

Research Article

Image Sequence Fusion and Denoising Based on 3D Shearlet Transform

Liang Xu, Junping Du, and Zhenhong Zhang

Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Junping Du; junpingdu@126.com

Received 22 December 2013; Accepted 8 February 2014; Published 17 March 2014

Academic Editor: Feng Gao

Copyright © 2014 Liang Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a novel algorithm for image sequence fusion and denoising simultaneously in 3D shearlet transform domain. In general, the most existing image fusion methods only consider combining the important information of source images and do not deal with the artifacts. If source images contain noises, the noises may be also transferred into the fusion image together with useful pixels. In 3D shearlet transform domain, we propose that the recursive filter is first performed on the high-pass subbands to obtain the denoised high-pass coefficients. The high-pass subbands are then combined to employ the fusion rule of the selecting maximum based on 3D pulse coupled neural network (PCNN), and the low-pass subband is fused to use the fusion rule of the weighted sum. Experimental results demonstrate that the proposed algorithm yields the encouraging effects.

1. Introduction

Video sensors have been applied extensively to video monitor and machine vision fields with the performance improvement and the reduced cost. In real applications, because of the effect of objects movement, occlusion, and illumination variation factors, single video sensor may not satisfy the requirements [1]. To get the complete information of a scene, multiple different sensors are employed simultaneously to capture the content of the same scene [2]. To utilize the information captured from multiple sensors sufficiently and efficiently, we need to combine the contents from different video sensors into an image sequence. This can be easily satisfied by the image sequence fusion methods, which can merge multiple image sequences from different sensors into a single image sequence containing all important information, eliminating redundancy and improving the availability of the information.

In the past decade, a variety of image fusion approaches have been developed for different applications [3–5]. The simplest fusion method is the weighted average of source images. However, the method is brittle and easy to lead to the contrast reduction and introduce artifacts. Now, fusion methods based on multiscale decomposition in the transform

domain are increasingly popular because of better robustness and reliability [6–8]. The most existing fusion methods are designed for static images; however, the fusion methods especially for image sequences and videos are seldom. The fusion methods for static images can even ensure the quality of single frame by the frame-by-frame fusion [9], but the temporal consistency and stability of the image sequence is hard to be preserved. So far, there are several fusion methods for image sequences and videos. By utilizing the information of temporal axis, in recent years, several state-of-the-art video fusion approaches [10, 11] have been developed. Based on the three-dimensional surfacelet transform, Zhang et al. [12] propose a video fusion framework which fuses multiframe images of input videos as a whole procedure. However, the method is still insufficient in extracting the spatial-temporal information from videos with dynamic background. Zhang et al. [13] propose a multisensor video fusion method based on three-dimensional uniform discrete curvelet transform (3D-UDCT) and spatial-temporal structure tensor. However, these schemes did not still obtain the satisfactory results due to the insufficient capability of representing motion information.

In addition, the most existing image fusion methods still only pay attention to combine the useful pixels of source

images and seldom consider dealing with the artifacts. The noises are easily introduced into the fusion image together with the useful pixels if the further processes are not performed for eliminating noises. There are some fusion methods that consider the enhancement and denoising when the fusion procedure is performed. Piella [14] presents a variational model to perform the fusion of input images while preserving the salient information and enhancing the contrast. Yang and Li [15] propose a sparse representation-based multifocus image fusion method which can simultaneously carry out denoising and fusion of noised source images.

In the paper, we propose a novel image sequence fusion algorithm which implements fusion and denoising simultaneously in 3D shearlet transform domain. Input image sequences are first decomposed by 3D shearlet transform into high- and low-pass subbands, and then the coefficients are merged and denoised in frequency domain. Finally, inverse 3D shearlet transform is applied on the fused coefficients to reconstruct the image sequence. Multiple frame images can be decomposed to different frequency scales once in 3D space by 3D shearlet, which has certain direction selectivity and can avoid aliasing and instability for the coefficients of neighbor frames. 3D shearlet considers the motion feature in temporal axis, such that the decomposed coefficients can approximate sufficiently the spatial-temporal features of image sequences in different scales. For merging coefficients, to preserve the consistency and stability of the interframe coefficients, we propose a spatial-temporal fusion rule based on 3D PCNN [16], which can extract the spatial-temporal information of the corresponding subbands from the neighbor frames. In addition, the most existing image fusion methods still only pay attention to combine the pixels of source images and seldom consider dealing with the noises. Here, we combine the fusion and denoising together and directly deal with noises on the coefficients using the recursive filter. Superior to the separate fusion and denoising based on multiscale transforms, our method will reduce the error only due to the need to perform the decomposition and reconstruction once.

The remainder of the paper is organized as follows. Section 2 reviews basic 3D shearlet transform theory in brief. Section 3 describes the proposed image sequence fusion and denoising algorithm in detail. Section 4 presents and discusses the experimental results. Section 5 concludes.

2. 3D Shearlet Transform

In this section, we briefly review theory and properties of 3D shearlet transform, which will be used in the rest of this paper (see [17] for details).

The shearlet approach inherits the general advantages of curvelets and surfacelets. During the last decade, to overcome the limitations of wavelets and other traditional methods, a new class of multiscale methods is introduced through a novel framework. The framework can effectively combine the standard multiscale decomposition and efficiently capture anisotropic features. The shearlet just belongs to part of the new class of multiscale methods. The shearlet representation is a multiscale pyramid of well-localized waveforms defined

at various locations and orientations. This representation can break through the limitations of traditional multiscale systems in dealing with multidimensional data.

The 3D shearlet transform is constructed by a shearlet system associated with the pyramidal regions. Three pyramidal regions P_1 , P_2 , and P_3 are obtained by partition of the Fourier space \widehat{R}^3 , defined as follows (shown in Figure 1):

$$\begin{aligned} P_1 &= \left\{ (\xi_1, \xi_2, \xi_3) \in R^3 : \left| \frac{\xi_2}{\xi_1} \right| \leq 1, \left| \frac{\xi_3}{\xi_1} \right| \leq 1 \right\}, \\ P_2 &= \left\{ (\xi_1, \xi_2, \xi_3) \in R^3 : \left| \frac{\xi_1}{\xi_2} \right| < 1, \left| \frac{\xi_3}{\xi_2} \right| \leq 1 \right\}, \\ P_3 &= \left\{ (\xi_1, \xi_2, \xi_3) \in R^3 : \left| \frac{\xi_1}{\xi_3} \right| < 1, \left| \frac{\xi_2}{\xi_3} \right| < 1 \right\}. \end{aligned} \quad (1)$$

