Performance comparison of different multi-resolution transforms for image fusion

Shutao Li*, Bin Yang, Jianwen Hu

College of Electrical and Information Engineering, Hunan University, Changsha 410082, China

A R T I C L E   I N F O
Article history:
Received 30 September 2009
Received in revised form 6 March 2010
Accepted 21 March 2010
Available online 27 March 2010

Keywords:
Image fusion
Multi-resolution transform
Wavelet
Curvelet
Contourlet

A B S T R A C T
Image fusion combines information from multiple images of the same scene to get a composite image that is more suitable for human visual perception or further image-processing tasks. In this paper, we compare various multi-resolution decomposition algorithms, especially the latest developed image decomposition methods, such as curvelet and contourlet, for image fusion. The investigations include the effect of decomposition levels and filters on fusion performance. By comparing fusion results, we give the best candidates for multi-focus images, infrared–visible images, and medical images. The experimental results show that the shift-invariant property is of great importance for image fusion. In addition, we also conclude that short filter usually provides better fusion results than long filter, and the appropriate setting for the number of decomposition levels is four.

1. Introduction

Today, imaging sensors of various types are widely used in military and civilian applications, such as battlefield surveillance, health-care applications, and traffic control. However, the information provided by different imaging sensors may be complementary and redundant. For example, visible image provides the outline of scene, while infrared image can show the existence of some special objects, such as concealed guns or people. To obtain an image that simultaneously contains the outline of scene as well as special objects for the convenience of human visual perception or for further image-processing tasks [1], image fusion can be used to integrate the information provided by individual sensors. In this paper, we concern with the fusion of three types of source images: multi-focus images, infrared–visible images, and medical images.

During the past two decades, many image fusion methods are developed [1–12]. According to the stage at which image information is integrated, image fusion algorithms can be categorized into pixel, feature, and decision levels [1]. The pixel-level fusion integrates visual information contained in source images into a single fused image based on the original pixel information [2]. In the past decades, pixel-level image fusion has attracted a great deal of research attention. Generally, these algorithms can be categorized into spatial domain fusion and transform domain fusion [3]. The spatial domain techniques fuse source images using local spatial features, such as gradient, spatial frequency, and local standard derivation [1]. For the transform domain methods, source images are projected onto localized bases which are usually designed to represent the sharpness and edges of an image [3]. Therefore, the transformed coefficients (each corresponds to a transform basis) of an image are meaningful in detecting salient features. Consequently, according to the information provided by transformed coefficients, one can select the required information provided from the source images to construct the fused image.

With the development of different transform bases, many kinds of multi-resolution transforms have been proposed and used for image fusion, including the pyramid decomposition [4,5,13], discrete wavelet (DWT) [6–8,14], stationary wavelet (SWT) [15,16], dual-tree complex wavelet (DTCWT) [18–21], curvelet (CVT) [22–28], contourlet (CT) [29–31], and nonsampled contourlet transform (NSCT) [32–34].

Zhang and Blum established a categorization of multiscale decomposition-based image fusion to achieve a high-quality digital camera image [7]. They focused mainly on fusing the multiscale decomposition coefficients. For this reason, only a few basic types were considered, i.e. the Laplacian pyramid transform, the DWT, and the discrete wavelet frame (DWF). Only visible images were considered in performance comparisons for digital camera application. Pajares and Cruz gave a tutorial of the wavelet-based image fusion methods [8]. They presented a comprehensive comparison of different pyramid merging methods, different resolution levels, and different wavelet families. Three fusion examples were provided, namely multi-focus images, multispectral-panchromatic remote sensing images, and functional–anatomical medical images.

Wavelets and related classical multiscale transforms conduct decomposition over a limited dictionary in which the two-
dimensional bases simply consist of all possible tensor products of one-dimensional basis functions. To solve this problem, some new multiscale transforms such as curvelet and contourlet are introduced [22–24,29]. The main motivation of these transforms is to pursue a “true” two-dimensional transform that can capture the intrinsic geometrical structure [22]. Various transforms have been used to explore the image fusion problem [25–28,30,31,33,34]; however, there is a lack of comprehensive comparison of these methods on different types of source images. In addition, notice that the general image fusion framework with pyramid decomposition and wavelet has been well studied [7,8]. In this paper, we investigate some recently developed multiscale image decomposition methods including the DWT, SWT, DTCWT, CVT, CT, and NSCT, especially different decomposition levels and filters using a general fusion rule.

The rest of this paper is organized as follows. In Section 2, the brief reviews of the DTCWT, CVT, CT, and NSCT are presented. In Section 3, we give the general image fusion framework using multiscale image decomposition. Section 4 presents details of numerical experiments and comprehensive discussions on the results. Finally, the main conclusions of this paper are given in Section 5.

2. Multi-resolution image decomposition

The multi-resolution transform investigated in this paper includes the DWT, SWT, DTCWT, CVT, CT, and NSCT. The DWT and SWT have been researched extensively, and their principles can be found in many literatures [14,15]; therefore, in this section, we will only briefly review the DTCWT, CVT, CT, and NSCT.

