Robust sparse representation based multi-focus image fusion with dictionary construction and local spatial consistency

Qiang Zhang , Tao Shi , Fan Wang , Rick S. Blum , Jungong Han

PII: S0031-3203(18)30213-9
DOI: 10.1016/j.patcog.2018.06.003
Reference: PR 6577

To appear in: *Pattern Recognition*

Received date: 31 October 2017
Revised date: 8 May 2018
Accepted date: 1 June 2018

Please cite this article as: Qiang Zhang , Tao Shi , Fan Wang , Rick S. Blum , Jungong Han , Robust sparse representation based multi-focus image fusion with dictionary construction and local spatial consistency, *Pattern Recognition* (2018), doi: 10.1016/j.patcog.2018.06.003

Highlight

- A RSR model is introduced for multi-focus image fusion.
- Local consistency among adjacent patches is considered in the fusion method.
- A dictionary is constructed for RSR by using “row-sparsity” constraint.
- The fusion method introduces few spatial artifacts to the fused image.
- The fusion method has high computation efficiency.

Robust sparse representation based multi-focus image fusion with dictionary construction and local spatial consistency

Qiang Zhang[a,b], Tao Shi[b], Fan Wang[b], Rick S. Blum[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]*Electrical and Computer Engineering Department, Lehigh University, Bethlehem, PA 18015, United States*

[d]*School of Comping and Communications, Lancaster University, Lancaster, LA1 4YW, U.K.*

Abstract: Recently, sparse representation-based (SR) methods have been presented for the fusion of multi-focus images. However, most of them independently consider the local information from each image patch during sparse coding and fusion, giving rise to the spatial artifacts on the fused image. In order to overcome this issue, we present a novel multi-focus image fusion method by jointly considering information from each local image patch as well as its spatial contextual information during the sparse coding and fusion in this paper. Specifically, we employ a robust sparse representation (LR_RSR, for short) model with a Laplacian regularization term on the sparse error matrix in the sparse coding phase, ensuring the local consistency among the spatially-adjacent image patches. In the subsequent fusion process, we define a focus measure to determine the focused and de-focused regions in the multi-focus images by collaboratively employing information from each local image patch as well as those from its 8-connected spatial neighbors. As a result of that, the proposed method is likely to introduce fewer spatial artifacts to the fused image. Moreover, an over-complete dictionary with small atoms that maintains good representation capability, rather than using the input data themselves, is constructed for the LR_RSR model during sparse coding. By doing that, the computational complexity of the proposed fusion method is greatly reduced, while the fusion performance is not degraded and can be even slightly improved. Experimental results demonstrate the validity of the proposed method, and more importantly, it turns out that our LR-RSR algorithm is more computationally efficient than most of the traditional SR-based fusion methods.

Key words: multi-focus image fusion, robust sparse representation, dictionary construction, spatial contextual information, spatial consistency.

1. Introduction

Due to the limited depth of field of optical lenses in conventional cameras, it is not often possible to obtain an image that contains all of the relevant objects in focus [1, 2]. As shown in Fig. 1, this issue can be effectively addressed by multi-focus image fusion, in which several images with different focus points (e.g., Fig. 1 (a) and Fig. 1(b)) are combined into a composite image (e.g., Fig. 1(c)) with full-focus.

Suppose at least one of the input images provides a focused version of the scene, the focused regions can be extracted from the given multi-focus input images and then preserved in the fused image, while all of the defocused regions should be discarded [1]. In addition, the fusion algorithm should not introduce any spatial artifacts or inconsistencies into the fused image. Finally, the fusion algorithm

*Corresponding author. Address: Lancaster University, InfoLab21, LA1 4YW, Lancaster, UK. Email address: jungong.han@lancaster.ac.uk (J. Han).

should have high computational efficiency, thereby facilitating real-world applications. In this paper, we will address the fusion of multi-focus images by using a robust sparse representation (RSR) model with dictionary construction and local spatial consistency, specifically designed to have high spatial consistency and computational efficiency.

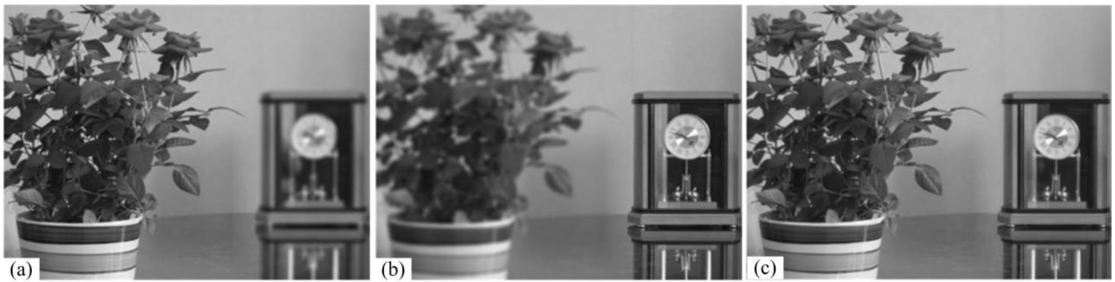

Fig. 1. Illustration of multi-focus image fusion. (a) Source image with focus on the flower; (b) Source image with focus on the clock; (c) Fused image with full-focus.

So far, many sparse representation-based (SR) methods have been presented for the fusion of multi-focus images [1, 2, 3, 4, 5, 6, 7, 8, 9]. A thorough review of these methods can be found in [10]. Rather than being fixed in advance as those in the traditional multi-scale transforms (MSTs), most of the over-complete dictionaries in SR are learned from a set of training images using some learning methods, such as K-SVD [11]. Compared with the fixed basis functions, these over-complete dictionaries contain richer basis atoms and are able to achieve more meaningful and stable representations of the source images. For this reason, SR-based image fusion methods generally outperform the traditional MST-based image fusion methods from both subjective and objective aspects [3, 4].

However, most of the existing SR-based fusion methods advocate the patch-based implementation. More specifically, each image patch is individually taken into account during sparse coding and fusion, giving rise to the spatial artifacts on the fused image. In order to reduce the spatial artifacts, a sliding window technology [3] is often employed in these methods, where the input images are divided into a larger number of patches overlapped with a fixed number of pixels (usually one pixel) along the horizontal and vertical directions, respectively. Owing to the overlap among image patches, the number of input patches to be fused is greatly large, which results in a *huge requirement of memory storage and increase of computational complexity*. In addition, some detailed information will be unavoidably lost in the fused image because of the overlap [1, 5].

In fact, images have strong local correlations among spatially adjacent patches. More precisely, for a multi-focus image, the spatially adjacent image patches are always synchronized in the sense that these patches are either all in-focus or are all out-focus in most cases. Motivated by the observation, the contextual information among spatially adjacent patches, instead of the sliding window approach, is employed to reduce the spatial artifacts of the fused image in this paper. In addition, we pay special attention to reducing the computational complexity of the proposed method in order to improve its utility for real world applications.

To this end, we employ a robust sparse representation (LR_RSR, for short) model with a Laplacian regularization term on the sparse error matrix during the sparse coding phase, which adequately considers the local consistency among the spatially-adjacent image patches. In the subsequent fusion process, we collaboratively employ information from each local image patch as well as those from its spatial neighbors to determine the focused and de-focused regions in the multi-focus images. By doing that, the spatial artifacts in the fused image may be obviously suppressed. Moreover, owing to the joint use of sparse errors from multiple spatially adjacent patches, a non-overlapping division of input images, rather than an overlapping division way as in most of SR-based fusion methods, may be adopted during the fusion process. This greatly reduces the requirement of memory storage and computational complexity of the proposed fusion method.

In addition, we will employ a learned dictionary with only a fixed small number of atoms but maintaining good representation capability, rather than the input data themselves as in [1], for the LR_RSR model during sparse coding. This will further greatly reduce the computational complexity of the proposed method, while the fusion performance is not degraded and even slightly improved. Experimental results demonstrate that the proposed method introduces fewer spatial artifacts to the fused images than most state-of-the-art methods. Especially, it is also shown to have higher computational efficiency than some traditional SR-based fusion methods.

In summary, the main contributions of this paper are as follows.

(1) We present a multi-focus image fusion algorithm based on a robust sparse representation (RSR) model, in which the spatial consistency among image patches is adequately considered during sparse coding and fusion. This is clearly different from most of the existing SR-based fusion methods, which usually treat each image patch independently.

(2) We employ a robust sparse representation (LR_RSR) model with a Laplacian regularization on the sparse error matrix during sparse coding. To the best of our knowledge, it is the first time that the Laplacian regularization is incorporated to SR-based image fusion. Moreover, we construct an

over-complete dictionary with small atoms while maintaining good representation capability for the LR_RSR model.

(3) We jointly employ local information (i.e., sparse reconstruction errors obtained by LR_RSR) of each image patch along with those from its spatially adjacent neighbors to determine the focused and defocused regions within an input multi-focus image during the fusion process.

The rest of the paper is organized as follows. Section 2 briefly reviews the SR-based fusion methods. Section 3 details the dictionary construction for RSR model. Section 4 describes the proposed fusion method in detail. Experimental results and conclusions are given in Section 5 and Section 6, respectively.

Notations

Throughout the paper, a vector is denoted by a lower-case letter, and a matrix is denoted by a capital letter. All elements of vectors and matrices are real-valued. Given a vector *x* and a matrix *X*, some notations used in this paper are listed in Table 1.

Table 1. List of vector and matrix related notations.

Symbols	Definition
$x(i)$	the *i*-th entry of the vector *x*
$X(i, j)$	the (i, j) -th entry of the matrix *X*
$X(i,:)$	the *i*-th row of the matrix *X*
$X(:, j)$	the *j*-th column of the matrix *X*
$\|x\|_2$	l_2-norm of the vector *x*, i.e., $\|x\|_2 = \sqrt{\sum_i x^2(i)}$
$\|X\|_{0,2}$	$l_{0,2}$-norm of the matrix *X*, i.e., the number of the non-zero rows in the matrix *X*
$\|X\|_{1,2}$	$l_{1,2}$-norm of the matrix *X*, i.e., $\|X\|_{1,2} = \sum_i \sqrt{\sum_j X^2(i, j)}$
$\|X\|_{2,0}$	$l_{2,0}$-norm of the matrix *X*, i.e., the number of the non-zero columns in the matrix *X*
$\|X\|_{2,1}$	$l_{2,1}$-norm of the matrix *X*, i.e., $\|X\|_{2,1} = \sum_j \sqrt{\sum_i X^2(i, j)}$
$\|X\|_F$	Frobenius-norm of the matrix *X*, i.e., $\|X\|_F = \sqrt{\sum_{i,j} X^2(i, j)}$
$\|X\|_\infty$	l_∞ -norm of the matrix *X*, i.e., the maximum absolute value of the entries in the matrix *X*
$(\cdot)^T$	transpose of a vector or a matrix

2. Related work

To date, numerous fusion algorithms have been presented for multi-focus images [12, 13, 14], wherein multi-scale transform-based (MST-based) image fusion algorithms are one of the most popular choices [15]. Various MST-based fusion methods have been discussed over the years, ranging from the early wavelet [16] and pyramid [17] transforms to the recently developed multi-scale geometric analysis approaches, such as curvelet [18], contourlet [19], and shearlet [20].

