

## Article

# A Novel Multi-Focus Image Fusion Method Based on Stochastic Coordinate Coding and Local Density Peaks Clustering

Zhiqin Zhu <sup>1</sup>, Guanqiu Qi <sup>2,\*</sup>, Yi Chai <sup>1</sup>, and Yinong Chen <sup>2</sup>

<sup>1</sup> State Key Laboratory of Power Transmission Equipment and System Security and New Technology, College of Automation, Chongqing University, Chongqing 400044, China; zhiqinzh@outlook.com (Z.Z.); chaiyi@outlook.com (Y.C.)

<sup>2</sup> School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ 85287, USA; yinong@asu.edu

\* Correspondence: guanqiuq@asu.edu; Tel.: +1-323-742-0133

Academic Editor: Carmen de Pablos Heredero

Received: 27 July 2016; Accepted: 3 November 2016; Published: 11 November 2016

**Abstract:** The multi-focus image fusion method is used in image processing to generate all-focus images that have large depth of field (DOF) based on original multi-focus images. Different approaches have been used in the spatial and transform domain to fuse multi-focus images. As one of the most popular image processing methods, dictionary-learning-based sparse representation achieves great performance in multi-focus image fusion. Most of the existing dictionary-learning-based multi-focus image fusion methods directly use the whole source images for dictionary learning. However, it incurs a high error rate and high computation cost in dictionary learning process by using the whole source images. This paper proposes a novel stochastic coordinate coding-based image fusion framework integrated with local density peaks. The proposed multi-focus image fusion method consists of three steps. First, source images are split into small image patches, then the split image patches are classified into a few groups by local density peaks clustering. Next, the grouped image patches are used for sub-dictionary learning by stochastic coordinate coding. The trained sub-dictionaries are combined into a dictionary for sparse representation. Finally, the simultaneous orthogonal matching pursuit (SOMP) algorithm is used to carry out sparse representation. After the three steps, the obtained sparse coefficients are fused following the max L1-norm rule. The fused coefficients are inversely transformed to an image by using the learned dictionary. The results and analyses of comparison experiments demonstrate that fused images of the proposed method have higher qualities than existing state-of-the-art methods.

**Keywords:** multi-focus image fusion; sparse representation; dictionary learning; local density peaks clustering

## 1. Introduction

High-quality images are widely used in different areas of a highly developed society. Following the development of cloud computing, more and more images are processed in a cloud [1,2]. High-quality images can increase the accuracy and efficiency of image processing. Due to the limitation of field depth in most of optical lenses, only the objects a certain distance away from the camera can be captured in focus and sharply, and other objects are out of focus and blurred. It usually takes multiple images of the same scene to enhance the robustness of image processing. However, viewing and analyzing a series of images separately is neither convenient nor efficient [3]. The multi-focus image fusion method is an effective way to resolve this problem by combining

complementary information from multiple images into a fused image, which is useful for human or machine perception [4,5].

During the past few years, many image fusion algorithms have been developed to integrate multi-focus images. In general, multi-focus fusion algorithms can be classified into two groups: spatial-domain fusion and transform-domain fusion [3,6–11]. The spatial-domain methods only need spatial information of images to carry out image fusion, without doing any type of transformation. The main principle of spatial-domain methods is to select those pixels or regions with higher clarity to construct the fused image, according to the image clarity measurement. Energy of Laplacian [8,12] and spatial frequency [3,6,11] are two typical focus measures used to decide the clarity of pixels or regions. The main limitations of spatial-domain fusion methods in generating desirable fused images are the misalignment of decision map within the boundary of focused objects, and the incorrect decision in locating sub-regions of focused or out-of-focused regions. To reduce these limitations, some spatial-domain techniques use the average weight of pixel values to fuse source images, instead of using binary decision [7]. Due to the weight construction method, these spatial-domain methods may lead to blurring edges, contrast decrease, and reduction of sharpness [6].

In contrast to spatial-domain fusion methods, transform-domain fusion methods first convert source images into a transform domain to obtain the corresponding transform coefficients. Then, the transformed coefficients are merged, according to the pre-defined fusion rule. Finally, the fused image is constructed by carrying out inverse transform of the fused coefficients. The most commonly used transform-domain fusion methods are based on multi-scale transform (MST). MST algorithms use the following methods, including discrete wavelet transform, gradient pyramid, dual tree complex wavelet transform, and so on. Recently, some novel transform-domain analysis methods have been proposed, such as curvelet transform [13], and nonsubsampling contourlet transform [9]. Although multi-scale transform coefficients can reasonably represent important features of an image, each transform has its own merits and limitations corresponding to the context of input images. Thus, it is difficult to select an optimal transform basis without apriori knowledge [14,15].

In recent years, sparse-representation based methods, as a subset of transform-domain fusion methods, have been applied to image fusion. Different from other MST methods, sparse-representation based methods usually use learned bases, which can adaptively change according to the input images without apriori knowledge. Due to the adaptively learning feature, sparse representation is an effective way to describe and reconstruct images and signals. It is widely applied to image denoising [16], image deblurring [17], image inpainting [18], super-resolution [19] and image fusion [20]. Yang and Li [21] first applied the sparse representation theory to image fusion field and also proposed a multi-focus image fusion method with an MST dictionary. Li and Zhang applied the morphologically filtering sparse feature to the matrix decomposition method to improve the accuracy of sparse representation in multi-focus image fusion [20]. Wang and Liu proposed an approximate K-SVD based sparse representation method for multi-focus fusion and exposure fusion to reduce the computation costs of sparse-representation based image fusion [22]. Nejati and Samavi proposed K-SVD dictionary-learning based sparse representation for the decision map construction of multi-focus fusion [6]. However, these aforementioned sparse-representation based methods do not take the high computation costs into account as K-SVD, and online dictionary learning. In recent years, many researchers have been devoted to speeding up dictionary learning for image fusion. Zhang and Fu [23] proposed a joint sparse-representation-based image fusion method. Their method had lower complexity than K-SVD. However, it still required a substantial amount of computations. Kim and Han [14] proposed a joint-clustering-based dictionary construction method for image fusion. The proposed method used K-means clustering to group the image patches before dictionary learning. The K-means clustering needed a number of cluster centers before clustering. However, in most cases, the number of cluster centers is difficult to estimate accurately before clustering.

This paper proposes a novel stochastic coordinate coding (SCC)-based image fusion framework integrated with local density peaks clustering. The proposed multi-focus image fusion framework consists of three steps. First, a local density peaks clustering method is applied in order to cluster image patches. The local density peaks based algorithm increases the accuracy of clustering, and does not need any presetting value for the input image data. Second, an SCC-based dictionary construction approach is proposed. The constructed dictionary not only obtains accurate descriptions of input images, but also dramatically decreases the costs of dictionary learning. Finally, the trained dictionary is used for the sparse representation of image patches, and max-L1 theory is implemented in the image fusion process. The key contributions of this paper can be elaborated as follows:

1. An integrated sparse representation framework for multi-focus image fusion is proposed that combines the local density peaks based image-patch clustering and stochastic coordinate coding.
2. An SCC-based dictionary construction method is proposed and applied to sparse representation process, which can obtain a more accurate dictionary and decrease the computation cost of dictionary learning.

The rest of this paper is structured as follows: Section 2 presents and specifies the proposed framework; Section 3 simulates the proposed solutions and analyzes experiment results; and Section 4 concludes this paper.

## 2. Framework

### 2.1. Introduction of Framework

The proposed framework for image fusion shown in Figure 1 has three main steps. All image patches are clustered into different groups in the first step. Then each image patch group is learned by a sub-dictionary using the SCC algorithm [24] and these sub-dictionaries are combined into an integrated dictionary. Finally, the learned dictionary is used for image fusion. The details of each algorithm and method will be explained in the following paragraphs.

