INFORMATION
TECHNOLOGY JOURNAL

Multi-focus Image Fusion Based on The Nonsubsampled Contourlet Transform and Dual-layer PCNN Model

[1,2]Guojiang Xin, [1]Beiji Zou, [1,3]Jianfeng Li and [1]Yixiong Liang
[1]School of Information Science and Engineering, Central South University, Changsha, 410083, China
[2]Department of Information and Technology, Hunan Radio and TV University, Changsha, 410004, China
[3]School of Mathematics and Computer Science, Jishou University, Jishou, 416000, China

Abstract: Image fusion is an important research field of image processing. How to get the best fusion quality is not an easy problem for the researchers. The nonsubsampled contourlet transform (NSCT) is a multi-resolution tool for image fusion. For NSCT, how to get a better focus measurement is an important research content. In this study, a new model named dual-layer PCNN model is proposed. It simulates human visual perception mechanism. The model not only takes into account local neighbor relativity each other but also takes into account the relativity between before and after layers. Compared with the other PCNN models, this model uses the Shannon information entropy to adaptively control its iteration process. Based on this model and NSCT, a new image fusion method is proposed. In the method, the source images are decomposed by NSCT firstly and then the dual-layer PCNN model and local energy match rule are used to select the coefficients. At last, the fused image is reconstructed by taking an inverse NSCT. The experimental results show that the dual-layer PCNN model is a good focus measurement for NSCT and the method proposed in this study has better fusion performance than the other classical methods.

Key words: Image fusion, nonsubsampled contourlet transform, dual-layer PCNN model, Shannon information entropy, focus measurement

INTRODUCTION

Information fusion (Fan and Li, 2006; Cheng *et al.*, 2006) is an important application to our daily life. As a branch of it, image fusion has many applications, for example, in robotics, intelligent manufacturing, medical diagnosis, remote sensing, intelligent transportation, military applications and other fields (Cvejic *et al.*, 2009). Sometimes people like to get the image that has more information, that is to say that people want to see more and clearer objects in one image. This leads to the appearance of image fusion. As an important research field of image processing, image fusion is a process to combine the information from two or more source images of the same scene to obtain an image with complete information. Through the information processing, fusion method can enhance the understanding of the scene and highlight the goal.

During the last decade, many methods for image fusion have been proposed. According to the different fusion phases, image fusion can be usually carried out in three levels: pixel level, feature level and decision level. The pixel level is the maturest one. Among the existing methods, most of them are in pixel level, these methods

are to operate directly on the source images, pixel-by-pixel, using operators such as the weighted averaging. However, this often leads to many undesired effects (e.g., loss of contrast). In recent years, many methods based on Multi-Scale Transforms (MST) have been proposed for image fusion (Li and Yang, 2010; Redondo *et al.*, 2008; Petrovic and Cootes, 2007). The basic idea of these methods is to perform a MST on each source image firstly and then employ some fusion rules to construct a composite multi-scale representation of the fused image. The fused image is finally reconstructed by taking an inverse MST. The multi-scale image decomposition and reconstruction tools and the fusion rules are the two most important components of MST-based image fusion algorithm. The MST methods include contourlet transform (Zhang and Guo, 2009; Do and Vetterli, 2005), curvelet transform (Choi *et al.*, 2005; Kirankumar and Devi, 2007), the discrete wavelet transform (DWT) (Li *et al.*, 1995; Xiaodong *et al.*, 2005), various pyramid algorithms (Burt, 1992; Liu *et al.*, 2001) etc. These methods' fusion results depend on the focus measurement to a great extent.

DWT-based method is a classical method for image fusion. For DWT-based method, it is a good tool for

---

Corresponding Author: Beiji Zou, School of Information Science and Engineering, Central South University,
Changsha, 410083, China Tel: 86-731-88877701 Fax: 86-731-88877701

sparsely representing one-dimensional piecewise smooth signals. However, for two-dimensional signals, it cannot effectively represent the line and the curve discontinuities. Therefore, they need more coefficients to represent the line or the curve discontinuities properly. Moreover, it can capture only limited directional information, so it cannot represent the direction of the edges accurately (Zhang and Guo, 2009).

The contourlet transform is proposed to solve this problem and it is a better method for two-dimensional signal's representation. Compared with the traditional DWT, the contourlet transform is not only with multi-scale and localization but also with multi-direction and anisotropy. As a result, the contourlet transform can represent edges and other singularities along curves much more efficiently. However, due to downsamplers and upsamplers presented in both the laplacian pyramid and the directional filter banks; the contourlet transform is not shift-invariant which causes pseudo-Gibbs phenomena around singularities. Therefore, based on the contourlet transform, Nonsub Sampled Contourlet Transform (NSCT) is presented which is fully shift-invariant. It leads to better frequency selectivity and regularity than the foremost contourlet transform.

Many methods based on NSCT for image fusion were proposed during the recent years. The key of these methods is to choose the best rule to fuse the efficient to get the best result. Therefore, how to get a better focus measurement for them is an important research content.

For human, we can easily choose the most useful information from different source images. Therefore, if a method based on human visual perception mechanism can be proposed, it can fuse the image more easily and get better fusion quality than most of the existing methods. In the recent years, some researchers have begun to attempt to propose some new fusion methods simulating human visual perception mechanism. Toet *et al.* (2010) proposed a method toward cognitive for image fusion. Some methods using region segmentation or block segmentation (Yang *et al.*, 2008; Li and Yang, 2008), they firstly segment the source images to several regions or blocks, then fuse the source images according to the similarity of the corresponding region or block. Sasikala and Kumaravel (2007) proposed a fusion method based on feature extraction. To some extent, they all simulate human visual perception mechanism.

In the recent years, benefitting from the research of human visual perception mechanism, many methods simulating human visual perception mechanism have been proposed, the PCNN (Pulse Coupled Neural Network) model is one of them and it is a simplified model of biological visual system. It is a good technique for image processing and has been applied in many fields of image processing, such as image segmentation, image compression, object recognization, image fusion, etc. Eckhorn *et al.* (1990) proposed the PCNN model and Eckhorn (1999) applied it in scene segmentation, getting a good segmentation result. Wang *et al.* (2010) proposed a new model whose name is dual-channel PCNN model and applied it in the image fusion. The fusion effect is good.