As a result of several successful applications in computer vision and image processing, sparse representation (SR) has also attracted more attention in multi-sensor image fusion, including multi-focus image fusion, in recent years [1-9, 21-24]. For example, in [3], Yang and Li first introduced SR [25] into image fusion, where the l_1-norm of the SR coefficient vector (i.e., the sum of the absolute values of SR coefficients) was employed as the activity level for each local image patch and the fused image was constructed using a maximum selection fusion rule. A SR model with dictionary learning was presented for multi-focus image fusion in [2], where the correlation between the sparse representation coefficients of input image patch and the pooled features obtained in the dictionary learning phase, instead of the simple l_1- or l_2-norm of the representation coefficient vector, was used as the activity level. In [4], a general image fusion framework was presented by combing MST and SR to simultaneously overcome the drawbacks of the MSR-based and SR-based fusion methods. They also presented an adaptive SR (ASR) model for simultaneous image fusion and denoising [21]. In [6], a group sparse presentation (GSR) model was presented to exploit the intrinsic structure among the atoms in different groups and applied to medical image fusion, where the non-zeros elements are forced to occur in clusters (i.e., group-sparsity) rather than appear randomly. Almost all these SR-based fusion methods are performed in a patch-based way. Alternatively, a newly merged convolutional SR (CSR) model was introduced to image fusion [5], which aims to achieve the SR of an entire image rather than a local image patch.

In [1], a robust SR (RSR) model was first presented to extract the detailed information in a set of multi-focus images by using a so-called sparse reconstruction error, instead of the conventional least-squared reconstruction error. Then a multi-task RSR (MRSR) model was presented for multi-focus image fusion by imposing a joint constraint on the reconstruction errors across all tasks. In the MRSR-based fusion method, information from each local image patch and those from its spatial neighbors (referred to as *its spatial contextual information*) were collaboratively employed to determine the focused and de-focused regions. Owing to the use of spatial context, block artifacts in the fusion results are greatly reduced and sometimes can even be eliminated.

Despite its great advances in terms of the performance, MRSR is computationally expensive, especially when the number of spatially adjacent patches of each image patch gets increased. In addition, the data is directly employed as the dictionary in the MRSR model. With an incremental size of each input image, the computational complexity of MRSR increases again, eventually leading the fusion algorithm to be computationally unaffordable for real world applications.

Similar to that in [1], we also consider the spatial context among image patches during the fusion process in this paper. But differently, we pay special attention to reducing the computational complexity of the proposed fusion method.

3. Dictionary construction for RSR model

Owing to the obvious superiority of RSR over the traditional SR [25], we also employ the RSR model [1] to achieve the sparse coding of each image patch. In addition to the RSR model, the employed over-complete dictionary also plays an important role for the fusion performance and computation efficiency of a multi-focus fusion method. In [1], the data themselves were simply employed as the dictionary during the sparse coding. Despite its excellent performance, the downside of such an approach is that the computational burden can be excessive for larger images if the number of dictionary atoms is propositional to the image size.

Alternatively, we will present a simple but efficient dictionary construction method for RSR. For that, we will first construct a set of data samples (or image patches), denoted by a matrix $Y=[y_1,y_2,...,y_N]\in R^{n\times N}$ of size $n\times N$, which are randomly selected from a set of training images. Here, n denotes the dimension of each data sample and N denotes the total number of image patches. Each column $y_i\in R^n$ of the matrix Y represents a data vector (i.e., a training image patch). Then we will find an optimal subset of the data samples set Y, rather than the whole set Y, to form the dictionary $D=[y_{i_1},y_{i_2},...,y_{i_M}]\in R^{n\times M}$ where $i_1,i_2,...,i_M\in\{1,2,...,N\}$, such that any column from Y can be well reconstructed by the subset D.

We will achieve this goal by first formulating the problem as the following robust "row-sparsity" optimization problem, similar to that in [26].

$$\min_{X,E}\|X\|_{0,2}+\lambda\|E\|_{2,0}\qquad s.t.\quad Y=YX+E\,. \tag{1}$$

Here, $X=[x_1,x_2,...,x_N]\in R^{N\times N}$ is the representation coefficient matrix to be sought, and each of its columns $x_i\in R^N$ denotes the representation coefficients for the data y_i. Note that YX denotes the authentic information contained in the data samples Y. $E\in R^{n\times N}$ is the sparse error matrix and denotes the corruptions or outliers within the data samples Y. The parameter $\lambda>0$ is to balance the effects of the two components in (1) and is experimentally set to 30 in this paper.

Similar to that in [27], the input training image patches themselves are also employed as the dictionary in (1). However, the goal of [27] is to achieve the sparse coding for the input data using a l_0 - or l_1 -norm minimization constraint, while our goal is to construct a dictionary for the RSR model by selecting only a small number of image patches with sufficient representation capability from the input training patches. Therefore, a "row-sparsity" (i.e., $l_{0,2}$-norm minimization) constraint is employed in (1).

When solving (1), the optimal solution of the representation coefficient matrix X^* may be incentivized to have some “zeros” rows because of the “row-sparsity” (i.e., $l_{0,2}$-norm minimization) constraint, which means that the corresponding data samples in the matrix Y are not used to reconstruct any data samples during the coding and thus cannot be selected as the dictionary atoms. In contrast, the data samples corresponding to the “non-zeros” rows in the matrix X^* have been used to reconstruct the other data samples. In fact, those data samples corresponding to the rows with larger energies (i.e., those row vectors with larger l_2-norm values) get higher weights during the coding phase, and can thus be deemed to be more important. Therefore, we will select those data samples corresponding to the row vectors of the optimal matrix X^* with the M largest l_2-norm values as the dictionary atoms, i.e., $D=[y_{i_1},y_{i_2},...,y_{i_M}]$ with $\|X^*(:,i_1)\|_2 \ge \|X^*(:,i_2)\|_2 \ge \cdots \ge \|X^*(:,i_M)\|_2 \ge \|X^*(:,j)\|_2, j \ne i_1,i_2,...,i_M$. Here, we experimentally set M to 128 or 256, which is far smaller than the total number of data samples N .

Next, we discuss the details of solving (1), which is a non-convex optimization problem and can be relaxed to the following convex one

$$\min_{X,E} \|X\|_{1,2} + \lambda \|E\|_{2,1} \quad s.t. \quad Y = YX + E . \tag{2}$$

The optimization problem in (2) is convex and can be solved by various methods. Here, we adopt the linearized alternating direction method with an adaptive penalty (LADMAP) [28, 29] considering that LADMAP has high computational efficiency and a convergence guarantee for such convex optimization problems as in (2). In addition, LADMAP can also ensure each sub-problem mentioned in (2) to have a closed-form solution. In LADMAP, an augmented Lagrangian function is first constructed by introducing a Lagrange multiplier to remove the equality constraint as

$$\begin{aligned} J &= \min_{X,E} \|X\|_{1,2} + \lambda \|E\|_{2,1} + \langle V, Y - YX - E \rangle + \frac{\mu}{2} \|Y - YX - E\|_F^2 \\ &= \min_{X,E} \|X\|_{1,2} + \lambda \|E\|_{2,1} + \frac{\mu}{2} \left\| Y - YX - E + \frac{V}{\mu} \right\|_F^2 \end{aligned} , \tag{3}$$

where V is a Lagrange multiplier and μ is a penalty parameter. $\langle A, B \rangle$ denotes the Euclidean inner product of the matrices *A* and *B*. Then the objective function in (3) is alternately minimized with respect to *X* and *E*, respectively, by fixing one or the other. Algorithm 1 gives the optimization algorithm for dictionary construction.

Algorithm 1: Optimization of RSR with "row-sparsity" constraint

```
Input: Sampling data $Y$ and parameter $\lambda$
Output: Representation coefficient and error matrices X and E
Initialize[1]: $X^0 = \mathbf{0}$, $E^0 = \mathbf{0}$, $\rho = 1.1$, $\mu = 10^{-6}$, $\mu_{\max} = 10^{10}$, $\varepsilon = 10^{-4}$
while not converged do
    (1) Fix X and update E:
        $E^{j+1} = \min_E \lambda \|E\|_{2,1} + \frac{\mu^j}{2}\left\|Y - YX^j - E + \frac{V^j}{\mu^j}\right\|_F^2 = \min_E \frac{\lambda}{\mu^j}\|E\|_{2,1} + \frac{1}{2}\left\|Y - YX^j - E + \frac{V^j}{\mu^j}\right\|_F^2 .$    (4)
        This sub-optimization problem has the following closed-form solution [30]:
        $E^{j+1}(:,i) = \begin{cases} \frac{\left(\|G(:,i)\|_2 - \lambda / \mu^j\right)}{\|G(:,i)\|_2} G(:,i), & \text{if } \|G(:,i)\|_2 \ge \lambda / \mu^j \\ 0, & \text{otherwise} \end{cases} ,$    (5)
        where $G = Y - YX^j + \frac{V^j}{\mu^j}$ .
    (2) Fix E and update X:
        $X^{j+1} = \min_X \|X\|_{1,2} + \frac{\mu^j}{2}\left\|Y - YX^j - E^{j+1} + \frac{V^j}{\mu^j}\right\|_F^2 = \min_X \|X\|_{1,2} + f(X) ,$    (6)
        where $f(X) = \frac{\mu^j}{2}\left\|Y - YX^j - E^{j+1} + \frac{V^j}{\mu^j}\right\|_F^2$ . To solve (6), the quadratic term $f(X)$ can be replaced by its first order approximation at the previous iteration by adding a proximal term [31], i.e.,
        $X^{j+1} = \min_X \|X\|_{1,2} + \frac{\eta\mu^j}{2}\left\|X - X^j\right\|_F^2 + \left\langle \nabla_X f\left(X^j\right), X - X^j \right\rangle = \min_X \|X\|_{1,2} + \frac{\eta\mu^j}{2}\left\|X - X^j + \frac{\nabla_X f\left(X^j\right)}{\eta\mu^j}\right\|_F^2 ,$    (7)
        where $\eta$ is set to $\eta = \|Y\|_2^2$ as in [31]. $\nabla_X f\left(X^j\right)$ is the partial differential of $f\left(X\right)$ with respect to X, and is computed by $\nabla_X f\left(X^j\right) = \mu^j Y^T\left(YX^j - Y + E^{j+1} - \frac{V^j}{\mu^j}\right)$. Then (6) has the following close-form solution [30]:
        $X^{j+1}(i,:) = \begin{cases} \frac{\left(\|H(i,:)\|_2 - \frac{1}{\eta\mu^j}\right)}{\|H(i,:)\|_2} H(i,:), & \text{if } \|H(i,:))\|_2 \ge \frac{1}{\eta\mu^j} \\ 0, & \text{otherwise} \end{cases} ,$    (8)
        where $H = X^j - \frac{1}{\eta}Y^T\left(YX^j - Y + E^{j+1} - \frac{V^j}{\mu^j}\right)$ .
    (3) Update the multiplier $V$ : $V^{j+1} = V^j + \mu^j\left(Y - YX^{j+1} - E^{j+1}\right)$
    (4) Update $\mu$ : $\mu^{j+1} = \min\left(\rho\mu^j, \mu_{\max}\right)$
    (5) Check the convergence conditions: $\left\|Y - YX^{j+1} - E^{j+1}\right\|_F \big/ \|Y\|_F \le \varepsilon, \left\|X^{j+1} - X^j\right\|_\infty \le \varepsilon, \left\|E^{j+1} - E^j\right\|_\infty \le \varepsilon$
end while
```

4. RSR-based multi-focus image fusion with local spatial consistency

In this section, we will first present a RSR model with Laplacian regularization (LR_RSR, for short) considering the local spatial consistency among image patches and then discuss how we apply it to the fusion of multi-focus images.

