

Research Article

Effective Multifocus Image Fusion Based on HVS and BP Neural Network

Yong Yang,^{1,2} Wenjuan Zheng,¹ and Shuying Huang³

¹ School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330013, China

² School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China

³ School of Software and Communication Engineering, Jiangxi University of Finance and Economics, Nanchang 330013, China

Correspondence should be addressed to Yong Yang; greatyangy@126.com

Received 17 July 2013; Accepted 19 December 2013; Published 6 February 2014

Academic Editors: P. Bifulco, C. Saravanan, K. Teh, and C. Zhang

Copyright © 2014 Yong Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The aim of multifocus image fusion is to fuse the images taken from the same scene with different focuses to obtain a resultant image with all objects in focus. In this paper, a novel multifocus image fusion method based on human visual system (HVS) and back propagation (BP) neural network is presented. Three features which reflect the clarity of a pixel are firstly extracted and used to train a BP neural network to determine which pixel is clearer. The clearer pixels are then used to construct the initial fused image. Thirdly, the focused regions are detected by measuring the similarity between the source images and the initial fused image followed by morphological opening and closing operations. Finally, the final fused image is obtained by a fusion rule for those focused regions. Experimental results show that the proposed method can provide better performance and outperform several existing popular fusion methods in terms of both objective and subjective evaluations.

1. Introduction

Due to a finite depth of field of optical lenses, it is usually impossible to get an image in which all relevant objects are in focus; that is, only those objects within the depth of field of the camera will be in focus, while other objects will be out of focus [1]. Consequently, in order to obtain an image with every object in focus, images taken from the same scene focusing on different objects need to be fused, that is, multifocus image fusion [2]. Image fusion refers to an image preprocessing technique that combines two or more source images that have been registered into a single image according to some fusion rules. Its aim is to integrate complementary and redundant information of multiple images coming from the same scene to form a single image that contains more information of the scene than any of the individual source images [3]. Multifocus image fusion is an important branch of this field. The fused image obtained then turns out to be more suitable for human/machine perception, segmentation, feature extraction, detection, or target recognition tasks [4].

Image fusion is generally performed at different levels of information representation, namely, pixel level, feature

level, and decision level [5]. Up to now, many multifocus image fusion techniques have been developed. Basically, the fusion technique can be categorized into spatial domain fusion and transform domain fusion [6]. The spatial domain-based methods directly select the clearer pixels or regions from source images in the spatial domain to construct the fused image [7, 8]. The basic idea of the transformed domain-based methods is to perform certain multiresolution decomposition on each source image, then integrate all these decompositions to obtain one combined representation according to some fusion rules, and finally reconstruct the fused image by performing the inverse transformation to the combined representation [9].

The simplest fusion method is to take the average of the source images pixel by pixel. The method is simple and suitable for real-time processing. However, it does not consider the correlation between the surrounding pixels and often leads to several undesired side effects such as reduced contrast [3]. In order to improve the quality of the fused image, the block-based multifocus image fusion methods have been proposed [7, 8]. These methods are shift-invariant, and all of the operations are performed in

the spatial domain, so they have high computational efficiency. However, they are also faced with some problems. The first problem is how to determine the suitable size of the subblock. These methods usually suffer from block effects which severely reduce the quality of the fused image if the size of the subblock is selected unreasonably. Another problem is that which evaluation criteria would be more suitable to measure the clarity of the subblocks. In recent years, various approaches based on multiscale transforms have been proposed, including pyramid transform and wavelet transform, such as the Laplacian pyramid [10], gradient pyramid [11], the ratio of low pass pyramid [12], discrete wavelet transform (DWT) [13–15], shift-invariant discrete wavelet transform (SIDWT) [16], curvelet transform [17], contourlet transform [18], and nonsubsampling contourlet transform (NSCT) [19]. Pyramid decomposition-based image fusion can achieve a good effect. However, the pyramid decomposition of the image is redundant decomposition. The information of the different decomposition layers is correlative, which makes it easy to reduce the stability of the algorithm. Generally, DWT is superior to the previous pyramid-based methods because of providing directional information and without carrying redundant information across different resolutions. Moreover, DWT has good locality of time frequency. However, these methods based on multiscale transforms are shift-variant; namely, their performance will quickly deteriorate when there is a slight camera/object movement or there is misregistration of the source images [7, 20]. Although the SIDWT [16] and NSCT [19] algorithms both can overcome the shortcoming mentioned above, the implementation of the algorithm is more complicated and more time-consuming. Besides, some information of the source images may be lost during the inverse multiresolution transform implementation [21]. Recently, pulse coupled neural network (PCNN) has also been introduced to the multifocus image fusion, as seen in literature [22, 23]. However, the PCNN technique is very complex and has too many parameters. In addition, it is long and time-consuming.

In order to overcome the shortcoming of the methods mentioned above, in this paper, we propose a pixel level multifocus image fusion method based on HVS and BP neural network. Firstly, three features including texture feature, local visibility, and local visual feature contrast are extracted based on HVS and are used to train the BP neural network. Secondly, the initial fused image is acquired using BP neural network followed by a consistency verification process. Then, in order to avoid yielding any artificial or erroneous information that may be introduced during the process of preliminary fusion, the focused regions in each source image are determined by a hybrid procedure. Finally, the fused image is obtained based on the focused regions and initial fused image. The experiments show that the performance of the proposed method is superior to several existing fusion methods.

