Fusing Images with Multiple Focuses using Support Vector Machines

Shutao Li[1,2] James T. Kwok[1] Yaonan Wang[2]

[1] Department of Computer Science
Hong Kong University of Science and Technology
Hong Kong, People's Republic of China
[2] College of Electrical and Information Engineering
Hunan University
Changsha, People's Republic of China

Abstract. Optical lenses, particularly those with long focal lengths, suffer from the problem of limited depth of field. Consequently, it is often difficult to obtain good focus for all the objects in the scene. One approach to address this problem is by performing image fusion, i.e., several pictures with different focus points are combined to a single image. This paper proposes a multifocus image fusion method based on the discrete wavelet frame transform and support vector machines. Experimental results show that the proposed method outperforms the conventional approach based on the discrete wavelet transform and maximum selection rule, particularly when there is slight camera/object movement or mis-registration of the source images.

1 Introduction

Optical lenses suffer from the problem of limited depth of field. According to the lens formula, only objects at one particular depth will be truly in focus for any fixed focus setting. Consequently, if one object in the scene is in focus, another object at a different distance from the lens will be out of focus and thus blurred. In general, many vision-related processing tasks, such as edge detection, image segmentation and stereo matching, can be more easily performed on focused images than on defocused ones. Hence, it is often advantageous if an everywhere-in-focus image can be recovered.

One approach to address this problem is by first estimating from the image the distance between the sensor and the objects in the scene. However, this requires a knowledge of the values of the camera parameters and/or a model of the camera point spread function, and thus involves a lot of mundane camera calibration. In this paper, we follow another approach by performing image fusion [5], in which several pictures with different focus points are combined to form a single image. If each relevant object is in good focus in at least one of the source images, the fused image will then hopefully have all these objects in focus.

The simplest fusion method just takes the pixel-by-pixel average of the source images. This, however, often produces undesirable side effects such as reduced contrast. To alleviate this problem, methods based on multiscale transforms (such as the popular discrete wavelet transform (DWT)) have been employed. The basic idea is to perform a multiresolution decomposition on each source image, then integrate all these decompositions to produce one composite representation, from which the fused image can be recovered by performing an inverse transform. However, one problem with these transforms is that their multiresolution decompositions, and consequently the fusion results, are not shift-invariant because of an underlying down-sampling process. Hence, their performance quickly deteriorates when there is slight camera/object movement or mis-registration of the source images.

Moreover, the *maximum selection* rule is usually used to fuse the multiresolution decompositions. Consider DWT as an example. After obtaining the sets of wavelet coefficients from the source images, corresponding coefficients are compared and the one with the largest absolute value is selected for use in the composite representation [2]. The rationale is that large absolute coefficients often correspond to salient features in the images. However, obviously this simple selection rule does not always work.

In this paper, we use the discrete wavelet frame transform (DWFT) [1] (also called the stationary wavelet transform [3]) to perform multiresolution decomposition, and then use the support vector machine (SVM) [4] to fuse the wavelet frame coefficients. DWFT is closely related to DWT but avoids down-sampling by using an overcomplete wavelet decomposition. Its resultant signal representation is thus both aliasing free and shift-invariant. The SVM, on the other hand, has outperformed many conventional pattern recognition approaches. The rest of this paper is organized as follows. Section 2 describes our proposed fusion scheme. Experimental results are presented in Section 3, and the last section gives some concluding remarks.

