Image Fusion Based on EM Algorithm and Discrete Wavelete Frame[1)]

LIU Gang JING Zhong-Liang SUN Shao-Yuan

(*Institute of Aerospace Information and Control, School of Electronic, Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai* 200030)

(E-mail: {lukelg, zljing, shysun}@sjtu.edu.cn)

Abstract The discrete wavelet transform has become an attractive tool for fusing multisensor images. This paper investigates the discrete wavelet frame transform. A major advantage of this method over discrete wavelet transform is aliasing free and translation invariant. The discrete wavelet frame (DWF) transform is used to decompose the registered images into multiscale representation with the low frequency and the high frequency bands. The low frequency band is normalized and fused by using the expectation maximization (EM) algorithm. The informative importance measure is applied to the high frequency band. The final fused image is obtained by taking the inverse transform on the composite coefficient representations. Experiments show that the proposed method is more effective than conventional image fusion methods.

Key words Image fusion, expectation maximization algorithm, discrete wavelet frame transform

1 Introduction

Image fusion is the process of generating a single combined image, which contains a more accurate description of the scene than the multiple images from different sources[1]. This fused image is usually more useful for human visual or machine perception. Most image fusion approaches are based on combining the multiscale decomposition. The basic idea is to perform a multiscale transform (MST) on all source images, and construct a composite multiscale representation of them. The fused image is then obtained by taking the inverse multiscale transform (IMST). Recently, estimation theory has been proposed to improve the efficiency of image fusion algorithms[2,3]. However, these approaches are all based on the assumption that the disturbance follows a Gaussian distribution. Since nature images actually follow a Gaussian scale mixture distribution in multiscale space, the Gaussian assumption might mistreat the useful signal as disturbance and hence degrade the final fused image quality.

In this paper we propose using the DWF[4,5] to decompose the image. When it is applied to image fusion, it provides a shift-independent fusion approach. After the decomposition, the EM algorithm is used to estimate the fused image in the low frequency band since this band is usually critical to the fused image. For other frequency bands, selection considering the informative importance measure is used to fuse the images. The final fused image is then computed by taking the inverse transform on the composite coefficient representation.

This paper is organized as follows. Section 2 briefly presents the DWF. The basic idea of the proposed image fusion algorithm is also given. Section 3 describes the details of the proposed EM algorithm for fusion of the low frequency band and Section 4 gives the selection mode for other frequency band. Computer simulations are reported in Section 5. Section 6 gives the conclusion.

2 Image fusion based on DWF

The multiscale process can be summarized as follows. First, the multiscale signal decomposition is applied to all input images to produce a multiscale high frequency band representation of the input imagery. Second, a composite high frequency band representation is built by combining the multiscale coefficients of all input imagery. The fused image is then computed by taking the inverse transform on the composite coefficient representation as shown in Fig. 1. It should be noted that in the combination process of the second step, important image visual information is used for combination in the high frequency band $w_t(j)$ and the EM algorithm is applied to fuse the low frequency band components $I_N(j)$. The subscript t denotes the scale and N denotes the total level number of DWF, $t = 1, \cdots, N$.

1) Supported by National Natural Science Foundation of P. R. China (60375008), Shanghai Key Scientific Project (02DZ15001), China Ph. D. Discipline Special Foundation (20020248029) and China Aviation Science Foundation (02D57003)

Received May 26, 2004; in revised form January 18, 2005

Since the discrete wavelet transform (DWT) is a shift-dependent process, the coefficients of the DWT vary with the position of the same pixel. This variation leads to some new source of error in the fusion process. In addition, when the pixels of the source image are corrupted by noise, the power of the wavelet coefficients will be spread. This results in a degraded fusion performance.

