Multi-sensor image fusion using discrete wavelet frame transform

Zhenhua Li (李振华), Zhongliang Jing (敬忠良), and Shaoyuan Sun (孙韶媛)

Institute of Aerospace Information and Control, School of Electrical and Information Engineering, Shanghai Jiaotong University, Shanghai 200030

Received March 24, 2004

An algorithm is presented for multi-sensor image fusion using discrete wavelet frame transform (DWFT). The source images to be fused are firstly decomposed by DWFT. The fusion process is the combining of the source coefficients. Before the image fusion process, image segmentation is performed on each source image in order to obtain the region representation of each source image. For each source image, the salience of each region in its region representation is calculated. By overlapping all these region representations of all the source images, we produce a shared region representation to label all the input images. The fusion process is guided by these region representations. Region match measure of the source images is calculated for each region in the shared region representation. When fusing the similar regions, weighted averaging mode is performed; otherwise selection mode is performed. Experimental results using real data show that the proposed algorithm outperforms the traditional pyramid transform based or discrete wavelet transform (DWT) based algorithms in multi-sensor image fusion.

OCIS code: 100.0100.

With the rapid improvement of sensor technology, numerous multi-sensor data are obtained in many fields such as remote sensing, medical imaging, machine vision, and military application. Multi-sensor images often contain complementary and redundant information about the region surveyed. Through combining registered images generated by different imaging systems, image fusion can produce new images with more complete information, which are more suitable for the purposes of human vision perception, object detection, and automatic target recognition. The system reliability can be improved by using the redundant information, and also the system capability can be improved by using the complementary information.

Multi-resolution decomposition is widely used in multi-sensor image fusion. It mainly includes pyramid transform in Refs. [1−3] (such as Laplacian pyramid, gradient pyramid, and the ratio-of-low-pass pyramid) and discrete wavelet transform (DWT) in Refs. [4−6]. A pyramid structure is an efficient way to implement multiscale representation. Each image in a pyramid is a filtered and sub-sampled copy of the previous images. DWT described in Refs. [7,8] can also decompose an image into several components, each of which captures information present at a given scale. In the filter theory, DWT decomposition of an image can be regarded as two directional filtering operations (on rows and columns) and subsampling by one of two factors. The pyramid transform or DWT yields a shift variant data representation that is not appropriate for multi-sensor image fusion. Compared with pyramid transform or DWT, discrete wavelet frame transform (DWFT) described in Refs. [9,10] avoids subsampling by using an overcomplete wavelet decomposition. This results in both aliasing free and translation invariant. Consequently, DWFT is more suitable for image fusion than pyramid transform or DWT. A generic framework of image fusion schemes was proposed[11], in which DWFT was considered to be one method of multi-resolution decomposition, and a region based fusion rule was proposed. In this paper, we propose a new multi-sensor image fusion scheme based on DWFT. The source images are firstly decomposed by DWFT. Corresponding coefficients are combined according to the region representations. Fused image is obtained by performing inverse DWFT. In our fusion scheme, a new region salience measure and a new region representation method are used to guide the fusion of multi-sensor images.

The block diagram of two-dimensional (2D) DWFT and inverse DWFT transform is shown in Fig. 1. For image I to be decomposed, we use $f(x,y)$ to denote the intensity of pixel (x,y). The DWFT coefficients of pixel (x,y) at level $i+1$ can be defined as

$$f_{i+1}(x,y)=\sum_{p,q}h_{\uparrow 2^i}(p-x)h_{\uparrow 2^i}(q-y)f_i(p,q), \quad (1)$$

$$w^1_{i+1}(x,y)=\sum_{p,q}h_{\uparrow 2^i}(p-x)g_{\uparrow 2^i}(q-y)f_i(p,q), \quad (2)$$

$$w^2_{i+1}(x,y)=\sum_{p,q}g_{\uparrow 2^i}(p-x)h_{\uparrow 2^i}(q-y)f_i(p,q), \quad (3)$$

$$w^3_{i+1}(x,y)=\sum_{p,q}g_{\uparrow 2^i}(p-x)g_{\uparrow 2^i}(q-y)f_i(p,q), \quad (4)$$

where $f_0(x,y)=f(x,y)$, h and g are 1D prototype filters, $h_{\uparrow 2^i}$ and $g_{\uparrow 2^i}$ are dilated versions of low-pass filter h and high-pass filter g, respectively. The reconstruction process at level i is defined as

