

Article

Dual-Tree Complex Wavelet Transform and Image Block Residual-Based Multi-Focus Image Fusion in Visual Sensor Networks

Yong Yang ^{1,*}, Song Tong ¹, Shuying Huang ² and Pan Lin ³

¹ School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China; E-Mail: tong-song91@126.com

² School of Software and Communication Engineering, Jiangxi University of Finance and Economics, Nanchang 330032, China; E-Mail: shuyinghuang2010@126.com

³ Key Laboratory of Biomedical Information Engineering of Education Ministry, Institute of Biomedical Engineering, Xi'an Jiaotong University, Xi'an 710049, China; E-Mail: linpan@mail.xjtu.edu.cn

* Author to whom correspondence should be addressed; E-Mail: greatyangy@126.com; Tel.: +86-791-8398-3830; Fax: +86-791-8398-3891.

External Editor: Neal N. Xiong

Received: 30 September 2014; in revised form: 11 November 2014 / Accepted: 14 November 2014 / Published: 26 November 2014

Abstract: This paper presents a novel framework for the fusion of multi-focus images explicitly designed for visual sensor network (VSN) environments. Multi-scale based fusion methods can often obtain fused images with good visual effect. However, because of the defects of the fusion rules, it is almost impossible to completely avoid the loss of useful information in the thus obtained fused images. The proposed fusion scheme can be divided into two processes: initial fusion and final fusion. The initial fusion is based on a dual-tree complex wavelet transform (DTCWT). The Sum-Modified-Laplacian (SML)-based visual contrast and SML are employed to fuse the low- and high-frequency coefficients, respectively, and an initial composited image is obtained. In the final fusion process, the image block residuals technique and consistency verification are used to detect the focusing areas and then a decision map is obtained. The map is used to guide how to achieve the final fused image. The performance of the proposed method was extensively tested on a number of multi-focus images, including no-referenced images, referenced images, and images with different noise levels. The experimental results clearly indicate that the proposed

method outperformed various state-of-the-art fusion methods, in terms of both subjective and objective evaluations, and is more suitable for VSNs.

Keywords: multi-focus image fusion; dual-tree complex wavelet transform; image block residual; visual sensor networks

1. Introduction

Visual sensor networks (VSNs), wherein camera-equipped sensor nodes can capture, process, and transmit visual information, are an important research area which has attracted considerable attention in recent years [1,2]. Camera sensors collect a huge amount of visual data, which are rich in information and offer tremendous potential when used in VSNs [3,4]. Due to the resource and bandwidth requirements in VSNs, how to process such data using generally low-power sensor nodes is becoming a problem [5,6]. Image fusion can fuse these pictures into a single image. The fused image has a complete description of the scene, and is more accurate and suitable for both visual perception and further processing. This processing can reduce the randomness and redundancy and save storage space and bandwidth on the network, and hence improve the transmission efficiency in VSNs [7].

Because of the limited depth of focus in optical lenses [8], it is difficult to acquire an image that contains all relevant focused objects and hence some objects are in focus, but others will be out of focus and, thus, blurred. In VSNs, we have the opportunity to extend the depth of focus by using image fusion techniques.

Recently, image fusion techniques have been performed either in the spatial domain or frequency domain [9]. Spatial domain techniques are carried out directly on the source images. The main spatial domain techniques are weighted average, principal component analysis (PCA), intensity-hue-saturation (IHS), and Brovey transform [10]. The fused images obtained by these methods have high spatial quality, but they usually overlook the high quality of spectral information and hence suffer from spectral degradation [11]. In 2002, Li *et al.* [12] introduced the artificial neural network (ANN) for use in multi-focus image fusion. However, the performance of the ANN depends on the sample images and this is not an appealing characteristic. In 2011, Tian *et al.* [13] proposed a bilateral gradient to perform image fusion. However, there were some erroneous selections of some blocks in the focus region due to noise or undesired effects. Since the actual objects usually contain structures at many different scales or resolutions and multi-scale techniques are similar to the human visual system (HVS), the multi-scale techniques have attracted increasing interest in image fusion [14].

So far, frequency domain methods have been explored by using multi-scale transforms, including pyramid transforms [15,16], wavelet transforms [17,18], fast discrete curvelet transform (FDCT) [19], complex wavelet transform (CWT) [20] and non-subsampled contourlet transform (NSCT) [21]. Due to the outstanding localization capability in both the time and frequency domains, wavelet analysis has become one of the most commonly used methods in frequency domain fusion [18]. However, wavelet analysis cannot represent the directions of the image edges accurately, which results in a lack of shift-invariance and the appearance of pseudo-Gibbs phenomena in the fused image [22]. To overcome these shortcomings of the wavelet transform, Da Cunha *et al.* [23] proposed the NSCT. This method can not only give the asymptotic optimal representation of contours, but also possesses shift-invariance, and

effectively suppresses pseudo-Gibbs phenomena [24]. In 2008, Qu *et al.* [25] proposed an image fusion method based on spatial frequency-motivated pulse coupled neural networks (SF_PCNN) in the NSCT domain, and proved that this method could extract more useful information and provide much better performance than typical fusion methods. However, the NSCT-based algorithm is time-consuming and of high complexity. The dual-tree complex wavelet transform (DTCWT) is also proposed for image fusion. This method deals with the shift variance phenomenon using parallel wavelet filtering with directionality support (six planes) [26]. Furthermore, this filtering provides a 1/2 sample delay in the wavelet branches of the dual trees allowing near-shift invariance and perfect reconstruction [27].

Multi-scale techniques can significantly enhance the visual effect, but in the focus area of the source image, the fused image clarity will have different degrees of information loss. This occurs because in the process of multi-scale decomposition and reconstruction, improper selection of fusion rules often causes the loss of useful information in the source image; this defect is almost impossible to completely avoid in the multi-scale based image fusion method [9].

In addition, it is noteworthy that there is a hypothesis whereby the source images are noise-free in most of the image fusion methods. These methods can always produce high-quality fused images with this assumption [28]. However, the images are often corrupted by noise that cannot be completely avoided during the image acquisition from VSNs [28].

Based on the above analysis, an image fusion method based on DTCWT with image block residual is proposed in this paper. The main differences and advantages of the proposed framework compared to other methods can be summarized as follows:

- (1) For VSNs, a novel multi-focus image fusion framework which combines the DTCWT and image block residual techniques is proposed. The framework is divided into visual contrast based DTCWT based initial fusion and block residual based final fusion processes.
- (2) In the visual contrast based DTCWT-based initial fusion process, the Sum-Modified-Laplacian (SML)-based visual contrast [29] and SML [30] are employed as the rules for low- and high-frequency coefficients in DTCWT domain, respectively. Using this model, the most important feature information is selected in the fused coefficients.
- (3) In the block residual-based final fusion process, the image block residuals technique and consistency verification are proposed to detect the focus area and then a decision map is obtained. The decision map is used to guide which block should be selected from the source images or the initial fused image. Using this model, we can select pixels from the focus areas of source images to avoid loss of useful information to the greatest extent.
- (4) Further, the proposed framework can provide a better performance than the various state-of-the-art fusion methods and is robust to noise. In addition, the proposed framework is efficient and more suitable for VSNs.

The rest of the paper is organized as follows: the related theories of the proposed framework are introduced in Section 2. The proposed fusion framework is described in Section 3. Experimental results and analysis are given in Section 4, and the concluding remarks are described in Section 5.

2. Preliminaries

This section covers two related concepts, including DTCWT and SML-based visual contrast, as described below.

2.1. Dual-Tree Complex Wavelet Transform (DTCWT)

Wavelet transforms provide a framework in which an image is decomposed, with each level corresponding to a coarser resolution band [17]. Once the wavelet transform is implemented, every second wavelet coefficient at each decomposition level is discarded, resulting in components that are highly dependent on their location in the subsampling vector and with great uncertainty as to when they occurred in time [27]. This unfavorable property is referred to as shift variance (throwing away 1 of every 2 samples). Coupled with the limitations in the direction, wavelet analysis thus cannot accurately represent the directions of the edges of images [22]. To overcome these shortcomings of DWT, in 1998 Kingsbury [26] proposed the dual-tree complex wavelet transform (DTCWT), an over-complete wavelet transform which provides both good shift invariance and directional selectivity. The DTCWT idea is based on the use of two parallel trees, one for the odd samples and the other one for the even samples generated at the first level. These trees provide the signal delays necessary for every level and hence eliminate aliasing effects and achieve shift invariance [31]. The horizontal and vertical sub-bands are divided into six distinct sub-bands which are ± 15 , ± 45 , ± 75 . The advantages of DTCWT, *i.e.*, shift invariance and directional sensitivity, give improved fusion results that outperform the discrete wavelet transform (DWT) [32]. Figure 1 shows the accumulated reconstructions from each level of DTCWT and DWT after four levels of decomposition. As shown the edge of the decomposed image (highlighted in the red rectangle), pseudo-Gibbs phenomena appear in DWT, while DTCWT shows smooth edges. That is because DTCWT can represent the line singularities and plane singularities more effectively than DWT. Detailed information about the design of DTCWT may be found in [31].

Figure 1. Accumulated reconstructions from each level of DTCWT (**top**) and DWT (**bottom**).



