

Article

Remote Sensing Image Fusion Based on Sparse Representation and Guided Filtering

Xiaole Ma ^{1,2}, Shaohai Hu ^{1,2,*}, Shuaiqi Liu ^{3,*}, Jing Fang ^{1,2} and Shuwen Xu ⁴

¹ Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China; maxiaole@bjtu.edu.cn (X.M.); fangjing@sdnu.edu.cn (J.F.)

² Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing Jiaotong University, Beijing 100044, China

³ College of Electronic and Information Engineering, Hebei University, Baoding 071002, China

⁴ Research Institute of TV and Electro-Acoustics, Beijing 100015, China; xu_sw@263.com

* Correspondence: shhu@bjtu.edu.cn (S.H.); shdkj-1918@163.com (S.L.); Tel.: +86-010-5168-8646 (S.H.)

Received: 18 January 2019; Accepted: 5 March 2019; Published: 8 March 2019



Abstract: In this paper, a remote sensing image fusion method is presented since sparse representation (SR) has been widely used in image processing, especially for image fusion. Firstly, we used source images to learn the adaptive dictionary, and sparse coefficients were obtained by sparsely coding the source images with the adaptive dictionary. Then, with the help of improved hyperbolic tangent function (\tanh) and $l_0 - \max$, we fused these sparse coefficients together. The initial fused image can be obtained by the image fusion method based on SR. To take full advantage of the spatial information of the source images, the fused image based on the spatial domain (SF) was obtained at the same time. Lastly, the final fused image could be reconstructed by guided filtering of the fused image based on SR and SF. Experimental results show that the proposed method outperforms some state-of-the-art methods on visual and quantitative evaluations.

Keywords: image fusion; sparse representation; hyperbolic tangent function; guided filter

1. Introduction

By making full use of the complementary information of the remote sensing images and other source images of the same scene, image fusion can be defined as the processing method for integrating this information together to obtain a fused image, which is more suitable for the human visual system [1]. Through image fusion, we can obtain one composite image, which contains more special features, and can provide more useful information. As a powerful tool for image processing, image fusion covers broad range of areas [2,3], such as computer vision, remote sensing, and so on [4].

Diversiform remote sensing image fusion methods have been proposed in recent years, which can be divided into three categories: Pixel-level fusion, feature-level fusion, and decision-level fusion [5]. Feature-level fusion mainly deals with the features of the source images, while decision-level fusion makes the decision after judging the information of the source images. Compared with the aforementioned levels, pixel-level fusion can serve more useful original information, although it has some shortcomings such as being time consuming. Despite complex computation, most researchers conduct image fusion based on pixel-fusion [6,7], such as the image fusion method based on the spatial domain, and the image fusion method based on the transform domain.

Recently, mainstream methods of image fusion have been based on the multi-scale transforms [8,9], such as image fusion based on object region detection and non-subsampled contourlet transform [10] and image fusion based on the complex shearlet transform with guided filtering [11]. For the image fusion method based on multi-scale transforms, the source images are represented by the fixed

orthogonal basis functions, and the fused image can be obtained by fusing the coefficients of different sub-bands together in the transform domain. Although the multi-scale geometric transform can represent most features of the image, which are always complex and diverse, there are some features that cannot be represented sparsely. Thus, it cannot represent all the useful features accurately by limited fixed transforms.

The rapidly developing sparse representation methods can not only more sparsely represent the source images, but also effectively extract the potential information hidden in the source images and produce more accurate fused images, compared with the multi-scale transforms [12–14]. Based on these findings, scholars apply sparse representation to image fusion. Mitianoudis [13] and Yang [14] laid the foundation for image fusion based on SR. Yu [15] applied sparse representation with K-singular value decomposition (K-SVD) to medical image fusion, Yang [16] applied sparse representation and multi-scale decomposition to remote sensing image fusion, and Yin [17] applied a novel sparse-representation-based method to multi-focus image fusion.

In the sparse model, the generation of the dictionary and sparse coding is crucial for the image fusion [18]. Although the fixed over-complete dictionary can realize good fusion results, it usually takes a lot of time to obtain the sparse coefficients, resulting in inefficiency. In this paper, adaptive dictionary learning [19,20] is adopted for its simplicity and convenience. Motivated by the multi-strategy fusion rule based sigmoid function in reference [21] and the characteristics of the hyperbolic tangent function, the multifarious rule based on \tanh and $l_0 - \max$ is proposed to fuse the sparse coefficients. Finally, by sparse reconstruction, the fused image based on SR is obtained, which is more suitable for the human visual system and subsequent image processing. However, there is more detailed information in the remote sensing images than other kinds of images. When performing image fusion by the method based on SR, it may lose some discontinuous edge features [22], which leads to the loss of some useful information of fused images. In addition, image fusion based on SR also ignores the spatial information, which can reflect the image structure more directly and accurately. As a result, we can simultaneously fuse the source remote sensing images by the method based on SR and SF, and obtain two different fused images, namely the fused image based on SR and the fused image based on SF. In this paper the two fused images above are processed by a guided filter to obtain the final image since a guided filter has good performance with edge preserving [23]. The main contributions of this paper can be summarized as follows.

(1) The learning of the dictionary is vital for sparse representation, and the adaptive dictionary of each source image can be generated in every step of dictionary learning. The final dictionary can be obtained by gathering together the sub-dictionaries. As a result, this work enriches the dictionary and can make the coefficients more sparsely.

(2) As is well known, the information in each source image is complementary and redundant. When fusing images to obtain the fused image, we need to consider the relationship between different source images. For the redundant information of the source images, the weighted rule would be better; on the other hand, the choose-max rule would result in a fused image with less block effect. Based on the above considerations and the characteristics of hyperbolic tangent function, the fusion rule based on \tanh and $l_0 - \max$ is proposed in this paper.

(3) The image fusion methods based on SR can obtain the fused image by sparsely coding the source images and fusing the sparse coefficients. However, it ignores the correlation of the image information in the spatial domain and loses some important detailed information of the source images. In this paper, we adopt the image fusion method based on SF and filter the fused image based on SR and SF by the guided filter. By making full use of the information in the spatial and the sparse representation domain, the fused image can reserve more information of the source images.

The rest of this paper is organized as follows. The theory of the sparse representation is introduced briefly in Section 2. Adaptive dictionary learning is presented in Section 2.1, and the proposed fusion rule is given in Section 2.2. The flow chart of the remote sensing image fusion method based on SR

and guided filtering is drawn in Section 3. In Section 4, some experiments and result analysis are done. Finally, conclusions are made in Section 5.

2. Sparse Representation

SR has been widely used in image processing, as one of the most powerful tools to represent signals especially image signals, such as image de-noising [24], image coding [25], object tracking [26], and image super resolution [27], etc.

In the SR model, the image is sparse and can be represented, or approximately represented, by one linear combination of a few atoms from the dictionary [14,28,29]. Suppose that the source image is \mathbf{I} , and the over-complete dictionary is $\mathbf{D} \in \mathbf{R}^{M \times k}$, the sparse representation model can be formulated as follows [16,22].

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \|\alpha\|_0 \text{ s.t. } \|\mathbf{I} - \mathbf{D}\alpha\|_2^2 \leq \varepsilon \quad (1)$$

where α denotes the sparse coefficients of the image and $\|\bullet\|_0$ denotes the l_0 -norm, respectively, which indicate the number of non-zero elements in the corresponding vector. Usually, $\|\alpha\|_0 \leq L \ll M$, and L is the maximal sparsity. ε indicates the limiting error.

For the image fusion method based on SR, there are two important steps: dictionary learning and sparse coding. Dictionary learning will be discussed in detail in Section 2.1. When performing sparse coding by orthogonal matching pursuit (OMP) [30] in this paper, Equation (1) can be replaced by Equation (2).

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \|\mathbf{I} - \mathbf{D}\alpha\|_2^2 + \mu \|\alpha\|_0 \quad (2)$$

where, μ is the penalty factor.