2 Proposed Fusion Scheme

Figure 1 shows a schematic diagram of our proposed multifocus image fusion method. Here, we consider the processing of just two source images (A and B), but the algorithm can be straightforwardly extended to handle more than two images. Moreover, the source images are assumed to have been registered. The algorithm consists of the following steps:

1. Decompose each source image by DWFT to d levels, resulting in a total of $3d$ details subbands and one approximation subband. Recall that the approximation subband is a low-pass filtered version of the original image. Hence, with a sufficient number of decompositions, the resultant approximation subband contains little information and so will not be used in constructing the feature vector in Step 2 (though it will still be used in reconstructing the fused image in Step 6).
2. For each details subband, compute at each pixel location a window-based activity level based on the wavelet coefficients. Specifically, consider the $b_1 b_2$

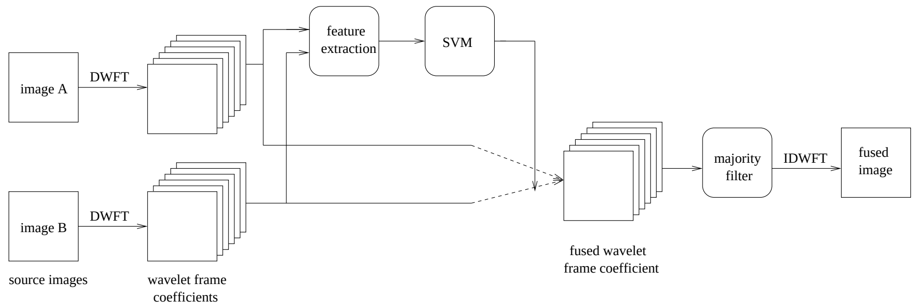


Fig. 1. Schematic diagram of the proposed fusion method.

details subband (where $b_1 b_2$ can be either HL, LH or HH) at decomposition level k, and with the wavelet coefficient at position (i,j) denoted by $D_{b_1 b_2, k}(i,j)$. Define a $p \times p$ window W_{ij} centered at (i,j), with weighting coefficients $w(u,v)$ satisfying $\sum_{(u,v)\in W_{ij}} w(u,v) = 1$. In Section 3, the following 3×3 window will be used:

$$\begin{bmatrix} \frac{1}{16} & \frac{1}{16} & \frac{1}{16} \\ \frac{1}{16} & \frac{1}{2} & \frac{1}{16} \\ \frac{1}{16} & \frac{1}{16} & \frac{1}{16} \end{bmatrix}.$$

The corresponding feature value is then computed as

$$A_{b_1 b_2, k}(i,j) = \sum_{(u,v)\in W_{ij}} w(u,v) |D_{b_1 b_2, k}(i,j)|. \tag{1}$$

Notice that this can be defined at *every* pixel location because the details subbands of DWFT are of the same size as the input image[3].

3. Train a SVM to determine whether the coefficients from A or B should be used in position (i,j) of the fused image. Here, we use the difference vector

$$\{A_{HL,k}(i,j) - B_{HL,k}(i,j), A_{LH,k}(i,j) - B_{LH,k}(i,j), A_{HH,k}(i,j) - B_{HH,k}(i,j)\}_k$$

as input, and the output is labeled according to

$$target(i,j) = \begin{cases} 1 & \text{if } A(i,j) \text{ is clearer than } B(i,j), \\ -1 & \text{otherwise.} \end{cases}$$

4. Perform testing of the trained SVM over the whole image. If the output of the SVM at a pixel location is positive, coefficients for all the details and approximation subbands of the fused image at this particular location will come from A, and vice versa. Denote the SVM's output at (i,j) by $out(i,j)$. Then, in other words, the fused coefficient at (i,j) is given by

$$C_{b_1 b_2, k}(i,j) = \begin{cases} A_{b_1 b_2, k}(i,j) & \text{if } out(i,j) > 0, \\ B_{b_1 b_2, k}(i,j) & \text{otherwise,} \end{cases}$$

[3] On the contrary, the sizes of the subbands in DWT decrease as the decomposition goes on, and so (1) would not be well-defined if DWT were used.

for all subbands $b_1 b_2 \in \{LL, HL, LH, HH\}$ and all decomposition levels k.

