Image fusion based on expectation maximization algorithm and steerable pyramid

Gang Liu (刘 刚)[1], Zhongliang Jing (敬忠良)[1], Shaoyuan Sun (孙韶媛)[1], Jianxun Li (李建勋)[1], Zhenhua Li (李振华)[1], and Henry Leung[2]

[1] *Institute of Aerospace Information and Control, School of Electronics and Information Technology, Shanghai Jiaotong University, Shanghai 200030*

[2] *ICT 451, Department of Electrical and Computer Engineering, University of Calgary, 2500, University Drive, N. W., Calgary, Alberta, Canada, T2N 1N4*

Received February 16, 2004

In this paper, a novel image fusion method based on the expectation maximization (EM) algorithm and steerable pyramid is proposed. The registered images are first decomposed by using steerable pyramid. The EM algorithm is used to fuse the image components in the low frequency band. The selection method involving the informative importance measure is applied to those in the high frequency band. The final fused image is then computed by taking the inverse transform on the composite coefficient representations. Experimental results show that the proposed method outperforms conventional image fusion methods.

OCIS codes: 100.2960, 100.7410, 070.6110, 110.3000.

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[1,2]. This fused image is usually more useful for human vision or machine perception. Most image fusion approaches are based on combining the multiresolution decomposition. The basic idea is to perform a multiresolution transform (MRT) on all source images, and construct a composite multiresolution representation of them. The fused image is then obtained by taking the inverse multiresolution transform (IMRT). Recently, estimation theory is proposed to improve the efficiency of image fusion algorithms[3,4]. However, these approaches are all based on the assumption that the disturbance follows a Gaussian distribution[5,6]. Since nature images actually follow a Gaussian scale mixture distribution in multiresolution 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 steerable pyramid[7,8] to decompose the image. When it is applied to image fusion, it provides a shift independent fusion approach. After the decomposition, the expectation maximization (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 the 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 steerable pyramid transform on the composite coefficient representation.

A steerable filters[7] of arbitrary orientation can be synthesized as a linear combination of a set of basis filters. For example, consider a two-dimensional, circularly symmetric Gaussian function in the Cartesian coordinate system. Using a bank of these steerable filters one can decompose the images into a pyramid called a steerable pyramid. It provides another degree of freedom to decompose image in arbitrary orientations as compared to the regular wavelet decompositions[1]. The multiresolution, steerable pyramid decomposition and reconstruction are then obtained by using a bank of these filters as shown in Fig. 1[8].

The multiresolution fusion process can be summarized as follows. In the first step, the multiresolution signal

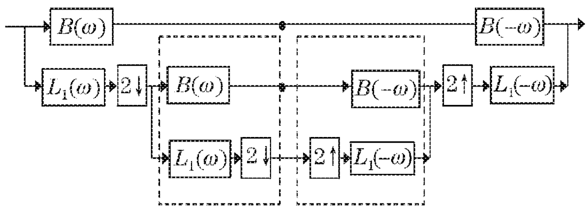

Fig. 1. Structure of steerable pyramid in frequency domain.

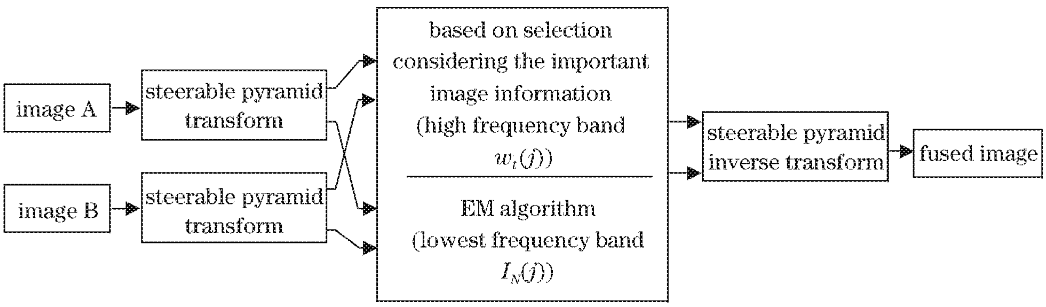

Fig. 2. The proposed multiscale image fusion scheme.

decomposition is applied to all input images, resulting in a multiresolution edge representation of the input imagery. Second, a composite multiresolution edge representation is built by a combination of the multiresolution coefficients of all input imagery. In the final step, the fused image is computed by an application of the inverse transform on the composite coefficient representation, as shown in Fig. 2.

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 ($t = 1, \cdots, N$) denotes the scale and N denotes the total level number of the steerable pyramid.