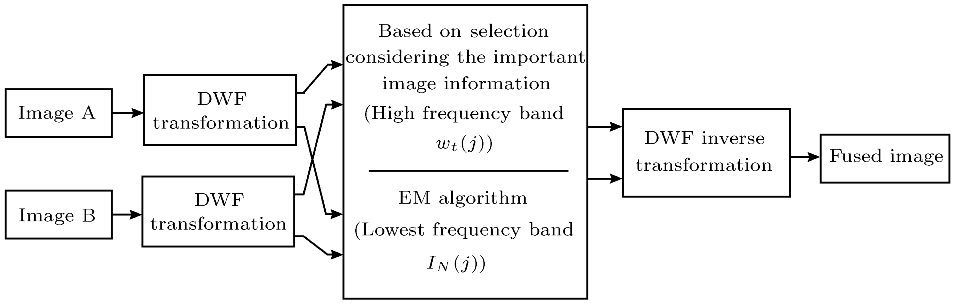


Fig. 1 The proposed multiscale image fusion scheme

To overcome this problem, the DWF is proposed here to provide a shift-invariant image fusion approach.

As in the DWT, each stage of the DWF splits the input sequence into the wavelet sequence $w_t(n)$ and the scale sequence $I_t(n)$ which are to be used as input for the next decomposition level. That is,

$$w_{t+1}(n) = \sum_k g(2^t k) I_t(n-k) \tag{1}$$

$$I_{t+1}(n) = \sum_k h(2^t k) I_t(n-k) \tag{2}$$

where $g(\cdot)$ and $h(\cdot)$ are the high and low frequency analysis filters, respectively, n denotes the pixel position, k represents the shift of the position, and t denotes the scale of the DWF. The zero level scale sequence is set to equal the input sequence $f(n)$, *i.e.*, $I_0(n) = f(n)$. In contrast to the standard DWT, the subsampling is dropped which results in a highly redundant wavelet representation. The analysis filters $g(2^t k)$ and $h(2^t k)$ at level t are obtained by inserting an appropriate number of zeros between the filter taps of the prototype filters $g(k)$ and $h(k)$.

The input sequence is reconstructed by the inverse DWF recursively as a convolution of the wavelet and scale sequences with the appropriate reconstruction filters $\tilde{g}(2^t k)$ and $\tilde{h}(2^t k)$:

$$I_t(n) = \sum_k \tilde{h}(2^t n - k) I_{t+1}(n) + \sum_k \tilde{g}(2^t n - k) w_{t+1}(n) \tag{3}$$

The reconstruction filters $\tilde{g}(k)$ and $\tilde{h}(k)$ can be achieved by using the relation of conjugate quadrature filter. That is,

$$\boldsymbol{H}(z)\tilde{\boldsymbol{H}}(z^{-1}) + \boldsymbol{G}(z)\tilde{\boldsymbol{G}}(z^{-1}) = 1 \tag{4}$$

where $\boldsymbol{H}, \tilde{\boldsymbol{H}}, \boldsymbol{G}$, and $\tilde{\boldsymbol{G}}$ are the frequency domain counterparts of the filters $\boldsymbol{h}, \tilde{\boldsymbol{h}}, \boldsymbol{g}$, and $\tilde{\boldsymbol{g}}$, respectively.

Due to the averaging inherent in both the discrete wavelet frame transform and the invert discrete wavelet frame transform, the filter coefficients from the standard DWT scheme must be slightly modified to obtain a perfect reconstruction, *i.e.*, an energy normalization has to be applied to the filters:

$$h^{DWF}(k) = \frac{1}{\sqrt{2}} h^{DWT}(k) \tag{5}$$

$$h^{DWF}(2^t k) = \frac{1}{\sqrt{2}} h^{DEF}(2^{t-1} k) \tag{6}$$

3 Fusion of the low frequency band by EM algorithm

The proposed EM method is applied to the low frequency band. The image model used is given by

$$I_i(j) = \alpha_i(j) S(j) + \beta_i(j) + \varepsilon_i(j), \quad i = 1, \cdots, q \tag{7}$$

where i represents the sensor index, j denotes the pixel location, $I_i(j)$ denotes the observed image value in the low frequency band and $S(j)$ represents the true signal. $\alpha_i(j) = \pm 1$ or 0 is the sensor selectivity factor. $\varepsilon_i(j)$ is the random noise which is modeled by a K-term mixture of Gaussian probability density function (pdf). That is,

$$f_{\varepsilon_i(j)}(\varepsilon_i(j)) = \sum_{k=1}^{K} \lambda_{k,i}(j) \frac{1}{\sqrt{2\pi\sigma_{k,i}^2(j)}} \exp\left(-\frac{\varepsilon_i(j)^2}{2\sigma_{k,i}^2(j)}\right) \tag{8}$$