2.1. Dual-tree complex wavelet

Since there is no subsampling process and the size of the filters increases in each scale, the SWT is computationally inefficient, especially in multiple dimensions [18]. In addition, the SWT only provides details in three directions for each scale. To overcome these problems, the DTCWT is proposed, which is approximately shift-invariant, directionally selective, and computationally efficient. Dual-tree of wavelet filters is used to obtain the real and imaginary parts of complex wavelet coefficients. A simple delay of one sample between the filters of the first level in each tree is conducted, and then odd-length and even-length linear-phase filters are used alternately. The filters in the two trees are just time-reverse of each other. Fig. 1 shows the practical implementation of the DTCWT on a 1D signal. $h_0(n), h_1(n)$ denote the low-pass/high-pass filter pair for the upper filter bank, and $g_0(n), g_1(n)$ denote the low-pass/high-pass filter pair for the lower filter bank. The DTCWT satisfies the property of approximate shift-invariance and directional selectivity in multiple dimensions. More details of the DTCWT are available in [18].

2.2. Curvelet

The DWT, SWT, and DTCWT cannot capture curves and edges of images well. More reasonable bases should contain geometrical structure information when they are used to represent images. Candès and Donoho proposed the curvelet transform (CVT) with the idea of representing a curve as a superposition of bases of various lengths and widths obeying the scaling law $\text{width} \approx \text{length}^2$ [22–24]. Two examples of the CVT bases are shown in Fig. 2a. Fig. 2b presents two examples of wavelet bases. From Fig. 2, it can be seen that the CVT is more suitable for the analysis of image edges, such as curve and line characteristics, than wavelet. The CVT is referred to as the “true” 2D transform. The discrete version implemented in this research uses a “wrapping” transform. The flowchart of the second generation of curvelet transform is presented in Fig. 3. Firstly, the 2D FFT is applied to the source image to obtain Fourier samples. Next, a discrete localizing window smoothly localizes the Fourier transform near the sheared wedges obeying the parabolic scaling. Then, the wrapping transformation is applied to re-index the data. Finally, the inverse 2D FFT is used to obtain the discrete CVT coefficients. More details can be found in [22].

2.3. Contourlet and nonsubsampled contourlet transform

Different from the CVT which is first developed in continuous domain and then is discretized for sampled data, contourlet transform (CT) starts with a discrete-domain construction [29]. The CT is also deemed as a “true” two-dimensional transform that can
capture the intrinsic geometrical structure of an image. Two filter banks are employed to implement the CT as shown in Fig. 4. The Laplacian pyramid is first used to capture the point discontinuities, and then a directional filter bank is used to link point discontinuities into linear structures. As the DWT, the CT also has no shift-invariant property because of the down-sampling operation. The nonsubsampled version of the CT, the NSCT, is implemented via the “à trous” algorithm [32]. It is built upon nonsampled multiscale pyramids and nonsampled directional filter banks; therefore, a fully shift-invariant version of the CT is achieved.

3. Generic framework for multiscale-based image fusion

In this paper, we make an assumption that there are just two source images A and B. It should be noticed that the multiscale methods can easily be extended to more source images [7]. Fig. 5 illustrates the generic image fusion framework based on multiscale image decomposition methods. The source images are firstly decomposed into low-frequency subbands and a sequence of high-frequency subbands in different scales and orientations. Then at each position in the transformed subbands, the value with the highest saliency is selected to construct the fused subbands. Finally, the fused image is obtained by applying inverse transform on the fused subbands.

There are two key issues in pixel-level image fusion algorithms, namely identifying the most important information in source images and transferring the salient information into the fused image. They are referred as activity-level measurement and coefficient combination, respectively, in many literatures [7–9]. The activity-level measurement is used to express the salience of each coefficient in image fusion methods based on multiscale decomposition. Generally, it is described by the absolute value of the corresponding transform coefficients. Other techniques of activity-level measurements include the square of the corresponding coefficient method, rank filter method, and spatial frequency method. They have been well discussed in [8]. In this paper, we choose the absolute value of the corresponding coefficient at each position as the activity-level:

\[ A_i(p) = |D_i(p)| \]

where \( D_i \) is the multiscale coefficient and \( p = (x,y,l,k) \) is the index of a particular coefficient; \( x \) and \( y \) indicate the spatial position in a given subband; and \( l \) and \( k \) indicate the scale and orientation of \( D_i \), respectively.