To simulate human visual perception mechanism for image fusion, we propose a new model named dual-layer PCNN model. Based on it, we propose a new method for image fusion, whose name is image fusion based on the NSCT and dual-layer PCNN model. In the method, the source image is firstly decomposed by the NSCT to get many different coefficients and then the dual-layer PCNN model and local energy match rule are applied to select the different coefficients. After that the inverse-NSCT is used to fuse the coefficients to reconstruct the image.

CONTOURLET TRANSFORM

Standard contourlet transform: Contourlet Transform (CT) is a kind of multi-resolution analysis tool. It is a multi-scale and multi-direction framework of digital image. In the transform, the multi-scale analysis and the multi-direction analysis are separated in a serial way. The Laplacian Pyramid (LP) is firstly used to capture the point discontinuities and the LP decomposition at each level generates a downsampled low-pass version of the original and the difference between the original and the prediction error. The process can be iterated on the coarse (downsampled low-pass) signal. Then the process is followed by a Directional Filter Bank (DFB) to link point discontinuities into the linear structures. The DFB avoids the modulation of the input image and has a simple rule for expanding the decomposition tree. The combination of LP and a DFB is a double filter bank structure. The overall result is an image expansion using basic elements like contour segments. The framework of CT is shown in Fig. 1.

NSCT: Compared with the foremost contourlet transform, the nonsubsampled pyramid structure and nonsub-sampled directional filter banks are employed in NSCT. The DFB is achieved by switching off the downsamplers/upsamplers in each two-channel filter bank in the DFB tree structure and upsampling the filters accordingly. As a result, NSCT is shift-invariant and leads to better frequency selectivity and regularity than the foremost contourlet transform. Figure 2 shows the decomposition framework of NSCT.
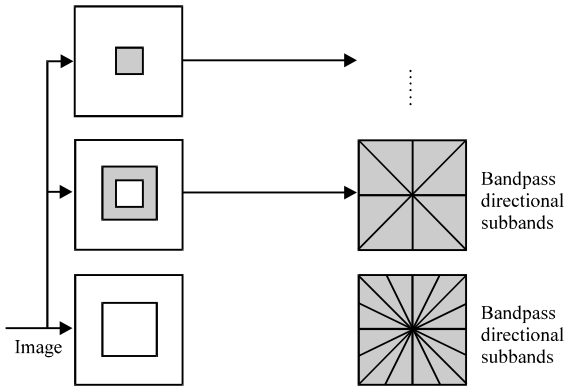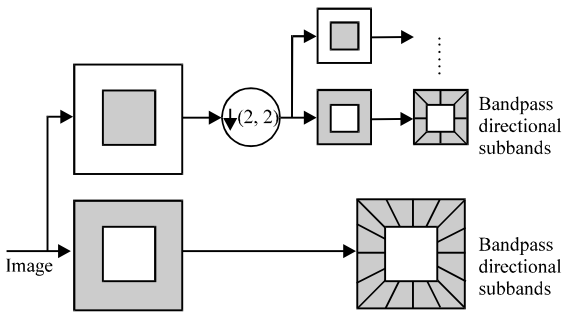
Fig. 1: The framework of contourlet transform

Fig. 2: The framework of NSCT

PCNN MODEL

Standard PCNN model: The PCNN model is a simplified model of biological visual system and it is a two-dimensional neural array model. It has been widely used in image processing. The PCNN model consists of three parts: the dentritic tree, the linking modulation and the pulse generator. Its mathematical model is expressed by Eq. 1-5:

$$F_{ij}(n) = e^{-\alpha_F}F_{ij}(n-1) + V_F\sum_{k,l}M_{ijkl}Y_{kl}(n-1) + S_{ij} \qquad (1)$$

$$L_{ij}(n) = e^{-\alpha_L}L_{ij}(n-1) + V_L\sum_{k,l}W_{ijkl}Y_{kl}(n-1) \qquad (2)$$

$$U_{ij}(n) = F_{ij}(n)(1+\beta L_{ij}(n)) \qquad (3)$$

$$E_{ij}(n) = e^{-\alpha_E}E_{ij}(n-1) + V_E Y_{ij}(n-1) \qquad (4)$$

$$Y_{ij}(n) = \begin{cases} 1 & U_{ij}(n) > E_{ij}(n-1) \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

Here, the dentritic tree is given by Eq. 1-2, the linking modulation is given by Eq. 3-4 and the pulse genetator is given by Eq. 5. The dentritic tree's role is to receive inputs from other neurons and external sources by two channels: F and L, F is the feeding channel and L is the linking channel. S is the external stimulus. The linking i and j refer to the pixel positions in image, k and l are the dislocation in a symmetric neighborhood around a pixel and n denotes the current iteration (discrete time step). The M and W are the constant synaptic weights and V_F, V_L and V_E are the magnitude scaling terms. α_F, α_L and α_E are the time delay constants of the neuron and β is the linking strength. Y is the pulse output.

However, the standard PCNN model is not very suitable for image fusion, because in PCNN model, each neuron has one input, usually multiple PCNN models are needed when applied to image fusion. This is not convenient and economical for a real system. Therefore, this shortcoming hampers PCNN's application in the field of image fusion. Is there any other way to modify the PCNN model to suit for the application of image fusion? The answer is yes, we propose a new PCNN model named dual-layer PCNN model based on the standard PCNN model.

Then, let's introduce the dual-layer PCNN model in detail.

The dual-layer PCNN model: Eckhorn has proposed a two-layer neuron model for scene segmentation. This two-layer neuron model is with feedback and top-down activity control. An additional negative feedback control keeps the two-layer network in an apropriate working range for high contrast region. Based on this idea, we proposed a new PCNN model whose name is dual-layer PCNN model, its structure is shown in Fig. 3.