The directionality of the shearlet systems is controlled through the use of shearing matrices. The 3D shearlet systems for $L^2(R^3)$ are a Parseval frame, which is obtained by using an appropriate combination of the systems of shearlets associated with the pyramidal regions P_d ($d = 1, 2, 3, l = (l_1, l_2) \in Z^2$). In this way, the 3D shearlet systems are defined as the collections consisting of the coarse-scale shearlets, the interior shearlets, and the boundary shearlets:

$$\begin{aligned} &\{\tilde{\psi}_{-1,k} : k \in Z^3\} \cup \{\tilde{\psi}_{j,l,k} : j \geq 0, l_1, l_2 = \pm 2^j, k \in Z^3\} \\ &\{\tilde{\psi}_{j,l,k,d} : j \geq 0, |l_1| < 2^j, |l_2| \leq 2^j, k \in Z^3, d = 1, 2, 3\}, \end{aligned} \quad (2)$$

where the shearing parameters l_1 and l_2 control the orientations of the support regions in 3D shearlet systems. Figure 2 illustrates a typical support region. The orientation of the support region is controlled by $l = (l_1, l_2)$. It can be seen that the support region is becoming more elongated as j increases.

A numerical implementation of the 3D discrete shearlet transform takes advantage of the sparsity properties of the corresponding continuous representation. The 3D digital shearlet transform can preserve the discrete integer lattice and enable a natural transition from the continuous to the discrete setting due to the use of shearing matrices rather than rotations. The 3D digital shearlet transform algorithm is implemented through a cascade of a multiscale decomposition and a directional filtering stage. The multiscale decomposition is first implemented using the Laplacian pyramid algorithm. And then the directional components are obtained using shearing matrices to control orientations in the pseudospherical domain.

3. Proposed Image Sequence Fusion and Denoising Algorithm

In this section, the proposed image sequence fusion and denoising algorithm is presented in detail. The main idea is that image sequence fusion and denoising are implemented simultaneously in 3D shearlet transform domain. The framework of the proposed algorithm is shown in Figure 3. For the clearness of the presentation, we assume that two registered image sequences are combined.

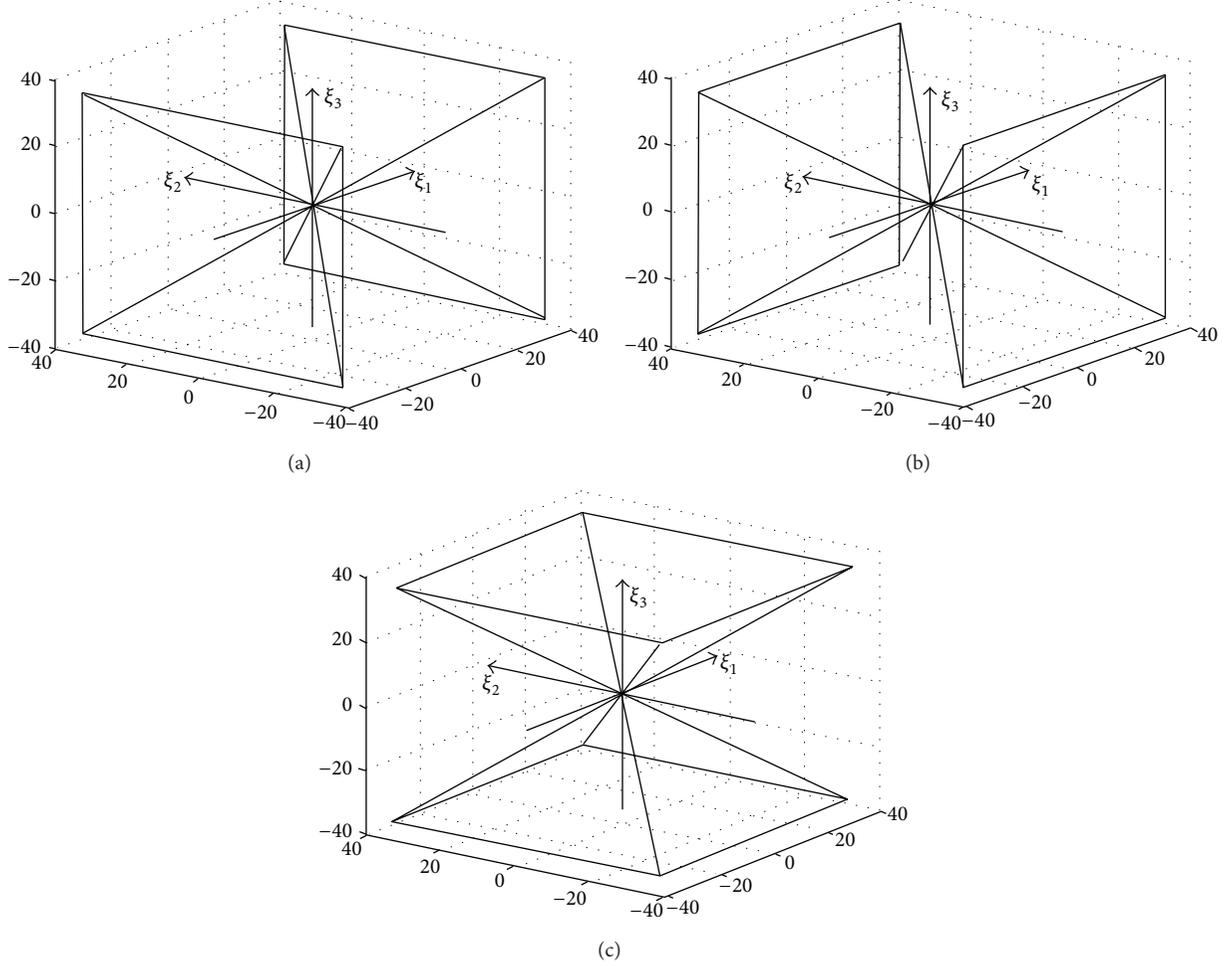


FIGURE 1: ((a)–(c)) The pyramidal regions P_1 , P_2 , and P_3 in frequency domain \widehat{R}^3 are illustrated.