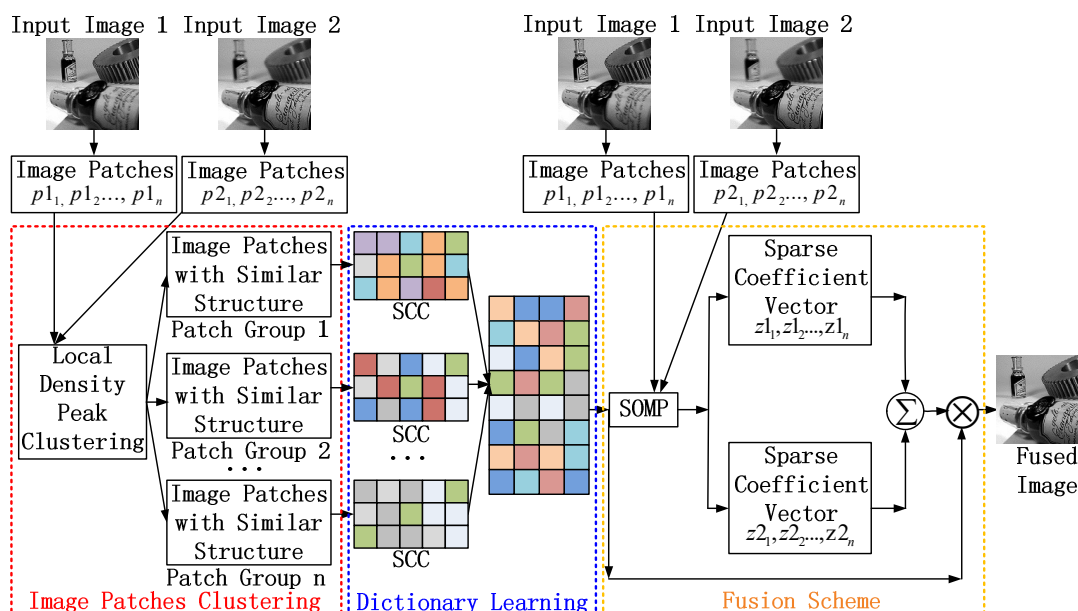


Figure 1. Proposed Image Fusion Framework.

### 2.2. Local Density Peaks Clustering

An image usually consists of different types of image patches. It is efficient to describe the underlying structure of each image patch by using specific sub-dictionary that describes different

types of image patches. This paper uses the local density peaks clustering method to classify image patches into a specific group by the similarity of structure [25,26]. Compared with other existing clustering methods, the local density peaks clustering method has two advantages. First, the method is insensitive to the start point (or initialization). Second, it does not need to know the number of clusters before clustering. Moreover, the basis of local density peaks clustering can be easily expressed by Euclidean distance between two different patches.

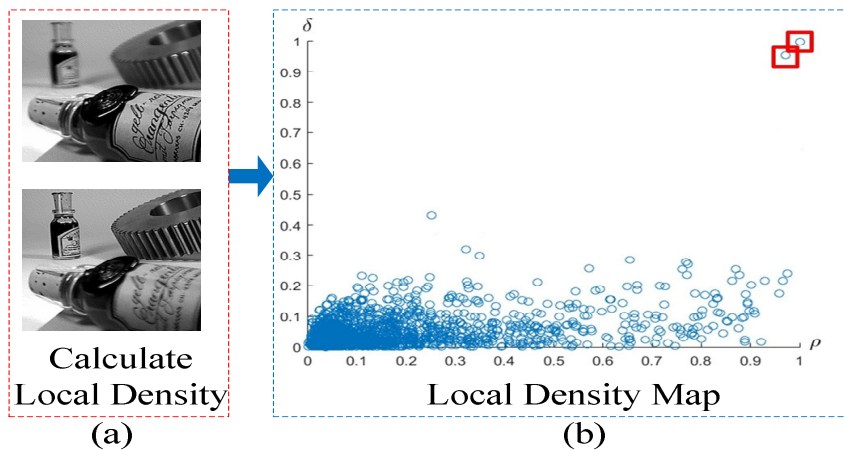
In Figure 2a, the local density  $\rho_i$  of each image patch  $i$  is calculated by using Equation (1).

$$\rho_i = \sum_j \chi(d_{ij} - d_c) , \quad (1)$$

when  $x \leq 0$ ,  $\chi(x) = 0$ , otherwise  $\chi(x) = 1$ .  $d_{ij}$  is the Euclidean distance between image patch  $i$  and  $j$ .  $d_c$  is a cutoff distance and usually set to the median value of  $d_{ij}$ . Basically,  $\rho_i$  equals the number of patches that are closer than  $d_c$  to patch  $i$ . The clustering algorithm is only sensitive to the relative magnitude of  $\rho_i$  from a different image patch, and robust with respect to the choice of  $d_c$ .

A distance  $\delta_i$  of each image patch is measured to find the cluster centers. The calculation equation is shown in Equation (2).  $\delta_i$  is the minimum distance between the image patch  $i$  and any other patch  $j$  with higher density. For the point with highest density,  $\delta_i = \max(d_{ij})$ . A local density map can be constructed by using  $\rho_i$  ( $x$ -axis) and the normalized  $\delta_i$  ( $y$ -axis,  $0 \leq \delta_i \leq 1$ ), which is shown in Figure 2b. The cluster centers are recognized that are boxed by dotted squares in Figure 2b, when the value of  $\delta_i$  is anomalously large. When the cluster centers are identified, the remaining image patches are clustered into the relatively nearest identified center.

$$\delta_i = \min_{j: \rho_j > \rho_i} (d_{ij}) , \quad (2)$$



**Figure 2.** Local Density and Distance Map of Fusion Images, (a) shows the local density calculation of each image patch; (b) shows the constructed local density map.

### 2.3. Dictionary Construction

In the clustering step, image patches with similar structure are classified into a few groups. To construct a more discriminative and compact dictionary, the SCC online dictionary learning algorithm [24] is used to learn a sub-dictionary for each cluster. Subsequently, the learned sub-dictionaries are combined to a new dictionary for the image sparse representation and restoration. The dictionary construction process is illustrated in Figure 1 as the dictionary learning step.



### 2.3.1. Sub-Dictionary Learning Approach

The SCC online dictionary learning algorithm [24] shown in Algorithm 1 extracts eigenvalues from each cluster and builds the corresponding sub-dictionary. The dictionary and sparse code are initialized and denoted as  $\mathbf{D}_1^1, \mathbf{z}_i^0 = 0, i = 1, 2, \dots, n$  respectively. The general expression of the sparse code is  $\mathbf{z}_i^k = 0, i = 1, 2, \dots, n, k = 1, 2, \dots, n$  and learning rate  $\eta_1^1 = 1$ . The number of epochs and the index of data points are represented as superscript  $k$  and subscript  $i$  respectively. It acquires an image patch  $\mathbf{x}_i$ , when  $k = 1$  and  $i = 1$ . The sparse code  $\mathbf{z}_i^k$  is updated by a few steps of coordinate descent (CD):

$$\mathbf{z}_i^k = CD(\mathbf{D}_i^k, \mathbf{x}_i, \mathbf{z}_i^{k-1}) , \quad (3)$$

The  $j$ -th coordinate  $z_{i,j}^{k-1}$  of  $\mathbf{z}_i^{k-1}$  ( $0 \leq j \leq m$ ) is updated as follows:

$$b_j \leftarrow (d_{i,j}^k)^T (\mathbf{x}_i - \mathbf{D}_i^k \mathbf{z}_i^{k-1}) + z_{i,j}^{k-1} , \quad (4)$$

$$z_{i,j}^{k-1} \leftarrow h_\lambda(b_j) , \quad (5)$$

where  $h_\lambda$  is a soft threshold shrinkage function [27,28] and  $b_j$  is a descent parameter that can be calculated by Equation (5). An updating cycle is equivalent to one step of coordinate descent. Dictionary  $\mathbf{D}$  is updated by stochastic gradient decent (SGD):

$$\mathbf{D}_{i+1}^k = P_{B_m}(\mathbf{D}_i^k - \eta_i^k \nabla_{\mathbf{D}_i^k} f_i(\mathbf{D}_i^k, \mathbf{z}_i^k)) , \quad (6)$$

where  $P$  denotes the projection operator, and  $B_m$  is the feasible set of  $\mathbf{D}$  that is defined as follows:  $B_m = \{\mathbf{D} \in \mathbb{R}^{m \times n} : \forall j = 1, \dots, m, \|d_j\|_2 \leq 1\}$ . The learning rate is an approximation of the inverse of the Hessian matrix  $\mathbf{H} = \sum_{k,i} \mathbf{z}_i^k (\mathbf{z}_i^k)^T$ . The gradient of  $\mathbf{D}_i^k$  can be obtained as follows:

$$\nabla_{\mathbf{D}_i^k} f_i(\mathbf{D}_i^k, \mathbf{z}_i^k) = (\mathbf{D}_i^k \mathbf{z}_i^k - \mathbf{x}_i) (\mathbf{z}_i^k)^T , \quad (7)$$

This allows  $i = i + 1$ . If  $i > n$ , set  $\mathbf{D}_1^{k+1} = \mathbf{D}_{n+1}^k, k = k + 1$  and  $i = 1$ . Then it repeats the previous calculating process. When  $k > m$ , the calculation stops,  $m$  is preset value and usually  $10 \leq m \leq 15$ . The SCC only runs a few steps of CD to update the sparse codes and SGD algorithm is conducted to update the dictionary.