The dual-layer PCNN model consists of two layers: layer 1 and layer 2. Layer 1 consists of two parts: part A and part B, each part represents a neuron's activity. In each part, there are also three parts: the dentritic tree, the linking modulation and the pulse generator which is as same as the standard PCNN model's. Layer 2 consists of one part and its structure is also as same as the standard PCNN model's.

The neurons in the layer 1 receive the input stimuli through the receptive fields. The stimuli include external inputs, their surrounding neurons' stimuli and the negative feedback from the layer 2's neurons; then, the stimuli are divided into two channels, one is feeding channel F, the other is linking channel L. In the modulation part, the linking input is weighted with β, then multiplied with the
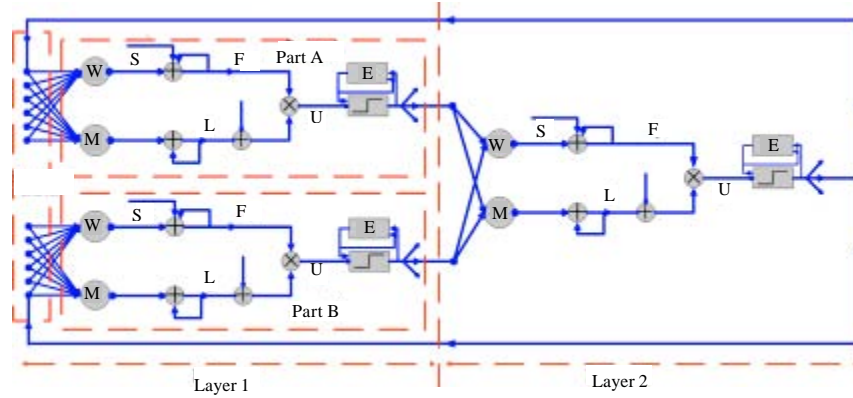
Fig. 3: The structure of dual-layer PCNN model

feeding input. The internal activity U is the output of the modulation part. In succession, the pulse generator compares U with a threshold E. If U is larger than E, the neuron will emit a pulse, it is also called "fire", otherwise, it will not "fire". The pulse output is sent to the layer 2's neurons. In layer 2, most processing is as same as layer 1, at last, the negative feedback is sent to layer 1's neurons. In the model, the negative feedback acts from each neuron in the layer 2 on feeding channel of neurons in layer 1.

The dual-layer PCNN model's mathematical model is shown below and it includes two parts: Layer 1's mathematical model and layer 2's mathematical model.

Layer 1's mathematical model:

$$F_{ij}^1(n) = e^{-\alpha_F^1}F_{ij}^1(n-1) + V_F^1\sum_{k,l}M_{ijkl}^1 Y_{kl}^2(n-1) + S_{ij}^1 \qquad (6)$$

$$L_{ij}^1(n) = e^{-\alpha_L^1}L_{ij}^1(n-1) + V_L^1\sum_{k,l}W_{ijkl}^1 Y_{kl}^2(n-1) \qquad (7)$$

$$U_{ij}^1(n) = F_{ij}^1(n)(1+\beta_1 L_{ij}^1(n)) \qquad (8)$$

$$E_{ij}^1(n) = e^{-\alpha_E^1}E_{ij}^1(n-1) + V_E^1 Y_{ij}^1(n-1) \qquad (9)$$

$$Y_{ij}^1(n) = \begin{cases} 1 & U_{ij}^1(n) > E_{ij}^1(n-1) \\ 0 & otherwise \end{cases} \qquad (10)$$

Layer 2's mathematical model:

$$F_{ij}^2(n) = e^{-\alpha_F^2}F_{ij}^2(n-1) + V_F^2\sum_{k,l}(M_{ijkl}^{2A}Y_{kl}^{1A}(n) + M_{ijkl}^{2B}Y_{kl}^{1B}(n)) + S_{ij}^2 \qquad (11)$$

$$L_{ij}^2(n) = e^{-\alpha_L^2}L_{ij}^2(n-1) + V_L^2\sum_{k,l}(W_{ijkl}^{2A}Y_{kl}^{1A}(n) + W_{ijkl}^{2B}Y_{kl}^{1B}(n)) \qquad (12)$$

$$U_{ij}^2(n) = F_{ij}^2(n)(1+\beta_2 L_{ij}^2(n)) \qquad (13)$$

$$E_{ij}^2(n) = e^{-\alpha_E^2}E_{ij}^2(n-1) + V_E^2 Y_{ij}^2(n-1) \qquad (14)$$

$$Y_{ij}^2(n) = \begin{cases} 1 & U_{ij}^2(n) > E_{ij}^2(n-1) \\ 0 & otherwise \end{cases} \qquad (15)$$

Here, the parameters superscript "1" or "2" denote the parameters themself belonging to layer 1 or layer 2. F^1 stands for the feeding channel for part A and part B in layer 1 and F^2 stands for the feeding channel in layer 2. L^1 stands for the linking channel for part A and part B in layer 1 and L^2 stands for the linking channel in layer 2. Y^1 is the pulse output of layer 1 and Y^2 is the negative feedback of layer 2. Other parameters' meaning is as same as the standard PCNN model's. In layer 2, Y^{1A} and Y^{1B} is the pulse output from part A and part B in layer 1. M^{2A}, M^{2B}, W^{2A} and W^{2B} are the constant synaptic weights corresponding to part A and B in layer 1.

IMAGE FUSION ALGORITHM

At first, the source images must be pre-processed to get the same resolution. After the pre-processing, the source images are decomposed by the NSCT to get many different coefficients, including high-frequency coefficients and low-frequency coefficients. Then different fusion rules (dual-layer PCNN model and local energy match rule) are used to select the coefficients. At last, the inverse NSCT is used to fuse the high-frequency coefficients and the low-frequency coefficients together to get the fused image. The flowchart is shown in Fig. 4.

