Abstract - A dynamic image fusion scheme for infrared and visible sequence based on region target detection is proposed in this paper. Target detection technique is employed to segment the source images into target and background regions. Different fusion rules are adopted respectively in target and background regions. A limitedly redundant discrete wavelet transform (LR DWT) method is introduced to achieve shift invariant multi-resolution representation of each source images. Fusion experiments on real world image sequences indicate that the proposed method is effective and efficient, which achieves better performance than the generic fusion method.

Keywords: Dynamic Image fusion, discrete wavelet transform, region target detection.

1 Introduction

Image fusion is a specialization of the more general topic of data fusion, dealing with image and video data [1]. It is the process by which multi-modality sensor imageries from same scene are intelligently combined into single view of the scene with extended information content. Image fusion has the important applications in the military, medical imaging, remote sensing, and security and surveillance fields. The benefits of image fusion include improved spatial awareness, increased accuracy in target detection and recognition, reduced operator workload and increased system reliability [2].

Image fusion processing must satisfy the following requirements, as described in [3]: Preserve (as far as possible) all salient information in the source images; do not introduce any artifacts or inconsistencies; be shift invariant; be temporal stable and consistent. The last two points are especially important in dynamic image fusion (or image sequences fusion) as human visual system is highly sensitive to moving artifacts introduced by the shift dependent fusion process [3].

Fusion process can be performed at different levels of information representation, sorted in ascending order of abstraction: signal, pixel, feature and symbol levels [4]. From the simplest weighted pixel averaging to more complicated multi-resolution (MR) method (including pyramidal schemes and wavelet schemes), pixel-based fusion methods were well researched [3]-[5],[10]. Recently, feature level fusion with region-based fusion scheme has been reported both qualitative and quantitative improvements over the pixel-based method as more intelligent semantic fusion rules can be considered based on actual features [11]-[16].

2 Generic Pixel-Based Image Fusion Scheme

The generic pixel-based fusion scheme is briefly reviewed, more details can be found in [3]-[8],[16]. Fig.1 illustrates generic wavelet fusion scheme, which can be divided into three steps as following: At first, all source images are decomposed by using multi-resolution method, which can be the pyramid transform (PT) [5],[6], discrete wavelet transform (DWT) [7]-[9], discrete wavelet frames (DWF) [3] or dual-tree complex wavelet transform [10] etc. Then the decomposition coefficients are fusion by applying a fusion rule, which can be a point-based maximum selection (MS) rule or more sophisticated area-based rules [6],[7]. Finally, the fused image is reconstructed by using the corresponding inverse transform on the fused coefficients.

3 The Region-based Target Detection Dynamic Image Fusion Scheme

For pixel-based approaches, the MR decomposition coefficient is treated independently (MS rule) or filtered by a small fixed window (area-based rule). However, the most applications of a fusion scheme are interested in features within the image, not in the actual pixels. Therefore, it seems reasonable to incorporate feature information into the fusion process [14]. A number of region-based fusion schemes have been proposed [11]-[16]. However, most of region-based schemes are designed for still image fusion, and every frame of each source sequence is processed individually in image sequences case. These methods do not take full advantage of the wealth of inter-frame-information within source sequences.

The novel region-based fusion scheme proposed for fusion of visible and infrared (IR) image sequences is
shown in Fig. 2, where the target detection (TD) techniques are introduced to segment target regions intelligently. For convenience, we assume both source sequences are registered well before fusion. Firstly, both the visible and IR sequences are enhanced by using preprocessing operator. Then each frame of the source sequences transformed by using a MR method (where the LR DWT, that is limited redundancy DWT, is adopted in this paper, see the Sec. 3.1). Simultaneously, the frames are segmented into object and background regions by using a TD method. Different fusion rules are adopted in target and background regions. Finally, the fused coefficients belong to each region are combined, and fused frames are reconstructed by using the corresponding inverse transform.

![Diagram](image_url)

**Fig. 2 Region-based IR and visible dynamic image fusion scheme**

### 3.1 The Limited Redundancy Discrete Wavelet Transform

It is well known that the standard DWT produces a shift dependent signal representation due to down-sampling operations in every sub-band, which results in a shift dependent fusion scheme, as described by Rockinger. To overcome the problem, Rockinger presents a perfectly shift invariant wavelet fusion scheme by using DWF. However, this method is computationally more expensive due to high redundancy \(2^{m} \times n\) for m-D and n-level decomposition) of the representation. Bull et al. further develop the wavelets fusion method by introducing DT CWT, which provides approximately shift invariance by introducing limited redundancy \(2^{m} : 1\) for m-D and any levels decomposition. However, the DT CWT employs two filters banks, which must be designed rigorously to achieve appropriate delays while satisfying perfect reconstruction (PR) conditions. Moreover, the decomposition coefficients from every tree should be regarded as the real or imaginary part of complex, increasing difficulty of subsequent processing in the fusion rule.