4.1 RSR with Laplacian regularization (LR_RSR)

[1] The initial values of these parameters are set as suggested in [30].

Given the over-complete dictionary $D \in R^{n \times M}$ constructed in the previous section, the existing RSR model in [1] can be computed by

$$\min_{X,E} \|X\|_0 + \lambda \|E\|_{2,0} \quad s.t. \quad Y = DX + E \ , \tag{9}$$

where $Y = [y_1, y_2, ..., y_N] \in R^{n \times N}$ is the observed data matrix (e.g., a multi-focus image in this paper), and each of its columns is a data vector (e.g., an image patch) $y_i \in R^n$. $X \in R^{M \times N}$ and $E^{n \times N}$ denote the representation matrix and error matrix, respectively.

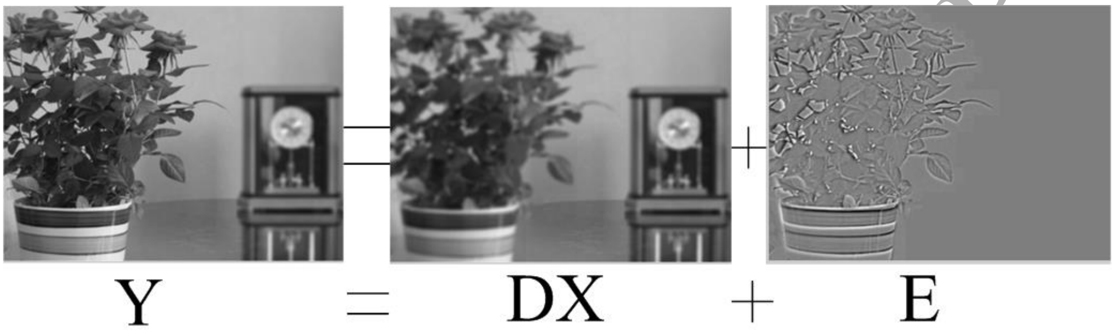

Fig. 2. Illustration of the decomposition of RSR on a multi-focus image (Credit to [1]). *Y* denotes an image focused on the flowerpot. *DX* denotes a fully defocused version of the original and *E* contains the details of the flowerpot.

The RSR model in (9) can be directly applied to the fusion of multi-focus images similar to many of the traditional SR-based fusion methods. Especially, as shown in Fig. 2, a multi-focus image can be decomposed into a blurred or fully-defocused entity plus a detailed entity, denoted by the reconstructed matrix DX and the error matrix E, respectively, by using the RSR model. In other words, the error matrix E, rather than the representation coefficients, contains the high frequency details in the multi-focus image and can thus be used to determine the focus measure of each multi-focus input image [1].

However, as shown in (9), the traditional RSR model considers each local image patch independently with no consideration of the local spatial consistency among image patches. As a result of that, some spatial artifacts will be easily introduced to the fused image in the subsequent fusion processing. In fact, images have strong local correlations among spatially adjacent patches. More exactly, for a multi-focus image, the spatially adjacent image patches have similar focus information, i.e., these patches are all in-focus or are all out-focus in most cases. Accordingly, they will also have similar sparse errors in the RSR model.

Motivated by the above observation, we present a new sparse representation model (LR_RSR, for short) by integrating a Laplacian regularization with respect to the sparse error matrix with the traditional RSR model in this paper as

$$\min_{X,E} \|X\|_0 + \lambda_1 \|E\|_{2,0} + \lambda_2 tr(ELE^T) \quad s.t. \quad Y = DX + E \,, \tag{10}$$

where λ_1 and λ_2 are two positive trade-off parameters to balance the three components. The Laplacian regularization term $tr(ELE^T)$ is defined by

$$tr(ELE^T) = \frac{1}{2}\sum_{i,j} \|E(:,i) - E(:,j)\|_2^2 \, \omega_{ij} \; . \tag{11}$$

The weight ω_{ij} implies the similarity between the *i*-th and *j*-th image patch and is computed by

$$\omega_{ij} = \begin{cases} \exp\left(-\dfrac{\|y_i - y_j\|_2^2}{2\sigma^2}\right), & \text{if } y_i \text{ and } y_j \text{ are spatially adjacent} \\ 0, & \text{otherwise} \end{cases} \; . \tag{12}$$

σ is a scalar parameter and is experimentally set to $\sqrt{0.5}$ in this paper. Based on these weights, an affinity matrix $W \in R^{N\times N}$ with $W(i,j) = \omega_{ij}$ and a diagonal degree matrix $\Delta^{N\times N}$ with $\Delta(i,i) = \sum_j W(i,j)$ are constructed. Then the Laplacian matrix *L* in (11) is computed by $L = \Delta - W$.

In general, the spatially adjacent patches with similar appearances will have similar representation coefficients as well as sparse errors in the RSR model. Accordingly, it might be more reasonable to introduce two Laplacian regularization terms with respect to the representation coefficient matrix and the sparse error matrix in (10), respectively. However, in this paper, the focus information of each local patch in a multi-focus image is determined by its sparse errors rather than its representation coefficients. Therefore, in (10), only one Laplacian regularization term with respect to the sparse error matrix is introduced for simplicity.

The Laplacian regularization with respect to sparse error matrix in the proposed LR-RSR model ensures the local consistency among the spatially-adjacent image patches. More specifically, each column y_i in the observed matrix Y in (10) denotes an image patch to be considered when LR-RSR is applied to multi-focus image fusion in our revised manuscript. The corresponding column $E(:,i)$ in the error matrix *E* denotes the sparse error for the *i*-th image patch. As shown in (12), if two spatially adjacent image patches y_i and y_j have similar appearances, the weight $\omega_{i,j}$ will be assigned to a high value. Then the difference between $E(:,i)$ and $E(:,j)$ will be forced to be a small value by minimizing the Laplacian regularization term $tr(ELE^T)$ in (10). In the subsequent fusion process, the two image patches will thus be seen as both in-focus or both out-focus. As a result of that, the spatial artifacts introduced into the fused image will be reduced to some extent.

```
**Algorithm 2**: Optimization of LR_RSR using LADMAP
**Input**: Observed data $Y$, over-complete dictionary $D$, and parameters $\lambda_1$, $\lambda_2$
**Output**: Representation coefficient and error matrices $X$ and $E$
**Initialize**: $X^0 = \mathbf{0}$, $E^0 = \mathbf{0}$, $\rho = 1.1$, $\mu = 10^{-6}$, $\mu_{\max} = 10^{10}$, $\varepsilon = 10^{-3}$
**while** not converged **do**
    (1) Fix E and update X:
        $X^{j+1} = \min_X \|X\|_1 + \frac{\mu^j}{2}\left\|Y - DX - E^j + \frac{V^j}{\mu^j}\right\|_F^2 = \min_X \|X\|_1 + g(X)$,    (15)
        where $g(X) = \frac{\mu^j}{2}\left\|Y - DX - E^j + \frac{V^j}{\mu^j}\right\|_F^2$. Similar to that in solving (6), this sup-optimization problem can be solved by replacing the quadratic term $g(X)$ with its first order approximation at previous iteration and a proximal term, i.e.,
        $X^{j+1} = \min_X \|X\|_1 + \frac{\eta_1\mu^j}{2}\|X - X^j\|_F^2 + \left\langle \nabla_X g\left(X^j\right), X - X^j \right\rangle = \min_X \|X\|_1 + \frac{\eta_1\mu^j}{2}\left\|X - X^j + \frac{\nabla_X g\left(X^j\right)}{\eta_1\mu^j}\right\|_F^2$,    (16)
        where $\eta_1$ is set to $\eta_1 = \|Y\|_2^2$ and $\nabla_X g\left(X^j\right)$ is computed by $\nabla_X g\left(X^j\right) = \mu^j D^T\left(DX^j - Y + E^{j+1} - \frac{V^j}{\mu^j}\right)$. Thus it has the following closed-form solution [32]:
        $X^{j+1} = S_{\frac{1}{\eta_1\mu^j}}\left(X^j - \frac{1}{\eta_1}D^T\left(DX^j - Y + E^{j+1} - \frac{V^j}{\mu^j}\right)\right)$,    (17)
        where the threshold function $S_\tau(x)$ is defined as
        $S_\tau(x) = \begin{cases} x - \tau, & \text{if } x > \tau \\ x + \tau, & \text{if } x < -\tau \\ 0, & \text{otherwise} \end{cases}$.    (18)
    (2) Fix X and update E:
        $E^{j+1} = \min_E \lambda_1\|E\|_{2,1} + \lambda_2 tr(ELE^T) + \frac{\mu^j}{2}\left\|Y - DX^{j+1} - E + \frac{V^j}{\mu^j}\right\|_F^2 = \min_E \|E\|_{2,1} + h(E)$,    (19)
        where $h(E) = \lambda_2 tr(ELE^T) + \frac{\mu^j}{2}\left\|Y - DX^{j+1} - E + \frac{V^j}{\mu^j}\right\|_F^2$. Similarly, this sub-optimization problem can be solved by
        $E^{j+1} = \min_E \lambda_1\|E\|_{2,1} + \frac{\eta_2}{2}\|E - E^j\|_F^2 + \left\langle \nabla_E h\left(E^j\right), E - E^j \right\rangle = \min_E \frac{\lambda_1}{\eta_2}\|E\|_{2,1} + \frac{1}{2}\left\|E - E^j + \frac{\nabla_E h\left(E^j\right)}{\eta_2}\right\|_F^2$,    (20)
        where $\eta_2$ is set to $\eta_2 = 1.02\left(2\lambda_2\|L\|_F^2 + \mu^j\right)$ as suggested in [33]. $\nabla_E h\left(E^j\right)$ is computed by $\nabla_E h(E^j) = 2\lambda_2 E^j L - \mu^j\left(Y - DX^{j+1} - E^j + \frac{V^j}{\mu^j}\right)$. Thus it has the following closed-form solution [30]:
        $E^{j+1}(:,i) = \begin{cases} \frac{\left(\|Q(:,i)\|_2 - \lambda/\mu^j\right)}{\|Q(:,i)\|_2}Q(:,i), & \text{if } \|Q(:,i)\|_2 \ge \lambda_1/\eta_2 \\ 0, & \text{otherwise} \end{cases}$,    (21)
        where $Q = E^j - \frac{\nabla_E h\left(E^j\right)}{\eta_2}$.
    (3) Update the multiplier $V$: $V^{j+1} = V^j + \mu^j\left(Y - DX^{j+1} - E^{j+1}\right)$
    (4) Update $\mu$: $\mu^{j+1} = \min\left(\rho\mu^j, \mu_{\max}\right)$
    (5) Check the convergence conditions: $\left\|Y - DX^{j+1} - E^{j+1}\right\|_F \big/ \|Y\|_F \le \varepsilon, \left\|X^{j+1} - X^j\right\|_\infty \le \varepsilon, \left\|E^{j+1} - E^j\right\|_\infty \le \varepsilon$
**end while**
```

Similar to the case in the previous section, the non-convex optimization problem in (10) can be solved as follows. First, it is relaxed to the following convex problem

$$\min_{X,E} \|X\|_1 + \lambda_1 \|E\|_{2,1} + \lambda_2 tr(ELE^T) \quad s.t. \quad Y = DX + E \,. \tag{13}$$

Then an augmented Lagrangian function is constructed by introducing a Lagrange multiplier to remove the equality constraint as

$$\begin{aligned} J &= \min_{X,E} \|X\|_1 + \lambda_1 \|E\|_{2,1} + \lambda_2 tr(ELE^T) + \langle V, Y - DX - E \rangle + \frac{\mu}{2} \|Y - DX - E\|_F^2 \\ &= \min_{X,E} \|X\|_1 + \lambda_1 \|E\|_{2,1} + \lambda_2 tr(ELE^T) + \frac{\mu}{2} \left\| Y - DX - E + \frac{V}{\mu} \right\|_F^2 \end{aligned} \,. \tag{14}$$

Finally, the optimization problem can be solved by using the LADMAP method [28, 29]. Algorithm 2 provides the optimization of LR_RSR in detail.