2.1. Adaptive Dictionary Learning

When fuse the source images by the methods based on SR, dictionary learning is one of important processes. To make full use of the image information, we generate a dictionary based on the source images themselves. And the generation of the adaptive dictionary can be changed into the iteration of the dictionary atoms. By the iteration process, it can realize dictionary learning with the over-complete dictionary based on the source images.

Since dictionary learning is more efficient for small image blocks, if the dictionary updating step is processed by the original source images directly, the sparsity would be seriously influenced. Thus optimal sparse coefficients cannot be obtained [29]. In order to solve this problem, we divided the source images into image blocks, which can replace the dictionary atoms for better dictionary learning. The improved dictionary generation method can not only obtain the optimal sparse representation but also accelerates the efficiency and accuracy of the SR algorithm. However, since we perform dictionary learning on the image block rather than the whole image, the reshaped vector on every atom is not very large and it reduces the computation cost.

K-singular value decomposition (K-SVD) [31] is one of the most used image fusion methods based on SR. Here, we adopt the K-SVD model on the sub-dictionary of the image block by the following iteration process:

$$\hat{\mathbf{D}}_{ij}^M = \operatorname{argmin}_{\mathbf{D}_{ij}^M, \alpha_{ij}^M} \sum_{ij} \left\| \mathbf{P}_{ij}^M - \hat{\mathbf{D}}_{ij}^M \alpha_{ij}^M \right\|_2^2 + \mu_{ij}^M \left\| \alpha_{ij}^M \right\|_0 \quad (3)$$

where ij denotes the position (i, j) in the image \mathbf{M} and \mathbf{P}_{ij}^M denotes the image block with the center pixel at the corresponding position (i, j) .

Then, we can obtain the adaptive dictionary of the source image \mathbf{M} shown in Equation (4).

$$\hat{\mathbf{D}}^M = \left\{ \hat{\mathbf{D}}_{ij}^M \right\} \quad (4)$$

At last, we can gather all the dictionaries of different source images by Equation (5), where n denotes the total number of the source images.

$$D = [D_1, D_2, \dots, D_n] \tag{5}$$

2.2. Fusion Rule Based on tanh and $l_0 - max$

As we all know, the fusion rules are vital for the final fusion results and for the sparse coefficients. In most cases, we always take the $l_1 - max$ rule to obtain the fused block vectors [7], where l_1 means the sum of absolute values of the vector elements. However, when there are noises or some unwanted pixels in the flat area of the source images, the unwanted portion will be included and lead to incorrect fusion [17]. The information in the source images is redundant and complementary for the image fusion shown in Figure 1. Figure 1a,b are one set of medical images, which contain complementary information, while Figure 1c,d are one set of multi-focus images, which contain redundant information. When the relationship of the image information is redundant, the weighted fusion rule is chosen, and the max fusion rule should be chosen for the complementary sparse coefficients [21]. The fused information would be lost and incomplete if the complementary information is multiplicative by the weighted factor. Based on these considerations, we proposed one new sparse coefficient fusion rule based on tanh and $l_0 - max$. We can obtain the fused coefficients by calculating $l_0 - norm$ and the weighting factor based on tanh.

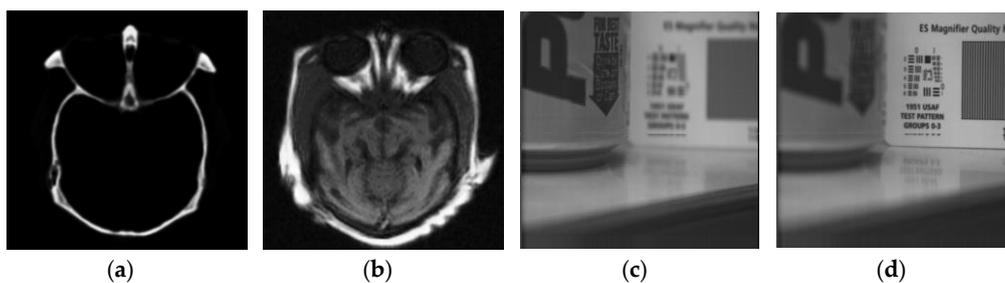


Figure 1. Images with different information: (a) CT image; (b) MRI image; (c) left-focus image; (d) right-focus image.

The hyperbolic tangent function is one of the hyperbolic functions, and derives from hyperbolic sine function and hyperbolic cosine function [32]. It can be calculated as follows:

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{6}$$

where the hyperbolic sine function and hyperbolic cosine function can be defined as Equations (7) and (8), respectively.

$$\sinh(x) = \frac{e^x - e^{-x}}{2} \tag{7}$$

$$\cosh(x) = \frac{e^x + e^{-x}}{2} \tag{8}$$

Figure 2 shows the different hyperbolic functions. From Figure 2a,b we can see that tanh is symmetrical around the origin point. As x increases, the difference between the value of the hyperbolic sine function and the hyperbolic cosine function narrows, and the value of $\tanh(x)$ changes from -1 to 1 . When there is redundant information in different source images and the weighted fusion rule is chosen, it would be better if different degrees of redundancy corresponded to different weights. Based on the aforementioned factors, we improve tanh shown in Figure 2c to obtain the weighted factor for fusing the sparse coefficients, and the corresponding equation is listed as Equation (9).

$$w_{ij} = \frac{1}{2} * [\tanh(a * (s_{ij} - 1)) + 1] \tag{9}$$

where s_{ij} denotes the sparse coefficient at the position (i, j) and w_{ij} denotes the corresponding weighted factor when adopting the fusion rule based on tanh. a denotes the sensitivity between the sparse coefficient and the weighted factor. According to the experiments on different image groups and values of the parameter a , we found that 3 is the best.

Compared with Figure 2b, the curve has a steeper slope in Figure 2c when s_{ij} is closer to 1, which means that the weighted factor is very sensitive to the sparse coefficients. When s_{ij} is near 0 or too large, the weighted factor w_{ij} is near 0 or 1, which means that the source images have complementary information, where the fusion rule based on $l_0 - \max$ is adopted.

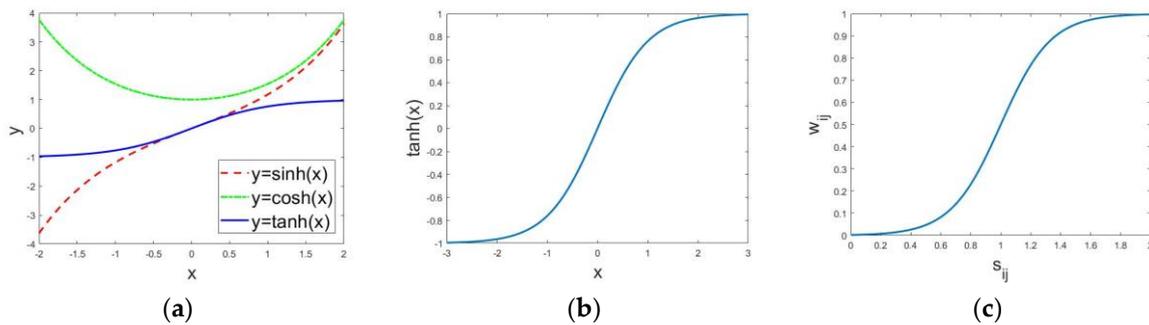


Figure 2. Different hyperbolic functions: (a) The hyperbolic functions; (b) tanh; (c) improved tanh.

Finally, we can obtain the fused sparse coefficients $\alpha_{F_{ij}}$ at the position (i, j) by Equation (10).

$$\alpha_{F_{ij}} = \begin{cases} w_{ij} * \alpha_{A_{ij}} + (1 - w_{ij}) * \alpha_{B_{ij}} & \text{if } \alpha_{A_{ij}} \& \alpha_{B_{ij}} \neq 0 \\ \max(\alpha_{A_{ij}}, \alpha_{B_{ij}}) & \text{else} \end{cases} \tag{10}$$

where $\alpha_{A_{ij}}$ and $\alpha_{B_{ij}}$ denotes the sparse coefficients in the source image A and B . $\alpha_{A_{ij}} \& \alpha_{B_{ij}} \neq 0$ means that both $\alpha_{A_{ij}}$ and $\alpha_{B_{ij}}$ are not zero. And w_{ij} can be calculated by Equation (9), where $s_{ij} = \alpha_{A_{ij}}$.

