Multifocus Image Fusion Using Artificial Neural Networks

Shutao LI
College of Electrical and Information Engineering, Hunan University, Changsha, China 410082

Yaonan WANG
College of Electrical and Information Engineering, Hunan University, Changsha, China 410082

Boris Lohmann
Institute of Automation Technology, University of Bremen, NW1, 28359, Bremen, Germany

Abstract

Due to the limited depth-of-focus of optical lenses (especially such with long focal lengths) it is often not possible to get an image which contains all relevant objects in focus. One possibility to overcome this problem is to take several pictures with different focus points and combine them together into a single frame, which contains the focused regions of all input images. This paper describes the application of artificial neural networks to pixel-level fusion of multifocus images taken from the same scene. The proposed fusion method exploits the pattern recognition capabilities of artificial neural networks. Moreover, the learning capability of neural networks makes it feasible to customize the image fusion process. The experimental results show that the proposed method can perform better than the wavelet transform based method in some situations.

1. Introduction

Due to the limited depth-of-focus of optical lenses (especially such with long focal lengths) it is often not possible to get an image which contains all relevant objects in focus. One possibility to overcome this problem is to take several pictures with different focus points and combine them together into a single frame, which contains the focused regions of all input images[1-3].

The simplest image fusion method is to take average of two (or multiple) original images pixel by pixel. However when this direct method is applied, the contrast of features uniquely presented in either of the images is reduced. In order to solve this problem, several sophisticated approaches based on multiscale transform, such as Laplacian pyramid, gradient pyramid, ratio-of-low-pass pyramid, morphological pyramid, and multiresolution wavelet transform have been proposed in recent years[4-8]. The basic idea of most of the multiscale transform-based methods is to perform multiresolution decomposition on each source image, then integrate a composite multiresolution representation from these. Subsequently, the fused image is reconstructed by performing an inverse multiresolution transform.

In this paper, an efficient pixel level multifocus image fusion algorithm based on artificial neural networks is proposed. The fusion method, originated from human visual perception principle, is suitable to merge images with diverse focuses.
Two spatially registered images with different focuses are decomposed into several blocks. Then, three features reflecting the clear level of every block are calculated. Finally, artificial neural networks are used to recognize the clear level of the corresponding blocks to decide which blocks should be used to construct the fusion result. The proposed method is

computationally simple and can be applied in real time. Compared to the method based on discrete wavelet transform (DWT), the proposed method can achieve better results in some situations both in visual and quantitative measure.

2. Neural network based image fusion approach

2.1. General structure

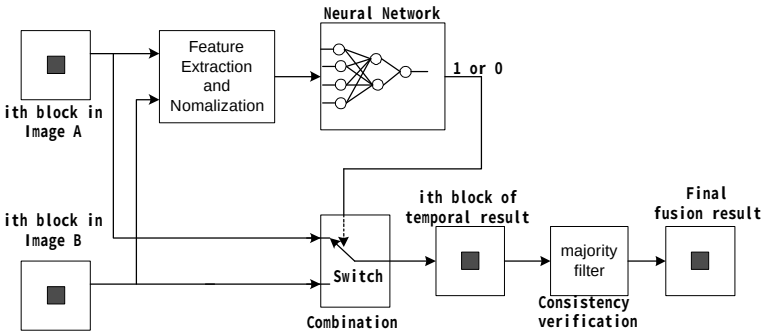


Figure 1. **Schematic diagram of the neural network multifocus image fusion method**

The schematic diagram of the proposed multifocus image fusion method is shown in Figure 1. The concrete fusion process is as follows.

Step 1. Decompose two registered images into several blocks with size of $M \times N$. Let A_i and B_i denote the *i*th blocks of image A and image B, respectively.