2.2. Sum-Modified-Laplacian Based Visual Contrast

For spatial domain based multi-focus image fusion, there are many typical focus measurements such as energy of image gradient (EOG), spatial frequency (SF), tenengrad, Laplacian energy and SML. In [33], the authors compared these measurements through extensive experiments, and the results proved that SML is the best measurement. In the transform domain, SML is also very efficient and can produce the best fused results [34]. The definition of SML is as follows:

$$SML(i, j) = \sum_{m=-M_1}^{M_1} \sum_{n=-N_1}^{N_1} [ML(i+m, j+n)], \tag{1}$$

where:

$$ML(i, j) = |2L(i, j) - L(i - step, j) - L(i + step, j)| + |2L(i, j) - L(i, j - step) - L(i, j + step)|, \tag{2}$$

where $(2M_1 + 1)(2N_1 + 1)$ is the size of the window, $step$ is a variable spacing between coefficients and always equal to 1 [33]. $L(i, j)$ denotes the pixel value of one coefficient located at (i, j) . According to physiological and psychological research, the human visual system is highly sensitive to the local contrast of the image rather than the value of the pixel [35]. To meet this requirement, local visual contrast is proposed [36]. Considering the excellent clear measurement of SML, the visual contrast based on SML [30] is introduced in this article. This scheme is defined as:

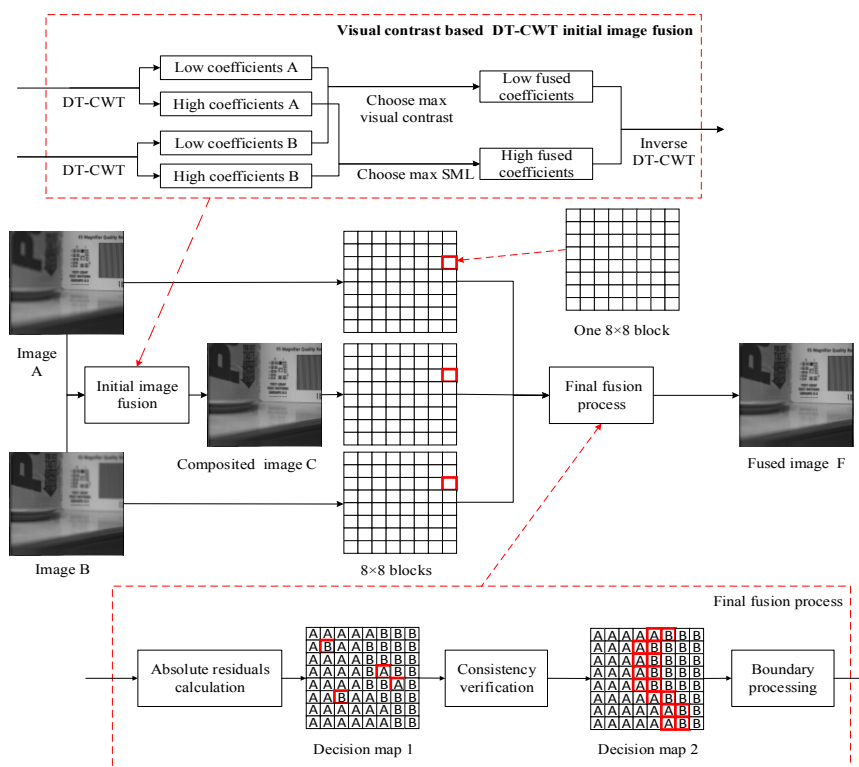
$$VC_{SML}(i, j) = \begin{cases} \frac{SML(i, j)}{\hat{L}(i, j)^{1+\alpha}}, & \text{if } \hat{L}(i, j) \neq 0 \\ SML(i, j), & \text{otherwise} \end{cases}, \tag{3}$$

where α is a visual constant representing the slope of the best-fitted lines through high-contrast data, which is determined by physiological vision experiments, and it ranges from 0.6 to 0.7 [34]. $\hat{L}(i, j)$ is the mean intensity value of the pixel (i, j) centered of the neighborhood window.

3. The Proposed Image Fusion Method

In this section, the proposed image fusion framework is depicted in Figure 2.

Figure 2. The framework of the proposed fusion scheme.



The framework includes two fusion procedures: the DTCWT based initial image fusion process and a hybrid final fusion process. The two processes are described as follows.

3.1. DTCWT Based Initial Image Fusion

In this subsection, the DTCWT based initial fusion, which applies new fusion rules for low- and high-frequency coefficients, respectively, is described in Algorithm 1.

Algorithm 1. DTCWT based initial image fusion

Input: Source images.

Step 1: Load the source multi-focus images.

Step 2: Perform three levels (how to set the level can be seen in literature [32]) DTCWT on source images to obtain two low frequency sub-bands and six high frequency sub-bands at each level. These sub-bands are denoted as:

$$\left\{ \left(L_t^X, H_{l,d}^X \right) \middle| X = A, B; l = 1, 2, 3; t = 1, 2; d = 1, \dots, 6 \right\}, \quad (4)$$

where L_t^X is the low frequency sub-bands in the t orientation, $H_{l,d}^X$ represent the high frequency sub-bands l level in the d orientation.

Step 3: Fuse low- and high-frequency sub-bands via the following different rules to obtain composite low- and high-frequency sub-bands. The detail of low coefficients and high coefficients fusion are discussed as follows.

Step 4: Perform three levels inverse DTCWT on the composited low- and high-frequency sub-bands to obtain the composited image.

Output: The initial composited image.

The low frequency coefficients represent the approximate information and contain the most energy of the source images. The widely used rule is to apply averaging methods to produce the fused coefficients. However, this rule will reduce the contrast in the fused images. According to [30], the SML-based visual contrast is a criterion which considers the nonlinear relationship between the contrast sensitivity threshold of HVS and the background luminance. Therefore, we introduce SML-based visual contrast to fuse the low frequency sub-bands and the fusion process is given as follows:

$$L_t^C(i, j) = \begin{cases} L_t^A(i, j), & \text{if } VC_{SML_t}^A(i, j) \geq VC_{SML_t}^B(i, j) \\ L_t^B(i, j), & \text{otherwise} \end{cases}, \quad (5)$$

where $VC_{SML_t}^A(i, j)$ and $VC_{SML_t}^B(i, j)$ are the visual contrast extracted from low frequency sub-bands at the t -th direction of the corresponding source image.

For the high frequency coefficients, the most popular fusion rule is to select the coefficients with larger absolute values, but this rule does not take any consideration of the surrounding pixels. The SML operator is developed to provide local measures of the quality of image focus [29]. In [33], it is proved that the SML is very efficient in the transform domain. Therefore, the composite rule for the high-frequency sub-bands is defined as:

$$H_{l,d}^C(i,j) = \begin{cases} H_{l,d}^A(i,j), & \text{if } \text{SML}_{l,d}^A(i,j) \geq \text{SML}_{l,d}^B(i,j) \\ H_{l,d}^B(i,j), & \text{otherwise} \end{cases}, \quad (6)$$

where $\text{SML}_{l,d}^A(i,j)$ and $\text{SML}_{l,d}^B(i,j)$ represent the SML extracted from high frequency sub-bands at the l -th scale and d -th direction of the corresponding source image.

3.2. The Final Fusion Process

After DTCWT-based initial image fusion, a novel hybrid method for obtaining the final fused image is proposed and the final fusion process is described in Algorithm 2. The detailed description of some key techniques and algorithms is as follows.

Algorithm 2. The final fusion process

Input: The source images and the initial composited image.

Step 1: Load the source images and the initial composited image.

Step 2: The source images and the initial fused image are partitioned into non-overlapping blocks.

Step 3: Calculate the values of image block residual between the source images and initial composited image.

Step 4: Compare the values of $RV^A(i_0, j_0)$ and $RV^B(i_0, j_0)$ (block residuals of A and B), and then obtain an initial decision map $M_1(i_0, j_0)$.

Step 5: Consistency verification is employed to eliminate the defects of $M_1(i_0, j_0)$ that caused by noise or undesired effects, and then get a modified decision map $CV(i_0, j_0)$ for next step.

Step 6: Get a final decision map $Z(i_0, j_0)$ by boundary processing. The final process can thus be illustrated as follows:

$$F(i_0, j_0) = \begin{cases} A(i_0, j_0), & \text{if } Z(i_0, j_0) > 0 \\ B(i_0, j_0), & \text{if } Z(i_0, j_0) < 0, \\ C(i_0, j_0), & \text{if } Z(i_0, j_0) = 0 \end{cases} \quad (7)$$

where $A(i_0, j_0)$, $B(i_0, j_0)$, $C(i_0, j_0)$, $F(i_0, j_0)$ denote the pixels of the block located at (i_0, j_0) of the source images, the composited image, and the final fused image, respectively.

Output: The final fused image.

3.2.1. Image Block Residual

In multi-focus images, the focused region is more informative [28]. It is easy to find that blocks in the focus area have greater similarity to the blocks of the initial fused image. Therefore, the image block residual is proposed to test the similarity, which is defined as:

$$RV^A(i_0, j_0) = \left| \sum_{i=N_0i_0-N_0+1}^{N_0i_0} \sum_{j=N_0j_0-N_0+1}^{N_0j_0} C(i,j)^2 - \sum_{i=N_0i_0-N_0+1}^{N_0i_0} \sum_{j=N_0j_0-N_0+1}^{N_0j_0} A(i,j)^2 \right| \quad (8)$$

$$RV^B(i_0, j_0) = \left| \sum_{i=N_0i_0-N_0+1}^{N_0i_0} \sum_{j=N_0j_0-N_0+1}^{N_0j_0} C(i,j)^2 - \sum_{i=N_0i_0-N_0+1}^{N_0i_0} \sum_{j=N_0j_0-N_0+1}^{N_0j_0} B(i,j)^2 \right| \quad (9)$$

where $N_0 \times N_0$ is the size of the blocks (how to set the size of block can be seen in literature [5,7]). $A(i, j)$, $B(i, j)$, and $C(i, j)$ are the pixel values of the source images and the composited image, respectively.

The block residual can be assumed as a similarity measure in image processing applications. Constructing an initial decision map by:

$$M_1(i_0, j_0) = \begin{cases} -1, & \text{if } RV^A(i_0, j_0) < RV^B(i_0, j_0) \\ 1, & \text{otherwise} \end{cases} \quad (10)$$

3.2.2. Consistency Verification

In a scene, we assume most blocks of a region are in the depth of focus of one source image. According to the theory of imaging, all the blocks in such region must be chosen from the image [5]. However, there can be some erroneous selection of some blocks due to noise or undesired effects. To remove these defects, consistency verification [37] is employed. If the center block comes from one image while the majority of the surrounding blocks come from the other one, the center sample is simply switched to the corresponding block in the other image. Here, consistency verification is applied in a 3×3 neighborhood window as in [7]. After consistency verification to $M_1(i_0, j_0)$, we get the modified decision map $CV(i_0, j_0)$ for boundary processing.

3.2.3. Boundary Processing

Assume one block contains both clear areas and blurred areas, so it is meaningless to judge the block is in the focus area or not. According to the theory of imaging, these blocks are always on the boundary of focused region. Therefore, we could choose these blocks by the equations below:

$$Z(i_0, j_0) = \max(\{M_2(i_0, j_0), M_2(i_0+1, j_0), M_2(i_0, j_0+1), M_2(i_0+1, j_0+1)\}) \\ + \min(\{M_2(i_0, j_0), M_2(i_0+1, j_0), M_2(i_0, j_0+1), M_2(i_0+1, j_0+1)\}), \quad (11)$$

in which:

$$M_2(i_0, j_0) = \begin{cases} 1, & \text{if } CV(i_0, j_0) > 0 \\ -1, & \text{otherwise} \end{cases} \quad (12)$$

If $M_2(i_0, j_0) = 1$, the block at (i_0, j_0) from image A is from the focus area, and if $M_2(i_0, j_0) = -1$, the block at (i_0, j_0) from image B is from the focus area. Equation (11) means if one of $M_2(i_0, j_0)$, $M_2(i_0+1, j_0)$, $M_2(i_0, j_0+1)$ and $M_2(i_0+1, j_0+1)$ equals to 1 and another one equals to -1 , the block at (i_0, j_0) is on the boundary of focused region.