3. The Proposed Image Fusion Method

An interesting remote sensing fusion method based on sparse representation and guided filtering is presented in this paper, and the framework can be seen in Figure 3. It mainly includes three image processing elements: image fusion based on SR, image fusion based on SF, and guided filtering. The adaptive dictionary was learned by the source images themselves, and the fused sparse coefficients was obtained by the dictionary and proposed fusion rule. Then, the fused image based on SR was reconstructed by the obtained adaptive dictionary and fused sparse coefficients. At the same time, we fused the source images obtained by the image fusion method based on SF such as the gradient fusion. As shown in Figure 3, the guided filter was finally adopted to guide the fused images based on SR and SF. Since there was more detailed information in the fused image based on SF, in the last part of the proposed method, we made the fused image based on SF as the guidance image, and the other fused image served as the input image.

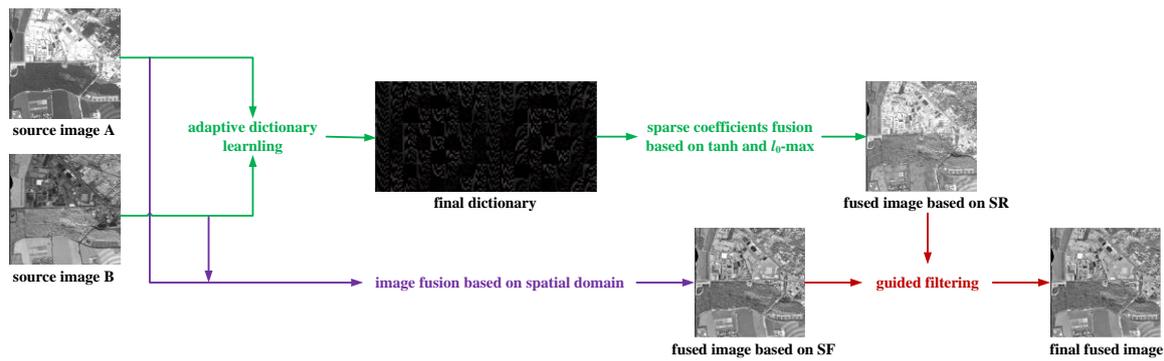


Figure 3. The framework of the proposed method.

4. The Experiments and Result Analysis

To testify the superiority of the proposed method, a series of experiments on the remote sensing and other source images were conducted in this section. We compared our method with some classical image fusion methods, including the multi-scale weighted gradient-based fusion (MWGF) [33], the image fusion with guided filtering (GuF) [34], image fusion based on Laplace transformation (LP) [35], multiresolution DCT decomposition for image fusion (DCT) [36], the image fusion algorithm in the nonsubsampling contourlet transform domain (NSCT) [37], image fusion with the joint sparsity model (SR) [1], and image fusion based on multi-scale and sparse representation (MST-SR) [8]. With adaptive dictionary learning, the size of every image block was 8×8 . Experiments conducted on dictionary learning of different source images showed that when the number of iterations was 3, it guaranteed the convergence and stability of the coefficients. In addition, the experiments in this paper were carried out by Matlab code on an Intel Core i5-2450M (Acer, Beijing, China) 2.50 GHz with 6 GB RAM.

4.1. Objective Valuation Indexes

To evaluate the experimental results more objectively, we adopted some objective valuation indexes [37] to evaluate the fused images by different image fusion methods, which included entropy (EN), spatial frequency (SF), $Q^{AB/F}$, and structural similarity (SSIM).

When we want to balance the wealth of information in one image, EN is a wonderful choice. The larger the value of EN in the fused image is, the more information does the image contain, which means better image fusion result. And EN can be summarized as Equation (11).

$$EN = - \sum_{i=0}^{L-1} p_i \times \log_2 p_i \tag{11}$$

where L denotes the total number of pixels included in the image and p_i denotes the probability distribution of pixels for each gray level.

SF can detect the total active of the fused image in the spatial domain and it denotes the expression ability of one image for minor detail contrast. The equation of SF is shown as follows:

$$SF(i, j) = \sqrt{(RF)^2 + (CF)^2} \tag{12}$$

where RF stands for the horizontal frequency while CF stands for the vertical frequency. And they can be calculated by Equations (13) and (14).

$$RF = \sqrt{\frac{1}{M \times N} \sum_{x=1}^M \sum_{y=2}^N [F(x, y) - F(x, y - 1)]^2} \tag{13}$$

$$CF = \sqrt{\frac{1}{M \times N} \sum_{x=2}^M \sum_{y=1}^N [F(x, y) - F(x - 1, y)]^2} \tag{14}$$

where F denotes the fused image with the size of $M \times N$.

While $Q^{AB/F}$ can balance how much the edge information of the source images A and B does the fused image contain by Sobel operator. It can be defined as Equation (15).

$$Q^{AB/F} = \frac{\sum_{\forall n,m} (Q_{n,m}^{AF} w_{n,m}^A + Q_{n,m}^{BF} w_{n,m}^B)}{\sum_{\forall n,m} (w_{n,m}^A + w_{n,m}^B)} \tag{15}$$

where $w_{n,m}^A = [g_A(n, m)]^L$, $w_{n,m}^B = [g_B(n, m)]^L$. Normally, L is one constant and the value is 1. Taking the source image A as an example, edge information retention value $Q_{n,m}^{AF}$ and edge strength information $g_A(n, m)$ can be calculated by Equations (16) and (17).

$$Q_{n,m}^{AF} = \Gamma_g \Gamma_\alpha [1 + e^{K_g(G_{n,m}^{AF} - \sigma_g)}]^{-1} [1 + e^{K_\alpha(A_{n,m}^{AF} - \sigma_\alpha)}]^{-1} \tag{16}$$

$$g_A(n, m) = \sqrt{s_A^x(n, m)^2 + s_A^y(n, m)^2} \tag{17}$$

where $\Gamma_g, K_g, \sigma_g, \Gamma_\alpha, K_\alpha, \sigma_\alpha$ are constant and they affect the sigmoid function together. $(G_{n,m}^{AF}, A_{n,m}^{AF}) = [(\frac{g_{n,m}^F}{g_{n,m}^A})^M, 1 - \frac{|\alpha_A(n,m) - \alpha_F(n,m)|}{\pi/2}]$ and $M = \begin{cases} -1 & \text{if } g_A(n, m) \leq g_F(n, m) \\ 1 & \text{otherwise} \end{cases}$. $\alpha_A(n, m) = \tan^{-1}[\frac{s_A^y(n, m)}{s_A^x(n, m)}]$ and $s_A^x(n, m), s_A^y(n, m)$ denote the convolution results of Sobel model with the center pixel at the position (n, m) in the horizontal and vertical directions with the source image A .

SSIM is the structural similarity between the source images and the fused image. And the equation of SSIM is as follows:

$$SSIM(A, B, F) = \frac{1}{2} (SSIM(A, F) + SSIM(B, F)) \tag{18}$$

where $SSIM(A, F)$ denotes SSIM of the source image A and fused image F , and so is $SSIM(B, F)$. More detail of their calculation is shown in Equations (19) and (20).

$$SSIM(A, F) = \frac{(2\mu_A\mu_F + C_1) \cdot (2\sigma_{AF} + C_2)}{(\mu_A^2 + \mu_F^2 + C_1)(\sigma_A^2 + \sigma_F^2 + C_2)} \tag{19}$$

$$SSIM(B, F) = \frac{(2\mu_B\mu_F + C_1) \cdot (2\sigma_{BF} + C_2)}{(\mu_B^2 + \mu_F^2 + C_1)(\sigma_B^2 + \sigma_F^2 + C_2)} \tag{20}$$

where μ_A, μ_B, μ_F denote the average of pixels of the image A, B and F , respectively. $\sigma_A^2, \sigma_B^2, \sigma_F^2$ denote the variance and σ_{AF}, σ_{BF} denote the joint variance. For the convenience of calculation, we make $C_1 = C_2 = 0$.