Step 2. Calculate features of the corresponding blocks in the two original images, and construct the normalized feature vector incident to neural networks. The features used to evaluate the clear level are spatial frequency, visibility, and edge. Let $\{SF_{A_i}, VI_{A_i}, EG_{A_i}\}$ and $\{SF_{B_i}, VI_{B_i}, EG_{B_i}\}$ are the three features value of A_i and B_i, respectively. Then the feature vector incident to neural networks is the normalization of $\{SF_{A_i}-SF_{B_i}, VI_{A_i}-VI_{B_i}, EG_{A_i}-EG_{B_i}\}$.

Step 3. Select some vector samples to train neural networks. The ideal classification results are

$$out_i = \begin{cases} 1 & \text{if } A_i \text{ is clearer than } B_i \\ 0 & \text{if } B_i \text{ is clearer than } A_i \end{cases} \quad (1)$$

Step 4. Use the trained neural networks to recognize the clear level of other image blocks.

Step 5. Construct the fused image according the classification results of neural networks. Let T_i is the *i*th block of the temporal fused image, then

$$T_i = \begin{cases} A_i & \text{if } out_i > 0.5 \\ B_i & \text{if } out_i \le 0.5 \end{cases} \quad (2)$$

Step 6. Verify the temporal fusion result using majority filter (which output 1 if the count of 1's outnumbers the count of 0's, and output 0 otherwise). Specially if the center block comes form image A while the majority of the surrounding blocks come from image B, the center block is switched to that of image B. In the implementation 3×3 window is applied.

2.2. Feature extraction

Three features are selected to reflect clear level of the decomposed blocks. Define block size is M×N.

2.2.1. Spatial frequency [9]

The row and column frequencies of an image block are given by

$$RF = \sqrt{\frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=1}^{N-1}[F(m,n)-F(m,n-1)]^2} \tag{3}$$

and

$$CF = \sqrt{\frac{1}{MN}\sum_{n=0}^{N-1}\sum_{m=1}^{M-1}[F(m,n)-F(m-1,n)]^2} \tag{4}$$

The total spatial frequency is then

$$SF = \sqrt{(RF)^2 + (CF)^2} \tag{5}$$

2.2.2. Visibility[10]

The visibility of an image block is defined as

$$VI = \sum_{m=1}^{M}\sum_{n=1}^{N}\omega(\mu)\bullet\frac{|f(m,n)-\mu|}{\mu} \tag{6}$$

where μ mean is the intensity mean value of the block, and

$$\omega(\mu) = (1/\mu)^{\alpha} \tag{7}$$

where α is a visual constant, which is ranging from 0.6 to 0.7.

2.2.3. Edge[11]

The edge detector proposed by Canny is used to every decomposed block, then a binary image is obtained. For example to block A_i,

$$BW_i = \text{canny}(A_i) \tag{8}$$

where the pixel on edge is set to one, or else to zero.

Let the number of pixel which equal to one to be the edge index of the decomposed block.

$$EG_i = \sum_{m=1}^{M}\sum_{n=1}^{N}BW_i \tag{9}$$

A image region with size of 64×64 selected from original 'Lena' image is shown in Figure 2(a). Figure 2(b) to Figure 2(e) show the Gaussian blurred versions of Figure 2(a). The blurring radiuses are 0.5, 0.8, 1.0 and 1.5, respectively. Three features of all these image regions are given in Table 1, from which it can be observed that the features values reduce along with the image blocks turn vague.

(a)Original image (b)Blurred with 0.5 (c) Blurred with 0.8 (d) Blurred with 1.0 (e) Blurred with 1.5

Figure 2. **Image region cut from 'Lena' and its blurred versions**

Similar experiment is implemented to an image region selected from original 'Peppers' image. Its blurred versions, shown in Figure 3(b) to Figure 3(e), are generated by Gaussian blurring with radiuses of 0.5, 0.8, 1.0, 1.5, respectively. Their feature vectors are shown in Table 2. It is clear the clearest image block corresponds with the highest feature values,

and the faintest image region has the lowest feature values.

