A Novel Multiple-Stage Multi-Focus Fusion Algorithm

ChenShen, Zhongliang JING*, HanPan
Institute of Aerospace Science and Technology, Shanghai Jiao Tong University
No. 800, Dong chuan Road, Min Hang District, Shanghai 200240, China
bluewind0815@sjtu.edu.cn  zljing@sjtu.edu.cn

Abstract—Image fusion is defined as the process of combining multiple source images into a single image retaining the significant details or features of each source image. Although existing multi-focus algorithms complete some improvement on the fusion performance, using each method alone may bring about geometric distortion which will be discussed in the following section. To solve this problem, this paper presents a novel algorithm that can effectively eliminate these effects caused by Discrete Wavelet Transform (DWT) and Spatial Frequency (SF). It comprises a multiple-stage fusion and a new fusion scheme on the low frequency of the DWT's coefficients. The fusion schemes to be proposed simultaneously calculates two intermediate images, and then merge the respectively pre-fused images into a new image. Objective measures and visual assessment on the extensive experiments show that the proposed approach outperforms some well-known methods.

Key words: Image Fusion, Discrete Wavelet Transform, Spatial Frequency, Multi-Focus

I. INTRODUCTION

Image fusion combines different aspects of information from multiple source images. Multi-focus image fusion is a process of combining two or more partially defocused images into a new image with all interested details of objects well displayed. To obtain an image with each object in focus, a multi-focus image fusion process is required to fuse the images taken under different focal length. There are three different fusion categories: pixel level, feature level and decision level [1]. Each of the levels depends on the stage where the fusion algorithm takes place. The multi-resolution fusion uses multi-scale transforms to analyze the information content of images for diverse purposes. These multi-scale fusion methods firstly calculate the multi-resolution decomposition coefficients on source images and then combine the coefficients with a composite representation and finally reconstruct fused image using inverse multi-resolution transform. These multi-scale transforms involve Laplacian pyramid [2-3], gradient pyramid [4], RoLP pyramid [5] and the Discrete Wavelet Transform (DWT) [6-8]. The fusion techniques based on image blocks aim to select sharper blocks using sharpness criterion such as Spatial Frequency (SF) [9-10] and local gradient.

Recently, many pixel based multi-focus image fusion algorithms have been presented. Some improved methods based on wavelet transform such as Dual Tree Complex Wavelet Transform (DT-CWT)[11] and Low Redundancy Discrete Wavelet Frame (LRDWF)[12] are shift-invariant but inevitably bring about distortion effects. Some recently presented papers use block-based fusion scheme, these algorithms aim to select sharper image blocks and the block size has been optimized using different strategies such as Genetic Algorithm (GA)[13] and Pulse Coupled Neural Network (PCNN)[14] to obtain the best fusion result. But the fused images using block-based methods are with bad continuity and produce jaggies and large block borders. For the algorithms concerning the spatial frequency or wavelet transform, a multi-focus image fusion based on spatial frequency and morphological operators was proposed in [15], a multi-focus image fusion using region segmentation and spatial frequency was presented in [16], a multi-focus image fusion based on stationary wavelet transform and extended spatial frequency measurement was given in [17], etc. However, these methods are essentially attached to either multi-resolution analysis or block-based methods.

The DWT fusion shows better local features in frequency domain, but traditional algorithms exhibit bad distortion effects along the edges and borders in the image. Since the blurred region has a relatively low SF value, the block-based method could easily discriminate the focused objects and the blurred ones.

As the fused image using SF rule inherits a lot from the region in focus and as well gains undesired block effect along the edge of the blurred region, a combination of DWT fusion and SF fusion is proposed in this paper to overcome the above-mentioned defects to some extent. The new approach uses a multiple fusion strategy in which a novel fusion algorithm of DWT low frequency using local adaptive weight applies. The original approach in the DWT low frequency subband aims to separate columns and rows of a pixel’s adjacent field block and calculate the statistical properties to obtain the final weight of the pixel. As the fusion scheme includes two phases of fusion, different fusion rules of the DWT low frequency subband are performed to coordinate with the specific phases, and the new low frequency fusion of DWT will be applied in the second phase. This paper presents two fusion paths in which the detailed algorithms will be illustrated in Section IV.
This paper firstly presents the basic concept of the DWT and SF and then outlines the proposed fusion scheme. In the later sections results of the risen scheme are presented with comparison of performance based on a variety of evaluation criteria.

