IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 4, 2017, accepted April 15, 2017, date of publication May 19, 2017, date of current version August 22, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2698217

A Hybrid Method for Multi-Focus Image Fusion Based on Fast Discrete Curvelet Transform

YONG YANG[1], (Senior Member, IEEE), SONG TONG[2], (Student Member, IEEE), SHUYING HUANG[3], (Member, IEEE), PAN LIN[4], (Member, IEEE), AND YUMING FANG[1], (Member, IEEE)

[1]School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China
[2]Department of Intelligence Science and Technology, Kyoto University, Kyoto 606-8501, Japan
[3]School of Software and Communication Engineering, Jiangxi University of Finance and Economics, Nanchang 330032, China
[4]Institute of Biomedical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Corresponding author: Shuying Huang (shuyinghuang2010@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61462031, Grant 61662026, and Grant 61473221, in part by the Key Project of Chinese Ministry of Education under Grant 211087, in part by the Natural Science Foundation of Jiangxi Province under Grant 20151BAB207033 and Grant 20161ACB21015, and in part by the Project of the Education Department of Jiangxi Province under Grant KJLD14031, Grant GJJ150461, and Grant GJJ150438. The work of S. Tong was supported by the China Scholarship Council under Grant 201608050017.

ABSTRACT This paper presents a fast discrete Curvelet transform (FDCT)-based technique for multi-focus image fusion to address two problems: texture selection in FDCT domain and block effect in spatial-based fusion. First, we present a frequency-based model by performing FDCT on the input images. Considering the human visual system characteristics, a union of pulse coupled neural network and sum-modified-Laplacian algorithms are proposed to extract the detailed information of frequencies. Then, we construct a hybrid spatial-based model. Unlike other spatial-based methods, we combine the image difference and the detailed information extracted from input images to detect the focused region. Finally, to evaluate the robustness of proposed method, we design a completed evaluation process considering the misregistration, noise error, and conditional focus situations. Experimental results indicate that the proposed method improves the fusion performance and has less computational complexity compared with various exiting frequency-based and spatial-based fusion methods.

INDEX TERMS Multi-focus image fusion, discrete Curvelet transform, block effect, human visual system.

I. INTRODUCTION

Visual sensor network (VSN), wherein camera-equipped sensor nodes can capture, process, and transmit visual information, is an important research topic attracting considerable attention in recent years [1]. However, the camera sensors collect a huge amount of visual data, which are rich in information and offer tremendous potential when used in VSN [2].

In practical systems, the huge amount of image data generated by image sensors have to be transmitted for various applications. This amount of data directly affects the power consumption of sensor nodes or, in other words, decreases the life time of the network. Therefore, in VSN, there are fusion centers used to integrate a large number of high rate image data. However, current image fusion approaches usually perform complex fusion operations, which will consume a significant portion of the battery power in a VSN. To meet the requirement of resource- and bandwidth-limited VSN, low-complexity image fusion techniques are much desired in practice [4].

As we know, the focused range of visible imaging system is limited. Thus, it is difficult to obtain all objects in the same scene clearly [5], [6]. Fig. 1(a) shows the basic principle of optical imaging characteristics [7]. Each point in a scene is projected onto a single point on the focal plane, leading to a focused image to be formed on it. However, if the sensor plane does not coincide with the focal plane, the image formed on the sensor plane will be a circular disk known as a blur circle with the diameter $2r$. Based on the property of geometrical optics imaging, the intensity distribution of the blur circle is supposed to be well-distributed. In reality, however, diffraction effects and characteristics of the system play a major role in forming the intensity distribution within the blur circle. After considering the effect of lens aberrations, the point spread function (PSF) of an optic system can be

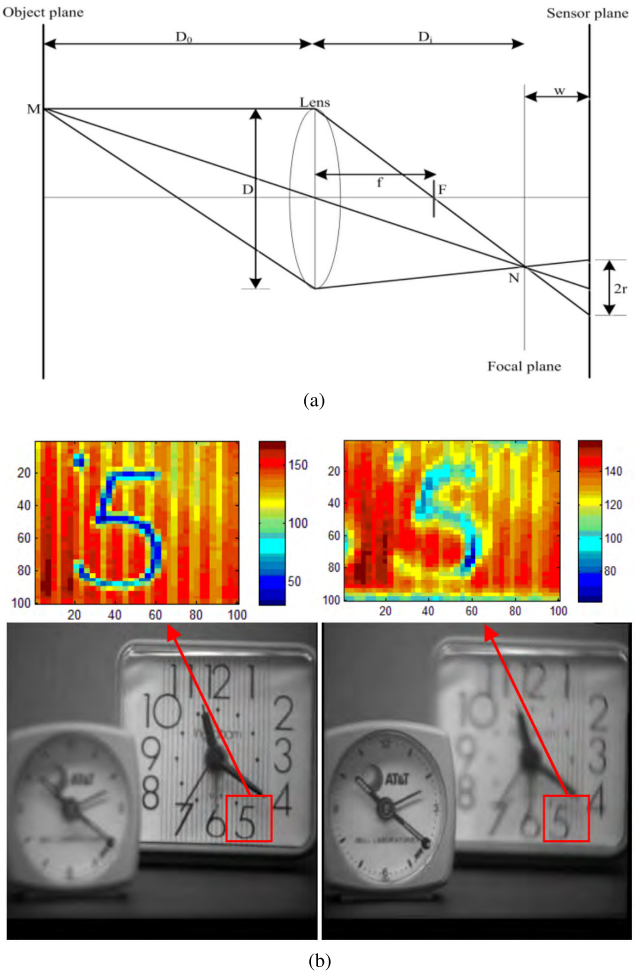


FIGURE 1. Multi-focus images. (a) The basic principle of optical imaging characteristics. (b) Two multi-focus images: the left one focuses on the big clock, the right one focuses on the small clock.

obtained by a 2D Gaussian function [7]:

$$h(x, y) = \exp\left(-\left(x^2 + y^2\right)/2\sigma^2\right)/2\pi\sigma^2, \qquad (1)$$

where x and y are image coordinates, σ is not only the standard deviation but also the spread parameter which is proportional to the blur radius r of the blur circle. Therefore, larger σ would result in the more blur image [6]. Fig. 1(b) shows two common test images with different focus settings, where the left image focuses on the big clock, and the right one focuses on the small clock. As shown in Fig. 1(b), when the big clock is focused, the selected block is clear and its corresponding surface is sharp, however, when the big clock is unfocused, the selected block is blurring and its corresponding surface is flat. Hence, the pseudo-color of the surface is related to the clarity of image block. A popular technique to overcome this problem is to utilize multi-focus image fusion techniques, in which one can obtain one image with all of the objects in focus by means of containing the best information from multiple original images [8].

Multi-focus image fusion is an important subfield of image fusion. Of course, with or without modification, many algorithms for merging multi-focus images can also be employed for other image fusion tasks such as visible-infrared image fusion and multi-modal medical image fusion [8], [9]. Generally, multi-focus image fusion algorithms can be divided into pixel level, feature level and decision level methods [10]. Pixel level image fusion methods can perform on the images in *spatial domain* and *frequency domain* [11]–[13].

The traditional spatial domain methods use the techniques of principal component analysis (PCA), and intensity-hue-saturation (IHS) [14], [15]. The fused images obtained by PCA and IHS methods have high spatial quality, but they usually overlook the high quality of spectral information and suffer from spectral degradation [16], [17]. Tian J. *et al.* proposed the bilateral gradient-based sharpness criterion (SCBG) based multi-focus image fusion method [3], which can be used to evaluate the local content (sharp) information well. Although the SCBG is of low computational complexity which suits VSN, the result always has *block effect* which is not suitable for both visual perception and further processing. In the result of SCBG-based method which is shown in Fig. 2(c), it is easy to find the *block effect* especially in the regions of red labels. The *block effect* is mainly caused by *two issues*: (1) The size of sub-blocks is difficult to determine. If the size is too large, it can easily lead to situations where one block contains both clear areas and blurred areas; if the size is too small, it is hard to judge the features of the sub-block, which is likely to cause sub-block selection error [18]. (2) The focusing properties of the sub-block are difficult to determine. Especially when the detail information of the block is not rich, it will easily cause sub-block selection error.

So far, frequency domain methods have been explored by using multi-scale transform, including Laplacian pyramid transform, Gradient pyramid transform [19], [20] and wavelet transform [21]. Due to the upstanding localize peculiarity in both time and frequency domain, wavelet analysis has become one of the most commonly used methods among the frequency domain fusion [22]. With the deepening of theoretical studies, methods have been proposed for image fusion that are better than wavelet transform, such as wavelet packet transform and wavelet frame transform. However, wavelet analysis cannot effectively represent the line and plane singularities of the images and thus cannot represent the edge directions in images accurately [23], which results in the wavelet family *lacking shift-invariance* and having *pseudo-Gibbs phenomena* in the fused image. Figs. 2(d)-(g) show a sample of fused result of discrete wavelet transform (DWT) and the pseudo-Gibbs phenomena.

To overcome these shortcomings of the wavelet transform, Da Cunha *et al.* [24] proposed the Non-subsampled Contourlet Transform (NSCT) for image fusion, which can possess shift-invariance and effectively suppress Pseudo-Gibbs phenomena but with *time-consuming* [25] and thus is not suitable for real-time application. Candes and Donoho [26] proposed the early Curvelet transform with strong anisotropy

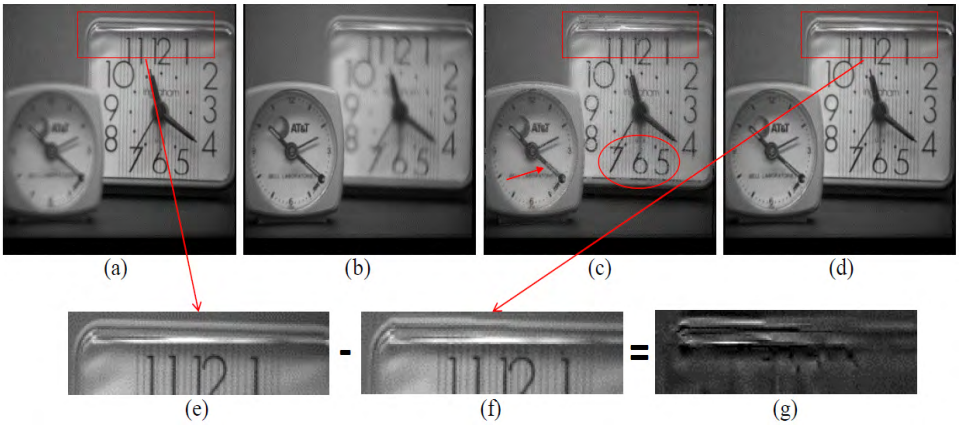


FIGURE 2. The fusion results of multi-focus images. (a-b) Multi-focus images: (a) focus on the right; (b) focus on the left. (c-d) Fusion results of (c) SCBG [3], and (d) discrete wavelet transform (DWT). (e-g) The analysis of pseudo-Gibbs phenomena: (e) the focus slicing of Fig. 2(a); (f) the same part of Fig. 2(d); (g) the image difference of Figs. 2(e) and (f).

and multi-direction. The Curvelet transform can not only work well on image edge, but also reduce nonlinear approximation error of the different curves with the smooth function [23]. Later, Candes and Donoho [27] and Candes *et al.* [28] designed an improved version of Curvelet transform for image fusion. The improved method is a fast discrete realization method for Curvelet transform, called fast discrete Curvelet transform (FDCT) [28]. The FDCT is simpler, faster and less redundant than existing proposals, which is reported on a series of practical successes, see [29]–[31]. However, the coefficients of FDCT are shrink which is more complex compared to the input images. This property of FDCT coefficients leads the frequency selection difficult which troubles the application of FDCT. Li *et al.* [32] used the average method for low frequency coefficients and larger absolute values for high frequency coefficients in FDCT domain. However, the common fusion rule will reduce the contrast in the fusion results and is not sensitive to the edges and directional features but signal noise [23].

