ELSEVIER

Information Fusion 2 (2001) 169–176

INFORMATION FUSION

www.elsevier.com/locate/inffus

Combination of images with diverse focuses using the spatial frequency

Shutao Li [a,b,*], James T. Kwok [b], Yaonan Wang [a]

[a] College of Electrical and Information Engineering, Hunan University, Changsha, China
[b] Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong, China

Received 18 July 2000; received in revised form 24 May 2001; accepted 15 June 2001

Abstract

Image fusion attempts to combine complementary information from multiple images of the same scene, so that the resultant image is more suitable for human visual perception and computer-processing tasks such as segmentation, feature extraction and object recognition. This paper presents an approach that fuses images with diverse focuses by first decomposing the source images into blocks and then combining them by the use of spatial frequency. The algorithm is computationally simple and can be implemented in real-time applications. Experimental results show that the proposed method is superior to wavelet transform based methods in both objective and visual evaluations. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Image fusion; Multisensor fusion; Spatial frequency; Wavelet transform

1. Introduction

Recently, image fusion has become an important topic in image analysis and computer vision [1,2,8,14]. Image fusion refers to image processing techniques that produce a new, enhanced image by combining images from two or more sensors. The fused image is then more suitable for human/machine perception, and for further image-processing tasks such as segmentation, feature extraction and object recognition.

The simplest image fusion method just takes the pixel-by-pixel gray level average of the source images. This, however, often leads to undesirable side effects such as reduced contrast. In recent years, various methods based on the multiscale transforms have been proposed to address this issue. The basic idea is to perform a multiresolution decomposition on each source image, then integrate all these decompositions to form a composite representation, and finally reconstruct the fused image by performing an inverse multiresolution transform. Examples of this approach include the Laplacian pyramid [3], the gradient pyramid [4], the ratio-of-low-pass pyramid [12], the morphological pyramid [10], and the wavelet transform [6,9,13].

In this paper, we consider situations where two or more objects in the scene are at different distances from the camera. As is typical with most inexpensive cameras, the image thus obtained will not be in focus everywhere, i.e., if one object is in focus, another one will be out of focus. However, by fusing images with different focus points, an image that is in focus everywhere can be obtained [11]. In this paper, an efficient pixel level image fusion algorithm based on the spatial frequency is proposed. This method is computationally simple and can be used in real-time applications. Moreover, the resultant fused images are both quantitatively and visually superior to those produced by the wavelet transform methods.

An important preprocessing step in image fusion is image registration, i.e., corresponding pixel positions in the source images must refer to the same location. In this paper, we will focus on the fusion issue and the source images are assumed to have already been registered.

The rest of this paper is organized as follows. A brief introduction to the spatial frequency will be given in Section 2. The proposed fusion scheme will be described in Section 3. Experimental results will be presented in Section 4, and the last section gives some concluding remarks.

* Corresponding author. Tel.: +852-2358-7027; fax: 852-2358-1477.
E-mail addresses: shutaoli@cs.ust.hk, shutao_li@yahoo.com.cn (S. Li), jamesk@cs.ust.hk (J.T. Kwok), yaonan@mail.hunu.edu.cn (Y. Wang).

PII: S1566-2535(01)00038-0

2. Spatial frequency

Spatial frequency measures the overall activity level in an image [7]. For an $M \times N$ image block F, with gray value $F(m,n)$ at position (m,n), the spatial frequency is defined as

$$SF = \sqrt{(RF)^2 + (CF)^2}, \tag{1}$$

where RF and CF are the row frequency

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

and column frequency

$$CF = \sqrt{\frac{1}{MN}\sum_{n=1}^{N}\sum_{m=2}^{M}[F(m,n) - F(m-1,n)]^2}, \tag{3}$$

respectively.

Fig. 1(a) shows a 64×64 image block extracted from the 'Lena' image. Fig. 1(b)–(e) show the degraded versions after blurring with a Gaussian of radius 0.5, 0.8, 1.0 and 1.5, respectively. As can be seen from Table 1, when the image gets more blurred, the spatial frequency diminishes accordingly. Another experiment on an image block extracted from the 'Peppers' image also produced similar results (Fig. 2 and Table 2). These demonstrate that the spatial frequency can be used to reflect the clarity of an image.