All sub-dictionaries  $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_n$  are learned by using SCC. These sub-dictionaries are used for describing the underlying structure of each image patches cluster.

### 2.3.2. Sub-Dictionary Combination

As a sub-dictionary for each cluster is learned, all sub-dictionaries are combined into a new dictionary  $\Phi$ .

$$\Phi = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_n] , \quad (8)$$

### 2.4. Fusion Scheme

The fusion scheme is shown in Figure 1 and the image fusion algorithm is shown in Algorithm 2. The learned dictionary is used for the estimation of coefficient vectors. For each image patch  $\mathbf{p}_i$ , a coefficient vector  $\mathbf{z}_i$  is estimated by the SOMP algorithm using the learned dictionary. Max-L1 rule [21] is conducted for coefficient fusion shown in Equation (17),

$$\mathbf{z}^i = \sum_{k=1}^m \mathbf{z}_k * O^k, \text{ where } \begin{cases} O^k = 1, \max(\|\mathbf{z}_1\|_1, \|\mathbf{z}_2\|_1, \dots, \|\mathbf{z}_m\|_1) = \|\mathbf{z}_k\|_1 \\ O^k = 0, \text{ otherwise} \end{cases} \quad (9)$$

where  $\mathbf{z}^i$  is the fused coefficient vector,  $\|\bullet\|_1$  is the  $l_1$  norm, and  $*$  is an elementary multiplication operation.

---

**Algorithm 1** Online Sub-dictionary Learning Algorithm
 

---

**Input:**

image patch  $\mathbf{x}_i$ , sparse code  $\mathbf{z}_i^k$ , learning rate  $\eta_1^1$ , running time  $m$

**Output:**

learned sub-dictionaries  $\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_n$

```

1: Initialize  $\mathbf{D}_1^1, \mathbf{z}_i^0=0$  ( $i=1,2,\dots,n$ ),  $\eta_1^1=1$ 
2: Acquire an image patch  $\mathbf{x}_i$ , when  $k=1$  and  $i=1$ 
3: while  $k \leq m$  do
4:   Update sparse code  $\mathbf{z}_i^k = CD(\mathbf{D}_i^k, \mathbf{x}_i, \mathbf{z}_1^{k-1})$ 
5:   Update  $j$ -th coordinate  $z_{i,j}^{k-1}$  of  $\mathbf{z}_i^{k-1}$  according to Equations (4) and (5)
6:   Update dictionary  $\mathbf{D}_{i+1}^k = P_{B_m}(\mathbf{D}_i^k - \eta_i^k \nabla_{\mathbf{D}_i^k} f_i(\mathbf{D}_i^k, \mathbf{z}_i^k))$ 
7:   Update  $i = i + 1$ 
8:   if  $i > n$  then
9:      $\mathbf{D}_1^{k+1} = \mathbf{D}_{n+1}^k, k = k + 1$ , and  $i = 1$ 
10:   end if
11: end while

```

---



---

**Algorithm 2** Image Fusion Algorithm
 

---

**Input:**

image patch  $\mathbf{p}_i$ , coefficient vector  $\mathbf{z}_i$

**Output:**

fused image  $I$

```

1: for  $i = 1; i \leq n; i++$  do
2:    $\mathbf{z}^i = \sum_{k=1}^m \mathbf{z}_k * O^k$  according to Equation 9
3: end for
4: for  $i = 1; i \leq n; i++$  do
5:    $\mathbf{I}^i = \mathbf{D}\mathbf{Z}^i$ 
6: end for

```

---

The fused coefficient vectors are restored to an image. The restoring process is based on Equation (10),

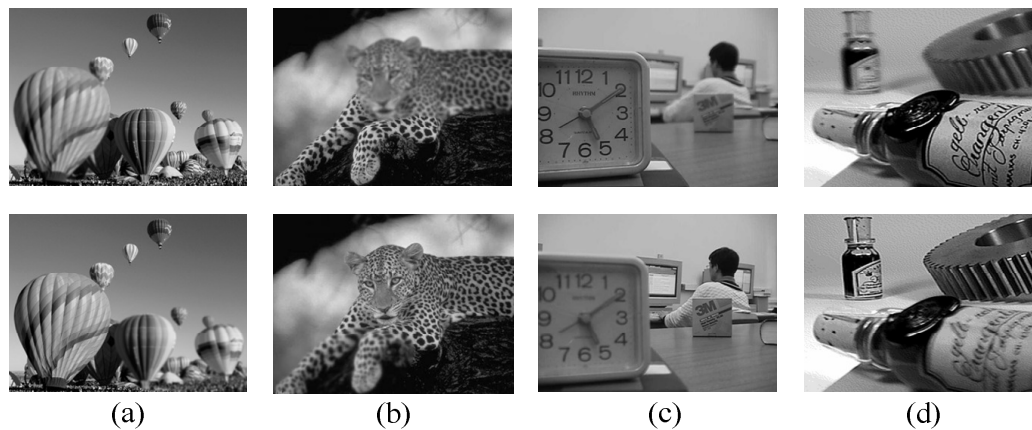
$$\mathbf{I}^i = \mathbf{D}\mathbf{Z}^i \quad (10)$$

where the  $\mathbf{Z}^i = \{\mathbf{z}_1^i, \mathbf{z}_2^i, \dots, \mathbf{z}_m^i\}$  is corresponding to image patches of the fused image and  $\mathbf{D}$  is the learned dictionary.

### 3. Experiments and Analyses

The proposed multi-focus image fusion method is applied to standard multi-focus images from public website [29]. All standard multi-focus images used in this paper are free to use for research purposes. The images from the image fusion library have the size of  $256 \times 256$  pixels or  $320 \times 240$  pixels. The fused images are evaluated by comparing them to the fused images of other existing methods. In this paper, four pairs of images are used as a case study to simulate the proposed multi-focus image fusion method. To simulate a real world environment, four pairs of images have

two scenes. One is an outdoor scene, such as a hot-air balloon and leopard shown in Figure 3a,b respectively. The other is an indoor scene, such as a lab and bottle shown in Figure 3c,d respectively. These four pairs of original images are from the same sensor modality. Since each image focuses on a different object, there are two images for each scene. The out-of-focus regions in the original images are blurred.



**Figure 3.** Four Source Image Pairs of Different Scenes for Multi-focus Fusion Experiments, (a)–(d) are source image pairs of hot-air balloon, leopard, lab, bottle respectively.

### 3.1. Experiment Setup

The quality of the proposed image fusion scheme is evaluated against the other seven popular multi-focus fusion methods. These multi-focus fusion methods consist of popular spatial-based image fusion methods, like Laplacian energy (LE) [8]; four states of art MST methods, including disperse wavelet transform (DWT) [30], dual-tree complex wavelet transform (DT-CWT) [31], curvelet (CVT) [13], non-subsampled contourlet (NSCT) [9]; two sparse representation methods, including sparse representation with a fixed DCT dictionary (SR-DCT) [21] and sparse representation with a trained dictionary by K-SVD (SR-KSVD) [32]. The objective evaluation of fused image includes edge intensity (EI) [33], edge retention  $Q^{AB/F}$  [34], mutual information (MI) [35,36], and visual information fidelity (VIF) [37]. Then it compares the dictionary construction time of the proposed method with K-SVD [32], which is the most popular dictionary construction method. All experiments are implemented using Matlab, version 2014a; MathWorks: Natick, MA, 2014. and Visual Studio, version Community 2013; Microsoft: Redmond, WA, 2013. mixed programming on an Intel(R) Core(TM), version i7-4720HQ; Intel: Santa Clara, CA, 2015. CPU @ 2.60GHz Laptop with 12.00 GB RAM.

#### 3.1.1. Edge Intensity

The quality of the fused image is measured by the local edge intensity  $L$  in image  $I$  [38]. It folds a Gaussian kernel  $G$  with the image  $I$  to get a smoothed image. Then it obtains the edge intensity image by subtracting the smoothed image from the original image. The spectrum of the edge intensities depends on the width of the Gaussian kernel  $G$ .

$$L = \max(0, I - G * I) \quad (11)$$

The fused image  $H$  is calculated by image  $L_j, j = 1, \dots, n$  and the weighted average of local edge intensities.

$$\mathbf{H}(x, y) = \sum_{j=1}^N w_j(x, y) \mathbf{L}_j(x, y), \quad (12)$$

$$w_j(x, y) = \frac{L_j(x, y)}{\sum_{k=1}^N L_k(x, y)} \quad (13)$$

### 3.1.2. Mutual Information

MI for images can be formalized as Equation (14) [35].

$$MI = \sum_{i=1}^L \sum_{j=1}^L h_{A,F}(i, j) \log_2 \frac{h_{A,F}(i, j)}{h_A(i) h_F(j)}, \quad (14)$$

where  $L$  is the number of gray-level,  $h_{R,F}(i, j)$  is the gray histogram of image A and F. The  $h_A(i)$  and  $h_F(j)$  are edge histogram of image A and F. For a fused image, the MI of the fused image can be calculated by Equation (15).

$$MI(A, B, F) = MI(A, F) + MI(B, F), \quad (15)$$

where  $MI(A, F)$  represents the MI value between input image A and fused image F;  $MI(B, F)$  represents the MI value of input image B and fused image F.

