the hardware's development. The fused results are compared with those obtained by the SG-, MWT-, DWT-, SWT-, CVT-, and NSCT-based methods. The experiment results demonstrate the feasibility and effectiveness of the proposed method. In this paper, only the DCT-based dictionary is studied. Note that there are still many other methods, such as learning-based methods, to construct the overcomplete dictionary. We will investigate the influence of different dictionaries on the fused results in the future.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their detailed review, valuable comments, and constructive suggestions.

REFERENCES

[1] Z. Zhang and R. S. Blum, "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," *Proc. IEEE*, vol. 87, no. 8, pp. 1315–1326, Aug. 1999.

[2] Z. Liu, D. S. Forsyth, M.-S. Safizadeh, and A. Fahr, "A data-fusion scheme for quantitative image analysis by using locally weighted regression and Dempster–Shafer theory," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 11, pp. 2554–2560, Nov. 2008.

[3] A. A. Goshtasby and S. Nikolov, "Image fusion: Advances in the state of the art," *Inf. Fusion*, vol. 8, no. 2, pp. 114–118, Apr. 2007.

[4] V. S. Petrovic and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 228–237, Feb. 2004.

[5] G. Pajares and J. Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.*, vol. 37, no. 9, pp. 1855–1872, Sep. 2004.

[6] G. Piella, "A general framework for multiresolution image fusion: From pixels to regions," *Inf. Fusion*, vol. 4, no. 4, pp. 259–280, Dec. 2003.

[7] S. T. Li, J. T. Kwok, I. W. Tsang, and Y. N. Wang, "Fusing images with different focuses using support vector machines," *IEEE Trans. Neural Netw.*, vol. 15, no. 6, pp. 1555–1561, Nov. 2004.

[8] I. De and B. Chanda, "A simple and efficient algorithm for multifocus image fusion using morphological wavelets," *Signal Process.*, vol. 86, no. 5, pp. 924–936, May 2006.

[9] A. Kubota, K. Aizawa, and T. Chen, "Reconstructing dense light field from array of multifocus images for novel view synthesis," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 269–279, Jan. 2008.

[10] H. A. Eltoukhy and S. Kavusi, "A computationally efficient algorithm for multi-focus image reconstruction," in *Proc. SPIE Electron. Imaging*, 2003, vol. 5017, pp. 332–341.

[11] S. T. Li, J. T. Kwok, and Y. N. Wang, "Combination of images with diverse focuses using the spatial frequency," *Inf. Fusion*, vol. 2, no. 3, pp. 169–176, Sep. 2001.

[12] P. L. Lin and P. Y. Huang, "Fusion methods based on dynamic-segmented morphological wavelet or cut and paste for multifocus images," *Signal Process.*, vol. 88, no. 6, pp. 1511–1527, Jun. 2008.

[13] S. T. Li and B. Yang, "Multifocus image fusion using region segmentation and spatial frequency," *Image Vis. Comput.*, vol. 26, no. 7, pp. 971–979, Jul. 2008.

[14] P. T. Burt and E. H. Andelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532–540, Apr. 1983.

[15] Z. Liu, K. Tsukada, K. Hanasaki, Y. K. Ho, and Y. P. Dai, "Image fusion by using steerable pyramid," *Pattern Recognit. Lett.*, vol. 22, no. 9, pp. 929–939, Jul. 2001.

[16] A. Toet, "Image fusion by a ratio of low-pass pyramid," *Pattern Recognit. Lett.*, vol. 9, no. 4, pp. 245–253, 1989.

[17] A. Toet, "A morphological pyramidal image decomposition," *Pattern Recognit. Lett.*, vol. 9, no. 3, pp. 255–261, May 1989.

[18] H. Li, B. Manjunath, and S. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57, no. 3, pp. 235–245, May 1995.

[19] P. Hill, N. Canagarajah, and D. Bull, "Image fusion using complex wavelets," in *Proc. 13th Brit. Mach. Vision Conf.*, Cardiff, U.K., 2002, pp. 487–496.

[20] B. Forster, D. van de Ville, J. Berent, D. Sage, and M. Unser, "Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images," *Microsc. Res. Tech.*, vol. 65, no. 1/2, pp. 33–42, 2004.

[21] T. Chen, J. P. Zhang, and Y. Zhang, "Remote sensing image fusion based on ridgelet transform," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. Symp.*, Seoul, Korea, 2005, vol. 2, pp. 1150–1153.

[22] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inf. Fusion*, vol. 8, no. 2, pp. 143–156, Apr. 2007.

[23] H. H. Song, S. Y. Yu, L. Song, and X. K. Yang, "Fusion of multispectral and panchromatic satellite images based on contourlet transform and local average gradient," *Opt. Eng.*, vol. 46, no. 2, pp. 020502-1–020502-3, Feb. 2007.

[24] O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform," in *Proc. Int. Conf. Image Process.*, Washington, DC, 1997, vol. 3, pp. 288–291.

[25] S. T. Li, J. T. Kwok, and Y. N. Wang, "Discrete wavelet frame transform method to merge Landsat TM and SPOT panchromatic images," *Inf. Fusion*, vol. 3, no. l, pp. 17–23, Mar. 2002.

[26] M. Beaulieu, S. Foucher, and L. Gagnon, "Multi-spectral image resolution refinement using stationary wavelet transform," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Vancouver, BC, Canada, 1989, vol. 6, pp. 4032–4034.

[27] L. D. Cunha and J. P. Zhou, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.

[28] B. Yang, S. T. Li, and F. M. Sun, "Image fusion using nonsubsampled contourlet transform," in *Proc. IEEE 4th Int. Conf. Image Graph.*, Chengdu, China, 2007, pp. 719–724.

[29] L. Tessens, A. Ledda, A. Pizurica, and W. Philips, "Extending the depth of field in microscopy through curvelet-based frequency-adaptive image

fusion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2007, pp. 1861–1864.
[30] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse coding for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, Jun. 1996.
[31] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constr. Approx.*, vol. 13, no. 1, pp. 57–98, Mar. 1997.
[32] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[33] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted norm minimization algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.
[34] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 3736–3745, Dec. 2006.
[35] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
[36] N. Mitianoudis and T. Stathaki, "Joint fusion and blind restoration for multiple image scenarios with missing data," *Comput. J.*, vol. 50, no. 6, pp. 660–673, Nov. 2007.
[37] C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, Feb. 2000.

Bin Yang received the B.S. degree from Zhengzhou University of Light Industry, Zhengzhou, China, in 2005. He is currently working toward the Ph.D. degree with Hunan University, Changsha, China.

His technical interests include image fusion and pattern recognition.

Shutao Li (M'07) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University. From May 2001 to October 2001, he was a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong. From November 2002 to November 2003, he was a Postdoctoral Fellow with the Royal Holloway College, University of London, Egham, U.K., working with Prof. J.-S. Taylor. From April 2005 to June 2005, he was a Visiting Professor with the Department of Computer Science, Hong Kong University of Science and Technology. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He has authored or coauthored more than 100 refereed papers. His professional interests are information fusion, pattern recognition, bioinformatics, and image processing.

Dr. Li served as a member of the Neural Networks Technical Committee from 2007 to 2008. He has won two Second-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006.