The rest of the paper is organized as follows. The related theory of the proposed method is described in Section 2. The fusion method that is based on HVS and BP neural network is introduced in Section 3. Experimental results

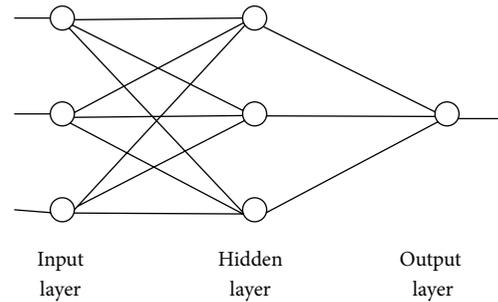


FIGURE 1: Architecture of BP neural network.

and performance analysis are presented and discussed in Section 4, and the last section gives some concluding remarks.

2. Related Theoretical Knowledge

2.1. BP Neural Network. BP neural network is a multilayer feed-forward neural network, which is one of the most widely used neural networks. The problem of multifocus fusion based on BP neural work can be considered as a classification problem, focused or blurred.

The basic BP neural network is a three-layer network, including input layer, hidden layer, and output layer. The architecture of BP neural network in the paper is shown in Figure 1. According to [24], we also adopt empirical formula to determine the number of nodes of the hidden layer, and the formula is defined as follows:

$$n_h = \text{sqrt}(n_i + n_o) + 1, \quad (1)$$

where n_h , n_i , and n_o are the number of nodes of the hidden layer, the number of nodes of the input layer, and the number of nodes of the output layer, respectively.

2.2. Features Extraction. In this paper, for each pixel, we extract three features based on the pixel centered of the 3×3 window to reflect its clarity. These are the texture features, local visibility, and local visual feature contrast.

2.2.1. Texture Features. Log-Gabor filter was designed in the log coordinate system which is more conducive to the texture feature extraction [25]. The main advantage of the log-Gabor functions is that it can construct filters with arbitrary bandwidth under the condition of maintaining the DC component 0, which reduces filters redundancy. Furthermore, log-Gabor filters are more in line with the HVS. Texture features (TF) based on amplitude information reflect the high and low frequency energy distribution of the images. Therefore, taking the advantages of the log-Gabor filters into account, texture features of the multifocus image based on amplitude information will be extracted using log-Gabor filters. 2D Log-Gabor filter is defined in the frequency domain as follows [26]:

$$H(f, \theta) = H_f \times H_\theta, \quad (2)$$

where H_f is radial component and H_θ is direction components. Specifically, the expressions are as follows:

$$H_f = \exp \left\{ \frac{-[\log(f/f_0)]^2}{2[\log(\sigma_f/f_0)]^2} \right\}, \quad (3)$$

$$H_\theta = \exp \left\{ -\frac{(\theta - \theta_0)^2}{2\sigma_\theta^2} \right\},$$

in which f_0 is the center frequency of filters, θ_0 is the direction of filters, and σ_f is a constant that controls radial filters bandwidth B_f . Consider

$$B_f = 2\sqrt{\frac{2}{\log 2}} \times \left| \log \left(\frac{\sigma_f}{f_0} \right) \right|. \quad (4)$$

In order to obtain log-Gabor filters with the same bandwidth, σ_f must be changed along with f_0 so that the value of

σ_f/f_0 is constant. σ_θ determines direction bandwidth B_θ . Consider

$$B_\theta = 2\sigma_\theta \sqrt{2 \log 2}. \quad (5)$$

2.2.2. Local Visibility. In the paper, we introduce the concept of the image visibility (VI), which is inspired from the HVS and defined as follows [27]:

$$VI = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N \left(\frac{1}{m_k} \right)^\alpha \times \frac{|I(i, j) - m_k|}{m_k}, \quad (6)$$

where m_k is the mean intensity value of the image, α is a visual constant ranging from 0.6 to 0.7, and $I(i, j)$ denotes the gray value of pixel at position (i, j) .

VI is more significant in multifocus image fusion than different sensor image fusion and the measurement has been successfully used in multifocus image fusion [27]. In the paper, in order to represent the clarity of a pixel, the local visibility (LVI) in spatial domain is proposed. The LVI is defined as

$$LVI(x, y) = \begin{cases} \frac{1}{(2m+1) \times (2n+1)} \sum_{i=-m}^m \sum_{j=-n}^n \left(\frac{1}{\bar{I}(x, y)} \right)^\alpha \times \frac{|I(x+i, y+j) - \bar{I}(x, y)|}{\bar{I}(x, y)} & \text{if } \bar{I}(x, y) \neq 0 \\ I(x, y) & \text{otherwise,} \end{cases} \quad (7)$$

where $(2m+1) \times (2n+1)$ is the size of neighborhood window and $\bar{I}(x, y)$ is the mean intensity value of the pixel (x, y) centered of the $(2m+1) \times (2n+1)$ window.

2.2.3. Local Visual Feature Contrast. The findings of psychology and physiology have shown that HVS is highly sensitive to changes in the local contrast of the image, but insensitive to real luminance at each pixel [28]. The local luminance contrast formula is defined as follows:

$$C = \frac{L - L_B}{L_B} = \frac{\Delta L}{L_B}, \quad (8)$$

where L is the local luminance and L_B is the local luminance of the background, namely, the low frequency component. Therefore, ΔL can be taken as the high frequency component. However, the value of single pixel is not enough to determine which pixel is focused without considering the correlation between the surrounding pixels. Therefore, to represent the salient features of the image more accurately the local visual feature (LVC) contrast in spatial domain is introduced, and is defined as

$$LVC(x, y) = \begin{cases} \left(\frac{1}{\bar{I}(x, y)} \right)^\alpha \times \frac{SML(x, y)}{\bar{I}(x, y)} & \text{if } \bar{I}(x, y) \neq 0 \\ SML(x, y) & \text{otherwise,} \end{cases} \quad (9)$$

where $\bar{I}(x, y)$ is the mean intensity value of the pixel (x, y) centered of the neighborhood window, α is a visual constant

ranging from 0.6 to 0.7, and the $SML(x, y)$ denotes the sum-modified-Laplacian (SML) located at (x, y) , and more details about SML can be found in [7].