The coefficients combining should integrate the visual information contained in all source images into the fused image without introduction of distortion or loss of information. However, this goal is almost impossible [2]. A more practical way is to integrate the faithful representation of the most important input information into the fused image. A general and effective image fusion rule for the multi-resolution-based methods is adopted in this paper. The low-frequency coefficients are fused by the average method, meaning the fused coefficient is the average of the corresponding coefficients of the source images. The high-frequency coefficients are fused by the approach of choosing absolute maximum. The technique of choosing absolute maximum can be formulated as:

\[
D_f(p) = \begin{cases} 
D_{i1}(p) & A_{i1}(p) > A_{i2}(p) \\
D_{i2}(p) & \text{otherwise}
\end{cases}
\]  

Based on the general rule, many new fusion rules, such as selection method [34], entropy method [17], and linear dependency method [28], have been developed. As indicated in the references, these new methods can improve fusion performance in some respects. For example, in Ref. [34], Zhang and Guo proposed a fusion rule to obtain contrast improvements for multi-focus image fusion. However, if all combinations of the multi-resolution transforms and the extend fusion rules are investigated, it is hard to obtain useful conclusions. We think that the conclusions of the general rules are also effective for more powerful rules, which will be verified by an experiment in Section 4.

There are two other issues, grouping method and consistency verification in the fusion process. Grouping method means that,
when determining the fused multiscale coefficients, a set of coefficients in other orientations and frequency subbands corresponding to the same pixel should be considered jointly. Consistency verification is based on the idea that a composite multiscale coefficient is unlikely to be generated in a completely different manner from all its neighbors. These two issues have been investigated in the literature [7,35], so they are not considered in this paper. In general, the grouping method and consistency verification can improve fusion result.

4. Experimental results and discussion

In this paper, we perform the experiments over 20 pairs of source images, as shown in Fig. 6. These images consist of three types of images, namely multi-focus images (eight pairs shown in Fig. 6a), infrared–visible images (eight pairs shown in Fig. 6b), and medical images (four pairs shown in Fig. 6c). Image registration, which geometrically aligns the source images, should be performed before image fusion. In our study, the source images are assumed to have been registered.

4.1. Evaluation criteria

Because the ideal fused image is often associated with specific tasks, universal quality evaluation of the fused image is difficult. Usually, image fusion results can be evaluated in subjective and objective means. Subjective methods are difficult to perform since they are based on psycho-visual testing and are expensive in terms of time, effort, and equipment required. Moreover, in most cases, there is little difference among fusion results. Therefore, subjective means are difficult to correctly evaluate the fusion results. For these reasons, many objective evaluation methods have been developed; however, so far, there is no universally accepted metric to objectively evaluate the image fusion results. In this paper, we quantitatively evaluate the fusion performance using five metrics, i.e., mutual information (MI) [36], QBF [37], QAB, QW, and Q3 [38,39], which have been proved to be effective to a great degree.

1. The mutual information [36] \( I_{AF} \) between the source image \( A \) and the fused image \( F \) is defined as follows:

\[
I_{AF} = \sum_{a,f} p_{AF}(a,f) \log \frac{p_{AF}(a,f)}{p_A(a)p_F(f)}
\]

where \( p_{AF} \) is the joint normalized histogram of \( A \) and \( F \), \( p_A \) and \( p_F \) are the normalized histogram of \( A \) and \( F \), and \( a \) and \( f \) represent the pixel value of the image \( A \) and \( F \), respectively. The mutual information \( I_{AF} \) between the source image \( A \) and the fused image \( F \) is similar to \( I_{AF} \). The mutual information between the source images \( A, B \), and the fused image \( F \) is the sum of \( I_{AF} \) and \( I_{BF} \), i.e.

\[
M_F = I_{AF} + I_{BF}
\]

2. The metric \( Q^{AF/B} [37] \) is defined as follows:

\[
Q^{AF/B} = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} \left( Q^{AF}(n,m)w_A(n,m) + Q^{BF}(n,m)w_B(n,m) \right)}{\sum_{n=1}^{N} \sum_{m=1}^{M} \left( w_A(n,m) + w_B(n,m) \right)}
\]

where \( Q^{AF}(n,m) = Q_{AF}^a(n,m)Q_{AF}^b(n,m) \); \( Q_{AF}^a(n,m) \) and \( Q_{AF}^b(n,m) \) are the edge strength and orientation preservation values, respectively; \( n, m \) represent the image location; and \( N, M \) are the size of images, respectively. \( Q^{AF}(n,m) \) is similar to \( Q^{BF}(n,m) \), \( w_A(n,m) \) and \( w_B(n,m) \) reflect the importance of \( Q^{AF}(n,m) \) and \( Q^{BF}(n,m) \), respectively. The dynamic range of \( Q^{AF/B} \) is \([0,1]\), and it should be as close to 1 as possible.

3. The metric \( Q_0 [38] \) between the source image \( A \) and the fused image \( F \) is defined as follows:

\[
Q_0(A,F) = \frac{2\sigma_{ag}}{\sigma_a^2 + \sigma_f^2} + 2\frac{\sigma_t}{\sigma_a^2 + \sigma_f^2 + \sigma_t^2}
\]

where \( \sigma_{ag} \) represents the covariance between \( A \) and \( F \); \( \sigma_a, \sigma_f \) denote the standard deviation of \( A \) and \( F \); and \( \sigma_t \) represents the mean value of \( A \) and \( F \), respectively. \( Q_0(A,B,F) \) is the average between \( Q_0(A,F) \) and \( Q_0(B,F) \), i.e.,

\[
Q_0(A,B,F) = (Q_0(A,F) + Q_0(B,F))/2
\]

Note that \(-1 \leq Q_0 \leq 1\), and it should be also as close to 1 as possible.