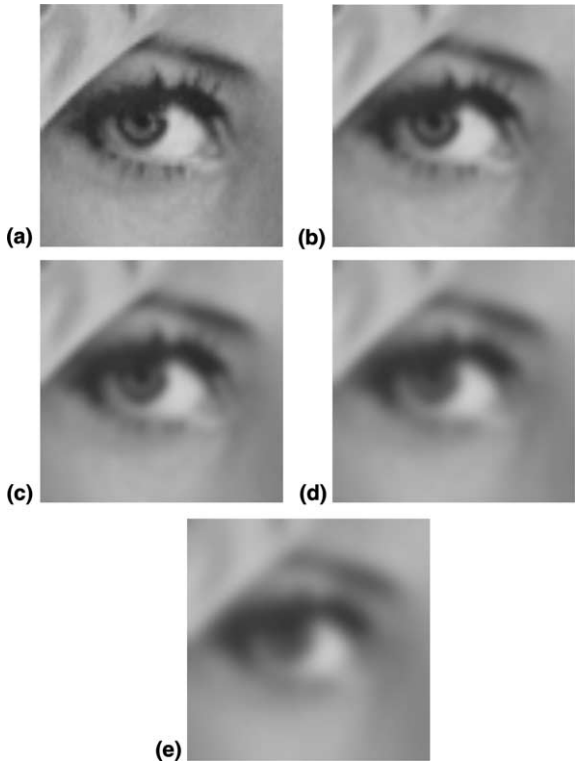

Fig. 1. Original and blurred versions of an image block extracted from the 'Lena' image: (a) original image; (b) radius = 0.5; (c) radius = 0.8; (d) radius = 1.0; (e) radius = 1.5.

Table 1
Spatial frequencies of the image blocks in Fig. 1

	Fig. 1(a)	Fig. 1(b)	Fig. 1(c)	Fig. 1(d)	Fig. 1(e)
SF	16.10	12.09	9.67	8.04	6.49

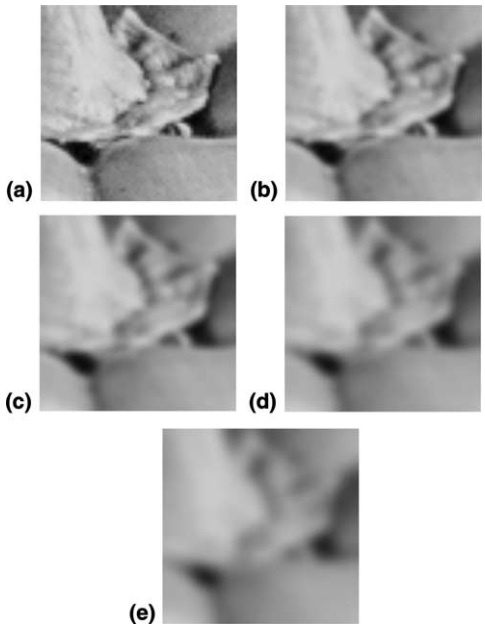

Fig. 2. Original and blurred versions of an image block extracted from the 'peppers' image: (a) original image; (b) radius = 0.5; (c) radius = 0.8; (d) radius = 1.0; (e) radius = 1.5.

Table 2
Spatial frequencies of the image blocks in Fig. 2

	Fig. 2(a)	Fig. 2(b)	Fig. 2(c)	Fig. 2(d)	Fig. 2(e)
SF	28.67	17.73	12.98	10.04	7.52

3. Multifocus image fusion

Fig. 3 shows a schematic diagram for the proposed multifocus image fusion method. Here, we only consider the fusing of two source images, though it can be extended straightforwardly to handle more than two images. The algorithm consists of the following steps:

1. Decompose the source images A and B into blocks of size $M \times N$. Denote the ith blocks of A and B by A_i and B_i, respectively.
2. Compute the spatial frequency of each block, and denote the spatial frequencies of A_i and B_i by SF_i^A and SF_i^B, respectively.

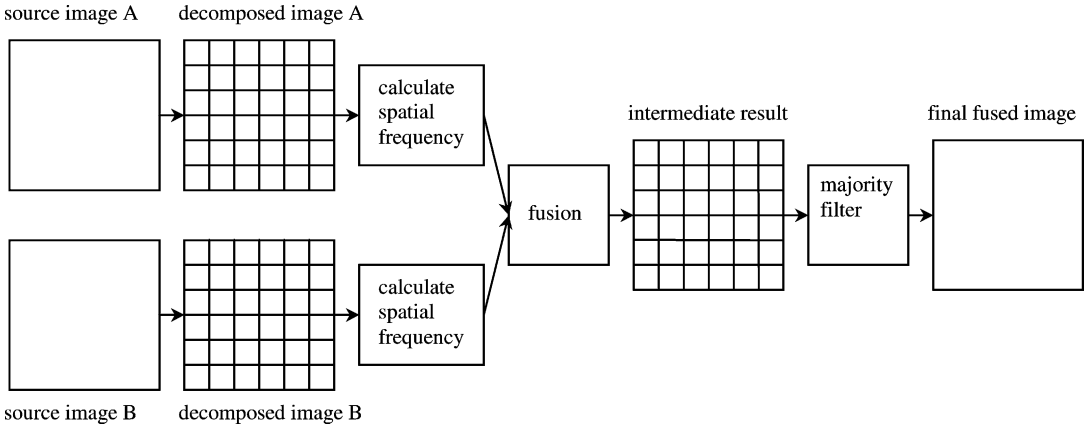

Fig. 3. Schematic diagram for multifocus image fusion.