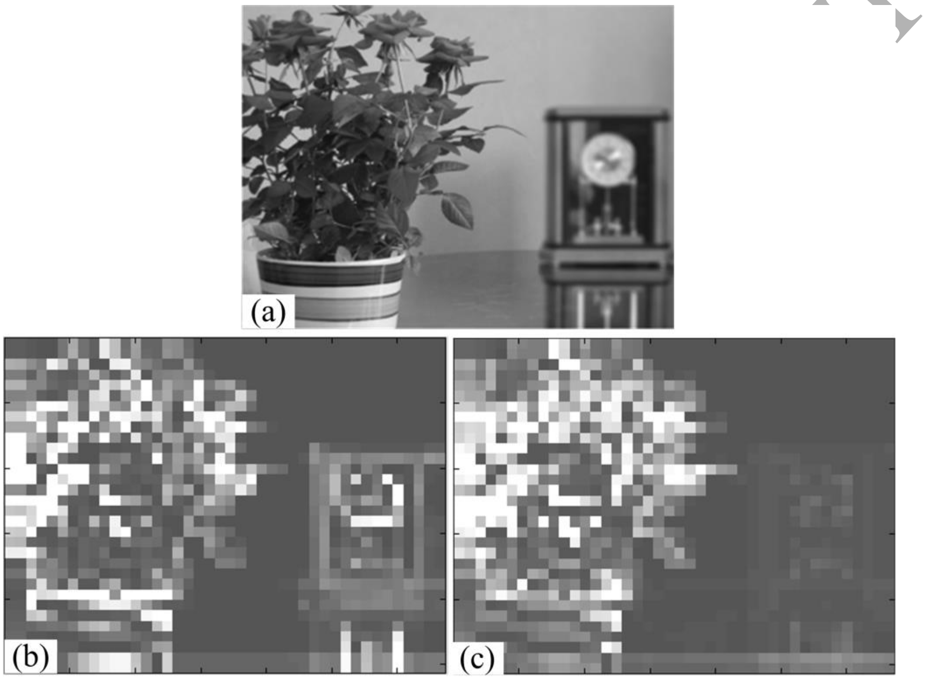

Fig. 3. Illustration of the validity of the Laplacian regularization term in the LR_RSR model. (a) An image with focus on the flower; (b) Sparse errors obtained using the traditional RSR model; (c) Sparse errors obtained using the LR-RSR model.

Fig.3 illustrates the validity of the Laplacian regularization term in the LR_RSR model. As shown in Fig. 3(b), parts of the clock regions also have higher sparse errors by using the traditional RSR model in addition to the flower regions. As a result of that, parts of the clock regions in Fig. 3(a) will be mistakenly determined to be in-focus in the subsequent fusion process by using the RSR model, thus introducing some spatial artifacts to the fused image. In contrast, as shown in Fig. 3(c), only the focused flower regions are forced to have high sparse errors and will be determined to be in-focus by using the LR-RSR model.

Furthermore, the introduction of the Laplacian regularization does not increase the computational complexity of the LR_RSR model. Similar to the traditional RSR, the major computational complexity of LR_RSR is the updating of the matrix X in (16), which requires computing the product of three matrices. As a result, LR_RSR has the same computational complexity as RSR. MRSR may also ensure the spatial consistency among image patches to some extent by imposing a joint sparsity constraint (i.e. $l_{2,1}$-norm minimization) on the reconstruction errors across all tasks. However, the joint sparsity constraint increases the computational complexity of MRSR.

More specifically, suppose the data matrix Y and dictionary D have sizes of $n \times N$ and $n \times M$, respectively. Then the coefficient matrix X has size $M \times N$. Thus, the computational complexities of RSR and LR_RSR are both $O(rnNM^2)$ by further considering the number of iterations r needed for convergence. While, the computational complexity of MRSR is about $O(rnKNM^2)$, where K denotes the number of spatially adjacent patches for each image patch to be considered. Here, the number of dictionary atoms M_k for all types of features in the MRSR model are assumed to be the same, i.e., $M_0 = M_1 = \cdots = M_{K-1} = M$. For this reason, we will apply LR_RSR to multi-focus image fusion in the following subsection.

4.2 **Multi-focus image fusion based on LR_RSR**

In this subsection, we will discuss the proposed multi-focus image fusion method in detail. In addition to LR_RSR, we will define a new focus measure by jointly employing information (i.e., the sparse errors obtained by LR_RSR) from each image patch along with information from the spatially-adjacent neighbors in the proposed fusion method to further reduce the introduction of spatial artifacts in the fused image.

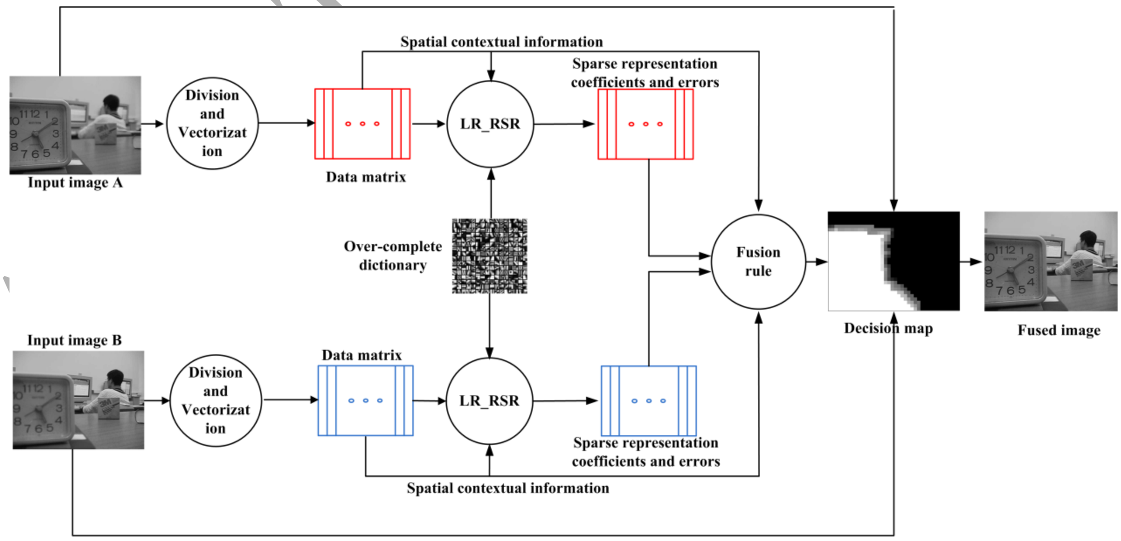
Spatial contextual information
Sparse representation coefficients and errors
Input image A
Division and Vectorization
Data matrix
LR_RSR
Over-complete dictionary
Fusion rule
Decision map
Fused image
Input image B
Division and Vectorization
Data matrix
LR_RSR
Sparse representation coefficients and errors
Spatial contextual information

Fig. 4. Diagram of the proposed fusion method.

Given the two input multi-focus images I_A and I_B that are assumed to have been well registered, the diagram of the proposed multi-focus image fusion algorithm based on LR_RSR is shown in Fig. 4 and described as follows.

(1) Divide the input images I_A and I_B into *N non-overlapping* image patches of the same size $p_x \times p_y$ pixels. Then two sets of image patches $\{I_i^A \mid i = 0,1,...,N-1\}$ and $\{I_i^B \mid i = 0,1,...,N-1\}$ are obtained from images I_A and I_B, respectively.

(2) Re-order each image patch as a vector of dimension $d = p_x \times p_y$ and construct the data matrices $Y_A = \left[y_0^A, y_1^A, ..., y_{N-1}^A\right]$ and $Y_B = \left[y_0^B, y_1^B, ..., y_{N-1}^B\right]$ for images I_A and I_B, respectively. y_i^A and y_i^B are vectors corresponding to the *i*-th image patches I_i^A and I_i^B of images I_A and I_B, respectively.

(3) Perform LR_RSR on Y_A and Y_B, respectively, using Algorithm 2 and then obtain their corresponding representation coefficient matrices X_A, X_B and error matrices E_A, E_B. In this step, a globally-trained dictionary *D* is employed, which is constructed from a set of training image patches by using Algorithm 1. As well, each image patch and one of its 8-connencted neighbors are seen as a pair of spatially-adjacent image patches during the computation of Laplacian regularization in (10).

(4) Define a decision map (i.e., a matrix) C with the same size of source images by using the sparse errors E_A and E_B. The values of the entries in the matrix C are in the range of $[0,1]$. “1” indicates that the fused pixels are directly selected from the source image I_A, while “0” means that the fused pixels are directly selected from the source image I_B. Otherwise, the fused pixels are the weighted average of the source images I_A and I_B. This step is one of the most important components in the proposed fusion method and will be further discussed soon in detail.

(5) Construct the finally fused image I_F by using the decision map C as

$$I_F(m,n) = I_A(m,n)C(m,n) + I_B(m,n)\big(1 - C(m,n)\big), \tag{22}$$

where $I_F(m,n)$, $I_A(m,n)$ and $I_B(m,n)$ denote the pixel values of the fused image I_F, input image I_A and input image I_B in location (m,n), respectively. Correspondingly, $C(m,n)$ denotes the (m,n) -th entry of the matrix (or the decision map) C .

In the following content, we will discuss the determination of the decision map C in detail.

First, define an initial decision map C' of the same size I_A or I_B, and divide the decision map C' into *N* patches of size $p_x \times p_y$ by using the same way as that in the division of source image I_A or I_B. Then obtain a set of decision map patches or sub-matrices $\{C_i' \mid i = 0,1,..., N-1\}$.