As we know, the initial fused image has extracted the most characteristics of the source images to some extent. On the boundary of the focused region, the blocks may contain both clear pixels and blurred pixels. To reduce the loss of useful information, we should choose the blocks of the initial fused image on the boundary of the focused region. Then, we can use the Equation (7) to get the fusion result.

4. The Experimental Results and Analysis

In this section, the evaluation index system and the experimental results and analysis are described.

4.1. Evaluation Index System

Generally speaking, the measure of image fusion quality can be divided into two stages: subjective metrics and objective metrics [38]. Subjective evaluations always depend on the human visual characteristics and the professional knowledge of observers, thus the processes are time-consuming and have poor repeatability [39]. Objective evaluations can be easily performed by computers, completely automatically, and generally evaluate the similarity between the fused and the referenced image or source images [40]. Therefore, besides visual observation, six objective performance metrics are used to demonstrate the superior performance of the proposed algorithm. The quantitative evaluations can be divided into two classes: objective evaluation of results of referenced images and objective evaluation of results of non-referenced images.

4.1.1. Objective Evaluation of Results of Referenced Images

- (1) *Root Mean Square Error (RMSE)*: The RMSE is the cumulative squared error between the fused image and the referenced image and it is defined by:

$$\text{RMSE} = \sqrt{\frac{\sum_{m=-M}^M \sum_{n=-N}^N [F(i+m, j+n) - R(i+m, j+n)]^2}{MN}} \quad (13)$$

where $M \times N$ is the size of fused image, $F(i, j)$ and $R(i, j)$ are the pixel value of the fused image and referenced image at the position (i, j) , respectively. The lower the value of RMSE, the better the fusion effect.

- (2) *Peak Signal to Noise Ratio (PSNR)*: The PSNR is a ratio between the maximum possible power of a signal and the power of noise that affects the fidelity [41]. The PSNR is formulated as:

$$\text{PSNR} = 10 \lg \left| \frac{L^2}{\text{RMSE}^2} \right| \quad (14)$$

where L is the dynamic range of the pixel values. Here, $L = 255$ is adopted. The greater the value of PSNR, the better the fusion effect.

- (3) *Structural Similarity-based Metric (SSIM)*: The SSIM is designed by modeling any image distortion as the combination of the loss of correlation and radiometric and contrast distortion [42]. SSIM between the fused image and the referenced image is defined as follows:

$$\text{SSIM}(R, F) = \frac{\left[2u_R u_F - (K_1 L)^2 \right] \left[2\sigma_{RF} + (K_2 L)^2 \right]}{\left[u_R^2 + u_F^2 + (K_1 L)^2 \right] \left[\sigma_R^2 + \sigma_F^2 + (K_2 L)^2 \right]} \quad (15)$$

in which u_R, u_F are mean intensity of the standard image and the fused image, respectively. $\sigma_R, \sigma_F, \sigma_{RF}$ are the variances and covariance, respectively [43]. K_1 and K_2 are small constants, and L is the dynamic range of the pixel values. Here, the default parameters: $K_1 = 0.01, K_2 = 0.03, L = 255$ is adopted [44]. The greater the value of SSIM, the better the fusion effect.

4.1.2. Objective Evaluation of Results of Non-Referenced Images

- (1) *Standard deviation*: The standard deviation can be used to estimate how widely spread the gray values in an image and defined as [45]:

$$\text{Std} = \sqrt{\left(\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (F(i, j) - \hat{\mu})^2 \right)}, \quad (16)$$

where $\hat{\mu}$ is the mean value of the fused image. The larger the standard deviation, the better the result.

- (2) *Mutual Information (MI)*: The MI can indicate how much information the fused image conveys about the source images [33]. MI between fusion image and the source images is defined as follows [46]:

$$\text{MI} = \text{MI}^{AF} + \text{MI}^{BF}, \quad (17)$$

in which:

$$\text{MI}^{AF} = \sum_{f=0}^L \sum_{a=0}^L p^{AF}(a, f) \log_2 \left(\frac{p^{AF}(a, f)}{p^A(a) p^F(f)} \right), \quad (18)$$

$$\text{MI}^{BF} = \sum_{f=0}^L \sum_{b=0}^L p^{BF}(b, f) \log_2 \left(\frac{p^{BF}(b, f)}{p^B(b) p^F(f)} \right), \quad (19)$$

where MI^{AF} and MI^{BF} denote the normalized MI between the fused image and the source image 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 fused image. $p^{AF}(a, f)$ and $p^{BF}(b, f)$ are the joint gray level histograms between the fused image and the source image A and B. The greater the value of MI, the better the fusion effect.

- (3) *Edge Based Similarity Measure ($Q^{AB/F}$)*: $Q^{AB/F}$ which proposed by Xydeas and Petrovic [47] gives the similarity between the edges transferred from the source images to the fused image. The definition is given as:

$$Q^{AB/F} = \frac{\sum_{i=1}^M \sum_{j=1}^N [Q^{AF}(i, j)w^A(i, j) + Q^{BF}(i, j)w^B(i, j)]}{\sum_{i=1}^M \sum_{j=1}^N [w^A(i, j) + w^B(i, j)]}, \quad (20)$$

in which:

$$Q^{AF}(i, j) = Q_a^{AF}(i, j)Q_g^{AF}(i, j), \quad (21)$$

$$Q^{BF}(i, j) = Q_a^{BF}(i, j)Q_g^{BF}(i, j), \quad (22)$$

where $w^A(i, j)$ and $w^B(i, j)$ are the corresponding gradient strengths for images A and B, respectively. $Q_a^{AF}(i, j)$ and $Q_g^{AF}(i, j)$ are the Sobel edge strength and orientation preservation values at location (i, j) for each source image [48]. The larger the value, the better the fusion effect result.

4.2. Experiments on Multi-Focus Image Fusion

The proposed fusion framework is tested on three sets of multi-focus image databases: non-referenced images, referenced images, and images with different noise. For all these image

databases, the results of the proposed fusion scheme are compared with those of the state-of-the-art wavelet transform-, FDCT-, DTCWT-, and NSCT-based methods. The low- and high-frequency coefficients of the wavelet transform-, FDCT-, and DTCWT (DTCWT-1)-based methods are merged by the widely used fusion rule of averaging and selecting larger absolute values of the coefficient (average-maximum rule), respectively. The DTCWT-2 method is the DTCWT-based initial image fusion technique proposed in this paper. In addition, we introduce two NSCT-based methods (NSCT-1 and NSCT-2) for comparison. NSCT-1 involves merging the coefficients by an average-maximum rule; NSCT-2 merges the coefficients by SF_PCNN [25], which was proposed by Qu *et al.* in [25] and can make full use of the characteristics of PCNN and has been proven to outperform PCNN in the NSCT domain in fusing multi-focus images. In order to perform a fair comparison, the source images are all decomposed into three levels for those methods, except for the FDCT-based method (five is proven in [9] to be the best level in the FDCT-based image fusion method). For the wavelet-based method, the images are decomposed using the DBSS (2, 2) wavelet. For implementing NSCT, “9-7” filters and “pkva” filters (how to set the filters can be found in [49]) are used as the pyramidal and directional filters, respectively. It should be noted that since image registration is out of scope of this paper, like all the literatures, in all cases we assume the source images have been in perfect registration. A thorough survey of image registration techniques can be found in [50]. All the experiments are implemented in MATLAB R2012a on an Intel (R) Core (TM) i3-2330M CPU @2.2 GHz computer (ASUS, Taipei, Taiwan) with 4 GB RAM.

4.2.1. Fusion of Referenced Multi-Focus Images

The first experiment is conducted using an extensive set of artificially generated images. Figure 3 shows the referenced Lena image, and two blurred artificial images that are generated by convolving the referenced image with an averaging filter centered on the left part and the middle part, respectively [51]. The fusion results of different images are shown in Figure 4a–g. For a clearer comparison, this paper has given the residual images [28] between the results and Figure 3b, and performed the following transform on each pixel of each residual image:

$$D^*(i, j) = 3 \times D(i, j) + 50 \quad (23)$$

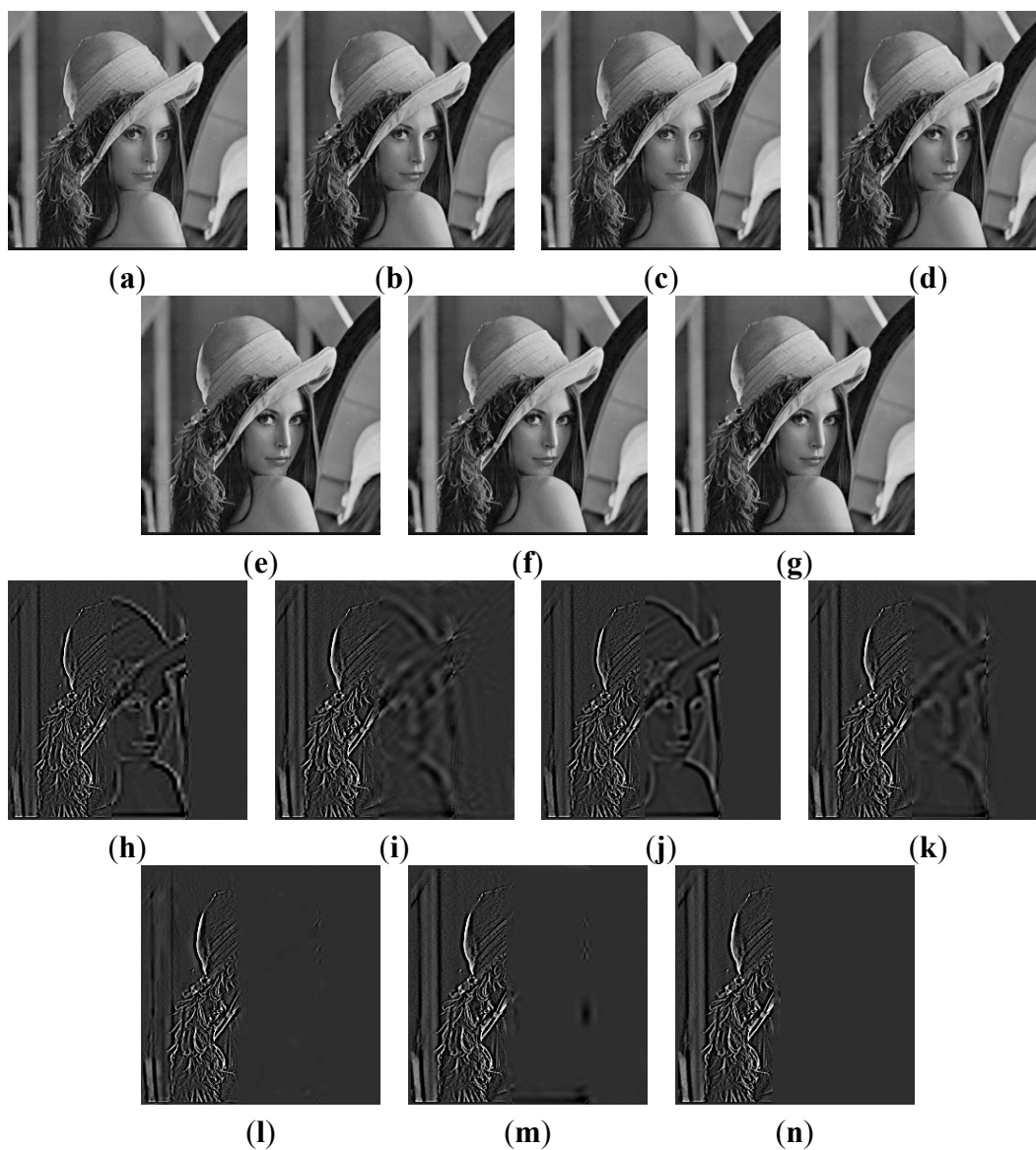
where the $D(i, j)$ is the pixel of residual image located at (i, j) . Figure 4h–n shows the sharpened residual images.