From the estimation theory viewpoint, the fusion process is to estimate true scene $S(j)$ from $I_i(j)$ for $i = 1, \cdots, q$. Estimation is usually performed in a neighborhood with j as the center. In this study, a neighborhood size $L = 5 \times 5$ is selected. Inside the neighborhood, we assume the model parameters $\beta(j), (\lambda_{k,l}(j), \sigma_{k,l}^2(j))$ and $\alpha_i(j)$ are constants. For simplicity, we drop the indices of these parameters in the sequel.

When the final estimates associated with coefficient j are produced using the iterative algorithm, the estimates produced for $\boldsymbol{\alpha}_i$, $\boldsymbol{\beta}_i$ and $\{\lambda_{1,i}, \cdots, \lambda_{K,i}; \sigma_{1,i}, \cdots, \sigma_{K,i}\}$ for $i = 1, \cdots, q$ will be assigned to the corresponding parameters for coefficient j from (7), the final estimates of the scene for all coefficients neighboring coefficients j are discarded, and the final estimate of the scene for coefficient j will be assigned to $S(j)$ from (7).

At the beginning of the recursive process, it is necessary to normalize the image data as (9), so as the normalized data is limited within the boundary [0,1]

$$I_i'(j) = (I_i(j) - \mu_i)/H_i \tag{9}$$

where I_i' and I_i are the standard low frequency bands of image and source image, μ_i is the mean of the low frequency band of the source image, H_i the gray level of the source image.

The recursive process is derived from the SAGE algorithm[6,7], similar to the development in [7].

When the parameters converge to a fixed range, we can calculate the true scene as

$$S'(l) = \frac{\displaystyle\sum_{i=1}^{q}\sum_{k=1}^{K} I_i(l)\alpha_i'^2 \frac{g_{k,i,l}(I_i(l))}{\sigma_{k,i}^2}}{\displaystyle\sum_{i=1}^{q}\sum_{k=1}^{K} \alpha_i'^2 \frac{g_{k,i,l}(I_i(l))}{\sigma_{k,i}'^2}} \tag{10}$$

where l indexes the coefficient location in the small region of neighboring j.

The fused low frequency band should preserve the imaging characteristic of the sensors so as to highlight the contrast of the image. In the process to calculate the final fused low frequency band, we do not invert the contrast polarity.

Initial values for the parameters are required to start the EM algorithm. A simple estimate for the lowest frequency band of the true scene $S(l)$ comes from the weighted average of the sensor images:

$$S(l) = \sum_{i=1}^{q} w_i I_i(l) \tag{11}$$

where $w_i = 1/q$, $i = 1, \cdots, q$.

The initialization scheme has worked very well for the cases we have studied. We have observed that the algorithm generally converged in less than 5 iterations in our experiments.

4 Selection of high frequency band by informative importance measure

In the pattern-selective fusion scheme, we propose a new measure to characterize important image information, which models the low-level image information processing in a way that resembles principal components of the early retinal process[8]. A quantitative estimation of this important information can be provided by the measure of uncertainty in pixel-to-neighbors interaction. Two sources of such uncertainty must be considered: luminance uncertainty and topological uncertainty.

$$PI_{\boldsymbol{X}}(m,n) = C(m,n)I(m,n) \tag{12}$$

where $PI_{\boldsymbol{X}}(m,n)$ indicate the important information of the wavelet frame coefficient, the subscript $\boldsymbol{X}$ denotes source images. It means the source image is image $\boldsymbol{A}$ when $\boldsymbol{X} = \boldsymbol{A}$. $C(m,n)$ is the absolute value of wavelets coefficient and also reflects the luminance uncertainty; $I(m,n)$ denotes the topological uncertainty.

$$C(m,n) = |w_{\boldsymbol{X}}(m,n)| \tag{13}$$

where $w_{\boldsymbol{X}}(m,n)$ is the high frequency coefficient. We neglect the subscript t that denotes the scale as Fig. 1.