### 3.1.3. $Q^{AB/F}$

The  $Q^{AB/F}$  metric is a gradient-based quality index to measure how well the edge information of source images conducted to the fused image [34]. It is calculated by:

$$Q^{AB/F} = \frac{\sum_{i,j} (Q^{AF}(i, j) w^A(i, j) + Q^{BF}(i, j) w^B(i, j))}{\sum_{i,j} (w^A(i, j) + w^B(i, j))}, \quad (16)$$

where  $Q^{AF} = Q_s^{AF} Q_0^{AF}$ ,  $Q_s^{AF}$  and  $Q_0^{AF}$  are the edge strength and orientation preservation values at location  $(i, j)$ .  $Q^{BF}$  can be computed similarly to  $Q^{AF}$ .  $w_A(i, j)$  and  $w_B(i, j)$  are the importance weights of  $Q^{AF}$  and  $Q^{BF}$  respectively.

### 3.1.4. Visual Information Fidelity

VIF is the novel full reference image quality metric. VIF quantifies the mutual information between the reference and test images based on natural scene statistics (NSS) theory and human visual system (HVS) model. It can be expressed as the ratio between the distorted test image information and the reference image information, the calculation equation of VIF is shown in Equation (17).

$$VIF = \frac{\sum_{i \in \text{subbands}} I(\overrightarrow{C^{N,i}}, \overrightarrow{F^{N,i}})}{\sum_{i \in \text{subbands}} I(\overrightarrow{C^{N,i}}, \overrightarrow{E^{N,i}})}, \quad (17)$$

where  $I(\overrightarrow{C^{N,i}}, \overrightarrow{F^{N,i}})$  and  $I(\overrightarrow{C^{N,i}}, \overrightarrow{E^{N,i}})$  represent the mutual information, which are extracted from a particular subband in the reference and the test images respectively.  $\overrightarrow{C^N}$  denotes N elements from a random field,  $\overrightarrow{E^N}$  and  $\overrightarrow{F^N}$  are visual signals at the output of HVS model from the reference and the test images respectively.

An average VIF value of each input image and integrated image is used to evaluate the fused image. The evaluation function of VIF for image fusion is shown in Equation (18) [37].

$$VIF(A, B, F) = \frac{VIF(A, F) + VIF(B, F)}{2} , \quad (18)$$

where  $VIF(A, F)$  is the VIF value between input image A and fused image F;  $VIF(B, F)$  is the VIF value between input image B and fused image F.

### 3.2. Image Quality Comparison

To show the efficiency of the proposed method, the quality comparison of fused images is demonstrated. Four pairs of multi-focused images of a hot-air balloon, leopard, lab, and bottle are employed for quality comparison. It compares the quality of fused image based on visual effect, the accuracy of focused region detection, and the objective evaluations. The different images are used to show the differences between fused images and corresponding source images. This paper increases the contrast and brightness of difference images for printable purposes. All difference images are adjusted by using the same parameters.

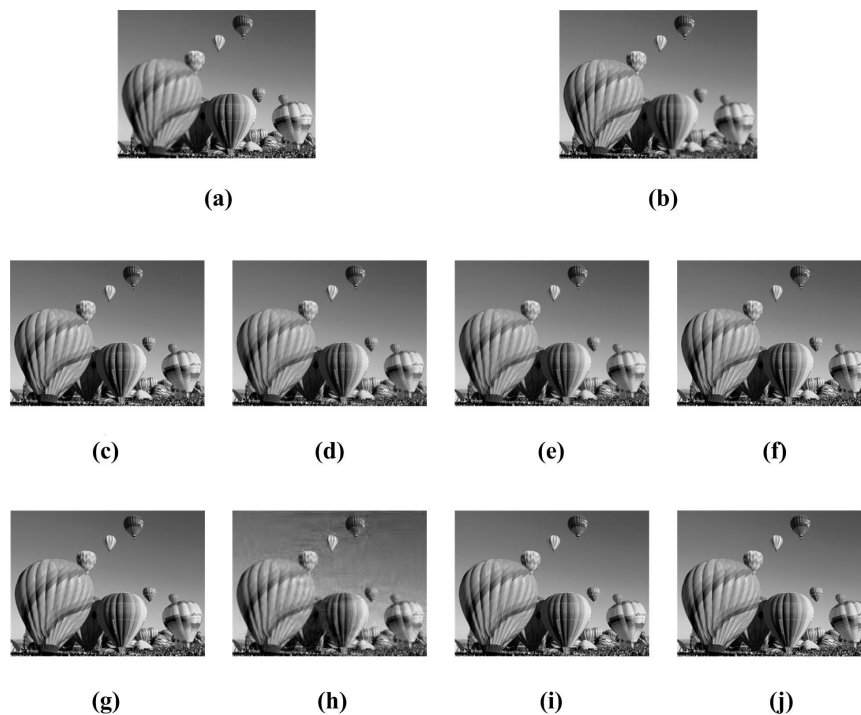
In the first comparison experiment, the "hot-air balloon" images are a pair of multi-focused images. The source multi-focused images are shown in Figure 4a,b. In Figure 4a, the biggest hot-air balloon on the bottom left is out of focus, the rest of the hot-air balloons are in focus. In contrast, in Figure 4b, the biggest hot-air balloon is in focus, but the rest of balloons are out of focus. LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD and the proposed method are employed to merge two multi-focused images into a clear one, respectively. The corresponding fusion results are shown in Figure 4c–j respectively. The difference images of LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD and the proposed method do the matrix subtraction with the source images shown in Figure 4a,b. The corresponding subtracted results are shown in Figure 5a–h,i–p respectively.

Figure 5a–h are the difference images between Figure 4a and Figure 4c–j, and Figure 5i–p are the difference images between Figure 4b and Figure 4c–j. The difference images of LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD and the proposed method are the matrix subtraction results of the corresponding fused images and source images shown in Figure 4a,b.

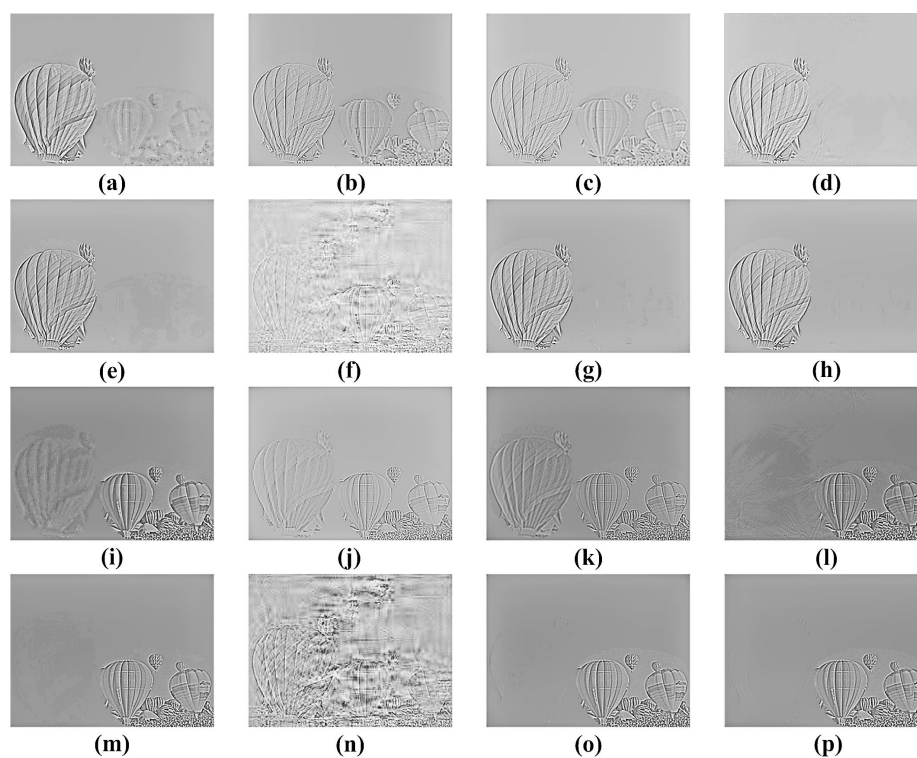
There are a lot of noises in Figure 4h, which are acquired by SR-DCT. The rest of integrated images in Figure 4 are similar. Difference images, that show hot-air balloons of LE, DWT, and DT-CWT on the left side respectively, do not get all the edge information in Figure 5a–c.