For the focused areas, the residual values between the fused images and the source image should be zero. According to the literature [9], in residual images, the lower residue the features, the more detail information in the focus region has been transferred to fused images and the better the corresponding method. From Figure 4, as mentioned above, the residual image between the result of the proposed method and Figure 3b is almost zero in the clear region. That is to say, the proposed method has extracted the most information from the clear region of Figure 3b.

Figure 3. Referenced and blurred Lena images. (a) Referenced image; (b) Blurred on left; (c) Blurred in the middle.

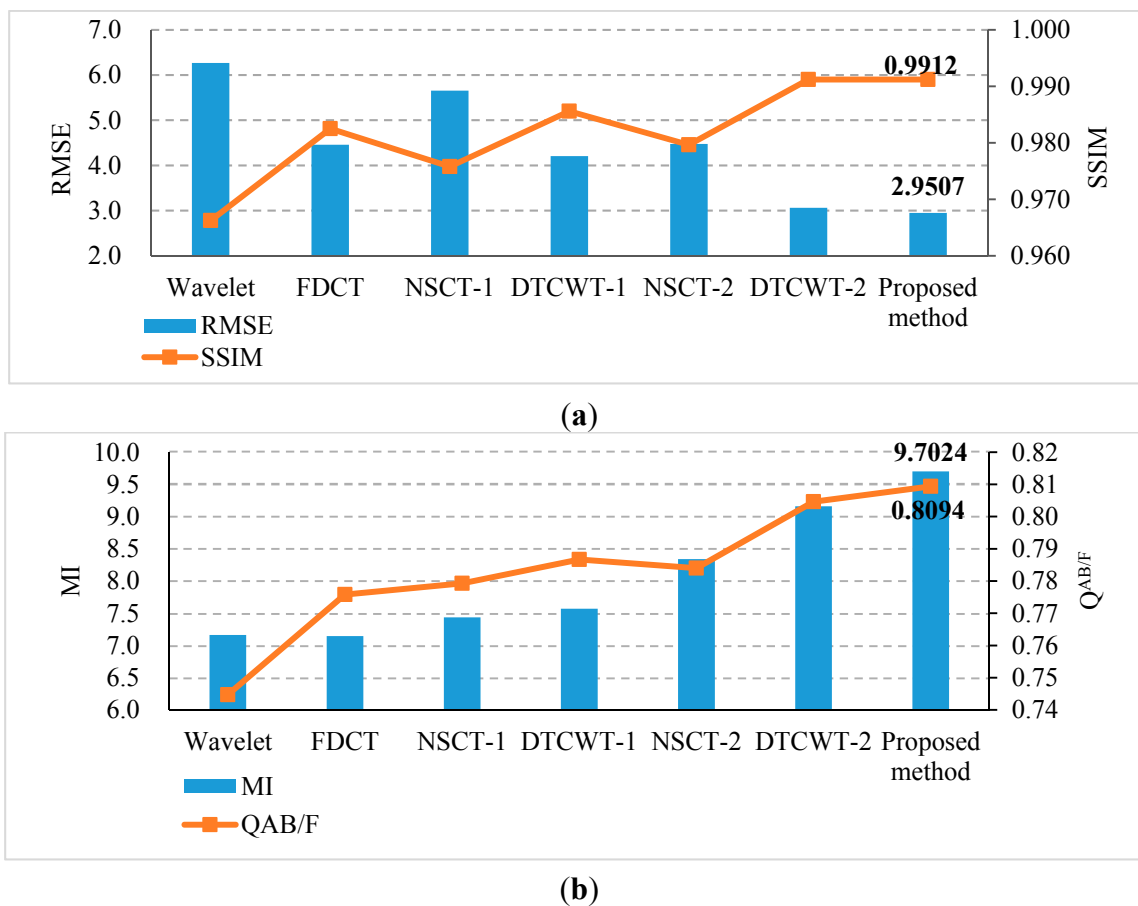


Figure 4. Fusion results of the Lena image. Fused images of: (a) Wavelet transform-based method; (b) FDCT-based method; (c) NSCT-1-based method; (d) DTCWT-1-based method; (e) NSCT-2-based method; (f) DTCWT-2-based method; (g) our proposed method; (h–n) the sharpened residual images between the seven results and Figure 3b.



RMSE and SSIM are the metrics that measure the error or similarity between a referenced image and a fused image. Using these metrics the effect of the methods in the fusion of referenced multi-focus images can be measured well. In addition, MI and $Q^{AB/F}$ are the state-of-the-art fusion performance metrics [25,40], which can be used to measure the fusion of no-reference multi-focus images. Figure 5 depicts the comparison on quantitative metrics of different methods; it is easy to find that the SSIM value of the proposed method is the largest (0.9912) and RMSE value is the smallest (2.9507). Both of them are the optimal values. That is to say, compared with other methods, the fused image of the proposed method is the closest to the referenced image. In addition, the MI and $Q^{AB/F}$ values of the proposed method results also outperform the other methods.

Figure 5. Comparison on (a) RMSE, SSIM and (b) MI, $Q^{AB/F}$ of different methods for Lena image.



For further comparison, we experimented with more standard images. Figure 6 shows the standard test image database. The database is the standard grayscale test images that have been downloaded from [52]. The sizes of the test images are all 512×512 . Each standard test image has been convolved with an averaging filter centered on the left part and right part, respectively. Figure 7 shows the typical examples of the fusion of artificial multi-focus images, the results of different methods and the corresponding residual images. From Figure 7, the residual images show that the proposed method results have extracted the most information from the focus region. Figure 8 shows the comparison on RMSE and SSIM of different methods for artificial multi-focus images. From Figure 8, it is seen that the results of the proposed method almost represent the referenced images with SSIM values

approaching 1. From Figure 8, it is easy to find that in 30 sets of images, the SSIM and RMSE values of the proposed method results outperform other methods.

Figure 6. Standard gray test image database.

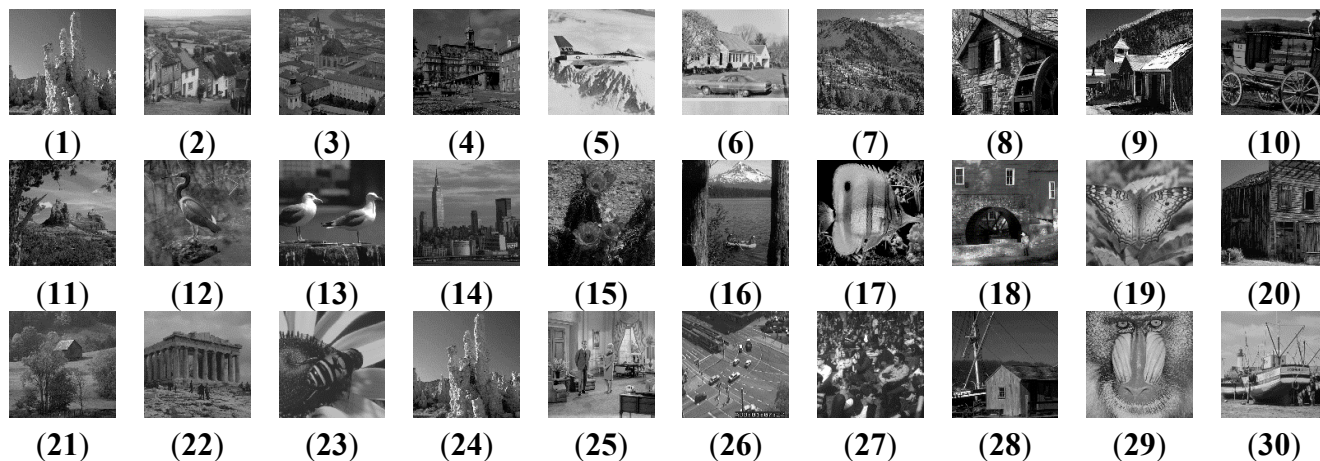


Figure 7. Two examples of the fusion of artificial multi-focus images. (a1,a2) Original images; (b1,b2) Images focused on right; (c1,c2) Images focused on left; (d1,d2) Wavelet results; (e1,e2) FDCT results; (f1,f2) NSCT-1 results; (g1,g2) DTCWT-1 results; (h1,h2) NSCT-2 results; (i1,i2) proposed method results; (j1–o1,j2–o2). The sharpened residual images between the six results and Figure 7b1,b2.