Secondly, assign each of its entries $C_i'(m,n)$ in the *i*-th decision map patch C_i' to “1” or “0” by comparing the focus measure value π_i^A of image patch I_i^A with the focus measure value π_i^B of image patch I_i^B as

$$C_i'(m,n) = \begin{cases} 1, & \text{if } \pi_i^A \geq \pi_i^B \\ 0, & \text{otherwise} \end{cases} . \tag{23}$$

Here, the focus measure value π_i^A is patch-based and is jointly determined by the sparse errors of image patch I_i^A as well as its 8-connencted spatial-adjacent neighbors, denoted by $\Gamma(I_i^A)$, as follows

$$\pi_i^A = \|E_A(:,i)\|_2 + \sum_{I_j^A \in \Gamma\left(I_i^A\right)} \|E_A(:,j)\|_2 . \tag{24}$$

Accordingly, the focus measure value π_i^B is determined by using the same way, i.e.,

$$\pi_i^B = \|E_B(:,i)\|_2 + \sum_{I_j^B \in \Gamma\left(I_i^B\right)} \|E_B(:,j)\|_2 . \tag{25}$$

It should be noted that *the sparse errors of current image patch and its spatially-adjacent neighbors, instead of the only sparse error of current image patch, are jointly employed to define the focus measure* in (24) and (25). This will reduce spatial artifacts, as shown in Fig. 5(e).

Thirdly, reconstruct the decision map C' by adding the patches $\{C_i' \mid i = 0,1,..., N-1\}$ to C' at their original spatial positions in C'. This can be seen as the reverse process of the division of C'.

Fourthly, refine the decision map C' by removing some "holes" with small areas to obtain a new decision map C''. Although the introduction of local consistency can reduce the artifacts to great extent, some isolated regions in-focus are still inevitably mistaken as de-focused ones. Similarly, some isolated regions out-focus are also mistakenly labeled as the focused ones. As a result of that, there will be some "holes" in the decision map C'. In this paper, those connected regions in C' whose numbers of entries are less than 5% of the total number of pixels in the input image are seen as isolated regions and are thus removed. This is simply achieved by re-assigning the entry values within these isolated regions as 1 minus their original values. The new decision map C'' is thus computed by

$$C''(m,n) = \begin{cases} 1 - C'(m,n), & \text{if } (m,n) \in \Psi_{C'} \\ C'(m,n), & \text{otherwise} \end{cases}, \tag{26}$$

where $\Psi_{C'}$ denotes the isolated regions in the decision map C'.

Finally, define some transitional regions between the focused regions and the defocused regions, and then construct the final decision map C. According to the decision map C'', each input image can be simply divided into two types of regions, i.e., focused regions and de-focused regions. For example, “1” means focused regions while “0” means de-focused regions for image I_A. In contrast, “1” and “0” mean de-focused regions and focused regions, respectively, for image I_B. However, as discussed in [34], the de-focused imaging system can be characterized by a low-pass filtering system. This indicates that it is a gradual process, rather than an abrupt process, from the focused (or de-focused) regions to the de-focused (or focused) regions. In other words, it is reasonable to define a transitional region between a focused region and a defocused region.

Therefore, in this paper, we will *divide each multi-focus input image into three types of regions (i.e., focused, de-focused and transitional regions), instead of two types of regions*. We simply take those patches in the decision map C'' as transitional regions, denoted by $\Upsilon_{C''}$, whose entries values are different from those of one of its 8-connected spatial neighbors. For these transitional regions, the fused image is computed as the weighted average of source images, instead of being simply selected from one of the source images. Here, the weights are also computed by using the focus measure values of these source image patches. Then the final decision map C is determined by

$$C(m,n) = \begin{cases} 1, & \text{if } \pi_i^A \geq \pi_i^B \ \& \ (m,n) \notin \Upsilon_{C''} \\ \dfrac{\pi_i^A}{\pi_i^A + \pi_i^B}, & (m,n) \in \Upsilon_{C''} \\ 0, & \text{if } \pi_i^A < \pi_i^B \ \& \ (m,n) \notin \Upsilon_{C''} \end{cases}, \tag{27}$$

where the index i in π_i^A or π_i^B is determined by the index of image patch I_i^A or I_i^B that the location (m,n) belongs to. By using the final decision map C, the fused image can be obtained by using (22).

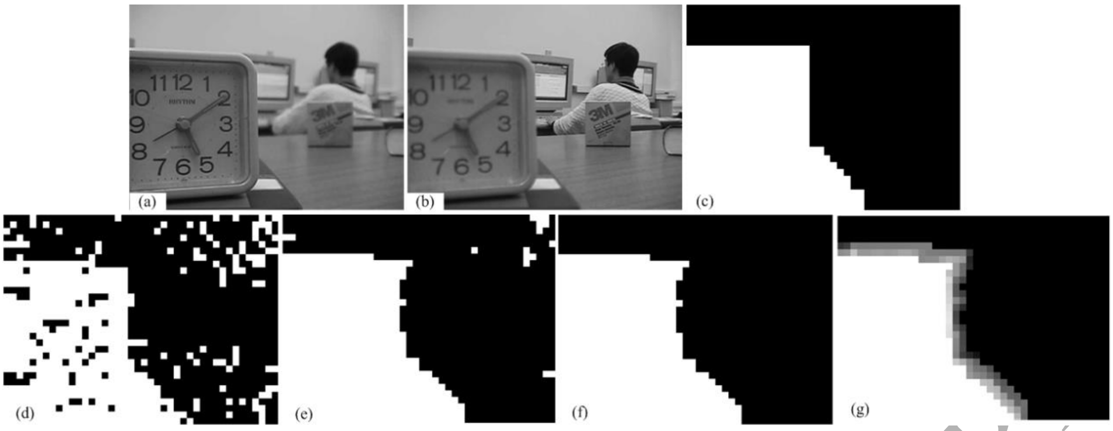

Fig. 5. Illustration of decision maps obtained by different methods. (a) Source image with focus on the 'clock'; (b) Source image with focus on the 'student'; (c) 'Ideal' decision map; (d) Decision map obtained by using the sparse error of each single image patch; (e) Decision map C' obtained by using the joint sparse errors of each image patch and its neighbors, i.e., (23); (f) Decision map C'' obtained by performing 'removing holes' on (e), i.e., (26); (g) Final decision map C by using (27), in which the gray regions denote the transitional regions.

Fig. 5 illustrates the decision maps obtained by different methods. As shown in Fig. 5(d), there are many isolated patches or "holes" in the decision map obtained by using the sparse error of each single image patch. In contrast, as shown in Fig. 5(e), these "holes" are greatly reduced by using the joint sparse errors of each image patch and its spatially-adjacent neighbors. This demonstrates the effectiveness of the proposed focus measures defined by (24) and (25). By further removing the remaining "holes" or isolated patches, the decision map can be closer to the 'ideal' decision map.

4.3 Computational Complexity of the proposed fusion method

The computational complexity of the proposed fusion method is fully dependent on that of the LR_RSR model. As discussed in Subsection 4.1, the computational complexity of LR_RSR is $O(rnNM^2)$, where *M* and *N* denote the numbers of the dictionary atoms and input image patches, respectively. *n* denotes the dimension of the dictionary atoms or input data. *r* is the number of iterations needed for convergence. Accordingly, the computational complexity of the proposed fusion method is also $O(rnNM^2)$, which demonstrates that the number of dictionary atoms M has a greater impact on the computational complexity of the proposed fusion method than other parameters.

When RSR and MRSR are applied to multi-focus image fusion, the input data themselves are simply employed as the dictionary in [1]. That is to say, the number of dictionary atoms *M* equals that of the data (or the input image patches) *N* in the RSR-based and MRSR-based fusion methods. As a result of that, the computational complexities of RSR and MRSR fusion methods in [1] are in fact about $O(rnN^3)$ and $O(rnKN^3)$, respectively. Here K denotes the number of spatially adjacent patches for each image patch to be considered. In addition, the number of dictionary atoms M (e.g., 256 in this paper) is usually smaller than the number of image patches N (e.g., 1200 for an image of size 320×240) in the proposed fusion method. Therefore, the proposed fusion method has greatly higher computation efficiency than the RSR-based and MRSR-based fusion method.

More importantly, due to the non-overlapping division of input images, the number of image patches N in the proposed fusion method is also much smaller than those (e.g., about 76800 for an image of size 320×240) in the traditional SR-based fusion methods, including the RSR-based one in [1], where an overlapping division way is usually adopted. Therefore, the proposed fusion method also has higher computational efficiency than those traditional SR-based fusion methods. This will be verified in the experimental part.

5. Experimental results and analysis

In this section, several sets of experiments are performed to verify the feasibility of the proposed multi-focus image fusion algorithm based on the LR_RSR. First, we discuss the validity of the constructed dictionary by using Algorithm 1. Then we discuss the impacts of some parameters on the fusion performance. Finally, several pairs of multi-focus images from two public databases are fused by using the proposed method and some state-of-the-art methods to demonstrate the validity of the proposed method.

5.1 Validity of the constructed dictionary

Here, we will discuss the impacts of different dictionaries on the fusion performance to show the validity of the proposed dictionary construction method. For that, 20,000 patches of size 8×8 are first randomly selected from a set of images with high resolution to construct the training data. These images are downloaded from http://r0k.us/graphics/kodak. Afterward, two sets of dictionaries with different parameters are constructed by using Algorithm 1. One set of dictionaries ($D_{\lambda=1}$, $D_{\lambda=10}$, $D_{\lambda=20}$, $D_{\lambda=30}$, $D_{\lambda=40}$, $D_{\lambda=50}$, $D_{\lambda=70}$ and $D_{\lambda=100}$, for short, respectively) are constructed by using the same number of atoms (i.e., *M*=256) but different values of the parameter λ (i.e., λ =1, 10, 20, 30, 40, 50, 70, 100, respectively). The other set of dictionaries ($D_{M=128}$, $D_{M=256}$, $D_{M=512}$, $D_{M=1024}$ and $D_{M=2048}$, for short, respectively) are constructed by using the same value of λ (i.e., $\lambda=30$) but different numbers of atoms (i.e., *M*=128, 256, 512, 1024, 2048, respectively). Finally, the multi-focus images in the previous Fig. 5(a) and Fig. 5(b) are fused using the proposed fusion method but with different dictionaries constructed above. In addition, the dictionary with 256 atoms learned by using the K-SVD method for the traditional SR model (D_{KSVD} , for short) and the normalized data themselves (D_{data} , for short) are also compared with our constructed dictionaries.

In order to subjectively evaluate the fusion performance by using different dictionaries, a fully focused ('ideal') image I_R is first created by visually extracting the focused regions from input images Fig. 5(a) and Fig. 5(b). Then the fused images are compared with the 'ideal' image by using the mean square error (*Emse*) and the difference coefficients (*dDC*). Smaller *Emse* and *dDC* values indicate higher fusion performance.

Table 2 and Table 3 present the fusion results obtained by using our proposed fusion method but with different dictionaries. Table 2 shows that the fusion performance varies with the parameter λ and achieves the best when λ is set to 30. Table 3 shows that better fusion performance can be obtained when using our constructed dictionaries (i.e., the first 5 dictionaries in Table 3) than the dictionaries D_{data} and D_{KSVD} . Further, the dictionary $D_{M=256}$ achieves the best fusion performance among the mentioned dictionaries. This demonstrates that dictionaries with only a few atoms (e.g., 256), carefully selected from among the 20,000 training data samples, have better representation capability than dictionaries with more atoms. By imposing the "row-sparsity" constraint on the representation coefficients, the data samples with the best representation capability can be selected from the training data. In particular, the constructed dictionary $D_{M=256}$ performs much better than the dictionary D_{KSVD} , although both of them have 256 dictionary atoms. This further demonstrates the effectiveness of our proposed dictionary construction method.