Based on above analysis, in this paper we present a hybrid method for multi-focus image fusion. We first design a frequency-based model to extract detailed information in a set of input images. In this model, the pulse coupled neural network (PCNN) and sum-modified-laplacian (SML) are proposed to be the rules for fusing of low- and high frequency coefficients in FDCT domain, respectively. Through this model, the most important feature information is selected into the fused coefficients. Secondly, we present a hybrid spatial-based model using image difference to detect the focus and defocus regions. Then the mathematical morphology technique is applied to repair the mistakes of detection. Finally, we present a completed robustness evaluation process for image fusion technique considering the misregistration, noise error and conditional focus situations. The main contributions of this paper are as follows:

(1) To solve the difficulty of frequency selection in FDCT domain. We present a fast frequency-based model based on FDCT, which employs PCNN and SML to fuse of low- and high frequency coefficients, respectively. This proposal can extract the major detailed information from input images with lower computational complexity compared with other frequency domain methods.

(2) We present a robust spatial-based model based on the detailed information and image difference. Since we got a good primary fused image above, we didn't need to detect the focus or detail information from bottom-up. We employed the image difference of primary fused image and inputs to precisely detect the focus/detailed information of original image and then got a decision map which represents the focus attributes.

(3) We design a completed robustness evaluation process including the misregistration, noise error and conditional focus situations for image fusion techniques.

II. RELATED WORK

A. FAST DISCRETE CURVELET TRANSFORM (FDCT)

The early Curvelet transform uses the theory of Ridgelet to analyze the image features. However, this algorithm is complex with high redundancy. To overcome that drawback, Candes *et al.* [28] proposed an improved version called FDCT. FDCT adopts local Fourier transform for the frequency domain decomposition [32]. Therefore, FDCT is of high speed and small redundancy. Here, we only briefly list the steps of the implementation of FDCT, and more details can refer to [28]. The implementation of FDCT [33] via wrapping is shown as follows:

(1) Apply two-dimensional Fast Fourier transform (2D-FFT) for a given image to obtain Fourier coefficients as:

$$\hat{f}[x, y], \quad for \ -\frac{N}{2} \le x, \ y \le \frac{N}{2}, \tag{2}$$

where $\hat{f}[x, y]$ is the Fourier coefficients obtained by 2D-FFT, representing the length of the side of square image.

(2) Resample (or interpolate) $\hat{f}[x, y]$ to obtain sampled values as, $\hat{f}[x, y - x \tan \Theta_l]$ for each scale j and angle l,

(a) (b)

FIGURE 3. Fast Discrete Curvelet decomposition. (a) The original image. (b) FDCT decomposition coefficients.

where $(x, y) \in P_i; P = (x, y) : n_{1,0} \leq x \leq n_{1,0} + L_{1,j}; n_{2,0} \leq y \leq n_{2,0} + L_{2,j}$, $(n_{1,0}, n_{2,0})$is the index of the pixel at the bottom-left of the rectangle, $L_{1,j} \approx 2^j, L_{2,j} \approx 2^{j/2}$.

(3) Multiply the interpolated (or sheared) object $\hat{f}[x, y - x \tan \Theta_l]$ with the parabolic window $\hat{U}[x, y]$, and obtain $\hat{f}[x, y] = \hat{f}[x, y - x \tan \Theta_l]\, \hat{U}[x, y]$.

(4) Wrap $\hat{f}[x, y]$ around the origin and then apply the inverse 2D-FFT to the wrapping result, hence collecting the discrete coefficients $C_{j_0}(x, y)$, $C_{j,r}(x, y)(j \geq j_0)$.

Fig.3 shows the Lena image and its FDCT decomposition coefficients with five levels. As shown in Fig. 3, the center area stores the low frequency (coarse scale) coefficients; the area of Cartesian Loop stores the high frequency (fine scale) coefficients; the outer the loop, the higher the coefficients. Since FDCT has better sparse representation ability than wavelet transform [32], FDCT based image fusion is able to extract more features like edges and textures than wavelet transform.

B. PULSE COUPLED NEURAL NETWORK

PCNN is a biologically inspired neural network based on the work by Eckhorn *et al.* [34]. It has been proved that PCNN is suitable for image fusion including spatial domain techniques and frequency domain methods [35], [36]. Each PCNN neuron consists of three parts: receptive field, modulation field and pulse generator. However, this structure is nonlinear, and with too many uncertain parameters, leading to the difficulty of the mathematical analysis to the network. In the PCNN-based image fusion, we use a simplified model as shown in Fig. 4(a). The neuron can be described as follows:

$$\begin{cases} F_n(x, y) = S_n(x, y) \\ L_n(x, y) = \exp(-\alpha_L) L_{n-1}(x, y) \\ \qquad\qquad + V_L \sum_{ab} W_{a,b}(x, y) Y_{a,b}(x, y) \\ U_n(x, y) = F_n(x, y)[1 + \beta(x, y) L_n(x, y)] \\ Y_n(x, y) = \begin{cases} 1, & U_n(x, y) > \theta_n(x, y) \\ 0, & otherwise \end{cases} \\ \theta_n(x, y) = \exp(-\alpha_\theta)\theta_{n-1}(x, y) + V_\theta Y_n(x, y), \end{cases} \tag{3}$$

where n denotes the current iteration, $S_n(x, y)$ denotes the pixel value of input image located at (x, y). $F_n(x, y)$ is the primary input from the neurons receptive fields. The indexes a and b refer to the dislocation in a symmetric neighborhood around the one pixel, θ is a dynamic neuron threshold. α_F, α_L and α_θ are the decay constants of the PCNN neuron. V_F, V_L and V_θ are the magnitude scaling terms. The constant $\beta(x, y)$ is the linking strength. $L_n(x, y)$ is the secondary input of lateral connections with neighboring neurons. The $W_{a,b}$ are the constant synaptic weight matrices for $L_n(x, y)$. The neuron will generate pulse when $U_n(x, y) > \theta_n(x, y)$. The sum of $Y_n(x, y)$ in iteration is called firing times $T_n(x, y)$ [37]. To represent image information, it is defined as follow:

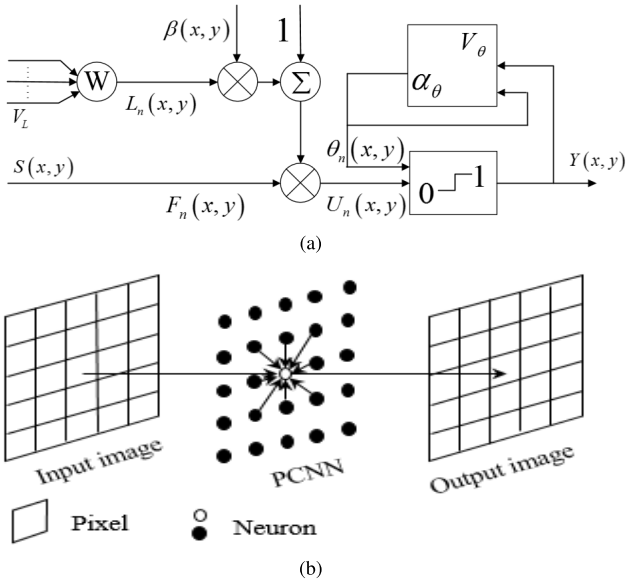


FIGURE 4. The schematic diagram of Pulse Coupled Neural Network (PCNN). (a) Simplified model of PCNN neuron. (b) Connection model of PCNN neuron.

$$T_n(x, y) = T_{n-1}(x, y) + Y_n(x, y), \tag{4}$$

The PCNN used for image fusion is a single-layer two-dimensional array of laterally linked pulse coupled neurons, as shown in Fig. 4(b) and all neurons are identical. The number of neurons in the network is equal to the number of pixels in the input image [38]. There is a one-to-one correspondence mapping between the image pixels and network neurons. Each neuron is connected with neighboring neurons in linking range [39]. The output of each neuron results in two states, namely firing (1 state) and non-firing (0 state). Pulse output will be delivered to adjacent neurons. If adjacent neurons have a similar intensity with the current neuron, they will fire together because of pulse coupled action [38]. In this case, we recall that neuron to capture the adjacent neurons. Finally, the neuron and the similar adjacent neurons will emit synchronous pulses. Thus, PCNN has the global coupling and pulse synchronization characteristics, which benefit image fusion because they make use of local image information.

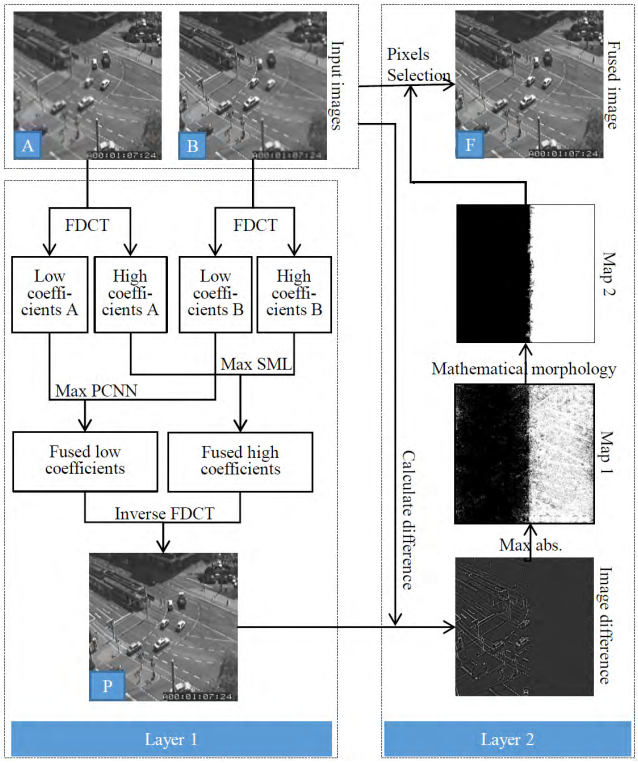


FIGURE 5. The schematic diagram of the proposed image fusion algorithm.

TABLE 1. The proposed frequency-based model.

Input:	Input images
Output:	The frequency-based fused image.
Step 1:	Load the input multi-focus images.
Step 2:	Perform 5 levels (how to set the level can be seen in [32]) FDCT on source images to obtain the frequency coefficients as follows: $\left\{C^A_{j_0}(x,y), C^A_{j,r}(x,y)\right\}, \left\{C^B_{j_0}(x,y), C^B_{j,r}(x,y)\right\} (j \geq j_0),$ (5) where $C^A_{j_0}(x,y), C^B_{j_0}(x,y)$ represent the low frequency coefficients, $C^A_{j,r}(x,y), C^B_{j,r}(x,y)$ represent the high frequency coefficients at j level and r orientation.
Step 3:	Fuse the low- and high-frequency coefficients by using different rules to obtain the fused low- and high-frequency subbands. The details of low coefficients and high coefficients fusion algorithm are discussed as follows.
Step 4:	Perform inverse FDCT with 5 levels on the fused low- and high-frequency coefficients.

III. PROPOSED MULTI-FOCUS IMAGE FUSION FRAMEWORK

The general framework of the proposed multi-focus image fusion scheme is depicted in Fig. 5. The method consists of two layers including frequency-based model and hybrid spatial-based model, which are described in detail as below.

