Exploring a unified low rank representation for multi-focus image fusion

Zhang, Qiang; Wang, Fan; Luo, Yongjiang; Han, Jungong

Published in:
Pattern Recognition

DOI:
10.1016/j.patcog.2020.107752

Publication date:
2020

Citation for published version (APA):
Zhang, Q., Wang, F., Luo, Y., & Han, J. (2020). Exploring a unified low rank representation for multi-focus image fusion. *Pattern Recognition*, [107752]. https://doi.org/10.1016/j.patcog.2020.107752

Exploring a Unified Low Rank Representation for Multi-focus Image Fusion

Qiang Zhang [a, b], Fan Wang [b], Yongjiang Luo [c,*], Jungong Han [d, *]

[a] *Key Laboratory of Electronic Equipment Structure Design, Ministry of Education, Xidian University, Xi'an, Shaanxi 710071, China*

[b] *Center for Complex Systems, School of Mechano-electronic Engineering, Xidian University, Xi'an, Shaanxi 710071, China*

[c] *School of Electronic Engineering, Xidian University, Xi'an, Shaanxi 710071, China*

[d] *Computer Science Department, Aberystwyth University, SY23 3FL, United Kingdom*

Abstract: Recent years have witnessed a trend that uses image representation models, including sparse representation (SR), low-rank representation (LRR) and their variants for multi-focus image fusion. Despite the thrilling preliminary results, existing methods conduct the fusion patch by patch, leading to insufficient consideration of the spatial consistency among the image patches within a local region or an object. As a result, not only the spatial artifacts are easily introduced to the fused image but also the "jagged" artifacts frequently arise on the boundaries between the focused regions and the de-focused regions, which is an inherent problem in these patch-based fusion methods. Aiming to address the above problems, we propose, in this paper, a new multi-focus image fusion method integrating super-pixel clustering and a unified LRR (ULRR) model. The entire algorithm is carried out in three steps. In the first step, the source image is segmented into a few super-pixels with irregular sizes, rather than patches with regular sizes, to diminish the "jagged" artifacts and meanwhile to preserve the boundaries of objects on the fused image. Secondly, a super-pixel clustering-based fusion strategy is employed to further reduce the spatial artifacts in the fused images. This is achieved by using a proposed ULRR model, which imposes the low-rank constraints onto each super-pixel cluster. This is apparently more reasonable for those images with complicated scenes. Moreover, a Laplacian regularization term is incorporated in the proposed ULRR model to ensure the spatial consistency among the super-pixels with the same cluster. Finally, a measure of focus for each super-pixel is defined to seek the focused as well as de-focused regions in the source image via jointly using representation coefficients and sparse errors derived from the proposed ULRR model. Extensive experiments have been conducted and the results demonstrate the superiorities of the proposed fusion method in diminishing the spatial artifacts in the fused image and the "jagged" boundary artifacts between the focused and de-focused regions, compared to the state-of-the-art fusion algorithms.

Keywords: Multi-focus image fusion, Super-pixel clustering, Unified low-rank representation, Spatial consistency.

1. Introduction

Owing to the limited field depth of optical imaging systems, it is usually difficult, if not impossible, to acquire an image with all the objects in-focus [1]. Hence, only parts of an image have sharp appearances while the others look relatively blurring, which brings great inconvenience for human visual perception and sometimes computer processing as well. A lot of technologies are available to remedy this situation, in which multi-focus image fusion is a simple yet efficient way to combine multiple images

* Corresponding author.
E-mail address: jungong.han@aber.ac.uk (J. Han), yjluo@mail.xidian.edu.cn (Yongjiang Luo).

that shoot the same scene at different focal points into a single image, on which all objects are clearly displayed [2].

There are two basic requirements for a multi-focus image fusion method. One is that the focused regions should be determined and extracted from the given multi-focus input images and then preserved into the fused image, while all the defocused regions should be discarded [1]. The other is spatial artifacts or inconsistencies should be introduced to the fused image as little as possible during the fusion. Hence, how to accurately identify the focused and de-focused regions given the source images and how to combine the focused regions organically are two open questions in multi-focus image fusion. Our answer here is a new multi-focus image fusion method that employs super-pixel clustering and unified low-rank representation model.

So far, a number of multi-focus image fusion methods have been presented. A thorough review of these methods can be found in [2]. Among these, the fusion methods based on image representation models, e.g., sparse representation (SR) [3, 4], low rank representation (LRR) [5, 6] and their different extensions [7, 8], have attracted considerable attention in recent years attributed to their flexibilities. Usually, most image representation-based fusion methods are implemented patch by patch. Concretely, they start by dividing the input images into patches with regular shapes and the same sizes, and then carry out the fusion at the patch level.

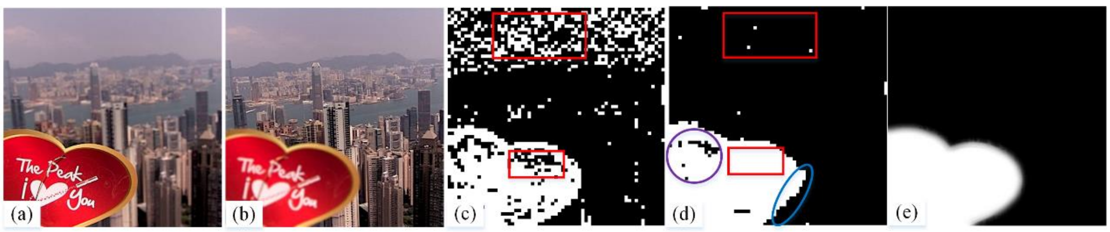

Fig. 1. Illustration of some decision maps. (a) and (b) Source images with focus on the front and the back, respectively; (c) Decision map obtained by [3] without sliding window, where the “white” and “black” points denote that the corresponding regions in the fused image are selected from Fig. 1(a) and Fig. 1(b), respectively; (d) Decision map obtained by [1] with spatial contextual information; (e) Decision map obtained by the proposed method.

However, most of these fusion methods just consider each image patch individually, which ignore the spatial consistency among those image patches within a local region or an object. As a result, some serious spatial artifacts appear on the focus decision maps (or the fused images), as shown in the red rectangular regions of Fig. 1(c).

For that, some fusion strategies have been proposed to suppress block artifacts and enhance the robustness against misregistration, among which the sliding window technology [3] is commonly employed. Despite its acceptable performance, sliding window usually leads to a huge requirement of memory storage as well as the increase of computational complexity. Alternatively, some spatial contexts or spatial consistency based strategies are presented in recent years [1, 7]. As displayed in the red rectangle regions of Fig. 1(d), these newly presented fusion strategies may diminish the spatial artifacts greatly. However, only the spatial consistency among those image patches within a local region is considered and the object area consistency among the patches within an object is ignored in these fusion strategies. Consequently, as shown in the purple circular region of Fig. 1(d), some patches may be still determined to have different focus information from those images in the same object.

In fact, an object in a multi-focus image is generally either wholly in-focus or out-of-focus due to the fact that the camera lens usually focuses on an object when taking a picture. Accordingly, those image patches within the same object may be similar in focus, i.e., they are all in-focus or all out-of-focus.

In addition to those spatial artifacts introduced in the fused image, "jagged" artifacts also arise frequently on the boundaries between the focused regions and the de-focused regions, as shown in the blue elliptical region of Fig. 1(d). This is an inherent problem in the patch-based fusion methods. Besides, it should be noted that most existing methods directly employ the intensity values as the feature for each patch, which are sensitive to the noise or illumination changes, especially, for smooth regions. As shown in the rectangle regions of Fig. 1(c) and Fig. 1 (d) again, those isolated regions usually appear in those smooth regions that containing few details.

In order to address such problems arising in those existing image representation-based fusion methods, we present a super-pixel clustering based multi-focus image fusion method via a unified low-rank representation (ULRR) model. First, the input images are segmented into some super-pixels with irregular shapes rather than patches with fixed shapes to reduce the "jagged" artifacts between the focused and de-focused regions and meanwhile to preserve the boundaries of objects in the fused image. As well, multiple types of features, including colors, edges and textures, are extracted for each super-pixel to boost the focus discrimination.

Secondly, the super-pixels having similar features in each source image are first grouped into different clusters. Then these clusters are represented by using a proposed ULRR model considering the low-rankness (or correlations) of the super-pixels within a cluster. The proposed ULRR model is a sort of improved version of the traditional LRR model [9] by incorporating a Laplacian regularization term with respect to the representation coefficients. Here, the Laplacian regularization term intends to enforce the spatially adjacent super-pixels from the same cluster to be similar in representation coefficients and thus end up having similar focus information.

Finally, a measure of focus (MOF) is defined for each super-pixel by engaging the representation coefficients and sparse errors obtained by ULRR to compute a focus decision map, which in turn guides the fusion procedure of various source images. As displayed in Fig. 1(e), the focused and de-focused regions can be well determined by using the proposed method, on which much fewer spatial artifacts appeared on the fused image, and the "jagged" artifacts between the focused and de-focused regions disappeared. Experimental results verify the superiorities of our proposed fusion method over some state-of-the-arts, even including some deep learning based methods, in diminishing the spatial artifacts in the fused image and the "jagged" boundary artifacts between the focused and de-focused regions.

The main contributions of this paper are highlighted as follows.

(1) A new multi-focus image fusion method based on clustering is proposed, where the spatial consistency among the local regions within an object is considered to reduce the spatial artifacts and to enhance the object area consistency in the fused image. This is different from that in [1] and [7], where only the local consistency among spatially-adjacent patches is considered.

(2) A unified low-rank representation (ULRR) model is proposed to capture the "intrinsic" low-rankness of each super-pixel cluster in our method, ensuring the spatial consistency among adjacent super-pixels within the same cluster or object. In addition, a new dictionary is constructed for ULRR.

(3) The proposed fusion method is implemented super-pixel by super-pixel, rather than in a patch based way as that in most existing SR and LRR based fusion methods, to reduce the "jagged" artifacts between the focused and de-focused regions and meanwhile to preserve the boundaries of objects in the fused image. Moreover, multiple types of features, including colors, edges, and textures, are extracted for each super-pixel to boost the focus discrimination. This is also beyond the traditional SR or LRR based fusion methods, where the intensity values are directly adopted as the features.

The rest of this paper is organized as follows. The related work is briefly introduced in Section 2, while the details of the proposed method are elaborated in Section 3. Experimental results as well as conclusions are given in Section 4 and Section 5, respectively.

2. Related works

So far, tremendous efforts have been devoted to multi-focus image fusion and numerous fusion algorithms have been presented, which fall into two groups: transform domain based methods and spatial domain based methods.

Among the former, multi-scale transform (MST) based fusion methods are the trends and have been discussed over the years [10, 11]. Different from transform domain based ones, spatial domain based methods directly extract the fused images (or image patches) from the source images via some measures of focus (MOFs) [12]. This usually caused many undesirable spatial artifacts. For that, some block and region based fusion methods have been presented in the past few years [13, 14, 15,]. Especially, in [15], a regional approach based on super-pixel segmentation and mean filtering was proposed. Currently, based on image matting [16], guided filtering [17], edge model [18] and conditional random field optimization [19], some new spatial domain based fusion methods have been presented to achieve state-of-the-art performance in information extraction and spatial consistency. A survey on these methods can be seen in [2, 20].

Recently, some new image representation models, such as sparse representation (SR) [3, 4, 21], low rank representation (LRR) [5, 6] and their variants [7, 8] have been employed to image fusion. For example, Yang *et al.* [3] took the first attempt in applying the SR theory to multi-sensor image fusion. In [21], Chen *et al.* introduced a multi-focus image fusion method based on clarity-enhanced image segmentation and regional sparse representation to strengthen its robustness against distortions that usually resulting from the pixel based coefficients selection. In our previous work [7], a robust sparse representation (RSR) based multi-focus image fusion was presented, where information from each local image patch and its spatial contextual information were jointly employed to determine the focused and de-focused regions. A multi-focus image fusion method based on dictionary learning and LRR was presented to achieve good performance in both global and local structures in [5]. The latent low-rank representation was used to extract the salient information of source images and guide the adaptive fusion of low-pass sub-images in [8]. A thorough review and discussion about these fusion methods can be seen in [22].