A new implementation of the DWT is introduced in this paper, which provides approximate shift invariance and nearly perfect reconstruction while preserving the properties of DWT: computational efficiency and easy implementation. Fig.3 shows decomposition and reconstruction scheme for the new transform in cascading form, which can be extended easily to 2-D by separable filtering along rows and then columns. For the limited redundancy (not more than 3:1 for 1-D) of the new transform, we use LR DWT (the limitedly redundant discrete wavelet transform) here to distinguish it from DWT and DWF.

![Diagram](image_url)

**Fig. 3 The LR DWT and inverse transform.**

### 3.2 Region Segmentation Algorithm

The target detection (TD) operator aims for segmenting both source frames into target regions, in which the significant information is included such as moving human and vehicle, and background regions. A novel target detection method is proposed in this paper based on the characteristics of IR imaging. At first, a region merging method \(^{[19]}\) is adopted to segment the initial IR frame. It is easy to find the target regions, which have high contrast with the neighboring background, in the segmented IR frame. A confidence measure \(^{[20]}\) for each candidate region is computed. It is very inefficient to compute the confidence measure for each candidate within every frame. Therefore, a model matching method is adopted to find the target regions in the subsequent frames. A target model is obtained by using intensity information of the target region in pre-frame. Not the whole but a small region in post-frame which correspond with (and is little larger than) the target region in pre-frame is matched. The initial detection operator based on segmentation and confidence measure will be repeated in case no target being detected in certain successive frames. The target detection in the visible sequence is similar to the IR sequence.

### 3.3 Fusion Rules in the Target Region

To preserve the full information as far as possible in the target region, a special fusion rule should be employed in object region. Assume that target detection gives \(M\) target maps: \(T_{IR} = \{t_{1}, t_{2}, ..., t_{M}\}\) in IR frame and \(N\) target region maps: \(T_{V} = \{v_{1}, v_{2}, ..., v_{N}\}\) in the corresponding visible frame.

The target map is down sampled by \(2^{m}\) (according to the resolution of decomposition coefficients) to give a decimated target map at each level. The target maps in both source frames are analyzed jointly \(T_{J} = T_{IR} \cup T_{V}\). The frame is segmented into three sets: single, overlapped target region sets and background region set. Overlapped
target regions are defined as \( T_0 = T_R \cap T_V \). Single target regions are all the target regions where no overlap \( T_j = T_R \cup T_O \). Clearly, there is \( T_j = T_R \cup T_O \). Background regions are defined as \( B = T_j \).

In the single target regions, fusion rule can be written as:

\[
\begin{align*}
&c_f(x,y) = \begin{cases} 
  c_v(x,y), & \text{if } (x,y) \in T_R \\
  c_i(x,y), & \text{if } (x,y) \in T_V 
\end{cases}
\end{align*}
\]

(1)

In a connected overlapped target region \( t \in T_O \), a similarity measure between two source is defined as:

\[
M(t) = 2 \cdot \left( \sum_{(x,y) \in T_I} I_v(x,y) \cdot I_i(x,y) \right) - \left( \sum_{(x,y) \in T_I} I_v(x,y) \right)^2 - \left( \sum_{(x,y) \in T_I} I_i(x,y) \right)^2
\]

(2)

where \( I_v \) and \( I_i \) denote IR and visible frames respectively. Then an energy index of the coefficients within the overlapped region is computed respectively in IR and visible frame,

\[
S_i(t) = \sum_{(x,y) \in T_i} c_i(x,y)^2
\]

(3)

where \( t \in T_O \) and \( i=ir,v \) means the IR and visible frame respectively. A threshold of similarity \( \alpha \) is introduced where \( \alpha \in [0,1] \) and normally \( \alpha = 0.85 \) is appropriate.

In case \( M(t) < \alpha \) the fusion rule in overlapped target region \( t \in T_O \) can be written as:

\[
\begin{align*}
&c_f(x,y) = \begin{cases} 
  c_v(x,y), & \text{if } S_v(t) \geq S_i(t) \\
  c_i(x,y), & \text{otherwise}
\end{cases}
\end{align*}
\]

(4)

In case \( M(t) \geq \alpha \), a weight average method is adopted:

\[
\begin{align*}
&c_f(x,y) = \begin{cases} 
  \sigma_{\min}(t) \cdot c_v(x,y) + \sigma_{\max}(t) \cdot c_i(x,y), & \text{if } S_v(t) \geq S_i(t) \\
  \sigma_{\min}(t) \cdot c_i(x,y) + \sigma_{\max}(t) \cdot c_v(x,y), & \text{if } S_v(t) < S_i(t)
\end{cases}
\end{align*}
\]

(5)

where the weights \( \sigma_{\min}(t) \) and \( \sigma_{\max}(t) \) can be obtained

\[
\begin{align*}
\sigma_{\min}(t) = & \frac{1}{2} \left( 1 - \frac{1-M(t)}{1-\alpha} \right) \\
\sigma_{\max}(t) = & 1 - \sigma_{\min}(t)
\end{align*}
\]

(6)

Finally, in the background regions, the simplest MS rule is adopted.