A. FREQUENCY-BASED MODEL

In this subsection, a novel FDCT based fusion, which applies different fusion rules for low- and high-frequency coefficients respectively, is described in Table 1.

1) LOW FREQUENCY FUSION RULE

The coefficients in the low frequency represent the approximate information and contain the most energy of source images. The widely used rule is to apply the averaging methods to produce the fused coefficients. However, this rule will reduce the contrast in the fused images. It is known that the Human Visual System (HVS) is highly sensitive to the edges of image and is insensitive to real luminance at independent positions [40]. The study shows that PCNN possesses the global coupling and pulse synchronization characteristics [41]. These characteristics for HVS based image fusion in frequency domain include low frequency [41]. Therefore, we propose PCNN to fuse the low frequency coefficients and the fusion process is defined as follows:

$$C^P_{j_0}(x,y) = \begin{cases} C^A_{j_0}(x,y), & if\ T^A_{n,j_0}(x,y) > T^B_{n,j_0}(x,y) \\ C^B_{j_0}(x,y), & if\ T^A_{n,j_0}(x,y) < T^B_{n,j_0}(x,y) \\ 0.5(C^A_{j_0}(x,y) + C^B_{j_0}(x,y)), & Otherwise, \end{cases} \quad (6)$$

where $T^A_{n,j_0}(x,y)$ and $T^B_{n,j_0}(x,y)$ are the firing times of low frequency coefficients of the corresponding source images; $C^P_{j_0}(x,y)$ represents the low frequency coefficients of the primary fused image.

2) HIGH FREQUENCY FUSION RULE

For the high frequency coefficients, the most popular fusion rule is to select the coefficients with larger absolute values. However, this rule does not take consideration of the surrounding pixels. Furthermore, this is inconsistent with the characteristics of the HVS, which is sensitive to the edges and directional features. According to the study [40] the modified-laplacian (ML) can effectively represent the salient features and sharp boundaries of image [42]. The SML operator represents the sum of the surrounding ML, which is developed to provide local measures of the quality of image focus [43]. It is proved that SML is more efficient than ML in frequency domain, particularly in high frequency [44]. Therefore, we propose to use SML as the high frequency fusion rule. The definition of SML is as follows:

$$SML(x,y) = \sum_{h=-H}^{H} \sum_{w=-W}^{W} [ML(x+h, y+w)], \quad (7)$$

where $[2H+1), (2W+1)]$ is the size of the window; $ML(x,y)$, representing the value of modified-laplacian located at (x,y), is calculated as:

$$\begin{aligned} ML(x,y) = {} & |2C(x,y) - C(x-s,y) - C(x+s,y)| \\ & + |2C(x,y) - C(x,y-s) - C(x,y+s)|, \end{aligned} \quad (8)$$

where s is a variable spacing between coefficients and is always equal to 1 [45]. $C(x,y)$ denotes the pixel value of one high coefficient located at (x,y). Therefore, the fusion rule for the high-frequency coefficients based SML is defined as:

$$C^P_{jr}(x,y) = \begin{cases} C^A_{jr}(x,y), & if\ SML^A_{jr}(x,y) \geq SML^B_{jr}(x,y) \\ C^B_{jr}(x,y), & Otherwise, \end{cases} \quad (9)$$

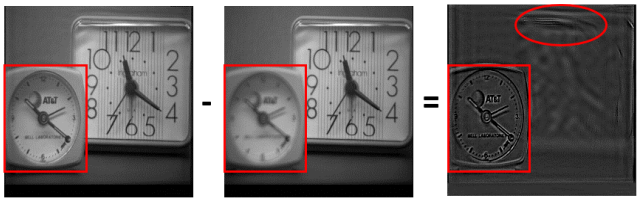


FIGURE 6. The image difference. (a) The FDCT-based fusion result, (b) One input image, (c) Image difference of Fig. 6(a) and (b).

where $SML_{jr}^{A}(x, y)$ and $SML_{jr}^{B}(x, y)$ are values of SML of high frequency coefficients at j level in the r orientation. $C_{jr}^{P}(x, y)$ represents the high frequency coefficients of the primary fused image at j level and r orientation.

B. HYBRID SPATIAL-BASED MODEL

After FDCT-based image fusion, we still find that there is some important details cannot be extracted by the frequency selection model. The reason is the scale shrinked in frequency domain and thus the frequency-feature extraction in FDCT domain cannot be fully successful in every coefficient. To address this problem, we present a spatial-based model using a novel process based on the first layer. Detailed description is described as below.

1) IMAGE DIFFERENCE

The image difference can be used for the fusion performance evaluation [6]. Fig. 6 shows the generation of image difference. From Fig. 6, we can easily find the defocus region (highlighted by the red rectangular frame) in image difference. The formula of image difference based onthe result of first layer is below:

$$\begin{aligned} DA(x, y) &= P(x, y) - A(x, y) \\ DB(x, y) &= P(x, y) - B(x, y), \end{aligned} \tag{10}$$

where $A(x, y)$, $B(x, y)$ and $P(x, y)$ denote the pixel located at (x, y) of source images and initial fused image, respectively. Then though comparing the absolute values of the difference image, a binary map $M_1(x, y)$ is achieved by:

$$M_1(x, y) = \begin{cases} 1, & if \ |DA(x, y)| \geq |DB(x, y)| \\ 0, & Otherwise. \end{cases} \tag{11}$$

2) MATHEMATICAL MORPHOLOGY TECHNIQUE

However, the binary map cannot fully represent the focus and defocus regions because of the defective frequency selection model. From Fig. 6(c), it can be seen that some information of focus region have been lost (highlighted by the red ellipse). According to the theory of imaging, the regions, either in focus or out of focus, are always continuous. Therefore, we design a final process based on mathematical morphology technique described as below,

$$\begin{aligned} M_a(x, y) &= bwareaopen(M_1(x, y), Th) \\ M_b(x, y) &= 1 - M_a(x, y) \\ M_2(x, y) &= bwareaopen(M_b(x, y), Th), \end{aligned} \tag{12}$$

where Th is a threshold that is set to remove the holes smaller than the threshold. Its value can be determined according to reference [35]. Mathematical morphology technique is employed to eliminate the defects of $M_1(x, y)$ which are caused by the noise or undesired effects and then get a modified decision map $M_2(x, y)$. The final process can thus be illustrated as below,

$$F(x, y) = \begin{cases} A(x, y), & if \ M_2(x, y) = 1 \\ B(x, y), & Otherwise, \end{cases} \tag{13}$$

where $F(x, y)$ denotes the pixel located at (x, y) of the fused image.

IV. THE EXPERIMENTAL RESULTS AND ANALYSIS

A. EVALUATION INDEX SYSTEM

In this study, we evaluate the performance of the proposed algorithm by both objective and subjective methods [46], [47]. The objective metrics used for are as follows.

1) ROOT MEAN SQUARE ERROR (RMSE)

RMSE represents the cumulative squared error between the fused and the referenced images [48]. which is defined as follows:

$$\mathbf{RMSE} = \sqrt{\frac{\sum_{x=0}^{M}\sum_{n=0}^{N}[F(x, y) - R(x, y)]^2}{MN}}, \tag{14}$$

where $M \times N$ is the size of fused image, $F(x, y)$ and $R(x, y)$ are the pixel values of the fused and referenced images at the position (x, y), respectively. The lower the value of RMSE is, the better performance the image fusion method is.

2) MUTUAL INFORMATION (MI)

MI can indicate how much information the fused image conveys from the source images [47], which is defined as follows [49]:

$$\mathrm{MI} = \mathrm{MI}^{AF} + \mathrm{MI}^{BF}, \tag{15}$$

in which

$$\begin{aligned} \mathrm{MI}^{AF} &= \sum_{f=0}^{L}\sum_{a=0}^{L} p^{AF}(a, f) \log\left(\frac{p^{AF}(a, f)}{p^{A}(a)p^{F}(f)}\right) \\ \mathrm{MI}^{BF} &= \sum_{f=0}^{L}\sum_{b=0}^{L} p^{BF}(b, f) \log\left(\frac{p^{BF}(b, f)}{p^{B}(b)p^{F}(f)}\right), \end{aligned} \tag{16}$$

where MI^{AF} and MI^{BF} denote the normalized MI between the fused image (F) and the source images (A and B); $a, b \ and \ f \in [0, L]$. $p^{A}(a)$, $p^{B}(b)$ and $p^{F}(f)$ are the normalized gray level histograms of source images and the fused image. $p^{AF}(a, f)$ and $p^{BF}(b, f)$ are the joint gray level histograms between the fused image F and the source images A and B. The larger the value of MI is, the better performance the image fusion method is.

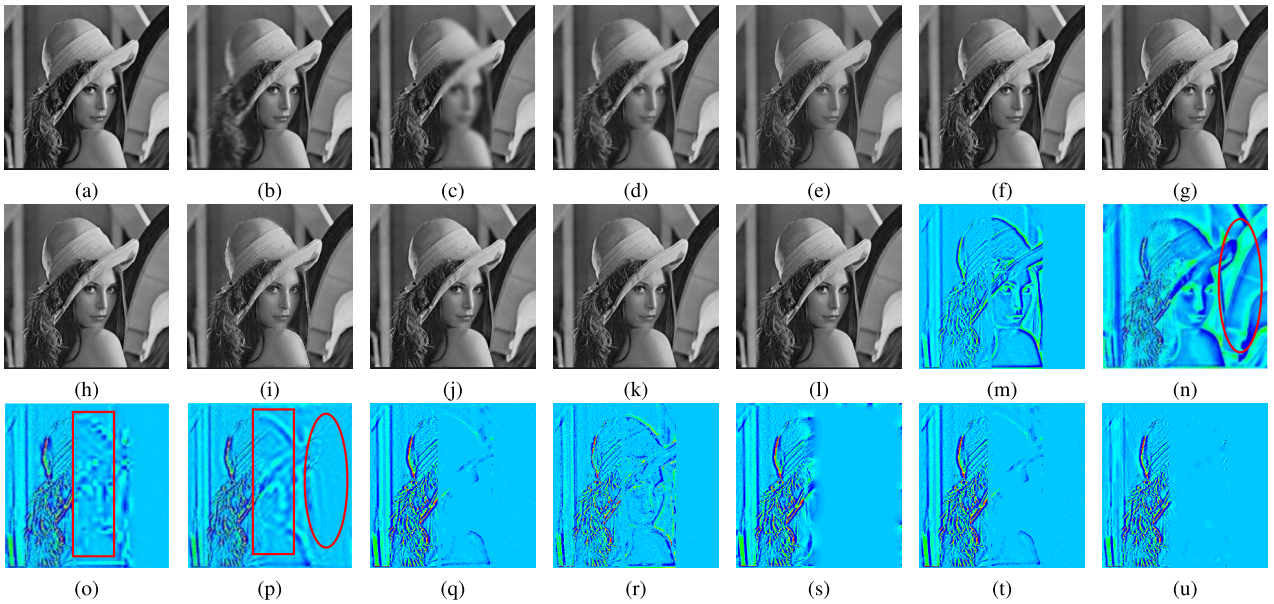


FIGURE 7. The data and fusion results of the artificial multi-focus images. (a-c) The experimental data: (a) the referenced image; (b) the image blurred on the left; (c) the image blurred in the middle. (d-l) The fusion results of different methods: Fused by (d) PCA method; (e) GP based method; (f) DWT based method; (g) FDCT based method; (h) NSCT-PCNN based method; (i) SCBG based method; (j) MWGF based method; (k) the primary method; (l) the proposed method. (m-u) Transformation of pseudocolor on the difference images between the nine fusion results Fig. 7(d)-(j) and Fig. 7(b).