Low-frequency coefficients fusion algorithm: By the NSCT analysis, we know that the NSCT divides the image's level into two parts: coarse and fine. From the
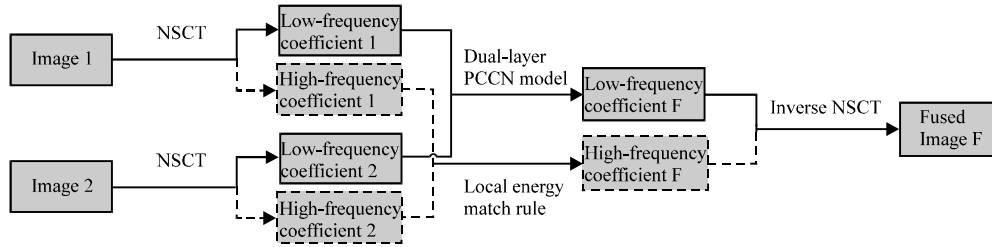
Fig. 4: The flowchart of image fusion

distribution of frequencies, the low-frequency coefficients are assigned to the coarse part and the high-frequency coefficients are assigned to the fine part. The spectrum and background information of the image are all included in the low-frequency coefficients through the NSCT while the detail information is in high-frequency coefficients.

The NSCT is close to the wavelet transform in low-frequency region, image component including main energy decides image contour. There are many methods to select the low-frequency coefficients, including maximum pixel method, minimum pixel method, computing the average pixel-pixel gray level value method, local region deviation method, etc.

Maximum pixel method, minimum pixel method and computing the average pixel-pixel gray-level value method do not take into account local neighbor relativity each other, so fusion result can not get good effect; local region deviation method takes into account local neighbor relativity each other but do not take into account image edge and definition.

Accounting to these lacks, we adopt dual-layer PCNN model to select the low-frequency coefficients. Because the dual-layer PCNN model simulates human visual perception mechanism. It not only takes into account the local neighbor relativity each other but also takes into account relativity between before and after layers.

There are four steps for the low-frequency coefficients selection with dual-layer PCNN model. To be brief, the simple description is shown firstly below:

Step 1: Coefficients adjustment and initialization
Step 2: PCNN iteration and fire
Step 3: Judging PCNN' iteration times
Step 4: Coefficients selection

Then, let's explain the steps in detail, respectively.

Coefficients adjustment and initialization: After NSCT decomposition, there may be N coefficients matrices of

same size in the same scale and same orientation; we also call them sub-bands or sub-images. As we know, because of the transform, there may be negative coefficients in the coefficients matrices; so the coefficients' sign must be adjusted firstly. All the negative coefficients should be adjusted to be positive. Because PCNN model just cares the difference between the coefficients, we add the absolute value of the minimum to all elements in the same scale in each matrix to adjust all the coefficients to be positive.

After the sign adjustment, normalize the N sub-bands to [0, 1] to get the neuron inputs of dual-layer PCNN model.

At the same time, initialize $F = 0$, $L = 0$, $U = 0$, $Y = 0$, $E = 20$ and set the value of the parameters: α_F, α_L, α_E, V_F, V_L, V_E.

PCNN iteration and fire: After getting the normalizd matrices, put the two images' coefficients matrices to the dual-layer PCNN model, then the model begins the iteration process. During each iteration process, in layer 1, part A and B fire, respectively, the pulse output of them is sent to layer 2 to stimulate the neurons' activity in layer 2. Then negative feedback from layer 2 is sent to layer 1 to stimulate neurons' activity in layer 1.

Judging PCNN' iteration times: The model can control the iteration times adaptively. In layer 2, the pulse output matrix's Shannon information entropy will be calculated during each iteration process, if the Shannon information entropy gets the maximum, the iteration will stop, the iteration times and the matrix are the best one and otherwise, the iteration process will be continued.

Shannon information entropy's mathematical model is expressed as:

$$H(P) = -P_1 \ln P_1 - P_0 \ln P_0 \qquad (16)$$

Here, P_1 and P_0 represent the probability of the pulse output will be 1 or 0, respectively.

Coefficients selection: Supposed H_A and H_B are the matrices of the firing times for part A and part B in layer 1, C_A and C_B are the coefficients matrices for part A and B in layer 1, C_F is the coefficients matrix after PCNN selection in layer 2. (i, j) represents the neuron in (i, j). The coefficients selection rule is expressed as:

$$C_F(i,j) = \begin{cases} C_A(i,j) & H_A(i,j) > H_B(i,j) \\ C_B(i,j) & H_A(i,j) < H_B(i,j) \\ (C_A(i,j) + C_B(i,j))/2 & H_A(i,j) = H_B(i,j) \end{cases} \quad (17)$$

High-frequency coefficients fusion algorithm: In the high-frequency region, there is more detail information in the coefficients, such as edge and texture features. In order to make full use of information in the neighborhood and cousin coefficients in the NSCT domain, the local energy match rule is adopted to select the high-frequency coefficients. At first, calculate each pixel's local energy in the high-frequency sub-bands. The local energy's formula is shown below:

$$E_{i,j}(p_1,p_2) = \sum_{m,n \in S_{M \times N}} \left| C_{i,j}(p_1 + m, p_2 + n) \right|^2 \quad (18)$$

Here, $S_{M \times N}$ denotes the regional window whose size is M×N, C is the coefficients in all scales and orientations. Local energy, rather than single pixel value, will be more reasonable to extract features of source images by utilizing neighbors' information. Large region energy means important image information. Note that the size of region energy map is equal to the size of each sub-band.

Then the energy match rule is defined as:

$$EM_{i,j}(p_1,p_2) = \frac{2 \sum_{m \in M, n \in N} \left| C_{i,j}^1(p_1 + m, p_2 + n) C_{i,j}^2(p_1 + m, p_2 + n) \right|}{E_{i,j}^1(p_1 + p_2) + E_{i,j}^2(p_1 + p_2)} \quad (19)$$

If $EM_{i,j} < T$ (T is the pre-set matching threshold), then the coefficients of the high-frequency sub-bands are selected by:

$$C_{i,j}^F(p_1,p_2) = \begin{cases} C_{i,j}^1(p_1,p_2) & E_{i,j}^1(p_1,p_2) \geq E_{i,j}^2(p_1,p_2) \\ C_{i,j}^2(p_1,p_2) & E_{i,j}^1(p_1,p_2) < E_{i,j}^2(p_1,p_2) \end{cases} \quad (20)$$