3. The Proposed Multifocus Image Fusion Method

3.1. Initial Fused Image Obtained by BP Neural Network. Figure 2 shows the schematic diagram of the proposed method for obtaining the initial fused image based on BP neural network. Here, we only consider the case of two-source-image fusion, though the method can be extended straightforwardly to handle more than two, with the assumption that the source images have always been registered.

The algorithm first calculates salient features of each pixel from each source image by averaging over a small window. Assume that there are two pixels (one from each source image) and BP neural network is trained to determine which one is in focus. Then the initial fused image is constructed by selecting the clearer pixel followed by a consistency verification process. Specifically, the algorithm consists of the following steps.

Step 1. Assume that there are two source images A and B . Denote the i th pixel pair by A_i and B_i , respectively.

Step 2. For each pixel, extract three features based on the pixel centered of the 3×3 window, which reflect its clarity (details

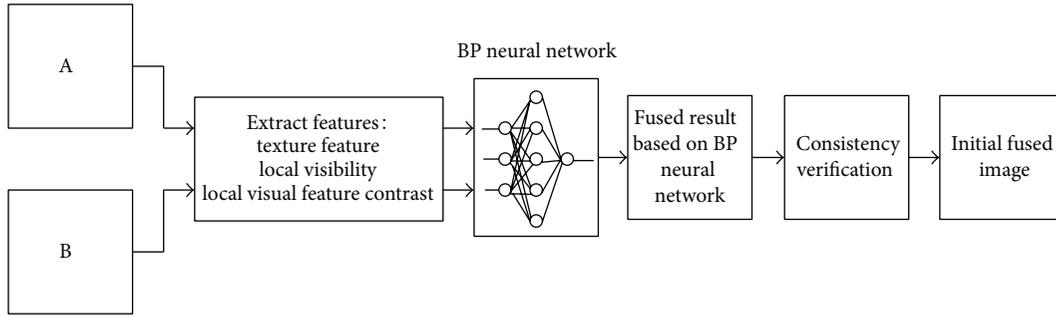


FIGURE 2: Schematic diagram of the BP neural network based fusion method.

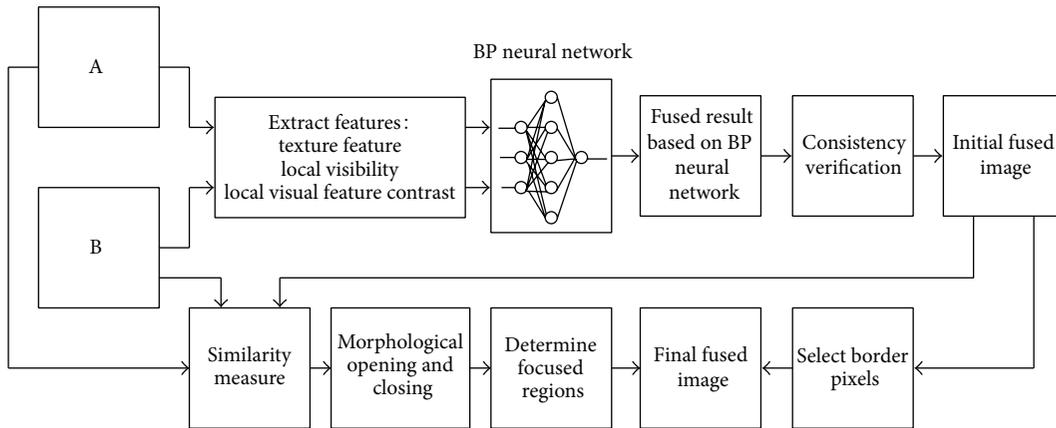


FIGURE 3: Schematic diagram of the proposed image fusion method.

in Section 2.2). Denote the feature vectors for A_i and B_i by $(TF_{A_i}, LVI_{A_i}, LVC_{A_i})$ and $(TF_{B_i}, LVI_{B_i}, LVC_{B_i})$, respectively.

Step 3. Train a BP neural network to determine which pixel is clearer. The difference vector $(TF_{A_i} - TF_{B_i}, LVI_{A_i} - LVI_{B_i}, LVC_{A_i} - LVC_{B_i})$ is used as input, and the output is labeled according to

$$\text{target}_i = \begin{cases} 1 & \text{if } A_i \text{ is clearer than } B_i, \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

Step 4. Perform simulation of the trained BP neural network on all pixel pairs. The i th pixel, F_i , of the fused image is then constructed as

$$F_i = \begin{cases} A_i & \text{if } \text{out}_i \geq 0.5 \\ B_i & \text{otherwise,} \end{cases} \quad (11)$$

where out_i is the BP neural network output using the i th pixel pair as corresponding input.

Step 5. Verify consistency of the result of the fusion obtained in Step 4. Especially, when the BP neural network decides that a particular pixel is to come from A but with the majority of its surrounding pixel from B , this pixel will be changed to come from B .

3.2. The Method for Obtaining Final Fused Image. In order to ensure that the pixels of the fused image come from the

focused regions of each source image, we need to identify the focused regions in each source image firstly. Then the fused image can be constructed by simply selecting pixels in those regions. And as for the boundary of focused regions, the corresponding pixel of the initial fused image is selected as the pixel of the final fused image. Therefore, we proposed the following flow chart for obtaining the final fused image as illustrated in Figure 3.