Considering the relationship of neighborhood coefficients, the sign of the high frequency coefficient is decided firstly.

$$sign(m,n) = sign(w_{\boldsymbol{X}}(m,n)) \tag{14}$$

If the high frequency coefficient is equal to or greater than zero, then *sign* is one, otherwise, *sign* is zero.

$$I(m,n) = p_{\boldsymbol{X}}(m,n)(1 - p_{\boldsymbol{X}}(m,n)) \tag{15}$$

where $p_{\boldsymbol{X}}(m,n)$ is the probability to find surrounding coefficients in the same state (sign) as the central coefficient at position (m,n).

Then, we can achieve fusion selective pattern:

$$w_F(m,n) = \begin{cases} w_A(m,n), & PI_A \geqslant PI_B \\ w_B(m,n), & \text{otherwise} \end{cases}, \quad (m,n) \in \boldsymbol{E} \tag{16}$$

where $\boldsymbol{A}$ is the source image $\boldsymbol{A}$, $\boldsymbol{B}$ is the source image $\boldsymbol{B}$ and $\boldsymbol{F}$ is the fused image $\boldsymbol{F}$.

Finally, when the fused lowest frequency band $\boldsymbol{S}_N$ and the fused high frequency band $\boldsymbol{w}_t$ are obtained, the final fused image can be achieved by performing the invert discrete wavelet frame transformation as Fig. 1. $t = 1, \cdots, N$.

5 Experiments

Performance measures are essential to determine the possible benefits of fusion as well as to compare results obtained with different algorithms. However, it is difficult to evaluate image fusion result objectively. Therefore, in this paper, three evaluation criteria are used for quantitatively assessing the performance of the fusion.

The first evaluation measure is the objective performance metric that was proposed by [9,10]. It models the accuracy with which visual information is transferred from the source images to the fused image. Important information is associated with edge information measured for each pixel. Correspondingly, by evaluating the relative amount of edge information that is transferred from the input images to the fused image, a measure of fusion performance is obtained. The larger objective performance metric means that more important information in the source images have been preserved.

Mutual information has been proposed for fusion evaluation. Given two images x_F and x_R, we define their mutual information as

$$Q(x_R; x_F) = \sum_{u=1}^{L}\sum_{v=1}^{L} h_{R,F}(u,v) \log_2 \frac{h_{R,F}(u,v)}{h_R(u)h_F(v)} \tag{17}$$

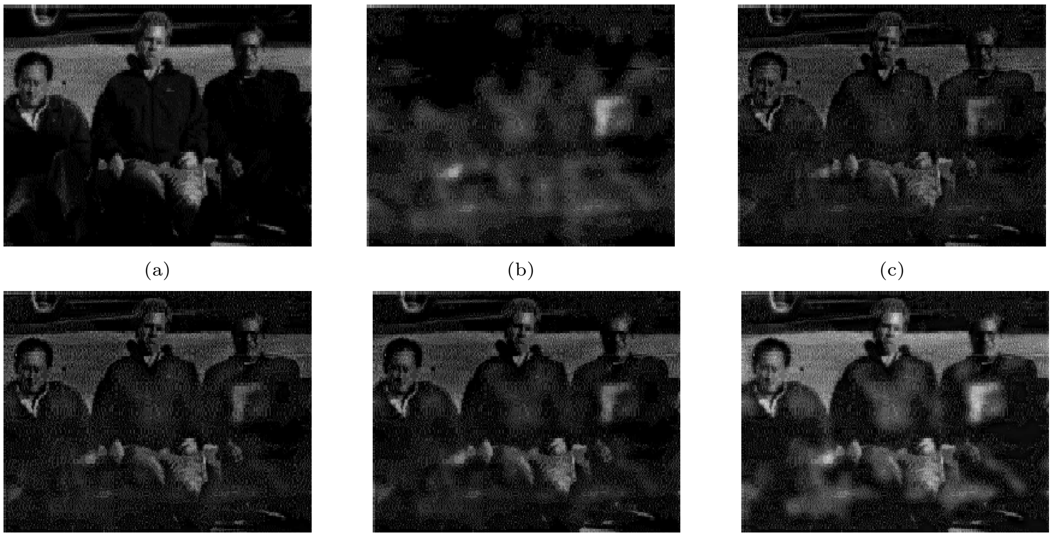
where x_R is the ideal reference, x_F the obtained fused image, h_R, h_F are the normalized graylevel histograms of x_R, x_F respectively, $h_{R,F}$ is the joint graylevel histogram of x_R and x_F, and L is the number of bins. We select $L = 100$. Thus, the higher the mutual information between x_R and x_F, the more likely x_F resembles the ideal x_R. Mutual information evaluation method may be modified into an objective measure according to [11]. This is the second evaluation measure.