3) EDGE BASED SIMILARITY MEASURE ($Q^{AB/F}$)

$Q^{AB/F}$ is proposed by Xydeas and Petrovic [50], which gives the similarity between the edges transferred from the input images to the fused image. Mathematically, the definition is given as:

$$Q^{AB/F} = \frac{\sum_{x=0}^{M}\sum_{y=0}^{N}[Q^{AF}(x,y)w^{A}(x,y) + Q^{BF}(x,y)w^{B}(x,y)]}{\sum_{x=0}^{M}\sum_{y=0}^{N}w^{A}(x,y) + w^{B}(x,y)]}, \tag{17}$$

in which

$$\begin{aligned} Q^{AF}(x,y) &= Q_a^{AF}(x,y)Q_g^{AF}(x,y) \\ Q^{BF}(x,y) &= Q_a^{BF}(x,y)Q_g^{BF}(x,y), \end{aligned} \tag{18}$$

where $w^A(x,y)$ and $w^B(x,y)$ are the corresponding gradient strengths for images A and B, respectively. $Q_a^{AF}(x,y)$, $Q_g^{AF}(x,y)$, $Q_a^{BF}(x,y)$ and $Q_g^{BF}(x,y)$ are the Sobel edge strength and orientation preservation values at location (x,y) for each source image. The larger the value the $Q^{AB/F}$ is, the better performance the image fusion method is.

B. EXPERIMENTS ON MULTI-FOCUS IMAGE FUSION

To evaluate the performance and robustness of the proposed image fusion approach, we use two sets of experiments: *general* and *conditional* testing. The test images are 2D gray images. The proposed method is compared with existing frequency-based and spatial-based methods. We use the following frequency-based methods of gradient pyramid (GP), DWT based methods, the FDCT [32] and NSCT-PCNN [25]; the spatial-based methods are also used in the experiments, including PCA and bilateral gradient-based sharpness criterion (SCBG) [3]. In addition, we compared our proposed method with a hybrid method based on Multi-scale weighted gradient-based fusion (MWGF) [51]. The NSCT-PCNN, SCBG and MWGF based fusion methods are open source codes offered by corresponding authors.

The high-frequency coefficients and low-frequency coefficients of GP and DWT are merged by the widely used fusion rule of selecting the coefficients with larger absolute values and the averaging rule (average-maximum rule), respectively. For DWT-based method, the images are decomposed using DBSS (2, 2) wavelet and the decomposition level is three. For implementing NSCT, the decomposition scale with 1, 2, or 8 directions and '9-7' filters and 'pkva' filters are used as pyramidal and directional filters respectively [52]. Parameters of PCNN is set as $\alpha_L = 0.06931, \alpha_\theta = 0.2, \beta = 0.2, V_L = 1.0, V_\theta = 20, W = \begin{bmatrix} 0.707 & 1 & 0.707 \\ 1 & 0 & 1 \\ 0.707 & 1 & 0.707 \end{bmatrix}$, and the maximal iterative number is $n = 200$ [25]. All the experiments are implemented in MATLAB R2014a on an Intel (R) Core (TM) i3-2330M CPU @2.2GHz computer (ASUS, Taipei, Taiwan) with 4 GB RAM.

1) GENERAL TESTING

The general testing includes two parts: a set of *artificial* and 5 sets of *natural* multi-focus test images.

a: FUSION OF ARTIFICIAL MULTI-FOCUS IMAGES

The first experiment is conducted by using a set of artificial multi-focus images. The sizes of these images are 256×256. Figs. 7(a)-(c) show the referenced Lena image and two blurred artificial images. The artificial images are generated

by convolving the referenced image with a blurring filter working on the left and the middle parts, respectively [49]. As shown in Figs. 7(a)-(c), the right part is the focus region of Fig. 7(b) and 7(c), simultaneously; the middle part is the focus region of Fig. 7(b) and the defocus region of Fig. 7(c); and the left part is the focus region of Fig. 7(c) and the defocus region of Fig. 7(b).

Subjective Evaluation: The fusion results of different methods are shown in Figs. 7(d)-(l). In comparison, we add the primary method with the fusion rule of selecting the low frequency by PCNN and high frequency by SML, respectively. For a clear comparison, we give the transformation of pseudocolor on the difference images between the fused results and Fig. 7(b). As seen in Fig. 7, the performance of different image fusion algorithms is different and the performance comparison is analyzed as follows.

(a) *For the right area,* the less information of the difference images is, the more information has been transferred to the fused images and hence the better performance the corresponding method is [35]. The results of the spatial domain methods including PCA and SCBG have less information than other transform-based methods. The reason is that the spatial domain method is a single-scale fusion method. For example, in PCA method, whichever the source images pixels are selected to the result, the result would be focused in the right area. However, for transform methods such as GP, and FDCT, we can find some information in the right parts (highlighted by red oval in Figs. 7(n) and 7(p), which is lost by the fusion rules. From these results, we can also clearly find that the results of the primary method can retain almost all the focus information as well. The reason is that the fusion rules of PCNN and SML can effectively select the focus information from different scales in FDCT domain and overcome the shortcoming of the fusion rules in [32].

(b) *For the middle area,* the less information of difference images is, the more information has been transferred to the fused images and hence the better performance the corresponding method is [35]. The results of the spatial domain methods show the outline of the original images. That is to say, these methods have lost the middle information of Fig. 7(b). This phenomenon appears in GP-based method as well. Meanwhile, the results of the DWT and FDCT methods lose a certain range of information. However, we can find little information from Figs. 7(s)-(u).

(c) *For the left area,* the more abundant the information of difference images is, the better performance the result is [35]. The results of the transform domain give the detailed features such as the hairs as shown in the Figs. 7(p)-(q),and 7(s)-(u). Conversely, the details of Figs. 7(m) and 7(r) are relatively less.

Therefore, through the comparative analysis above, we can find that the primary method is better than the DWT and FDCT based methods. It is proven that the coefficient selection methods proposed in this paper can overcome the shortcoming of current methods and extract most focus information from source images. The proposed algorithm and MWGF are superior to other 7 methods since almost all of the useful information of the source images has been transferred to the fused images.

Objective Evaluation: In order to better evaluate the performance of the proposed method, quantitative assessments are then carried out. Table 2 depicts the comparison results of different methods. It is easy to find that the RMSE value of the proposed method is the smallest (3.2131) among the compared methods. That is to say, compared with other methods, the fused image from the proposed method is the most close to the referenced image. Meanwhile, the RMSE value of the primary method is the third smallest (4.3967), which demonstrates that the coefficient selection method proposed in this paper can extract more focus information from source images to make the fused image more close to the referenced image and overcome the shortcoming of existing methods. Compared with FDCT, NSCT-PCNN, SCBG and MWGF methods, the RMSE values of the proposed method are decreased by 27.90%, 28.24%, 62.51% and 15.11%, respectively. In addition, the values of MI and $Q^{AB/F}$ of the proposed method also outperform other eight methods, as shown in Table 2.

TABLE 2. Comparison on quantitative evaluations of different methods for artificial multi-focus images.

Methods	RMSE	MI	$Q^{AB/F}$
PCA	8.7932	7.5317	0.7197
GP	10.8855	5.5334	0.7240
DWT	4.6122	7.1554	0.7639
FDCT[32]	4.4565	7.1449	0.7758
NSCT-PCNN[25]	4.4775	8.3407	0.7840
SCBG[3]	8.5711	9.5428	0.7044
MWGF[51]	3.7852	8.7219	0.7932
FDCT-PCNN (Ours)	4.3967	7.2013	0.7762
Proposed method	**3.2131**	**9.8013**	**0.8039**

From the above analysis, it is clearly to find that the proposed method not only preserves the most characteristics and information of the source images, but also improves the definition and the spatial quality better than the existing methods, which can be verified by the optimum values of objective criteria. Meanwhile, the performance of the primary method result is improved by the appropriate coefficient selection method compared to the method in [32].

b: FUSION OF NATURAL MULTI-FOCUS IMAGES

The second experiment is conducted on 5 sets of non-referenced multi-focus test images. These images are shown in Fig. 8. The sizes of Fig. 8 (a)-(b) are 256 × 256, Fig. 8 (c) are 640 × 480, Fig. 8 (d) are 480 × 360, Fig. 8 (e) are 512 × 364. The comparison on quantitative evaluation of different methods for natural multi-focus images is shown in Table 3. The values of MI and $Q^{AB/F}$ of the proposed method results are all better than other methods.

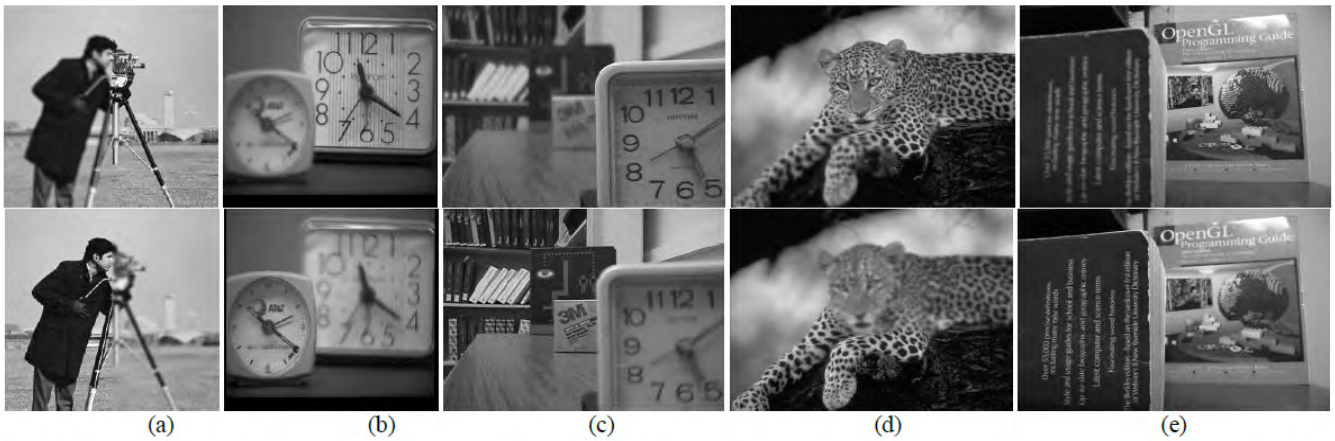


FIGURE 8. Natural multi-focus image database. (a) Cameraman images; (b) Clock images; (c) Desk images; (d) Leopard images;(e) Book images.

TABLE 3. Comparison on quantitative evaluations of different methods for natural multi-focus images.

Input	Indices	PCA	GP	DWT	FDCT [32]	NSCT-PCNN [25]	SCBG [3]	MWGF[51]	Proposed method
Cameraman	MI	7.2343	6.0235	6.8442	7.2117	7.8205	8.6715	8.2056	**8.9030**
	$Q^{AB/F}$	0.6784	0.7003	0.6965	0.7136	0.7137	0.6422	0.7370	**0.7404**
Clock	MI	7.2860	6.3777	6.5325	6.6894	7.8537	8.4652	7.9504	**8.6515**
	$Q^{AB/F}$	0.6639	0.6878	0.6786	0.6954	0.7181	0.5946	0.7365	**0.7395**
Desk	MI	5.9845	5.7187	5.5308	5.8684	6.3811	8.0455	7.7665	**8.3815**
	$Q^{AB/F}$	0.5217	0.6414	0.6326	0.6740	0.6717	0.5387	0.7286	**0.7297**
Leopard	MI	8.7721	6.3837	8.6843	8.9055	10.1345	10.7734	10.1408	**10.9038**
	$Q^{AB/F}$	0.7691	0.7769	0.8033	0.8125	0.8177	0.7434	0.8195	**0.8256**
Book	MI	7.7417	6.1279	7.0503	7.5758	8.4036	8.9884	8.9353	**9.3188**
	$Q^{AB/F}$	0.7005	0.7096	0.7205	0.7338	0.7385	0.6698	**0.7571**	0.7569
Mean	MI	7.4037	6.1263	6.9284	7.2502	8.1187	8.9888	8.5997	**9.2317**
	$Q^{AB/F}$	0.6667	0.7032	0.7063	0.7259	0.7319	0.6377	0.7557	**0.7584**

TABLE 4. Comparison on quantitative evaluations of different methods for misregistration multi-focus images.