3.2.1. Detection of the Focused Regions. The pixels of the source images with higher similarity to the corresponding initial fused image pixels can be considered to be located in the focused regions. Thus, the focused regions in each source image can be determined by this method. In the paper, we adopt root mean square error (RMSE) [14] to measure the similarity between the source images and the initial fused image. Specifically, the algorithm of the detection of focused regions consists of the following steps.

Step 1. Calculate the RMSE of each pixel within $(2m + 1) \times (2n + 1)$ window between the source images and the initial fused image. Assume that A and B are two source images and F is the initial fused image. The formulas are defined as follows, respectively. In order to acquire the best fusion effect, we have tried different window sizes and found that the fusion effect is best when the size of the window is 5×5 or 7×7 .

Step 2. Compare the values $RMSE_{AF}(x, y)$ and $RMSE_{BF}(x, y)$ to determine which pixel is in focus. The decision diagram, which is a binary image, will be constructed as follows:

$$Z(x, y) = \begin{cases} 1 & \text{if } RMSE_{AF}(x, y) < RMSE_{BF}(x, y) \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

where “1” in Z indicates that the pixel at position (x, y) in source image A is in focus; conversely, the pixel in source image B is in focus, which indicates that the pixel with smaller $RMSE(x, y)$ value is more possible in focus.

Step 3. In order to determine all the focused pixels and avoid the misjudgement of pixels, morphological opening and closing with small square structuring element and connected domain are employed. Opening, denoted as $Z \circ b$, is that Z is eroded firstly by the structure element b followed by dilation of the result by b . It can smooth the contours of the object and remove narrow connections and small protrusions. Like the opening, closing can also smooth the contours of the object. However, the difference is that closing can join narrow gaps and fill the hole which is smaller than the structure element b . Closing is dilation by b followed by erosion by b and is denoted as $Z \bullet b$. In fact, those small holes are usually generated by the misjudgement of pixels. What was worse, the holes larger than b are hard to remove simply using opening and closing operators. Therefore, a threshold TH should be set to remove the holes smaller than the threshold but larger than b . Then opening and closing are again used to smooth the contours of the object. Finally, the focused regions of each source image can be acquired, which can be more uniform and have well connected regions.

As for the structure element b and the TH , they can be determined according to the experimental results. In the paper, the structure element b is a 7×7 matrix with logical 1. In order to remove small and isolated areas which are misjudged, two different thresholds are set. The first threshold is set to be 20000 to remove areas which are focused in image B but misjudged as blurred. The second threshold is set to be 3000 to remove those areas which are focused in image A but misjudged as blurred.

3.2.2. *Fusion of the Focused Regions.* The final fused image FF can be acquired according to the fusion rules that are as follows:

$$FF(x, y) = \begin{cases} A(x, y) & \text{if } ZZ(x, y) == 1, \text{ count}(x, y) \\ & = (2m + 1) \times (2n + 1) \\ B(x, y) & \text{if } ZZ(x, y) == 0, \text{ count}(x, y) = 0 \\ F(x, y) & \text{otherwise,} \end{cases} \quad (13)$$

where

$$\text{count}(x, y) = \sum_{i=-m}^{i=m} \sum_{j=-n}^{j=n} ZZ(x + i, y + j), \quad (14)$$

ZZ is the modified Z matrix of Step 3 in Section 3.2.1, $A(x, y)$, $B(x, y)$, $F(x, y)$, and $FF(x, y)$ denote the gray value

of pixel at position (x, y) of the source images (A and B), the initial fused image F , and the final fused image FF , respectively, and $(2m + 1) \times (2n + 1)$ is the size of slipping window; $\text{count}(x, y) = (2m + 1) \times (2n + 1)$ suggests that the pixel at position (x, y) in image A is in focus and will be selected as the pixel of the final fused image FF directly. On the contrary, $\text{count}(x, y) = 0$ indicates that the pixel at the position coming from image B is focused and can be chosen as the pixel of the final fused image FF . Other cases, namely, $0 < \text{count}(x, y) < (2m + 1) \times (2n + 1)$, imply that the pixel at position (x, y) is located in the boundary of focused regions, and the corresponding pixel of the initial fused image F is selected as the pixel of the final fused image FF .

4. Experimental Results and Performance Analysis

4.1. *Experimental Setup.* In this section, the first step we should do is to train the BP neural network. The training experiment is performed on the standard popular widely used “lena” image, which is a 256-level image with all in focus. We then artificially produce three out-of-focus images blurred with Gaussian radius of 0.5, 1.0, and 1.5, respectively. A training set with a total of $4 \times 256 \times 256$ pixel pairs is formed. The three features of each pixel, TF, LVI, and LVC, are extracted with $\alpha = 0.65$. In addition, we artificially produce a pair of out-of-focus images shown in Figures 4(a) and 4(b), which are acquired by blurring the left part and the middle part of the original image using the Gaussian function, respectively. To evaluate the advantage of the proposed fusion method, experiments are performed on three sets of source images as shown in Figures 4, 5, and 6, respectively, including one set of source images produced artificially and two sets of source images acquired naturally. Their sizes are 256×256 , 256×256 , and 640×480 , respectively. These images all contain multiple objects at different distances from the camera and only those objects within the depth of field of the camera will be focused, while other objects naturally will be out of focus when taken. For example, Figure 5(a) is focused on testing card, while Figure 5(b) is focused on the pepsi can.

In order to compare the performance of the proposed fusion method, these multifocus images are also performed using the conventional and classical methods, such as taking the average of the source images pixel by pixel, the gradient pyramid method [11], the DWT-based method, and the SIDWT-based method [16]. The decomposition level of the multiscale transform is 4 layers. The wavelet basis of the DWT and SIDWT is DBSS (2, 2) and Haar, respectively. The fusion rules of lowpass subband coefficients and the highpass subband coefficients are the “averaging” scheme and the “absolute maximum choosing” scheme, respectively.