Similarly, Figure 5i–k shows that the biggest hot-air balloons of LE, DWT, and DT-CWT on the bottom left, respectively, are not totally focused. Due to the misjudgement of focused areas, the fused "hot-air balloon" images of LE, DWT, DT-CWT, and SR-DCT have shortcomings. Compared with source images, the rest of the methods do great job in identifying the focused area. To further compare the quality of fused images, objective metrics are used.

Table 1 shows the objective evaluations. Compared with the rest of the image fusion methods, the proposed method SR-SCC gets the largest value in MI and VIF. LE and DT-CWT get the largest value in EI and  $Q^{AB/F}$  respectively, but they provide an inaccurate decision in detecting the focused region. The proposed method has the best overall performance of multi-focus image fusion in the "hot-air balloon" scene among all eight methods, according to the quality of fused image, accuracy of locating focused region, and objective evaluations.



**Figure 4.** Fused Images of "Hot-air Balloon" by Different Methods, (a,b) are source images, (c–j) are fused images of LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD, and proposed SR-SCC respectively.



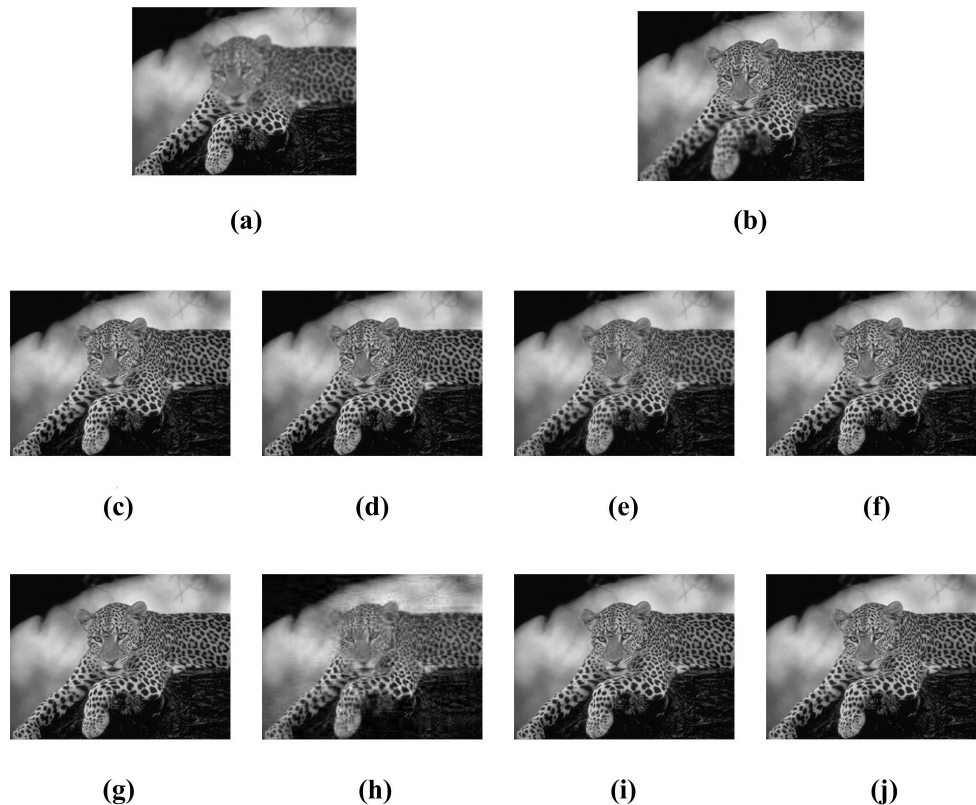
**Figure 5.** Difference Images of "Hot-air Balloon" by Different Methods, (a–h) are differences images between source image in Figure 4a and fused images of LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD, and proposed SR-SCC in Figure 4c–j respectively, (i–p) are differences images between source image in Figure 4b and corresponding fused images in Figure 4c–j respectively.

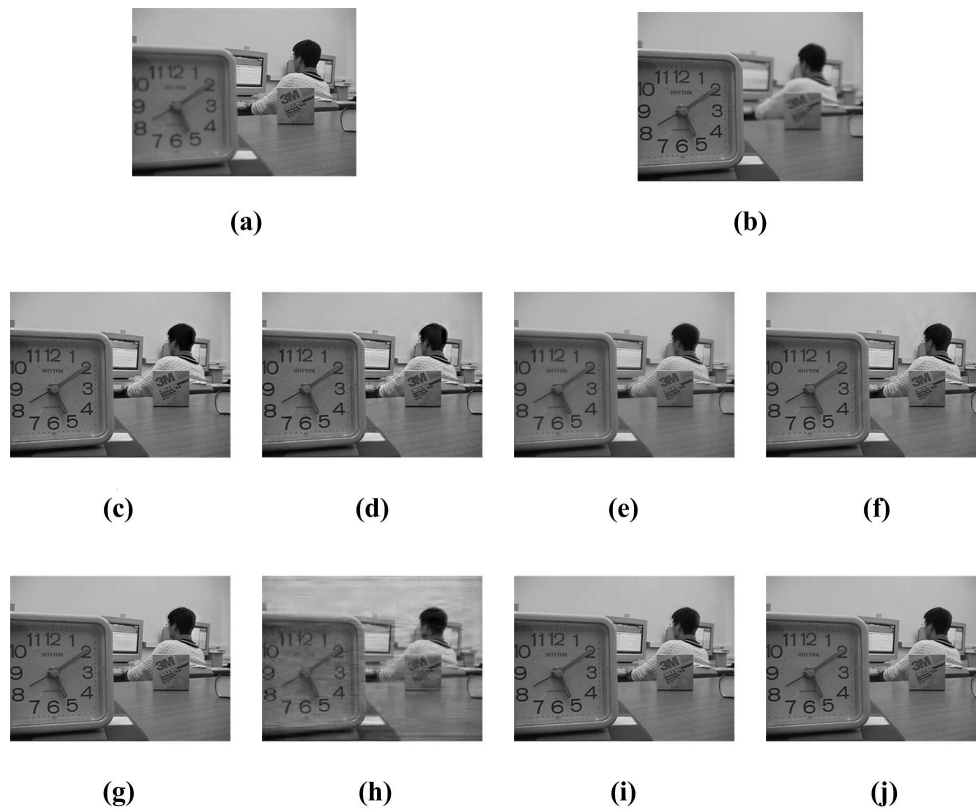


**Table 1.** Objective Evaluations of Multi-focus "Hot-air Balloon" Fusion Experiments.

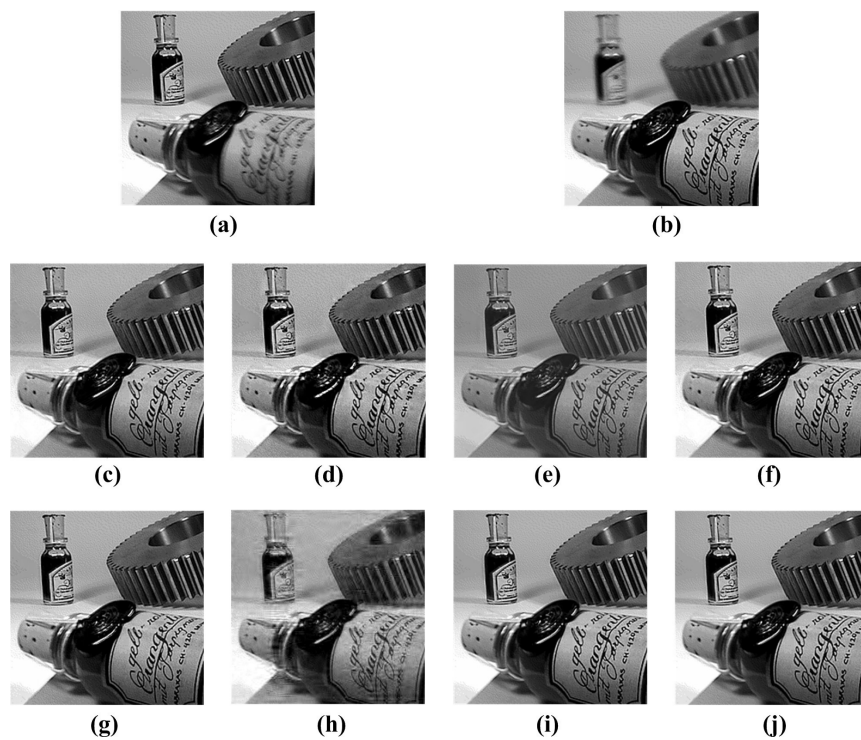
Method	EI	$Q^{AB/F}$	MI	VIF
LE	<b>70.5069</b>	0.9843	7.1359	0.7829
DWT	64.2923	0.9465	6.8532	0.7463
DT-CWT	63.5688	<b>0.9878</b>	6.8828	0.7771
CVT	69.6241	0.9812	7.1579	0.8113
NSCT	69.4987	0.9850	7.6092	0.8187
SR-DCT	65.9602	0.7488	3.9220	0.4695
SR-KSVD	69.5190	0.9854	7.8824	0.8234
SR-SCC	69.5625	0.9852	<b>7.9117</b>	<b>0.8242</b>