Table 1. **Feature vectors of the image regions in Figure2**

	Figure2(a)	Figure2(b)	Figure2(c)	Figure2(d)	Figure2(e)
SF	16.10	12.09	9.67	8.04	6.49
VI	0.0069	0.0066	0.0062	0.0059	0.0055
EG	269	243	225	183	181

(a)Original image (b)Blurred with 0.5 (c) Blurred with 0.8 (d) Blurred with 1.0 (e) Blurred with 1.5

Figure 3. **Image region cut from 'peppers' and its blurred versions**

Table 2. **Feature vectors of the image regions in Figure3**

	Figure3(a)	Figure3(b)	Figure3(c)	Figure3(d)	Figure3(e)
SF	28.67	17.73	12.98	10.04	7.52
VI	0.0067	0.0063	0.0060	0.0057	0.0054
EG	329	310	274	260	216

From the experimental results, it can be seen that the three features can reflect the clear level of a still image region. In the situation of combination of images with diverse focuses, the objective is to obtain an image with focus everywhere from multiple images with respective focus. So, it is natural to use these features as clear level evaluation criteria to fuse the multifocus images.

2.3. Neural network

Artificial neural networks are simplified models originally designed to mimic the behavior of biological nervous systems. In this paper, two supervised neural networks, i.e. probabilistic neural network (PNN) and radial basis functions (RBF) are employed [12].

3. Experimental results

3.1. Experimental data set and performance measure

To quantitatively evaluate the performance of the proposed method, multiple images with different focuses are generated through blurring reference image with different focuses. The suitable candidate image should contain two objects with different distances to the camera. Firstly, one object is blurred to generate one image. Then, the other object is blurred to generate the other image. From the image shown in Figure 4(a), whose size is 480×640, the pair of distorted source images shown in Figure 4(b) and Figure 4(c) are generated by Gaussian blurring with radiuses of 2.

The multifocus images, shown in Figure 5-7, are used as benchmarks to subjective comparison between the proposed

method and the method based on DWT. Figure 5(a) focuses on the clock and Figure 5(b) on the student. In Figure 6(a) and Figure 6(b) the focuses are the Pepsi can and the testing card, respectively. Figure 7(a) and Figure 7(b) are two original images with different focuses.

Two evaluative measures, i.e. root mean square error (RMSE) and mutual information (MI) are used [3].

3.2. Objective evaluation of the proposed method

The training set is obtained manually by selecting some image blocks from the two original images. In order to apply the neural networks, the feature vectors must be converted into the range of the network. To make this operation, the feature vectors are scaling to the range [0,1]. After the neural networks are trained over, they are used to classify the entire image. According to the classification results, the fusion image is constructed by Equation (2).

The PNN consists of three neurons in the input layer, 60 neurons in hidden layer and one neuron in output layer. The best results were obtained with $\sigma = 0.09$.

The architecture of RBF neural network used for classification is 3-16-1. Several values for the width of units were examined. The best results were obtained with $\sigma = 2.5$.

The fusion results of Figure 4(b) and Figure 4(c) using the two different neural networks are shown in Figure 4(d) and Figure 4 (e), respectively. The decomposed block is with size of 32×32. The objective evaluations of the results are shown in Table 3. The objective evaluation of DWT-based method is also given. In the DWT-based method, wavelet filters are chosen as 'coif5', and the decomposition level is 5. Region-based activity measurement is employed to reflect the active level of decomposed coefficients. Coefficient combining method is the choosing maximum scheme. Window-based verification is used to consistency verification. These three selections are optimal according to the experimental results shown in [3].

3.3. Comparison to DWT-based method

The proposed method is visually compared with the DWT-based method. The fusion results obtained by the proposed algorithm and DWT are shown in Figure 5-7 (c) and (d), respectively.