4.2. *Evaluation Criteria.* In general, the evaluation methods of image fusion can be categorized into subjective methods and objective methods. However, observer personal visual differences and psychological factors will affect the results of image evaluation. Furthermore, in most cases, it is difficult for us to perceive the difference among fusion results. Therefore,



FIGURE 4: Original images and fused images of "lena": (a) focus on the right; (b) focus on left and right sides; (c) fused image using average; (d) fused image using gradient pyramid; (e) fused image using DWT; (f) fused image using SIDWT; (g) fused image using the proposed method; (h) the initial focused region; (i) the modified focused region.

the subjective evaluation of the fused results is always incomprehensive. Hence, in addition to the subjective evaluation, we also adopt several metrics to objectively evaluate the image fusion results and quantitatively compare the different fusion methods in the paper.

4.2.1. Mutual Information (MI). The mutual information MI_{AF} between the source image A and the fused image F is defined as follows:

$$MI_{AF} = \sum_{k=0}^{L-1} \sum_{i=0}^{L-1} p_{AF}(k, i) \log_2 \frac{p_{AF}(k, i)}{p_A(k) \times p_F(i)}, \quad (15)$$

where p_{AF} is the jointly normalized histogram of A and F , p_A and p_F are the normalized histograms of A and F , L is the gray level of the image, and k and i represent the pixel value of the images A and F , respectively. The mutual information MI_{BF} between the source image B and the fused image F is similar to MI_{AF} . The mutual information between the source images A , B and the fused image F is defined as follows:

$$MI_F^{AB} = MI_{AF} + MI_{BF}. \quad (16)$$

The metric reflects the total amount of information that the fused image F contains about source images A and B . The larger the value is, the more the information is obtained from the original image and the better the fusion effect is.

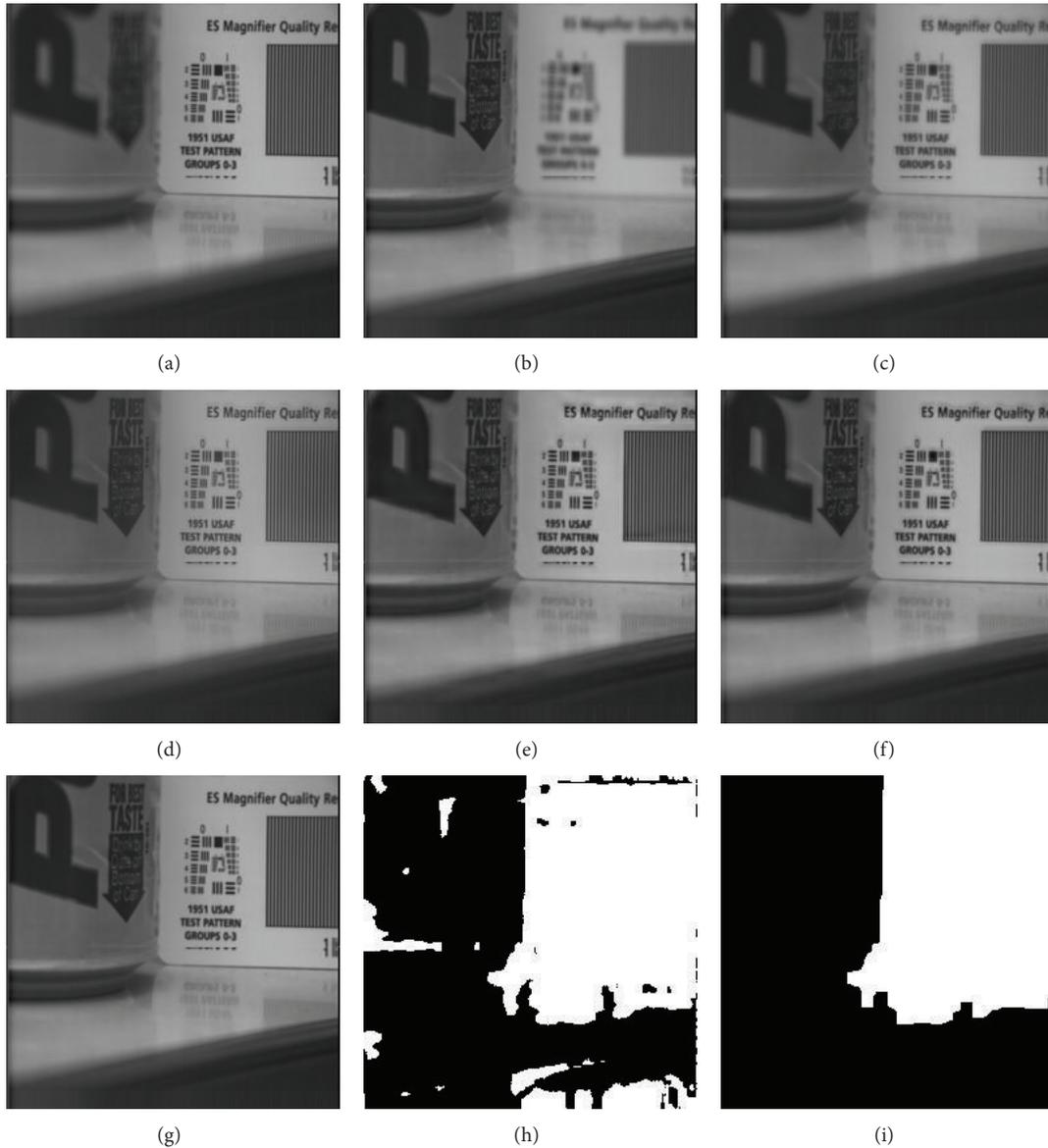


FIGURE 5: Original and fused image of “pepsi”: (a) focus on the right; (b) focus on the left; (c) fused image using average; (d) fused image using gradient pyramid; (e) fused image using DWT; (f) fused image using SIDWT; (g) fused image using the proposed method; (h) the initial focused region; (i) the modified focused region.