4 Experimental Results and Analysis

The proposed fusion scheme was examined on real world dynamic images (image sequences). Fig.4 shows some results of fusion of spatially registered visual (visible light) and infrared (IR) (thermal 3-5 micro m) image sequences consisting of 32 subsequent frames of each sensor. Visual inspection of the results from each method shows that the proposed dynamic scheme is superior to the generic scheme in case of same transform method.

To quantify temporal stability and consistency, we adopt the quantitative measure \( I(S_v, S_i, F) \) proposed by Rockinger, more details can be found in [3]. We computed the average mutual information (AMI) over the 31 set of inter-frame-differences (IFDs) using each scheme with each wavelet fusion method. In Table 1, DB4 means the Daubechies wavelets of order 4, BIOR4.4 denotes the biorthogonal wavelets, where the two “4” are orders of synthesis and analysis filters respectively, Q-shift9 designed by Kingsbury [21]. “Generic” and “Proposed” means using the generic wavelet fusion scheme (fusion frames one by one) and the proposed region-based dynamic fusion method. Clearly, the measure results in Table 1 are consistent with the conclusion of visual inspection. In addition, the LR DWT method performs better than the DWT and DT CWT methods in both generic and dynamic schemes, and it is comparable with the DWF method in respects of temporal stability and consistency while using lower computational cost because of lower redundancy of the LR DWT.

<table>
<thead>
<tr>
<th>Fusion Scheme</th>
<th>DWT DB4</th>
<th>DWT BIOR4.4</th>
<th>DT CWT Q-shift9</th>
<th>LR DWT DB4</th>
<th>LR DWT BIOR4.4</th>
<th>DWF DB4</th>
<th>DWF BIOR4.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>0.5233</td>
<td>0.5453</td>
<td>0.6362</td>
<td>0.8534</td>
<td>0.8624</td>
<td>0.9066</td>
<td>0.9173</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.6021</td>
<td>0.6144</td>
<td>0.6809</td>
<td>0.9062</td>
<td>0.9081</td>
<td>0.9245</td>
<td>0.9201</td>
</tr>
</tbody>
</table>

Figure 4 fused results of visual (visible light) and infrared (IR) image sequences

(c) - (f): Fused frames using the generic scheme with DWT (DB4), DT CWT (Q-shift9), LR DWT (DB4) and DWF (DB4), respectively. (g)-(j) corresponding fused frames using the proposed region-based target detection dynamic fusion scheme.

The original IR and visible images, which are available online at www.imagefusion.org, are kindly supplied by Alexander Toet of the TNO Human Factors Research Institute.
Table 2. The quantitative evaluation for different fusion methods of visual and IR image sequences.

<table>
<thead>
<tr>
<th>Method</th>
<th>AG</th>
<th>En</th>
<th>SF</th>
<th>AMI</th>
<th>CE</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGA</td>
<td>1.71</td>
<td>6.61</td>
<td>3.74</td>
<td>4.25</td>
<td>1.58</td>
</tr>
<tr>
<td>LPT</td>
<td>1.97</td>
<td>6.72</td>
<td>4.48</td>
<td>3.80</td>
<td>1.72</td>
</tr>
<tr>
<td>CPT</td>
<td>2.09</td>
<td>6.72</td>
<td>4.66</td>
<td>3.82</td>
<td>1.72</td>
</tr>
<tr>
<td>GPT</td>
<td>1.62</td>
<td>6.51</td>
<td>3.27</td>
<td>3.61</td>
<td>1.42</td>
</tr>
<tr>
<td>DWT</td>
<td>2.35</td>
<td>6.64</td>
<td>5.08</td>
<td>3.62</td>
<td>1.69</td>
</tr>
<tr>
<td>LR DWT</td>
<td>1.80</td>
<td>6.61</td>
<td>3.85</td>
<td>4.13</td>
<td>1.58</td>
</tr>
</tbody>
</table>

In Table 2, AG means Average degree, En denotes Entropy, SF means spatial frequency, AMI denotes average mutual information, and CE means Cross-entropy. All of these quantitative Evaluation Indexes are described in reference [22].

5 Conclusion

In this paper, a dynamic image fusion scheme for infrared and visible sequence based on region target detection is proposed. Target detection technique is employed to segment the source images into target and background regions. Different fusion rules are adopted respectively in target and background regions. A limitedly redundant discrete wavelet transform (LR DWT) method is introduced to achieve shift invariant multi-resolution representation of each source images. Fusion experiments on real world image sequences indicate that the proposed method is effective and efficient, which achieves better performance than the generic fusion method.

Acknowledgments

This work has been jointly supported by Shanghai-NRC International Co-operate Project (05SN07118), Basic Research Project (A1420060161) and the Project (9140A010606JW0305).

References