The third evaluation measure is entropy:

$$EN = -\sum_{i}^{H} p_i \ln p_i \tag{18}$$

where p_i is the probability when the pixel number of gray level is i.

Figs. 2 (a) and (b) are a visual image and in MMW image employed in Concealed Weapon Detection. We can find that a weapon is concealed in the third person from Fig. 2(b). The discrete wavelet[1], steerable pyramid[12], Laplacian pyramid[1] and the method the paper proposes are performed separately on the fusion process. In all cases, we perform a 3-level decomposition. Fusion result is demonstrated as Fig. 2 (c)∼(f). It is clear that the proposed method outperformes the others. The quantity of the evaluation measurs are exhibited in Table 1 in terms of numbers of the entropy, pixel mutual information, and edge mutual information.



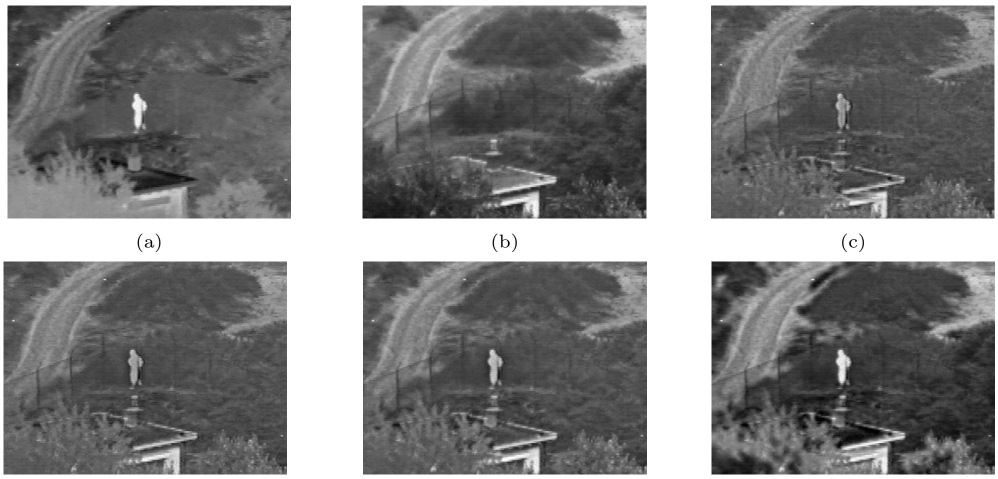
(a) (b) (c) (d) (e) (f)

Fig. 2 Fused result of visual image (a), and MMW image (b), the fusion result employing DB3 wavelet (c), steerable pyramid (d), Laplacian pyramid (e) and the proposed method (f)

Table 1 Fusion result of visual image and MMW image

Evaluation measures	Entropy	Pixel mutual information	Edge mutual information
DB3 wavelet	4.0415	0.0526	0.5465
Steerable pyramid	4.0247	0.0609	0.5864
Laplacian pyramid	4.0135	0.0621	0.6611
The proposed method	4.5912	0.1340	0.6864

Fig. 3 (b) demonstrates a visual image of a scene of which the background is the road, grass-land, fence and house. But one can hardly find the person in it (see the infrared image Fig. 3 (a)). Again, the discrete wavelet, steerable pyramid, Laplacian pyramid and the proposed method are performed separately on the fusion process. In all cases, we perform a 3-level decomposition. The fusion result as shown in Fig. 3 (c)~(f) demonstrate that the proposed method outperformes the others. The evaluation measure, are exhibited in Table 2 in terms of numbers of the entropy, pixel mutual information, and edge mutual information.