4.2.2. $Q^{AB/F}$. The metric $Q^{AB/F}$ evaluates the sum of edge information preservation values and is defined as follows:

$$Q^{AB/F} = \left(\sum_{m=1}^M \sum_{n=1}^N (Q^{AF}(m, n) \times \omega_A(m, n) + Q^{BF}(m, n) \times \omega_B(m, n)) \right) \times \left(\sum_{m=1}^M \sum_{n=1}^N (\omega_A(m, n) + \omega_B(m, n)) \right)^{-1} \quad (17)$$

where $Q^{AF}(m, n) = Q_g^{AF}(m, n)Q_\alpha^{AF}(m, n)$, $Q_g^{AF}(m, n)$ and $Q_\alpha^{AF}(m, n)$ are the edge strength and orientation preservation values, respectively, $Q^{BF}(m, n)$ is similar to $Q^{AF}(m, n)$, and $\omega_A(m, n)$ and $\omega_B(m, n)$ are weights to measure the importance of $Q^{AF}(m, n)$ and $Q^{BF}(m, n)$, respectively, The dynamic range of $Q^{AB/F}$ is $[0, 1]$ and it should be as close to 1 as possible, and for the “ideal fusion”, $Q^{AB/F} = 1$. In addition, (m, n) represents the pixel location, and M and N are the size of images, respectively.

The $Q^{AB/F}$ metric reflects the quality of visual information obtained from the fusion of input images. Therefore, the larger the value, the better the performance.

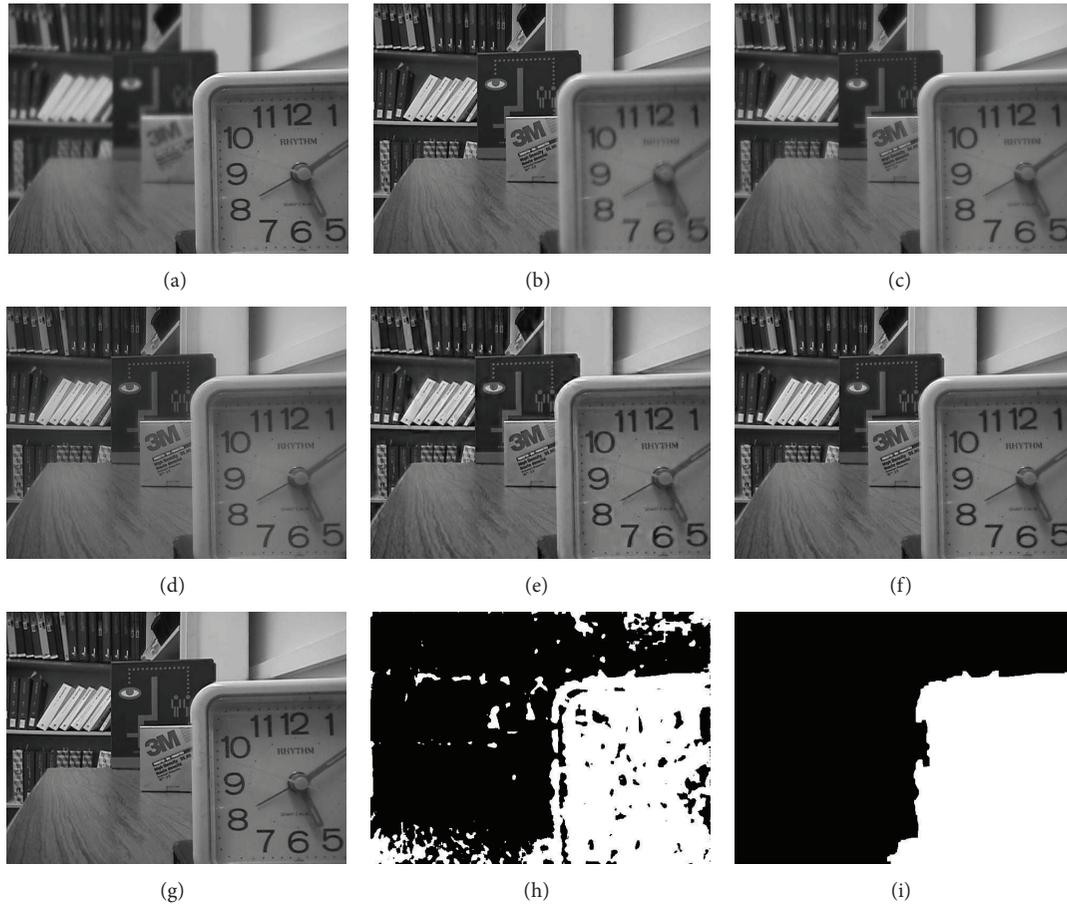


FIGURE 6: Original and fused image of “disk”: (a) focus on the right; (b) focus on the left; (c) fused image using average; (d) fused image using gradient pyramid; (e) fused image using DWT; (f) fused image using SIDWT; (g) fused image using the proposed method; (h) the initial focused region; (i) the modified focused region.