II. WAVELET TRANSFORM AND SPATIAL FREQUENCY

The general procedure of wavelet transform is shown in Figure 1. Given an \( M \times N \) input image \( I \) we generate two \( (M/2) \times N \) images, \( I_L \) and \( I_H \), by separately filtering and down-sampling the rows in \( I \) using a low-pass filter \( L \) and a high-pass filter \( H \). We repeat the process by filtering and down-sampling the columns in \( I_L \) and \( I_H \) using the filters \( L \) and \( H \). The output are four \( (M/2) \times (N/2) \) images \( I_{LL} \), \( I_{HL} \), \( I_{IH} \) and \( I_{HH} \), where \( I_{LL} \) is a low-frequency approximation of \( I \), and \( I_{HL} \), \( I_{IH} \) and \( I_{HH} \) are high-frequency detail images which represent vertical (V), horizontal (H) and diagonal (D) structures in \( I \) \( [8][18] \).

![Figure 1](image1.png)

Figure 1. It shows an \( M \times N \) input image \( I \) decomposed into three \( (M/2) \times (N/2) \) detail images \( I_{LL}, I_{HL} \) and \( I_{HH} \) and one \( (M/4) \times (N/4) \) approximation image \( I_{LL} \). The image \( I_{LL} \) is further decomposed into three \( (M/4) \times (N/4) \) approximation image \( F_{LL} \) and \( I_{HL}, I_{IH} \) and \( I_{HH} \) are detail images which represent vertical (V), horizontal (H) and diagonal (D) structures in \( I \).

The Wavelet Transform is firstly performed on each source images, and then a fusion decision map is generated based two fusion rules. The fused wavelet coefficients can be constructed through the wavelet coefficients of the source images according to the fusion path. Finally the fused image is obtained by performing the Inverse Discrete Wavelet Transform (IDWT).

The Spatial Frequency measures the overall activity level in an image \( [10][16] \). For an \( M \times N \) image \( F \), with the gray value at pixel position \( (m,n) \) denoted by \( F(m,n) \), its spatial frequency is defined as

\[
SF = \sqrt{RF^2 + CF^2},
\]

where \( RF \) and \( CF \) are the row frequency

\[
RF = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} (F(m,n) - F(m,n-1))^2,
\]

and column frequency

\[
CF = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} (F(m,n) - F(m-1,n))^2,
\]

respectively. The result of the spatial frequency will be derived via block-based computation.

III. THE ASSESSMENT OF THE BLOCK-BASED AND WAVELET-BASED MULTI-FOCUS FUSION ALGORITHM

According to the reconstruction property of the wavelet transform, the final fused image keeps a nice continuity after multi-scale reconstruction in despite of ringing and distortion artifacts. These artifacts are generated by down-sampling, redundant coefficients, inaccurate selection of coefficients and other process in wavelet-based analysis. Although some recently presented multi-scale transforms are shift-invariant and with low redundancy, the distortion artifacts cannot be perfectly eliminated. Since the multi-focus source images usually don’t have obvious bounds between clear region and blurred region, the block-based fusion methods may bring about jaggies, large mutations and blocks in the fused image. However, the block-based selection keeps the highest fidelity from the clear region. Recently presented papers merely use one of the two strategies, then the above-mentioned drawbacks may not achieve a better solution. Figure 2 clearly shows these geometric distortion and blocking artifacts.

![Figure 2](image2.png)

Figure 2. (a) Fused image using DWT (b) Fused image using SF (c) Difference image between (a) and source image (d) Difference image between (b) and source image (e) Grid pattern of (c) (f) Grid pattern of (d).