Methods	MI	$Q^{AB/F}$
PCA	7.1223	0.5906
GP	6.6818	0.6861
DWT	6.6042	0.6810
FDCT [32]	6.8235	0.7053
NSCT-PCNN [25]	7.7695	0.7099
SCBG [3]	8.5788	0.5804
MWGF[51]	8.4459	0.7515
Proposed method	**8.8424**	**0.7526**

Compared with the FDCT based method, the values of MI and $Q^{AB/F}$ of the proposed method are increased by 40.64% and 5.85%, respectively. From this analysis, we can also conclude that the proposed method provides the best performance and outperforms the other 7 algorithms.

2) CONDITIONAL TESTING

To test the robustness of proposed technique, we present completed evaluations which consider three conditions: (1) *Misregistration*: Some works may satisfy well for input image with prefect registration. However, in case of image misregistration, isolated local patches with significant details, may also be selected during the fusion process. This leads to undesirable effects in the fused image, for example, block artifacts or spatial inconsistency. As well, misregistration is often prevalent in some local regions within the set of multi-focus images because of their different focal points or object movement in VSN. Thus, the first conditional testing is on a set of misregistration multi-focus images. (2) *Noisy environment*: In VSN, the low-cost imaging sensors always caused image with noise. Some techniques always select the noise as the details for mistake. Thus, the second conditional testing is on a set of multi-focus images with different noise levels. (3) *Conditional focus situations*: In VSN, the shapes and focus levels of multi-focus images are always affected by the environment changed. This condition may influence the fusion result, as well. Thus, the third conditional testing is on two sets of multi-focus images with varied focus shapes and levels.

a: FUSION OF MISREGISTRATION MULTI-FOCUS IMAGES

The experimental data and results are given in Fig. 9. The sizes of Fig. 9 are 640×480. As shown in Figs. 9(a)-(b), it is noticed that there is a slight shift of the student's head between the source images (marked by the red arrows). As shown in Figs. 9(c)-(j), one can obviously find that the fused image of the proposed method and MWGF are more

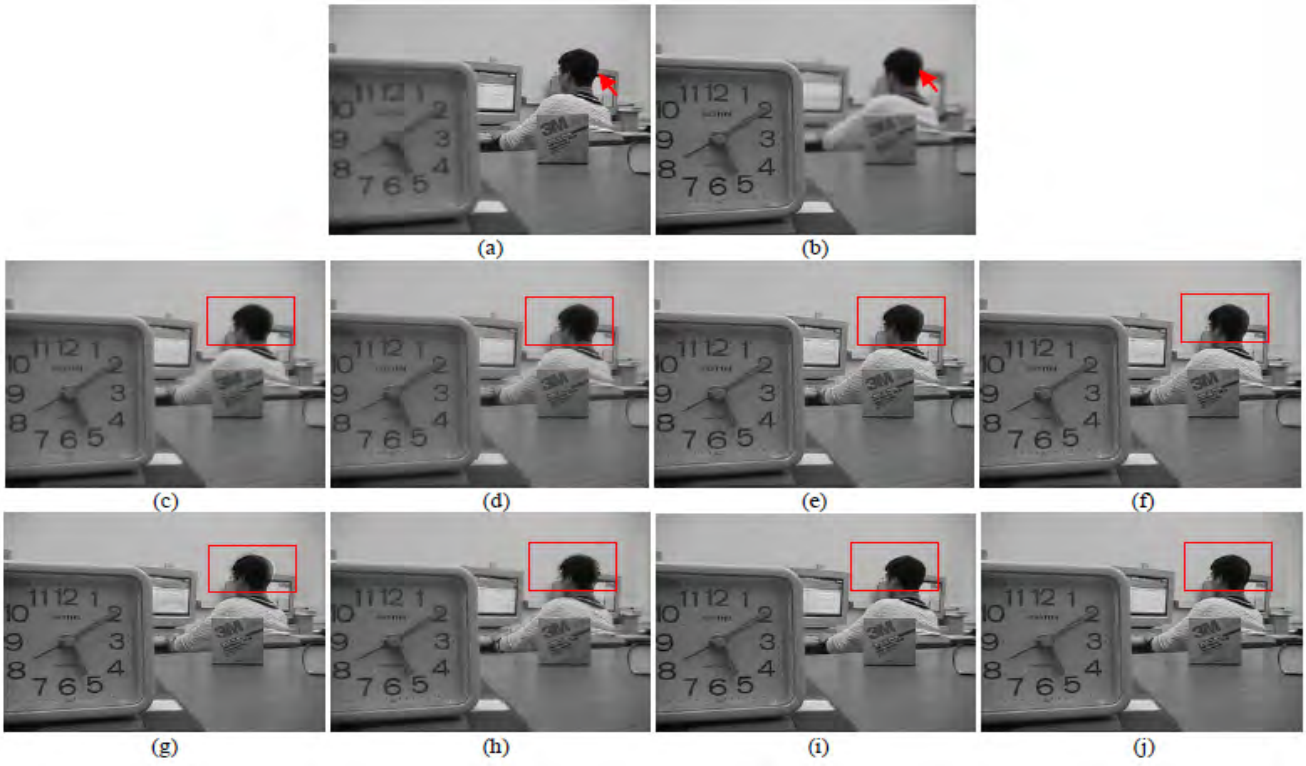


FIGURE 9. The data and results of misregistration multi-focus images. (a-b) The experimental data: (a) the Lab image focused on background; (b) the Lab image focused on foreground. (c-j) The fusion results of different methods: Fused by (c) PCA method; (d) GP based method; (e) DWT based method; (f) FDCT based method; (g) NSCT-PCNN based method; (h) SCBG based method; (i) MWGF based method; (j) the proposed method.

FIGURE 10. Samples of multi-focus images in different level of noisy environment. (a) 3%, (b) 7%, (c) 11%, (d) 15%.

natural than other methods in the region of the hair (highlighted by the red squares). Table 4 depicts the comparison on MI and $Q^{AB/F}$ of different methods. The MI and $Q^{AB/F}$ values of the proposed method are 8.8424 and 0.7526, respectively. Compared with FDCT, NSCT-PCNN, SCBG and MWGF methods, the MI and $Q^{AB/F}$ values of the proposed method increase 29.59% and 6.71%, 13.81% and 6.01% , 3.07% and 29.67%, 8.84% and 0.14%, respectively. Therefore, we can find that our proposed method is outperform other methods when original images have misregistration.

b: FUSION OF IMAGES IN DIFFERENT LEVEL OF NOISY ENVIRONMENT

To evaluate the noise-robustness of the proposed method, the four groups of experiments are conducted on the artificial multi-focus images as shown in Figs. 7(b)-(c), which are additionally corrupted with different Gaussian noise, with a standard deviation of 3%, 7%, 11%, and 15%, respectively as shown in Fig. 10. Fig. 11 shows the fusion results of Lena images with 7% noise for example.

A quantitative evaluation comparison of the different methods for Lena images with varied noise is shown in Fig. 12. From Figs. 12(a)-(b), we can find that the MI and $Q^{AB/F}$ values of the proposed method also outperform other methods. That is to say, in different noise environments, the results of the proposed method can get more useful information from original images. From Fig. 12(c), the RMSE values of the proposed method results are all the second lowest in the 8 methods. Although the RMSE of the proposed method is a little higher than MWGF method, the values of MI and $Q^{AB/F}$ are obvious bigger. Through comprehensive analysis, we can find that the proposed method can provide the best performance in different noisy environments.

c: FUSION OF MULTI-FOCUS IMAGES WITH CONDITIONAL FOCUS SITUATIONS

This experiment is conducted on multi-focus images with varied focus shapes and levels.

Experiments on Multi-Focus Images With Varied Focus Shapes: Fig. 13 shows the standard test image database and the corresponding artificial multi-focus images with varied focus shapes. We design three different focus shapes

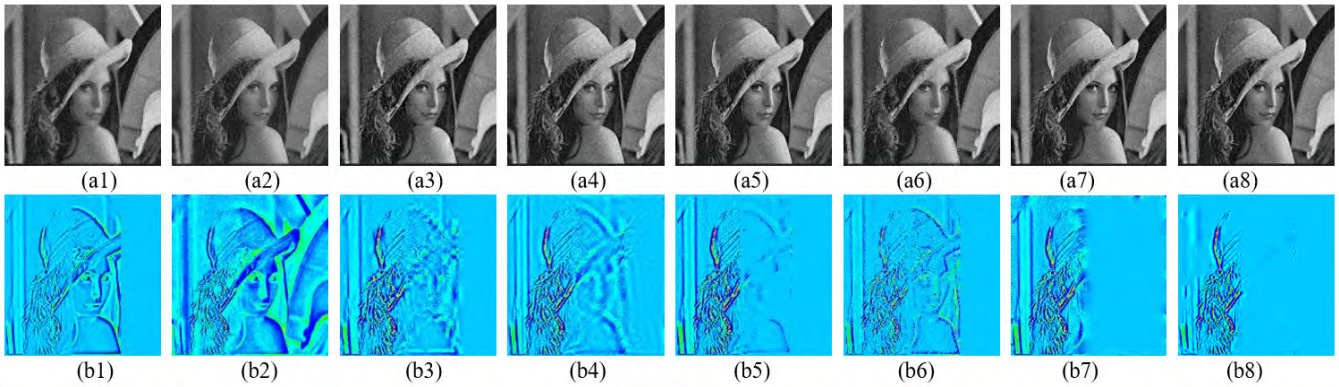
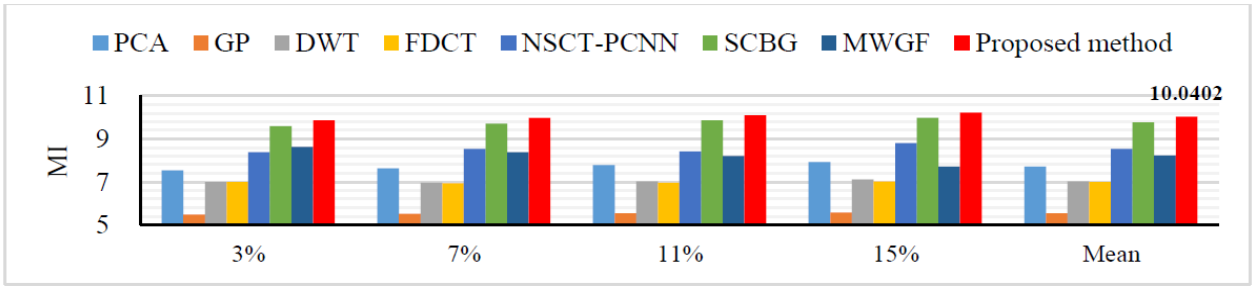
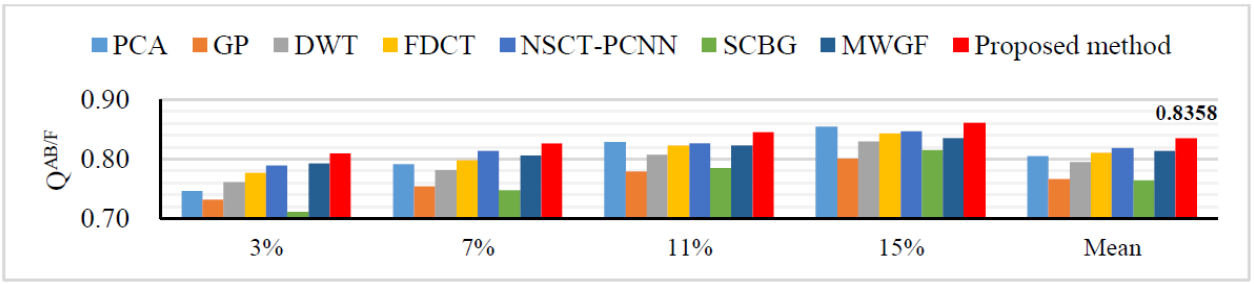