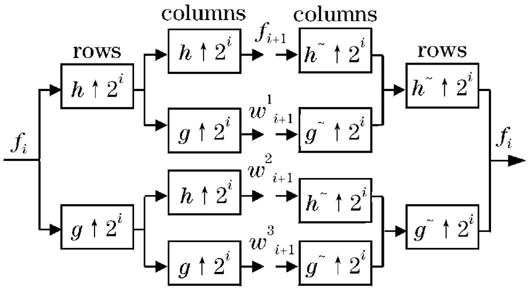

Fig. 1. The block diagram of DWFT and inverse DWFT.

$$f_i(x,y) = \sum_{p,q} \tilde{h}_{\uparrow 2^i}(p-x)\tilde{h}_{\uparrow 2^i}(q-y) f_{i+1}(p,q)$$
$$+ \sum_{p,q} \tilde{h}_{\uparrow 2^i}(p-x)\tilde{g}_{\uparrow 2^i}(q-y) w^1_{i+1}(p,q)$$
$$+ \sum_{p,q} \tilde{g}_{\uparrow 2^i}(p-x)\tilde{h}_{\uparrow 2^i}(q-y) w^2_{i+1}(p,q)$$
$$+ \sum_{p,q} \tilde{g}_{\uparrow 2^i}(p-x)\tilde{g}_{\uparrow 2^i}(q-y) w^3_{i+1}(p,q), \quad (5)$$

where $\tilde{h}$ and $\tilde{g}$ are 1D synthesis filters, $\tilde{h}_{\uparrow 2^i}$ and $\tilde{g}_{\uparrow 2^i}$ are dilated versions of $\tilde{h}$ and $\tilde{g}$, respectively.

A DWFT with N decomposition levels will have a total of $3 \cdot N + 1$ subbands. All of them are of the same size as the source image. The nth subband coefficient of pixel (x,y) is denoted by DWF(x,y,n), where $n = 1,2,\cdots,3\cdot N+1$.

Before the fusion process by combining the source DWFT coefficients, image segmentation in Ref. [12] is performed on source images A and B. We denote R_A (or R_B) as the region representation (different label values representing different regions) used to label the segmented image of A (or B). A shared region representation R used to label both A and B is obtained by overlapping two segmented images: each intersection area belongs to a different region. The salience of a region is measured by the region's mean square deviation (σ) and abrupt change magnitude (τ). For region $r_\mathrm{A} \in R_\mathrm{A}$, the region mean square deviation of r_A is defined as

$$\sigma_\mathrm{A}(r_\mathrm{A}) = \sqrt{\frac{1}{\mathrm{Num}(r_\mathrm{A})} \sum_{(x,y)\in r_\mathrm{A}} [f_\mathrm{A}(x,y) - E_\mathrm{A}(r_\mathrm{A})]^2}, \quad (6)$$

where $f_\mathrm{A}(x,y)$ is the intensity of pixel (x,y) of image A in r_A, $\mathrm{Num}(r_\mathrm{A})$ is the total number of pixels in r_A, $E_\mathrm{A}(r_\mathrm{A})$ is the mean intensity of all pixels in region r_A, $E_\mathrm{A}(r_\mathrm{A}) = \frac{1}{\mathrm{Num}(r_\mathrm{A})} \sum_{(x,y)\in r_\mathrm{A}} f_\mathrm{A}(x,y)$. Regions having large region mean square deviations often contain much information in the sense of image detail. For region $r_\mathrm{A} \in R_\mathrm{A}$, the region abrupt change magnitude of r_A is defined as

$$\tau_\mathrm{A}(r_\mathrm{A}) = \frac{P(r_\mathrm{A})}{\mathrm{Num}(r_\mathrm{A})}$$
$$\times \frac{1}{T} \sum_{\mathrm{nr}} (|E_\mathrm{A}(r_\mathrm{A}) - E_\mathrm{A}(nr)| + |\sigma_\mathrm{A}(r_\mathrm{A}) - \sigma_\mathrm{A}(nr)|), \quad (7)$$

where 'nr' is a neighborhood region of region r_A in R_A, T is the total number of the neighborhood regions around r_A, $P(r_\mathrm{A})$ is the perimeter of region r_A and is calculated as the total number of pixels in the boundary of region r_A. Regions having large abrupt change magnitude values often contain much important information such as targets being detected.