4. The metric \( Q_{wv} [39] \) among images \( A, B, \) and \( F \) is defined as follows:

\[
Q_{wv}(A,B,F) = \sum_{w} c(w)\langle i(w)Q_0(A,F|w) + (1 - i(w))Q_0(B,F|w) \rangle
\]

where \( i(w) \) represents the relative salience of \( A \) compared to \( B \) in the same window \( w \), and \( c(w) \) denotes the normalized salience of the window \( w \). More details about the metric \( Q_{wv} \) are available in Ref. [39].
5. The metric $Q_E$ [39] is defined as follows:

$$Q_E(A, B, F) = Q_W(A, B, F) \cdot Q_W(A', B', F)^a$$

where $A'$, $B'$, $F'$ are the corresponding edge images of $A$, $B$, $F$, respectively. Parameter $a$ which is set to 1 in this paper reflects the contribution of the edge images compared to the original images.

These five metrics given above evaluate the amount of information transferred from source images into the fused image, but there exist differences among them. Mutual information reflects the statistical dependence of two random variables from information theory viewpoint. Especially for image fusion, it measures the similarity of image intensity distribution of the corresponding image pair. The metric $Q_{AB/F}$ evaluates the amount of edge information transferred from source images into fused image. The metrics $Q_0$, $Q_W$, and $Q_E$ integrate characteristics of the human visual system. The metric $Q_0$ evaluates the degree of distortion of the fused image. It combines three factors of image distortion related to the human visual system, i.e., loss of correlation, luminance distortion, and contrast distortion. The metric $Q_W$ further takes the salience of information into account. The metric $Q_E$ contains visual information and edge information, simultaneously. In addition, the larger value for the above metrics means the better fusion result.

4.2. Quantitative analysis

We compare different numbers of decomposition levels and filters for each multi-resolution transform with the generic image fusion framework in Fig. 5. For each multi-resolution transform, the number of decomposition levels from one to four is considered. When the number of decomposition levels is too large, one coefficient in coarse resolutions responds to a large group of pixels in the fused image. Therefore, an error in coarse resolutions has a great effect on the final fused image, producing artificial distortions. Fig. 7 shows the fusion results obtained using the wavelet basis ‘bior1.3’ with the number of decomposition levels from one to seven in the fusion scheme of wavelet. It can be found that the block effect occurs when the number of decomposition levels is larger than four. One region showing the block effect is labeled with a red rectangle. In addition, large decomposition levels require that the size of image is also large; therefore, the number of decomposition levels that is larger than four is not considered in the following experiment.

4.2.1. Fusing with the DWT

To compare the performance of the DWT using different numbers of decomposition levels and wavelet bases, we consider four wavelet families: Daubechies (db$N$, $N = 1, 3, 6, 10, 13$), Symlets (sym$N$, $N = 3, 6, 10, 13$), Coiflets (coif$N$, $N = 1–5$), Biorthogonal (bi$or[N, N]$, $N, N = 1.3, 2.2, 3.5, 4.4, 6.8$), and Reverse Biorthogonal (rbio[N, N], $N, N = 1.3, 2.2, 3.5, 4.4, 6.8$). For each wavelet basis, the number of decomposition levels from one to four is considered. The best results, labeled in bold, obtained by the DWT for each metric are given in Table 1. All the reported values are the average results of all images of the same type. On the whole, when the mean of a fusion result is large, its corresponding standard deviation is relatively small. This table illustrates that the biorthogonal and Daubechies filters show good performance. Moreover, it can be seen that the short filter usually provides better performance than the long filter.
In addition, large numbers of decomposition levels show good fusion result for all the metrics except for the metric $MI$. It can be concluded that three or four levels of decomposition is a good setting. This is because that the metrics $Q_{AB}$, $Q_{W}$, $Q_{E}$, and $Q_{0}$, mainly measure the amount of salient information transferred from source images into the fused image, which makes larger decomposition levels show better performance for these metrics. However, one level of decomposition shows the best performance for the metric $MI$. This is because that, when one level of decomposition is employed, low-frequency components combined by average contain the main information in the source images. So the fused image contains more average information from the source images, which makes that the metric $MI$ value are larger for lower levels of decomposition.