With the development of deep learning, some multi-focus image fusion methods based on deep neural networks, e.g., convolutional neural networks (CNNs), have been proposed. Early CNN works [23, 24] view the determination of each image patch to be in-focus or out-of-focus as a classification problem. Later, some end-to-end networks are introduced for multi-focus image fusion [25, 26, 27]. Recently, several ensemble learning based multi-focus image fusion methods [28, 29] were presented, where an ensemble of three CNNs were trained on three datasets to predict the decision maps without the need of post-processing steps. Although these deep learning based methods may achieve satisfactory performance, a massive amount of training data with labels are required to train such networks. This is a challenging work for multi-focus image fusion.

3. Proposed method

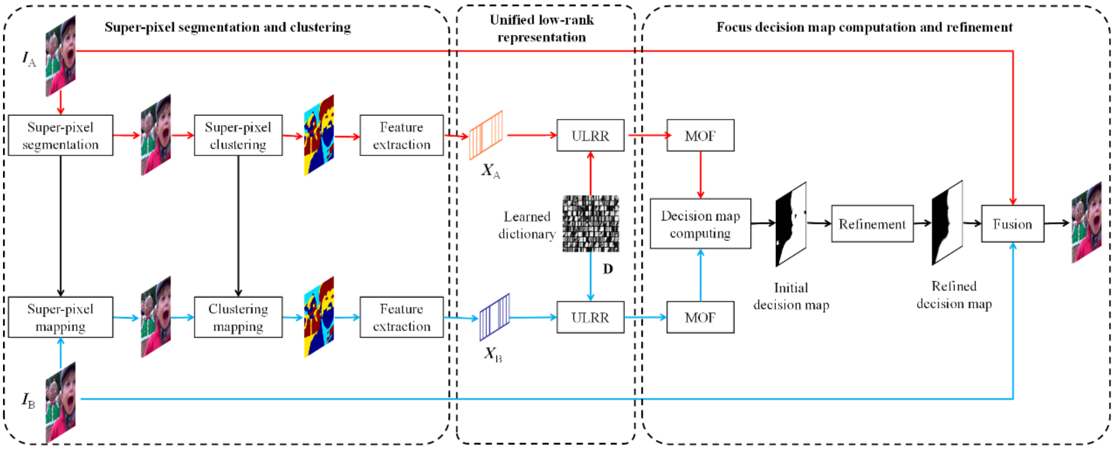

Fig. 2. Diagram of the proposed multi-focus image fusion algorithm.

In this paper, only two source images are taken into account, and the images are supposed to have been well registered in advance. Fig. 2 depicts the diagram of the proposed fusion method, which consists of the following components: (1) Super-pixel segmentation and clustering; (2) Unified low-rank representation; (3) Focus decision map computation and refinement. Based on the focus decision map, the fused image is thus obtained. In addition, a local compact dictionary will be constructed from the average image of each pair of source images, when the source images are decomposed using the proposed ULRR model. In the following subsection, we will describe each component in detail.

3.1. Super-pixel segmentation and clustering

3.1.1. Super-pixel segmentation

Super-pixels group perceptually similar pixels to create visually meaningful entities while heavily reducing the number of primitives for subsequent processing steps [30]. Since they were first named in 2003 [31], super-pixels have been widely applied to many computer vision tasks, including image fusion [15]. Compared with image patches of regular shapes, super-pixels can preferably preserve the boundary of objects in an image. Considering that, we adopt super-pixels, instead of image patches, in our proposed fusion method. So far, many super-pixel algorithms have been proposed, among which, linear spectral clustering (LSC) [32] is shown to achieve higher visual compactness and boundary adherence for natural images but with lower computational costs. Considering that, we adopt LSC for super-pixel segmentation in this paper.

As well, all the source images to be fused should be segmented into the same results so that their corresponding super-pixels can be properly merged in the subsequent fusion process. A commonly used way is to perform super-pixel segmentation on the average image of source images [15], and then map the super-pixel segmentation results to each source image. For multi-focus source images, some undesirable results may be obtained, especially for those transitional regions between focused and de-focused regions. For example, as shown in the rectangle region of Fig. 3(a), parts of the hands in the focused regions have been grouped into the same super-pixel with some de-focused regions. Accordingly, parts of these regions will be mistakenly determined to be focused or de-focused ones and the border of the hands will be destroyed in the fused image.

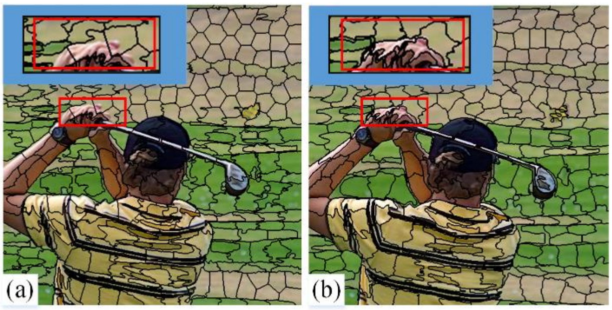

Fig. 3. Illustration of super-pixel segmentation results by performing LSC on different input images. (a) On the average image of multi-focus source images; (b) On one of the multi-focus source images.

Alternatively, we will perform super-pixel segmentation on one of the source images, rather than on the average image, in order to obtain more accurate boundaries between focused and de-focused regions. As shown in the rectangle region of Fig. 3(b), each part of the hands is grouped into a super-pixel. In the subsequent fusion process, the border of the hands will be well preserved.

In summary, given a pair of source images, denoted by I_A and I_B, respectively, we first perform LSC on I_A to obtain a set of super-pixels $\{sp_{A,i} \mid i = 1, 2, \cdots, N\}$, where N denotes the total number of super-pixels and is experimentally set to 350 in this paper. Then we map the segmentation results on I_B and obtain $\{sp_{B,i} \mid i = 1, 2, \cdots, N\}$.

3.1.2. Super-pixel clustering

In general, a super-pixel only denotes a regional atom without any perceptual meaning. Accordingly, as shown in Fig. 4 (c), each object in an image and the background may be constructed by many super-pixels with similar features. In real applications, we usually focus the lens on one object in the scene when taking a picture. As a result of that, the super-pixels within the same object in a multi-focus image may be all in-focus or all out-of-focus with a large probability. When the fusion is directly performed on super-pixels, the super-pixels within the same object may be mistakenly determined to have different focus information from the others. Some spatial artifacts may thus be easily introduced to the fused image.

Fig. 4. Super-pixel clusters. (a) and (b) Source images with focus on the front and the back, respectively; (c) Results of super-pixel segmentation; (d) Results of super-pixel clustering.

In order to address such problem, we will first group the super-pixels in each source image into different clusters and then consider the spatial consistency among the super-pixels within the same cluster to introduce fewer spatial artifacts to the fused image. Similar to that in super-pixel segmentation, we just group the super-pixels in one of the source images and then map the clustering results to the other source image to ensure that the two multi-focus source images have the same clustering results.

In this paper, because of its popularity and simplicity, we adopt the k-means algorithm [33] to achieve the super-pixel clustering, where the averaging RGB color values of all the pixels in each super-pixel are employed as the super-pixel feature. Specifically, given the two source images $I_{A/B}$ [1] and their corresponding two sets of super-pixels $\{sp_{A/B,i} \mid i = 1, 2, \cdots, N\}$, two sets of super-pixel clusters $\{C_{A/B,k} \mid k = 1, 2, \cdots, K\}$ are obtained, where K denotes the number of clusters and will be discussed in the experimental part. And each cluster $C_{A/B,k}$ contains N_k super-pixels, i.e. $C_{A/B,k} = \{sp_{A/B,k,i} \mid i = 1, 2, ..., N_k\}$. As shown in Fig. 4 (d), each object in the image is segmented into only a fewer number of clusters, which will facilitate the consistency among the super-pixels within the same object in the subsequent fusion process.

3.1.3. Feature extraction

In most SR or LRR based fusion methods, pixel intensity values are often directly employed as the

---

[1]The symbol A/B in $I_{A/B}$ denotes A or B, i.e., $I_{A/B}$ means I_A or I_B. In the following contents, the definition of similar symbols is the same.

features, which are sensitive to noise or illumination changes. Some regions, especially those smooth regions, are easily mistakenly determined to be in-focus or out-of-focus. In view of this, we extract multiple types of features, including colors, edges and textures, rather than just the intensity, for each super-pixel in our proposed fusion method. Specifically, feature extraction for each super-pixel and super-pixel cluster can be described as follows.

(1) For each pixel $p_{A/B,j}$ in one of the source images, construct its feature vector $\mathbf{v}_{A/B,j} \in R^d$ of dimension $d = 44$, including colors, edge and texture features. RGB color values as well as HIS (Hue, Saturation, Intensity) components are extracted for each pixel, producing 6-dimensional color features. For edge features, high pass filter, discrete wavelet and several edge operators (LOG, Prewitt, Sobel, *et al.*) are performed onto the image, yielding 18-dimension filter responses at each location. Texture features, which are constituted by the gray level co-occurrence matrix [34] of each super-pixel, contain a total of 20-dimensional features, including contrast, energy, homogeneity, dissimilarity and difference entropy.

(2) Construct the feature vector $\mathbf{x}_{A/B,i} \in R^d$ for each super-pixel $sp_{A/B,i}$ by averaging all the feature vectors of pixels contained in the current super-pixel, i.e., $\mathbf{x}_{A/B,i} = \frac{1}{N_{sp_i}} \sum_{p_j \in sp_i} \mathbf{v}_{A/B,j}$, where N_{sp_i} stands for the total number of pixels contained in the super-pixel $sp_{A/B,i}$.

(3) Construct the feature matrix $\mathbf{X}_{A/B,k} \in R^{d \times N_k}$ for each super-pixel cluster $C_{A/B,k}$ by using all of the vectors of the super-pixels in the same cluster, i.e.,

$$\mathbf{X}_{A,k} = [\mathbf{x}_{A,k,1}, \mathbf{x}_{A,k,2}, \cdots, \mathbf{x}_{A,k,N_k}], \tag{1}$$

$$\mathbf{X}_{B,k} = [\mathbf{x}_{B,k,1}, \mathbf{x}_{B,k,2}, \cdots, \mathbf{x}_{B,k,N_k}], \tag{2}$$

where $\mathbf{x}_{A/B,k,i}$ denotes the feature vector of the *i*-th super-pixel $sp_{A/B,k,i}$ in the cluster $C_{A/B,k}$.

3.2. Proposed unified low-rank representation (ULRR) for super-pixel clusters

As shown in Fig. 4 (d), each super-pixel cluster represents a part of an object or a local region having similar appearances in the scene. Therefore, the super-pixels from the same cluster are likely to be similar. Accordingly, the feature matrix $\mathbf{X}_k$ [2] for each super-pixel cluster constructed in the previous subsection 3.1.3 has "intrinsic" property of low-rankness. Therefore, a low-rank representation (LRR) model is a natural choice for capturing the "intrinsic" low-rankness of each super-pixel cluster in our proposed fusion method.

As a powerful analytical tool, LRR [9] intends to recover low-rank structures from the data corrupted by sparse but strong noise. We may directly perform LRR on the feature matrix $\mathbf{X}_k$, but this would ignore the spatial consistency among the super-pixels within the same cluster.

As discussed in the previous subsection 3.1.2, the spatially adjacent super-pixels residing within the same cluster may have similar focus information, i.e., they may be all in-focus or all out-of-focus. Therefore, these super-pixels will have similar "intrinsic" property, i.e., they may have similar representation coefficients via LRR. Motivated by that, we think of a unified low-rank representation (ULRR) model by incorporating a Laplacian regularization term with respect to the representation coefficients into the traditional LRR model to capture the low-rankness of each super-pixel cluster.