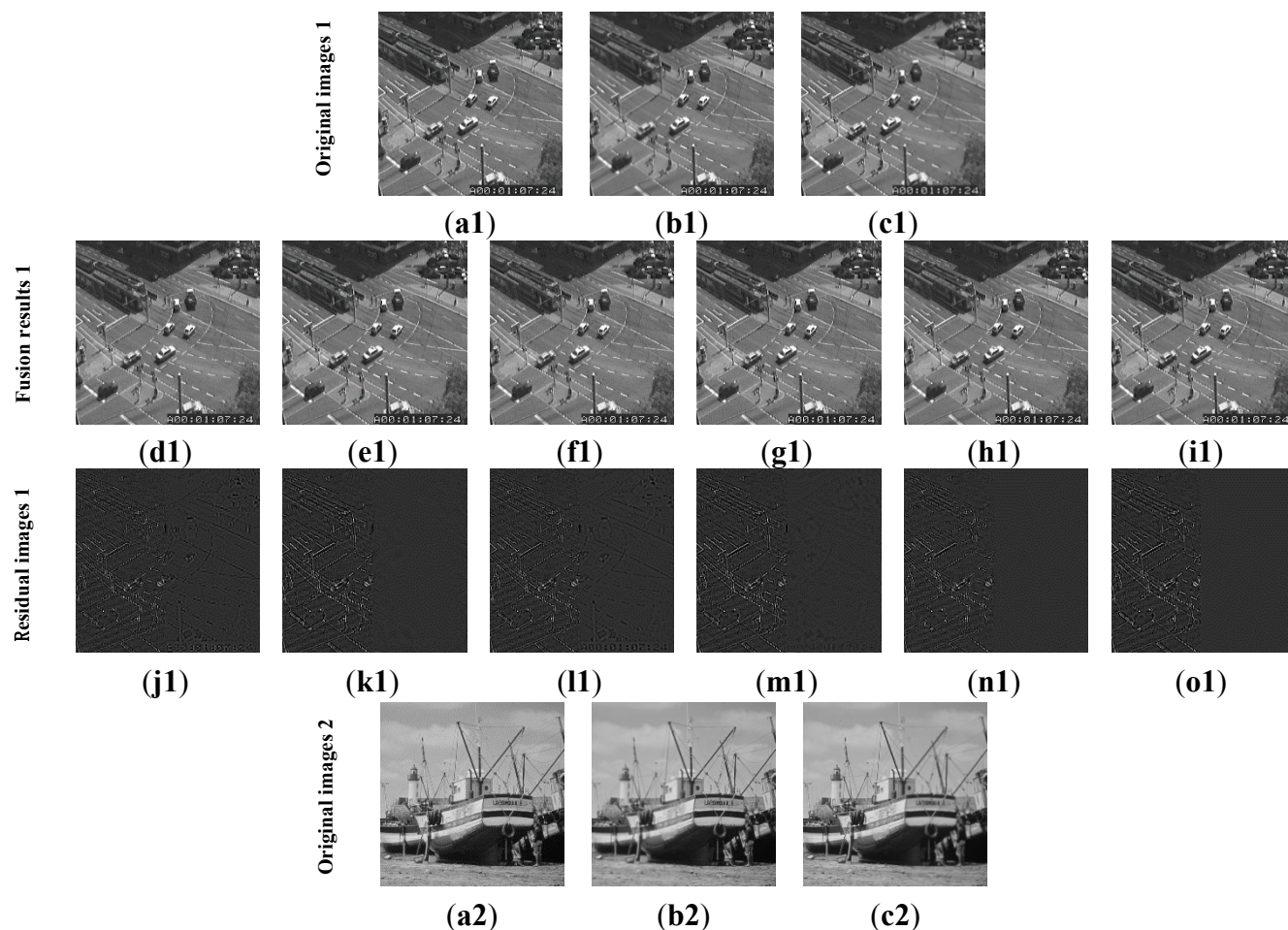


Figure 7. Cont.

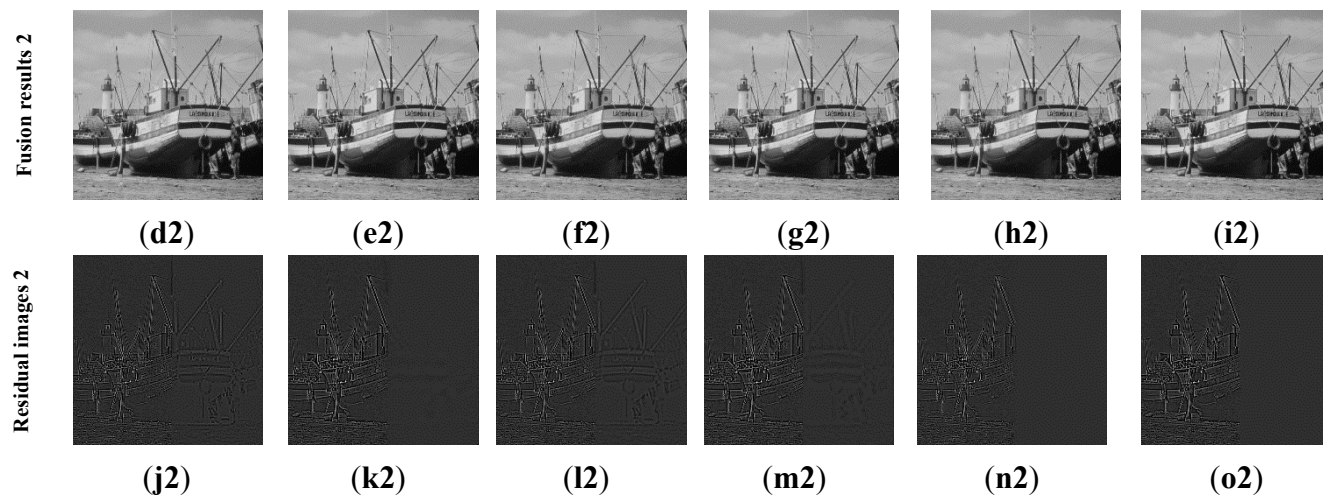
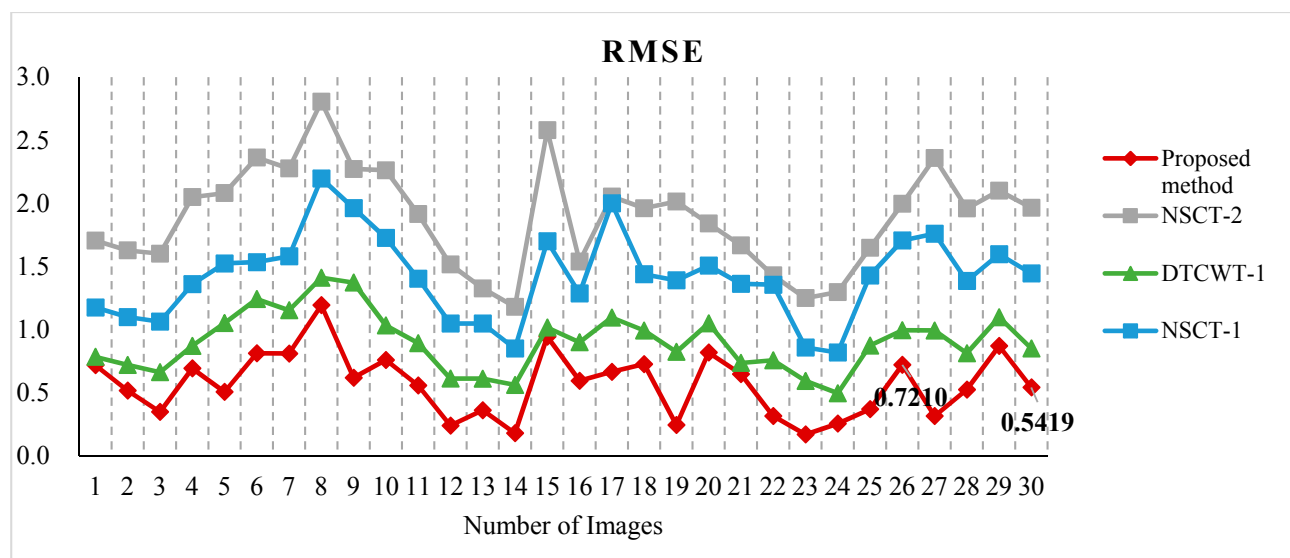
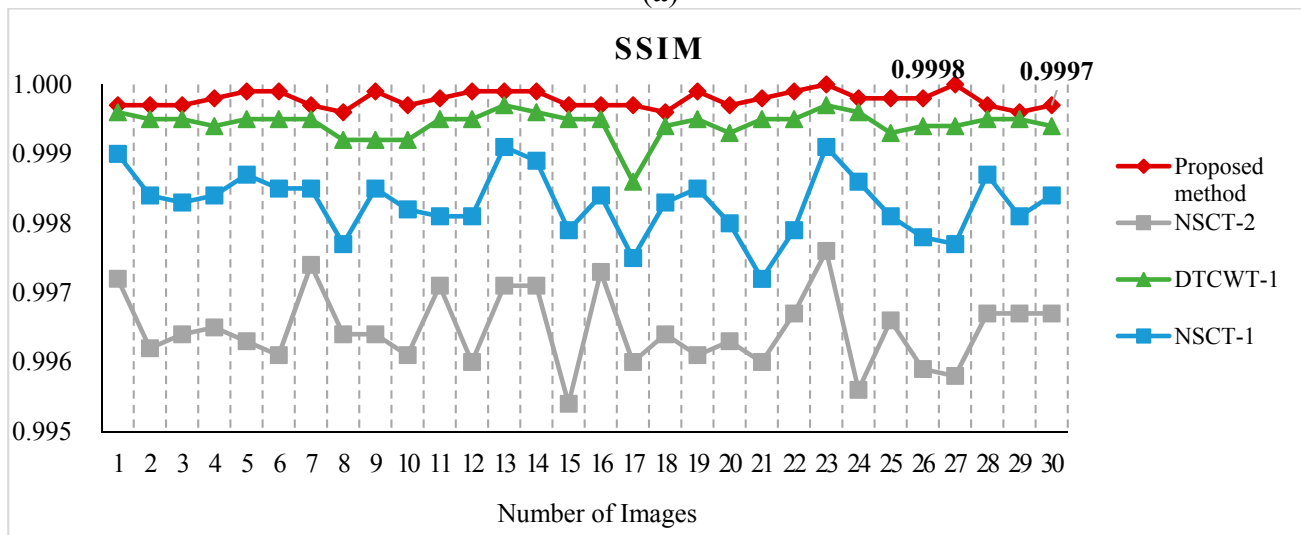


Figure 8. Comparison on (a) RMSE and (b) SSIM of different methods for artificial multi-focus images.



(a)



(b)

4.2.2. Fusion of Images in a Noisy Environment

To evaluate the noise-robustness of the proposed framework, the second experiment is conducted on two blurred “Lena” images as shown in Figure 3b–c additionally corrupted with different Gaussian noise, with a standard deviation of 1%, 3%, 5%, 7% and 9%, respectively. Figure 9 shows the blurred images with 5% noise and the corresponding results of different methods. A quantitative evaluation comparison of the different methods for artificial images with different noise is shown in Figure 10. From Figure 10a, the PSNR values of the proposed method results are all the highest of the six methods. That is to say, in different noise environments, the results of the proposed method are closer to the referenced image than other methods. From Figure 10b,c, the MI and $Q^{AB/F}$ values of the proposed method also outperform other methods. From the above analysis, we can observe that the proposed method provides the best performance and outperforms the other algorithms, even in noisy environments.