We determine the overall best results for each image type by comprehensive consideration of all five evaluation metrics instead of one or several metrics. For multi-focus images, there are always three metrics for the ‘bior6.8’ filter with four levels, which is superior to those of other cases. Therefore, on the whole, the ‘bior6.8’ filter with four levels of decomposition shows the best fusion results for multi-focus images. And the ‘db1’ filter with three levels provides the best fusion results for infrared–visible images and medical images. The ‘db1’ wavelet shows good fusion results for all three types of images. This is due to the fact that

<table>
<thead>
<tr>
<th>Image</th>
<th>Filters</th>
<th>Levels</th>
<th>$MI$ (mean/std)</th>
<th>$Q_{AB}$ (mean/std)</th>
<th>$Q_{W}$ (mean/std)</th>
<th>$Q_{E}$ (mean/std)</th>
<th>$Q_{0}$ (mean/std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-focus</td>
<td>rbio1.3</td>
<td>1</td>
<td>2.4496/0.2420</td>
<td>0.6323/0.0965</td>
<td>0.7751/0.1119</td>
<td>0.5902/0.1709</td>
<td>0.7371/0.1207</td>
</tr>
<tr>
<td></td>
<td>bior6.8</td>
<td>4</td>
<td>2.4126/0.2471</td>
<td>0.6866/0.0691</td>
<td>0.8376/0.0843</td>
<td>0.7130/0.1358</td>
<td>0.7206/0.1095</td>
</tr>
<tr>
<td></td>
<td>db13</td>
<td>4</td>
<td>2.3961/0.2429</td>
<td>0.6798/0.0679</td>
<td>0.846/0.0606</td>
<td>0.7233/0.1222</td>
<td>0.7190/0.1109</td>
</tr>
<tr>
<td></td>
<td>db1</td>
<td>3</td>
<td>2.4218/0.2369</td>
<td>0.6511/0.0846</td>
<td>0.8323/0.0966</td>
<td>0.6615/0.1551</td>
<td>0.7769/0.1071</td>
</tr>
<tr>
<td>Infrared–visible</td>
<td>rbio1.3</td>
<td>1</td>
<td>2.0581/0.1982</td>
<td>0.4301/0.0828</td>
<td>0.6288/0.1075</td>
<td>0.4222/0.1334</td>
<td>0.5586/0.0946</td>
</tr>
<tr>
<td></td>
<td>bior2.2</td>
<td>4</td>
<td>2.0182/0.1966</td>
<td>0.5062/0.1362</td>
<td>0.6841/0.0917</td>
<td>0.4868/0.1162</td>
<td>0.5741/0.0605</td>
</tr>
<tr>
<td></td>
<td>db1</td>
<td>3</td>
<td>2.0274/0.2007</td>
<td>0.4885/0.1139</td>
<td>0.7217/0.0831</td>
<td>0.4939/0.1164</td>
<td>0.6214/0.0650</td>
</tr>
<tr>
<td>Medical</td>
<td>bior4.4</td>
<td>1</td>
<td>2.5933/0.0775</td>
<td>0.3912/0.0902</td>
<td>0.4588/0.0288</td>
<td>0.1816/0.0130</td>
<td>0.2469/0.1840</td>
</tr>
<tr>
<td></td>
<td>db1</td>
<td>3</td>
<td>2.4492/0.1202</td>
<td>0.4657/0.0406</td>
<td>0.5473/0.0229</td>
<td>0.2226/0.0303</td>
<td>0.2793/0.0928</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image</th>
<th>Filters</th>
<th>Levels</th>
<th>$MI$ (mean/std)</th>
<th>$Q_{AB}$ (mean/std)</th>
<th>$Q_{W}$ (mean/std)</th>
<th>$Q_{E}$ (mean/std)</th>
<th>$Q_{0}$ (mean/std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-focus</td>
<td>coif1</td>
<td>3</td>
<td>2.4677/0.2520</td>
<td>0.7088/0.0609</td>
<td>0.8549/0.0607</td>
<td>0.7384/0.1072</td>
<td>0.7639/0.0995</td>
</tr>
<tr>
<td></td>
<td>bior2.2</td>
<td>4</td>
<td>2.4510/0.2429</td>
<td>0.7140/0.0575</td>
<td>0.8596/0.0634</td>
<td>0.7443/0.1158</td>
<td>0.7555/0.1004</td>
</tr>
<tr>
<td></td>
<td>db1</td>
<td>3</td>
<td>2.4585/0.2437</td>
<td>0.7113/0.0690</td>
<td>0.8468/0.0636</td>
<td>0.7147/0.1128</td>
<td>0.7725/0.0968</td>
</tr>
<tr>
<td>Infrared–visible</td>
<td>bior2.2</td>
<td>1</td>
<td>2.0623/0.2012</td>
<td>0.4682/0.0988</td>
<td>0.6343/0.0987</td>
<td>0.4265/0.1268</td>
<td>0.5620/0.0879</td>
</tr>
<tr>
<td></td>
<td>db1</td>
<td>4</td>
<td>2.0314/0.1994</td>
<td>0.5354/0.1352</td>
<td>0.7026/0.0854</td>
<td>0.5012/0.1300</td>
<td>0.6009/0.0544</td>
</tr>
<tr>
<td></td>
<td>bior2.2</td>
<td>4</td>
<td>2.0293/0.1975</td>
<td>0.5330/0.1341</td>
<td>0.7180/0.0756</td>
<td>0.5207/0.1189</td>
<td>0.5994/0.0586</td>
</tr>
<tr>
<td>Medical</td>
<td>rbio1.3</td>
<td>1</td>
<td>2.6319/0.0836</td>
<td>0.4609/0.0737</td>
<td>0.4915/0.0413</td>
<td>0.2184/0.0277</td>
<td>0.2637/0.1839</td>
</tr>
<tr>
<td></td>
<td>coif1</td>
<td>4</td>
<td>2.3785/0.1062</td>
<td>0.5970/0.0359</td>
<td>0.5810/0.0321</td>
<td>0.3123/0.0399</td>
<td>0.2918/0.1376</td>
</tr>
<tr>
<td></td>
<td>coif1</td>
<td>2</td>
<td>2.5028/0.0889</td>
<td>0.5420/0.0311</td>
<td>0.5193/0.0137</td>
<td>0.2710/0.0191</td>
<td>0.3005/0.2308</td>
</tr>
</tbody>
</table>
the ‘db1’ is high contrast basis. These results will be listed in Table 7 for global comparison.