Similarly, the source images of the other three comparison experiments, as "leopard", "lab" and "bottle", are shown in Figures 6, 7 and 8a,b respectively. In a set of source images, two images (a) and (b) focus on different items. The source images are fused by LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD, and the proposed method to get a totally focused image, and the corresponding fusion results are shown in Figures 6, 7 and 8c–j respectively. The difference between fused images and their corresponding source images are shown in Figures 9, 10 and 11a–h,i–p respectively.

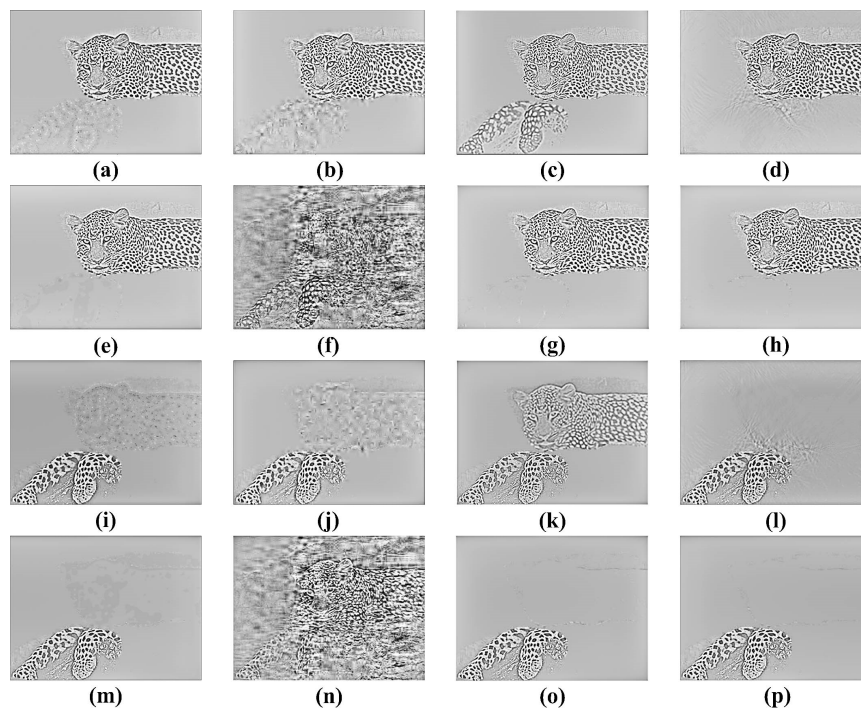
**Figure 6.** Fused Images of "Leopard" by Different Methods, (a,b) are source images, (c–j) are fused images of LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD, and proposed SR-SCC respectively.



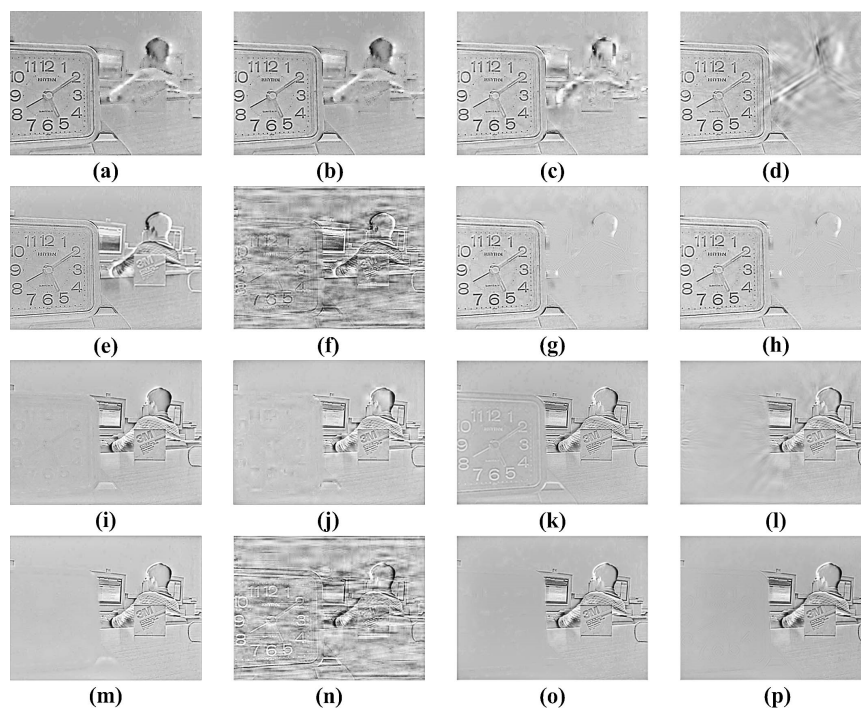
**Figure 7.** Fused Images of "Lab" by Different Methods, (a,b) are source images; (c–j) are fused images of LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD, and proposed SR-SCC respectively.



**Figure 8.** Fused Images of "Bottle" by Different Methods, (a,b) are source images; (c–j) are fused images of LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD, and proposed SR-SCC respectively.

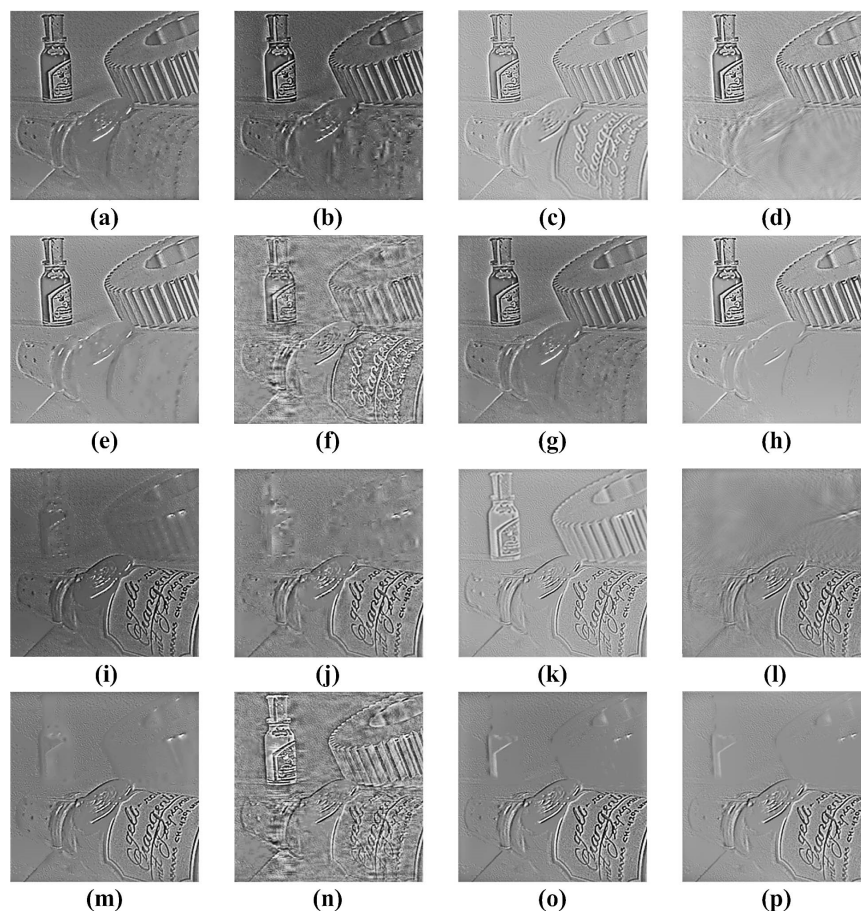


**Figure 9.** Difference Images of "Leopard" by Different Methods, (a–h) are differences images between source image in Figure 6a and fused images of LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD, and proposed SR-SCC in Figure 6c–j respectively, (i–p) are differences images between source image in Figure 6b and corresponding fused images in Figure 6c–j respectively.