The larger all the indexes above are, the better the fused image is. What's more, when obtaining the adaptive dictionary by the proposed method, there is slight deviation of the final results. We adopt the mean of the evaluation values in three times.

4.2. Large Scale Image Fusion of Optical and Radar Images

Figure 4 shows one SAR image of the harbor around Oslo with a size of 1131×942 and the registered optical image on a large scale for the whole scenery [38]. Due to the use of the high-resolution digital elevation model (DEM), the optical image fits onto the signatures of the buildings very well. Figure 4c,d are partially enlarged details of Figure 4a,b at the position of the red rectangle in Figure 4a. Figures 5 and 6 are the corresponding fused images obtained by the methods above, and partially enlarged views.

Since the optical image in Figure 4b is colorful, we processed the image fusion in the RGB dimension separately. Although the visual effect of Figure 6a is better, there was a greater color contrast in Figure 5a, which introduced some incorrect information in the left corner. In Figure 6, the partially enlarged detail images of Figure 5d by DCT and Figure 5f by SR are very blurred which seriously affects the fused images. Compared with Figure 6g, the left corner in Figure 6h contains more information of the remote sensing image in Figure 5c, which indicates that the fused image by our method is better.

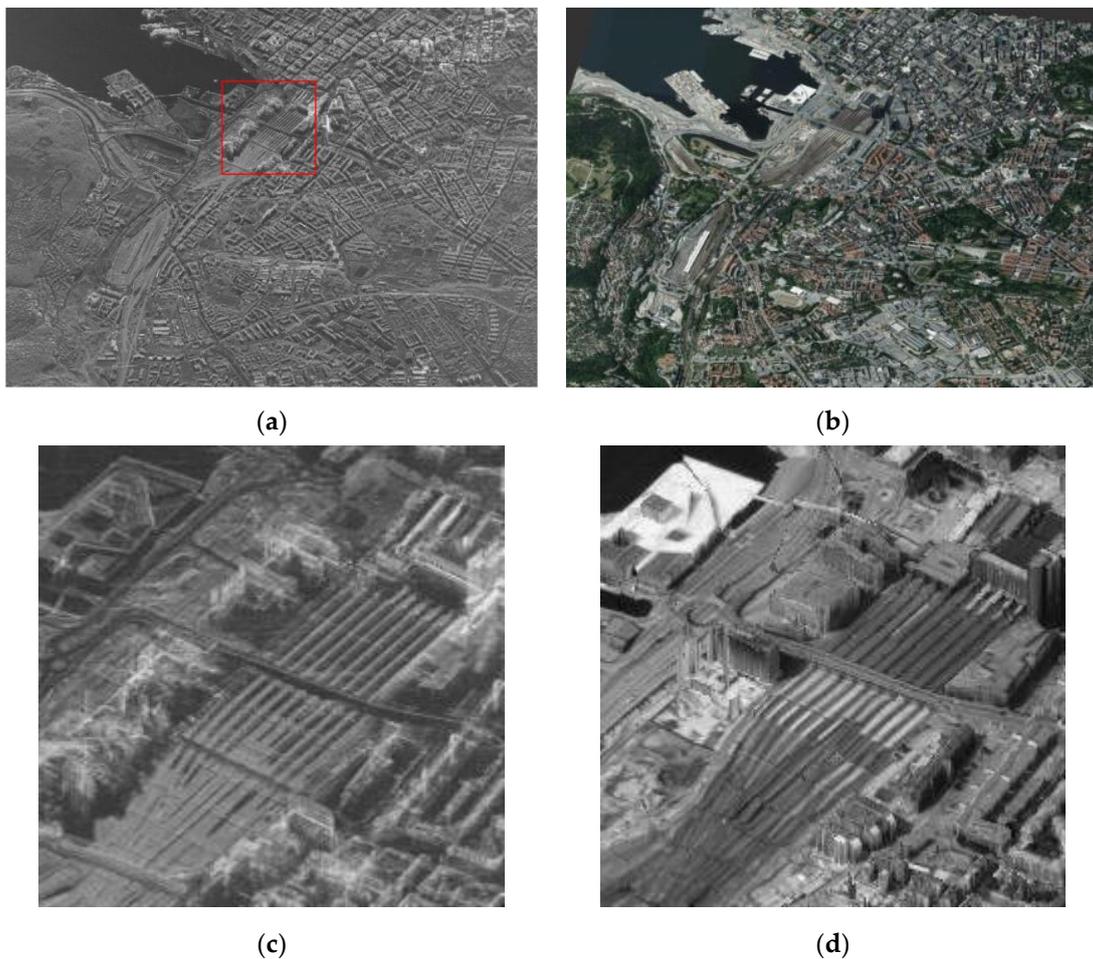


Figure 4. The large scale images: (a) TerraSAR-X staring spotlight image of Oslo; (b) optical image of Oslo; (c) part of (a); (d) part of (b).

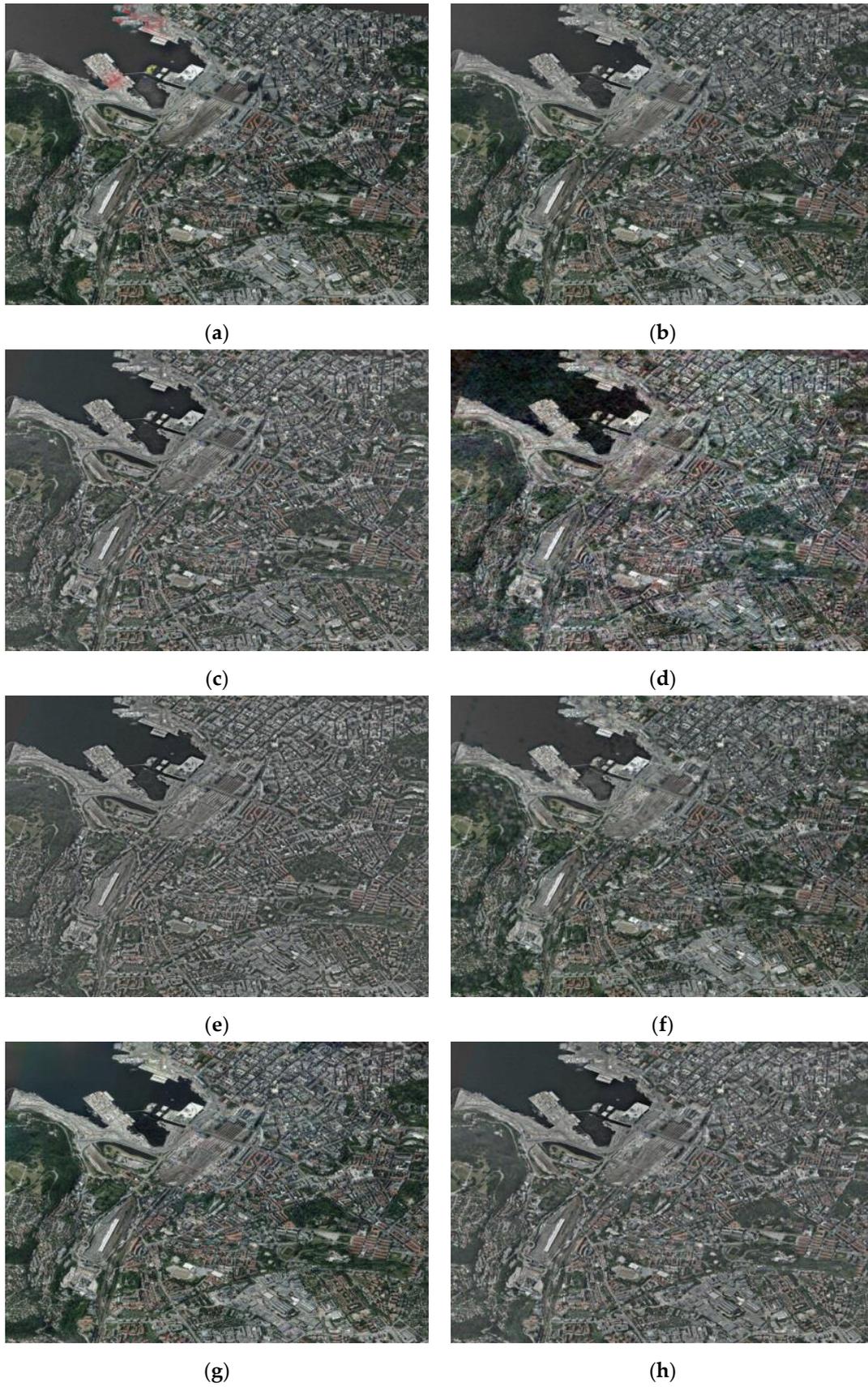


Figure 5. The fused images of Figure 4: (a) MWGF; (b) GuF; (c) LP; (d) DCT; (e) NSCT; (f) SR; (g) MST-SR; (h) the proposed.