3.2.1. Unified low-rank representation model

Given a dictionary $\mathbf{D} \in R^{d \times M}$ with M atoms of dimension d and the feature matrix $\mathbf{X}_k \in R^{d \times N_k}$ ($k = 1, 2, ..., K$) for each super-pixel cluster C_k, the unified low-rank representation model is mathematically defined by

$$\min_{\substack{\mathbf{Z}_1, \cdots \mathbf{Z}_k \\ \mathbf{E}_1, \cdots \mathbf{E}_k}} \sum_{k=1}^{K} \left\| \mathbf{Z}_k \right\|_* + \alpha \left\| \mathbf{E} \right\|_{2,1} + \beta \, \mathrm{tr}(\mathbf{Z} \mathbf{L} \mathbf{Z}^T), \;\; s.t. \;\; \mathbf{X}_k = \mathbf{D} \mathbf{Z}_k + \mathbf{E}_k \;, k = 1, 2, \cdots, K \;, \tag{3}$$

[2]In this subsection, we remove the symbol A/B from $\mathbf{X}_{A/B,k}$ for generality.

where $\mathbf{DZ}_k$ denotes the "intrinsic" low-rank part contained in the matrix $\mathbf{X}_k$. $\mathbf{Z}_k \in R^{M\times N_k}$ denotes the representation coefficient matrix to be sought. $\mathbf{E}_k \in R^{d\times N_k}$ represents the error or noise part. $\|\mathbf{Z}_k\|_*$ indicates the nuclear norm of the matrix $\mathbf{Z}_k$ and is a convex relaxation of the rank function. The matrices $\mathbf{Z}\in R^{M\times N}$ and $\mathbf{E}\in R^{d\times N}$ are constructed by $\mathbf{Z}=[\mathbf{Z}_1,\mathbf{Z}_2,...,\mathbf{Z}_K]$ and $\mathbf{E}=[\mathbf{E}_1,\mathbf{E}_2,...,\mathbf{E}_K]$, respectively. $\|\mathbf{E}\|_{2,1}$ denotes the $l_{2,1}$-norm of the matrix $\mathbf{E}$. Minimizing $\|\mathbf{E}\|_{2,1}$ enforces the matrix $\mathbf{E}$ to have column-sparsity. α and β are two positive trade-off parameters to balance the effect of each part.

The Laplacian regularization term $\mathrm{tr}(\mathbf{ZLZ}^T)$ in Eq. (3) is defined by

$$\mathrm{tr}(\mathbf{ZLZ}^T)=\frac{1}{2}\sum_{i,j}\|\mathbf{z}_i-\mathbf{z}_j\|_2^2\,\omega_{i,j}\ , \tag{4}$$

where $\mathbf{z}_i$ denotes the i-th column of $\mathbf{Z}$. The weight $\omega_{i,j}$ refers to the similarity between the i-th and j-th super-pixels sp_i and sp_j, and is computed by

$$\omega_{i,j}=\begin{cases}\exp(-\dfrac{\|\mathbf{x}_i-\mathbf{x}_j\|_2^2}{2\sigma^2}), & \text{if } sp_i \text{ and } sp_j \text{ are spatially adjacent and belong to the same cluster}\\ 0, & \text{otherwise}\end{cases}. \tag{5}$$

Here, $\mathbf{x}_i$ and $\mathbf{x}_j$ are the feature vectors of sp_i and sp_j, respectively. σ is a scalar parameter and we experimentally set it to $\sqrt{0.5}$. Given these weights, an affinity matrix $\mathbf{W}\in R^{N\times N}$ with its (i, j)-th entry $\mathbf{W}_{i,j}=\omega_{i,j}$ and a diagonal degree matrix $\mathbf{\Lambda}\in R^{N\times N}$ with its i-th diagonal element $\mathbf{\Lambda}_{i,i}=\sum_j \mathbf{W}_{i,j}$ can be constructed. The Laplacian matrix $\mathbf{L}\in R^{N\times N}$ is then defined as $\mathbf{L}=\mathbf{\Lambda}-\mathbf{W}$.

Eq. (3) presents a convex optimization problem that can be solved by various methods. For that, we first convert it to the below equivalent one by involving some auxiliary variables in this paper:

$$\min_{\substack{\mathbf{Z}_1,\cdots\mathbf{Z}_k,\\ \mathbf{E}_1,\cdots\mathbf{E}_k}}\sum_{k=1}^{K}\|\mathbf{J}_k\|_*+\alpha\|\mathbf{E}\|_{2,1}+\beta\,\mathrm{tr}(\mathbf{HLH}^T),\ \ s.t.\ \mathbf{X}_k=\mathbf{DZ}_k+\mathbf{E}_k\ , \mathbf{Z}_k=\mathbf{J}_k, \mathbf{Z}=\mathbf{H}\ . \tag{6}$$

To solve it, a linearized alternating direction method with adaptive penalty (LADMAP) [35] is adopted, which requires minimizing the following augmented Lagrangian function

$$L=\sum_{k=1}^{K}\left(\left\|\mathbf{J}_k\right\|_*+\left\langle\mathbf{Y}_{1,k},\mathbf{X}_k-\mathbf{D}\mathbf{Z}_k-\mathbf{E}_k\right\rangle+\left\langle\mathbf{Y}_{2,k},\mathbf{Z}_k-\mathbf{J}_k\right\rangle+\frac{\mu}{2}\left\|\mathbf{X}_k-\mathbf{D}\mathbf{Z}_k-\mathbf{E}_k\right\|_F^2+\frac{\mu}{2}\left\|\mathbf{Z}_k-\mathbf{J}_k\right\|_F^2\right)$$
$$+\alpha\left\|\mathbf{E}\right\|_{2,1}+\beta\,\mathrm{tr}(\mathbf{H}\mathbf{L}\mathbf{H}^T)+\left\langle\mathbf{Y}_3,\mathbf{Z}-\mathbf{H}\right\rangle+\frac{\mu}{2}\left\|\mathbf{Z}-\mathbf{H}\right\|_F^2, \quad (7)$$

where Lagrange multipliers $\mathbf{Y}_{1,k}$, $\mathbf{Y}_{2,k}$ $(k=1,2,...,K)$ and $\mathbf{Y}_3$ help to remove the equality constraint in Eq. (6). $\mu>0$ is a penalty term. $\langle\mathbf{A},\mathbf{B}\rangle$ represents the Euclidean inner product of $\mathbf{A}$ and $\mathbf{B}$. This problem can thus be minimized with respect to $\mathbf{Z}_k$ (or $\mathbf{Z}$), $\mathbf{E}_k$ (or $\mathbf{E}$), $\mathbf{Y}_{1,k}$, $\mathbf{Y}_{2,k}$ $(k=1,2,...,K)$, $\mathbf{Y}_3$, and $\mathbf{H}$, respectively. Algorithm 1 briefly summarizes how we calculate the proposed ULRR, and more details are explained in Appendix A.

Algorithm 1: Solving ULRR via LADMAP

Input: Observed data $\mathbf{X}_k\,(k=1,2,...,K)$, dictionary $\mathbf{D}$, and parameters α and β

Output: $\mathbf{Z}$ and $\mathbf{E}$

Initialization: $\mathbf{Z}^0=\mathbf{0}$, $\mathbf{E}^0=\mathbf{0}$, $\mathbf{J}_k^0=\mathbf{0}$, $\mathbf{H}^0=0$, $\mathbf{Y}_1^0=0$, $\mathbf{Y}_2^0=0$, $\mathbf{Y}_3^0=0$, $\mu^0=10^{-6}$, $\mu_{\max}=10^6$, $\varphi=1.1$

While not converged **do**

(1) Fix the others and update $\mathbf{J}_k\,(k=1,2,..,K)$ using Eq. (A2);

(2) Fix the others and update $\mathbf{H}$ using Eq. (A4);

(3) Fix the others and update $\mathbf{Z}_k\,(k=1,2,..,K)$ and $\mathbf{Z}$ using Eq. (A6);

(4) Fix the others and update $\mathbf{E}$ using Eq. (A8);

(5) Update the multipliers $\mathbf{Y}_{1,k}$, $\mathbf{Y}_{2,k}\,(k=1,2,..,K)$ and $\mathbf{Y}_3$:

$$\mathbf{Y}_{1,k}^{i+1}=\mathbf{Y}_{1,k}^{i}+\mu(\mathbf{X}_k-\mathbf{D}\mathbf{Z}_k^{i+1}-\mathbf{E}_k^{i+1}),\quad \mathbf{Y}_{2,k}^{i+1}=\mathbf{Y}_{2,k}^{\ i}+\mu^i(\mathbf{Z}_k^{i+1}-\mathbf{J}_k^{i+1}),\ \mathbf{Y}_3^{i+1}=\mathbf{Y}_3^{i}+\mu^i(\mathbf{Z}^{i+1}-\mathbf{H}^{i+1});$$

(6) Update μ:

$$\mu^{i+1}=\min(\mu^i\varphi,\mu_{\max});$$

(7) Check the convergence conditions:

$$\max_k\left\|\mathbf{X}_k-\mathbf{D}\mathbf{Z}_k^{i+1}-\mathbf{E}_k^{i+1}\right\|_\infty<\varepsilon,\ \left\|\mathbf{Z}^{i+1}-\mathbf{Z}^i\right\|_\infty<\varepsilon,\ \text{and}\ \left\|\mathbf{E}^{i+1}-\mathbf{E}^i\right\|_\infty<\varepsilon;$$

where $\|\cdot\|_\infty$ denotes the l_∞-norm of a matrix.

end while

3.2.2. Dictionary construction

In addition to the ULRR model, the dictionary is also crucial to fusion success. The original feature matrices (e.g., $\mathbf{X}_{A,k}$ or $\mathbf{X}_{B,k}$) from each source image may be directly employed as the dictionary [9] for ULRR. However, it is difficult to maintain the fairness of focus measure for the corresponding super-pixels from different source images. Alternatively, an adaptive dictionary is constructed from an image obtained by averaging each pair of source images in our proposed fusion method when decomposing $\mathbf{X}_{A,k}$ and $\mathbf{X}_{B,k}$. Moreover, as discussed in [36], the dictionary with fairly low-rank is more desirable for LRR. Considering that, we will perform a Gaussian filtering on the average image before constructing the dictionary. Specifically, the dictionary for ULRR is constructed as in Algorithm 2.

Algorithm2: Dictionary construction

(1) For each pair of source images I_A and I_B, an image $\overline{I}_{AB}$ is obtained by averaging the source images;

(2) A blurred average image $\overline{I}'_{AB}$ is obtained by performing a Gaussian filtering with kernel size of 8×8 on $\overline{I}_{AB}$.

(3) The super-pixel segmentation result for I_A (or I_B) is mapped to $\overline{I}'_{AB}$, obtaining a set of super-pixels $\left\{sp_{AB,i} \mid i=1,2,...,N\right\}$.

(4) The features $\left\{\mathbf{x}_{AB,i} \in R^d \mid i=1,2,...,N\right\}$ are extracted for the super-pixels $\left\{sp_{AB,i} \mid i=1,2,...,N\right\}$ by using the same way as in Subsection 3.1.3. A feature matrix is thus constructed by $\mathbf{X}_{AB}=\left[\mathbf{x}_{AB,1},\mathbf{x}_{AB,2},...,\mathbf{x}_{AB,N}\right]\in R^{d\times N}$.

(5) A set of eigenvalues $\{\lambda_i \mid \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0, i=1,2,...,N\}$, sorted in descending order, and their corresponding eigenvectors $\{\boldsymbol{\rho}_i \mid i=1,2,...,N\}$ are obtained by performing principal component analysis (PCA) on the matrix $\mathbf{X}_{AB}$.

(6) A compact dictionary with M atoms of dimension d is constructed by the eigenvectors $\{\boldsymbol{\rho}_i \mid i=1,2,...,M\}$ corresponding to the first M largest eigenvalues, i.e., $\mathbf{D}=[\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\cdots,\boldsymbol{\rho}_M]\in \mathrm{R}^{d\times M}$, where M is experimentally set to 64 in this paper.

As discussed above, a dictionary is adaptively constructed from each pair of multi-focus source images. The adaptability will improve the representation ability of each constructed dictionary, which will be validated in the subsequent experimental part. In addition, the number of atoms in the dictionary gets reduced by using PCA. This also speeds up computation of the proposed fusion method.

3.3. Focus decision map computation and refinement

Given the constructed dictionary $\mathbf{D}$ and a set of feature matrices $\left\{\mathbf{X}_{A,k} | k=1,2,..,K\right\}$ extracted from a source image I_A, a representation coefficient matrix $\mathbf{Z}_A$ and a sparse error matrix $\mathbf{E}_A$ are obtained by solving Eq. (3). Similarly, a representation coefficient matrix $\mathbf{Z}_B$ and a sparse error matrix $\mathbf{E}_B$ for the source image I_B are obtained.

As shown in Fig. 5, the super-pixels in the focused regions normally have larger representation coefficient magnitudes as well as sparse errors, especially larger representation coefficient magnitudes, than those super-pixels in the de-focused regions. Therefore, the focused and de-focused regions in a multi-focus image can be determined by jointly using the representation coefficients and sparse errors.