From the Figure 2 above we can obtain some intuitive perception. The wavelet-based fusion does well in visual perception, but it is with severe geometric distortion and low fidelity. The fused image using block-based method has some very flat regions in difference image, and these regions are the perfect duplicates from the clear regions of source image. However, the blocking effects are unfavorable to visual assessment.
IV. THE PROPOSED FUSION SCHEME

In this section, the details of fusion rule are discussed. The fusion algorithm is decided according to the goal of fusion. There are two proposed rules using a combination of DWT and SF. In order to simplify the process of the fusion rule using two-dimensional DWT, an assumption that there are only two multi-focus source images is supposed in the fusion scheme. All of the fusion rules described below will be easily extended in case of three source images or more.

First of all, $A$ and $B$ are assumed as source images focused on different regions relatively compared with the reference image. The fusion scheme is divided into a two-stage process including DWT fusion and SF fusion.

On the first stage, the DWT fusion and SF fusion is simultaneously executed and two fused images will be acquired respectively as $F_{DWT}(x,y)$ and $F_{SF}(x,y)$. The fused image $F_{DWT}(x,y)$ is derived by dealing with the decomposed wavelet coefficients and $F_{SF}(x,y)$ is derived by directly dealing with image blocks.

In the first DWT fusion, the low frequency coefficients are fused using local gradient-based weighted average scheme. Let $f^A(x,y)$ and $f^B(x,y)$ be the DWT low frequency coefficients of image $A$ and $B$. Similarly, let $g^A(x,y)$ and $g^B(x,y)$ be the high frequency coefficients of image $A$ and $B$. Consider an $M \times N$ window ($5 \times 5$ in this paper) centered at location $(x,y)$, calculate its average gradient which can be described as

$$AG = \frac{1}{M \times N} \sum_{i=-(M-5)/2}^{(M-5)/2} \sum_{j=-(N-5)/2}^{(N-5)/2} \sqrt{\Delta g^A(x+i,y+j)^2 + \Delta g^B(x+i,y+j)^2},$$

where $\Delta g^A(x+i,y+j) \text{ and } \Delta g^B(x+i,y+j)$ denote the first order difference at location $(x+i,y+j)$ along x-axis and y-axis direction respectively. Assuming $AG^A$ and $AG^B$ represent the average gradient of $f^A(x,y)$ and $f^B(x,y)$ at location $(x,y)$, then the fused low frequency sub-image can be written as

$$f^F(x,y) = \frac{AG^A}{AG^A + AG^B} \cdot f^A(x,y) + \frac{AG^B}{AG^A + AG^B} \cdot f^B(x,y),$$

For DWT high frequency coefficients, Burt and Kolczynski have proposed a window-based fusion rule [19], and it can be suitable for dealing with multi-focus image fusion. This algorithm is summarized below. Choose a small neighborhood window of $M \times N$ ($5 \times 5$ in this paper) centered on point $(x,y)$ of both $A$ and $B$, calculate its weighted local energy as salience measure of it

$$S^I(x,y) = \sum_{[l=-L\ldots L],[t=-T\ldots T]} w(I,t) \cdot [g^I(x+l,y+t)]^2, \ \ I = A,B,$$

where $w(l,t)$ denotes a coefficient template,

$$\sum_{[l=-L\ldots L],[t=-T\ldots T]} w(l,t) = 1 \text{ and } 2L+1=M, \ 2T+1=N.$$  

Then the local normalized correlation of the window is defined as

$$M_{AG}^{IL}(x,y) = \frac{\sum_{[l=-L\ldots L],[t=-T\ldots T]} w(I,t) \cdot g^I(x+l,y+t) \cdot g^B(x+l,y+t)}{S^A(x,y) + S^B(x,y)}, \ \ (7)$$