Else by:

$$C_{i,j}^F(p_1,p_2) = \begin{cases} (1-\beta)C_{i,j}^1(p_1,p_2) + \beta C_{i,j}^2(p_1,p_2), & \text{if } E_{i,j}^1(p_1,p_2) \geq E_{i,j}^2(p_1,p_2) \\ \beta C_{i,j}^1(p_1,p_2) + (1-\beta)C_{i,j}^2(p_1,p_2), & \text{if } E_{i,j}^1(p_1,p_2) < E_{i,j}^2(p_1,p_2) \end{cases} \quad (21)$$

Here:

$$\beta = \frac{1}{2} - \frac{1}{2} \cdot \frac{1 - EM_{i,j}(p_1,p_2)}{1 - T}$$

RESULT ANALYSIS

To verify the performance of the proposed method, many multi-focus images were selected to do the experiments. Because of the lack of space, only two sets of experimental results are presented here (two sets of 256-level images, namely letters and alarm clock). The two sets of test source images used in the experiments are shown in Fig. 5. In the experiment, we use three scales of
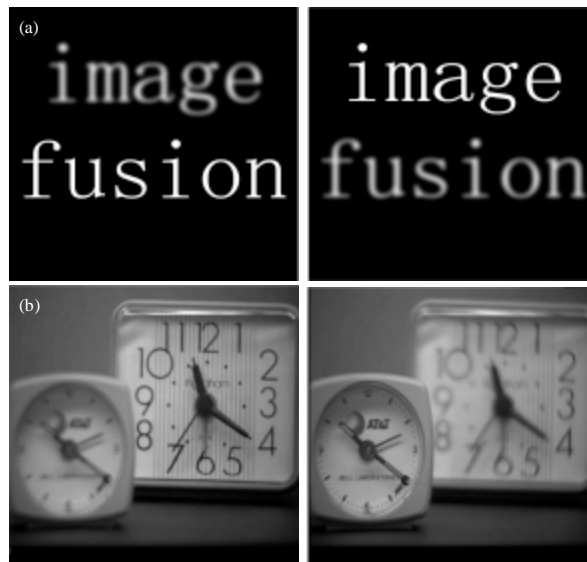

Fig. 5: Multi-focus source images; (a) letters and (b) alarm clock

decomposition for NSCT and two scales of high-frequency coefficients are obtained. For the coarse scale, there are two directions and for the fine, there are eight directions. Thus, for each source image, 11 subimages with the same size as the source image are generated. The experiments were implemented on an Intel (R) Core (TM) i3 M330 2.13 GHz computer with 2.0 GB RAM. The simulation software is Matlab 2006.

To do the comparison, we contrast with the classical fusion methods such as average, DWT with DBSS, maximum, minimum, contrast pyramid, gradient pyramid, laplacian pyramid, morphological pyramid, PCA, SIDWT with Haar, standard PCNN. In this work, we use both qualitative and quantitative analysis to test the fused results. The qualitative analysis judges the quality of the fused results visually and the quantitative analysis includes RMSE and SSIM indices.

Figure 6 and 7 are the experimental results. In each figure, there are twelve fused results obtained by different methods. Again, for chearer comparison, parts of the reference image and fusion results are extracted and put into Fig. 9-12 show the performance of different methods.

In Fig. 6a-*l* to 8a-*l* represent different methods which are average, DWT with DBSS, maximum, minimum, contrast pyramid, gradient pyramid, laplacian pyramid, morphological pyramid, PCA, SIDWT with Haar, standard PCNN, our method, respectively.

Qualitative analysis: From Fig. 6 and 7, we can see that the fused results of different methods are not same, results (d), (e), (g), (j) and (l) are clearer than the others, results (b) and (f) have the shadows of the objects, results (a), (h) and (i) have the halo, results (c) and (k) are darker than the others.

Though the results (d), (e), (g), (j) and (l) are clearer than the others, as illustrated in Fig. 8, from the partially enlarging region labeled with red line of the results, we can see that there is still little difference between them. The edge of letters is blurry in (d) and (j) and there are dark dots in the letters in (e) and (g). The fused result of our method looks the best.

Quantitative analysis: In addition to visual analysis, we conducted a quantitative analysis. To evaluate the results, we choose different objectivity evaluation indices. The evaluation indices of these methods are used to evaluate the fusion performance quantitatively. They are the Root Mean Squared Error (RMSE) index and the structural similarity (SSIM) index.
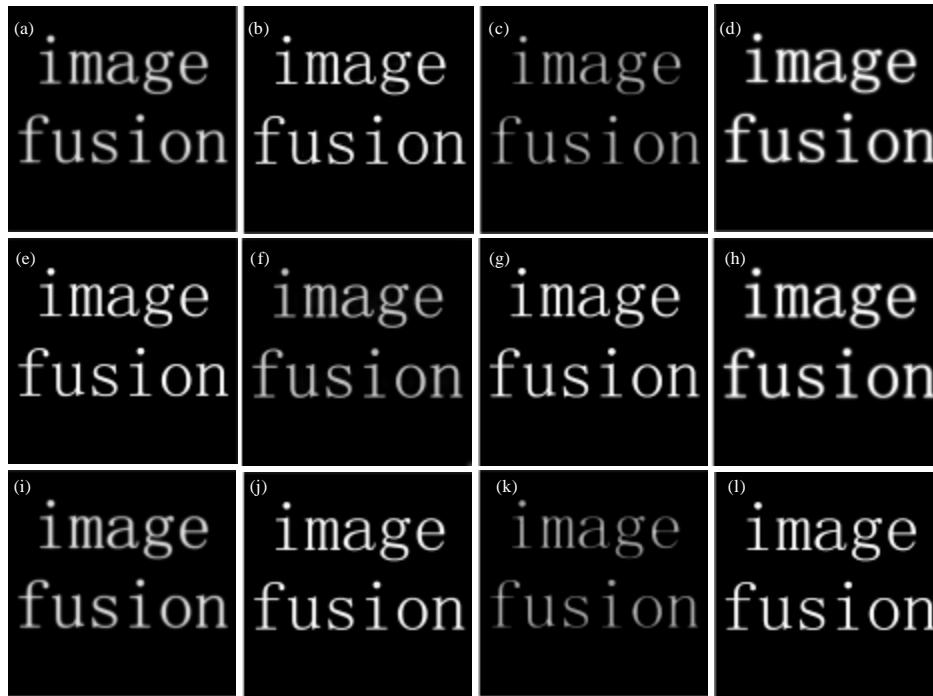