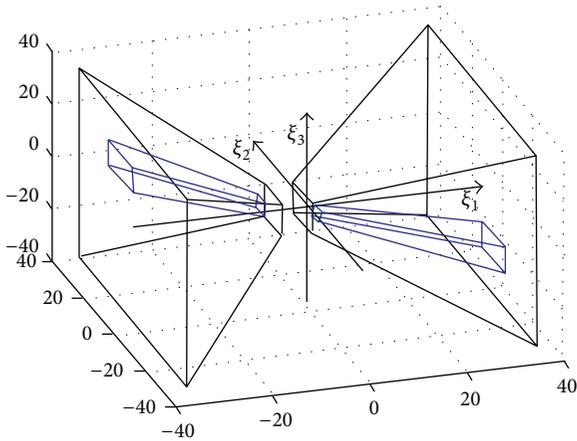


FIGURE 2: Frequency support of a representative shearlet function $\psi_{j,l,k}$, inside the pyramidal region P_1 .

Two groups of image sequences Va and Vb are decomposed by 3D shearlet transform into two groups of high-pass subbands and two groups of low-pass subbands for

several frames. Then, high-pass subbands are denoised by the recursive filter (RF) [18]. Next, the fused subband coefficients of all frames are obtained to employ a 3D PCNN-based spatial-temporal saliency fusion rule. Finally, the fused image sequence is reconstructed by inverting 3D shearlet transform on the fused coefficients of all frames. The steps of the proposed algorithm are presented as follows.

Step 1. Input image sequences are transformed using 3D shearlet transformation to the frequency domain and produce the low- and high-pass subbands $\{C_{j_0}^{Va}(x, y, t), C_{j,l}^{Va}(x, y, t)\}$ and $\{C_{j_0}^{Vb}(x, y, t), C_{j,l}^{Vb}(x, y, t)\}$ for several frames, where $C_{j_0}(x, y, t)$ denotes the low-pass subband coefficients at the coarsest scale of the t th frame and $C_{j,l}(x, y, t)$ denotes the high-pass subband coefficients at the j th scale and in the l th direction.

Step 2. For the high-pass subband coefficients, the recursive filter is performed on the coefficients of each frame for eliminating noises. Then, 3D PCNN is used to compute the spatial-temporal activity levels of the denoised high-pass coefficients to obtain the spatial-temporal activity maps,

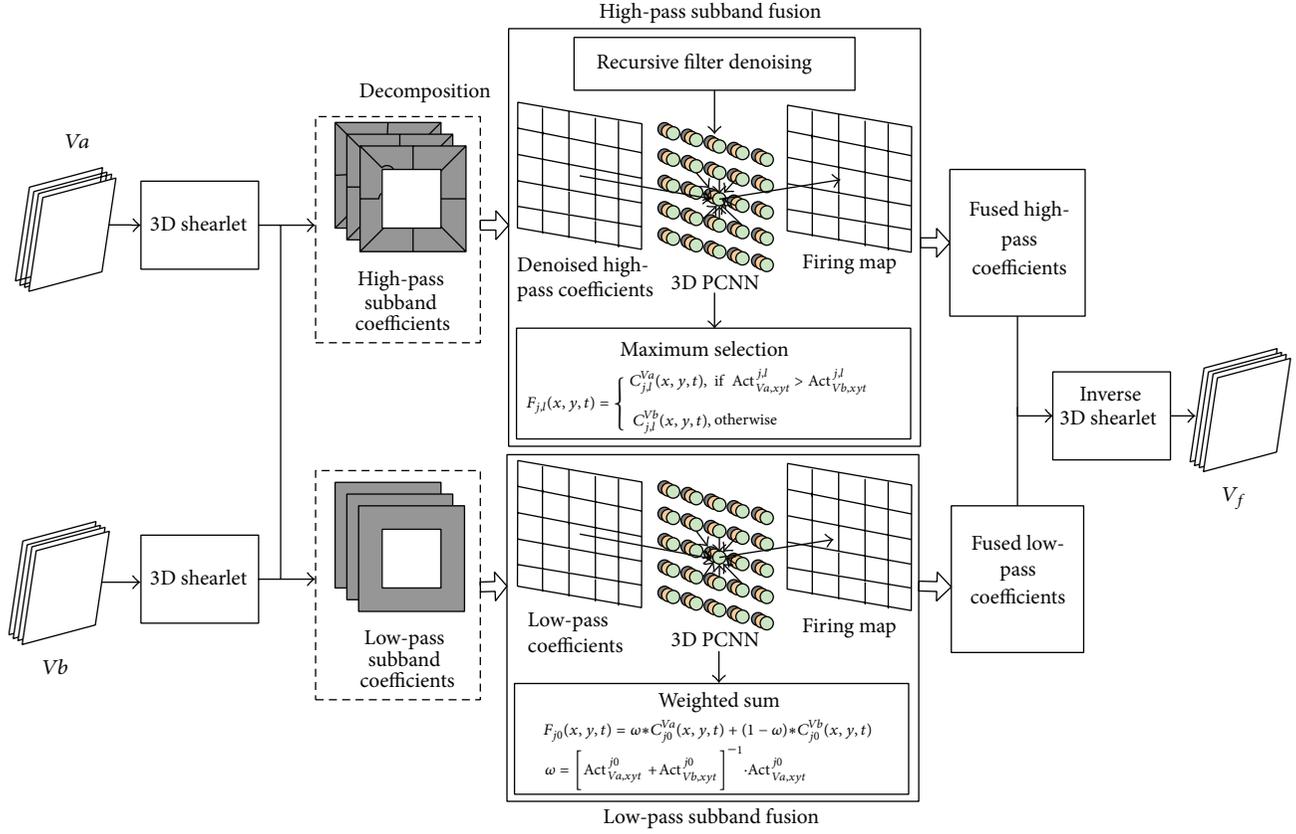


FIGURE 3: The diagram of the proposed image sequence fusion and denoising algorithm.

which are employed to merge the fused high-pass subband by the selecting maximum.

Step 3. The low-pass subband coefficients are merged using a spatial-temporal energy weighted fusion rule based on the activity maps yielded by 3D PCNN.

Step 4. Apply the inverse 3D shearlet transform to the fused coefficients $\{F_{j_0}(x, y, t), F_{j,l}(x, y, t)\}$ for all frames and then obtain the fused image sequence V_f .

3.1. 3D PCNN-Based Spatial-Temporal Fusion. This section discusses the proposed 3D PCNN-Based Spatial-Temporal Fusion in detail. Based on the experimental observations of synchronous pulse bursts in cat and monkey visual cortex, a novel biological neural network called PCNN is developed. PCNN neuron consists of receptive field, modulation field, and pulse generator [19, 20]. Different from the traditional neural network, PCNN is a feedback network and does not need to be trained. A secondary receptive field of PCNN, known as the linking field, integrates inputs from adjacent neurons to modulate the primary feeding field. In image processing, 2D PCNN is a single layer and two-dimensional connection neural network. Considering the spatial features of 2D image plane, 2D PCNN utilizes the output of the spatial neighborhood pixels as the inner input of next iteration. The similar neurons in PCNN generate pulses simultaneously to

compensate effectively the spatial incoherence and the slight amplitude changes, so that PCNN can measure the salient object regions completely. It has been successfully used to deal with the single static image. However, if 2D PCNN is applied directly on an image sequence frame by frame, it may not extract the temporal motion information.