(a)



(b)



(c)



(d)



(e)



(f)

Figure 6. Cont.

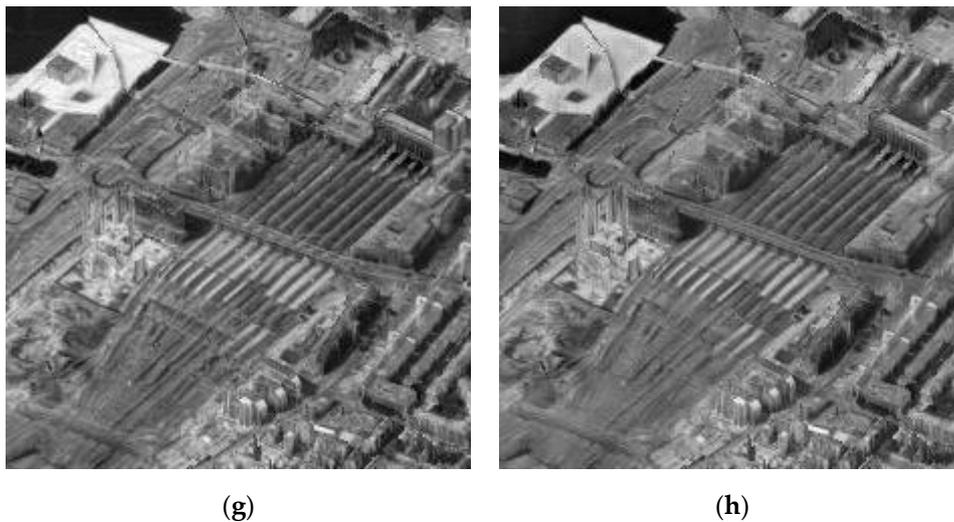


Figure 6. Part of Figure 5: (a) MWGF; (b) GuF; (c) LP; (d) DCT; (e) NSCT; (f) SR; (g) MST-SR; (h) the proposed.

Table 1 shows the corresponding index values of the fused images in Figure 5 and the best values are in bold. From Table 1, we can see the image fusion methods based on the spatial domain such as MWGF and GuF have big ability to preserve the spatial frequency, and MWGF has a better value of $Q^{AB/F}$. However, the visual result of MWGF is the worst. $Q^{AB/F}$ of the proposed method ranks third among the compared methods, which is worse than the methods based on the spatial domain. This explains why we adopt the image fusion method based on the spatial domain and guide it with the fused image-based SR in this paper. The values of EN, SF, and SSIM of the fused image obtained by the proposed method are better, which indicates that the proposed method has a better ability to fuse the remote sensing image.

Table 1. The evaluation index values of fused images in Figure 5.

Methods Indexes	MWGF	GuF	LP	DCT	NSCT	SR	MST-SR	The Proposed
EN	7.3411	7.2547	7.2606	7.2158	7.1843	7.2711	7.3742	7.5879
SF	31.1834	30.1662	32.4682	30.1223	31.5884	30.2674	32.4878	33.9360
$Q^{AB/F}$	0.6313	0.6111	0.5753	0.3754	0.5453	0.5705	0.5767	0.5794
SSIM	0.5798	0.5981	0.5990	0.5410	0.5910	0.5829	0.5966	0.6246

4.3. Image Fusion of Remote Sensing Images

To testify the effectiveness and universality of the proposed method, the classical image pairs shared by Durga Prasad Bavirisetti (<https://sites.google.com/view/durgaprasadbavirisetti/datasets>) are used to test the performance of the fused algorithms. The dataset contains rich remote sensing images and we conduct our experiments on different kinds of image pairs, which contain the forest with greater high-frequency information, rivers with low-frequency information, and so on. To save space, we only show the four groups and the results analysis. The four groups include rich information with different types and are representative in the dataset, shown in Figure 7. Figures 8–11 are the fused images obtained by the diverse compared methods of the different source images.

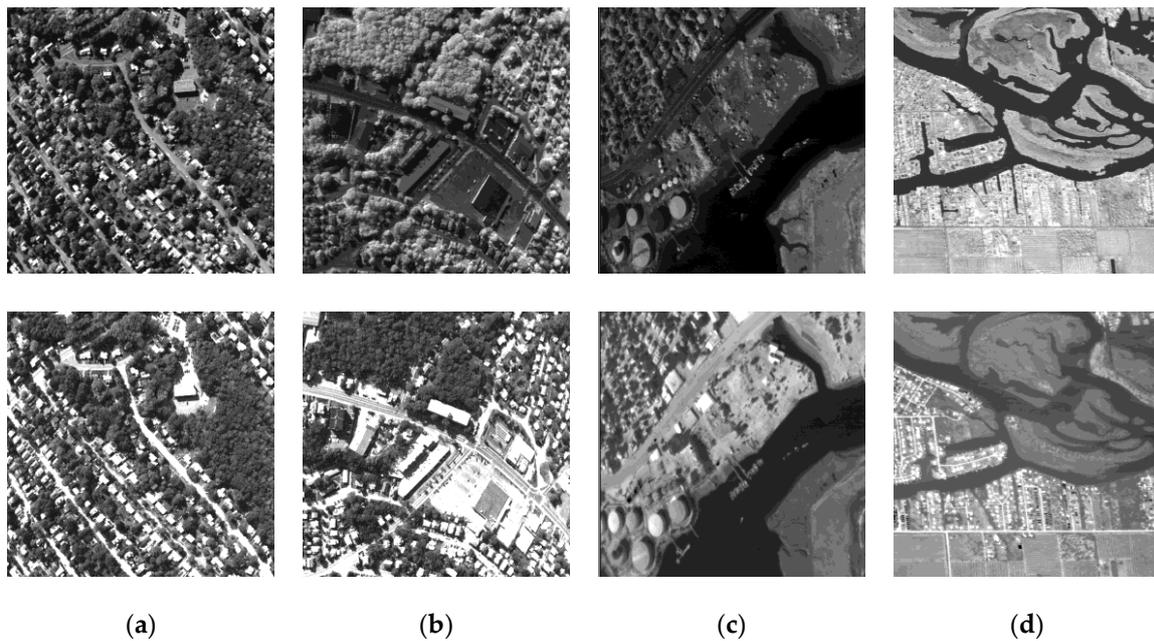


Figure 7. The source remote sensing images: (a) Group 1; (b) Group 2; (c) Group 3; (d) Group 4.

Figure 7a,b are forests and rural areas with fewer buildings, of which the top view is sharper and has richer detailed information. From Figure 8, we can see that the trees in Figure 8a–e is more darker than Figure 8f–g and has less information in the second line of Group 1 in Figure 7, which indicates that the image fusion based on SR is more powerful than the methods based on the spatial domain and transform domain. And there are some artificial textures in the roof of Figure 8f. Above all, the fused image of Figure 8h obtained by the proposed has better visual effect.