At each position $(x,y)$, assume a threshold $\alpha \in [0,1]$, if the match measure $M_{AG}^{IL}(x,y)$ satisfies $M_{AG}^{IL}(x,y) \leq \alpha$, then the selected coefficient can be written as

$$g^F(x,y) = \begin{cases} g^A(x,y), & \text{if } S^A(x,y) \geq S^B(x,y) \\ g^B(x,y), & \text{if } S^A(x,y) < S^B(x,y) \end{cases}, \ \ (8)$$

else if $M_{AG}^{IL}(x,y) > \alpha$, the combined result will be implemented as

$$g^F(x,y) = \begin{cases} \omega_{\min} \cdot g^A(x,y) + \omega_{\max} \cdot g^B(x,y), & \text{if } S^A(x,y) \geq S^B(x,y) \\ \omega_{\min} \cdot g^A(x,y) + \omega_{\max} \cdot g^B(x,y), & \text{if } S^A(x,y) < S^B(x,y) \end{cases}, \ \ (9)$$

where $\omega_{\min} = \frac{1}{2} - \frac{1}{2} \left( 1 - \frac{M_{AG}^{IL}(x,y)}{1 - \alpha} \right)$ and $\omega_{\max} = 1 - \omega_{\min}$.

As for $F_{SF}(x,y)$, it is obtained via the above mentioned spatial frequency method. Spatial frequency rule is applied through these steps[10]:

1. Decompose the source images into $k$ blocks of size $M \times N$.
2. Compute the above-mentioned spatial frequency for each block.
3. Compare the spatial frequency of two corresponding blocks $A_k$ and $B_k$, then construct the $k$th block $F_k$ of the fused image as

$$F_k^{SF} = \begin{cases} A_k, & \text{if } SF_{k}^{A} > SF_{k}^{B} + \text{Threshold} \\ B_k, & \text{if } SF_{k}^{A} < SF_{k}^{B} - \text{Threshold} \end{cases}, \ \ (10)$$

On the second stage, the intermediate images $F_{DWT}(x,y)$ and $F_{SF}(x,y)$ will be fused again to generate the final fused $F(x,y)$ using two schemes respectively:

Rule A: Generate the final fused image using spatial frequency rule with blocks of relatively higher resolution.

Rule B: Generate the final fused image using a novel DWT fusion method at a relatively low decomposition level.

The reason why we do not use the same fusion scheme of low frequency subband described in the first stage is that the intermediate image is with higher sharpness and exhibit some ringing effects generated by the DWT fusion, and the region with ringing effects will gain local gradient values which makes mistakes in clear region detection. If we still use the gradient features to detect clear region, these bad ringing effects will still be preserved in the finally fused image.

The novel wavelet-based fusion algorithm uses a new low frequency fusion strategy based on neighborhood operation and adaptive weights. Given two low frequency sub-images $f^A(x,y)$ and $f^B(x,y)$ of intermediate images $F_{DWT}(x,y)$ and $F_{DWT}(x,y)$, choose a small neighborhood window of $N \times N$ ($5 \times 5$ in this paper) centered on point $(x,y)$, then we regard this window as $N \times N$ matrix $W$. The matrix $W$ can be partitioned by columns and rows as.
frequency sub-image can be written as

$$W = (p_1, p_2, p_3, \ldots, p_N) = [q_1, \ldots, q_N].$$

Consider the deviation matrix of column vectors $p_i$ $(i=1, 2, \ldots, N)$ as

$$P = \frac{1}{N} \sum_{i=1}^{N} (p_i - \bar{p})(p_i - \bar{p})^T,$$  \hspace{1cm} (12)

where $\bar{p}$ denotes the mean vector of $p_i$. Similarly, we have

$$Q = \frac{1}{N} \sum_{i=1}^{N} (q_i - \bar{q})(q_i - \bar{q})^T,$$  \hspace{1cm} (13)

Let $\Lambda_i$ $(i=1, 2, \ldots, N)$ be the eigenvalues of matrix $P$, then we calculate the trace of matrix $P$ and use the fact that

$$Tr(P) = \frac{1}{N} \sum_{i=1}^{N} \|p_i - \bar{p}\|^2 = \sum_{i=1}^{N} \Lambda_i,$$  \hspace{1cm} (14)