To adapt PCNN to deal with image sequences, 2D PCNN is extended to 3D PCNN utilizing the correlation of neighbor frames. 3D PCNN has been used to the segmentation of stereo images successfully [16]. Here, 3D PCNN is employed to measure the activity energy of the coefficients from 3D shearlet decomposition. Let $C_{j,l}(x, y, t)$ indicate the coefficient located at (x, y) in the j th scale at the l th direction and at t th frame. $C_{j,l}(x, y, t)$ in each subband is inputted to 3D PCNN as the external feeding input. Both the last output of the spatial neighbor pixels and the corresponding output of the neighbor frames are used as the inner linking input. In this way, 3D PCNN can extract the spatial-temporal information sufficiently in an image sequence. The 3D PCNN is defined as follows:

$$\begin{aligned}
 F_{xyt}^{j,l} [n] &= C_{j,l} (x, y, t) \\
 L_{xyt}^{j,l} [n] &= \exp(-\alpha_L) L_{xyt}^{j,l} [n-1] \\
 &+ V_L \sum_{pqr} W_{xyt,pqr}^{j,l} Y_{xyt,pqr}^{j,l} [n-1]
 \end{aligned}$$

$$\begin{aligned}
U_{xyt}^{j,l} [n] &= F_{xyt}^{j,l} [n] * (1 + \beta L_{xyt}^{j,l} [n]) \\
\theta_{xyt}^{j,l} [n] &= \exp(-\alpha_\theta) \theta_{xyt}^{j,l} [n-1] + V_\theta Y_{xyt}^{j,l} [n-1] \\
Y_{xyt}^{j,l} [n] &= \begin{cases} 1, & \text{if } U_{xyt}^{j,l} [n] > \theta_{xyt}^{j,l} [n] \\ 0, & \text{otherwise,} \end{cases}
\end{aligned} \tag{3}$$

where the coefficient $C_{j,l}(x, y, t)$ is inputted to the feeding input $F_{xyt}^{j,l}$. The linking input $L_{xyt}^{j,l}$ is equal to the sum of neurons firing times in linking range, where α_L indicates the decay constants and V_L is the amplitude gain. $W_{xyt,pqr}$ is the weighted coefficient (p, q, r point out the size of linking range in 3D PCNN). $Y_{xyt}^{j,l}[n-1]$ is the output of the neuron from the previous iteration. The internal state signal $U_{xyt}^{j,l}$ is obtained by modulating $F_{xyt}^{j,l}$ and $L_{xyt}^{j,l}$, where β is the linking strength. $\theta_{xyt}^{j,l}$ is the threshold, where α_θ is the decay constants and V_θ is the amplitude gain. n indicates the iteration times. If $Y_{xyt}^{j,l} = 1$, the neuron will generate a pulse, called one firing. If $Y_{xyt}^{j,l} = 0$, the neuron will not generate a pulse.

In the paper, the 3D PCNN can be used to measure the spatial-temporal saliency of the coefficients. Through combining the coefficients of the corresponding scales and directions from neighbor frames, we construct a 3D volume whose size is $M \times N \times T$. In this way, a neuron model of 3D PCNN is constructed, where each of the coefficients is the external input of 3D PCNN. The neuron with the maximum coefficient value is first fired. Following, the similar neurons from the 3D space constructed by the internal linking matrix W are motivated to produce synchronization pulse through pulse propagation. The generated pulse sequence $Y[n]$ forms a 3D binary sequence, which contains the saliency information of images, for example, regions, edges, textures, and so forth.

In applications, the firing times are generally employed to represent image information. Firing times $\text{Act}_{xyt}^{j,l}[n]$ can be computed by accumulating all pixels before and including the present iteration, written as follows:

$$\text{Act}_{xyt}^{j,l} [n] = \text{Act}_{xyt}^{j,l} [n-1] + Y_{xyt}^{j,l} [n], \tag{4}$$

where $\text{Act}_{xyt}^{j,l}[n]$ is often used to indicate the total firing times in n iteration. Here, the firing times indicate the activity energy of the coefficients.

The high-pass subbands of 3D shearlet decomposition contain abundant detail information, for example, lines, edges, contours, and so forth. To preserve the detail components in fused images, we propose a fusion rule of spatial-temporal selecting maximum based on 3D PCNN for the high-pass subbands. Coefficients with large firing times are selected as the fused coefficients. Consequently, the fused high-pass subband coefficients located at (x, y) in the j th

scale at the l th direction and at t th frame denoted as $F_{j,l}(x, y, t)$ are defined as

$$F_{j,l}(x, y, t) = \begin{cases} C_{j,l}^{Va}(x, y, t), & \text{if } \text{Act}_{Va,xyt}^{j,l} > \text{Act}_{Vb,xyt}^{j,l} \\ C_{j,l}^{Vb}(x, y, t), & \text{otherwise.} \end{cases} \tag{5}$$

The low-pass subband of 3D shearlet decomposition in the coarsest scale contains the main energy of source images and denotes abundant structural information. The fusion rule of the low-pass subband employs a spatial-temporal weighted fusion rule based on firing times of 3D PCNN. The fused coefficients of low-pass subbands denoted as $F_{j_0}(x, y, t)$ employ a weighted fusion rule based on firing times of 3D PCNN on coefficients $C_{j_0}^{Va}(x, y, t)$ and $C_{j_0}^{Vb}(x, y, t)$, which are defined as

$$F_{j_0}(x, y, t) = \omega * C_{j_0}^{Va}(x, y, t) + (1 - \omega) * C_{j_0}^{Vb}(x, y, t), \tag{6}$$

where ω is the weight of coefficients and $\text{Act}_{xyt}^{j_0}$ is computed by (3) and (4):

$$\omega = \left[\text{Act}_{Va,xyt}^{j_0} + \text{Act}_{Vb,xyt}^{j_0} \right]^{-1} \cdot \text{Act}_{Va,xyt}^{j_0}. \tag{7}$$

Finally, apply the inverse 3D shearlet transform to the fused coefficients of N frames $\{F_{j_0}(x, y, t), F_{j,l}(x, y, t)\}$, and then obtain the fused image sequence V_f .