4.2.2. Fusing with the SWT

To compare the performance of the SWT, we consider the same wavelet families and number of decomposition levels as the DWT. Table 2 gives the best results of each metric for the SWT. From this table, it can be seen that standard deviations are close for the same metric, which indicates that the stability is similar for different filters and numbers of decomposition levels. The ‘bior2.2’ filter with four levels of decomposition provides the best result for multi-focus images and infrared–visible images. The ‘coif1’ filter with four levels of decomposition obtains the best fusion results for medical images. Similar to the DWT, the biorthogonal and Daubechies filters with large decomposition levels show good fusion results. In addition, the ‘coif1’ filter provides nice fusion results especially for medical images. This is because the ‘coif1’ is a short filter.

4.2.3. Fusing with the DTCWT

The DTCWT employs a dual-tree of real wavelet filters to generate the real and imaginary parts of transform coefficients [18]. Note that the filter of the first-level decomposition is different from that of other-levels of decomposition. We compare four categories of first-level filters: Antonini 9-7, LeGall 5-3, near symmetric 5-7, and near symmetric 13-19 tap filters, respectively. These names of filters are abbreviated as ‘9-7’, ‘5-3’, ‘5-7’, and ‘13-19’, respectively. Four categories of quarter sample shift orthogonal filters are used for other-levels of decomposition, which are 10-10, 14-14, 16-16, and 18-18 tap filters, respectively. Two 10-10 tap filters are used, and the difference between them is that one has only 6-6 nonzero taps while the other has 10-10 nonzero taps. All five names of filters are abbreviated to ‘q_6’, ‘q_a’, ‘q_b’, ‘q_c’, and ‘q_d’, respectively. All the possible groups of the two filters families are compared in this paper.

Table 3 lists the best results in terms of each metric for the DTCWT. From this table, it can be seen that the filters ‘5-3’ and ‘5-7’ for the first-level and the filter ‘q_6’ for the other-levels show good fusion performance. We know that these filters are short. The filter ‘5-7’ and ‘q_6’ with four levels of decomposition provides the best results for multi-focus images. The filter ‘5-3’ and ‘q_6’ with four levels of decomposition shows the best fusion performance for infrared–visible images. The filter ‘5–3’ and ‘q_a’ with three levels of decomposition gives the best fusion results for medical images. Similar to the DWT, it can be found that standard deviations are close for the same metric from Table 3, which indicates that the stability is similar for different filters and numbers of decomposition levels.

4.2.4. Fusing with the CVT

Table 4 lists the best fusion results of each metric by the CVT. These results are obtained with 4, 3, 3 levels of decomposition for multi-focus images, infrared–visible images, and medical images, respectively. However, similar to the DWT, one level of decomposition shows the best performance for the metric MI. Moreover, most of the standard deviations are small when the fusion results are good in terms of mean. Since line is an important feature of image that can be captured by the curvelet effectively, the fusion performance of the curvelet is stable, making the standard deviation small.

4.2.5. Fusing with the CT

The implementation of the CT is based on pyramid filtering and orientation filtering as discussed in Section 2. We compare four categories of pyramid filters, i.e., ‘9–7’, ‘5–3’, ‘Burt’, and ‘adder’. And four classes of orientation filters, i.e., ‘haar’, ‘5–3’, ‘9–7’, ‘adder’, are compared. We investigate all groups of two-sort filters in this paper. The number of orientation of the DWT is only three, while that of the CT can be set to the integer exponent of 2. The decomposition orientations are set to [4], [4, 8], (4, 8), and (4, 8, 8, 16), respectively, for four decomposition levels in this paper. For example [4, 8, 8] means that the source images are decomposed into three levels, and the number of orientation from coarse to fine resolution are 4, 8, 8, respectively.