3. Compare the spatial frequencies of two corresponding blocks A_i and B_i, and construct the ith block F_i of the fused image as

$$F_i = \begin{cases} A_i, & SF_i^A > SF_i^B + TH, \\ B_i, & SF_i^A < SF_i^B - TH, \\ (A_i + B_i)/2, & \text{otherwise.} \end{cases} \quad (4)$$

Here, TH is a user-defined threshold, and $(A_i + B_i)/2$ means taking the pixel-by-pixel gray level average of A_i and B_i.

4. Verify and correct the fusion result in step 3: specifically, if the center block comes from A but the majority of its surrounding blocks are from B, then this center block will be changed to be from B, and vice versa. In the implementation, we use a majority filter together with a 3×3 window.

4. Experimental results

4.1. Setup for quantitative evaluation

Experiments have been performed on two sets of images. Each original (reference) image (Figs. 4(a) and 5(a)) contains two or more objects, and are in good focus everywhere. Their sizes are 128×128 and 480×640, respectively. We artificially produce a pair of out-of-focus images from each reference image, by first blurring one object to obtain an image, and then blurring another object to produce a second image. Figs. 4(b) and (c) show the two images obtained by blurring Fig. 4(a) with a Gaussian of radius 0.3 and 0.5, respectively. Similarly, Figs. 5(b) and (c) are obtained by blurring Fig. 5(a) with a Gaussian of radius 6 and 2, respectively.

The root mean square error (RMSE) is used as the evaluation criterion. For reference image R and fused image F (both of size $I \times J$), the RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{I} \sum_{j=1}^{J} [R(i,j) - F(i,j)]^2}{I \times J}}, \quad (5)$$

where $R(i,j)$ and $F(i,j)$ are the pixel values at position (i,j) of R and F, respectively. Mutual information has also been used. But the results are similar to RMSE and so will not be reported here.

4.2. Effects of the block size and threshold

Table 3 shows the RMSEs' obtained on fusing images in Figs. 4(b) and (c) with different block sizes and thresholds. The best result (Fig. 4(d)) corresponds to the use of 8×8 blocks and a threshold of 1.75. Table 4 shows the results on fusing images in Figs. 5(b) and (c),

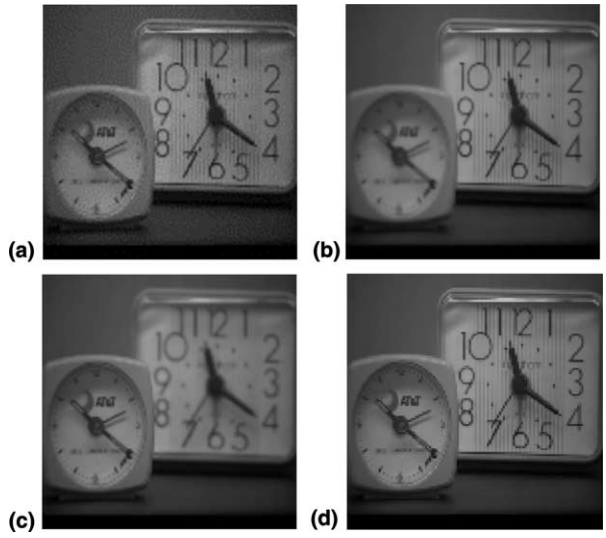

Fig. 4. The reference image, blurred images and fusion result: (a) reference image; (b) blurred image, with focus on the right; (c) blurred image, with focus on the left; (d) fused image (block size = 8×8, $TH = 1.75$).