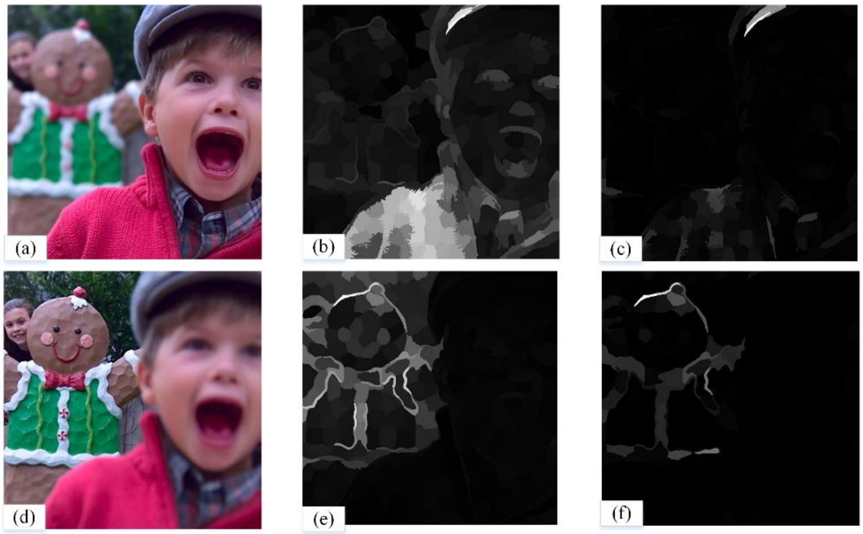

Fig. 5. Illustration of the ULRR results on a pair of multi-focus source images. (a), (d) Source images with focus on the front and the back, respectively; (b), (e)Representation coefficients obtained by ULRR for (a) and (d), respectively; (c), (f) Sparse errors obtained by ULRR for (a) and (d), respectively. For better displaying, each super-pixel in the source image is replaced by the l_2-norm of its corresponding column vector in the representation coefficient matrix and spare error matrix.

For that, a measure of focus (MOF) for the i-th super-pixel $sp_{A/B,i}$ is first defined by:

$$MOF_{A/B,i} = \eta \left\| \mathbf{z}_{A/B,i} \right\|_2 + (1-\eta) \left\| \mathbf{e}_{A/B,i} \right\|_2 , \tag{8}$$

where $\mathbf{z}_{A/B,i}$ and $\mathbf{e}_{A/B,i}$ are the i-th column of $\mathbf{Z}_{A/B}$ and $\mathbf{E}_{A/B}$, respectively. $\left\| \cdot \right\|_2$ denotes the l_2-norm of a vector. η is experimentally set to 0.95 in this paper.

Then an initial focus decision map Υ of the same size as the source images is defined and its each element $\Upsilon(x,y)$ is computed by

$$\Upsilon(x,y) = \begin{cases} 1, & (x,y) \in sp_{A/B,i} \ \& \ MOF_{A,i} \geq MOF_{B,i} \\ 0, & \text{otherwise} \end{cases} . \tag{9}$$

Fig. 6(c) illustrates an initial focus decision map obtained by using Eq. (9). It can be obviously found that most of the focused regions and de-focused regions can be accurately determined by using the proposed MOF defined by Eq. (8). In addition, the boundaries between the focused regions and the de-focused regions are naturally preserved, and few “jagged” artifacts are introduced because of the super-pixel segmentation.

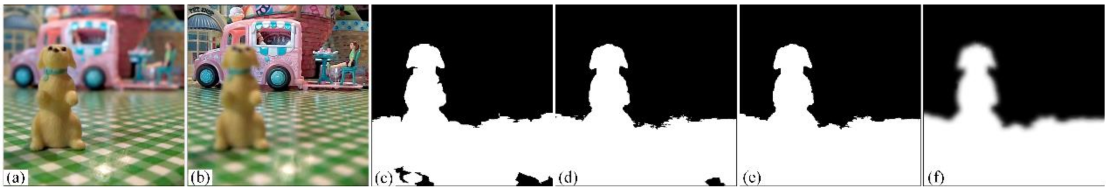

Fig 6. Illustration of the decision maps obtained by different post-processing. (a) and (b) A pair of multi-focus images with focus on the front and the back, respectively; (c) Initial focus decision map obtained by using Eq. (9); (d) Decision map after image matting; (e) Decision map after removing “holes”; (f) Final decision map after guided filtering.

In spite of that, some isolate regions still exist, as shown in Fig. 6(c). To address such problems, some post-processing is further performed on the initial focus map, which includes: (1). Image matting [37] to refine the boundary between the focused and de-focused region; (2). Removing holes [7] to erase small isolate regions, i.e., a region smaller than an area threshold is reversed in the binary initial decision map; (3). Guided filtering [17] to reduce the spatial artifacts between focused and de-focused regions. After that, a refined focus decision map Υ' is finally obtained.

Fig. 6 (d) indicates that the boundary accuracy between the focused and de-focused regions is improved to some extent by using image matting. Some isolated regions are also eliminated. After removing holes, some small isolate regions are further removed, as shown in Fig. 6(e). Finally, as shown in Fig. 6(f), some gradual transitional regions are generated between the focused and de-focused regions, which makes the boundaries look more natural.

3.4. Fusion

Given the refined focus decision map Υ', the fused image I_F can thus be obtained by using a ‘weighted averaging’ scheme, i.e.,

$$I_F(x,y) = \Upsilon'(x,y) I_A(x,y) + (1 - \Upsilon'(x,y)) I_B(x,y) . \tag{10}$$

Here, the refined focus decision map Υ' is used as the weighted map. In summary, the proposed fusion method can be described in Algorithm 3.

Algorithm 3:The proposed multi-focus image fusion method based on super-pixel clustering and ULRR

(1) Perform super-pixel segmentation on the source images as described in Subsection 3.1.1, and obtain two sets of super-pixels $\{sp_{A,i} \mid i = 1, 2, ..., N\}$ and $\{sp_{B,i} \mid i = 1, 2, ..., N\}$;

(2) Perform super-pixel clustering on $\{sp_{A,i} \mid i = 1, 2, ..., N\}$ and $\{sp_{B,i} \mid i = 1, 2, ..., N\}$ as described in Subsection 3.1.2, and

obtain two sets of super-pixel clusters $\{C_{A,k} \mid k=1,2,...,K\}$ and $\{C_{B,k} \mid k=1,2,...,K\}$;

(3) Construct the feature matrix for each super-pixel as described in Subsection 3.1.3 and obtain two sets of feature matrices $\{\mathbf{X}_{A,k} \mid k=1,2,...,K\}$ and $\{\mathbf{X}_{B,k} \mid k=1,2,...,K\}$;

(4) Construct the dictionary by using Algorithm2.

(5) Perform ULRR on $\{\mathbf{X}_{A,k} \mid k=1,2,...,K\}$ and $\{\mathbf{X}_{B,k} \mid k=1,2,...,K\}$ by using Eq. (3), and obtain representation coefficient and sparse error matrices $\{\mathbf{Z}_A, \mathbf{E}_A\}$ and $\{\mathbf{Z}_B, \mathbf{E}_B\}$ for source images I_A and I_B, respectively;

(6) Compute the focus decision map Υ' by using the matrices $\{\mathbf{Z}_A, \mathbf{Z}_B, \mathbf{E}_A, \mathbf{E}_B\}$ as described in Subsection 3.3;

(7) Construct the fused image I_F by using Eq. (10).

4. Experiment results and analysis

Extensive experiments are conducted to verify the performance of the proposed multi-focus image fusion algorithm, which are organized as: 1) the impacts of several important parameters on the proposed method are investigated; 2) the validities of the constructed dictionaries and the proposed ULRR are carried out; 3) comparisons against some state-of-the-art methods on two public databases; (4) extension to the fusion of triple multi-focus images; (5) some discussions on the proposed fusion method.

4.1. Parameters setting

Here, we use ten pairs of multi-focus source images, which are manually generated from the images in Fig. 7 [38], to investigate how the parameters, including the cluster number *K*, and the trade-off parameters α and β in Eq. (3), affect the proposed method. For that, image matting [37] is performed on each image in Fig. 7 to extract the foreground object regions and the background regions, respectively. Then the foreground regions and the background regions are blurred using a 'Gaussian' low-pass filter, respectively, to obtain a pair of multi-focus images. Thus, ten pairs of manually generated multi-focus source images are obtained. Finally, several fused images are obtained from each pair of these multi-focus images by using the proposed method with different parameters. These fused images are compared against their corresponding 'ideal' images in Fig. 7 using the metrics like mean square error (*MSE*) and difference coefficients (*DC*) [1]. Smaller *MSE* and *DC* values imply higher fusion performance.

The experimental results in Table 1 demonstrate that the fusion performance achieves desirable when the number of clusters *K* is set to 4 or 5. Similarly, the experimental results in Table 2 indicate that the proposed fusion method achieves the best when α is set to 0.05. The performance increases with the decrease of β and keeps almost unchanged when β achieves 0.002. In the following experiments, we set the parameters K, α and β to 4, 0.05, 0.002, respectively.

Fig.7. Original images that used to generate the multi-focus images.

Table 1. Fusion performance with different values of K on the 10 pairs of manually generated multi-focus images.

K	2	3	4	5	6	8
MSE	12.8483	12.5746	11.9715	11.9666	12.3555	13.5860
DC	0.0122	0.0118	0.0115	0.0114	0.0115	0.0120

Table 2. Fusion performance with different values of α and β on the 10 pairs of manually generated multi-focus images.

	α with β=0.002					β with α=0.05				
	0.03	0.04	0.05	0.06	0.08	0.01	0.005	0.002	0. 001	0.0001
MSE	13.5656	13.5622	11.9715	13.4647	13.6323	13.3931	12.5915	11.9715	11.9715	11.9715
DC	0.0119	0.0119	0.0115	0.0125	0.0125	0.0125	0.0118	0.0115	0.0115	0.0115

4.2. Validity of the constructed dictionary

In this subsection, we will investigate the impacts of different dictionaries on the fusion results to verify the constructed dictionary in our proposed fusion method. To do so, six dictionaries are constructed for fusion. The first three dictionaries ($\mathbf{D}_{Ksvd}^{Global}$, $\mathbf{D}_{Ksvd}^{Ave}$ and $\mathbf{D}_{Ksvd}^{Blur}$, for short) are constructed by using K-SVD [39], and the other three dictionaries ($\mathbf{D}_{PCA}^{Global}$, $\mathbf{D}_{PCA}^{Ave}$ and $\mathbf{D}_{PCA}^{Blur}$, for short) are constructed by using PCA. Especially, $\mathbf{D}_{Ksvd}^{Global}$ and $\mathbf{D}_{PCA}^{Global}$ are globally learned from a set of nature images with high spatial resolutions and have 256 dictionary atoms. $\mathbf{D}_{Ksvd}^{Ave}$ and $\mathbf{D}_{PCA}^{Ave}$ are adaptively constructed from the average image of each pair of source images and have 64 dictionary atoms. For that, the source images are first averaged and the feature matrix for the average image is extracted afterwards. Then a dictionary is learned from the feature matrix by using K-SVD or PCA. $\mathbf{D}_{Ksvd}^{Blur}$ and $\mathbf{D}_{PCA}^{Blur}$ (i.e., the employed dictionary in our proposed fusion method) are adaptively constructed from the blurred average image of each pair of source images and also have 64 dictionary atoms.

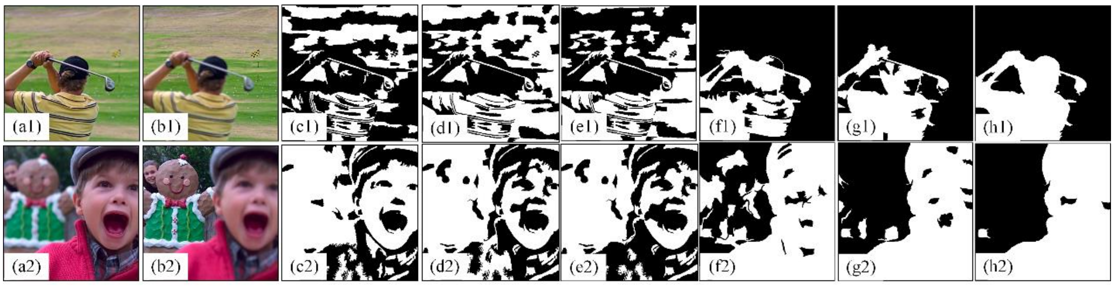

Fig. 8. Fusion results by using different dictionaries. (a1) and (b1) A pair of source images with focus on the front and the back, respectively; (c1) ~ (h1) Initial focus decision maps for (a1) and (b1) obtained by using $\mathbf{D}_{Ksvd}^{Global}$, $\mathbf{D}_{Ksvd}^{Ave}$, $\mathbf{D}_{Ksvd}^{Blur}$, $\mathbf{D}_{PCA}^{Global}$, $\mathbf{D}_{PCA}^{Ave}$ and $\mathbf{D}_{PCA}^{Blur}$, respectively; (a2) ~ (h2) Another pair of source images and their initial focus decision maps obtained by using different dictionaries.