For each metric, the best fusion results obtained by the CT-based method are listed in Table 5. The pyramid filter ‘9–7’ and

<table>
<thead>
<tr>
<th>Table 3</th>
<th>The best results for the DTCWT.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td>First</td>
</tr>
<tr>
<td>Multi-focus</td>
<td>5-3</td>
</tr>
<tr>
<td></td>
<td>5-7</td>
</tr>
<tr>
<td></td>
<td>5-7</td>
</tr>
<tr>
<td></td>
<td>5-7</td>
</tr>
<tr>
<td>Infrared–visible</td>
<td>5-3</td>
</tr>
<tr>
<td></td>
<td>5-7</td>
</tr>
<tr>
<td></td>
<td>5-3</td>
</tr>
<tr>
<td>Medical</td>
<td>5-3</td>
</tr>
<tr>
<td></td>
<td>5-3</td>
</tr>
<tr>
<td></td>
<td>5-3</td>
</tr>
<tr>
<td></td>
<td>5-3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>The best results for the CVT.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td>Levels</td>
</tr>
<tr>
<td>Multi-focus</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Infrared–visible</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Medical</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>
the orientation filter ‘9-7’ with {4, 8, 16} levels of decomposition perform the best for multi-focus images. The pyramid filter ‘5-3’ and the orientation filter ‘9-7’ with {4, 8, 16} levels of decomposition obtain the best fusion results for infrared–visible images. The pyramid filter ‘ladder’ and the orientation filter ‘haar’ with {4} levels of decomposition provide the best fusion performance for medical images. These results are listed in Table 7 for global comparison.

### 4.2.6. Fusing with the NSCT

The implementation of the NSCT is also based on pyramid filtering and the orientation filtering. Different from the CT, these filters are upsampled in each scale. Four categories of pyramid filters, i.e., ‘9-7’, ‘maxflat’, ‘pyr’, and ‘pyrexc’ are compared, while twelve categories of orientation filters, i.e., ‘haar’, ‘McClellan’, ‘17-17’, ‘9-9’, ‘dvmmlp’, ‘7-9’, ‘ladder’, ‘ideal’, ‘dmaxflat7’, ‘dmaxflat6’, ‘dmaxflat5’, and ‘dmaxflat4’, are compared. We investigate all forty-eight groups of two-sort filters in this paper. For each filter grouping, the setting of decomposition levels is the same as the CT.

For each metric, the best fusion results obtained by NSCT-based methods are listed in Table 6. For pyramid filters, we can see that ‘pyrexc’ performs the best. For orientation filters, ‘7-9’ provides good fusion results for the multi-focus images, and ‘ideal’, ‘vk’, ‘dmaxflat4’, and ‘dmaxflat6’ give better fusion results than the other filters for the infrared–visible images and medical images. From this table, it can be seen that the standard deviations are close for the same metric and type of image, which means that the stability among these results is similar. The pyramid filter ‘9-7’ and the orientation filter ‘7-9’ with {4, 8, 16} levels of decomposition perform the best for multi-focus images. The pyramid filter ‘pyrexc’ and the orientation filter ‘7-9’ with {4, 8, 16} levels of decomposition do the best for infrared–visible images. The pyramid filter ‘pyrexc’ and the orientation filter ‘vk’ with {4, 8, 16} levels of decomposition provide the best fusion results for medical images.

Comparing Tables 5 and 6, it can be seen that the variation of standard deviation of the CT is larger than that of the NSCT among the same metric and type of image. This is because the CT is shift-invariant, which causes large error in the image reconstruction process due to the upsampling operator.

### 4.2.7. Global comparison

Table 7 lists the overall best results of each type of transform. The standard deviation is not shown in this table due to lack of space. From this table, it can be found that the SWT generally shows better performance than the DWT. This is because the SWT is shift-invariant. Moreover, we can also observe that short filter performs usually better than long filter. For example, the filters ‘db1’, ‘bior2.2’, ‘coif1’, ‘5-7’, and ‘q_6’ obtain the best results. This is because pixel-level image fusion mainly concerns the fusion of local information in the source images, making the short filter to have more advantages. In addition, short filter consumes less time than long filter.

When comparing evaluation metrics, the values for multi-focus images are usually larger than those of infrared–visible images and medical images. This is because the nature of these three image types is different. The source images of multi-focus are generated by the same imaging mechanism. The infrared–visible and medical multi-modal images are generated by different imaging mechanisms. Therefore, the similarity of multi-focus images is larger than that of the other two image types, which makes that the evaluation metrics of multi-focus images have larger values.

Table 7 shows that the DTCWT obtains better results than the DWT. This is because the DTCWT is nearly shift-invariant and...
can capture more orientation information than the DWT. When compared with the SWT, DTCWT performs better for multi-focus images but worse for infrared–visible images and medical images. The CVT shows better performance than the DWT for multi-focus images but worse than the DWT for infrared–visible images and medical images, because the CVT can better capture edge and line features than the DWT, making the CVT better for multi-focus images. However, infrared–visible images and medical images belong to multi-modality images. This causes the difference among the source images to be large, consequently making the difference of edges and lines large as well. If different edges and lines exist in the same location, aliasing occurs in the fusion results. Therefore, the CVT shows worse performance for multi-modality images. Moreover, from Table 7, we can find that the CVT provides worse results than the SWT and the DTCWT. This is because the CVT is shift-variant.