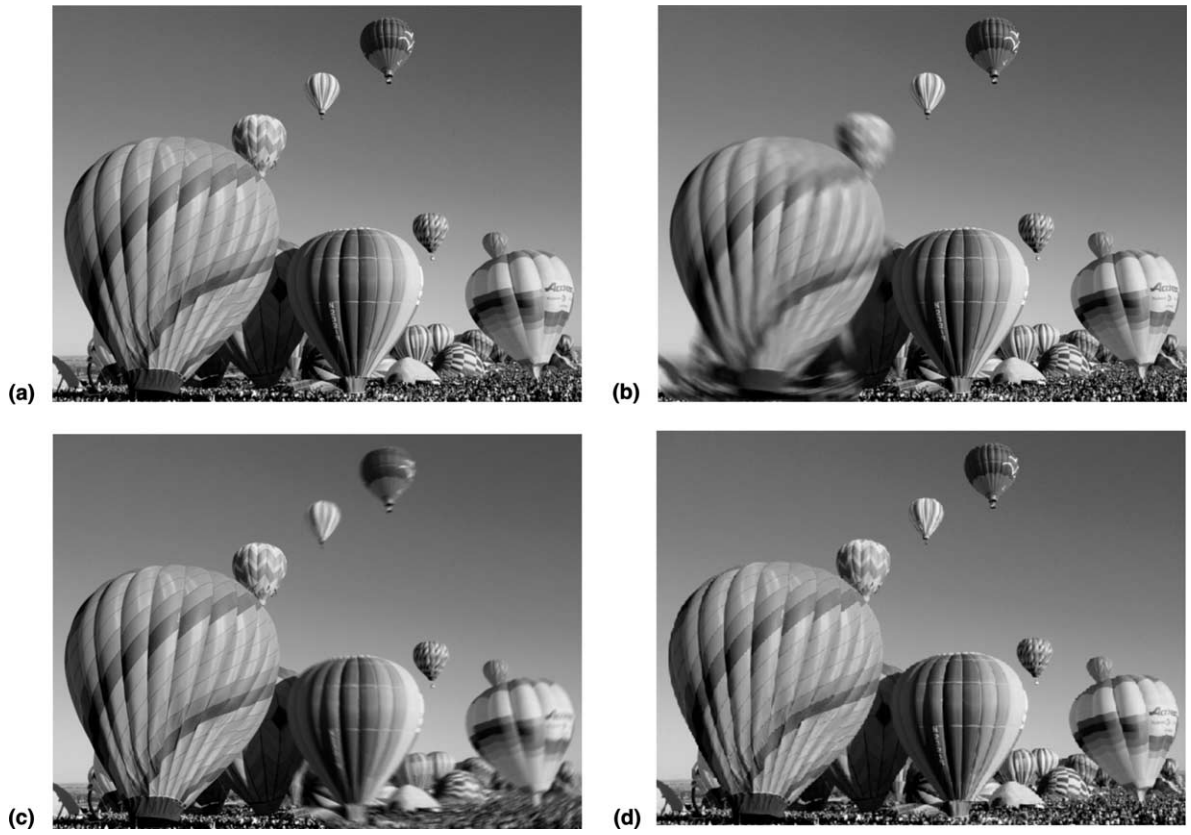

Fig. 5. The reference image, blurred images and fusion result: (a) reference image; (b) blurred image, with focus on the right; (c) blurred image, with focus on the left; (d) fused image (block size $= 32 \times 32$, $TH = 0.50$).

Table 3
Effects on fusing images in Figs. 4(b) and (c) with different block sizes and thresholds[a]

TH	Block size								
	4×4	4×8	8×8	8×16	16×16	16×32	32×32	32×64	64×64
0.00	0.8064	0.6339	**0.5166**	0.6902	0.5549	1.3832	1.8363	1.9682	1.9682
0.25	0.8038	0.6136	**0.4940**	0.6699	0.5550	1.3832	1.8363	1.9682	1.9682
0.50	0.7821	0.6104	**0.4634**	0.6710	0.5550	1.3856	1.3264	1.9682	1.9682
0.75	0.7467	0.6503	**0.4822**	0.6821	0.5550	1.3883	1.3352	1.9682	1.9682
1.00	0.7495	0.5916	**0.4841**	0.6367	0.5706	1.3883	1.3428	1.9793	1.9682
1.25	0.7462	0.5924	**0.4718**	0.6307	0.5903	1.3990	1.3428	1.9793	1.9682
1.50	0.7462	0.5924	**0.4718**	0.6307	0.5903	1.3990	1.3428	1.9793	1.9682
1.75	0.7394	0.6069	**0.4493**	0.6384	0.6295	1.4137	1.3428	1.9793	1.9682
2.00	0.7498	0.6234	**0.4597**	0.6352	0.6400	1.4267	1.3428	1.9793	1.9682
2.25	0.7556	0.6265	**0.4764**	0.6633	0.6626	1.4267	1.3428	1.9793	1.9682
2.50	0.7300	0.5838	**0.4896**	0.7014	0.6716	1.4267	1.3428	1.6003	1.9682
2.75	0.7588	0.5997	**0.5316**	0.7014	0.6716	1.4267	1.3428	1.6003	1.9682
3.00	0.7586	0.6545	**0.5580**	0.7014	0.7034	1.4267	1.3428	1.6003	1.6709

[a] Numbers in bold indicate the lowest RMSE obtained over different block sizes for a given threshold.

and the best result (Fig. 5(d)) corresponds to the use of 32×32 blocks and a threshold of 0.5.

As can be seen from Tables 3 and 4, the optimal block size is image-dependent for a given threshold. In general, using a block size too large or too small is undesirable, and this is demonstrated here with the second set of images (Figs. 5(b) and (c)). Figs. 6(a)–(c) show parts of the fused images obtained with block sizes of 4×4, 32×32 and 120×128, respectively, at a threshold value of 0.5. With 4×4 blocks, the fused image contains saw-tooth edges. On using 120×128 blocks, the fused image suffers from uneven gray level distribution. Experiment with a block size of 32×32 yields the best result in this case.