5. Perform consistency verification on the fusion result obtained in Step 4. Specifically, if the SVM decides that the coefficient of a particular pixel is to come from A but the majority of its surrounding coefficients are from B, then this coefficient will also be switched to come from B. In the implementation, a majority filter together with a 5×5 window are used.
6. Finally, recover the fused image by performing the inverse DWFT.

3 Experiments

In this experiment, a reference image with good focus everywhere is used. We then produce a pair of out-of-focus images by blurring this reference image, and perform restoration using various fusion schemes. Because of the availability of the reference image, quantitative comparison of the fusion schemes is possible. In the following, two evaluative criteria, namely,

1. the root mean squared error (RMSE) between the reference image R and the fused image F, and
2. the mutual information $MI = \sum_{i_1=1}^{L} \sum_{i_2=2}^{L} h_{R,F}(i_1, i_2) \log_2 \frac{h_{R,F}(i_1,i_2)}{h_R(i_1) h_F(i_2)}$ ($h_{R,F}$ is the normalized joint gray level histogram of images R and F, h_R, h_F are the normalized marginal histograms of the two images, and L is the number of gray levels),

will be used. Notice that MI measures the reduction in uncertainty about the reference image due to the knowledge of the fused image, and so a larger MI is preferred.

A linear SVM with $C = 1000$ is used. Preliminary studies show that nonlinear SVMs do not perform better here. For training, two pairs of 30×30 regions are selected from the two out-of-focus images. In one pair of these regions, the first image is clearer than the second image, and the reverse is true for the second pair. Each pixel generates one training sample for the SVM, and so the training set has a total of 1800 training samples. For the DWFT, we use the biorthogonal B-spline wavelet basis with 3 decomposition levels.

3.1 Registered Images

The first experiment is performed on an 256-level, 480×640 image (Figure 2(a)), with good focus everywhere. We blur the left and right halves separately to obtain the images in Figures 2(b) and 2(c). For comparison purposes, we also perform fusion using DWT with the maximum selection rule (DWT+max_slct), and DWFT with the maximum selection rule (DWFT+max_slct).

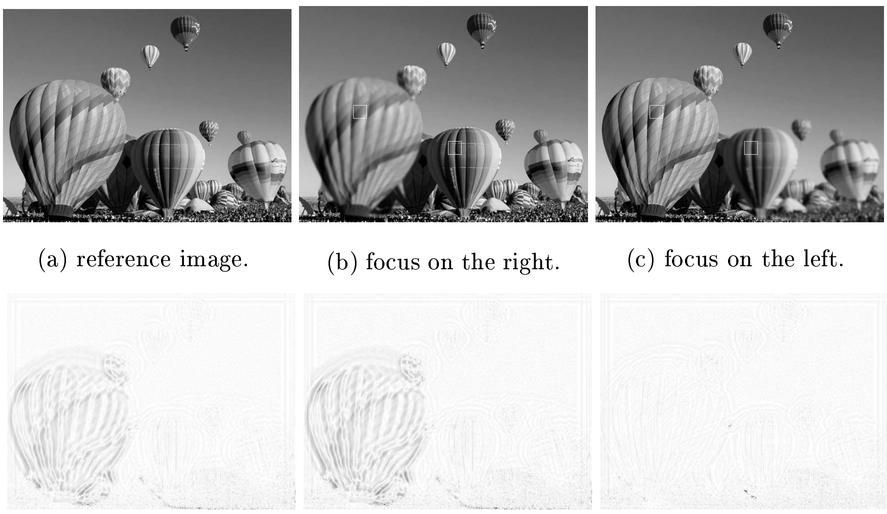
Table 1 compares the various fusion results quantitatively, and a visual comparison by examining the differences between the fused images and the reference image is shown in Figure 2. As can be seen, the fused image produced by DWFT+SVM is basically a combination of the focused parts in the source images. Results based on using maximum selection are much inferior.

Table 1. Performance on fusing the two images in Figures 2(b) and 2(c).