Figure 9. An example of the fusion for noise image. (a,b) Blurred Lena images with 5% noise; (c) Wavelet result; (d) FDCT result; (e) NSCT-1 result; (f) DTCWT-1 result; (g) NSCT-2 result; (h) proposed method result; (i–n) the sharpened residual images between the six results and Figure 9a.

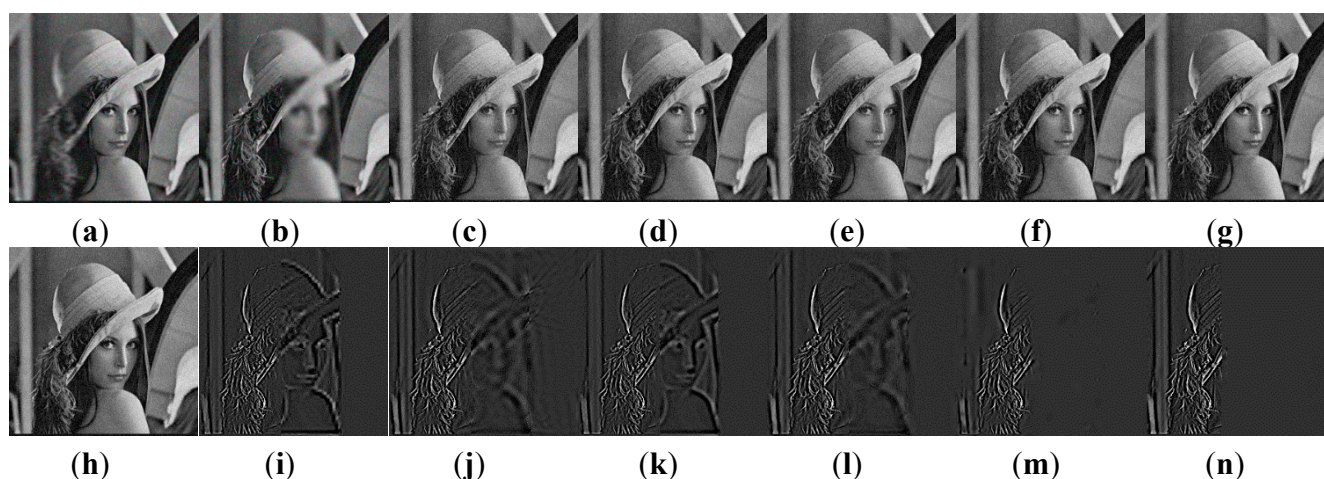
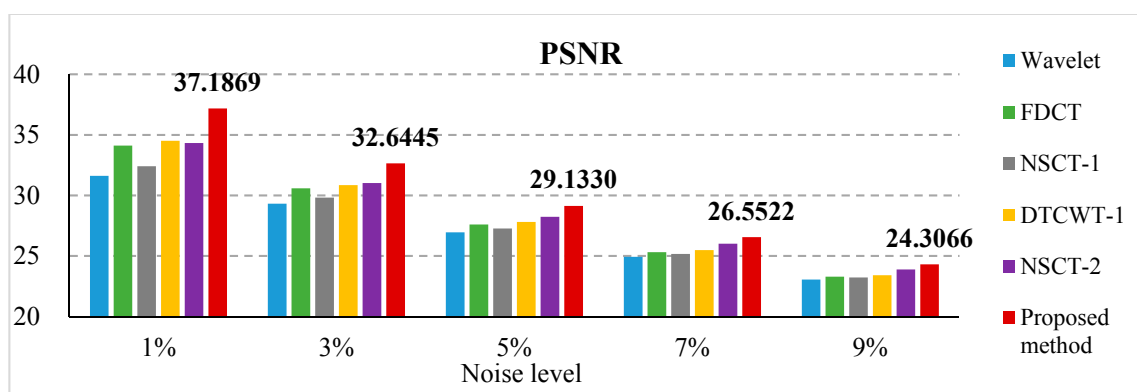
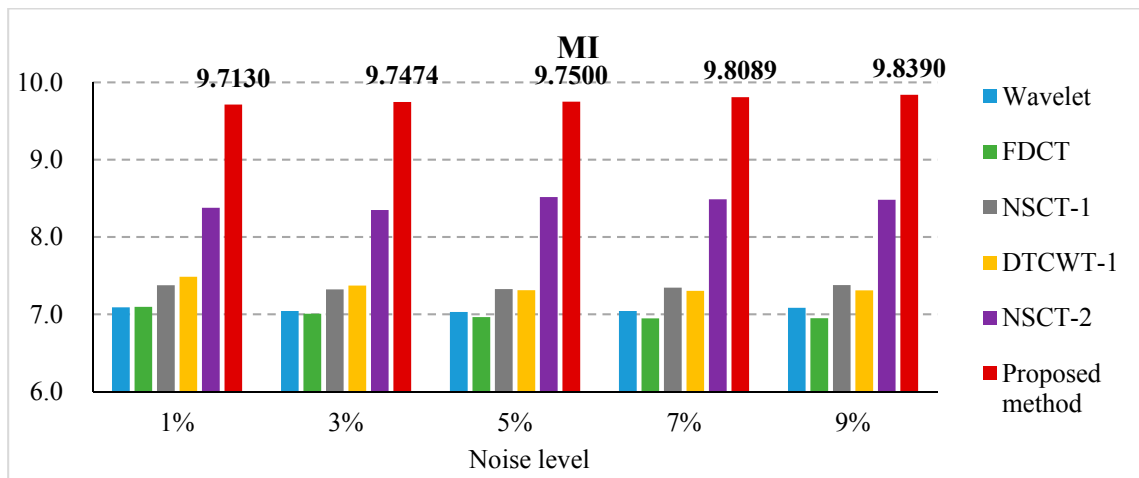


Figure 10. Comparison on (a) PSNR; (b) MI and (c) $Q^{AB/F}$ of different methods for multi-focus image with different noise.

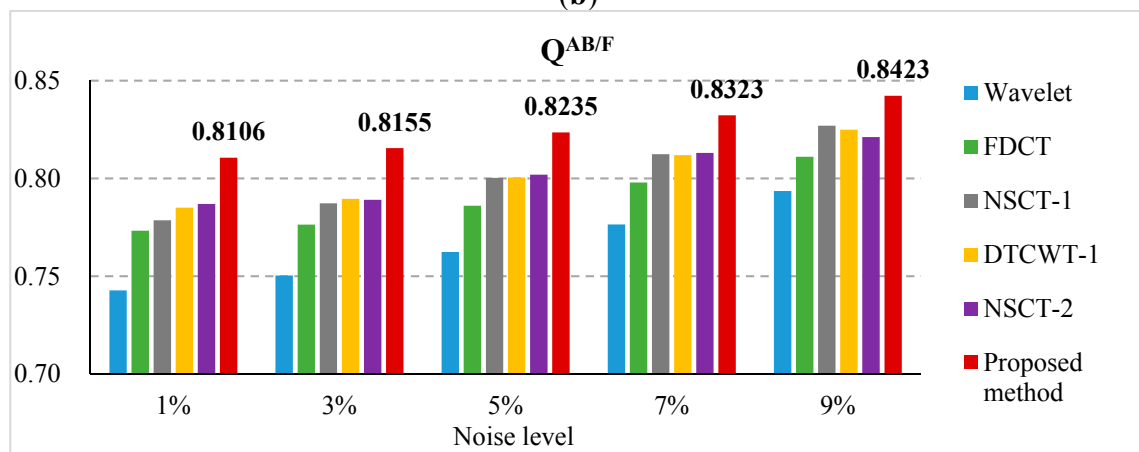


(a)

Figure 10. Cont.



(b)



(c)

4.2.3. Fusion of Non-Referenced Multi-Focus Images

The third experiment is conducted on four sets of non-referenced multi-focus test images. These images are shown in Figure 11. The size of Figure 11a,b is 256×256 , while the others are 480×640 . The fusion results are shown in Figure 12a1–f1, a2–f2 and their sharpened residual images between the fused result and image focused on background are shown in Figure 12g1–l1, g2–l2. From the residual images, it is easy to find in the focused background areas that the proposed method has extracted more information than other methods and almost all of the useful information in the focus area of the source images has been transferred to the fused image.

These experiments have no referenced images. Therefore, we used standard deviation, MI and $Q^{AB/F}$ to evaluate how much clear or detailed information of the source images is transferred to the fused images. The comparison of the quantitative evaluation of different methods for non-referenced multi-focus images is shown in Figure 13. Comparing the average values of the standard deviation, MI and $Q^{AB/F}$ of DTCWT 1 method and the proposed method, we can find that their corresponding values have been increased 1.37%, 29.50% and 8.26%, respectively. From Figure 13, the values of standard deviation, MI and $Q^{AB/F}$ of the proposed method result are all better than those of the other methods.

From the above analysis, we can also observe that the proposed scheme provides the best performance and outperforms the other algorithms.

Figure 11. Non-referenced multi-focus image database (**top**: focused on background; **bottom**: focused on foreground). Sets of (a) Pepsi image; (b) Clock image; (c) Disk image; (d) Lab image.

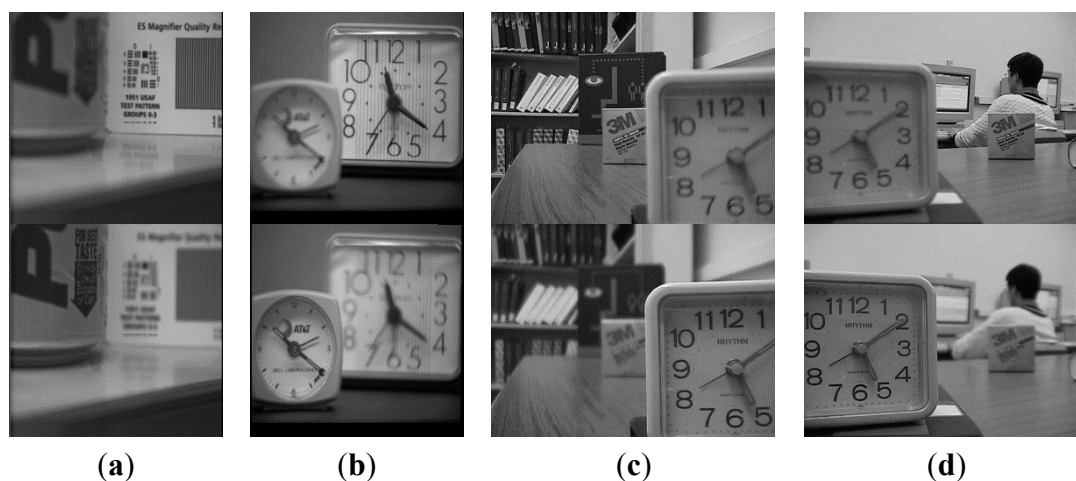


Figure 12. Fusion results of non-referenced multi-focus images. (a1,a2) Wavelet results; (b1,b2) FDCT results; (c1,c2) NSCT-1 results; (d1,d2) DTCWT-1 results; (e1,e2) NSCT-2 results; (f1,f2) Proposed method results; (g1–l1,g2–l2), The sharpened residual images between the six results and source image (image focused on background).