Fig. 6: Fusion results by different methods (letters)
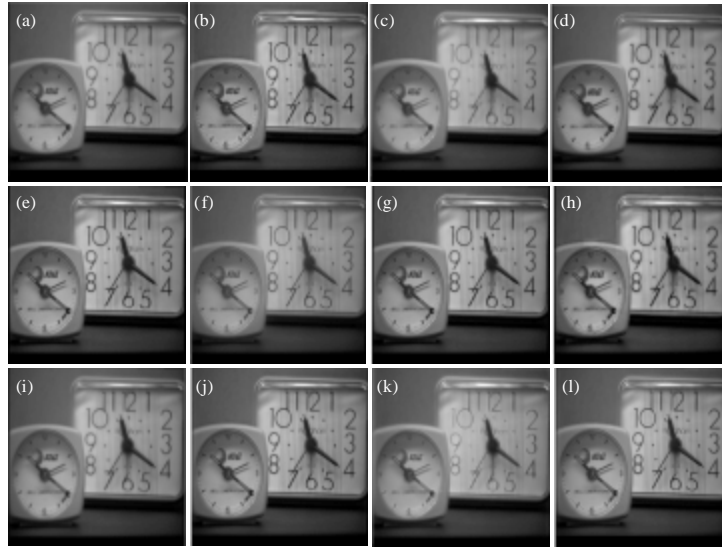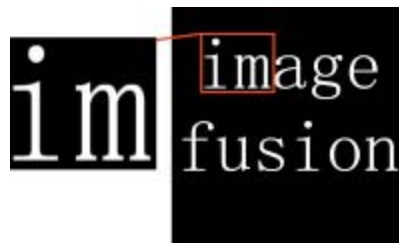
Fig. 7: Fusion results by different methods (alarm clock)

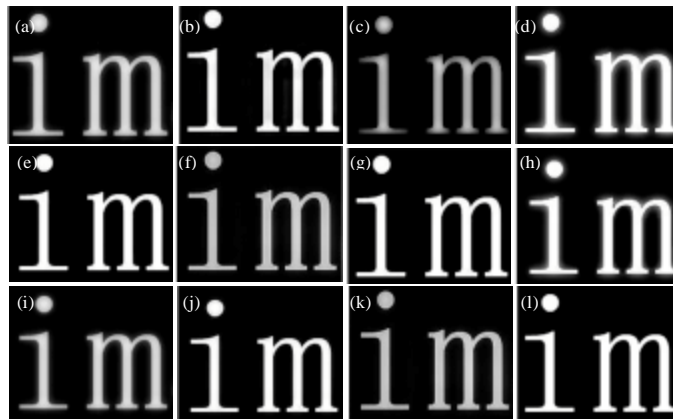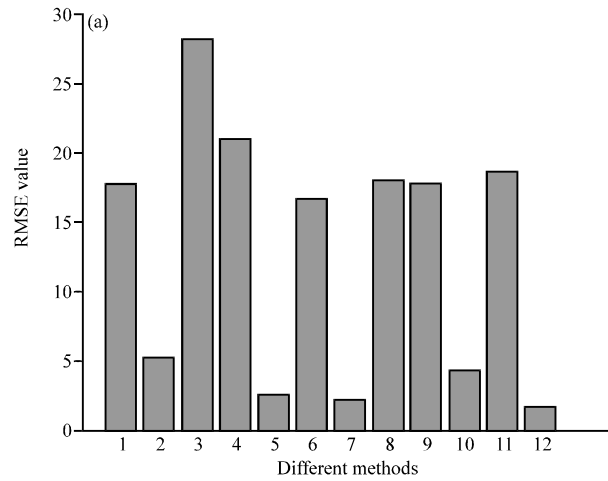The partially enlarging region of reference image

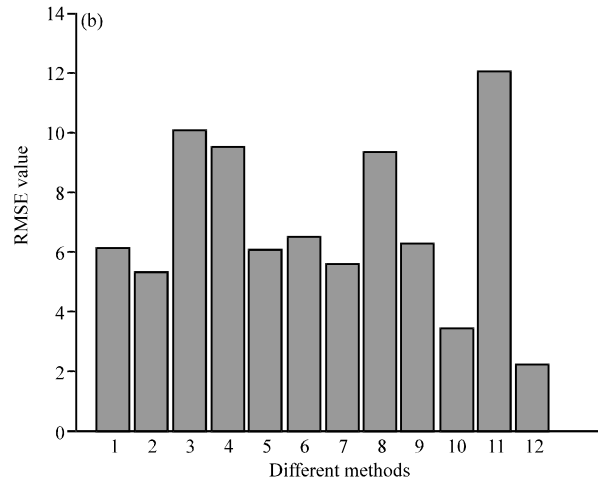Fig. 8: The partially enlarging image of fusion results by different methods (letters)

The Root Mean Squared Error (RMSE) index: RMSE is used to evaluate the performance of the focus measures. Here, R and F represent reference image and fused image respectively. RMSE's mathematical model is shown as follow:

$$RMSE = \sqrt{\frac{1}{M \times N} \sum_{i=i}^{M} \sum_{j=1}^{N} [F(i,j) - R(i,j)]^2} \qquad (22)$$

Usually, the value of RMSE is less, the performance for the fusion method is better. Figure 9, 10 show the

Number	1	2	3	4
Method	Average	DWT with DBSS	Maximum	Miniimum
Number	5	6	7	8
Method	Contrast pyramid	Gradient pyramid	Laplacian pyramid	Morphlogical pyramid
Number	9	10	11	12
Method	pca	SIDWT with haar	PCCN	Ours