Fig. 8 illustrates the fusion results when using different dictionaries. It is clear that using the dictionaries constructed by K-SVD could not lead the proposed fusion method to achieve desirable results. Furthermore, it is also obvious that the proposed fusion method can achieve better results by using $\mathbf{D}_{PCA}^{Blur}$ than by using $\mathbf{D}_{PCA}^{Ave}$ and $\mathbf{D}_{PCA}^{Global}$. This indicates that the representation coefficients and the sparse errors, especially the representation coefficients, deduced from the proposed ULRR model can better capture the "intrinsic" focus characteristics of a multi-focus image under a dictionary constructed from blurred images than those constructed from clear images.

4.3. Validity of the proposed ULRR model for multi-focus image fusion

In order to test the validity of the proposed ULRR model for multi-focus image fusion, three versions (ULRR_v1, ULRR_v2, ULRR_v3, for short, respectively) of our proposed fusion method just with different LRR models are performed on two pairs of multi-focus source images, shown in Fig. 9. In ULRR_v1, the traditional LRR model [9] is employed, which is directly performed on the feature matrices $\mathbf{X}_{A/B}$ constructed from source image super-pixels rather than on the feature matrices $\left\{\mathbf{X}_{A/B,k} \mid k=1,2,...,K\right\}$ constructed from the source super-pixels clusters. In ULRR_v2, the proposed ULRR model in Eq. (3) without the Laplacian regularization term is employed. In ULRR_v3, i.e., the proposed fusion method, the proposed ULRR model in Eq. (3) with the Laplacian regularization term is employed.

As shown in Fig. 9 (c1) and Fig. 9 (c2), many isolated regions existed in the focus decision maps obtained by using ULRR_v1. Differently, the isolated regions are greatly reduced in the focus decision maps obtained by using ULRR_v2, as shown in Fig. 9 (d1) and Fig. 9 (d2). Especially, as shown in Fig. 9 (e1) and Fig. 9 (e2), the isolated regions are significantly reduced in the focus decision maps when using ULRR_v3. This indicates that performing ULRR on the super-pixel clusters can better capture the "intrinsic" focus information of different regions in a multi-focus image than directly performing LRR on super-pixels. This owes to the consideration of super-pixel clusters in the proposed method via ULRR, especially the spatial consistency among the super-pixels within the same cluster via the Laplacian regularization term in ULRR.

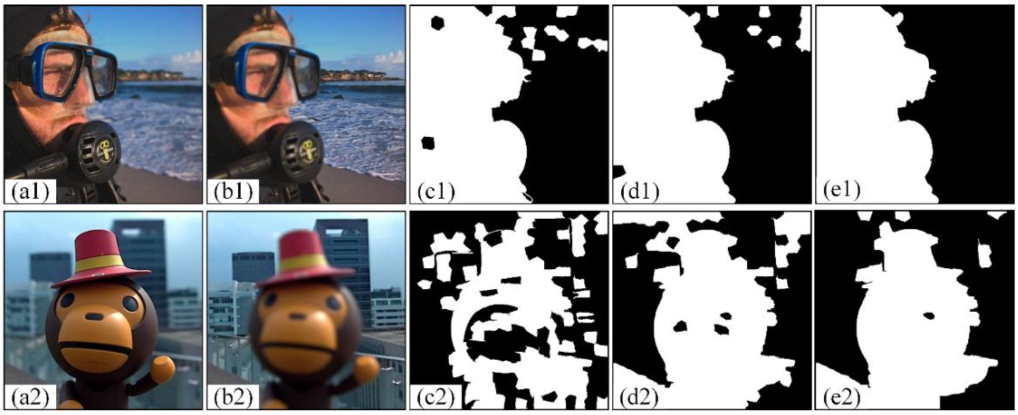

Fig. 9. Illustration of the validity of the proposed ULRR model. (a1) and (b1) A pair of source images with focus on the front and the back, respectively; (c1) ~ (e1) Initial focus decisions maps for (a1) and (b1) obtained by ULRR_v1, ULRR_v2 and ULRR_v3, respectively; (a2) ~ (e2) Another pair of source images and their initial focus decision maps obtained by using different models.

4.4. Comparisons with traditional fusion methods

Here, we compare our method (ULRR, for short) with another 6 traditional state-of-the-art methods, including GFF [17], IM [16], SPixel [15], LR_RSR [7], SRCF [4], and DL_LRR [5]. The public Lytro Dataset in [4] including 20 pairs of multi-focus images and a smaller dataset (SMD, for short) including

10 pairs of multi-focus images that are collected from different kinds of literature are employed to test different fusion methods, which are illustrated in Fig. 10 and Fig. 11, respectively.

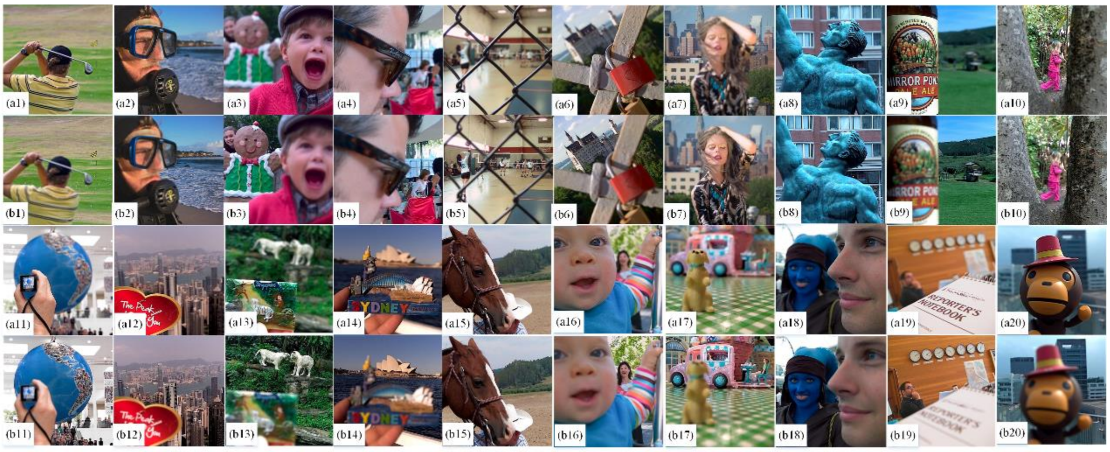

Fig. 10. Lytro Dataset. (a1) ~ (a10) The first 10 input images with the focus on the front part; (b1) ~ (b10) The corresponding input images with the focus on the back part; (a11) ~ (a20) The remaining 10 input images with the focus on the front part; (b11) ~ (b20) The corresponding input images with the focus on the back part.

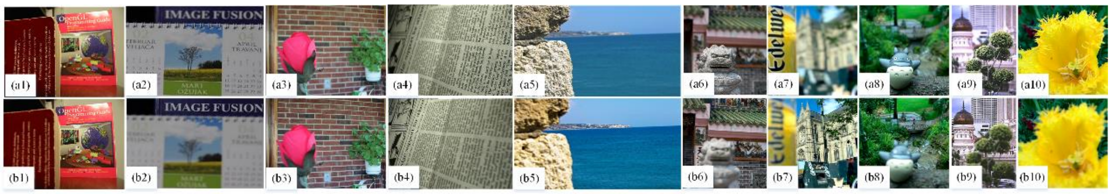

Fig. 11. SMD Dataset. (a1) ~ (a10) The 10 input images with focus on the left (or front) part; (b1) ~ (b10) The corresponding 10 input images with focus on the right (or back) part.

In order to quantitatively compare the fusion performance of different methods, six metrics are used, including mutual information based fusion metric *FMI* [40], universal image quality index based metric Q_{uiqi} [41], quaternion based color image fusion quality metrics Q_{ssim} [42] and Q_4 [43], and phase consistency based metrics Q_{PC} [44] and $ZNCC\_PC$ [45]. The former four metrics measure how well the original information, such as entropy information and structures, from the source images, have been preserved in the fused images. The last two metrics may evaluate different methods in spatial consistency to some extent. Larger values of these metrics are more desirable for a fusion method.

Fig. 12 and Fig. 13 illustrate some fusion results on Lytro Dataset and SMD Dataset obtained by some of the fusion methods mentioned above[3]. In addition, the initial decision maps obtained by using our proposed ULRR method without any post-processing steps are also illustrated in Fig. 12 (i) and Fig. 13(i). Table 3 and Table 4 report the average quantitative results on the two datasets, respectively.

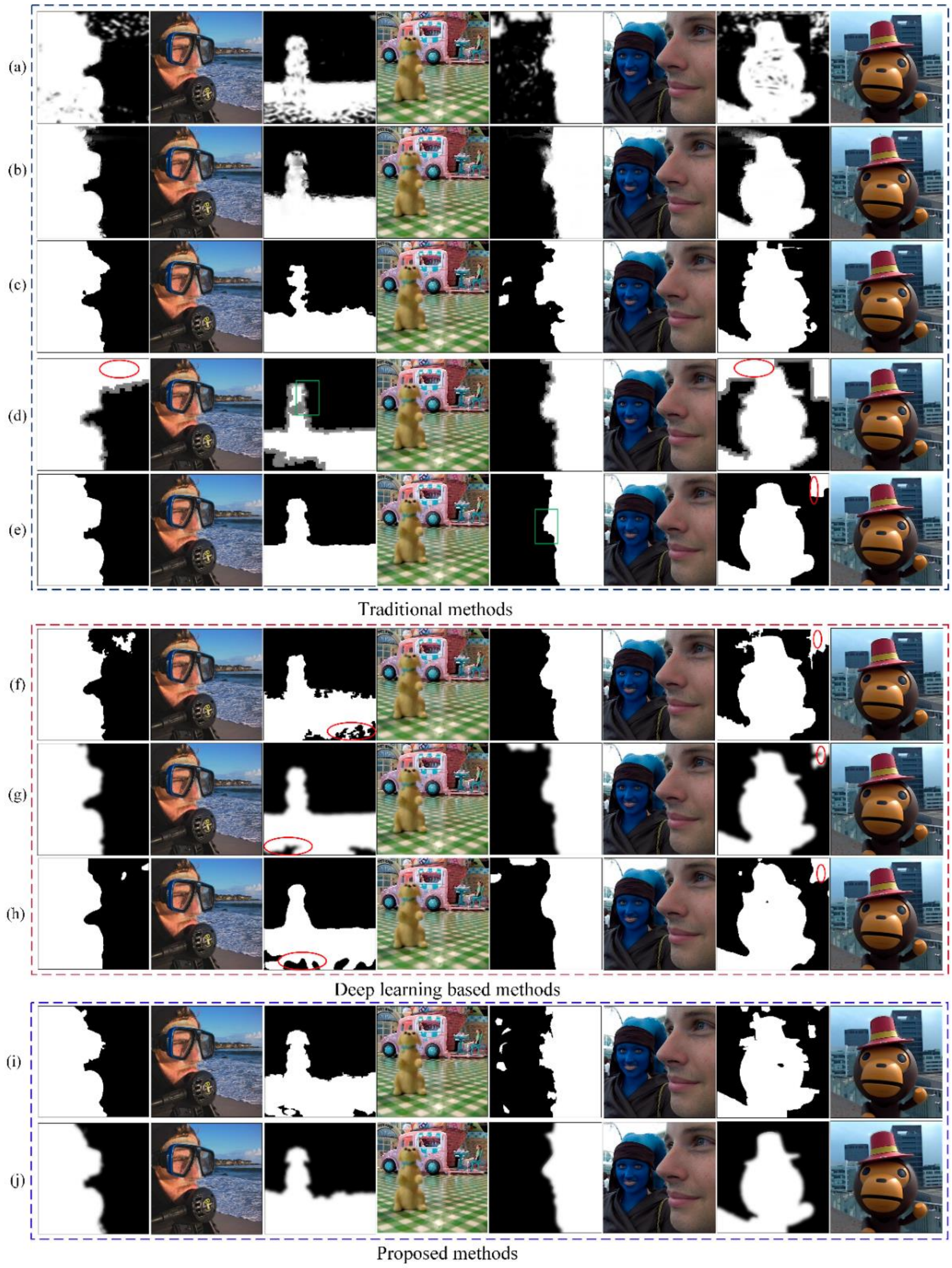

Fig. 12. Some fusion results on Lytro Dataset. (a) GFF; (b) IM; (c) SPixel; (d) LR_RSR; (e) SRCF; (f) p_CNN; (g) CNN; (h) EN_CNN; (i) ULRR without post-processing steps; (j) ULRR. The decision maps in (h) and (i) are initial ones without using any

[3]Given a pair of multi-focus images, DL_LRR directly outputs the finally fused images without using focus decision maps. Therefore, the visual results obtained by DL_LRR are not provided in Fig. 12 and Fig. 13.

post-processing steps. The remaining decision maps are the final ones after the post-processing steps.