3.2. Recursive Filter Denoising. The previous section introduces the fusion method of the decomposed coefficients. When input images contain some noises, the fused image will also introduce noise artifacts if the coefficients are only directly merged. So we need first to deal with the coefficients for denoising. This section presents mainly the recursive filter method [18] which is used as the denoising of the coefficients. The recursive filter is a real-time edge-preserving smoothing filter. Comparing with the separate fusion and denoising based on multiscale transform, the fusion and denoising simultaneously on the coefficients decomposed by 3D shearlet transform will reduce the error caused by the decomposition and reconstruction. In addition, the denoising filter is run on the coefficients of different scales and directions such that it enhances the robustness of the algorithm.

The low-pass subband contains the main energy of images and denotes the structural information. The high-pass subbands contain the abundant details, for example, lines, edges, and contours. In general, the noises appear in the high-pass subbands, so here the recursive filter is only performed on the high-pass subband coefficients to obtain the denoised coefficients $\tilde{C}_{j,l}$, which are defined as follows:

$$\tilde{C}_{j,l} = \text{RF}(C_{j,l}), \tag{8}$$

where RF indicates the recursive filter. When the noises are eliminated, the details need to be preserved on the high-pass

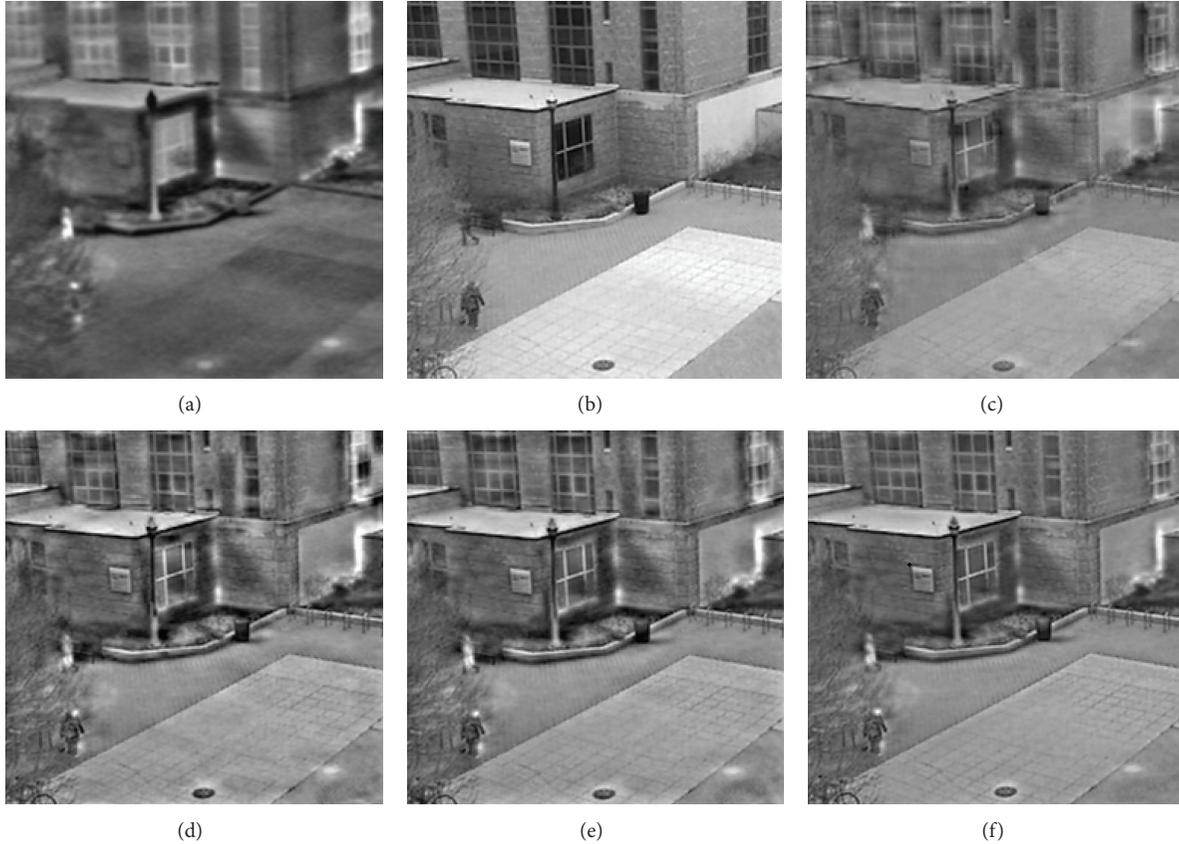


FIGURE 4: Comparisons of different fusion methods. (a) A frame of visible light image sequence; (b) corresponding frame of infrared image sequence; and fused frame images using (c) DWT, (d) 3D DWT, (e) 3D DTCWT, and (f) proposed 3DShearlet-3DPCNN methods.

subbands. This can be easily satisfied by the edge-preserving filter, called the recursive filter:

$$J[n] = (1 - a^d)I[n] + a^d J[n - 1], \quad (9)$$

where $a \in [0, 1]$ is a feedback coefficient, $I[n]$ is the value of the n th coefficient of the input high-pass subband, $J[n]$ is the n th coefficient of the filtered high-pass subband, and d is the distance between neighborhood coefficients of the high-pass subband. As d increases, a^d goes to zero, stopping the propagation chain and preserving details on high-pass subbands. More details about the recursive filter can refer to [18].

The recursive filter is performed in all N frame images for obtaining the denoised high-pass coefficients. Following, the fused coefficients can be obtained through merging the denoised coefficients.

4. Experiments and Analysis

In this section, the proposed image sequence fusion algorithm based on 3D shearlet and 3D PCNN (named as 3DShearlet-3DPCNN) is tested on several groups of image sequences. For comparison, besides the fusion scheme proposed in this paper, another three fusion algorithms, the discrete wavelet transform based (DWT), 3D DWT based, and

3D dual-tree complex wavelet transform based (3D DTCWT) methods, are used to fuse the same image sequences. All of these methods use the average sum and absolute maximum selection schemes for merging low- and high-pass subband coefficients, respectively. The decomposition level of all of the transforms is three. It is assumed that source image sequences have been registered. The tested collections contain the clear and the noised image sequences.