4.2.3. Correlation Coefficient (CORR). Correlation coefficient between the fused image F and the standard reference image R is defined as follows:

$$\begin{aligned} \text{CORR} &= \left(\sum_{m=1}^M \sum_{n=1}^N (R(m, n) - \bar{R}(m, n))(F(m, n) - \bar{F}(m, n)) \right) \\ &\times \left(\sum_{m=1}^M \sum_{n=1}^N (R(m, n) - \bar{R}(m, n))^2 \right. \\ &\quad \left. \times \sum_{m=1}^M \sum_{n=1}^N (F(m, n) - \bar{F}(m, n))^2 \right)^{-1/2}, \end{aligned} \quad (18)$$

where $\bar{R}(m, n)$ and $\bar{F}(m, n)$ represent the pixel gray average value of the standard reference image R and fused image F , respectively.

The metric reflects the degree of correlation between the fused image and the standard reference image. The larger the value is, the better the fusion effect is.

4.2.4. Root Mean Squared Error (RMSE). Root mean square error (RMSE) between the fusion image F and the standard reference image R is defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{m=1}^M \sum_{n=1}^N (R(m, n) - F(m, n))^2}{M \times N}}. \quad (19)$$

The metric is used to measure the difference between the fused image and the standard reference image. The smaller the value is, the better the fusion effect is.

4.3. Fusion of Artificial Test Images. The experiment is performed on a pair of “lena” multifocus images as shown in Figures 4(a) and 4(b). The initial and modified detected focused regions are shown in Figures 4(h) and 4(i), respectively. The white pixels in Figure 4(i) indicate that corresponding pixels from Figure 4(a) are in focused regions, while the black pixels suggest that corresponding pixels from Figure 4(b) are in focused regions. By comparison, we can observe that the detected focused regions of Figure 4(i) are better than those of Figure 4(h); for example, there are some misdetrcted focused regions in the right side of Figure 4(h), whereas they are correctly detected in Figure 4(i) because the right side

TABLE 1: Performance comparison of different fusion algorithms in Figure 4.

Fusion algorithms	MI	$Q^{AB/F}$	CORR	RMSE
Average method	7.4946	0.72072	0.98987	8.8023
Gradient pyramid	5.102	0.72693	0.98876	13.791
DWT	7.1207	0.76948	0.99784	4.0532
SIDWT	7.3712	0.77171	0.99553	5.7595
Proposed method	9.8067	0.81128	0.99962	2.8699

of it is almost totally white. The fusion results obtained by the previous five different methods are shown in Figures 4(c)–4(g), respectively. It can be found that the results of the pixel averaging and gradient pyramid method have a poor contrast compared to those of the DWT-based method, the SIDWT-based method, and the proposed method. However, it is difficult for us to perceive the difference among the results of the DWT-based method, the SIDWT-based method, and the proposed method according to the subjective evaluation. Therefore, to objectively evaluate these five fusion methods, quantitative assessments of the five fusion results are needed. The results of the quantitative assessments are shown in Table 1. As can be seen from Table 1, MI, $Q^{AB/F}$, and CORR values of the proposed method are higher and RMSE value is less than those of the other methods, which means that by using our proposed method, the best quantitative evaluation results have been achieved.

4.4. Fusion of Real Digital Camera Images. The experiments carried out in this section are performed on two sets of source images acquired naturally as shown in Figures 5(a)–5(b) and Figures 6(a)–6(b), respectively. The initial and modified detected focused regions of those two sets of source images are shown in Figures 5(h)–5(i) and Figures 6(h)–6(i), respectively. The fused images obtained by using pixel averaging method, gradient pyramid method, DWT-based method, the SIDWT-based method, and the proposed method on these two sets of source images are shown in Figures 5(c)–5(g) and Figures 6(c)–6(g), respectively. From the fusion results, we can easily observe that fusion effects acquired based on the pixel averaging and gradient pyramid are not satisfactory and with poor contrast. For example, the regions of the testing card in Figures 5(c)–5(d) are not clear, but they are clear in Figures 5(e)–5(g). But it is difficult to discriminate the difference among the results of the DWT-based method, the SIDWT-based method, and the proposed method by subjective evaluation, so we need to do objective evaluation. However, it should be noted that the reference image is usually not available for real multifocus images, so only the two evaluation criteria including the MI and $Q^{AB/F}$ are used to objectively compare the fusion results. The quantitative comparison of the five methods for fusion of these two sets of source images is shown in Tables 2 and 3, respectively. As can be seen from the two tables, we can find that the MI and $Q^{AB/F}$ values of the proposed method are significantly higher than those of the other methods. It should be noted that we have carried out experiments on other multifocus images, and their results are identical to these two

TABLE 2: Performance comparison of different fusion algorithms in Figure 5.

Fusion algorithms	MI	$Q^{AB/F}$
Average method	7.2941	0.64922
Gradient pyramid	5.9768	0.67975
DWT	6.4442	0.68264
SIDWT	6.82160	0.70890
Proposed method	9.1957	0.75904

TABLE 3: Performance comparison of different fusion algorithms in Figure 6.

Fusion algorithms	MI	$Q^{AB/F}$
Average method	5.9845	0.52143
Gradient pyramid	5.3657	0.63792
DWT	5.3949	0.64323
SIDWT	5.8380	0.67620
Proposed method	8.3105	0.73806

examples, so we did not mention all of them here. Therefore, the results of subjective and objective evaluation presented here can verify that the performance of the proposed method is superior to those of the other methods.

5. Conclusions

By combining the idea of the correlation between the neighboring pixels and BP neural networks, a novel multifocus image fusion method based on HVS and BP neural network is proposed in the paper. Three features which are based on HVS and can reflect the clarity of a pixel are extracted and used to train a BP neural network to determine which pixel is clearer. The clearer pixels are combined to form the initial fused image. Then the focused regions are detected by judging whether pixels from the initial fused image are in the focused regions or not. Finally the final fused image is obtained with the help of the technique of focused region detection by a certain fusion rule. The results of subjective and objective evaluation of several experiments show that the proposed method outperforms several popular widely used fusion methods. In the future, we will focus on improving the robustness of the method for noise.