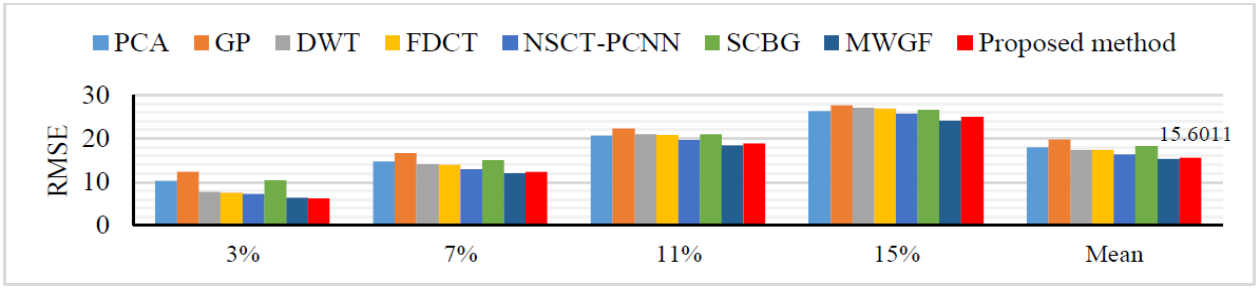
FIGURE 11. The example of the fusion of multi-focus images with 7% noise. (a1-a8) The fusion results of different methods: Fused by (a1) PCA method; (a2) GP based method; (a3) DWT based method; (a4) FDCT based method; (a5) NSCT-PCNN based method; (a6) SCBG based method; (a7) MWGF based method; (a8) the proposed method. (b1-b8) Transformation of pseudocolor on the difference images between the eight results and the top image of Fig. 10(b).



(a)



(b)



(c)

FIGURE 12. Comparison on quantitative evaluations of different methods for multi-focus images in different noisy environments.

(including left-right focused, middle-two-sided focused and center-around focused) into the standard images. Each class including four image pairs as shown in Fig. 13. These images are generated by convolving the test image with a 9×9 averaging filter. The sizes of the test images are 512×512.

Fig. 14 shows the results of different methods of an example of artificial multi-focus images with varied focus shapes, and the corresponding difference images with pseudocolor. From the difference images, it is clear to observe that the FDCT based method can extract more information from focus

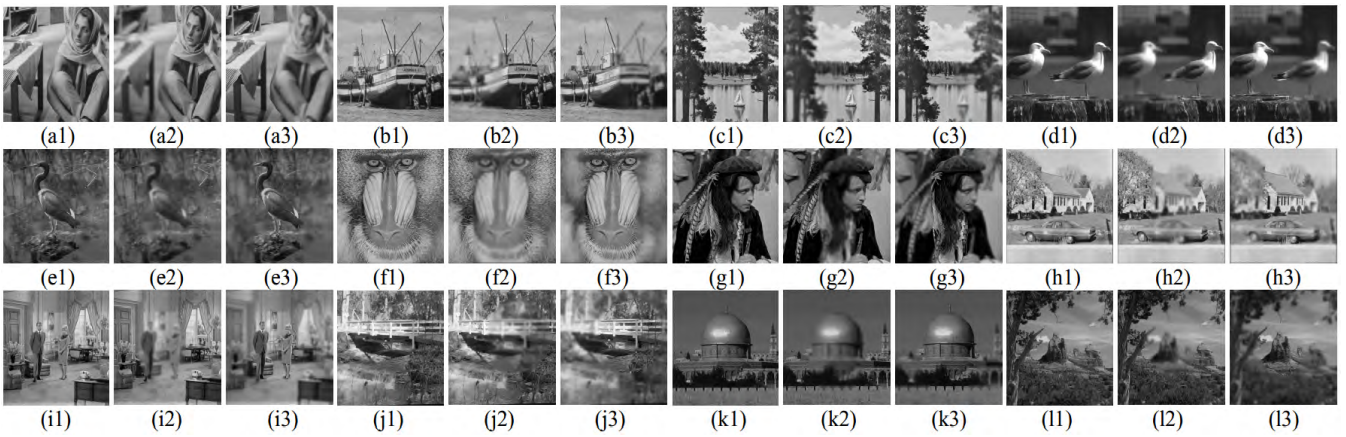


FIGURE 13. Dataset of multi-focus images with varied focus shapes. (a1)-(l1) Standard gray test images. (a2)-(d2) focus on right; (a3)-(d3) focus on left. (e2)-(h2) focus on the two sides; (e3)-(h3) focus in the middle. (i2)-(l2) focus on the around; (i3)-(l3) focus on the center.

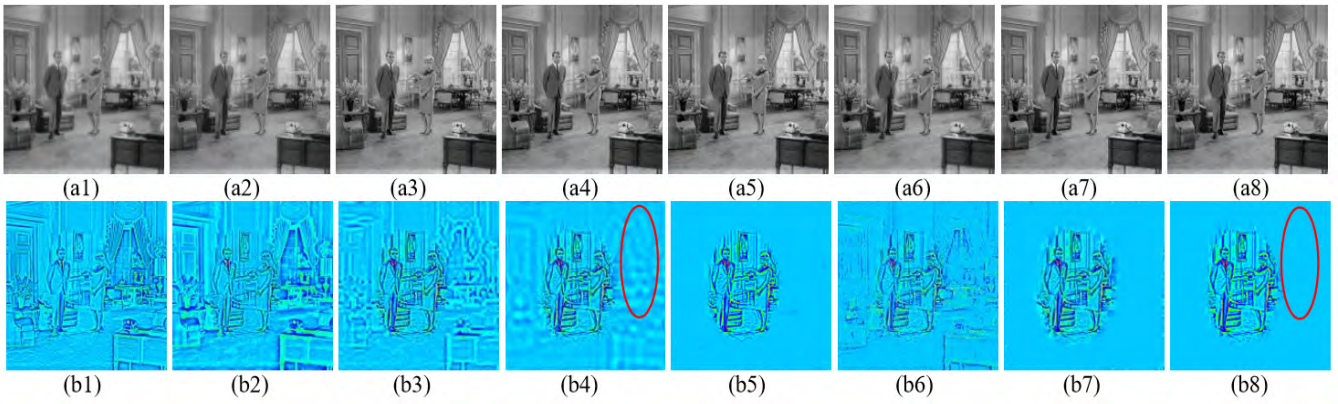


FIGURE 14. The example of the fusion of multi-focus images with varied focus shapes. (a1-a8) The fusion results of different methods: Fused by (a1) PCA method; (a2) GP based method; (a3) DWT based method; (a4) FDCT based method; (a5) NSCT-PCNN based method; (a6) SCBG based method; (a7) MWGF based method; (a8) the proposed method. (b1-b8) Transformation of pseudocolor on the difference images between the eight results and Fig. 13(i2).

TABLE 5. Execution time of different methods for Lena images.

Methods	Execution time (s)
FDCT [32]	**2.1697**
NSCT-PCNN [25]	72.4169
SCBG [3]	12.9979
MWGF [51]	38.8934
Proposed method	9.1889

region than GP and DWT based methods and the proposed method can obtain better performance than FDCT (highlighted by the red ellipse area) and SCBG methods.

Fig. 15 shows the comparison on MI, $Q^{AB/F}$ and RMSE of different methods for varied focus shape images. Averagely, comparing with FDCT, NSCT-PCNN, SCBG and MWGF based methods, the RMSE values of the proposed method are decreased by 32.56%, 40.81%, 78.97% and 62.16%, respectively. From Fig. 15, it is seen that the result of the proposed method is the most close to the standard images as its RMSE value is the smallest. In addition, it is easy to find that for 12 sets of images, the MI, $Q^{AB/F}$ and RMSE values of the proposed method results outperform other fusion results.

Experiments on Multi-Focus Images With Varied Focus Levels: Fig. 16 shows the traffic image and its corresponding multi-focus images with varied focus levels. Fig. 17 shows the results of different methods of the typical examples of the fusion of traffic images with 7×7 averaging filter, and the corresponding difference images with pseudocolor. From the difference images, it is clear to see that the proposed method can almost extract all of the focus area of source images and outperform other fusion methods.

Fig. 18 shows the comparison on MI, $Q^{AB/F}$ and RMSE of different methods for traffic images with varied focus level. As shown in Fig. 18, the MI, $Q^{AB/F}$ and RMSE values of the proposed method are all optimal except one $Q^{AB/F}$ value shown in Fig. 18(b). Although one $Q^{AB/F}$ value of the proposed method (with 3×3 averaging filter) is a little smaller than PCA based method, the corresponding values of RMSE and MI are obvious bigger than PCA based method. Hence, from the overall assessments, we can conclude that the proposed method is better than other fusion methods.

Finally, we compare the computational complexity of the proposed method with the state-of-the-art methods.