Between Figure 5(a) and Figure 5(b) there is some motion of the students head. Because DWT yields a shift variant signal representation, i.e., a simple integer shift of the input signal will usually result in a nontrivial modification of the wavelet transform coefficients. Thus, the image fusion scheme based on DWT will also be shift dependent. From the fusion result of DWT based method, it can be seen that the areas around the student's head are worse than that of the original image shown in Figure 6(b). In the fusion result of our proposed method, that area is very clear. But, because the decomposed blocks are with rectangle, the fusion result of the proposed method on the cross over of the upright corner of the clock and the upleft corner of the monitor is worse than that of DWT based method.

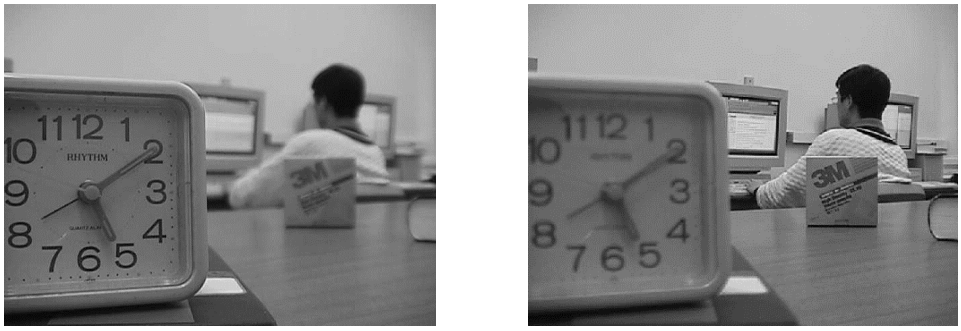
Because Figure 6(a) and Figure 6(b) are a little unregistered, the fused result based on DWT is visually bad. But the fused result obtained by the proposed method is very good overall. And it should be note that the margin of the clock is some uneven.

Comparing Figure 7(c) and Figure 7(d), we can found that the small area around the characters on the test card in Figure 7(d) is vague, whereas that in Figure 7(c) is clearer.

(a) Reference image (all focus); (b) Image focused on right; (c) Image focused on left;

(d) Fusion image using PNN; (e) Fusion image using RBF; (f) Fusion image using wavelet-based method

Figure 4. **Reference image, the blurred images and the fusion resul**t



(a) Image focused on the clock; (b) Image focused on the student;

(c) Fused image using PNN; (d) Fused image using DWT (db40 with level of 5).

Figure 5 **Source images (size = 480 × 640) and fusion results**

It can be seen from the experimental results that if the contained out-focus objects are not intersected, i.e. the objects

have certain distance each other, even though there are some motion between the corresponding objects or the original images are not stringently registered, the proposed method still can achieve good fusion performance.

Table 3. **Objective fusion performance of Fig4(a) and Figure4(b)**

	PNN	RBF	Wavelet-based
RMSE	0.2634	0.2580	1.5342
MI	7.3940	7.2289	6.2200

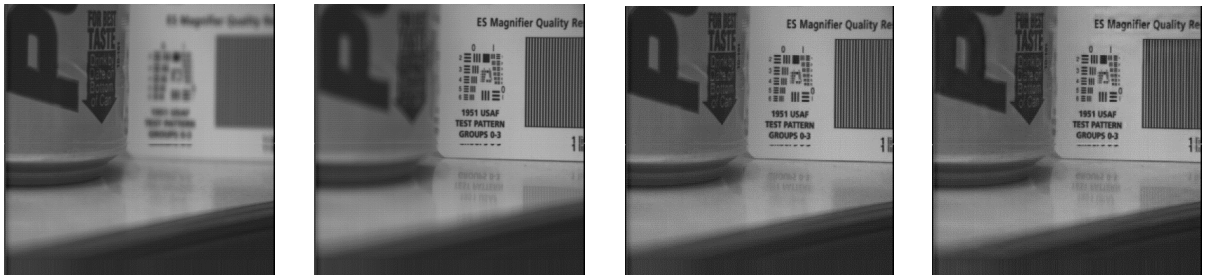
(a) Image focused on the left;

(b) Image focused on the right;

(c) Fusion result using PNN;

(d) Fusion result using DWT (db8 with level of 5).