Table 2. Fusion results using the dictionaries with different values of λ .

Dictionary	$D_{\lambda=1}$	$D_{\lambda=10}$	$D_{\lambda=20}$	$D_{\lambda=30}$	$D_{\lambda=40}$	$D_{\lambda=50}$	$D_{\lambda=70}$	$D_{\lambda=100}$
Emse	2.4449	2.1839	2.1940	2.1929	2.2126	2.2205	2.2320	2.2344
dDC	0.0137	0.0128	0.0127	0.0127	0.0127	0.0128	0.0128	0.0129

Table 3. Fusion results using the dictionaries with different numbers of atoms *M*.

Dictionary	$D_{M=128}$	$D_{M=256}$	$D_{M=512}$	$D_{M=1024}$	$D_{M=2048}$	D_{data}	D_{KSVD}
Emse	2.1945	2.1929	2.2582	2.2966	2.5222	2.6676	3.0549
dDC	0.0128	0.0127	0.0129	0.0131	0.0138	0.0146	0.0150

5.2 Fusion parameter impacts

In this subsection, we still employ the input multi-focus images in Fig. 5(a) and Fig. 5(b) to test the impacts of some parameters, including λ_1 and λ_2 in (10) or (13), σ in (12), and patch sizes

$p_x \times p_y$, on the fusion performance.

Table 4. Fusion results by using the proposed method with different values of λ_2 .

λ_2 with $\lambda_1 = 1, p_x = p_y = 8, \sigma = \sqrt{0.5}$	0.001	0.01	0.1	1	10	100	1000	10000
Emse	3.0279	2.7759	2.3954	2.1929	2.2105	2.3485	2.3535	2.9409
dDC	0.0154	0.0145	0.0135	0.0127	0.0127	0.0137	0.0138	0.0152

Table 5. Fusion results by using the proposed method with different values of σ .

σ with $\lambda_1 = \lambda_2 = 1, p_x = p_y = 8$	$\sqrt{0.2}$	$\sqrt{0.3}$	$\sqrt{0.4}$	$\sqrt{0.5}$	$\sqrt{0.6}$	$\sqrt{0.7}$	1.0
Emse	2.9620	2.9613	2.3442	2.1929	2.3300	2.9573	2.9569
dDC	0.0151	0.0151	0.0133	0.0127	0.0132	0.0150	0.0150

Experimental results demonstrate that the fusion performance remains nearly unchanged when the parameter λ_1 is within the range of [0.001, 300]. When λ_1 is larger than 300, the fusion performance will be greatly degraded. In contrast, the fusion performance varies continuously with the parameter λ_2 and is best when λ_2 is set to 1, which is shown in the Table 4. As shown in Table 5, the fusion performance also varies with the parameter σ and achieves the best when σ is set to $\sqrt{0.5}$. Similar to those in the traditional SR and RSR fusion methods, better fusion results can be obtained when the sizes of image patches are set to 8×8 . Therefore, we will set $p_x = p_y = 8$, $\lambda_1 = \lambda_2 = 1$ and $\sigma = \sqrt{0.5}$ in the following experiments.

5.3 Validity of the proposed fusion method

Several pairs of multi-focus images from two public databases are employed to thoroughly demonstrate the validity of the proposed fusion method LR_RSR. Fig. 6 provides the ten pairs of multi-focus images from the first database, which are downloaded from http://home.ustc.edu.cn/~liuyu1. Fig. 7 illustrates the twenty pairs of multi-focus images from the second database, which are downloaded from https://www.researchgate.net/publication/291522937_Lytro_Multi-focus_Image_Dataset.

We will compare our proposed method LR_RSR with some state-of-the-art methods, including SR [3], adaptive SR (ASR) [21], NSCT_SR [4], convolutional SR (CSR) [5], RSR [1], MRSR [1], neighbor distance (ND) [12], NSCT [4], homogeneity similarity (HS) [35], image matting (IM) [36] and deep convolutional neural network (DCNN) [37]. It should be noted that DCNN is a deep-learning-based fusion method. The mutual information (*MI*) quality metric [38], gradient preservation quality metric Q_G [39], two-phase congruency-based fusion quality metric *ZN_CC* [40] and the Q_{PC} metric [41] are employed to subjectively evaluate the different fusion methods. The former two metrics *MI* and Q_G are used to evaluate the different fusion methods based on information extraction, while the latter two metrics *ZN_CC* and Q_{PC} are used to evaluate different fusion methods based on spatial consistency. Larger values of these metrics mean better fusion performance.

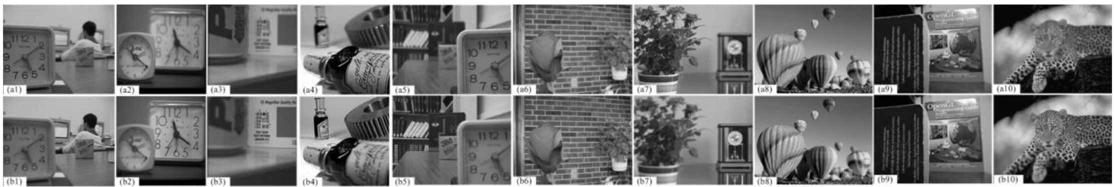

Fig. 6. Ten pairs of multi-focus images in the first database. The top row contains 10 input images with the focus on the left part, and the bottom row contains the corresponding input images with the focus on the right part.

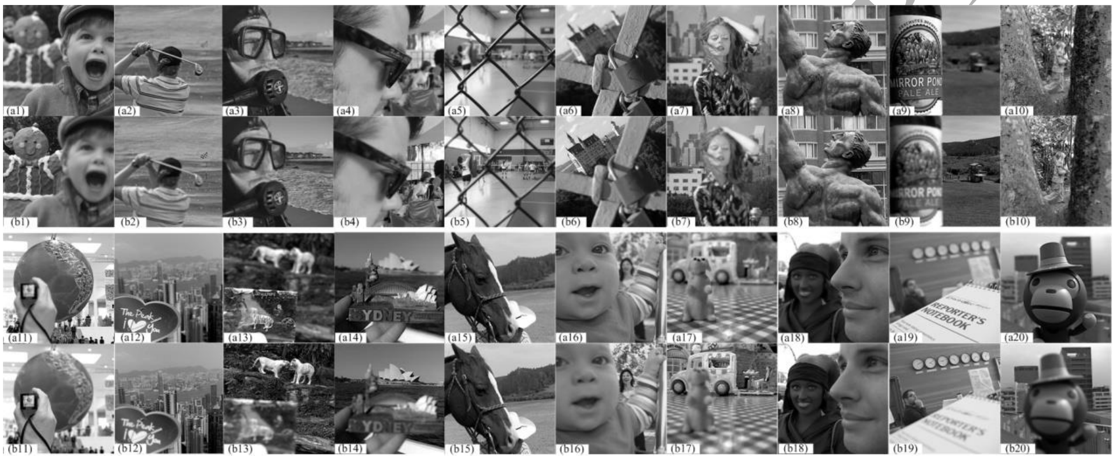

Fig. 7. Twenty pairs of multi-focus images in the second database. The first top row contains the first 10 input images with the focus on the front part, and the second row contains the corresponding input images with the focus on the back part. The third row contains the remaining 10 input images with the focus on the front part, and the bottom row contains the corresponding input images with the focus on the back part.

Fig. 8 illustrates some fusion results on the first pair of multi-focus images in Fig. 6(a1) and Fig. 6(b1) (i.e., Fig. 5(a) and Fig. 5(b)) obtained by using different fusion methods. In order to better compare different fusion methods visually, in Fig. 9, we also provide the normalized difference images [1] between each of the fused images in Fig. 8 and one of the input images in Fig. 6(b1).

As shown in Fig. 8, all of the fusion methods mentioned here seem to perform well for Fig. 6(a1) and Fig. 6(b1). However, a more careful comparison in Fig. 9 indicates that LR_RSR, DCNN and MRSR perform better than others. As shown in Fig. 9(a) ~ Fig. 9(i), there are many residual errors between each of the fused images and the input image Fig. 6(b1). This indicates that the fused images obtained by these methods do not completely come from the focused regions of the input images and thus introduce serious spatial artifacts, especially on the borders of the head of the student. In contrast, the residual errors between each of the fused images obtained by the other three methods are greatly reduced. This demonstrates that MRSR, DCNN and our proposed LR_RSR method can more accurately determine the focused and defocused regions of the input images. As shown in circle regions of Fig. 9(j) ~ Fig. 9(l), few residual errors exist on the borders of head regions. This also indicates that MRSR, DCNN and LR_RSR, especially the latter two methods, introduce fewer spatial artifacts to the fused images.

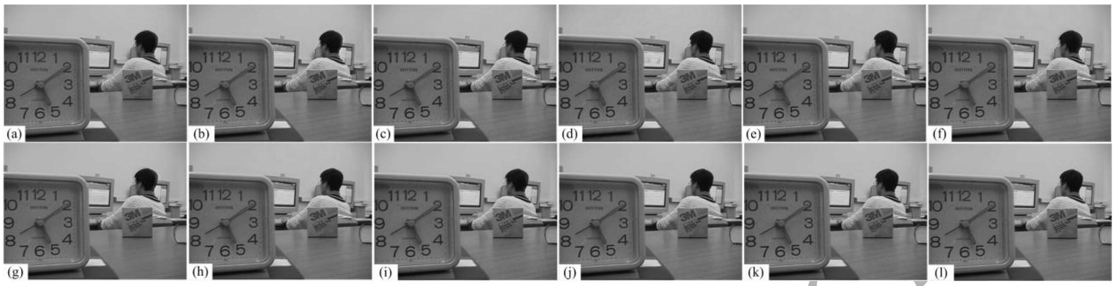

Fig. 8. Fusion results for Fig. 6(a1) and (b1). (a) ND; (b) NSCT; (c) HS; (d) MF; (e) SR; (f) NSCT_SR; (g) ASR; (h) CSR; (i) RSR; (j) MRSR; (k) DCNN; (l) LR_RSR.

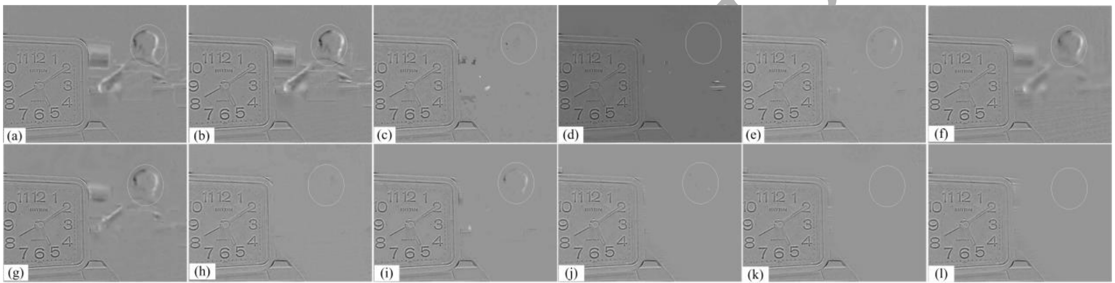

Fig. 9. Normalized difference images between Fig. 6(b1) and each of the fused images in Fig. 8. (a) ND; (b) NSCT; (c) HS; (d) MF; (e) SR; (f) NSCT_SR; (g) ASR; (h) CSR; (i) RSR; (j) MRSR; (k) DCNN; (l) LR_RSR.