**Figure 10.** Difference Images of "Lab" by Different Methods, (a–h) are differences images between source image in Figure 7a and fused images of LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD, and proposed SR-SCC in Figure 7c–j respectively, (i–p) are differences images between source image in Figure 7b and corresponding fused images in Figure 7c–j respectively.





**Figure 11.** Difference Images of "Bottle" by Different Methods, (a–h) are differences images between source image in Figure 8a and fused images of LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD, and proposed SR-SCC in Figure 8c–j respectively, (i–p) are differences images between source image in Figure 8b and corresponding fused images in Figure 8c–j respectively.

Objective metrics of multi-focus "leopard", "lab", and "bottle" fusion experiments are shown in Tables 2–4 respectively to evaluate the quality of fused images.

- **Multi-focus "leopard" fusion:** The proposed method SR-SCC achieves the largest value in MI and VIF. LE obtains the largest value in EI index, but it makes inaccurate decision in detecting the focused region. SR-KSVD shows great performance in  $Q^{AB/F}$ , and the result of proposed method is only 0.0002 smaller than SR-KSVD. According to the quality of visual image, the accuracy of focused region, and objective evaluations, the proposed method does a better job than the rest of the methods.
- **Multi-focus "lab" fusion:** The proposed method SR-SCC achieves the largest value in  $Q^{AB/F}$  and VIF. DWT obtains the largest value of EI index, but it cannot distinguish the correct focused areas. SR-KSVD has the best performance in MI. The proposed method and SR-KSVD show great performance of visual effect in focused area, distinguishing focused area, and objective evaluation. Compared with SR-KSVD, the proposed method dramatically reduces computation costs in dictionary construction. So the proposed method has the best overall performance among all comparison methods.
- **Multi-focus "bottle" fusion:** DWT obtains the largest value in EI, but it does not get an accurate focused area. The proposed method achieves SR-SCC with the largest values of the rest of the objective evaluations. So the proposed method has the best overall performance compared with other methods in the "bottle" scene.

**Table 2.** Objective Evaluations of Multi-focus "Leopard" Fusion Experiments.

Method	EI	$Q^{AB/F}$	MI	VIF
LE	<b>94.1343</b>	0.9443	6.7677	0.7920
DWT	94.8093	0.9430	6.1765	0.7629
DT-CWT	80.2643	0.9470	6.2083	0.7636
CVT	93.5736	0.9359	6.4109	0.8068
NSCT	93.6767	0.9447	7.1940	0.8245
SR-DCT	76.9935	0.7343	3.3002	0.7291
SR-KSVD	93.5276	<b>0.9450</b>	7.4607	0.8269
SR-SCC	93.6448	0.9448	<b>7.5420</b>	<b>0.8283</b>

**Table 3.** Objective Evaluations of Multi-focus "Lab" Fusion Experiments.

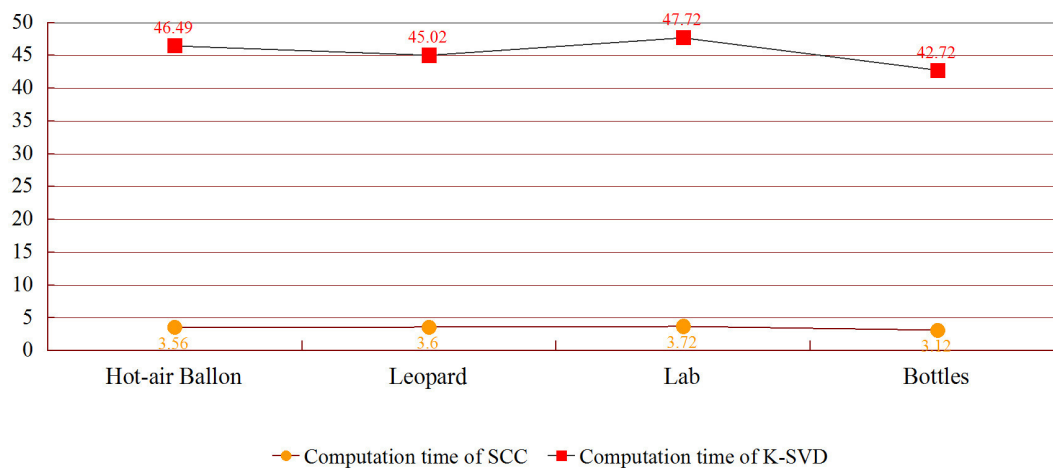
Method	EI	$Q^{AB/F}$	MI	VIF
LE	55.0926	0.7964	5.0724	0.6299
DWT	<b>56.7262</b>	0.8795	4.6665	0.5885
DT-CWT	44.6470	0.8196	5.0999	0.6025
CVT	53.5711	0.8438	4.7937	0.6367
NSCT	53.5958	0.8055	5.3111	0.6625
SR-DCT	48.8765	0.6332	3.1122	0.3385
SR-KSVD	53.0461	0.8791	<b>5.7345</b>	0.6766
SR-SCC	53.0304	<b>0.8826</b>	5.6656	<b>0.6893</b>

### 3.3. Dictionary Construction Time Comparison

As shown in the previous subsection, the fused images of different multi-focus fusion methods are compared by objective evaluations. Dictionary-learning based image fusion methods, including SR-KSVD and the proposed SCC method, achieve the best performance. However, the dictionary construction process usually takes a very long time. The efficiency of dictionary construction is an important feature of image fusion method. Both K-SVD [39] and the proposed SCC are sparse-representation based dictionary learning methods. So it compares the dictionary construction time of K-SVD and the proposed SCC. K-SVD is one of the most popular dictionary learning methods in recent years. It uses an iterative algorithm to reduce dictionary learning errors and can describe the underlying structure of the image perfectly. To verify the low computation cost of the proposed method, four pairs of images are used for testing computation time. The time consumption of K-SVD and SCC are shown in Figure 12 and Table 5. SCC uses low computation times, that are marked in bold, in four group experiments. The experimental results demonstrate that SCC has a much better performance of computation time than K-SVD.

**Table 4.** Objective Evaluations of Multi-focus "Bottle" Fusion Experiments.

Method	EI	$Q^{AB/F}$	MI	VIF
LE	128.0512	0.8478	4.1059	0.4113
DWT	<b>131.5033</b>	0.8355	3.5696	0.3779
DT-CWT	96.3171	0.8548	3.8049	0.4035
CVT	122.5616	0.8448	3.8799	0.4678
NSCT	69.4987	0.9850	7.6092	0.8187
SR-DCT	122.4806	0.8518	4.1766	0.4784
SR-KSVD	128.4125	0.8625	4.8410	0.5118
SR-SCC	128.5106	<b>0.8629</b>	<b>4.8901</b>	<b>0.5291</b>



**Figure 12.** Computation Time Comparison.

**Table 5.** Time Consumption Comparison.

Method	Hot-Air Balloon	Leopard	Lab	Bottle
K-SVD	46.49	45.02	47.72	42.72
SCC	3.56	3.60	3.72	3.12

#### 4. Conclusions

This paper proposed an integrated image fusion framework based on online dictionary learning. Compared with traditional image fusion methods, the integrated framework had two major improvements. First, it introduced a local density-based clustering method in sparse representation, which had high performance in clustering without any apriori knowledge. Second, an online dictionary learning algorithm was used to extract discriminative features that enhanced the efficiency of image fusion. The proposed method was compared with seven existing algorithms LE, DWT, DT-CWT, CVT, NSCT, SR-DCT, SR-KSVD, and SR-SCC using four source image pairs and objective metrics. Experimental results demonstrated that the proposed method was significantly superior to other methods in terms of subjective and objective evaluation. This means that the fused images of the proposed method had better quality than other methods. Compared with other sparse-representation based methods, the proposed method had high efficiency in generating fused images.

Although the proposed solution had a good performance in image fusion, many optimizations are still worth doing in the follow-up research. The parallel processing and the use of multiple graphics processing units (GPUs) will be considered to improve the efficiency of proposed solution. Denoising techniques will also be applied to the proposed solution to enhance the quality of the fused image.

**Acknowledgments:** We would like to thank the supports by National Natural Science Foundation of China (Nos.61374135, 61203321,61302041), China Central Universities Foundation (106112013CDJZR170005), Chongqing Special Funding in Postdoctoral Scientific Research Project (XM2013007) and Chongqing Funding in Postgraduate Research Innovation Project (CYB14023).