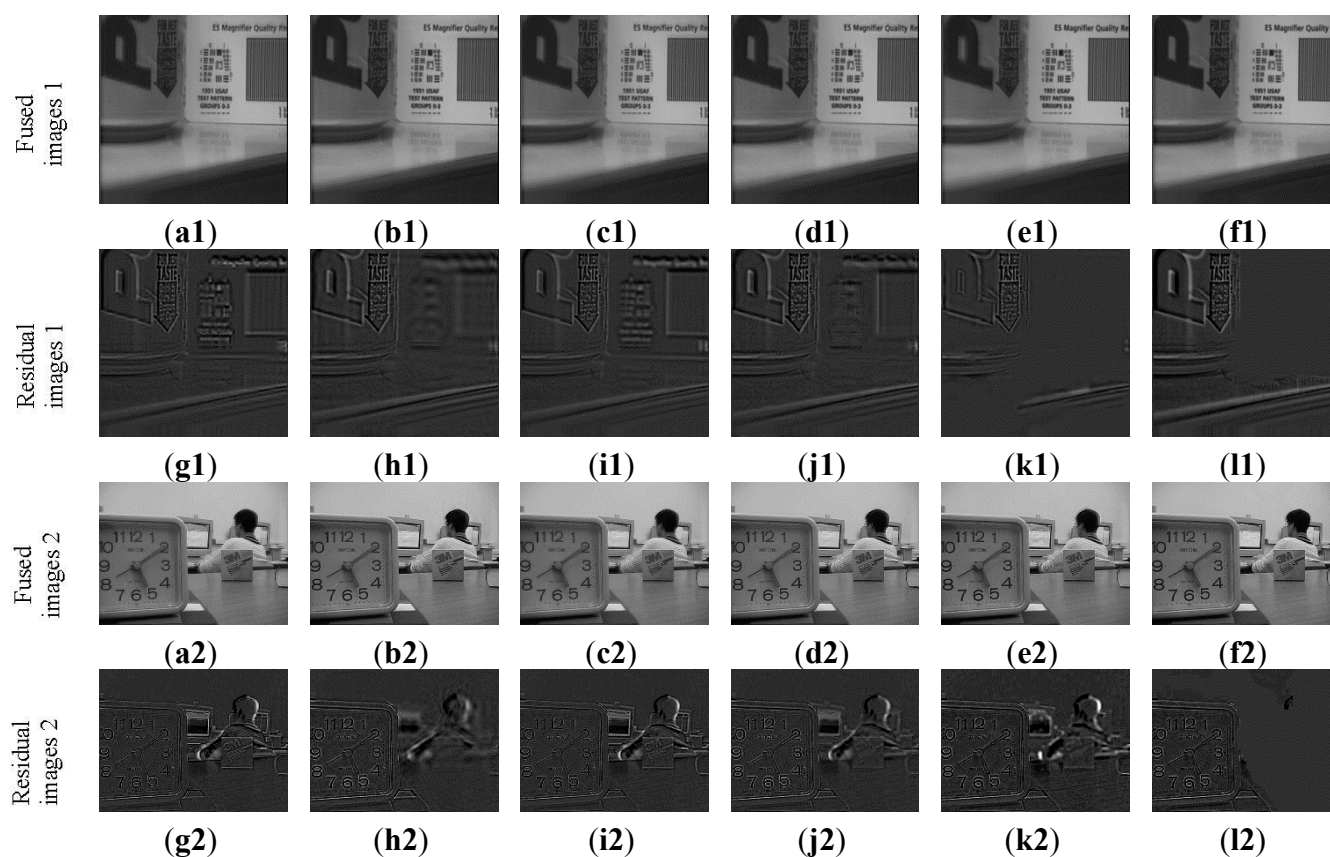
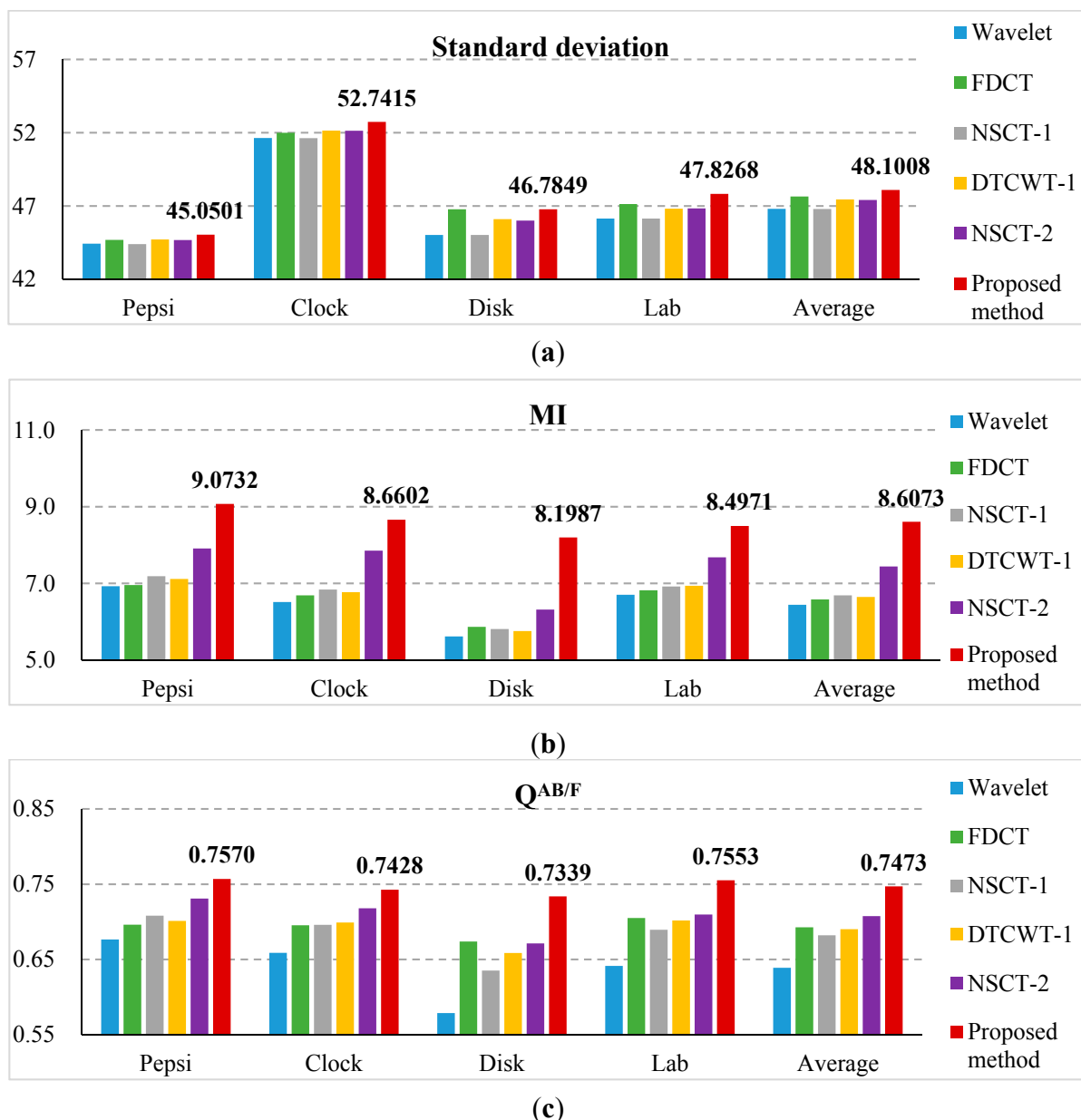


Figure 13. Comparison on (a) standard deviation; (b) MI and (c) $Q^{AB/F}$ of different methods for non-referenced multi-focus image.



Finally, we compare the time-consumption of the proposed method with the DTCWT- and NSCT- based methods. These methods can provide both good shift invariance and directional selectivity. A comparison of the execution times of the DTCWT-, NSCT-based methods and the proposed method for the Pepsi image sets is presented in Table 1.

Table 1. Execution time of different methods for Pepsi images.

Fusion Scheme	NSCT-1	DTCWT	NSCT-2	Proposed Method
Execution time (s)	14.4024	0.3273	32.6268	3.3020

The execution times are calculated by running all the codes in MATLAB R2012a on the same computer. From Table 1, it is easy to find that the NSCT-based methods (including NSCT-1 and

NSCT-2 methods) are time-consuming and of high complexity, and the efficiency of the proposed method and the DTCWT-based method is higher than that of the NSCT-based methods. Although the proposed method needs a little more execution time than the DTCWT method, the fusion performance including the above qualitative and quantitative evaluations of our method have been proved better than the DTCWT method. Hence, from the overall assessments, we can conclude that the proposed method is effective and suitable for real time applications.

5. Conclusions

In this paper, a novel DTCWT-based multi-focus image fusion framework is proposed for VSNs. The potential advantages include: (1) DTCWT is more suitable for image fusion because its advantages such as multi-directionality and shift-invariance; (2) use of the visual contrast and SML as fusion rules to strengthen the effect of DTCWT; and (3) image block residual and consistency verification are proposed to get a fusion decision map to guide the fusion process, which not only reduces the complexity of the procedure, but also increases the reliability and robustness of the fusion results. In the experiments, extensive sets of multi-focus images, including no-referenced images, referenced images, and images with different noise, are fused by using state-of-the-art methods and the proposed framework. In the experiments with the no-referenced images, comparing with simple DTCWT-based methods, the MI and $Q^{AB/F}$ values of the proposed method have been increased 29.50% and 8.26%, respectively. The experimental results have shown the superior performance of the proposed fusion framework. In the future, we plan to design a pure C++ platform to reduce the time cost and validate our framework in multi-sensor image fusion problems like infrared and visible light image fusion.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61262034, No. 61462031 and No. 61473221), by the Key Project of Chinese Ministry of Education (No. 211087), by the Doctoral Fund of Ministry of Education of China (No. 20120201120071), by the Natural Science Foundation of Jiangxi Province (No. 20114BAB211020, No. 20132BAB201025), by the Young Scientist Foundation of Jiangxi Province (No. 20122BCB23017), by the Science and Technology Application Project of Jiangxi province (No. KJLD14031), and by the Science and Technology Research Project of the Education Department of Jiangxi Province (No. GJJ14334).