Table 6. Performance of different SR-based fusion methods on the first database. Scores for the 10 image pairs in this database are averaged.

Method	*MI*	Q_G	*ZNCC_PC*	Q_{PC}	$T_{avg}(s)$
ND	3.9255	0.7438	0.9403	0.6541	2.0
NSCT	3.8461	0.7420	0.9458	0.6673	0.7
HS	4.8726	0.7629	0.9433	0.6858	0.8
MF	4.7915	0.7638	0.9597	0.6828	37.8
SR	4.5100	0.7583	0.9508	0.6793	15.9
NSCT_SR	4.0559	0.7496	0.9507	0.6735	12.6
ASR	3.7429	0.7244	0.8879	0.6321	42.0
CSR	4.0694	0.7396	0.9131	0.6656	43.1
RSR	4.7241	0.7707	0.9670	0.6977	880.1
MRSR	4.8530	0.7737	0.9686	0.7009	39.6
DCNN	4.8150	0.7681	0.9721	0.7005	43.8
LR_RSR	4.8127	0.7687	0.9752	0.6993	4.6

This visual comparison among different fusion methods is consistent with the quantitative results in Table 6, which also demonstrates that MRSR, DCNN and LR_RSR perform better than other methods in information extraction as well as in spatial consistency. Although it performs competitively with MRSR and DCNN, the proposed LR_RSR method has significantly higher computational efficiency than MRSR and DCNN. For all of the test images in the first database, the average computational time T_{avg} of LR_RSR is only about one-tenth that of MRSR and DCNN. This owes to the smaller number of dictionary atoms and the non-overlap of input images employed in LR_RSR.

Fig. 10 illustrates the fusion results on the input images in Fig. 7(a1) and Fig. 7(b1) obtained by all of the mentioned fusion methods. The difference images between each of the fused images in Fig. 10 and the input image in Fig. 7(b1) are also provided in Fig. 11 to better illustrate the validity of the proposed method. The average quantitative results on all of the twenty pairs of input images in the second database are given in Table 7. Similar conclusions follow from these visual and quantitative fusion results.

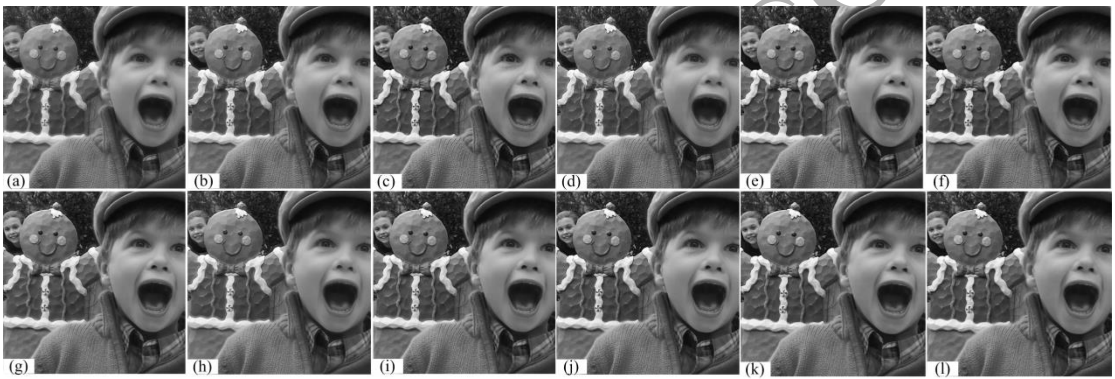

Fig. 10. Fusion results for Fig. 7(a1) and (b1). (a) ND; (b) NSCT; (c) HS; (d) MF; (e) SR; (f) NSCT_SR; (g) ASR; (h) CSR; (i) RSR; (j) MRSR; (k) DCNN; (l) LR_RSR.

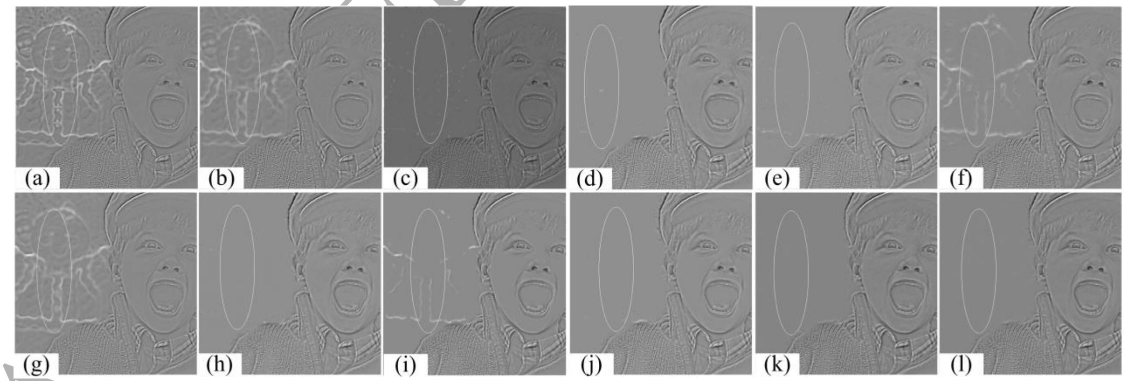

Fig. 11. Normalized difference images between Fig. 7(b1) and each of the fused images in Fig. 10. (a) ND; (b) NSCT; (c) HS; (d) MF; (e) SR; (f) NSCT_SR; (g) ASR; (h) CSR; (i) RSR; (j) MRSR; (k) DCNN; (l) LR_RSR.

As shown in Fig. 11(a) ~ Fig. 11(i), larger residual errors exist between each of the fused images and the input image in Fig. 7(b1). In contrast, fewer residual errors exist in the elliptical regions of Fig. 11(j) ~ Fig. 11(l). This demonstrates fewer spatial artifacts are introduced to the fused images obtained by MRSR, DCNN, and LR_RSR. Accordingly, the three methods outperform others in spatial consistency.

The quantitative results in Table 7 demonstrate that our proposed LR_RSR method performs competitively with the deep-learning-based fusion method DCNN and better than all of the other methods, including SR, NSCT_SR, ASR, CSR, RSR, and MRSR, in terms of information extraction as well as in terms of spatial consistency. In addition, LR_RSR has the highest computational efficiency among the SR-based methods. The average computational time T_{avg} of LR_RSR is about 1/15 and 1/4 that of that for MRSR and DCNN for this database, respectively.

Table 7. Performance of different fusion methods on the second database. Scores for the 20 image pairs in this database are averaged.

Method	*MI*	Q_G	*ZNCC_PC*	Q_{PC}	$T_{avg}(s)$
ND	5.3882	0.7323	0.8988	0.5500	11.5
NSCT	5.6332	0.7389	0.9079	0.5730	1.9
HS	9.7445	0.7588	0.9175	0.6031	4.9
MF	9.6883	0.7630	0.9227	0.6083	43.6
SR	8.5003	0.7585	0.9181	0.6009	1272.0
NSCT_SR	7.1512	0.7512	0.9159	0.5943	54.3
ASR	5.0913	0.7001	0.8439	0.5131	401.1
CSR	6.8324	0.7384	0.9068	0.5878	161.9
RSR	8.9146	0.7541	0.9171	0.5974	598.6
MRSR	9.0576	0.7548	0.9241	0.6082	529.2
CNN	9.2878	0.7589	0.9283	0.6093	121.5
LR_RSR	9.1986	0.7583	0.9298	0.6098	31.6

5.4 Fusion of more than two multi-focus images

The proposed fusion method can also be applied to the fusion of more than two multi-focus images by simple extension. Suppose that there are total *R* images $\{I_r \mid r = 1, 2, .., R\}$ to be fused. After the division and sparse coding of input images, a set of decision maps $\{C^r \mid r = 1, 2, ..., R\}$ of the same size as those of input images are defined. Similar to (27) in the Subsection 4.2, each entry $C^k(m,n)$ in the *k*-th decision map C^k may be computed by

$$C^k(m,n) = \begin{cases} 1, & \text{if } k = \max_r \pi_i^r \ \& \ (m,n) \notin \Upsilon \\ \dfrac{\pi_i^k}{\sum_{r=1}^{R} \pi_i^r}, & (m,n) \in \Upsilon \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

where π_i^r denotes the focus measure value for the *i*-th patch I_i^r of input image I_r. Here, the index *i* is determined by the index of image patch I_i^r that the location (m,n) belongs to. Υ denotes the transitional regions and may be determined in a way similar to that in Subsection 4.2. The fused image is thus obtained by

$$I_F\left(m,n\right)=\sum_{r=1}^{R} I_r\left(m,n\right)C^r\left(m,n\right) \tag{29}$$

Fig. 12 illustrates the fusion of three multi-focus images[2], which demonstrates that all focus regions of input images are integrated into the final fusion result with no obvious artifacts.

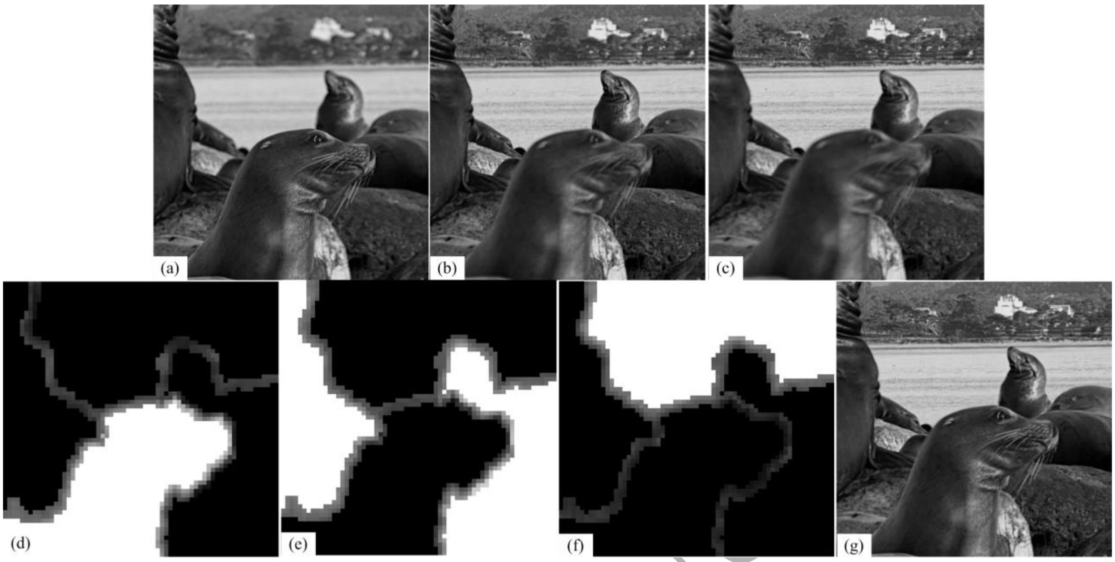

Fig.12. Application of the proposed method for more than two multi-focus images. (a)-(c) Source images with focus on the front, middle and back, respectively; (d)-(e) Decision maps for (a)-(c), respectively; (g) Fused image.