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

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

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

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

The fusion process of multiple images is in fact estimation true scene $S(j)$. The estimation process is performed in a neighborhood region of j. The region size must be carefully chosen to be large enough to allow good estimation but small enough to allow our assumptions to be reasonable. Here, we choice a size 5×5 region, and assume the model parameters $\beta(j)$, $[\lambda_{k,l}(j), \sigma_{k,l}^2(j)]$, and $\alpha_i(j)$ is constant. Here, we perform the EM fusion process on the lowest frequency part.

At the beginning of the recursive process, it is necessary to normalize the image data

$$I_i'(j) = [I_i(j) - \mu]/H, \tag{3}$$

where I_i' and I_i is the normalized image and source image, μ is the mean of the source image, and H is the gray level and is defined as the highest gray value of the source image.

The recursive process is derived from the space-alternating generalized EM (SAGE) algorithm[9], similar to the development in Ref. [10]. When the parameters converge to a fixed range, we can calculate the true scene

$$S'(l) = \frac{\sum_{i=1}^{q}\sum_{k=1}^{K} I_i(l)\alpha_i'^2 \frac{g_{k.i.l}[I_i(l)]}{\sigma_{k.i}^2}}{\sum_{i=1}^{q}\sum_{k=1}^{K} \alpha_i'^2 \frac{g_{k.i.l}[I_i(l)]}{\sigma_{k.i}^2}}, \tag{4}$$

where the dummy variable l is the scalar in the region sized 5×5, surrounding the jth pixel in the source images; $g_{k.i.l}(I_i(l))$ is the the condition probability density in the term of the ith sensor and the kth term of K-term mixture of Gaussian PDF.

In the pattern-selective fusion scheme, we proposed 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[11]. A quantitative estimation of this important information can be provided by the measure of uncertainty in pixel-to-neighbors interaction. The sources of both luminance uncertainty and topological uncertainty must be considered,

$$PI_X(m,n) = C(m,n) \cdot I(m,n), \tag{5}$$

where $PI_X(m,n)$ indicate the important information of the steerable pyramid coefficient, the subscript X denote source images. It means the source image is image A when $X = \mathrm{A}$. $C(m,n)$ is the absolute value of steerable pyramid coefficient and also reflects the luminance uncertainty, $I(m,n)$ denotes the topological uncertainty

$$C(m,n) = |w_X(m,n)|, \tag{6}$$

where $w_X(m,n)$ is the high frequency coefficient, we neglect the subscript t that denote the scale as Fig. 1. Considering the relationship of neighborhood coefficients, the sign of the high frequency coefficient is decided firstly

$$\mathrm{sign}(m,n) = \mathrm{sign}[w_X(m,n)]. \tag{7}$$

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

$$I(m,n) = p_X(m,n) \cdot [1 - p_X(m,n)], \tag{8}$$

where $p_X(m,n)$ is 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

$$w_F(m,n) = \begin{cases} w_\mathrm{A}(m,n) & PI_\mathrm{A} \geq PI_\mathrm{B} \\ w_\mathrm{B}(m,n) & \text{otherwise} \end{cases} \quad (m,n) \in E, \tag{9}$$

where A denotes the source image A, B denotes the source image B, and F is the fused image.

Finally, when the fused lowest frequency band S_N and the fused high frequency band w_t ($t = 1, \cdots, N$) is obtained, the final fused image can be achieved by performing the invert steerable pyramid transform as Fig. 1.

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 Xydeas and Petrovic[12]. Mutual information has been proposed for fusion evaluation. The second evaluation method may be modified into an objective measure according to Ref. [13]. The third evaluation measure is entropy.

Since multiple sensor images are always corrupted by noise or have register errors, the study on the robustness of image fusion system becomes important[15]. This paper proposed a measure to evaluate the robustness by using the measures described above. We call E as the relative difference

Table 1. 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	5.0015	0.2987	0.4761

Table 2. Fusion Result of Noise Corrupted IR Image and Noise Corrupted Visual Image

Evaluation Measures	Entropy		Pixel Mutual Information		Edge Mutual Information	
	Measure	E	Measure	E	Measure	E
DB3 Wavelet	5.0671	0.0963	0.3580	0.8035	0.2739	0.2635
Steerable Pyramid	5.0620	0.0987	0.3801	0.8126	0.3310	0.2179
Laplacian Pyramid	5.0853	0.0826	0.4205	0.6536	0.3541	0.2744
The Proposed Method	5.0754	0.0148	0.3821	0.2792	0.3834	0.1947

Table 3. Fusion Result of IR Image and Visual Image Corresponding with Register Error 0.5 Pixel