In Table 7, the SWT, DTCWT, and NSCT, which are shift-invariant, perform better than the DWT, CVT, and CT. Therefore, it can be concluded that the performance of image fusion is greatly affected by the shift-variance resulting from the subsampling operator, which is not only for the misregistered images but also for strictly registered source images. This is because some important information is lost in the subsampling process. For shift-variant transform, the error will be enlarged in the image reconstruction process due to the upsampling operator. The complexity and memory requirement, which depend mainly on the methods of image decomposition, should also be considered. The time and memory needed for shift-invariant transforms are much larger. For example, the SWT-based method uses roughly $3 \times$ levels times more memory than that of the DWT-based method. The complexity and memory requirement of the NSCT is much larger because of multi-direction; however, with the fast development of hardware, it will not be a big problem in the future.

For multi-focus images, both the DTCWT and the NSCT present good results due to their properties of shift-invariance and multi-direction. For infrared–visible images, the NSCT obtains the best results except for the metric $Q_0$, but it is merely slightly worse than the DWT. Therefore, as a whole, the NSCT shows the best performance for infrared–visible images. For medical images, the NSCT obtains the best results except for the metric $MI$. So the NSCT is the best method of multi-resolution image decomposition among all that have been investigated above. This is mainly because the NSCT, which has better frequency selectivity and regularity, is a flexible multiscale, multi-directional, and shift-invariant image decomposition method. Due to lack of space, only three representative best fusion results are given in Fig. 8. These results show that the important features of the source images have been transformed into the fused images. For instance, Fig. 8a is in focus everywhere.

In addition, the number of multi-resolution decomposition level is also important for the generic image fusion framework. From Table 7, it can be seen that most of the best results are obtained with four levels of decomposition. It is a trade-off between the capability of catching spatial details and the sensitivity to noise and transform artifacts.

Moreover, we think that the optimal setting of filter and decomposition level obtained by the general fusion rule can achieve good fusion results for other fusion rules. To verify this viewpoint, we perform some experiments with the NSCT using the selection

![Fig. 8. Three representative best fusion results of the NSCT-based method. (a) Fused multi-focus image. (b) Fused infrared–visible image. (c) Fused medical image.](image-url)
fusion rule [34] on all the multi-focus source images (eight pairs) used in this paper. The experimental results are listed in Table 8. All the reported values are the average results of all source images. The second row lists the best result in terms of each metric for the selection rule. The third row lists the fusion results of the fusion rule [34] using the setting suggested in this paper. It can be easily found that there is only slight improvement. This indicates that the conclusions for the general rule are effective for the selection rule.

5. Conclusions

In this paper, we compare the image fusion performance of six multi-resolution transforms with different filters and different numbers of decomposition levels. For each multi-resolution transform, the optimal settings are presented for multi-focus images, infrared-visible images, and medical images, respectively. Then, these optimal settings are compared against each other globally. The experimental results indicate that the appropriate setting for the number of decomposition levels is four. It is a trade-off between the capability of catching spatial details and the sensitivity to noise and transform artifacts. When the number of decomposition levels is too large, one coefficient in coarse resolutions responds to a large group of pixels of fused image. Therefore, an error in coarse resolutions has a great effect on final fused image. Some errors inevitably occur in the process of fusion, producing some artificial distortion. Large decomposition levels give rise to fusion methods that are sensitive to noise. Moreover, large decomposition levels consume more time and have higher memory requirements. When the number of decomposition levels is too small, spatial details cannot be captured well.

In addition, the experimental results indicate that the NSCT performs usually the best, followed by the DTCWT and the SWT, which provide similar results. Four out of all five metrics of the NSCT for infrared-visible images and medical images are the best, while three out of all five metrics of the NSCT for multi-focus images are the best. Therefore, the advantages of the NSCT for multi-modality images are more evident than that for single-modality images. The CVT performs better than the DWT for multi-focus images, while the DWT presents better results than CVT for the infrared-visible images and medical images; therefore, the fusion performance of multi-resolution methods is affected by different types of images. The CT is the worst one. Moreover, the property of shift-invariance is important for image fusion, not only for mis-registered images but also for strictly registered source images. In addition, short filter usually provides better fusion results than long filter. This is because long filters may smooth details and generate diffusion effect.

Acknowledgements

The authors would like to thank the editor and anonymous reviewers for their detailed review, valuable comments, and constructive suggestions. The authors would also like to thank Professor Arthur Ardeshir Goshhtasby (Wright State University) for his insightful suggestions and constructive comments on this work. This paper is supported by the National Natural Science Foundation of China (No. 60871096 and 60835004), the Ph.D. Programs Foundation of Ministry of Education of China (No. 200805320006), the Key Project of Chinese Ministry of Education (2009-120), and the Open Projects Program of National Laboratory of Pattern Recognition.

References