Conflict of Interests

The authors declare no conflict of interests.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and constructive suggestions. This work was supported by the National Natural Science Foundation of China (no. 60963012 and no. 61262034), by the Key Project of Chinese Ministry of Education (no. 211087), by the Natural Science Foundation of Jiangxi Province (no. 20114BAB211020 and no. 20132BAB201025), by the Young Scientist Foundation of Jiangxi Province (no. 20122BCB23017), and by the Science and Technology Research Project of the Education Department of Jiangxi Province (no. GJJ13302).

References

- [1] P. Shah, S. N. Merchant, and U. B. Desai, "Multifocus and multispectral image fusion based on pixel significance using multiresolution decomposition," *Signal, Image and Video Processing*, vol. 7, pp. 95–109, 2013.
- [2] V. Aslantas and R. Kurban, "Fusion of multi-focus images using differential evolution algorithm," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8861–8870, 2010.
- [3] R. Benes, P. Dvorak, M. Faundez-Zanuy, V. Espinosa-Duró, and J. Mekysk, "Multi-focus thermal image fusion," *Pattern Recognition Letters*, vol. 34, pp. 536–544, 2013.
- [4] J. Dong, D. Zhuang, Y. Huang, and J. Fu, "Advances in multi-sensor data fusion: algorithms and applications," *Sensors*, vol. 9, no. 10, pp. 7771–7784, 2009.
- [5] S. Li and B. Yang, "Hybrid multiresolution method for multi-sensor multimodal image fusion," *IEEE Sensors Journal*, vol. 10, no. 9, pp. 1519–1526, 2010.
- [6] S. Li, B. Yang, and J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," *Information Fusion*, vol. 12, no. 2, pp. 74–84, 2011.
- [7] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 493–500, 2007.
- [8] S. Li, J. T. Kwok, and Y. Wang, "Multifocus image fusion using artificial neural networks," *Pattern Recognition Letters*, vol. 23, no. 8, pp. 985–997, 2002.
- [9] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "Multi-focus image fusion for visual sensor networks in DCT domain," *Computers and Electrical Engineering*, vol. 37, no. 5, pp. 789–797, 2011.
- [10] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.
- [11] P. J. Burt, "A gradient pyramid basis for pattern selective image fusion," in *Proceedings of the Society for Information Display Conference*, pp. 467–470, 1992.
- [12] A. Toet, "Image fusion by a ration of low-pass pyramid," *Pattern Recognition Letters*, vol. 9, no. 4, pp. 245–253, 1989.
- [13] G. Pajares and J. M. de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognition*, vol. 37, no. 9, pp. 1855–1872, 2004.
- [14] Y. Zheng, E. A. Essock, B. C. Hansen, and A. M. Haun, "A new metric based on extended spatial frequency and its application to DWT based fusion algorithms," *Information Fusion*, vol. 8, no. 2, pp. 177–192, 2007.
- [15] Y. Yang, D. S. Park, S. Huang, and N. Rao, "Medical image fusion via an effective wavelet-based approach," *Eurasip Journal on Advances in Signal Processing*, vol. 2010, Article ID 579341, 13 pages, 2010.
- [16] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Transactions on Image Processing*, vol. 4, no. 11, pp. 1549–1560, 1995.
- [17] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Information Fusion*, vol. 8, no. 2, pp. 143–156, 2007.
- [18] L. Yang, B. L. Guo, and W. Ni, "Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform," *Neurocomputing*, vol. 72, no. 1–3, pp. 203–211, 2008.
- [19] Q. Zhang and B.-L. Guo, "Multifocus image fusion using the nonsubsampling contourlet transform," *Signal Processing*, vol. 89, no. 7, pp. 1334–1346, 2009.
- [20] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 884–892, 2010.
- [21] W. Wu, X. M. Yang, Y. Pang, J. Peng, and G. Jeon, "A multifocus image fusion method by using hidden Markov model," *Optics Communications*, vol. 287, pp. 63–72, 2013.
- [22] Z. Wang, Y. Ma, and J. Gu, "Multi-focus image fusion using PCNN," *Pattern Recognition*, vol. 43, no. 6, pp. 2003–2016, 2010.
- [23] D. Agrawal and J. Singhai, "Multifocus image fusion using modified pulse coupled neural network for improved image quality," *IET Image Processing*, vol. 4, no. 6, pp. 443–451, 2010.
- [24] F. Zhang and H. Y. Chang, "Employing BP neural networks to alleviate the sparsity issue in collaborative filtering recommendation algorithms," *Journal of Computer Research and Development*, vol. 43, pp. 667–672, 2006.
- [25] P. F. Xiao and X. Z. Feng, *Segmentation and Information Extraction of High-Resolution Remote Sensing Image*, Beijing Science Press, 2012.
- [26] R. B. Huang, F. N. Lang, and Z. Shi, "Log-Gabor and 2D semi-supervised discriminant analysis based face image retrieval," *Application Research of Computers*, vol. 29, pp. 393–396, 2012.
- [27] Y. Chai, H. Li, and Z. Li, "Multifocus image fusion scheme using focused region detection and multiresolution," *Optics Communications*, vol. 284, no. 19, pp. 4376–4389, 2011.
- [28] H.-F. Li, Y. Chai, and X.-Y. Zhang, "Multifocus image fusion algorithm based on multiscale products and property of human visual system," *Control and Decision*, vol. 27, no. 3, pp. 355–361, 2012.