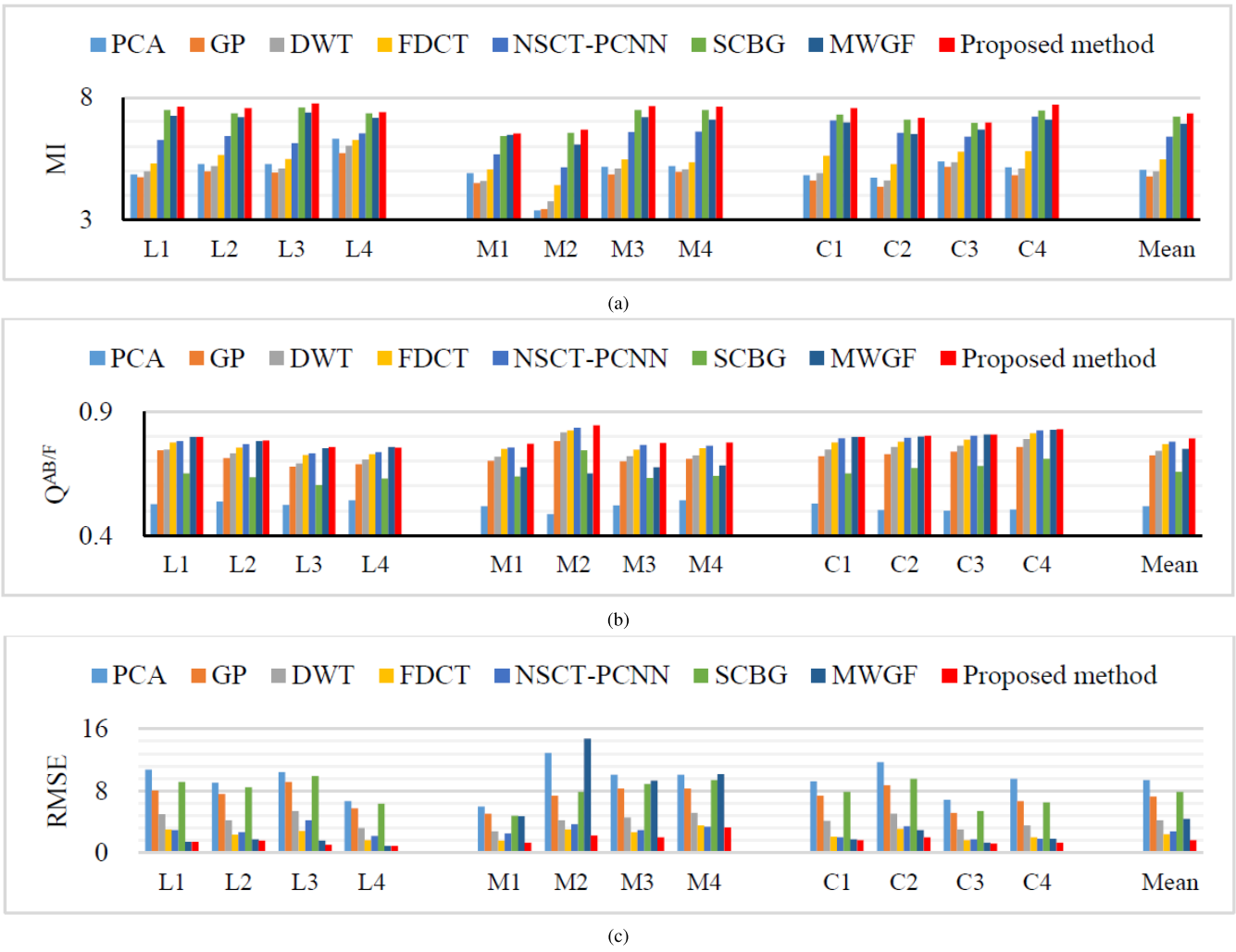


FIGURE 15. Comparison on quantitative evaluations of different methods for multi-focus images with varied focus shapes. L1-L4 represent the left-right focus images showed in Figs. 13(a)-(d), respectively; M1-M4 represent the middle-two sides focus images showed in Figs. 13(e)-(h), respectively; C1-C4 represent the center-around focus images showed in Figs. 13(i)-(l), respectively.

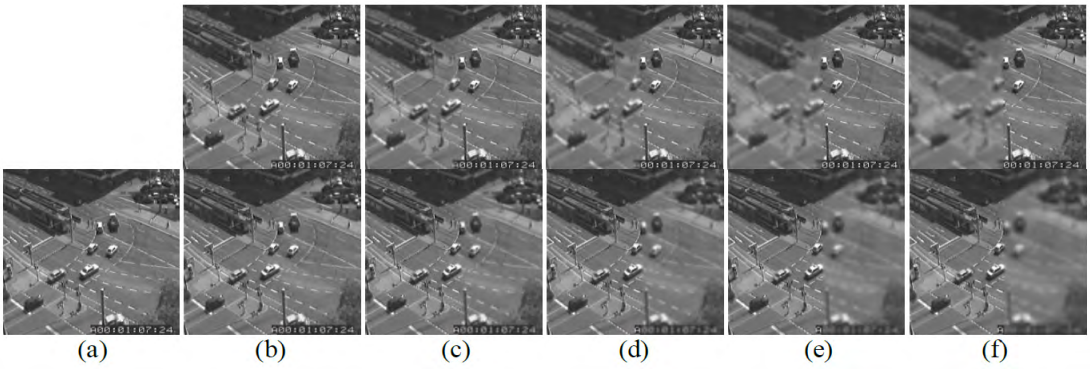


FIGURE 16. Multi-focus images with varied focus levels. (a) Standard traffic image; the multi-focus images are generated for traffic image by convolving the test image with (b) 3×3, (c) 5×5, (d) 7×7, (e) 9×9, (f) 11×11 averaging filters. (top: focus on the right; bottom: focus on the left)

A comparison of the execution time of FDCT, NSCT-PCNN and SCBG based methods and the proposed method for the Lena image sets is presented in Table 5. The execution time is calculated by running all the codes on the same computer. From Table 5, it can be seen that the NSCT-PCNN method is time-consuming, and the computational complexity of proposed and the FDCT-based method is lower than those of the NSCT-PCNN and SCBG methods. Although the

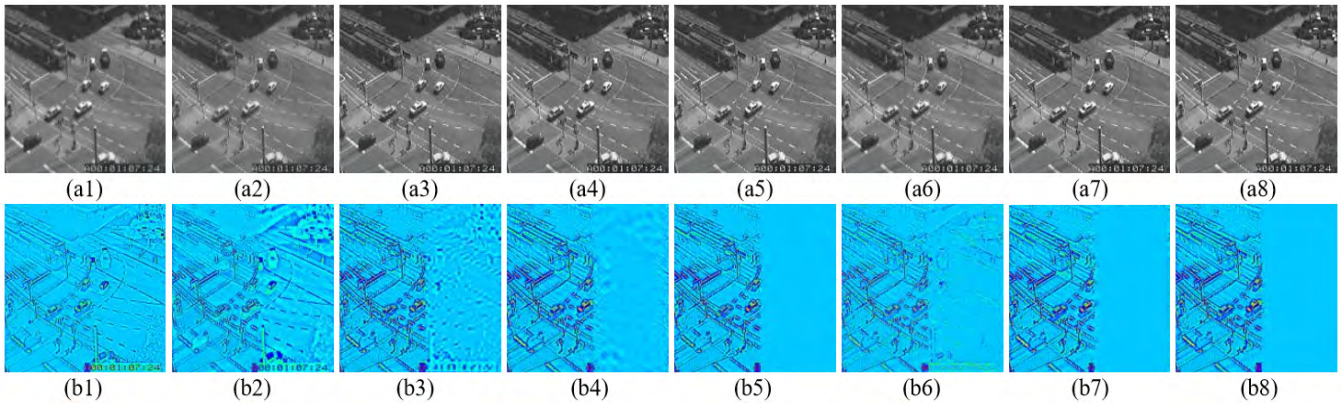


FIGURE 17. Fusion of multi-focus images with 7×7 averaging filter. (a1-a8) The fusion results of different methods: Fused by (a1) PCA method; (a2) GP based method; (a3) DWT based method; (a4) FDCT based method; (a5) NSCT-PCNN based method; (a6) SCBG based method; (a7) MWGF based method; (a8) the proposed method. (b1-b8) Transformation of pseudocolor on the difference images between the eight results and Fig. 16(d) (top image).

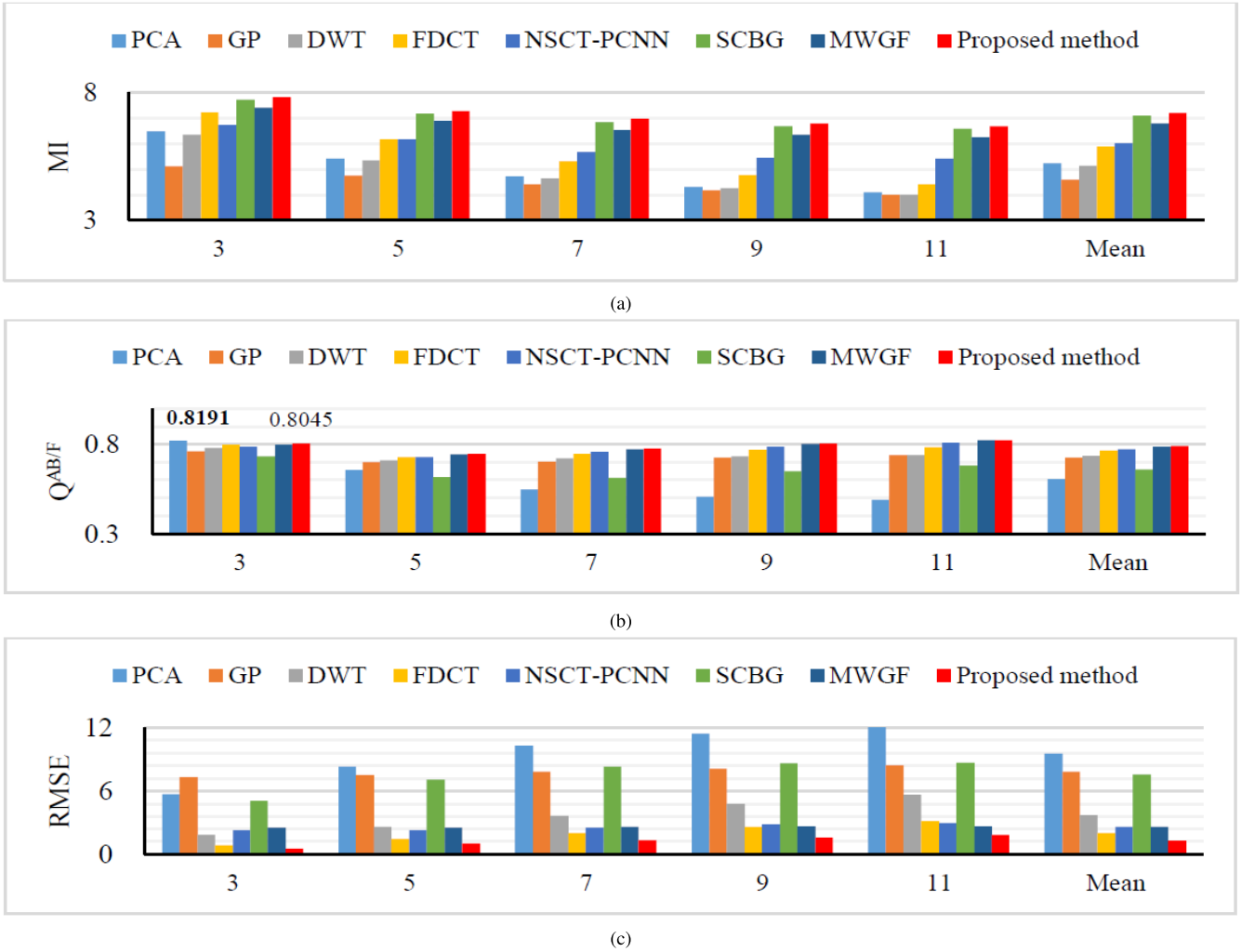


FIGURE 18. Comparison on quantitative evaluations of different methods for multi-focus images with varied focus level.

proposed method needs a little more execution time than the FDCT method, the fusion performance including the above qualitative and quantitative evaluations of our method have been proved to be better than the FDCT method.

V. CONCLUSION

In this paper, a FDCT based hybrid multi-focus image fusion framework is presented. The potential advantages of the proposed method include: (1) FDCT is more suitable for image fusion in real-time applications because of merits of multi-direction and fast properties; (2) The union fusion rules of the PCNN and SML can be used to strengthen the effect of FDCT; (3) Using the hybrid fusion procedure based on the image difference and taking advantage of result of first layer can not only reduce the computational complexity of the method, but also increase the reliability and robustness of the fusion results; (4) To evaluate the robustness of image fusion techniques, extensive sets of multi-focus images, including misregistration, noise error and conditional focus images are designed for comprising existing fusion methods. Various experiments can clearly prove the superior performance of the proposed fusion framework. In addition, the proposed method is more efficient than popular widely used NSCT-PCNN and SCBG based methods, which means our method is more suitable for VSN. The future work will focus on extending the proposed method for color multi-focus images and other image fusion tasks such as visible-infrared image fusion and multi-modal medical image fusion.