	DWT+max_slct	DWFT+max_slct	DWFT+SVM
RMSE	0.7634	0.7073	0.3433
MI	6.3607	6.4488	6.8980

3.2 Effect of Mis-registration

In this Section, we artificially shift the source image in Figure 2(b) by two pixels to the right, and the source images are consequently not perfectly registered. Figure 3 shows the differences between the fused images and the source images (Figure 2(c) and the shifted version of Figure 2(b)). Again, the fused image produced by DWFT+SVM is much superior. Also, notice that mis-registration has caused significant performance deterioration for the DWT-based method.



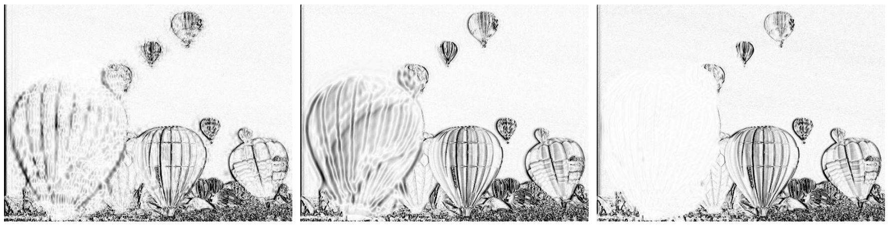
(a) reference image. (b) focus on the right. (c) focus on the left.

(d) DWT+max_slct. (e) DWFT+max_slct. (f) DWFT+SVM.

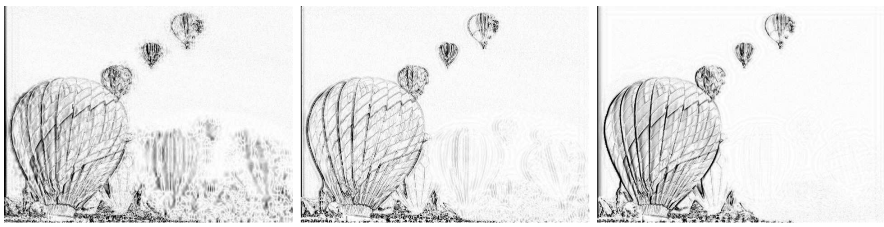
Fig. 2. Source images ((a)-(c)) and difference images ((d)-(f)) for the experiment in Section 3.1. The training set is selected from regions marked by the rectangles in (b) and (c).

4 Conclusion

In this paper, we use DWFT and SVM for multifocus image fusion. DWFT is advantageous over DWT in that the decomposed signal representation, and



(a) DWT+max_slct. (b) DWFT+max_slct. (c) DWFT+SVM.



(d) DWT+max_slct. (e) DWFT+max_slct. (f) DWFT+SVM.

Fig. 3. Difference images ((a)-(c) are matches with Figure 2(c), while (d)-(f) are matches with the shifted version of Figure 2(b)) for the experiment in Section 3.2.

consequently the fusion result, is shift-invariant. SVM, on the other hand, outperforms the simple maximum selection rule in determining which source image is in better focus at a particular pixel location. Experiments demonstrate that the proposed method outperforms, both visually and quantitatively, the conventional approach based on DWT and maximum selection rule. The improvements are particularly significant when there is slight camera/object movement or misregistration of the source images.

References

1. A. Laine and J. Fan. Frame representations for texture segmentation. *IEEE Transactions on Image Processing*, 5(5):771–780, 1996.
2. H. Li, B.S. Manjunath, and S.K. Mitra. Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3):235–245, May 1995.
3. G.P. Nason and B.W. Silverman. The stationary wavelet transform and some statistical applications. In A. Antoniadis and G. Oppenheim, editors, *Lecture Notes in Statistics*, volume 103, pages 281–300. Springer-Verlag, New York, 1995.
4. V. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.
5. Z. Zhang and R.S. Blum. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. *Proceedings of the IEEE*, 87(8):1315–1326, August 1999.