(a) (b) (c) (d) (e) (f)

Fig. 3 Fuse result of IR image (a) and visual image (b), the fusion result employing DB3 wavelet (c), Steerable pyramid (d), Laplacian pyramid (e) and the proposed method (f)

Table 2 Fusion result of IR image and visual image

Evaluation measures	Entropy	Pixel mutual information	Edge mutual information
DB3 wavelet	4.6220	0.1985	0.3719
Steerable pyramid	4.6071	0.2097	0.4232
Laplacian pyramid	4.6974	0.2543	0.4880
The proposed method	4.9835	0.2898	0.4848

6 Conclusions

In this paper, an image fusion method is proposed based on the EM (expectation maximum) algorithm and discrete wavelet frame for merging multiple sensor images. Experimental results indicate that the proposed method outperforms the discrete wavelet transform and existing wavelet frame transform. We also used three different evaluation measures of image fusion system to illustrate that the proposed method is more effective than the existing image fusion methods.

References

1 Zhang Z, Blum R S. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *Proceedings of the IEEE*, 1999, **87**(8): 1315~1326

2 Sharma R K. Probabilistic model-based multisensor image fusion. PHD dissertation, Oregon Graduate Institute, Portland, OR. 1999. 39~52

3 Yang J, Blum R S. A statistical signal processing approach to image fusion for concealed weapon detection. In: Proceedings of IEEE International Conference on Image Processing, Rochester, NY: International Conference on Image Processing, 2002. 513~516

4 Unser M. Texture classification and segmentation using wavelet frames. *IEEE Transactions on Image Processing*, 1995, **4**(11): 1549~1560

5 Rockinger O, Fechner T. Pixel-level image fusion: The case of image sequences. Proceedings of SPIE – The International Society for Optical Engineering, California, USA: San Jose, 1998. **3374**: 378~388

6 Fessler J A, Hero A O. Space-alternating generalized expectation-maximization algorithm. *IEEE Transactions on Signal Processing*, 1994, **42**(10): 2664~2677

7 Blum R S, Kozick R J, Sadler B M. An adaptive spatial diversity receiver for non-Gaussian interference and noise. *IEEE Transaction on Signal Processing*, 1999, **47**(8): 2100~2111

8 Diamant E. Image information content assessment and elicitation. *Electrical and Electronics Engineers in Israel*, 2002. 66~68

9 Petrovic V S, Xydeas C S. Sensor noise effects on signal-level image fusion performance. *Information Fusion*, 2003, **4**(3): 167~183

10 Xydeas C S, Petrovic V S. Objective image fusion performance measure. *Electronics Letters*, 2000, **36**(4): 308~309

11 Qu G, Zhang D, Yan P. Information measure for performance of image fusion, *Electronics Letters*, 2002, **38**(7): 313~315

12 Liu Z, Tsukada K, Hanasaki K, Ho Y K, Dai Y P. Image fusion by using steerable pyramid. *Pattern Recognition Letters*, 2001, **22**(9): 929~939

LIU Gang Received his master degree from Northeastern University, Liaoning, China, in 2002. He is a Ph. D. candidate in control science and engineering at Shanghai Jiaotong University. His research interests include multisensor data fusion, image fusion, and digital signal processing.

JING Zhong-Liang Received his bachelor, master, and Ph. D. degrees from Northwestern Polytechnical University, Xi′an, China in 1983, 1988, and 1994, respectively. Now he is Cheung Kong Professor of the School of Electronic, Information and Electrical Engineering, Shanghai Jiaotong University. His research interests are intelligent information processing, information fusion, target tracking, stochastic neuro-fuzzy systems, high performance motion control, and aerospace control and information processing.

SUN Shao-Yuan Received her Ph. D degree from Nanjing Science & Technology University, Nanjing, China in 2002. Her research interests include multisensor data fusion, image fusion, and digital signal processing.