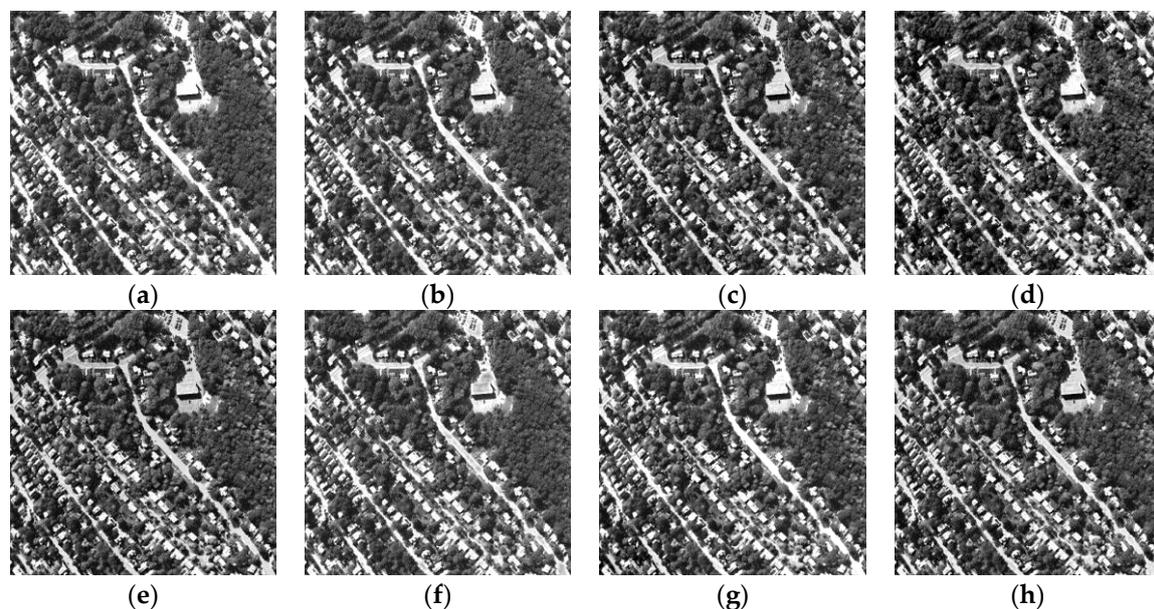


Figure 8. The fused images of Group 1: (a) MWGF; (b) GuF; (c) LP; (d) DCT; (e) NSCT; (f) SR; (g) MST-SR; (h) the proposed.

Compared with Group 2, there are some suburbs next to the forests in Group 1. And the contrast in Figure 9c,e,h looks better. From the roofs in the fused images shown in Figure 9, the flat area and edges in Figure 9h obtained by the proposed method look more comfortable and are more suitable

for we human visual system, which indicates that the proposed method has powerful ability to fuse remote sensing images.

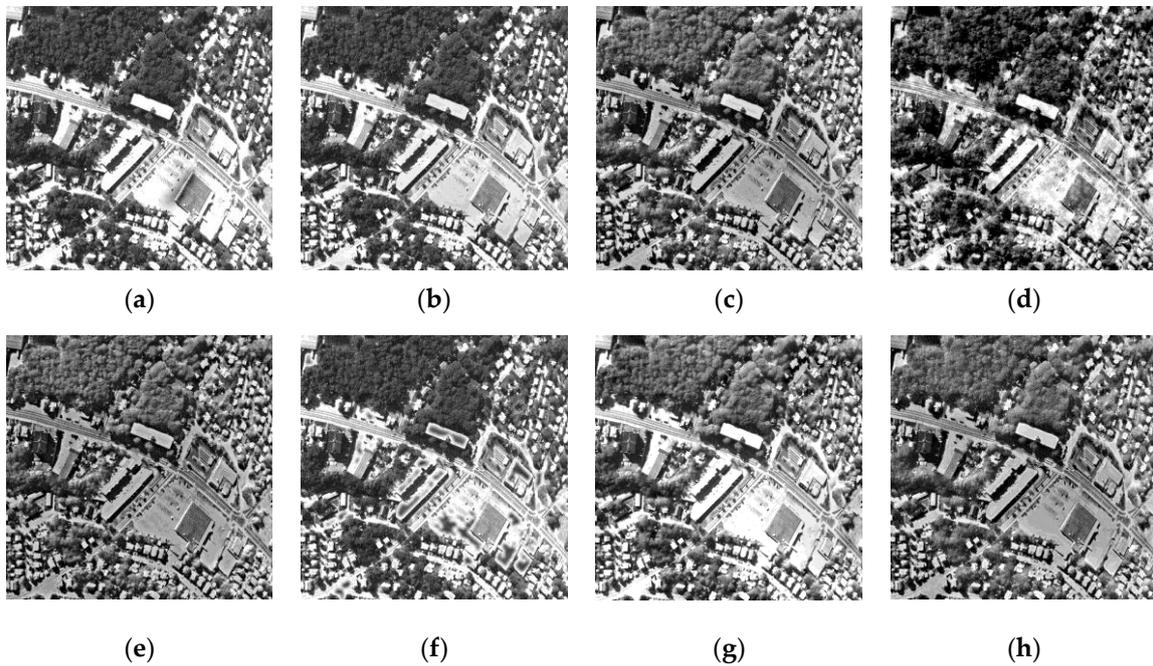


Figure 9. The fused images of Group 2: (a) MWGF; (b) GuF; (c) LP; (d) DCT; (e) NSCT; (f) SR; (g) MST-SR; (h) the proposed.

There are some river and coastal area in Group 3. And by comparing the fused images in Figure 10, the center in Figure 10a looks very bad and some areas in Figure 10g are too bright, which have the strong exposure. From these figures, we can see that there is less artificial texture in Figure 10h, which means the fused image obtained by the proposed method have better visual result.

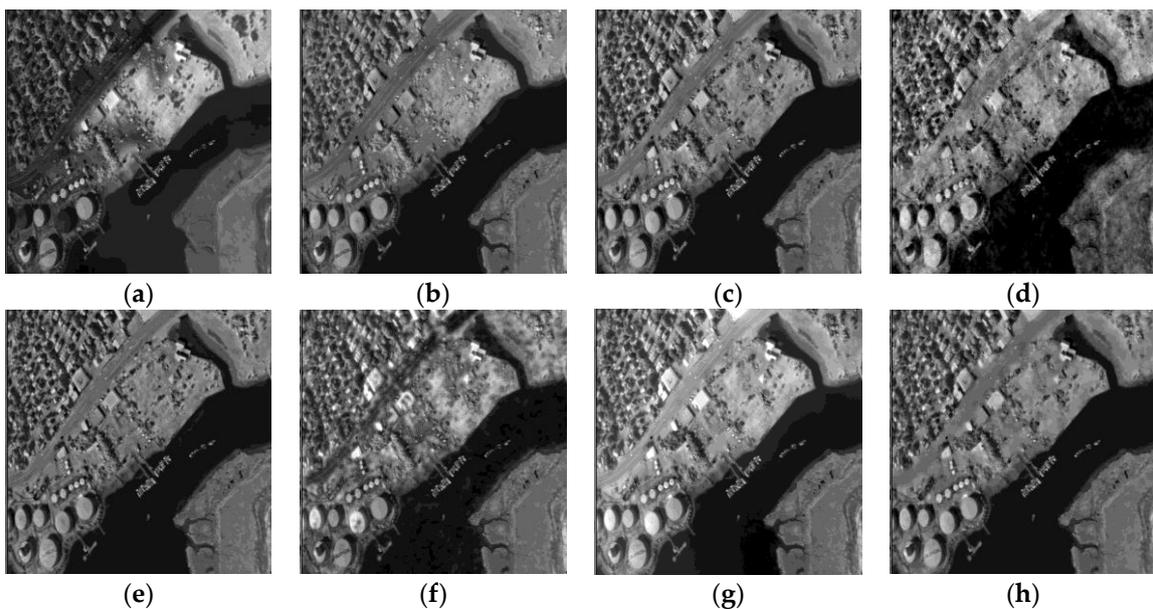


Figure 10. The fused images of Group 3: (a) MWGF; (b) GuF; (c) LP; (d) DCT; (e) NSCT; (f) SR; (g) MST-SR; (h) the proposed.

Group 4 is one set of classic multi-sensor image pair, which can be found in most of papers about remote sensing image fusion. By comparing the bottoms of the fused images in Figure 10, we can find that there are some unwanted spots and artificial texture in Figure 10d, and the small round black area is very blurred or even lost in Figure 10a–c,f. Since the rivers display as black areas like wide line or curve in the fused images, it has worst visual effect in Figure 10f, of which the detailed information has been lost. As a result, the fused image in Figure 10h looks more comfortable for our eyes and the proposed method has better ability to fuse remote sensing images.