**Author Contributions:** Zhiqin Zhu and Guanqiu Qi conceived and designed the experiments; Zhiqin Zhu and Guanqiu Qi performed the experiments; Zhiqin Zhu and Guanqiu Qi analyzed the data; Zhiqin Zhu contributed reagents/materials/analysis tools; Zhiqin Zhu and Guanqiu Qi wrote the paper; Yi Chai and Yinong Chen provided technical support and revised the paper.

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



## References

1. Tsai, W.; Qi, G. DICB: Dynamic Intelligent Customizable Benign Pricing Strategy for Cloud Computing. In Proceedings of the 5th IEEE International Conference on Cloud Computing, Honolulu, HI, USA, 24–29 June 2012; pp. 654–661.
2. Tsai, W.; Qi, G.; Chen, Y. Choosing cost-effective configuration in cloud storage. In Proceedings of the 11th IEEE International Symposium on Autonomous Decentralized Systems, ISADS 2013, Mexico City, Mexico, 6–8 March 2013; pp. 1–8.
3. Li, S.; Kang, X.; Hu, J.; Yang, B. Image matting for fusion of multi-focus images in dynamic scenes. *Inf. Fusion* **2013**, *14*, 147–162.
4. Zhu, Z.; Qi, G.; Chai, Y.; Yin, H.; Sun, J. A Novel Visible-infrared Image Fusion Framework for Smart City. *Int. J. Simul. Process Model.* **2016**, in press.
5. Liu, Z.; Chai, Y.; Yin, H.; Zhou, J.; Zhu, Z. A novel multi-focus image fusion approach based on image decomposition. *Inf. Fusion* **2017**, *35*, 102–116.
6. Nejati, M.; Samavi, S.; Shirani, S. Multi-focus image fusion using dictionary-based sparse representation. *Inf. Fusion* **2015**, *25*, 72–84.
7. Kumar, B.K.S. Image fusion based on pixel significance using cross bilateral filter. *Signal Image Video Process.* **2015**, *9*, 1193–1204.
8. Bai, X.; Zhang, Y.; Zhou, F.; Xue, B. Quadtree-based multi-focus image fusion using a weighted focus-measure. *Inf. Fusion* **2015**, *22*, 105–118.
9. Kong, X.L.H.L.Z.Y.Y. Multifocus image fusion scheme based on the multiscale curvature in nonsubsamped contourlet transform domain. *Opt. Eng.* **2015**, *54*, 1–15.
10. Lai, Y.; Chen, Y.; Liu, Z.; Yang, Z.; Li, X. On monitoring and predicting mobile network traffic abnormality. *Simul. Model. Pract. Theory* **2015**, *50*, 176–188.
11. Chai, Y.; Li, H.; Li, Z. Multifocus image fusion scheme using focused region detection and multiresolution. *Opt. Commun.* **2011**, *284*, 4376–4389.
12. Tsai, W.T.; Qi, G. Integrated fault detection and test algebra for combinatorial testing in TaaS (Testing-as-a-Service). *Simul. Model. Pract. Theory* **2016**, *68*, 108–124.
13. Bhutada, G.G.; Anand, R.S.; Saxena, S.C. Edge preserved image enhancement using adaptive fusion of images denoised by wavelet and curvelet transform. *Dig. Signal Process.* **2011**, *21*, 118–130.
14. Kim, M.; Han, D.K.; Ko, H. Joint patch clustering-based dictionary learning for multimodal image fusion. *Inf. Fusion* **2016**, *27*, 198–214.
15. Keerqinhu; Qi, G.; Tsai, W.; Hong, Y.; Wang, W.; Hou, G.; Zhu, Z. Fault-Diagnosis for Reciprocating Compressors Using Big Data. In Proceedings of the Second IEEE International Conference on Big Data Computing Service and Applications, BigDataService 2016, Oxford, UK, 29 March–1 April 2016; pp. 72–81.
16. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K.O. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095.
17. Dong, W.; Zhang, L.; Shi, G.; Wu, X. Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization. *IEEE Trans. Image Process.* **2011**, *20*, 1838–1857.
18. Xu, Z.; Sun, J. Image Inpainting by Patch Propagation Using Patch Sparsity. *IEEE Trans. Image Process.* **2010**, *19*, 1153–1165.
19. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image super-resolution as sparse representation of raw image patches. In Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, AK, USA, 24–26 June 2008.
20. Li, H.; Li, L.; Zhang, J. Multi-focus image fusion based on sparse feature matrix decomposition and morphological filtering. *Opt. Commun.* **2015**, *342*, 1–11.
21. Yang, B.; Li, S. Multifocus Image Fusion and Restoration With Sparse Representation. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 884–892.
22. Wang, J.; Liu, H.; He, N. Exposure fusion based on sparse representation using approximate K-SVD. *Neurocomputing* **2014**, *135*, 145–154.
23. Zhang, A.; Fu, Y.; Li, H.; Zou, J. Dictionary learning method for joint sparse representation-based image fusion. *Opt. Eng.* **2013**, *52*, 532–543.

24. Lin, B.; Li, Q.; Sun, Q.; Lai, M.; Davidson, I.; Fan, W.; Ye, J. Stochastic Coordinate Coding and Its Application for Drosophila Gene Expression Pattern Annotation. *arXiv* **2014**, arXiv:1407.8147.
25. Rodriguez, A.; Laio, A. Clustering by fast search and find of density peaks. *Science* **2014**, *344*, 1492–1496.
26. Tsai, W.; Colbourn, C.J.; Luo, J.; Qi, G.; Li, Q.; Bai, X. Test algebra for combinatorial testing. In Proceedings of the 8th IEEE International Workshop on Automation of Software Test, AST 2013, San Francisco, CA, USA, 18–19 May 2013; pp. 19–25.
27. Combettes, P.L.; Wajs, V.R. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **2005**, *4*, 1168–1200.
28. Wu, W.; Tsai, W.; Jin, C.; Qi, G.; Luo, J. Test-Algebra Execution in a Cloud Environment. Proceedings of 8th IEEE International Symposium on Service Oriented System Engineering, SOSE 2014, Oxford, UK, 7–11 April 2014; pp. 59–69.
29. Image Fusion Examples. Available online: <http://www.imagefusion.org> (accessed on 27 July 2016).
30. Zheng, Y.; Essock, E.A.; Hansen, B.C.; Haun, A.M. A new metric based on extended spatial frequency and its application to DWT based fusion algorithms. *Inf. Fusion* **2007**, *8*, 177–192.
31. Anantrasirichai, N.; Achim, A.; Bull, D.; Kingsbury, N. Mitigating the effects of atmospheric distortion using DT-CWT fusion. In Proceedings of the 19th IEEE International Conference on Image Processing (ICIP), Orlando, FL, USA, 30 September–3 October 2012; pp. 3033–3036.
32. Chen, H.; Huang, Z. Medical Image Feature Extraction and Fusion Algorithm Based on K-SVD. In Proceedings of the 9th IEEE International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), Guangzhou, China, 8–10 November 2014; pp. 333–337.
33. Dixon, T.D.; Canga, E.F.; Nikolov, S.G.; Troscianko, T.; Noyes, J.M.; Canagarajah, C.N.; Bull, D.R. Selection of image fusion quality measures: objective, subjective, and metric assessment. *J. Opt. Soc. Am. A* **2007**, *24*, B125–B135.
34. Qu, G.; Zhang, D.; Yan, P. Information measure for performance of image fusion. *Electron. Lett.* **2002**, *38*, 313–315.
35. Xydeas, C.; Petrović, V. Objective image fusion performance measure. *Electron. Lett.* **2000**, *36*, 308–309.
36. Tsai, W.T.; Qi, G.; Zhu, Z. Scalable SaaS Indexing Algorithms with Automated Redundancy and Recovery Management. *Int. J. Softw. Inform.* **2013**, *7*, 63–84.
37. Sheikh, H.; Bovik, A. Image information and visual quality. *IEEE Trans. Image Process.* **2006**, *15*, 430–444.
38. Block, M.; Schaubert, M.; Wiesel, F.; Rojas, R. Multi-Exposure Document Fusion Based on Edge-Intensities. In Proceedings of the 2009 10th International Conference on Document Analysis and Recognition, Catalonia, Spain, 26–29 July 2009; pp. 136–140.
39. Aharon, M.; Elad, M.; Bruckstein, A. SVDD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *Trans. Signal Proc.* **2006**, *54*, 4311–4322.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).