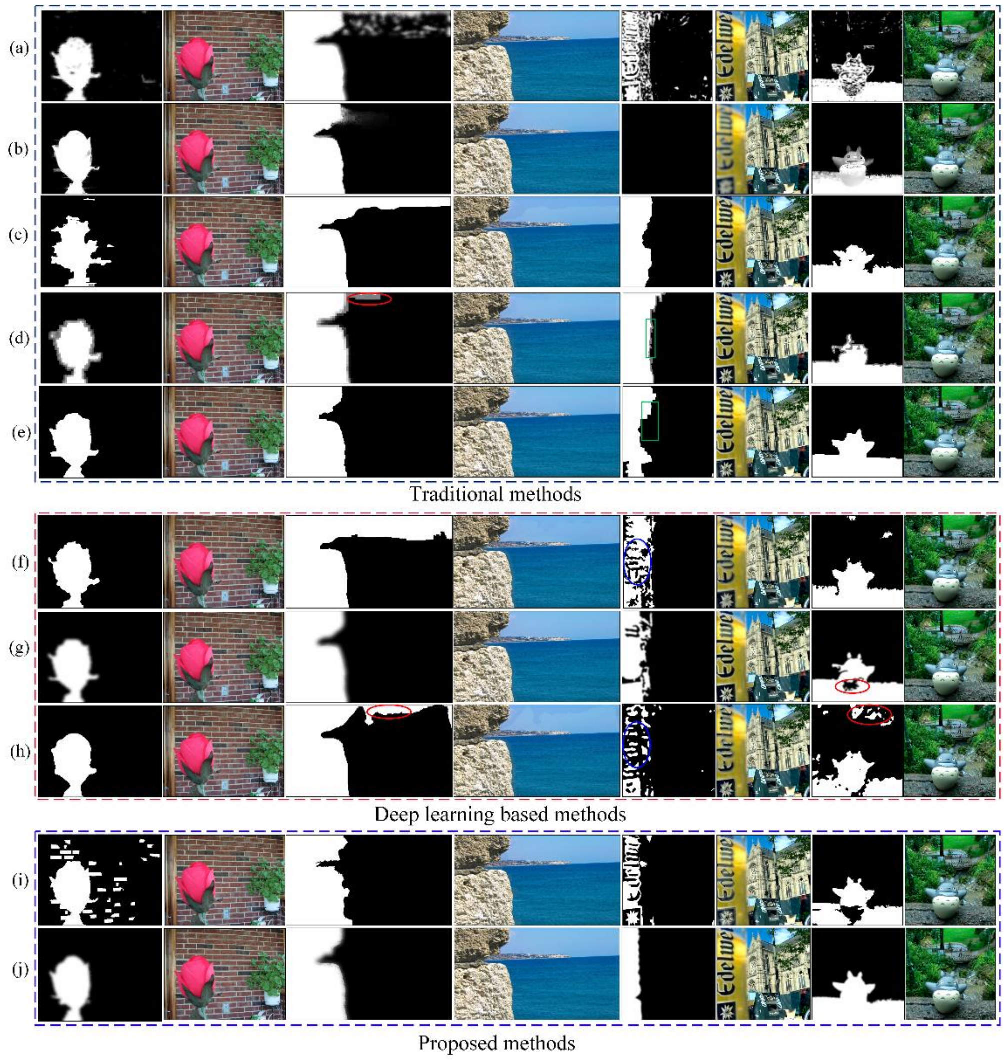

Fig. 13. Some fusion results on SMD Dataset. (a) GFF; (b) IM; (c) SPixel; (d) LR_RSR; (e) SRCF; (f) p_CNN; (g) CNN; (h) EN_CNN; (i) ULRR without post-processing steps; (j) ULRR. The decision maps in (h) and (i) are initial ones without using any post-processing steps. The remaining decision maps are the final ones after the post-processing steps.

From Fig. 12 and Fig. 13, the following facts can be easily observed. Plenty of spatial artifacts appear on the fused images obtained by GFF. IM usually introduces some spatial artifacts on the boundaries between focused and de-focused regions. In most cases, SPixel cannot accurately determine the boundaries between the focused and de-focused regions, although few spatial artifacts are introduced in their decision maps. LR_RSR and SCRF introduce fewer spatial artifacts in their decision maps. But some regions, especially those smooth regions as observed in the red elliptical regions in Fig. 12 (d) and (e), are mistakenly labeled as out-of-focus (or in-focus) by the two methods. As discussed in the earlier part of Section 1, this may be due to the fact that only the intensity values of the input images are employed as the features in LR_RSR and SRCF. Moreover, some “jagged” artifacts exist in the boundaries between the focused and de-focused regions (e.g., the green rectangle regions) of Fig. 12(d), Fig. 12(e), Fig. 13(d) and Fig. 13(e).

Differently, as shown in Fig. 12(j) and Fig. 13(j), almost no isolate regions exist in the decision maps obtained by ULRR. This indicates that the focused and de-focused regions in the source images are better determined by ULRR than the other methods. Accordingly, fewer spatial artifacts are involved in the fused images by using our proposed fusion method. Moreover, no “jagged” artifacts exist in the decision maps obtained by ULRR. The boundaries between the focused and de-focused regions in Fig. 12(j) and Fig. 12(j) look closer to the boundaries of the objects in the source images. The comparisons between Fig. 12 (i) and (j), Fig. 13(i) and (j), indicate that the post-processing steps can partially benefit to the improvements of our proposed method.

Table 3. Averaging performance of different traditional fusion methods on Lytro Dataset.

Methods	FMI	Q_{uiqi}	Q_{ssim}	Q_4	Q_{PC}	$ZNCC\_PC$
GFF	1.4153	0.9092	0.8838	0.9754	0.6817	0.9273
IM	1.4102	0.8932	0.8759	0.9736	0.6225	0.9078
SPixel	1.4194	0.9048	0.8818	0.9746	0.6736	0.9248
LR_RSR	1.4198	0.9072	0.8838	0.9751	0.6763	0.9270
SRCF	1.4195	0.9105	0.8818	0.9747	0.6790	0.9276
DL_LRR	1.4199	0.9015	0.8829	0.9748	0.6347	0.9077
ULRR	1.4197	0.9117	0.8844	0.9756	0.6819	0.9317

Table 4. Averaging performance of different traditional fusion methods on SMD Dataset.

Methods	FMI	Q_{uiqi}	Q_{ssim}	Q_4	Q_{PC}	$ZNCC\_PC$
GFF	1.1831	0.9053	0.7824	0.9103	0.6213	0.9189
IM	1.1931	0.9024	0.7777	0.9053	0.5948	0.9243
SPixel	1.1871	0.9070	0.7779	0.9058	0.6213	0.9287
LR_RSR	1.1852	0.9079	0.7802	0.9076	0.6297	0.9288
SRCF	1.1873	0.9141	0.7783	0.9057	0.6234	0.9283
DL_LRR	1.1860	0.8850	0.7767	0.9055	0.5223	0.8743
ULRR	1.1891	0.9119	0.7815	0.9086	0.6314	0.9356

The quantitative results in Table 3 and Table 4 are in line with the visual results above, which demonstrates that ULRR significantly outperforms the other methods in terms of Q_{PC} and $ZNCC\_PC$. This indicates that our proposed fusion method performs the best in spatial consistency, compared to those methods mentioned here, and fewer spatial artifacts have been introduced to the fused images by using ULRR than by the other methods. Table 3 also demonstrates that ULRR performs the best on Lytro Dataset in terms of Q_{uiqi}, Q_{ssim} and Q_4. Table 4 demonstrates that ULRR always achieves the top two performance on SMD Dataset in terms of FMI, Q_{uiqi}, Q_{ssim} and Q_4. This indicates that, in addition to spatial consistency, our proposed fusion method can also achieve better performance in information extraction than the other methods in most cases.

4.5 Comparisons with deep learning based methods

In addition to those traditional methods, three deep learning (DL) based fusion methods, including CNN [23], p_CNN [24] and EN_CNN [28], are compared with our proposed fusion method. Some visual results on Lytro Dataset and SMD Dataset are also illustrated and provided in Fig. 12 and Fig. 13, respectively. The quality results on the two datasets are provided in Table 5 and Table 6, respectively.

As shown in the first columns of Fig. 12 and Fig. 13, these DL based methods can generally achieve desirable fusion results. Especially, EN_CNN can accurately determine the focused and de-focused regions without using any post-processing steps, thanks to the strong abilities of CNNs for image representation and feature extraction. However, as shown in the red elliptical regions of Fig. 12 and Fig. 13, some smooth regions are also mistakenly determined to be in-focus (or out-of-focus) by these DL based methods. For some regions with abundant textures (e.g., the blue elliptical regions in Fig. 13), these DL based fusion methods could not determine the focused and de-focused regions uniformly. Differently, our proposed fusion method can completely determine the focused and de-focused regions in most cases, as illustrated in Fig. 12 and Fig. 13.

The experimental results in Table 5 and Table 6 indicate that ULRR achieves comparable performance with these DL based fusion methods in terms of FMI , Q_{uiqi} , Q_{ssim} and Q_4 . In terms of Q_{PC} and $ZNCC\_PC$, ULRR performs competitively with CNN and outperforms p_CNN and ES_CNN by a clear margin. This indicates that, even in information extraction, our proposed method still performs competitively with these DL based ones, but in spatial consistency, our proposed method is clearly superior to most of these DL based ones. This further verifies the validity of our proposed fusion method in the reduction of spatial artifacts.

Table 5. Averaging performance of different deep learning based fusion methods on Lytro Dataset.

Methods	FMI	Q_{uiqi}	Q_{ssim}	Q_4	Q_{PC}	$ZNCC\_PC$
CNN	1.4195	0.9111	0.8839	0.9753	0.6851	0.9307
p_CNN	1.4208	0.9091	0.8823	0.9747	0.6749	0.9260
ES_CNN	1.4207	0.9087	0.8816	0.9746	0. 6628	0. 9247
ULRR	1.4197	0.9117	0.8844	0.9756	0.6819	0.9317

Table 6. Averaging performance of different deep learning based fusion methods on SMD Dataset.