Similarly, we use the aforementioned objection evaluation indexes to value the fused images in Figures 8–11 and the objective values are shown in Tables 2–5. As shown in Tables 2 and 3, the algorithm proposed in this paper has obtained the best results for Group 1 and Group 2 in Figure 7. This fully demonstrates that the proposed method has a better ability to perform remote sensing image fusion. Compared with Group 1 and Group 2, there is more low frequency information and less detail and edges in Group 3 and Group 4. However, the proposed method is more suitable for the images with great detail. As a result, the SSIM of the fused image by NSCT is better than others in Table 4, but other values of the proposed method are satisfactory. All these values demonstrate that the proposed method performs better in terms of remote sensing image fusion.

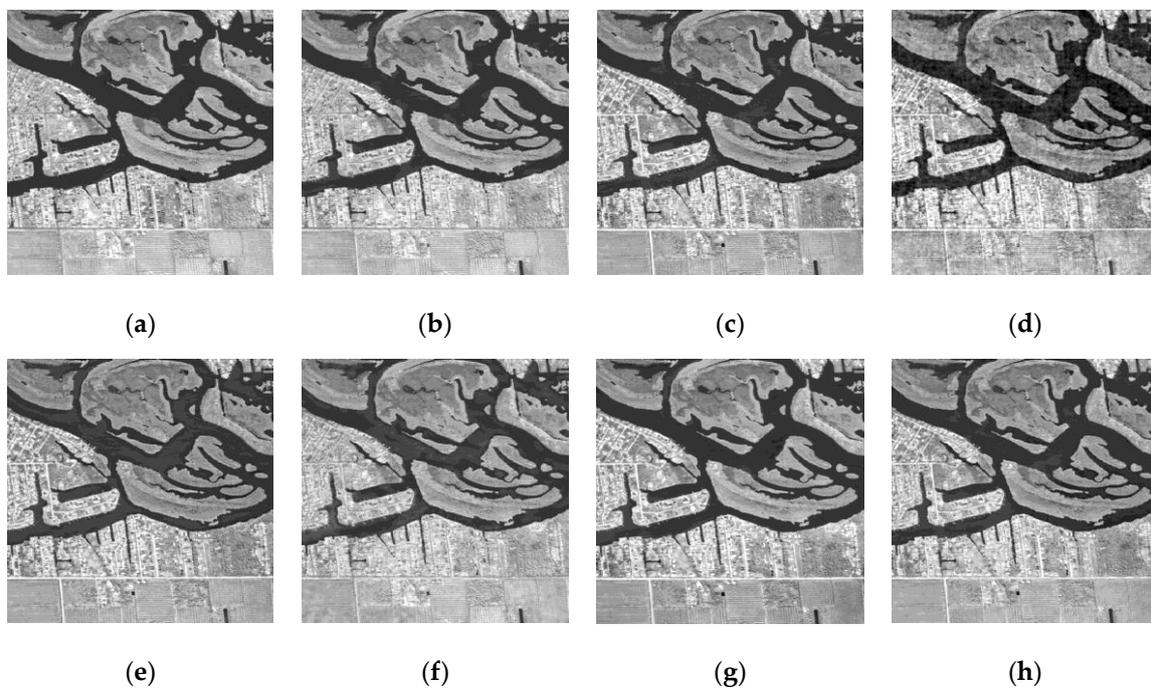


Figure 11. The fused images of Group 4: (a) MWGF; (b) GuF; (c) LP; (d) DCT; (e) NSCT; (f) SR; (g) MST-SR; (h) the proposed.

Table 2. The evaluation index values of fused images in Figure 8.

Methods Indexes	MWGF	GuF	LP	DCT	NSCT	SR	MST-SR	The Proposed
EN	7.1741	7.4931	7.6687	7.6006	7.6935	7.4351	7.5925	7.7076
SF	54.1839	53.6728	55.0006	55.2435	54.6567	53.9868	54.8697	56.3409
$Q^{AB/F}$	0.7030	0.7104	0.7110	0.6105	0.7143	0.7089	0.7073	0.7174
SSIM	0.7933	0.7976	0.8077	0.7759	0.8170	0.7916	0.8060	0.8235

Table 3. The evaluation index values of fused images in Figure 9.

Methods Indexes	MWGF	GuF	LP	DCT	NSCT	SR	MST-SR	The Proposed
EN	6.9778	7.6668	7.8936	7.5537	7.8154	7.3634	7.6580	7.8946
SF	53.6844	53.3379	53.8149	53.2507	53.3216	53.1757	54.1247	54.1590
Q ^{AB/F}	0.6668	0.6687	0.6310	0.4783	0.6190	0.6379	0.6346	0.6710
SSIM	0.6283	0.6314	0.6340	0.5433	0.6172	0.5969	0.6429	0.6438

Table 4. The evaluation index values of fused images in Figure 10.

Methods Indexes	MWGF	GuF	LP	DCT	NSCT	SR	MST-SR	The Proposed
EN	7.2073	7.0657	6.9885	6.9246	6.9359	7.2902	7.3556	7.5855
SF	15.4453	16.4780	18.7239	19.5068	18.3164	18.3152	18.9748	19.8665
Q ^{AB/F}	0.5491	0.5640	0.5574	0.3619	0.5411	0.4741	0.5658	0.5774
SSIM	0.6641	0.6954	0.6898	0.4789	0.6958	0.6168	0.6797	0.6826

Table 5. The evaluation index values of fused images in Figure 11.

Methods Indexes	MWGF	GuF	LP	DCT	NSCT	SR	MST-SR	The Proposed
EN	6.0932	7.1620	7.3451	7.3081	7.3317	7.1194	7.0635	7.6861
SF	27.8713	27.0044	29.2465	28.9847	28.2961	25.7550	29.2601	29.6650
Q ^{AB/F}	0.6473	0.6318	0.5857	0.4143	0.5653	0.5586	0.5886	0.5968
SSIM	0.6528	0.6647	0.6736	0.5548	0.6871	0.6590	0.6733	0.6930

5. Conclusions

Due to the good performance of sparse representation and the rich information in the spatial domain, this paper presents one new remote sensing image fusion method based on sparse representation and guided filtering. It also makes full use of the redundant and complementary information of different source images. Experimental results show that our method is more suitable for the human visual system and has better objective evaluation index values. However, the proposed image fusion method is very powerful for the details such as image edges. Although remote sensing images have rich detailed information, it would be inefficient if there is much more low frequency information than high frequency information. How to overcome this shortcoming will be investigated in future work.

Author Contributions: X.M. wrote the original draft and she performed the experiments with J.F., S.L. wrote the review & editing. S.H. provided Funding acquisition, and he also provided resources with S.X.

Funding: This research was funded by the Natural Science Foundation of China under Grant No. 61572063 and No. 61401308; Natural Science Foundation of Hebei Province under Grant No. F2016201142 and No. F2016201187; Science Research Project of Hebei Province under Grant No. QN2016085. Opening Foundation of Machine vision Engineering Research Center of Hebei Province under Grant 2018HBMV02, Science Research Project of Hebei Province under Grant QN2016085, Natural Science Foundation of Hebei University under Grant 2014-303, Post-graduate's Innovation Fund Project of Hebei University under Grant hbu2018ss01.