REFERENCES

[1] B. Yue, S. Wang, X. Liang, L. Jiao, and C. Xu, "Joint prior learning for visual sensor network noisy image super-resolution," *Sensors*, vol. 16, no. 3, p. 288, 2016.
[2] Y. Charfi, N. Wakamiya, and M. Murata, "Challenging issues in visual sensor networks," *IEEE Wireless Commun.*, vol. 16, no. 2, pp. 44–49, Apr. 2009.
[3] J. Tian, L. Chen, L. Ma, and W. Yu, "Multi-focus image fusion using a bilateral gradient-based sharpness criterion," *Opt. Commun.*, vol. 284, no. 1, pp. 80–87, 2011.
[4] Y. A. V. Phamila and R. Amutha, "Discrete cosine transform based fusion of multi-focus images for visual sensor networks," *Signal Process.*, vol. 95, no. 2, pp. 161–170, 2014.
[5] Z. Chen and S. Muramatsu, "Multi-focus image fusion based on multiple directional lots," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. E98.A, no. 11, pp. 2360–2365, 2015.
[6] Q. Zhang and M. D. Levine, "Robust multi-focus image fusion using multi-task sparse representation and spatial context," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2045–2058, May 2016.
[7] V. Aslantas, "A depth estimation algorithm with a single image," *Opt. Exp.*, vol. 15, no. 8, pp. 5024–5029, 2007.
[8] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.
[9] Y. Yang, Y. Que, S. Huang, and P. Lin, "Multiple visual features measurement with gradient domain guided filtering for multisensor image fusion," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 4, pp. 691–703, Apr. 2017.
[10] E. Andreas, C. L. Pagliari, and E. A. B. da Silva, "Multiscale image fusion using the undecimated wavelet transform with spectral factorization and nonorthogonal filter banks," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1005–1017, Mar. 2013.
[11] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 884–892, Apr. 2010.
[12] M. Nejati, S. Samavi, and S. Shirani, "Multi-focus image fusion using dictionary-based sparse representation," *Inf. Fusion*, vol. 25, pp. 72–84, Sep. 2015.
[13] J. Xiao, T. Liu, Y. Zhang, B. Zou, J. Lei, and Q. Li, "Multi-focus image fusion based on depth extraction with inhomogeneous diffusion equation," *Signal Process.*, vol. 125, pp. 171–186, Aug. 2016.
[14] Q. Wang, Y. Shen, Y. Zhang, and J. Q. Zhang, "A quantitative method for evaluating the performances of hyperspectral image fusion," *IEEE Trans. Instrum. Meas.*, vol. 52, no. 4, pp. 1041–1047, Aug. 2003.
[15] V. Aslantas and R. Kurban, "Fusion of multi-focus images using differential evolution algorithm," *Expert Syst. Appl.*, vol. 37, pp. 8861–8870, Dec. 2010.
[16] Y. Yang, D. S. Park, S. Huang, and N. Rao, "Medical image fusion via an effective wavelet-based approach," *EURASIP J. Adv. Signal Process.*, vol. 2010, pp. 579341:1–579341:13, 2010.
[17] R. Singh and A. Khare, "Multiscale medical image fusion in wavelet domain," *Sci. World J.*, vol. 2013, pp. 521034:1–521034:10, 2013.
[18] X. Bai, M. Liu, Z. Chen, P. Wang, and Y. Zhang, "Multi-focus image fusion through gradient-based decision map construction and mathematical morphology," *IEEE Access*, vol. 4, pp. 4749–4760, 2016.
[19] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, and I. Pippi, "Multispectral fusion of multisensor image data by the generalized Laplacian pyramid," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. Symp.*, vol. 2. Jun./Jul. 1999, pp. 1183–1185.
[20] V. S. Petrović and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 228–237, Feb. 2004.
[21] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57, no. 3, pp. 235–245, 1995.
[22] L. Han, L. Shi, Y. Yang, and D. Song, "Thermal physical property-based fusion of geostationary meteorological satellite visible and infrared channel images," *Sensors*, vol. 14, no. 6, pp. 10187–10202, 2014.
[23] Y. Yang, S. Tong, S. Huang, and P. Lin, "Dual-tree complex wavelet transform and image block residual-based multi-focus image fusion in visual sensor networks," *Sensors*, vol. 14, no. 12, pp. 22408–22430, 2014.
[24] A. L. da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.
[25] X.-B. Qu, J.-W. Yan, H.-Z. Xiao, and Z.-Q. Zhu, "Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain," *Acta Autom. Sin.*, vol. 34, no. 12, pp. 1508–1514, 2008.
[26] E. J. Candès and D. Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges," in *Curves and Surfaces*, C. Rabut, A. Cohen, and L. L. Schumaker, Eds. Nashville, TN, USA: Vanderbilt Univ. Press, 2000.
[27] E. J. Candès and D. Donoho, "Recovering edges in ill-posed inverse problems: Optimality of curvelet frames," *Ann. Statist.*, vol. 30, no. 3, pp. 784–842, 2002.
[28] E. J. Candès, L. Demanet, D. Donoho, and X. Ying, "Fast discrete curvelet transforms," *Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861–899, Sep. 2006.
[29] C. Amiot, C. Girard, J. Chanussot, J. Pescatore, and M. Desvignes, "Curvelet based contrast enhancement in fluoroscopic sequences," *IEEE Trans. Med. Imag.*, vol. 34, no. 1, pp. 137–147, Jan. 2015.
[30] Y. Alaudah and G. AlRegib, "A curvelet-based distance measure for seismic images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 4200–4204.
[31] S. Elaiwat, M. Bennamoun, F. Boussaid, and A. El-Sallam, "3-D face recognition using curvelet local features," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 172–175, Feb. 2014.
[32] H. Li, L. Guo, and H. Liu, "Research on image fusion based on the second generation curvelet transform," *Acta Opt. Sin.*, vol. 26, no. 5, pp. 657–662, 2006.
[33] S. Li and B. Yang, "Multifocus image fusion by combining curvelet and wavelet transform," *Pattern Recognit. Lett.*, vol. 29, no. 9, pp. 1295–1301, 2008.
[34] R. Eckhorn, H. J. Reitboeck, M. Arndt, and P. Dicke, "Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex," *Neural Comput.*, vol. 2, no. 3, pp. 293–307, 1990.
[35] H. Li, Y. Chai, and Z. Li, "Multi-focus image fusion based on nonsubsampled contourlet transform and focused regions detection," *Optik-Int. J. Light Electron Opt.*, vol. 124, no. 1, pp. 40–51, 2013.
[36] Z. Baohua, L. Xiaoqi, and J. Weitao, "A multi-focus image fusion algorithm based on an improved dual-channel PCNN in NSCT domain," *Optik-Int. J. Light Electron Opt.*, vol. 124, no. 20, pp. 4104–4109, 2013.
[37] Z. Wang, Y. Ma, F. Cheng, and L. Yang, "Review of pulse-coupled neural networks," *Image Vis. Comput.*, vol. 28, no. 1, pp. 5–13, 2010.

[38] T. Lindblad and J. M. Kinser, *Image Processing using Pulse-Coupled Neural Networks: Applications in Python*. London, U.K.: Springer-Verlag, 1998.
[39] Z. Wang, Y. Ma, and J. Gu, "Multi-focus image fusion using PCNN," *Pattern Recognit.*, vol. 43, no. 6, pp. 2003–2016, 2010.
[40] Y. Chai, H. F. Li, and J. F. Qu, "Image fusion scheme using a novel dual-channel PCNN in lifting stationary wavelet domain," *Opt. Commun.*, vol. 283, no. 19, pp. 3591–3602, 2010.
[41] Y. Chai, H. F. Li, and M. Y. Guo, "Multifocus image fusion scheme based on features of multiscale products and PCNN in lifting stationary wavelet domain," *Opt. Commun.*, vol. 284, no. 5, pp. 1146–1158, 2011.
[42] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 493–500, 2007.
[43] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, Aug. 1994.
[44] X.-B. Qu, J.-W. Yan, and G.-D. Yang, "Multifocus image fusion method of sharp frequency localized contourlet transform domain based on sum-modified-laplacian," *Opt. Precision Eng.*, vol. 17, no. 5, pp. 1203–1212, 2009.
[45] Y. Chai, H. Li, and X. Zhang, "Multifocus image fusion based on features contrast of multiscale products in nonsubsampled contourlet transform domain," *Optik-Int. J. Light Electron Opt.*, vol. 123, no. 7, pp. 569–581, 2012.
[46] X. Y. Luo, J. Zhang, and Q. H. Dai, "Saliency-based geometry measurement for image fusion performance," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 4, pp. 1130–1132, Apr. 2012.
[47] G. Bhatnagar, Q. M. J. Wu, and Z. Liu, "Directive contrast based multimodal medical image fusion in NSCT domain," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1014–1024, Aug. 2013.
[48] Y. Yang, W. Wan, S. Huang, F. Yuan, S. Yang, and Y. Que, "Remote sensing image fusion based on adaptive ihs and multiscale guided filter," *IEEE Access*, vol. 4, pp. 4573–4582, 2016.
[49] B. Yang and S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Inf. Fusion*, vol. 13, no. 1, pp. 10–19, 2012.
[50] C. S. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, 2000.
[51] Z. Zhou, S. Li, and B. Wang, "Multi-scale weighted gradient-based fusion for multi-focus images," *Inf. Fusion*, vol. 20, pp. 60–72, Nov. 2014.
[52] S. Li, B. Yang, and J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," *Inf. Fusion*, vol. 12, pp. 74–84, Apr. 2011.

YONG YANG (M'13–SM'16) received the Ph.D. degree in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 2005. From 2009 to 2010, he was a Post-Doctoral Research Fellow with Chonbuk National University, Jeonju, South Korea. He is currently a Full Professor and the Vice Dean of the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. His current research interests include image fusion and segmentation, medical image processing and analysis, and pattern recognition. He received the title of Jiangxi Province Young Scientist in 2012.

SONG TONG (S'16) received the M.E. degrees in logistics engineering from the JiangXi University of Finance and Economics, Nanchang, China, in 2014. He is currently pursuing the Ph.D. degree from the Department of Intelligence Science and Technology, Kyoto University, Japan. His current research interests include human perception, cognition, computer vision, and cognitive neuroscience.

SHUYING HUANG (M'14) received the Ph.D. degree in computer application technology from the Ocean University of China, Qingdao, China, in 2013. She is currently an Associate Professor with the School of Software and Communication Engineering, Jiangxi University of Finance and Economics, Nanchang, China. Her current research interests include image and signal processing, and pattern recognition.

PAN LIN (M'16) received the Ph.D. degree in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 2005. From 2007 to 2011, he was a Post-Doctoral Researcher with the Center for Mind/Brain Sciences, University of Trento, Italy. From 2013 to 2014, he was a Research Fellow with Mclean Hospital, Harvard Medical School, Boston, MA, USA. Since 2011, he has been an Associate Professor with the Institute of Biomedical Engineering, Xi'an Jiaotong University. His current research interests include multimodal MRI imaging, medical image processing, image processing, and pattern recognition.

YUMING FANG (M'13) received the B.E. degree from Sichuan University, Chengdu, China, the M.S. degree from the Beijing University of Technology, Beijing, China, and the Ph.D. degree from Nanyang Technological University, Singapore. He is currently a Professor with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China. He has authored or co-authored over 90 academic papers in international journals and conferences in the areas of multimedia processing. His research interests include visual attention modeling, visual quality assessment, image retargeting, computer vision, and 3-D image/video processing. He serves as an Associate Editor of the IEEE Access and is on the Editorial Board of *Signal Processing: Image Communication*.