Number	1	2	3	4
Method	Average	DWT with DBSS	Maximum	Miniimum
Number	5	6	7	8
Method	Contrast pyramid	Gradient pyramid	Laplacian pyramid	Morphlogical pyramid
Number	9	10	11	12
Method	pca	SIDWT with haar	PCCN	Ours

Fig. 9: RMSE performance of single fusion result by different methods: (a) letters and (b) alarm clock

RMSE performance of different methods. The RMSE performance of the single fused result (letters and alarm clock) is shown in Fig. 9a and b, respectively. Figure 10 is the contrast results of ten images between our method and the others, in which, Fig. 10a is the contrast result between our method and contrast pyramid, DWT with DBSS, laplacian pyramid, SIDWT with Haar, Fig. 10b is the contrast result between our method and average, gradient pyramid, maximum, minimum, morphological

pyramid, PCA, standard PCNN. The values of RMSE of the methods in Fig. 10a are generally smaller than those in Fig. 10b. From the contrast results, we can see that the RMSE value of our method is smaller than the other methods.

The structural similarity (SSIM) index: The structural similarity (SSIM) index's mathematical model is shown as follow:
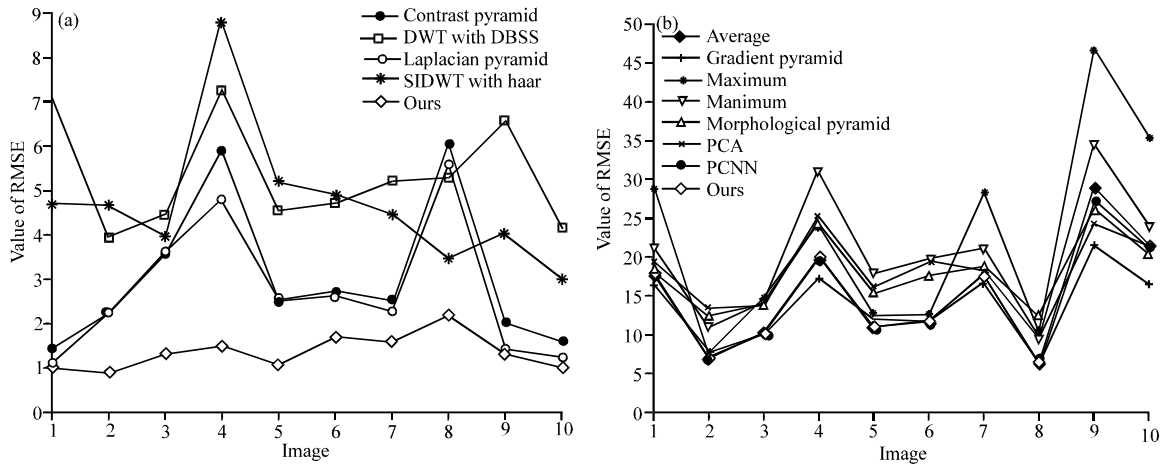
Fig. 10: RMSE performance of fusion results of ten images by different methods
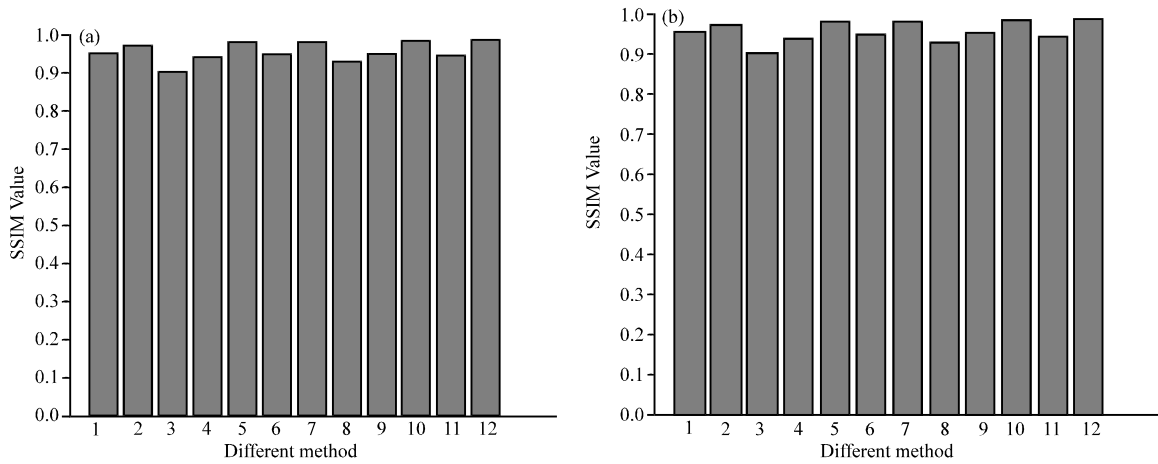

Fig. 11: SSIM performance of single fusion result by different methods: (a) letters and (b) alarm clock 1: Average, 2: DWT with DBSS, 3: Maximum, 4: Minimum, 5: Contrast pyramid, 6: Gradient pyramid, 7: Laplacian pyramid, 8: Morphological pyramid, 9: pca, 10: SIDWT with haar, 11: PCNN and 12: Ours
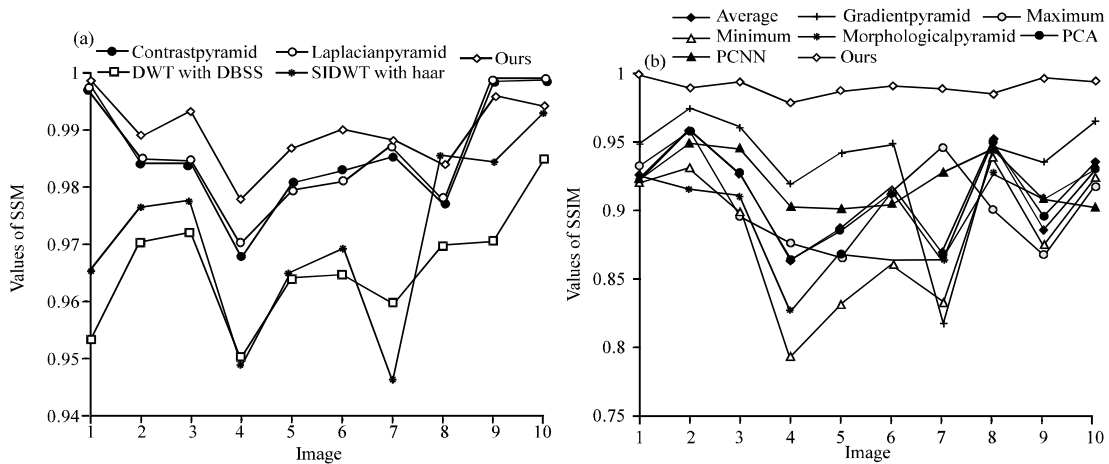