4.1. Fusion Results of Clear Image Sequences. The first experiment is a pair of visible light and infrared image sequences without noises. A pair of frames from input image sequences and four fused images produced by DWT, 3D DWT, 3D DTCWT, and proposed 3DShearlet-3DPCNN are shown in Figure 4. Figure 4(a) is an infrared image frame from an input image sequence, and Figure 4(b) is a visible light source image frame from the corresponding visible light image sequence. From Figure 4, we can observe that Figure 4(c) is the worst result, which has the lower contrast, and the motion objects tend to blur. Figures 4(d)–4(f), produced by 3D DWT, 3D DTCWT, and proposed 3DShearlet-3DPCNN, present the better results, which have higher contrast, clearer motion objects. However, Figures 4(d) and 4(e) have some distortion near the windows. The edges of the windows appear warp. The middle wall between windows introduces the dark regions.

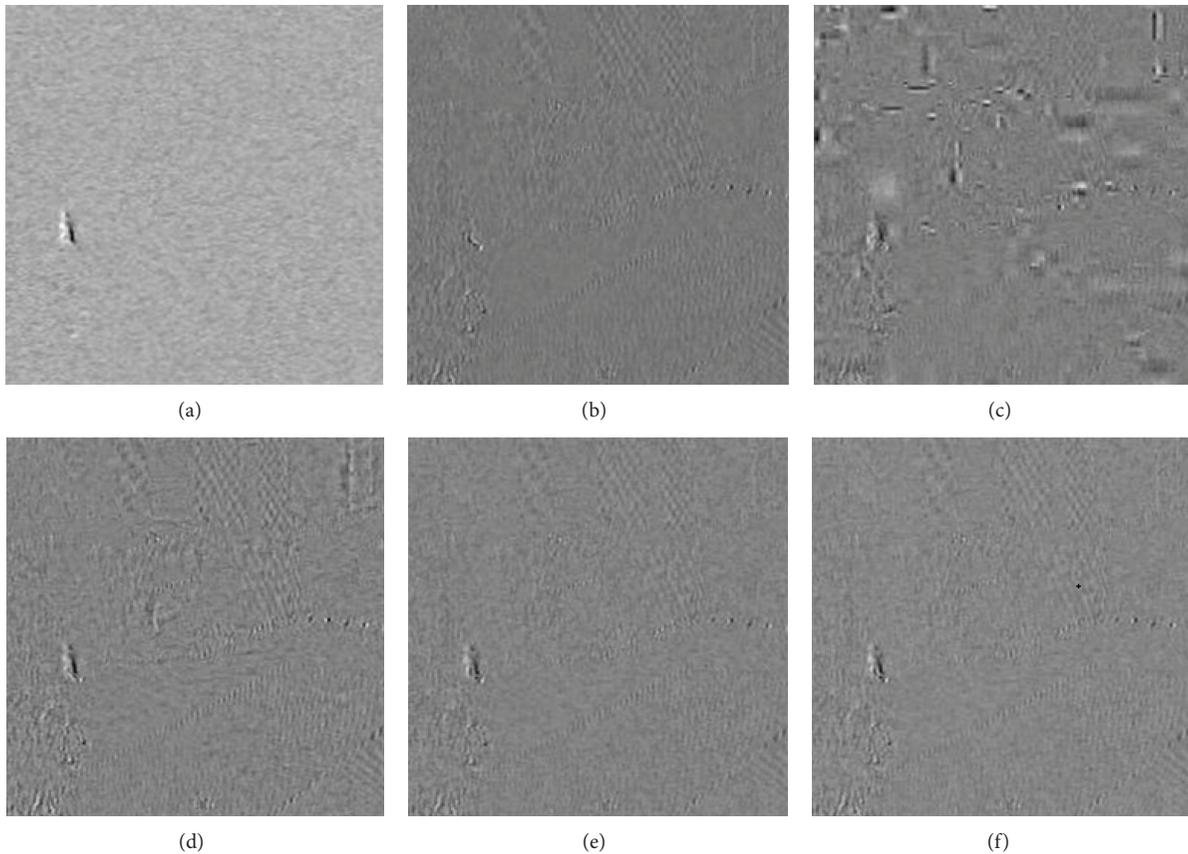


FIGURE 5: One set of IFD images of input and fused image sequences corresponding to (a)–(f), IFD images between the current frame images in Figures 4(a)–4(f), and their corresponding preframe images, respectively.

In contrast, the proposed method (Figure 4(f)) yields the best result. This comparison reveals that the proposed fusion approach effectively determines complementary or redundant information between input image frames. It can preserve all the useful information of the input image frames while avoiding artifacts. In addition, to evaluate the performance of image sequence fusion methods in temporal stability and consistency, a clearer comparison is made by examining the interframe-difference (IFD) between the current frame and the preframe, shown in Figure 5. One set of IFD images for source and fused frames between the current frames in Figure 4 and their corresponding preframes are shown in Figure 5. Obviously, the IFDs in Figure 5(c) introduce many artifacts, which exist neither in Figure 5(a) nor in Figure 5(b). In Figure 5(d), artifacts are greatly reduced. In Figures 5(e) and 5(f), we cannot see nearly any artifacts. This further demonstrates that 3D DTCWT and proposed fused image frames have better temporal stability and consistency.

Figure 6 shows one pair of input image frames and the corresponding fused results. Comparisons of different fused results (Figures 6(c)–6(f)) and the frames fused using DWT, 3D DWT, and 3D DTCWT (Figures 6(c)–6(e)) are not clear enough and have lower contrast. In particular, artifacts around the man were also introduced in Figure 6(c). The image frame fused using the proposed approach (Figure 6(f)) is obviously clearer and has stronger contrast. Experimental

results further demonstrate that the proposed algorithm can effectively improve the quality of the fused image sequence.

For more accurate comparison, besides the visual analysis, the performance of fusion algorithms needs to be further measured by objective quantitative analysis tools. Here two metric tools were used: the spatial-temporal gradient preservation based video fusion performance metric ($DQ^{AB/F}$) [21] and the mutual information of IFD images (IFD_MI) [9]. $DQ^{AB/F}$ indicates how much the spatial-temporal information is extracted and transferred into the fused image sequence. IFD_MI reflects the performance of the image sequence fusion method in temporal stability and consistency. The higher the values are for the two metrics, the better are the fusion results.

The measurement results of the different fusion methods for the image frames in Figure 4 and Figure 6 are shown in Figure 7. The results show that the DWT method has the lowest scores. Compared with the DWT, 3D DWT and 3D DTCWT methods have the improved results. The proposed method represents the highest performance in the listed fusion approaches for $DQ^{AB/F}$ and IFD_MI metrics. The quantitative results are consistent with the visual analysis, illustrating that the proposed fusion method has the superior performance while it can preserve the temporal stability and consistency for image sequences.