Author Contributions

S.H. conceived and designed the experiments; S.T. performed the experiments; Y.Y. and S.T. analyzed the data and wrote the manuscript; P.L. modified the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Liu, Y.; Xiong, N.; Zhao, Y.; Vasilakos, A.V.; Gao, J.; Jia, Y. Multi-layer clustering routing algorithm for wireless vehicular sensor networks. *IET Commun.* **2010**, *4*, 810–816.
2. Ahamed, S.I.; Wang, W.; Hong J. Recent Advances in Energy-Efficient Sensor Networks. *Int. J. Distrib. Sens. Netw.* **2013**, *2013*, 329867.
3. Charfi, Y.; Wakamiya, N.; Murata, M. Challenging issues in visual sensor networks. *IEEE Wirel. Commun.* **2009**, *16*, 44–49.
4. Xiong, N.; Vasilakos, A.V.; Yang, L.T.; Pedrycz, W.; Zhang, Y.; Li, Y.S. A resilient and scalable flocking scheme in autonomous vehicular networks. *Mob. Netw. Appl.* **2010**, *15*, 126–136.
5. Phamila, Y.; Amutha, R. Discrete Cosine Transform based fusion of multi-focus images for visual sensor networks. *Signal Process.* **2014**, *95*, 161–170.
6. Li, H.; Xiong, N.; Park, J.H.; Cao, Q. Predictive control for vehicular sensor networks based on round-trip time-delay prediction. *IET Commun.* **2010**, *4*, 801–809.
7. Haghghat, M.B.A.; Aghagolzadeh, A.; Seyedarabi, H. Multi-focus image fusion for visual sensor networks in DCT domain. *Comput. Electron. Eng.* **2011**, *37*, 789–797.
8. Lee, I.H.; Mahmood, M.T.; Choi T.S. Robust Depth Estimation and Image Fusion Based on Optimal Area Selection. *Sensors* **2013**, *13*, 11636–11652.
9. Li, H.; Chai, Y.; Li, Z. A new fusion scheme for multi-focus images based on focused pixels detection. *Mach. Vis. Appl.* **2013**, *24*, 1167–1181.
10. Wang, J.; Lai, S.; Li, M. Improved image fusion method based on NSCT and accelerated NMF. *Sensors* **2012**, *12*, 5872–5887.
11. Pradhan, P.S.; King, R.L.; Younan, N.H.; Holcomb, D.W. Estimation of the number of decomposition levels for a wavelet-based multiresolution multisensor image fusion. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 3674–3686.
12. Li, S.; Kwok, J.T.; Wang, Y. Multi-focus image fusion using artificial neural networks. *Pattern Recognit. Lett.* **2002**, *23*, 985–997.
13. Tian, J.; Chen, L.; Ma L.; Yu, W. Multi-focus image fusion using a bilateral gradient-based sharpness criterion. *Opt. Commun.* **2011**, *284*, 80–87.
14. Singh, R.; Khare, A. Multiscale Medical Image Fusion in Wavelet Domain. *Sci. World J.* **2013**, *2013*, 521034.
15. Petrovic, V.S.; Xydeas, C.S. Gradient-based multiresolution image fusion. *IEEE Trans. Image Process.* **2004**, *13*, 228–237.
16. Aiazzi, B.; Alparone, L.; Barducci, A.; Baronti, S.; Pippi, I. Multispectral fusion of multisensor image data by the generalized Laplacian pyramid. In Proceeding of the IEEE 1999 International Geoscience and Remote Sensing Symposium, Hamburg, Germany, 28 June–2 July 1999; pp. 1183–1185.
17. Dong, J.; Zhuang, D.; Huang Y.; Fu, J. Advances in multi-sensor data fusion: Algorithms and applications. *Sensors* **2009**, *9*, 7771–7784.
18. Yang, Y.; Huang, S.Y.; Gao, J.F.; Qian, Z.S. Multi-focus Image Fusion Using an Effective Discrete Wavelet Transform Based Algorithm. *Meas. Sci. Rev.* **2014**, *14*, 102–108.

19. Li, H.; Guo, L.; Liu, H. Research on Image Fusion Based on the Second Generation Curvelet Transform. *Acta Opt. Sin.* **2006**, *26*, 657–662.
20. Drajić, D.; Cvejić, N. Adaptive fusion of multimodal surveillance image sequences in visual sensor networks. *IEEE Trans. Consum. Electr.* **2007**, *53*, 1456–1462.
21. Nirmala, D.E.; Vignesh, R.K.; Vaidehi, V. Multimodal Image Fusion in Visual Sensor Networks. In Proceeding of the 2013 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 17–19 January 2013; pp. 1–6.
22. Singh, R.; Khare, A. Fusion of multimodal medical images using Daubechies complex wavelet transform-A multiresolution approach. *Inform. Fusion* **2014**, *19*, 49–60.
23. Da Cunha, A.L.; Zhou, J.; Do, M.N. The nonsubsampling contourlet transform: Theory, design, and applications. *IEEE Trans. Image Process.* **2006**, *15*, 3089–3101.
24. Yang, Y.; Tong S.; Huang, S.Y.; Lin, P. Log-Gabor Energy Based Multimodal Medical Image Fusion in NSCT Domain, *Comput. Math. Meth. Med.* **2014**, *2014*, 835481.
25. Qu, X.B.; Yan, J.W.; Xiao, H.Z.; Zhu, Z.Q. Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampling contourlet transform domain. *Acta Autom. Sin.* **2008**, *34*, 1508–1514.
26. Kingsbury, N. The dual-tree complex wavelet transform: A new technique for shift-invariance and directional filters. In Proceeding of the IEEE Digital Signal Processing Workshop, Bryce Canyon, UT, USA, 9–12 August 1998; pp. 120–131.
27. Ioannidou, S.; Karathanassi, V. Investigation of the dual-tree complex and shift-invariant discrete wavelet transforms on Quickbird image fusion. *IEEE Geosci. Remote Sens. Lett.* **2007**, *4*, 166–170.
28. Li, H.; Chai, Y.; Yin, H.; Liu, G. Multi-focus image fusion and denoising scheme based on homogeneity similarity. *Opt. Commun.* **2012**, *285*, 91–100.
29. Yang, Y.; Zheng, W.; Huang, S.Y. Effective Multi-focus Image Fusion Based on HVS and BP Neural Network. *Sci. World J.* **2014**, *2014*, 281073.
30. Nayar, S.K.; Nakagawa, Y. Shape from focus. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 824–831.
31. Kingsbury, N. Image processing with complex wavelets. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **1999**, *357*, 2543–2560.
32. Kannan, K.; Perumal, A.S.; Arulmozhi, K. Optimal decomposition level of discrete, stationary and dual tree complex wavelet transform for pixel based fusion of multi-focused images. *Serb. J. Electron. Eng.* **2010**, *7*, 81–93.
33. Huang, W.; Jing, Z. Evaluation of focus measures in multi-focus image fusion. *Pattern Recognit. Lett.* **2007**, *28*, 493–500.
34. Bhatnagar, G.; Jonathan Wu, Q.M.; Liu, Z. Human visual system inspired multi-modal medical image fusion framework. *Expert Syst. Appl.* **2013**, *40*, 1708–1720.
35. Gerhard, H.E.; Wichmann, F.A.; Bethge, M. How sensitive is the human visual system to the local statistics of natural images. *PLoS Comput. Biol.* **2013**, *9*, 1–15.
36. Toet, A.; van Ruyven, L.J.; Valetton, J.M. Merging thermal and visual images by a contrast pyramid. *Opt. Eng.* **1989**, *28*, 789–792.

37. Redondo, R.; Šroubek, F.; Fischer, S.; Cristobal, G. Multi-focus image fusion using the log-Gabor transform and a multisize windows technique. *Inform. Fusion* **2009**, *10*, 163–171.
38. Li, H.; Manjunath, B.; Mitra, S. Multisensor image fusion using the wavelet transform. *Graph. Models Image Process.* **1995**, *57*, 235–245.
39. Piella, G.; Heijmans, H. A new quality metric for image fusion. In Proceeding of the 2003 International Conference on Image Processing, Amsterdam, The Netherlands, 14–17 September 2003; pp. 173–176.
40. Bhatnagar, G.; Jonathan Wu, Q.M.; Liu, Z. Directive contrast based multimodal medical image fusion in NSCT domain. *IEEE Trans. Multimed.* **2013**, *15*, 1014–1024.
41. Indhumadhi, N.; Padmavathi, G. Enhanced image fusion algorithm using laplacian pyramid and spatial frequency based wavelet algorithm. *Int. J. Comput. Sci. Eng.* **2011**, *1*, 298–303.
42. Wang, Z.; Bovik, A.; Sheikh H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612.
43. Li, S.; Kang, X.; Hu, J.; Yang, B. Image matting for fusion of multi-focus images in dynamic scenes. *Inform. Fusion* **2013**, *14*, 147–162.
44. Lu, H.; Zhang L.; Serikawa, S. Maximum local energy: An effective approach for multisensor image fusion in beyond wavelet transform domain. *Comput. Math. Appl.* **2012**, *64*, 996–1003.
45. Zhang, X.; Li, X.; Feng, Y. A new image fusion performance measure using Riesz transforms. *Opt. Int. J. Light Electron Opt.* **2014**, *125*, 1427–1433.
46. Yang, B.; Li, S. Pixel-level image fusion with simultaneous orthogonal matching pursuit. *Inform. Fusion* **2012**, *13*, 10–19.
47. Xydeas, C.S.; Petrović, V. Objective image fusion performance measure. *Electron. Lett.* **2000**, *36*, 308–309.
48. Zhang, B.; Zhang, C.T.; Yuanyuan, L.; Wu, J.S.; Liu, H. Multi-focus image fusion algorithm based on compound PCNN in Surfacelet domain. *Opt. Int. J. Light Electron Opt.* **2014**, *125*, 296–300.
49. Li, S.; Yang, B.; Hu, J. Performance comparison of different multi-resolution transforms for image fusion. *Inform. Fusion* **2011**, *12*, 74–84.
50. Zitova, B.; Flusser J. Image registration methods: A survey. *Image Vis. Comput.* **2003**, *21*, 977–1000.
51. Tian, J.; Chen, L. Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure. *Signal Process.* **2012**, *92*, 2137–2146.
52. Dataset of Standard 512×512 Grayscale Test Images. Available online: <http://decsai.ugr.es/cvg/CG/base.htm> (accessed on 26 November 2014).