On the other hand, the threshold setting has a relatively mild effect on fusion performance. From Tables 3

Table 4
Effects on fusing images in Figs. 5(b) and (c) with different block sizes and thresholds[a]

TH	Block size									
	4 × 4	4 × 8	8 × 8	8 × 16	16 × 16	16 × 32	32 × 32	32 × 40	40 × 40	40 × 64
0.0	3.1640	1.9681	1.2564	0.7050	0.5864	**0.2107**	0.2212	0.4901	0.4313	0.8010
0.25	3.1546	1.9186	1.2524	0.6516	0.5883	**0.2099**	0.2178	0.4212	0.4354	0.6836
0.50	3.1228	1.8898	1.2688	0.6682	0.5978	0.2403	**0.1862**	0.4239	0.4418	0.6882
0.75	3.0728	1.8772	1.2569	0.7033	0.5538	0.2759	**0.2388**	0.4440	0.4597	0.6975
1.0	3.0711	1.8356	1.2484	0.6796	0.5868	0.3141	**0.2813**	0.4846	0.4789	0.7121
1.25	3.0529	1.8403	1.2563	0.6849	0.6058	0.3526	**0.3206**	0.5311	0.5292	0.7243
1.50	3.0478	1.8488	1.2765	0.7187	0.6386	0.4175	**0.3776**	0.5587	0.5571	0.8525
1.75	3.0048	1.7744	1.2638	0.7393	0.6575	0.4412	**0.3977**	0.5587	0.5722	0.8525
2.00	2.9863	1.7737	1.2375	0.7528	0.6763	0.4412	**0.4023**	0.5787	0.5950	0.8837
2.25	2.9544	1.7611	1.2441	0.7949	0.6927	0.4644	**0.4199**	0.7129	0.6128	0.9114
2.50	2.9498	1.7643	1.2717	0.8071	0.7120	0.5012	**0.4625**	0.7289	0.7831	0.9114
2.75	2.9503	1.7718	1.2833	0.8179	0.7417	**0.5170**	0.6449	0.7611	0.8276	0.9186
3.00	2.9368	1.7539	1.3050	0.8523	0.7676	**0.5558**	0.6898	0.8044	0.8276	0.9525

[a] Numbers in bold indicate the lowest RMSE obtained over different block sizes for a given threshold.

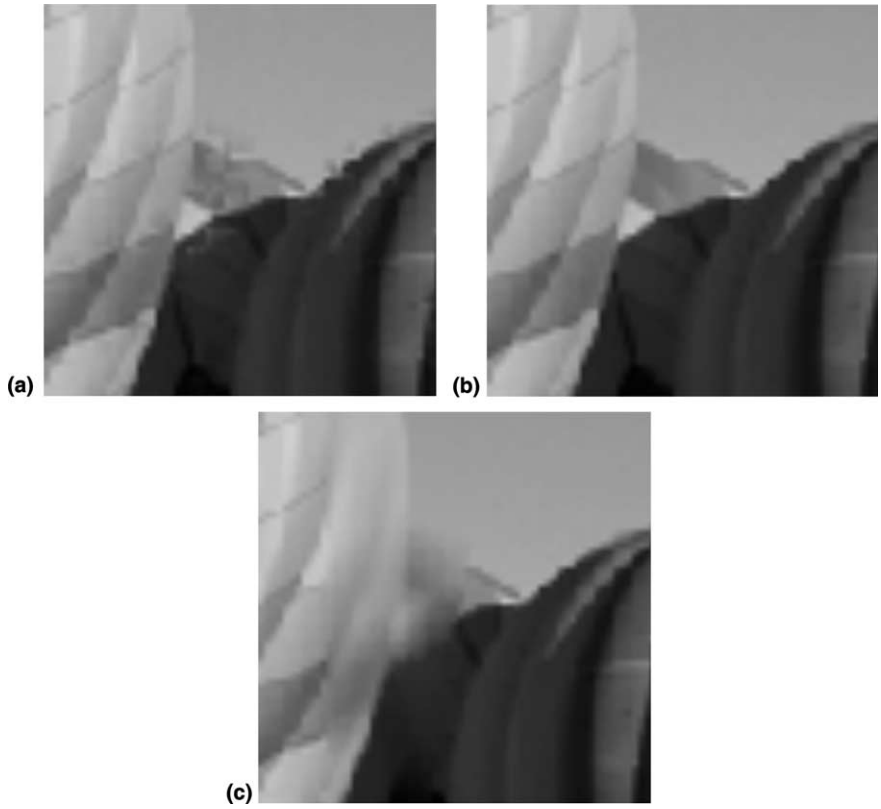

Fig. 6. Effects of using different block sizes on fusion performance: (a) block size = 4 × 4; (b) block size = 32 × 32; (c) block size = 120 × 128.