From the equation above, it can be easily found that the trace of matrix $P$ indicates mean square Euclidean distance of columns which describes the distribution of each column vector around their mean vector. Thus with larger trace or larger sum of eigenvalues, the column vectors tend to have a relatively "rough" distribution around their mean vector. Taking rows of $W$ into consideration, we define the overall significance level of point $(x,y)$ as

$$Sig'(x,y) = Tr(P) \cdot Tr(Q)$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} \|p_i - \bar{p}\|^2 \cdot \sum_{i=1}^{N} \|q_i - \bar{q}\|^2, \hspace{1cm} I = A,B$$  \hspace{1cm} (15)

If the neighborhood window of point $(x,y)$ is with relatively high sharpness, its columns and rows will gain "rough" distribution around their mean vectors, that means a larger value of $Sig(x,y)$, and vice versa. Then the fused low frequency sub-image can be written as

$$f^E(x,y) = \frac{Sig^A(x,y)}{Sig^A(x,y) + Sig^B(x,y)} \cdot f^A(x,y)$$

$$+ \frac{Sig^B(x,y)}{Sig^A(x,y) + Sig^B(x,y)} \cdot f^B(x,y).$$  \hspace{1cm} (16)

As for high frequency coefficients, the fusion scheme will still conform to above-mentioned algorithm. Experiments show that this new low frequency fusion algorithm performs the best with high resolution sub-images, thus with a low decomposition level of the DWT, the approximate sub-images have high resolution and the fusion result will be the best.

The two proposed fusion rules on the second stage are used to partially eliminate distortion and blocking effects brought by the DWT and SF on the first stage. These artifacts come from the fusion on the first stage. Then after the fusion on the second stage, blurred regions will be replaced by relatively clear regions and geometric distortions will be alleviated.

Figure 3 represents the schematic diagram for the entire fusion rule:

![Figure 3: Entire fusion process using DWT and SF.](image)

**V. EXPERIMENT RESULTS**

The concrete formulation implemented on the fusion process using the DWT and SF is shown as follows. During the experiment period, two different images focused at two different objects are provided for fusion. These two images have been spatially registered. The experiment involves three tests. First test uses artificially blurred images with reference images while the last two tests use naturally blurred images with no reference images.

In order to compare the proposed rules and traditional ones, we employ three additional multi-resolution based fusion algorithms and one block-based method which are the DWT, the SF, the Dual Tree-Complex Wavelet Transform(DT-CWT) [20] and the Low Redundancy Discrete Wavelet Frame(LRDWF) [12]. All parameters here are set to reach the best performance. For the multi-resolution based methods DWT, DT-CWT and LRDWF, the coefficients are combined by choosing maximum absolute values of transformation coefficients. The wavelet function 'db9' is chosen in DWT fusion method and the decomposition level is 4 for the best performance. The basis 'db4' is chosen in DT-CWT and LRDWF methods and decomposition level is 4 for DT-CWT and LRDWF to obtain better results. For the traditional SF method, block size will be chosen adaptively for the best performance. The evaluation criteria used here are RMSE (Root Mean Square Error), PSNR (Peak Signal-to-Noise Ratio), $Q^AB$ [21][22] and MI (Mutual Information)[23].

The first implementation of artificial blurred images with resolution of $512 \times 512$ is shown in Figure 4. For rule A, images are divided into $16 \times 16$ blocks on the first stage and $32 \times 32$ blocks on the second stage in the process of the SF fusion, and the DWT decomposition level is 5. As for rule B, the decomposition level of the DWT is 6 on the first stage and
is changed to 1 on the second stage, and images are divided into $16 \times 16$ blocks on the first stage using SF.

![Figure 4](image)

Figure 4. Test 1: source images and fused images. (a) Reference image (b) Source image focused on the foreground (c) Source image focused on the background (d) Fused image using DWT (e) Fused image using SF (f) Fused image using DT-CWT (g) Fused image using LRDWF (h) Fused image by taking fusion rule A (i) Fused image by taking fusion rule B.

A comparison of the results in terms of the evaluation criteria is shown in Table 1.