Methods	FMI	Q_{uiqi}	Q_{ssim}	Q_4	Q_{PC}	$ZNCC\_PC$
CNN	1.1879	0.9135	0.7815	0.9082	0.6357	0.9318
p_CNN	1.1894	0.8970	0.7808	0.9067	0.5960	0.9102
ES_CNN	1.1902	0.8949	0. 7773	0. 9053	0. 5892	0. 9152
ULRR	1.1891	0.9119	0.7815	0.9086	0.6314	0.9356

4.6. Fusion of more than two multi-focus images

The proposed fusion method can be easily extend to fuse more than two multi-focus images. Suppose that there are total S images $I_s(s=1,2,...,S)$ to be fused. For that, similar to Eq. (9) in the Subsection 3.3, the initial decision map Υ_s for the s-th source image is determined by

$$\Upsilon_s(x,y)=\begin{cases}1, & (x,y)\in sp_{s,i}\ \ \&\ \ s=\max\limits_{j} MOF_{j,i} \\ 0, & otherwise\end{cases}, \tag{11}$$

where $MOF_{j,i}$ denotes the measure of focus for the i-th super-pixel $sp_{j,i}$ in the j-th image. After some post-processing, the final decision map Υ'_s is obtained and the fused image I_F is obtained by

$$I_F(x,y) = \sum_{s=1}^{S} \Upsilon'_s(x,y) I_s(x,y) \ . \tag{12}$$

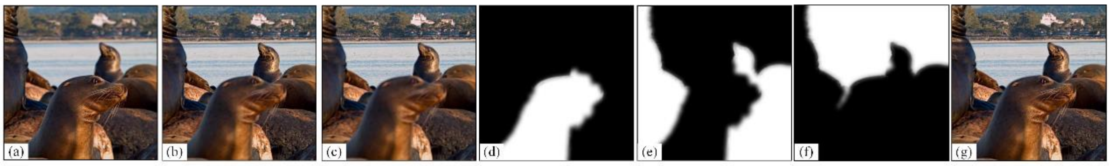

Fig.14.Illustration of the triple multi-focus image fusion. (a),(b) and (c) A set of triple multi-focus source images with the focus on the front, middle and back, respectively; (d), (e) and (f) The decision maps for (a),(b) and (c), respectively; (g) Fused image.

Fig. 14 illustrates the fusion of a set of three multi-focus images, which are also provided in the Lytro Dataset [4]. Similarly, the fusion results demonstrate that all of the focused regions within the input images can be effectively combined into the fused images without the introduction of obvious spatial artifacts.

4.7 Fusion of gray-scale multi-focus images

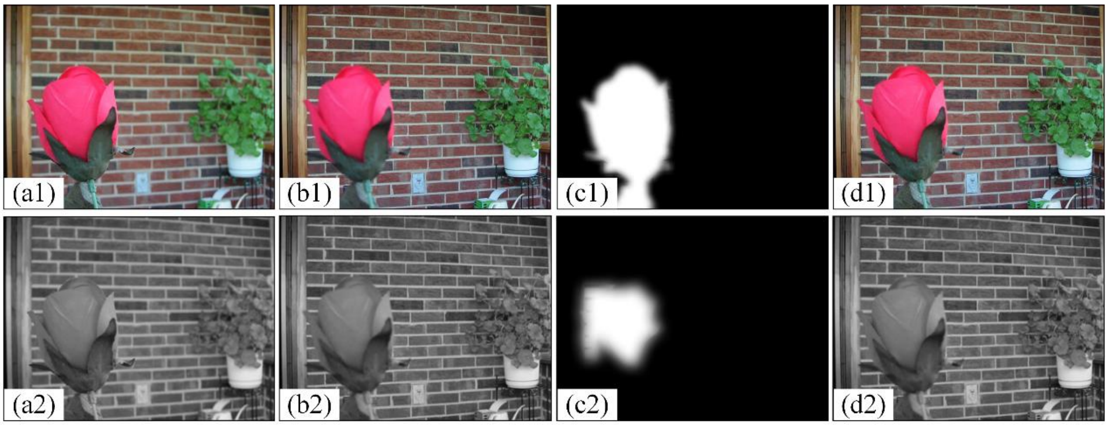

Fig. 15. Illustration of the fusion results on a pair of color multi-focus images and their gray-scale versions by using our proposed method. (a1) and (b1) A pair of color multi-focus images with the focus on the left and right parts, respectively; (c1) Focus decision map for (a1) and (b1); (d1) Fusion result on (a1) and (b1); (a2) and (b2) Gray-scale versions of (a1) and (b1), respectively; (c2) Focus decision map for (a2) and (b2); (d2) Fusion result on (a2) and (b2).

We have also tried to apply our proposed fusion method to fuse gray-scale multi-focus images. Fig. 15 illustrates the fusion results of a pair of color multi-focus images and their gray-scale versions by using our proposed fusion method. As shown in Fig. 15, we find that the fusion results on gray-scale multi-focus images are not satisfactory although the fusion results on the color multi-focus images are desirable. This may owe to the feature extraction, super-pixel segmentation and clustering modules in our proposed method, which heavily depend on the color information of source images.

4.8 Discussion

In this subsection, we will discuss two issues. One is about the computational complexities of different methods while the other is about the superiorities of our proposed methods over current DL based fusion methods.

With respect to the first issue, Table 7 provides the average computational time T of different methods on Lytro Dataset. Here, all of the traditional methods and some of the DL based methods (e.g., CNN and p_CNN) are tested in Matlab R2013b environment on a PC with an Intel i7 CPU and 32 GB of RAM. ES_CNN is tested on an NVIDIA 1080Ti GPU with 11 G memory.

Table 7. Averaging computational time of different methods on Lytro Dataset.

Methods	GF	IM	SPixel	LR_RSR	SRCF	DL_LRR	CNN	p_CNN	ES_CNN	ULRR
T(s)	0.57	2.49	57.49	25.50	13.27	4443.15	124.38	439.41	366.37	71.00

As shown in Table 7, ULRR has higher computational complexity than most of the other traditional fusion methods, such as LR_RSR and SRCF. This may owe to the part of feature extraction for each super-pixel in our proposed method. As shown in Table 8, for a pair of multi-focus images with size of 520×520, the running time of our proposed method is about 73 seconds, among which feature extraction for each super-pixel takes about 82% of the total time. Despite that, Table 7 also demonstrates that ULRR has higher computational efficiency than those DL based fusion methods. Especially, the average computational time of ULRR is about half that of CNN, although the two methods perform competitively in spatial consistency as well as in information extraction.

Table 8. Computational time of different modules in our proposed method for a pair of multi-focus images of size 520×520.

Module	Super-pixel segmentation	Feature extraction	Super-pixel clustering	Dictionary construction	ULRR decomposition	Post-processing	Total
Time(s)	0.47	60.56	1.51	0.12	3.32	7.47	73.45
Percentage(%)	0.64	82.45	2.06	0.16	4.52	10.17	100

Regarding the second issue, as discussed in Subsection 4.5, our proposed fusion method performs competitively and even better than some DL based ones. A part of the reason might be that the training data for these DL based fusion methods, which are manually generated by just performing different Gaussian filters on the original images, could not fully simulate the multi-focus characteristics of an image. Also, similar to most existing SR band LRR based fusion methods, these DL fusion methods usually perform fusion on image patches of fixed shapes independently, thus ignoring the spatial consistency among adjacent patches and degrading the fusion performance to some extent. Differently, the spatial consistency among adjacent super-pixels and the object area consistency among the super-pixels within an object are jointly considered in our proposed fusion method. Moreover, post-processing steps also contribute to the improvement of our fusion performance.

5. Conclusion

In this paper, we present a novel multi-focus image fusion algorithm based on super-pixel clustering and a unified low-rank representation (ULRR) model. Owing to the use of super-pixels of irregular sizes, the “jagged” artifacts between the focused and de-focused regions, which arises from the patch based fusion methods, can be effectively eliminated. Thanks to the use of multiple types of features, the focus information for the smooth regions as well as those regions with rich details can be well determined by the proposed fusion method. By further using super-pixel clustering and considering the object consistency among the super-pixels within the same cluster via the proposed ULRR model, the spatial artifacts in the fused images are greatly reduced and even eliminated by the proposed fusion method. Experimental results demonstrate that the proposed fusion method outperforms some state-of-the-arts, even including some deep learning based methods, in terms of visual and quantitative evaluations, especially in the reduction of spatial artifacts or in spatial consistency.

Finally, it should be also noted that the high fusion performance of our proposed method is at the cost of high computational complexity. Moreover, the proposed method works well for the color multi-focus images but it does not perform well for the gray-scale multi-focus images. In future, we will explore how to reduce the computational complexity of our proposed method and how to modify our proposed method to the fusion of gray-scale multi-focus images.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant No. 61773301.

Appendix A

(1) Update $\mathbf{J}_k$

$$\begin{aligned}\mathbf{J}_k^{i+1} &= \arg\min_{\mathbf{J}_k} \|\mathbf{J}_k\|_* + \left\langle \mathbf{Y}_{2,k}^i, \mathbf{Z}_k^i - \mathbf{J}_k \right\rangle + \frac{\mu^i}{2}\left\|\mathbf{Z}_k^i - \mathbf{J}_k\right\|_F^2 \\ &= \arg\min_{\mathbf{J}_k} \frac{1}{\mu^i}\|\mathbf{J}_k\|_* + \frac{1}{2}\left\|\mathbf{J}_k - \left(\mathbf{Z}_k^i + \frac{\mathbf{Y}_{2,k}^i}{\mu^i}\right)\right\|_F^2\end{aligned} \quad . \tag{A1}$$

The sub-optimization has the following closed-form solution:

$$\mathbf{J}_k^{i+1} = SVT_{\frac{1}{\mu^i}}\left(\mathbf{Z}_k^i + \frac{1}{\mu^i}\mathbf{Y}_{2,k}^i\right), \tag{A2}$$

where $SVT_\delta(\boldsymbol{\varphi})$ denotes the Singular Value Thresholding (SVT) operation on the matrix $\boldsymbol{\varphi}$ with the threshold δ.

(2) Update $\mathbf{H}$

$$\begin{aligned}\mathbf{H}^{i+1} &= \arg\min_{\mathbf{H}} \beta\,\mathrm{tr}(\mathbf{HLH}^T) + \left\langle \mathbf{Y}_3^i, \mathbf{Z}^i - \mathbf{H} \right\rangle + \frac{\mu^i}{2}\left\|\mathbf{Z}^i - \mathbf{H}\right\|_F^2 \\ &= \arg\min_{\mathbf{H}} \beta\,\mathrm{tr}(\mathbf{HLH}^T) + \frac{\mu^i}{2}\left\|\mathbf{Z}^i - \mathbf{H} + \frac{1}{\mu^i}\mathbf{Y}_3^i\right\|_F^2\end{aligned} \quad , \tag{A3}$$

The optimization problem in Eq. (A3) has the flowing closed-form solution:

$$\mathbf{H}^{i+1} = \left(\mathbf{Z}^i + \frac{\mathbf{Y}_3^i}{\mu^i}\right)\left(2\times\frac{\beta}{\mu^i}\times\mathbf{L} + \mathbf{I}\right)^{-1}. \tag{A4}$$

(3) Update $\mathbf{Z}$ ($\mathbf{Z}_k$)

$$
\begin{aligned}
\mathbf{Z}^{i+1} &= \arg\min_{\mathbf{Z}} \sum_{k=1}^{K} \left(\left\langle \mathbf{Y}_{1,k}^{i}, \mathbf{X}_k - \mathbf{D}\mathbf{Z}_k - \mathbf{E}_k^{i} \right\rangle + \left\langle \mathbf{Y}_{2,k}^{i}, \mathbf{Z}_k - \mathbf{J}_k^{i+1} \right\rangle + \frac{\mu^i}{2} \left\| \mathbf{X}_k - \mathbf{D}\mathbf{Z}_k - \mathbf{E}_k^{i} \right\|_F^2 + \frac{\mu^i}{2} \left\| \mathbf{Z}_k - \mathbf{J}_k^{i+1} \right\|_F^2 \right) \\
&\quad + \left\langle \mathbf{Y}_3^{i}, \mathbf{Z} - \mathbf{H}^{i+1} \right\rangle + \frac{\mu^i}{2} \left\| \mathbf{Z} - \mathbf{H}^{i+1} \right\|_F^2 \\
&= \arg\min_{\mathbf{Z}} \left\langle \mathbf{Y}_1^{i}, \mathbf{X} - \mathbf{D}\mathbf{Z} - \mathbf{E}^{i} \right\rangle + \left\langle \mathbf{Y}_2^{i}, \mathbf{Z} - \mathbf{J}^{i+1} \right\rangle + \left\langle \mathbf{Y}_3^{i}, \mathbf{Z} - \mathbf{H}^{i+1} \right\rangle \\
&\quad + \frac{\mu^i}{2} \left\| \mathbf{X} - \mathbf{D}\mathbf{Z} - \mathbf{E}^{i} \right\|_F^2 + \frac{\mu^i}{2} \left\| \mathbf{Z} - \mathbf{J}^{i+1} \right\|_F^2 + \frac{\mu^i}{2} \left\| \mathbf{Z} - \mathbf{H}^{i+1} \right\|_F^2 \\
&= \arg\min_{\mathbf{Z}} \frac{\mu^i}{2} \left\| \mathbf{X} - \mathbf{D}\mathbf{Z} - \mathbf{E}^{i} + \frac{\mathbf{Y}_1^{i}}{\mu^i} \right\|_F^2 + \frac{\mu^i}{2} \left\| \mathbf{Z} - \mathbf{J}^{i+1} + \frac{\mathbf{Y}_2^{i}}{\mu^i} \right\|_F^2 + \frac{\mu^i}{2} \left\| \mathbf{Z} - \mathbf{H}^{i+1} + \frac{\mathbf{Y}_3^{i}}{\mu^i} \right\|_F^2
\end{aligned} \tag{A5}
$$