Table 5
Results on fusing images in Figs. 5(b) and (c) using the wavelet transform[a]

Wavelet	Decomposition level						
	1	2	3	4	5	6	7
Db4	7.3326	5.0334	2.9996	2.1537	2.0655	2.0652	**2.0598**
Db10	7.4248	5.0764	2.8370	2.0179	1.9483	1.9408	**1.9400**
Coif5	7.5434	5.0950	2.7625	1.8964	1.7911	1.7887	**1.7821**
Sym8	7.4810	5.0769	2.9147	1.9943	1.9044	1.9036	**1.9016**
Bior3.5	9.1313	8.0289	5.3517	3.3348	2.9505	**2.8909**	**2.8909**

[a] Numbers in bold indicate the lowest RMSE obtained over different decomposition levels for a given basis.

and 4, threshold values in the range 0.5–2.0 often produce good results. Moreover, notice from Eq. (4) that if the threshold is set too large, the algorithm degenerates to pixel-by-pixel gray level averaging.

4.3. Comparison with wavelet transform based methods

In this section, we compare the proposed method with the wavelet transform method. Five wavelet filter banks, Daubechies 'db4' and 'db10', Coiflets 'coif5', Symlets 'sym8', and Biorthogonal 'bior3.5' [5], with decomposition levels from 1 to 7, are used to fuse images in Figs. 5(c) and (d). Following [13], we employ a region-based activity measurement for the active level of the decomposed wavelet coefficients, a maximum selection rule for coefficient combination, together with a window-based consistency verification scheme. The optimal RMSE in Table 5, obtained by using the 'coif5' wavelet filter bank at a decomposition level of 7, is still inferior to most of the results in Table 4.

4.4. Subjective evaluation

In this section, we use images (Figs. 7–9) that contain multiple objects at different distances from the camera. One or more objects thus naturally become(s) out of focus when the image is taken. For example, the focus in Fig. 7(a) is on the Pepsi can, while that in Fig. 7(b) is on the testing card. The 'true' gray value of each pixel is, however, not available and so only a subjective comparison is intended here.

Results on using the wavelet based approach and the proposed algorithm for fusing Figs. 7(a) and (b) are shown in Figs. 7(c) and (d), respectively. Notice that the tiny area above the string 'Re' in the upper right corner of Fig. 7(c) is blurred, while that in Fig. 7(d) is clear. Also, the table edge in Fig. 7(d) is smooth, as in the source images, while that in Fig. 7(c) contains wrinkles. Similar phenomena can also be observed in Figs. 8 and 9.

5. Conclusion

In this paper, we propose a new approach for multifocus images fusion by using the spatial frequency. The method is computationally simple and can be used in real-time applications. Extensive experiments on studying the fusion performance with different block sizes and thresholds have been made. Comparison with wavelet transform based methods