Figure 6 **Source images (size = 480×640) and fusion results**



(a) (b) (c) (d)

Figure 7 **Source images (size = 512×512) and fusion results**

(a) Image focused on the Pepsi can; (b) Image focused on the testing card; (c) Fused result using PNN; (d) Fused result using DWT (coif5 with level of 5).

4. Conclusions

This paper describes the application of neural networks to pixel-level fusion of multiple out focus images taken from the same scene. The proposed fusion method exploits the pattern recognition capabilities of artificial neural networks. Three features, i.e. spatial frequency, visibility, and edge, reflecting the clear level of image blocks are extracted and fed into the trained neural networks. The fused results are constructed according to the classification results of neural networks. Two kinds of neural networks, i.e., PNN and RBF are used. Experimental results show that the PNN achieves better performance. Compared to the DWT-based method, it can be seen that if the two objects in the two out focus images are not intersected, the proposed method can perform better than the DWT-based method. The advantages of the proposed method is even the objects in the original images have some motions or the original images are not stringently registered, it still perform well. And the proposed method can perform real time.

In this paper, the decomposed image blocks are rectangle, so it can not correctly reflect the clear level of the intersection area between the out focus objects. To decompose the multifocus image into blocks with odd shapes instead of rectangle need further study. In that situation, because the intersection area can be segmented into different regions, the neural network based fusion method should always perform better than the DWT- based method.

References

[1] H. Li, B. S. Manjunath and S. K. Mitra, “Multisensor image fusion using the wavelet transform”, *Graphical Models and Image Processing*, 1995, vol.57, no.3, pp.235-245.

[2] W. B. Seales and S. Dutta, “Everywhere-in-focus image fusion using controllable cameras”, *Proceedings of SPIE*, 1996, vol.2905, pp.227-234.

[3] 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*, 1999, vol.87, no.8, pp.1315-1326.

[4] P. J. Burt and E. H. Andelson, “The Laplacian pyramid as a compact image code”, *IEEE Transactions on Communications*, 1983, vol.31, no.4, pp.532-540.

[5] P. J. Burt and R. J. Lolczynski, “Enhanced image capture through fusion”, in Proceedings of the 4th International Conference on Computer Vision, pp.173-182, May, 1993, Berlin, Germany.

[6] G. K. Matsopoulos, S. Marshall, and J.Brunt, “ Multiresolution morphological fusion of MR and CT images of the human brain”, *Proceedings of IEE, Vision, Image and Signal Processing*, 1994, vol.141, no.3, pp.137-142.

[7] A. Toet, L. J.Van Ruyven, and J. M.Valeton, “Merging thermal and visual images by a contrast pyramid”, *Optical Engineering*, 1989, vol.28, no.7, pp.789-792.

[8] A. Y. David, “Image merging and data fusion by means of the discrete two-dimensional wavelet transform”, *Journal of Optical Society of America*, 1995, vol.12, no.9, pp.1834-1841.

[9] A. M. Eskicioglu and P. S. Fisher, “Image quantity measures and their performance” *IEEE Transactions on Communications*, 1995,vol.43, no.12, pp.2959-2965.

[10] J. W. Huang, Y. Q. Shi, and X. H. Dai, “A segmentation-based image coding algorithm using the features of human vision system”, *Journal of Image and Graphics*, 1999, vol.4(A), no.5, pp.400-404.

[11] J. Canny, “A computational approach to edge detection”, *IEEE Transactions on Pattern Recognition and Machine Analysis*, 1986, vol.8, no.6, pp.679-698.

[12] S. Maleki, M. Amin Zia, A. R. Mirzai, et al, “Application of neural networks to the segmentation of MRI: comparison of different networks”, *Proceedings of SPIE*, 1997, vol.3164, pp.161-168.