For region $r_\mathrm{A} \in R_\mathrm{A}$, the region salience of r_A is defined as

$$S_\mathrm{A}(r_\mathrm{A}) = \varpi_1 \cdot \tau_\mathrm{A}(r_\mathrm{A}) + \varpi_2 \cdot \sigma_\mathrm{A}(r_\mathrm{A}), \quad (8)$$

where ϖ_1 and ϖ_2 are weights, $\varpi_1 + \varpi_2 = 1$. For region $r_\mathrm{B} \in R_\mathrm{B}$, the region salience $S_\mathrm{B}(r_\mathrm{B})$ is calculated in the same way as the computation of $S_\mathrm{A}(r_\mathrm{A})$.

The fusion process is guided by the shared region representation R. For region $r \in R$, the region match measure between images A and B is defined as

$$M_\mathrm{AB}(r) = \frac{1}{3} \cdot \left\{ \frac{2 \cdot \sum_{(x,y)\in r} f_\mathrm{A}(x,y) \cdot f_\mathrm{B}(x,y)}{\sum_{(x,y)\in r} (f_\mathrm{A}(x,y))^2 + \sum_{(x,y)\in r} (f_\mathrm{B}(x,y))^2} \right.$$
$$\left. +2 - \frac{|E_\mathrm{A}(r) - E_\mathrm{B}(r)|}{E_\mathrm{A}(r) + E_\mathrm{B}(r)} - \frac{|\sigma_\mathrm{A}(r) - \sigma_\mathrm{B}(r)|}{\sigma_\mathrm{A}(r) + \sigma_\mathrm{B}(r)} \right\}. \quad (9)$$

To determine whether selection or averaging will be used in the fusion process, $M_\mathrm{AB}(r)$ is compared to a threshold α. For region $r \in R$, if $M_\mathrm{AB}(r)$ is less than or equal to α, then selection mode is implemented as

$$\mathrm{DWF_F}(x,y,n) = \begin{cases} \mathrm{DWF_A}(x,y,n), & \text{if } S_\mathrm{A}(r) \ge S_\mathrm{B}(r) \\ \mathrm{DWF_B}(x,y,n), & \text{if } S_\mathrm{A}(r) < S_\mathrm{B}(r) \end{cases},$$
$$\text{for} \quad (x,y) \in r, n = 1,2,\cdots,3\cdot N+1, \quad (10)$$

where $\mathrm{DWF_F}(x,y,n)$ is the composite coefficient, 'F' indicates the fused image, $\mathrm{DWF_A}(x,y,n)$ and $\mathrm{DWF_B}(x,y,n)$ are the source coefficients of images A and B, respectively, $S_\mathrm{A}(r) = S_\mathrm{A}(r_\mathrm{A})$ subjected to $r \subseteq r_\mathrm{A}$ and $r_\mathrm{A} \in R_\mathrm{A}$, and $S_\mathrm{B}(r) = S_\mathrm{B}(r_\mathrm{B})$ subjected to $r \subseteq r_\mathrm{B}$ and $r_\mathrm{B} \in R_\mathrm{B}$. For region $r \in R$, if $M_\mathrm{AB}(r)$ is greater than α, then the source images are similar in region r and the weighted averaging mode is performed on fusing the source coefficients in region r. The weights used for averaging are defined as

$$\begin{cases} \varpi_\mathrm{min} = \frac{1}{2}\left(1 - \frac{1 - M_\mathrm{AB}(r)}{1-\alpha}\right) \\ \varpi_\mathrm{max} = 1 - \varpi_\mathrm{min} \end{cases}. \quad (11)$$

In the weighted averaging mode, the composite coefficient is the weighted average of the source coefficients and is calculated as

$$\mathrm{DWF_F}(x,y,n) =$$
$$\begin{cases} \varpi_\mathrm{max} \cdot \mathrm{DWF_A}(x,y,n) + \varpi_\mathrm{min} \cdot \mathrm{DWF_B}(x,y,n), \\ \quad \text{if} \quad S_\mathrm{A}(r) \ge S_\mathrm{B}(r) \\ \varpi_\mathrm{min} \cdot \mathrm{DWF_A}(x,y,n) + \varpi_\mathrm{max} \cdot \mathrm{DWF_B}(x,y,n), \\ \quad \text{if} \quad S_\mathrm{A}(r) < S_\mathrm{B}(r) \end{cases},$$
$$\text{for} \quad (x,y) \in r, n = 1,2,\cdots,3\cdot N+1. \quad (12)$$

After the composite coefficients are obtained, inverse DWFT transform is performed to get the fused image F.