where $\mathbf{X} = \left[\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_K\right]$, $\mathbf{Z} = \left[\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_K\right]$, $\mathbf{Y}_1^i = \left[\mathbf{Y}_{1,1}^i, \mathbf{Y}_{1,2}^i, \cdots, \mathbf{Y}_{1,K}^i\right]$, $\mathbf{Y}_2^i = \left[\mathbf{Y}_{2,1}^i, \mathbf{Y}_{2,2}^i, \cdots, \mathbf{Y}_{2,K}^i\right]$, $\mathbf{J}^{i+1} = \left[\mathbf{J}_1^{i+1}, \mathbf{J}_2^{i+1}, \cdots, \mathbf{J}_K^{i+1}\right]$ and $\mathbf{E}^i = \left[\mathbf{E}_1^i, \mathbf{E}_2^i, \cdots, \mathbf{E}_K^i\right]$. Eq. (A5) is a convex function and has the following optimal solutions:

$$
\mathbf{Z}^{i+1} = \left(\mathbf{D}^{\mathrm{T}}\mathbf{D} + 2\mathbf{I}\right)^{-1} \times \left[\mathbf{D}^{T} \left(\mathbf{X} - \mathbf{E}^{i} + \frac{\mathbf{Y}_1^{i}}{\mu^i} \right) + \mathbf{J}^{i+1} + \mathbf{H}^{i+1} - \frac{\mathbf{Y}_2^{i}}{\mu^i} - \frac{\mathbf{Y}_3^{i}}{\mu^i} \right], \tag{A6}
$$

(4) Update $\mathbf{E}$

$$
\begin{aligned}
\mathbf{E}^{i+1} &= \arg\min_{\mathbf{E}} \alpha \left\| \mathbf{E} \right\|_{2,1} + \sum_{k=1}^{K} \left(\left\langle \mathbf{Y}_{1,k}^{i}, \mathbf{X}_k - \mathbf{D}\mathbf{Z}_k^{i+1} - \mathbf{E}_k \right\rangle + \frac{\mu^i}{2} \left\| \mathbf{X}_k - \mathbf{D}\mathbf{Z}^{i+1} - \mathbf{E}_k \right\|_F^2 \right) \\
&= \arg\min_{\mathbf{E}} \alpha \left\| \mathbf{E} \right\|_{2,1} + \frac{\mu^i}{2} \left\| \mathbf{X} - \mathbf{D}\mathbf{Z}^{i+1} - \mathbf{E} + \frac{\mathbf{Y}_1^{i}}{\mu^i} \right\|_F^2
\end{aligned}, \tag{A7}
$$

where $\mathbf{E} = \left[\mathbf{E}_1, \mathbf{E}_2, \cdots, \mathbf{E}_K\right]$. This sub-optimization problem has the following closed-form solution [35]:

$$
\mathbf{E}^{i+1}(:, i) = \begin{cases} \dfrac{\left(\left\| \mathbf{G}(:, i) \right\|_2 - \dfrac{\alpha}{\mu^i} \right)}{\left\| \mathbf{G}(:, i) \right\|_2} \mathbf{G}(:, i), & \text{if } \left\| \mathbf{G}(:, i) \right\|_2 \geq \dfrac{\alpha}{\mu^i} \\ 0, & \text{otherwise} \end{cases}, \tag{A8}
$$

where $\mathbf{G} = \mathbf{X} - \mathbf{D}\mathbf{Z}^{i+1} + \frac{\mathbf{Y}_1^i}{\mu^i}$. $\mathbf{E}(:, i)$ and $\mathbf{G}(:, i)$ denote the *i*-th column of $\mathbf{E}$ and $\mathbf{G}$, respectively.

References

[1]. Q. Zhang, M. D. Levine, Robust multi-focus image fusion using multi-task sparse representation and spatial context, IEEE Transactions on Image Processing 25 (5) (2016) 2045-2058.

[2]. S. Li, X. Kang, L. Fang, J. Hu, H. Yin, Pixel-level image fusion: A survey of the state of the art, Information Fusion 33 (2017) 100-112.

[3]. B. Yang, S. Li, Multifocus image fusion and restoration with sparse representation, IEEE Transactions on Instrumentation and Measurement 59 (4) (2010) 884-892.

[4]. M. Nejati, S. Samavi, S. Shirani, Multi-focus image fusion using dictionary-based sparse representation, Information Fusion 25 (2015) 72-84.

[5]. H. Li, X. Wu, Multi-focus image fusion using dictionary learning and low-rank representation, International Conference on Image and Graphics 10666 (2017) 675-686.

[6]. H. Li, X. He, D. Tao, Y. Tang, R. Wang, Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning, Pattern Recognition 79 (2018) 130-146.

[7]. Q. Zhang, T. Shi, F. Wang, R. S. Blum, J. Han, Robust sparse representation based multi-focus image fusion with dictionary construction and local spatial consistency, Pattern Recognition 83 (2018) 299-313.

[8]. B. Cheng, L. Jin, G. Li, General fusion method for infrared and visual images via latent low-rank representation and local non-subsampled shearlet transform, Infrared Physics & Technology 92 (2018) 68-77.

[9]. G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Transaction on Pattern Analysis and Machine Intelligence 35 (1) (2013) 171-184.

[10]. F. Kou, Z. Li, C. Wen, W. Chen, Edge-preserving smoothing pyramid based multi-scale exposure fusion, Journal of Visual Communication and Image Representation 53 (2018) 235-244.

[11]. H. Zhao, Z. Shang, Y. Tang, B. Fang, Multi-focus image fusion based on the neighbor distance, Pattern Recognition 46 (3) (2013) 1002-1011.

[12]. W. Huang, Z. Jing, Evaluation of focus measures in multi-focus image fusion, Pattern Recognition Letters 28 (2007) 493-500.

[13]. S. Li, B. Yang, Multifocus image fusion using region segmentation and spatial frequency, Image and Vision Computing 26 (2008) 971-979.

[14]. Y. Liu, J. Jin, Q. Wang, Y. Shen, X. Dong, Region level based multi-focus image fusion using quaternion wavelet and normalized cut, Signal Processing, 97 (2014) 9-30.

[15]. J. Duan, L. Chen, C. L. P. Chen, Multifocus image fusion using superpixel segmentation and superpixel-based mean filtering, Applied Optics 55 (36) (2016) 10352-10362.

[16]. S. LI, X. Kang, J. Hu, B. Yang, Image matting for fusion of multi-focus images in dynamic scenes, Information Fusion 14 (2013) 147-162.

[17]. S. Li, X. Kang, J. Hu, Image fusion with guided filtering, IEEE Transactions on Image Processing 22 (7) (2013) 2864-2875.

[18]. Y. Chen, J. Guan, W. Cham, Robust multi-focus image fusion using edge model and multi-matting, IEEE Transactions on

Image Processing 27 (3) (2018) 1526-1541.

[19]. O. Bouzos, I. Andreadis, N. Mitianoudis, Conditional random field model for robust multi-focus image fusion, IEEE Transactions on Image Processing 28 (11) (2019) 5636-5648.

[20]. B. Meher, S. Agrawal, R. Panda, A. Abraham, A survey on region based image fusion methods, Information Fusion 48 (2019) 119-132.

[21]. L. Chen, J. Li, C. L. Chen, Regional multifocus image fusion using sparse representation, Optics Express 21 (4) 2013 5182-5197.

[22]. Q. Zhang, Y. Liu, R.S. Blum, J. Han, D. Tao, Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: a review, Information Fusion 40 (2018) 57-75.

[23]. Y. Liu, X. Chen, H. Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, Information Fusion 36 (2017) 191-207.

[24]. H. Tang, B. Xiao, W. Li, G. Wang, Pixel convolutional neural networks for multi-focus image fusion, Information Science 433 (2018) 125-141.

[25]. W. Zhao, D. Wang, H. Lu, Multi-focus image fusion with a natural enhancement via joint multi-level deeply supervised convolutional neural network, IEEE Transactions on Circuits and Systems for Video Technology 29 (4) (2019) 1102-1115.

[26]. H. T. Mustafa, F. Liu, J. Yang, Z. Khan, Q. Huang, Dense multi-focus fusion net: A deep unsupervised convolutional network for multi-focus image fusion, In: Internal Conference on Artificial Intelligence and Soft Computing, 2019, pp. 153-163.

[27]. H. Ma, J. Zhang, S. Liu, Q. Liao, Boundary aware multi-focus image fusion using deep neural network, In: IEEE International Conference on Multimedia and Expo, 2019, pp. 1150-1155.

[28]. M. Amin-Naji, A. Aghagolzadeh, M. Ezoji, Ensemble of CNN for multi-focus image fusion, Information Fusion 51 (2019) 201-214.

[29]. M. Amin-Naji, A. Aghagolzadeh, M. Ezoji, CNNs hard voting for multi-focus image fusion, Journal of Ambient Intelligence and Humanized Computing 11 (2020) 1749-7969.

[30]. D. Stutz, A. Hermans, B. Leibe, Superpixels: An evaluation of the state-of-the-art, Computer Vision and Image Understanding 166 (2018) 1-27.

[31]. X. Ren, J. Malik, Learning a classification model for segmentation, In: International Conference on Computer Vision, 2003, pp. 10-17.

[32]. J. Chen, Z. Li, B. Huang, Liner spectral clustering superpixel, IEEE Transactions on Image Processing 26 (7) (2017) 3317-3329.

[33]. S. P. Lloyd, Least squares quantization in PCM, IEEE Transaction on Information Theory 28 (2) (1982) 129-137.

[34]. F. R. Siqueira, W. R. Schwartz, H. Pedrini, Multi-scale gray level co-occurrence matrices for texture description, Neurocomputing 120 (2013) 336-345.

[35]. Z. C. Lin, R. S. Liu, Z. X. Su, Linearized alternating direction method with adaptive penalty for low-rank representation, In: Advances in neural information processing systems, 2011, pp. 612-620.

[36]. G. Liu, Q. Liu, P. Li, Blessing of dimensionality: recovering mixture data via dictionary pursuit, IEEE Transaction on Pattern Analysis and Machine Intelligence 39 (1) (2017) 47-60.

[37]. A. Levin, D. Lischinski, Y. Weiss, A closed-form solution to natural image matting, IEEE Transaction on Pattern Analysis and Machine Intelligence 30 (2) (2008) 228-242.

[38]. http://r0k.us/graphics/kodak.

[39]. M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing over-complete dictionaries for sparse representation, IEEE Transaction on Signal Processing 54 (11) (2006) 4311-4322.

[40]. M.B.A. Haghighat, A. Aghagolzadeh, H. Seyedarabi, A non-reference image fusion metric based on mutual information of image features, Computers & Electrical Engineering 37 (2011) 744-756.

[41]. G. Piella, H. J. Heijmans, A new quality metric for image fusion, In: International Conference on Image Processing, 2003.

[42]. A. Kolaman, O. Yadidpecht, Quaternion structural similarity: A new quality index for color images, IEEE Transactions on Image Processing 21(4) (2012) 1526-1536.

[43]. L. Alparone, S. Baronti, A. Garzelli, F. Nencini, A global quality measurement of pan-sharpened multispectral imagery, IEEE Geoscience and Remote Sensing Letters 1 (4) (2004) 313-317.

[44]. J. Zhao, R. Laganiere, Z. Liu, Performance assessment of combinative pixel-level image fusion based on an absolute feature measure, International Journal of Innovative Computing, Information and Control 3(6A) (2007) 1433-1447.

[45]. Z. Liu, D. S. Forsyth, R. Laganiére, A feature-based metric for the quantitative evaluation of pixel-level image fusion, Computer Vision and Image Understanding 109 (1) (2008) 56-68.