Acknowledgments: Some of the images adopted in experiments are downloaded from the website of <https://sites.google.com/view/durgaprasadbaviriseti/datasets>. This work was supported by the High-Performance Computing Center of Hebei University. We also thank Qu XiaoBo, Zhou Zhiqiang, Li Shutao, and Liu Yu for their shared codes of image fusion. We also thank the Editor and Reviewers for the efforts made in processing this submission and we are particularly grateful to the reviewers for their constructive comments and suggestions which help us improve the quality of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhu, Z.; Yin, H.; Chai, Y.; Li, Y.; Qi, G. A novel multi-modality image fusion based on image decomposition and sparse representation. *Inf. Sci.* **2018**, *432*, 516–529. [[CrossRef](#)]
2. Anandhi, D.; Valli, S. An algorithm for multi-sensors image fusion using maximum a posteriori and nonsubsampling contourlet transform. *Comput. Electr. Eng.* **2017**, *65*, 139–152. [[CrossRef](#)]
3. Gao, Z.; Zhang, C. Texture clear multi-modal image fusion with joint sparsity model. *Optik* **2017**, *130*, 255–265. [[CrossRef](#)]
4. Hu, S.; Yang, D.; Liu, S.; Ma, X. Block-matching based multimodal medical image fusion via PCNN with SML. In Proceedings of the 2016 IEEE 13th International Conference on Signal Processing (ICSP), Chengdu, China, 6–10 November 2016; pp. 13–18.
5. Li, H.; Chai, Y.; Ling, R.; Yin, H. Multifocus image fusion scheme using feature contrast of orientation information measure in lifting stationary wavelet domain. *J. Inf. Sci. Eng.* **2013**, *29*, 227–247.
6. Ghassemian, H. A review of remote sensing image fusion methods. *Inf. Fusion* **2016**, *32*, 75–89. [[CrossRef](#)]
7. Zhang, J.; Feng, X.; Song, B.; Li, M.; Lu, Y. Multi-focus image fusion using quality assessment of spatial domain and genetic algorithm. In Proceedings of the Conference on Human System Interactions, Krakow, Poland, 25–27 May 2008; pp. 71–75. [[CrossRef](#)]
8. Liu, Y.; Liu, S.; Wang, Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fusion* **2015**, *24*, 147–164. [[CrossRef](#)]
9. Li, H.; Qiu, H.; Yu, Z.; Li, B. Multifocus image fusion via fixed window technique of multiscale images and non-local means filtering. *Signal Process.* **2017**, *138*, 71–85. [[CrossRef](#)]
10. Meng, F.; Song, M.; Guo, B.; Shi, R.; Shan, D. Image fusion based on object detection and non-subsampling contourlet transform. *Comput. Electr. Eng.* **2017**, *62*, 375–383. [[CrossRef](#)]
11. Liu, S.; Shi, M.; Zhu, Z.; Zhao, J. Image fusion based on complex-shearlet domain with guided filtering. *Multidimens. Syst. Signal Process.* **2017**, *28*, 207–224. [[CrossRef](#)]
12. Donoho, D.L. Compressed Sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]
13. Mitianoudis, N.; Stathaki, T. Pixel-based and region-based image fusion schemes using ICA bases. *Inf. Fusion* **2007**, *8*, 131–142. [[CrossRef](#)]
14. Yang, B.; Li, S. Multifocus image fusion and restoration with sparse representation. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 884–892. [[CrossRef](#)]
15. Yu, N.; Qiu, T.; Liu, W. Medical image fusion based on sparse representation with K-SVD. In Proceedings of the World Congress on Medical Physics and Biomedical Engineering, Beijing, China, 26–31 May 2012; pp. 550–553. [[CrossRef](#)]
16. Yang, Y.; Wu, L.; Huang, S.; Wan, W.; Que, Y. Remote sensing image fusion based on adaptively weighted joint detail injection. *IEEE Access* **2018**, *6*, 6849–6864. [[CrossRef](#)]
17. Yin, H.; Li, Y.; Chai, Y.; Liu, Z.; Zhu, Z. A novel sparse-representation-based multi-focus image fusion approach. *Neurocomputing* **2016**, *216*, 216–229. [[CrossRef](#)]
18. Zong, J.; Qiu, T. Medical image fusion based on sparse representation of classified image patches. *Biomed. Signal Process. Control* **2017**, *34*, 195–205. [[CrossRef](#)]
19. Zhu, Z.; Chai, Y.; Yin, H.; Li, Y.; Liu, Z. A novel dictionary learning approach for multi-modality medical image fusion. *Neurocomputing* **2016**, *214*, 471–482. [[CrossRef](#)]
20. Elad, M.; Yavneh, I. A plurality of sparse representation is better than the sparsest one alone. *IEEE Trans. Inf. Theory* **2009**, *55*, 4701–4714. [[CrossRef](#)]
21. Luo, X.; Zhang, Z.; Zhang, C.; Wu, X. Multi-focus image fusion using HOSVD and edge intensity. *J. Visual Commun. Image Represent.* **2017**, *45*, 46–61. [[CrossRef](#)]
22. Ma, X.; Hu, S.; Liu, S. SAR image de-noising based on invariant K-SVD and guided filter. *Remote Sens.* **2017**, *9*, 1311. [[CrossRef](#)]
23. He, K.; Sun, J.; Tang, X. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1397–1409. [[CrossRef](#)]
24. Liu, S.; Liu, M.; Li, P.; Zhao, J.; Zhu, Z.; Wang, X. SAR image de-noising via sparse representation in shearlet domain based on continuous cycle spinning. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2985–2992. [[CrossRef](#)]

25. Zhang, X.; Sun, J.; Ma, S.; Lin, Z.; Zhang, J.; Wang, S.; Gao, W. Globally variance-constrained sparse representation and its application in image set coding. *IEEE Trans. Image Process.* **2018**, *27*, 3753–3765. [[CrossRef](#)] [[PubMed](#)]
26. Sitani, D.; Subramanyam, A.V.; Majumdar, A. Online single and multiple analysis dictionary learning-based approach for visual object tracking. *J. Electron. Imag* **2019**, *28*, 013004. [[CrossRef](#)]
27. Peng, L.; Yang, J. Single-image super resolution via hashing classification and sparse representation. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 1923–1927. [[CrossRef](#)]
28. Vanika, S.; Prerna, K.; Angshul, M. Class-wise deep dictionary learning. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 1125–1132. [[CrossRef](#)]
29. Zhang, Z.; Xu, Y.; Yang, J.; Li, X.; Zhang, D. A survey of sparse representation: Algorithms and applications. *IEEE Access* **2015**, *3*, 490–530. [[CrossRef](#)]
30. Pati, Y.C.; Rezaifar, R.; Krishnaprasad, P.S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In Proceedings of the 27th Annual Asilomar Conference on Signals Systems and Computers, Pacific Grove, CA, USA, 1–3 November 1993; pp. 40–44. [[CrossRef](#)]
31. Elad, M.; Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **2007**, *15*, 3736–3745. [[CrossRef](#)]
32. Wang, G. The inverse hyperbolic tangent function and Jacobian sine function. *J. Math. Anal. Appl.* **2017**, *448*, 498–505. [[CrossRef](#)]
33. Zhou, Z.; Li, S.; Wang, B. Multi-scale weighted gradient-based fusion for multi-focus images. *Inf. Fusion* **2014**, *20*, 60–72. [[CrossRef](#)]
34. Li, S.; Kang, X.; Hu, J. Image fusion with guided filtering. *IEEE Trans. Image Process.* **2013**, *22*, 2864–2875. [[CrossRef](#)]
35. Zhao, P.; Liu, G.; Hu, C.; Huang, H.; He, B. Medical image fusion algorithm based on the Laplace-PCA. In Proceedings of the 2013 Chinese Intelligent Automation Conference, CIAC 2013, Yangzhou, Jiangsu, China, 23–25 August 2013; Volume 256, pp. 787–794. [[CrossRef](#)]
36. Shreyamsha Kumar, B.K.; Swamy, M.N.S.; Ahmad, M.O. Multiresolution DCT decomposition for multifocus image fusion. In Proceedings of the 2013 26th IEEE Canadian Conference on Electrical and Computer Engineering, CCECE 2013, Regina, SK, Canada, 5–8 May 2013; pp. 1–4. [[CrossRef](#)]
37. Qu, X.; Yan, J.; Xiao, H.; Zhu, Z. Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampling contourlet transform domain. *Acta Autom. Sin.* **2008**, *34*, 1508–1514. [[CrossRef](#)]
38. Available online: http://www.dlr.de/hr/en/desktopdefault.aspx/tabid-2434/3770_read-32515 (accessed on 6 March 2019).