Evaluation Measures	Entropy		Pixel Mutual Information		Edge Mutual Information	
	Measure	E	Measure	E	Measure	E
DB3 Wavelet	4.5519	0.0152	0.0163	0.9179	0.3814	0.0255
Steerable Pyramid	4.5455	0.0134	0.0419	0.8002	0.4407	0.0414
Laplacian Pyramid	4.6265	0.0151	0.0662	0.7397	0.4964	0.0172
The Proposed Method	4.9883	0.0026	0.2549	0.1466	0.4671	0.0189

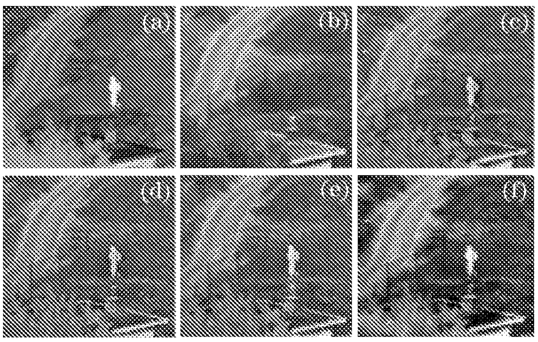

Fig. 3. Fusion result of IR image (a) and visual image (b), the fusion result employing DB3 wavelet (c), steerable pyramid (d), Laplacian pyramid (e), and the proposed method (f).

$$E = \frac{|Q' - Q|}{Q}, \tag{10}$$

where Q' denotes the evaluation measure of image fusion corrupted by noise or having register errors, Q denotes the evaluation measure of image fusion having no noise effect.

Figure 3(a) demonstrates an infrared (IR) image. Figure 3(b) demonstrates a visual image of a scene in which the background is road, grass land and fence as well as house but hardly find the person appeared in Fig. 3(a). Figures 3(c)−(f) demonstrate the fusion results employing the discrete wavelet[1], steerable pyramid based on traditional fusion scheme[14], Laplacian pyramid[1], and the proposed method. In all cases, we perform a 3-level decomposition.

Table 2 is fusion result of visual image and IR image of Fig. 3 corrupted by noise, the mean of the noise is 0 and the variance is 0.01. It indicates the relative difference is small, that is to say, the proposed method is more robust than other methods. Table 3 indicates the fusion result of IR image and visual image corresponding with register error 0.5 pixel.

In this paper, an image fusion method is proposed based on the EM algorithm and steerable pyramid for merging multiple sensor images. Experimental results indicate that the proposed method outperforms the discrete wavelet transform and traditional steerable pyramid transform. A relative difference measure is proposed to evaluate the robustness of image fusion system. The measure also illustrate that the proposed method is more robust than traditional image fusion methods.

This work was supported by National "863" Project of China (No. 2001AA135091), the 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). G. Liu's e-mail address is lukelg@sjtu.edu.cn.

References

1. Z. Zhang and R. S. Blum, Proceedings of IEEE **87,** 1315 (1999).
2. P. J. Burt and R. J. Kolczynski, in *Proceedings of 4th Int. Conf. Computer Vision* 171 (1993).
3. R. K. Sharma, *Probabilistic model-based multisensor image fusion* Ph.D Dissertation (Oregon Graduate Insti-

tute, Portland, October 1999).
4. J. Yang and R. S. Blum, in *Proceedings of IEEE International Conference on Image Processing* **1,** I/513 (2002).
5. J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, IEEE Trans. Image Processing **12,** 1338 (2003).
6. M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, Applied and Computational Harmonic Analysis **11,** 89 (2001).
7. W. T. Freeman and E. H. Adelson, IEEE Trans. Pattern Anal. Mach. Intell. **13,** 891 (1991).
8. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, IEEE Trans. Infrom. Theory **38,** 587 (1991).
9. J. A. Fessler and O. A. Hero, IEEE Trans. Signal Processing **42,** 2664 (1994).
10. R. S. Blum, R. J. Kozick, and B. M. Sadler, IEEE Trans. Signal Processing **47,** 2100 (1999).
11. E. Diamant, Proc. SPIE **5014,** 20 (2003).
12. C. S. Xydeas and V. Petrovic, Electron. Lett. **36,** 308 (2000).
13. G. H. Qu, D. L. Zhang, and P. F. Yan, Electron. Lett. **38,** 313 (2002).
14. Z. Liu, K. Tsukada, K. Hanasaki, Y. K. Ho, and Y. P. Dai, Pattern Recognition Lett. **22,** 929 (2001).
15. V. S. Petrovic and C. S. Xydeas, Information Fusion **4,** 167 (2003).