Two experiments were conducted to compare our DWFT-based algorithm with pyramid transform based algorithm in Ref. [1] and discrete wavelet transform based algorithm in Ref. [4]. The results of experiment 1 are shown in Fig. 2. In experiment 1, the source images to be fused as Figs. 2(a) and (b) are infrared image and visual image (size: 460 × 346), respectively. The results of experiment 2 are shown in Fig. 3. In experiment 2, the source images to be fused as Figs. 3(a) and (b) are computerized tomography (CT) image and magnetic resonance (MR) image (size: 346×346), respectively. In the region representations of Fig. 2 or Fig. 3, different colors represent different regions. The "9-7" biorthogonal filters are used for DWT and DWFT. When performing the algorithms in Refs. [1,4], 3×3 window is used to calculate the salience of each coefficient after decomposition. The total decomposition level is 3 in all these experiments.

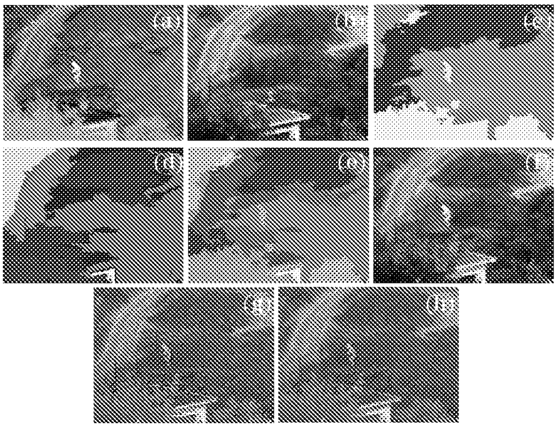

Fig. 2. Fusion results of infrared and visual images. (a) Infrared image to be fused; (b) visible image to be fused; (c) the region representation of (a); (d) the region representation of (b); (e) the shared region representation by overlapping (c) and (d); (f) fused image using our algorithm; (g) fused image using pyramid transform based algorithm in Ref. [1]; (h) fused image using DWT based algorithm in Ref. [4].

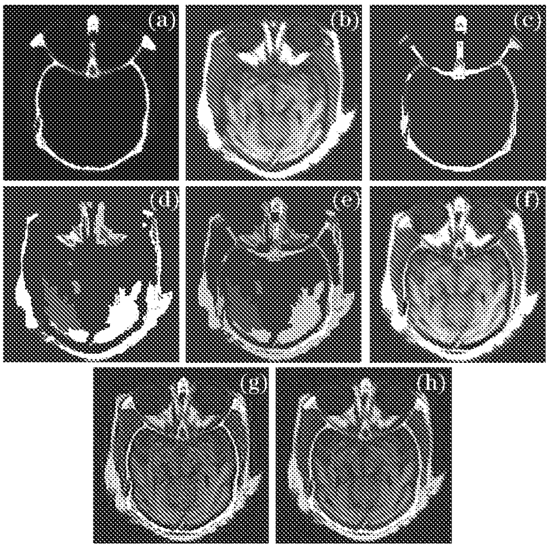

Fig. 3. Fusion results of CT and MR images. (a) CT image to be fused; (b) MR image to be fused; (c) the region representation of (a); (d) the region representation of (b); (e) the shared region representation by overlapping (c) and (d); (f) fused image using our algorithm; (g) fused image using pyramid transform based algorithm in Ref. [1]; (h) fused image using DWT based algorithm in Ref. [4].

Two objective evaluation methods are employed to evaluate the performance of each image fusion algorithm:

1) Entropy (H)

$$H = -\sum_{f=0}^{L} p(f) \log_2 p(f), \tag{13}$$

where $p(f)$ is the normalized histogram of the fused image F, L is the maximum intensity, in our experiments L is equal to 255. The entropy is used to measure the overall information in the fused image. The larger the value is, the better the fusion result we get.