6. Conclusion

In this paper, we present a RSR-based multi-focus image fusion algorithm with local spatial consistency, in which information from each local image patch and those from its spatially-adjacent neighbors are jointly considered during sparse coding and in the definition of activity level in the fusion process. As a result, the proposed method is shown to be competitive with and even superior to some state-of-the-art methods in terms of spatial consistency as well as information extraction.

We employ a robust sparse representation (LR_RSR) model with a Laplacian regularization term on the sparse error matrix, instead of using the MRSR model, to achieve the sparse coding of each image patch. Additionally, we construct an over-complete dictionary with small atoms while maintaining good representation capability, rather than employing the data themselves, for the LR_RSR model during sparse coding. Owing to the LR_RSR model and the constructed dictionary, the proposed fusion method is shown to have higher computational efficiency than most of the traditional SR-based fusion methods, including the MRSR-based method in [1]. In future work, we will extend the LR_RSR model to deal with other applications such as action recognition [42][43], image retrieval [44][45], and visual saliency detection [46][47][48].

[2] The input images are provided in the Lytro Dataset.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant No. 61773301, by Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2016JM6008), by the Fundamental Research Funds for the Central Universities under Grant No. JBZ170401, and by the National Science Foundation under Grant No. ECCS-1405579.

References

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

[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]. Y. Liu, S. P. Liu, Z. F. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," Information Fusion 24 (2015) 147-164.

[5]. Y. Liu, X. Chen, R. K. Ward, Z. J. Wang, "Image fusion with convolutional sparse representation," IEEE Signal Processing Letters 23 (12) (2016) 1882-1886.

[6]. S. Li, H. Yin, L Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," IEEE Transactions on Biomedical Engineering 59 (12) (2012) 3450-3459.

[7]. Y. Yao, P. Guo, X. Xin, Z. Jiang, "Image fusion by hierarchical joint sparse representation," Cognitive Computation 6 (3) (2014) 281-292.

[8]. H. Yin, Y. Li, Y. Chai, Z. Liu, Z. Zhu, "A novel sparse-representation-based multi-focus image fusion approach," Neurocomputing 216 (2016) 216-229.

[9]. R. Ibrahim, J. Alirezaie, P. Babyn, "Pixel level jointed sparse representation with RPCA image fusion algorithm," in: Proceedings of International Conference on Telecommunications and Signal Processing, 2015, pp. 592-595.

[10]. 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.

[11]. M. Aharon, M. Elad, A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for spare representation," IEEE Transactions on Signal Processing 54 (11) (2006) 4311-4322.

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

[13]. S. Pertuz, D. Puig, M. A. Garcia, "Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images," IEEE Transactions on Image Processing 22 (3) (2013) 1242-1251.

[14]. J. Xiao, T. Liu, Y. Zhang, B. Zou, J. Lei, Q. Li, "Multi-focus image fusion based on depth extraction with inhomogeneous

diffusion equation," Signal Processing 125(2016) : 171-186.

[15]. S. Li, B. Yang, J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," Information Fusion 12 (2) (2011) 74-84.

[16]. Y. Liu, J. Jin, Q. Wang, Y. Shen, X. Dong, "Novel focus region detection method for multifocus image fusion using quaternion wavelet," Journal of Electronic Imaging 22 (2) (2013) 02317-1-02317-17.

[17]. J. Du, W. Li, B. Xiao, Q. Nawaz, "Union Laplacian pyramid with multiple features for medical image fusion," Neurocomputing 194 (2016) 326-339.

[18]. S. Sulochana, R. Vidhya, R. Manonmani, "Optical image fusion using support value transform (SVT) and curvelets," Optik 126 (18) (2015) 1672-1675.

[19]. A. Lutz, M. Ginasiracusa, N. Messer, S. Ezekiel, E. Blasch, M. Alford, "Optimal multi-focus contourlet-based image fusion algorithm selection," Proceedings of SPIE 9841 (2016) 98410E-1-98410E-8.

[20]. B. Zhang, X. Lu, H. Pei, H. Liu, Y. Zhao, W. Zhou, "Multi-focus image fusion algorithm based on focused region extraction, " Neurocomputing, 174, Part B (2016): 733-748.

[21]. Y. Liu, Z. F. Wang, "Simultaneous image fusion and denoising with adaptive sparse representation," IET Image Processing 9 (5) (2015) 347-357.

[22]. B. Yang, S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," Information Fusion 13 (1) (2012) 10-19.

[23]. M. Kim, D. K. Han, H. Ko, "Joint patch clustering-based dictionary learning for multimodal image fusion," Information Fusion 27 (2016) 198-214.

[24]. J. Wang, J. Peng, X. Feng, G. He, J. Fan, "Fusion method for infrared and visible images by using non-negative sparse representation," Infrared Physics & Technology 67 (2014) 477-489.

[25]. J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, "Robust face recognition via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009) 210-227.

[26]. Y. Cong, J. Yuan, J. Liu, "Abnormal event detection in crowded scenes using sparse representation," Pattern Recognition 46 (7) (2013) 1851-1864.

[27]. E. Elhamifar, R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (11) (2013) 2765-2781.

[28]. Z. Lin, R. Liu, Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in: Advances in Neural Information Processing System, 2011, pp. 1-9.

[29]. X. Ren, Z. Lin, "Linearized alternating direction method with adaptive penalty and warm starts for fast solving transform invariant low-rank textures, " International Journal of Computer Vision 104 (1) (2013) 1-14.

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

[31]. Y. Zhang, Z. Jiang, L.S. Davis, "Learning structured low-rank representations for image classification," in: Proceedings of

the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 676-683.

[32]. J. F. Cai, E. J. Candès, Z. Shen, “A singular value thresholding algorithm for matrix completion,” SIAM Journal on Optimization 20 (4) (2010) 1956-1982.

[33]. M. Yin, J. Gao, Z. Lin, “Laplacian regularized low-rank representation and its applications,” IEEE Transactions on Pattern Analysis & Machine Intelligence 38 (3) (2016) 504-517.

[34]. Q. Zhang, B. L. Guo, “Multifocus image fusion using the nonsubsampled contourlet transform,” Signal Processing 89 (7) (2009) 1334-1346.

[35]. H. Li, Y. Chai, H. Yin, G. Liu, “Multifocus image fusion and denoising scheme based on homogeneity similarity,” Optics Communications 285 (2) (2012) 91-100.

[36]. S. Li, X. Kang, J. Hu, B. Yang, “Image matting for fusion of multi-focus images in dynamic scenes,” Information Fusion 14 (2) (2013) 147-162.

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

[38]. G. H. Qu, D. L. Zhang, P. F. Yan, “Information measure for performance of image fusion,” Electronics Letters 38 (7) (2002) 313-315.

[39]. V. Xydeas, V. Petrovic, “Objective image fusion performance measure,” Electronic Letters 36 (4) (2000) 308-309.

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

[41]. Q. Zhang, Z. Ma, L. Wang, “Multimodality image fusion by using both phase and magnitude information,” Pattern Recognition Letters 34 (2) (2013) 185-193.

[42]. B. Zhang, Y. Yang, C. Chen, L. Yang, J. Han, and L. Shao, “Action recognition using 3d histograms of texture and a multi-class boosting,” IEEE Transactions on Image Processing, 26 (10) (2017) 4648–4660.

[43]. J. Han, E. Pauwels, P. de Zeeuw, and P. de With, “Employing a RGB-D sensor for real-time tracking of humans across multiple re-entries in a smart environment,” IEEE Transaction on Consumer Electronics, 58 (2) (2012) 255–263.

[44]. Y. Guo, G. Ding, L. Liu, J. Han, and L. Shao, “Learning to hash with optimized anchor embedding for scalable retrieval,” IEEE Transactions on Image Processing, 26 (3) (2017) 1344–1354.

[45]. Y. Guo, G. Ding, and J. Han, “Robust quantization for general similarity search,” IEEE Transactions on Image Processing, 27(2) (2018) 949–963.

[46]. X. Yao, J. Han, D. Zhang, and F. Nie, “Revisiting co-saliency detection: A novel approach based on two-stage multi-view spectral rotation co-clustering,” IEEE Transactions on Image Processing, 26(7) (2017) 3196–3209.

[47]. D. Zhang, J. Han, L. Jiang, S. Ye, and X. Chang, “Revealing event saliency in unconstrained video collection,” IEEE Transactions on Image Processing, 26 (4) (2017) 1746–1758.

[48]. J. Han, E. Pauwels, and P. de Zeeuw, “Fast saliency-aware multi-modality image fusion,” Neurocomputing 111 (2013) 70–80.

Qiang Zhang received the B.S. degree in automatic control, the M.S. degree in pattern recognition and intelligent systems, and the Ph.D. degree in circuit and system from Xidian University, China, in 2001,2004, and 2008, respectively. He was a Visiting Scholar with the Center for Intelligent Machines, McGill University, Canada. He is currently a professor with the Automatic Control Department, Xidian University, China. His current research interests include image processing, pattern recognition.

Tao Shi received the B.S. degree in Automatic Control from Xi'an Polytechnic University, Xi'an, China, in 2015. He is currently pursuing his M.S. degree in Control Engineering at Xidian University, Xi'an, China. His current research interests include computer vision and image fusion.

Fan Wang received the B. S. degree in Geographic Information System from Henan University, China, in 2010. He is currently working towards the M. S. degree in Control Engineering at Xidian University, Xian. His current research interests include computer vision and image fusion.

Rick S. Blum received a B.S. in Electrical Engineering from the Pennsylvania State University in 1984 and his M.S. and Ph.D. in Electrical Engineering from the University of Pennsylvania in 1987 and 1991. From 1984 to 1991 he was a member of technical staff at General Electric Aerospace in Valley Forge, Pennsylvania and he graduated from GE's Advanced Course in Engineering. Since 1991, he has been with the Electrical and Computer Engineering Department at Lehigh University in Bethlehem, Pennsylvania where he is currently a professor and holds the Robert W. Wieseman Endowed Professorship in Electrical Engineering. His research interests include signal processing for smart grid, communications, sensor networking, radar, image fusion and sensor processing. He was on the editorial board for the Journal of Advances in Information Fusion of the International Society of Information Fusion. He was an associate editor for IEEE Transactions on Signal Processing and for IEEE Communications Letters. He has edited special issues for IEEE Transactions on Signal Processing, IEEE Journal of Selected Topics in Signal Processing and IEEE Journal on Selected Areas in Communications. He was a member of the SAM Technical Committee (TC) of the IEEE Signal Processing Society. He was a member of the Signal Processing for Communications TC of the IEEE Signal Processing Society and was a member of the Communications Theory TC of the IEEE Communication Society. He was on the awards Committee of the IEEE Communication Society.

Dr. Blum is a Fellow of the IEEE, a former IEEE Signal Processing Society Distinguished Lecturer, an IEEE Third Millennium Medal winner, a member of Eta Kappa Nu and Sigma Xi, and holds several patents. He was awarded an ONR Young Investigator Award and an NSF Research Initiation Award. His IEEE Fellow Citation "for scientific contributions to detection, data fusion and signal processing with multiple sensors" acknowledges contributions to the fields of sensor processing and sensor networking.

Jungong Han is a tenured faculty member with the School of Computing and Communications at

Lancaster University, Lancaster, UK. Previously, he was a faculty member with the Department of Computer and Information Sciences at Northumbria University, UK. His research interests include computer vision, image processing, machine learning, and artificial intelligence.