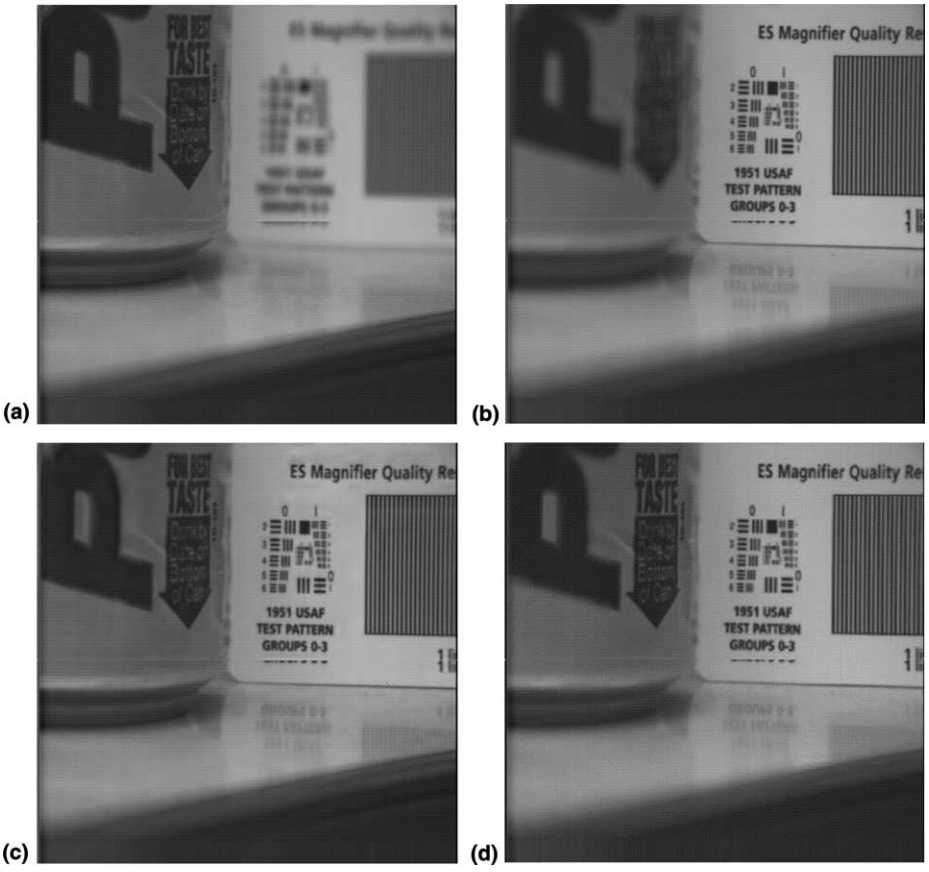

Fig. 7. The 'Pepsi' source images (size = 512 × 512) and fusion results: (a) focus on the Pepsi can; (b) focus on the testing card; (c) fused image using the wavelet transform (coif5 with decomposition level of 6); (d) fused image using the proposed algorithm (block size = 32 × 32, $TH = 1.0$).

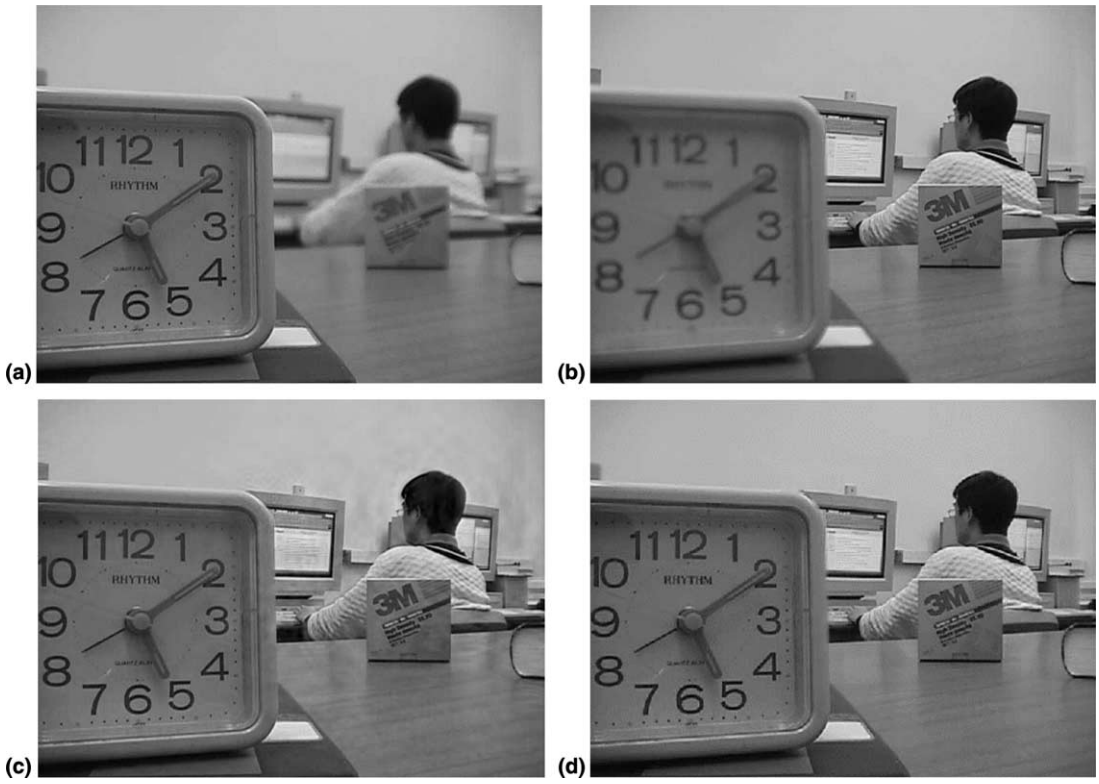

Fig. 8. The 'Lab' source images (size = 480 × 640) and fusion results: (a) focus on the clock; (b) focus on the student; (c) fused image using the wavelet transform (db8 with decomposition level of 5); (d) fused image using the proposed algorithm (block size = 40 × 40, $TH = 0.75$).