Fig. 12: SSIM performance of fusion results of ten images by different methods

$$SSIM = \frac{(2\mu_R\mu_F - C_1)(2\sigma_{RF} + C_2)}{(\mu_R^2 + \mu_F^2 + C_1)(\sigma_R^2 + \sigma_F^2 + C_2)} \qquad (23)$$

Here, $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$ where, L is the dynamic range of the pixel values (255 for 8-bit grayscale images); K_1, $K_2 \ll 1$ are small constants. SSIM index describes the similarity of two inputs. Larger value shows both inputs more similar. The values of SSIM index in the paper are computed with default parameters. Figure 11 and 12 show the SSIM performance of different methods. The SSIM performance of the single fused result (letters and alarm clock) is shown in Fig. 11a and b, respectively. Figure 12a and b are the contrast results of ten images between our method and the other methods, among them, Fig. 12a is the contrast result between our method and contrast pyramid, DWT with DBSS, laplacian pyramid, SIDWT with Haar, Fig. 12b is the contrast result between our method and average, gradient pyramid, maximum, minimum, morphological pyramid, PCA, standard PCNN. From Fig. 11, we can see that SIDWT with Haar has the largest SSIM value. However, From Fig. 12, we can see, in the most time, the SSIM value of our method is larger than the other methods.

CONCLUSION

This study presents a new model named dual-layer PCNN model; this model simulates human visual perception mechanism. Based on it, a novel image fusion method is proposed. Compared with the other fusion methods based on PCNN model, it has the adaptive ability and can automatically select the best coefficients. This method could better describe the edge direction of images and analyze the feature of images better. According to the experimental results, on the vision, our method acquires better fusion result than the other classical methods. In objective evaluation criteria, our method's fusion characteristic is better than the other classical methods.

ACKNOWLEDGMENT

REFERENCES

Burt, P.J., 1992. A gradient pyramid basis for pattern-selective image fusion. Proc. Soc. Inform. Display Conf., 23: 467-470.

Cheng, L., L. Kong and X. Li, 2006. Solving information fusion problems on unreliable evidential sources with generalized DSmT. J. Applied Sci., 6: 1581-1585.

Choi, M., R.Y. Kim, M.R. Nam and H.O. Kim, 2005. Fusion of multispectral and panchromatic satellite images using the curvelet transform. IEEE Geosci. Remote Sens. Lett., 2: 136-140.

Cvejic, N., T. Seppanen and S.J. Godsill, 2009. A nonreference image fusion metric based on the regional importance measure. IEEE J. Selected Topics Signal Process., 3: 212-221.

Do, M.N. and M. Vetterli, 2005. The contourlet transform: An efficient directional multi-resolution image representation. IEEE Trans. Image Process., 14: 2091-2106.

Eckhorn, R., 1999. Neural mechanisms of scene segmentation: Recordings from the visual cortex suggest basic circuits for linking field models. IEEE Trans. Neural Networks, 10: 464-479.

Eckhorn, R., H.J. Reitboeck, M. Arndt and P. Dicke, 1990. Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. Neural Comput., 2: 293-307.

Fan, E.L.Y. and X. Li, 2006. Information fusion technology with application to the localization of pipeline leakage based on negative pressure wave detection. J. Applied Sci., 6: 2010-2013.

Kirankumar, Y. and S.S. Devi, 2007. Medical image fusion transforms-2d approach. J. Medical Sci., 7: 870-874.

Li, H., B.S. Manjunath and S.K. Mitra, 1995. Multisensor image fusion using the wavelet transform. Graphic. Models Image Process., 57: 235-245.

Li, S. and B. Yang, 2008. Multifocus image fusion using region segmentation and spatial frequency. Image Vision Comput., 26: 971-979.

Li, S. and B. Yang, 2010. Hybrid multiresolution method for multisensor multimodal image fusion. IEEE Sensors J., 10: 1519-1526.

Liu, Z., K. Tsukada, K. Hanasaki, Y.K. Ho and Y.P. Dai, 2001. Image fusion by using steerable pyramid. Pattern Recognition Lett., 22: 929-939.

Petrovic, V. and T. Cootes, 2007. Objectively adaptive image fusion. Inform. Fusion, 8: 168-176.

Redondo, R., S. Fischer, F. Sroubek, G. Cristobal, 2008. A 2D wigner distribution-based multisize windows technique for image fusion. J. Visual Commun. Image Representation, 19: 12-19.

Sasikala, M. and N. Kumaravel, 2007. A comparative analysis of feature based image fusion methods. Inform. Technol. J., 6: 1224-1230.

Toet, A., M.A. Hogervorst, S.G. Nikolov, J.J. Lewis, T.D. Dixon, D.R. Bull and C.N. Canagarajah, 2010. Towards cognitive image fusion. Inform. Fusion, 11: 95-113.

Wang, Z., Y. Ma and J. Gu, 2010. Multi-focus image fusion using PCNN. Pattern Recognit., 43: 2003-2016.

Xiaodong, Z., H. Xinghan and W. Ming, 2005. Object edge smoothing by wavelet transformation. Inform. Technol. J., 4: 451-455.

Yang, C., J.Q. Zhang, X.R. Wang and X. Liu, 2008. A novel similarity based quality metric for image fusion. Inform. Fusion, 9: 156-160.

Zhang, Q. and B.L. Guo, 2009. Multifocus image fusion using the nonsubsampled contourlet transform. Signal Process., 89: 1334-1346.