FIGURE 6: Comparisons of different fusion methods. (a) A frame of visible light image sequence; (b) corresponding frame of infrared image sequence; and fused frame images using (c) DWT, (d) 3D DWT, (e) 3D DTCWT, and (f) proposed 3DShearlet-3DPCNN methods.

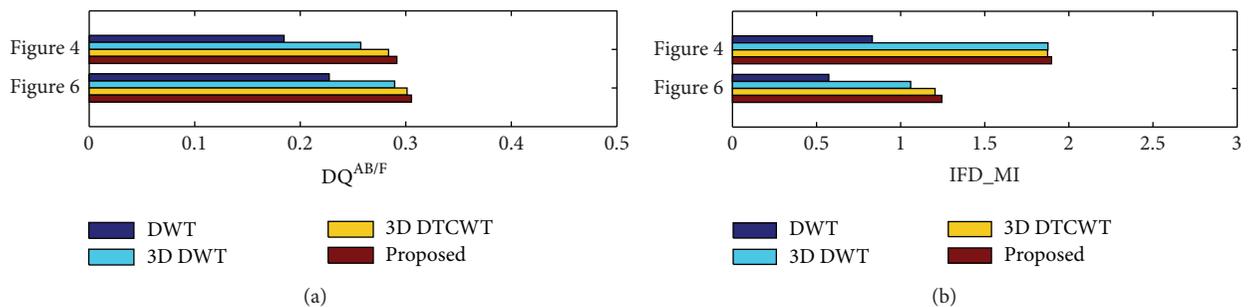


FIGURE 7: Quality measures for the different fusion methods.

4.2. Fusion Results of Noised Image Sequences. In previous discussion, the fusion results of different algorithms for the clear image sequences have been analyzed by both visual and objective metric tools. In the section, the denoising and fusion results of different algorithms for the noised image sequences are compared. Our method is compared to the 3D DTCWT fusion method followed by 3D DTCWT denoising [22], named as 3D DTCWT-FD.

Figure 8 presents the fusion and denoising results yielded by our method and 3D DTCWT-FD method from a pair of real world noise image sequences. Observing the source image frames (Figures 8(a) and 8(b)), the visible light image frame (Figure 8(a)) is clear and the infrared image frame

(Figure 8(b)) contaminated by noises. It can be seen that the fusion image frames (Figures 8(c) and 8(d)) have eliminated the noises. However, the 3D DTCWT-FD result (Figure 8(c)) becomes blurred due to over-smoothing. The proposed denoising and fusion method yields the optimal result which nearly removes all the noises and preserves the high contrast.

In addition, to evaluate the performance of different fusion and denoising schemes, $DQ^{AB/F}$, IFD_MI, and peak signal to noise ratio (PSNR) metrics are used. Table 1 shows the results of $DQ^{AB/F}$ and IFD_MI metrics for Figure 8. We can see that the proposed method gets the higher scores in both $DQ^{AB/F}$ and IFD_MI. This indicates that the proposed

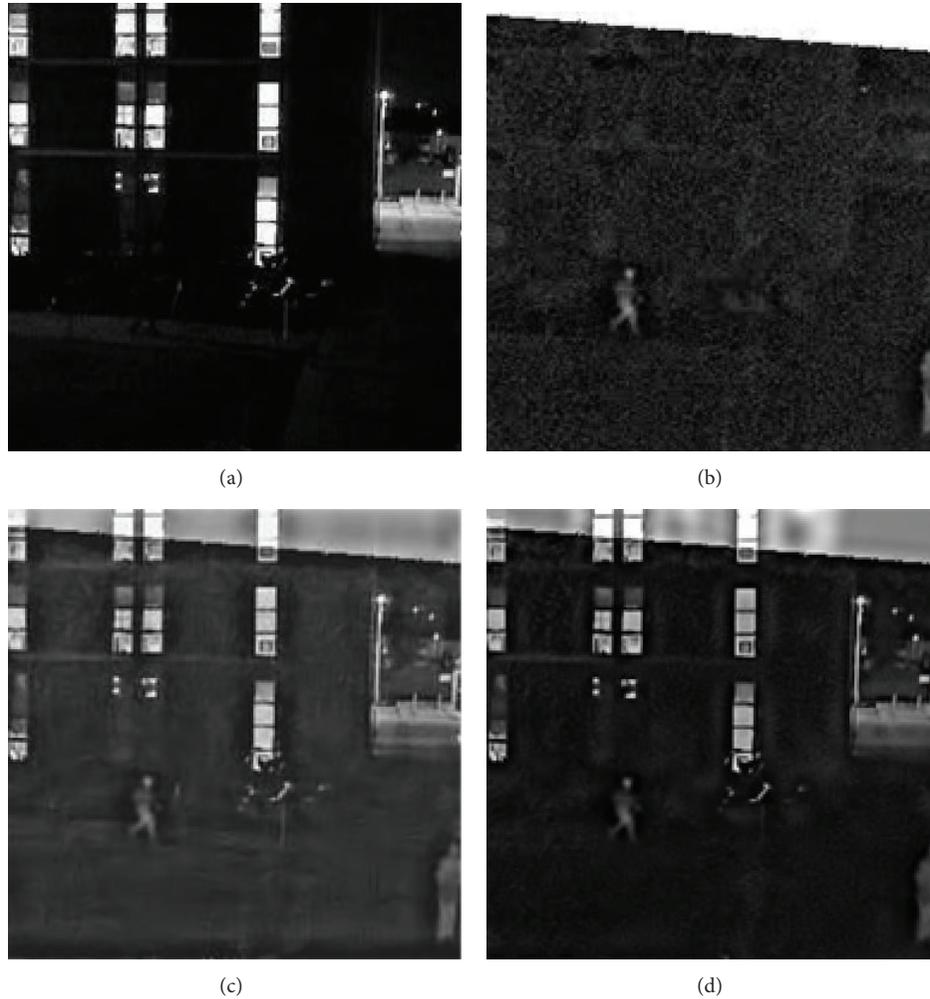


FIGURE 8: Dublin source image frames and denoised and fused results: (a) a visible light image frame; (b) an infrared image frame; and fused image frames using (c) 3D DTCWT-FD and (d) proposed 3DShearlet-3DPCNN.

fusion and denoising simultaneously method contains all the useful information from source images, while eliminating the noises. Figure 9 presents the PSNR results for 1000 frame fusion images associated with Figure 8. Here, to get the PSNR results, the clear visible light source image frames are used as the reference images. From Figure 9, we can observe that the results of the proposed method have higher PSNR scores which indicate that our method yields better the denoised images. Consequently, from the comparisons of $DQ^{AB/F}$, IFD_MI, and PSNR metrics, the experiments further demonstrate that the proposed image sequence fusion and denoising simultaneously method yields the satisfactory results, which merge the important complementary information and remove the artifacts.