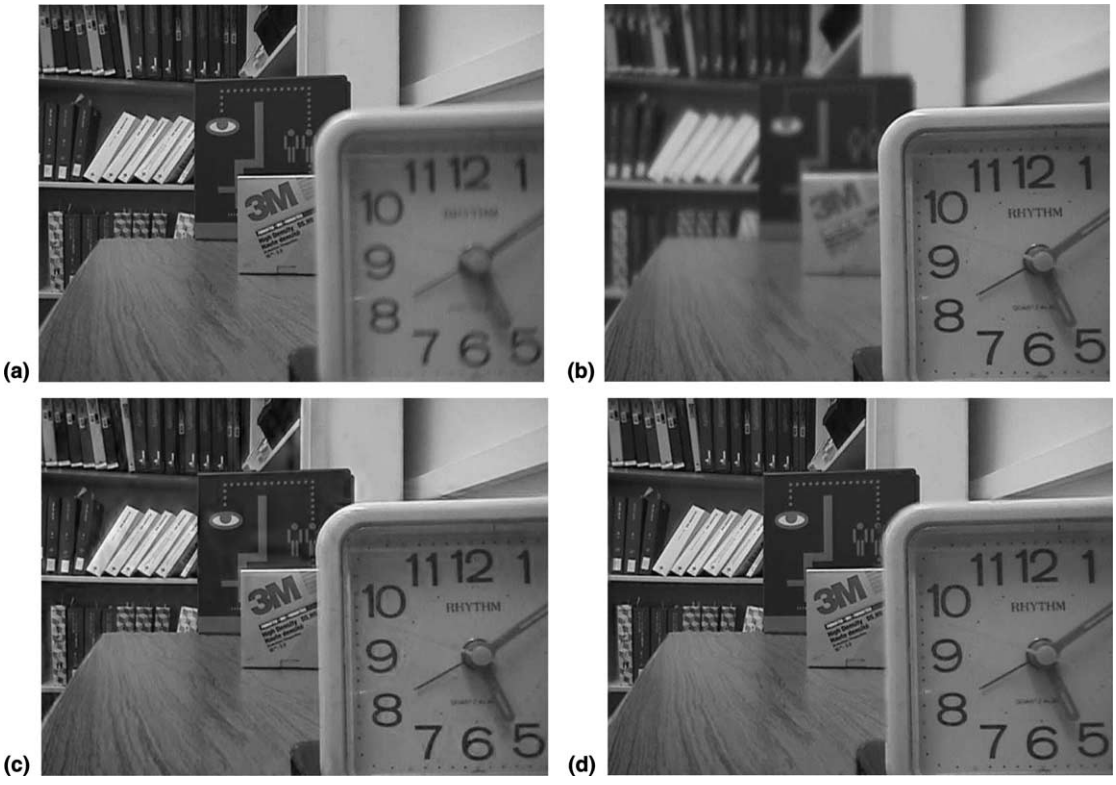

Fig. 9. The 'Disk' source images (size = 480 × 640) and fusion results: (a) focus on the left; (b) focus on the right; (c) fused image using the wavelet transform (db8 with decomposition level of 5); (d) fused image using the proposed algorithm (block size = 32 × 32, $TH = 0.75$).

shows that the proposed method is superior both quantitatively and visually.

A number of issues still need to be addressed in the future. First, adaptive methods for choosing the block size and threshold need further investigation. We will also study how the proposed method can be extended to other image fusion tasks, such as the fusion of visual images with thermal or millimeter-wave images.

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive comments on an earlier version of this paper, and also to thank Prof. Rick S. Blum for providing the multifocus images.

References

[1] J.K. Aggarwal, Multisensor Fusion for Computer Vision, Springer, Berlin, 1993.
[2] A. Akerman, Pyramid techniques for multisensor fusion, Proc. SPIE 2828 (1992) 124–131.
[3] P.T. Burt, E.H. Andelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540.
[4] P.T. Burt, R.J. Kolczynski, Enhanced image capture through fusion, in: Proceedings of the 4th International Conference on Computer Vision, Berlin, Germany, May 1993, pp.173–182.
[5] I. Daubechies, Ten lectures on wavelets, SIAM, CBMS-NSF Lecture Notes, vol. 61, 1992.
[6] D.A. Yocky, Image merging and data fusion by means of the discrete two-dimensional wavelet transform, J. Opt. Soc. Am. A: Opt., Image Sci. Vision 12 (9) (1995) 1834–1841.
[7] A.M. Eskicioglu, P.S. Fisher, Image quantity measures and their performance, IEEE Trans. Commun. 43 (12) (1995) 2959–2965.
[8] D.L. Hall, J. Llinas, An introduction to multisensor data fusion, Proc. IEEE 85 (1) (1997) 6–23.
[9] H. Li, B.S. Manjunath, S.K. Mitra, Multisensor image fusion using the wavelet transform, Graphical Models Image Process. 57 (3) (1995) 235–245.
[10] G.K. Matsopoulos, S. Marshall, J.N.H. Brunt, Multiresolution morphological fusion of MR and CT images of the human brain, Proc. IEE: Vision, Image Signal Process. 141 (3) (1994) 137–142.
[11] W.B. Seales, S. Dutta, Everywhere-in-focus image fusion using controllable cameras, Proc. SPIE 2905 (1996) 227–234.
[12] A. Toet, L.J. van Ruyven, J.M. Valeton, Merging thermal and visual images by a contrast pyramid, Opt. Eng. 28 (7) (1989) 789–792.
[13] Z. Zhang, R.S. Blum, A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application, Proc. IEEE 87 (8) (1999) 1315–1326.
[14] P.K. Varshney, Multisensor data fusion, Electron. Commun. Eng. J. 9 (6) (1997) 245–253.