<table>
<thead>
<tr>
<th>Test 1</th>
<th>Method</th>
<th>$RMSE$</th>
<th>$PSNR$</th>
<th>$Q^{RB}$</th>
<th>$MI$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DWT</td>
<td>1.8108</td>
<td>42.9733</td>
<td>0.6844</td>
<td>7.4633</td>
</tr>
<tr>
<td></td>
<td>SF</td>
<td>1.3451</td>
<td>45.5555</td>
<td>0.7098</td>
<td>9.2419</td>
</tr>
<tr>
<td></td>
<td>DT-CWT</td>
<td>1.2888</td>
<td>43.3666</td>
<td>0.6988</td>
<td>7.8237</td>
</tr>
<tr>
<td></td>
<td>LRDWF</td>
<td>1.5033</td>
<td>42.3718</td>
<td>0.6956</td>
<td>7.6043</td>
</tr>
<tr>
<td></td>
<td>Rule A</td>
<td>1.2675</td>
<td>46.0715</td>
<td>0.9601</td>
<td>12.6606</td>
</tr>
<tr>
<td></td>
<td>Rule B</td>
<td>1.2836</td>
<td>45.9624</td>
<td>0.9556</td>
<td>12.5157</td>
</tr>
</tbody>
</table>

Test 2 illustrates an experiment on naturally blurred images with resolution of $1280 \times 960$. For rule A, images are divided into $32 \times 32$ blocks on the first stage and $64 \times 64$ blocks on the second stage in the process of SF fusion, and DWT decomposition level is 5. As for rule B, the decomposition level of DWT is 4 on the first stage and is changed to 1 on the second stage, and images are divided into $64 \times 64$ blocks on the first stage using SF.

Another test results on naturally blurred images with resolution of $512 \times 512$ is show in Figure 6. For rule A, images are divided into $32 \times 32$ blocks on the first stage and $64 \times 64$ blocks on the second stage in the process of SF fusion, and DWT decomposition level is 5. As for rule B, the decomposition level of DWT is 5 on the first stage and is changed to 2 on the second stage, and images are divided into $32 \times 32$ blocks on the first stage using SF.

![Figure 5](image)

![Figure 6](image)
The objective performance assessments are shown in Table 2.

**TABLE 2. COMPARISON OF FUSION RESULTS FOR IMAGES WITHOUT REFERENCES**

<table>
<thead>
<tr>
<th>Method</th>
<th>( Q^{AB/F} )</th>
<th>MI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWT</td>
<td>0.67475</td>
<td>5.9826</td>
</tr>
<tr>
<td>SF</td>
<td>0.70525</td>
<td>8.6629</td>
</tr>
<tr>
<td>DT-CWT</td>
<td>0.68213</td>
<td>6.5843</td>
</tr>
<tr>
<td>LRDWF</td>
<td>0.67422</td>
<td>6.4981</td>
</tr>
<tr>
<td>Proposed rule A</td>
<td>0.94830</td>
<td>9.8104</td>
</tr>
<tr>
<td>Proposed rule B</td>
<td>0.94569</td>
<td>9.9081</td>
</tr>
<tr>
<td>Test 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWT</td>
<td>0.74867</td>
<td>6.3184</td>
</tr>
<tr>
<td>SF</td>
<td>0.79816</td>
<td>7.6934</td>
</tr>
<tr>
<td>DT-CWT</td>
<td>0.77735</td>
<td>6.7656</td>
</tr>
<tr>
<td>LRDWF</td>
<td>0.77357</td>
<td>6.6787</td>
</tr>
<tr>
<td>Proposed rule A</td>
<td>0.94309</td>
<td>10.0023</td>
</tr>
<tr>
<td>Proposed rule B</td>
<td>0.93577</td>
<td>9.9143</td>
</tr>
</tbody>
</table>

It can be easily found from the above pictures that the proposed approaches show better visual effects than wavelet-based methods and SF fusion rules in terms of distortion effects and de-blocking. Some local magnifications below may clearly present the amelioration of these visual effects.

**REFERENCES**


[8] Yocky, D. A., Image merging and data fusion by means of the discrete two dimensional wavelet transform, Journal of the Optical Society of