5. Conclusion

The paper proposes a novel algorithm for image sequence fusion and denoising based on 3D shearlet transform. The most existing image fusion methods do not deal with the

TABLE 1: Quality measures for the different fusion methods.

Data set	Metric	3D DTCWT-FD	Proposed
Figure 8	$DQ^{AB/F}$	0.1841	0.1903
	IFD_MI	0.7294	1.4470

artifacts when the fusion procedure is performed. If source images contain noises, the noises may be also transferred into the fusion image together with useful pixels. Therefore, we propose that the recursive filter is first performed on the high-pass subbands to obtain the denoised high-pass coefficients in 3D shearlet transform domain. The high-pass subbands are then combined to employ the fusion rule of the selecting maximum based on 3D PCNN, and the low-pass subband is fused to use the fusion rule of the weighted sum. In this way, the fusion and denoising simultaneously can be achieved for the noised image sequences. Experiments demonstrate that the proposed method improves greatly the quality of the fused image sequence.

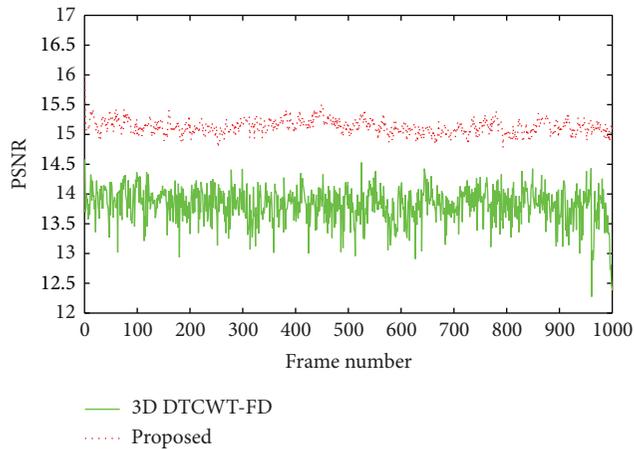


FIGURE 9: PSNR results of the fused and denoising image sequence associated with Figure 8 (frame number: 1000). The red dotted line represents the PSNR scores of our method. The green solid line indicates the PSNR values of the 3D DTCWT-FD.

Conflict of Interests

The authors do not have any actual or potential conflict of interests.

Acknowledgments

This work was supported by the National Basic Research Program of China (973 Program) 2012CB821200 (2012CB821206) and the National Natural Science Foundation of China (no. 61320106006).

References

- [1] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: a review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.
- [2] G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognition*, vol. 37, no. 9, pp. 1855–1872, 2004.
- [3] N. Mitianoudis and T. Stathaki, "Pixel-based and region-based image fusion schemes using ICA bases," *Information Fusion*, vol. 8, no. 2, pp. 131–142, 2007.
- [4] L. Yang, B. L. Guo, and W. Ni, "Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform," *Neurocomputing*, vol. 72, no. 1–3, pp. 203–211, 2008.
- [5] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884–892, 2010.
- [6] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Information Fusion*, vol. 4, no. 4, pp. 259–280, 2003.
- [7] V. S. Petrović and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Transactions on Image Processing*, vol. 13, no. 2, pp. 228–237, 2004.
- [8] L. Xu, J. P. Du, Q. Hu, and Q. P. Li, "Feature-based image fusion with a uniform discrete curvelet transform," *International Journal of Advanced Robotic Systems*, vol. 10, pp. 1–11, 2013.
- [9] O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform," in *Proceedings of the International Conference on Image Processing*, vol. 3, pp. 288–291, Santa Barbara, Calif, USA, October 1997.
- [10] G. Xiao, K. Wei, and Z. Jing, "Improved dynamic image fusion scheme for infrared and visible sequence based on image fusion system," in *Proceedings of the 11th International Conference on Information Fusion*, pp. 1–6, Cologne, Germany, July 2008.
- [11] X. H. Zhang, M. Tang, Z. C. Jiang, and R. F. Tong, "Video fusion oriented object motion removing," *Journal of Computational Information Systems*, vol. 8, no. 6, pp. 2317–2324, 2012.
- [12] Q. Zhang, L. Wang, Z. Ma, and H. Li, "A novel video fusion framework using surfacelet transform," *Optics Communications*, vol. 285, no. 13–14, pp. 3032–3041, 2012.
- [13] Q. Zhang, Y. L. Chen, and L. Wang, "Multisensor video fusion based on spatial-temporal salience detection," *Signal Processing*, vol. 93, no. 9, pp. 2485–2499, 2013.
- [14] G. Piella, "Image fusion for enhanced visualization: a variational approach," *International Journal of Computer Vision*, vol. 83, no. 1, pp. 1–11, 2009.
- [15] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884–892, 2010.
- [16] N. Chou, J. Wu, J. B. Bingren, A. Qiu, and K.-H. Chuang, "Robust automatic rodent brain extraction using 3-D pulse-coupled neural networks (PCNN)," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2554–2564, 2011.
- [17] P. S. Negi and D. Labate, "3D discrete shearlet transform and video processing," *IEEE Transactions on Image Processing*, vol. 21, no. 6, pp. 2944–2954, 2012.
- [18] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Transactions on Graphics*, vol. 30, no. 4, article 69, 2011.
- [19] R. Eckhorn, H. J. Reitboeck, M. Arndt, and P. Dicke, "Feature linking via synchronization among distributed assemblies: simulations of results from cat visual cortex," *Neural Computation*, vol. 2, no. 3, pp. 293–307, 1990.
- [20] J. L. Johnson and M. L. Padgett, "PCNN models and applications," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 480–498, 1999.
- [21] V. Petrovic, T. Cootes, and R. Pavlovic, "Dynamic image fusion performance evaluation," in *Proceedings of the 10th International Conference on Information Fusion*, pp. 1–7, Québec, Canada, July 2007.
- [22] I. W. Selesnick and K. Y. Li, "Video denoising using 2D and 3D dual-tree complex wavelet transforms," in *Wavelets: Applications in Signal and Image Processing X*, vol. 5207 of *Proceedings of SPIE*, no. 2, pp. 607–618, San Diego, Calif, USA, August 2003.