2) Mutual Information (MI)

Mutual information measure proposed in Ref. [13] can estimate how much information is obtained from the source images. The larger the value is, the better the fusion result we get. Let $p_{\mathrm{A}}(a)$ and $p_{\mathrm{B}}(b)$ be the normalized histograms of images to be fused A and B, respectively, $p_{\mathrm{F}}(f)$ is the normalized histograms of fused image F, and $p_{\mathrm{FA}}(f,a)$(or $p_{\mathrm{FB}}(f,b)$) is the joint histogram of images F and A (or B). The joint histogram, for example $p_{\mathrm{FA}}(f,a)$, is defined as

$$p_{\mathrm{FA}}(f,a) = \frac{1}{N \cdot M} \sum_{n=1}^{N} \sum_{m=1}^{M} h_{\mathrm{FA}}(I_{\mathrm{F}}(m,n), I_{\mathrm{A}}(m,n)),$$

$$a = 0,1,\cdots,L, b = 0,1,\cdots,L, \tag{14}$$

where $h_{\mathrm{FA}}(I_{\mathrm{F}}(m,n), I_{\mathrm{A}}(m,n))$ equals to 1 if $I_{\mathrm{F}}(m,n) = f$ and $I_{\mathrm{A}}(m,n) = a$, otherwise it equals to 0, $I_{\mathrm{F}}(m,n)$ (or $I_{\mathrm{A}}(m,n)$) is the intensity of pixel (m,n) in image F (or A), L has the same meaning as which in Eq. (13), M is the height of the source images, and N is the width of the source images. The mutual information between the fused image (F) and the source images (A and B) is defined as

$$\mathrm{MI}_{\mathrm{F}}^{\mathrm{AB}} = I_{\mathrm{FA}} + I_{\mathrm{FB}}, \tag{15}$$

where $I_{\mathrm{FA}} = \sum_{f=0}^{L} \sum_{a=0}^{L} p_{\mathrm{FA}}(f,a) \log_2 \left(\frac{p_{\mathrm{FA}}(f,a)}{p_{\mathrm{F}}(f) \cdot p_{\mathrm{A}}(a)} \right)$ and $I_{\mathrm{FB}} = \sum_{f=0}^{L} \sum_{b=0}^{L} p_{\mathrm{FB}}(f,b) \log_2 \left(\frac{p_{\mathrm{FB}}(f,b)}{p_{\mathrm{F}}(f) \cdot p_{\mathrm{B}}(b)} \right)$.

Performance evaluation is shown in Table 1. From this table, we can conclude that the performance of our algorithm is better than that of these traditional pyramid based or DWT based multi-sensor image fusion algorithms.

Table 1. Performance Evaluation

Algorithm		DWFT	DWT	Pyramid
Exp. 1	H	7.0668	6.3431	6.2512
	MI	4.3732	1.3637	1.4797
Exp. 2	H	6.7018	6.2034	6.1025
	MI	4.6293	2.6157	3.2427

This work was jointly supported by the National '863' Program of China (No. 2001AA135091), National Natural Science Foundation of China (No. 60375008), Shanghai Key Scientific Project (No. 02DZ15001), China Ph.D Discipline Special Foundation (No. 20020248029), and China Aviation Science Foundation (No. 02D57003). Z. Li's e-mail is randy_lee@sjtu.edu.cn.

References

1. P. J. Burt and R. J. Kolczynski, in *IEEE Proc. of 4th Int. Conf. on Computer Vision* 173 (1993).
2. A. Toet, Pattern Recognition **9,** 245 (1989).
3. P. J. Burt, in *SPIE Proc. of the Society for Information Display Conference* **467,** (1992).
4. H. Li, B. S. Manjunath, and S. K Mitra, Graphical Models and Image Processing **57,** 235 (1995).
5. P. K. Varshney, H. M. Chen, L. C. Ramac, M. Uner, D. Ferris, and M. Alford, in *IEEE Proc. Int. Conf. Image Processing* **3,** 532 (1999).
6. L. C. Ramac, M. K. Uner, and P. K. Varshney, Proc. SPIE **3376,** 110 (1998).
7. S. G. Mallat, IEEE Trans. Pattern Analysis and Machine Intelligence **11,** 674 (1989).
8. S. G. Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, San Diego, California, 1998).
9. A. Laine and J. Fan, IEEE Trans. Image Processing **5,** 771 (1996).
10. M. Unser, IEEE Trans. Image Processing **4,** 1549 (1995).
11. Z. Zhang and R. S. Blum, Proceedings of IEEE **87,** 1315 (1999).
12. N. Frank and N. Richard, in *Proc. IEEE Conf. Computer Vision and Pattern Recognition* **2,** 19 (2003).
13. G. Qu, D. Zhang, and P. Yan, Electron. Lett. **38,** 313 (2002).