

Multi-Focus Image Fusion Via Coupled Sparse Representation and Dictionary Learning

Rui Gao and Sergiy A. Vorobyov

Abstract

We address the multi-focus image fusion problem, where multiple images captured with different focal settings are to be fused into an all-in-focus image of higher quality. Algorithms for this problem necessarily admit the source image characteristics along with focused and blurred feature. However, most sparsity-based approaches use a single dictionary in focused feature space to describe multi-focus images, and ignore the representations in blurred feature space. Here, we propose a multi-focus image fusion approach based on coupled sparse representation. The approach exploits the facts that (i) the patches in given training set can be sparsely represented by a couple of overcomplete dictionaries related to the focused and blurred categories of images; and (ii) merging such representations leads to a more flexible and therefore better fusion strategy than the one based on just selecting the sparsest representation in the original image estimate. By jointly learning the coupled dictionary, we enforce the similarity of sparse representations in the focused and blurred feature spaces, and then introduce a fusion approach to combine these representations for generating an all-in-focus image. We also discuss the advantages of the fusion approach based on coupled sparse representation and present an efficient algorithm for learning the coupled dictionary. Extensive experimental comparisons with state-of-the-art multi-focus image fusion algorithms validate the effectiveness of the proposed approach.

Index Terms

Coupled sparse representations, dictionary learning, image fusion, multi-focus image.

R. Gao is with Aalto University, Dept. Signal Processing and Acoustics, FI-00076, AALTO, Finland. She is also with Northeastern University, Dept. Computer Application Technology, Shenyang 110819, China. E-mail: rui.gao@aalto.fi

S. A. Vorobyov is with Aalto University, Dept. Signal Processing and Acoustics, FI-00076, AALTO, Finland. E-mail: svor@ieee.org

I. INTRODUCTION

Over the last several decades, considerable attention has been given to the multi-focus image fusion problem [1]–[5]. Multi-focus image fusion is an effective post-processing technique for combining multiple images captured with different focal distances into an all-in-focus image, without sacrificing image quality, and at the same time without using specialized optic sensors [6]–[8]. The problem is of high importance in many fields, ranging from remote sensing to medical imaging [9]–[12], especially for addressing the demand for cost minimization of optical sensors/cameras. The majority of the existing literature on the topic can be categorized into two basic classes of approaches [13]: the *spatial frequency-based* and *transform domain-based* approaches. In the first class, the methods such as image fusion based on Laplacian pyramid (LP) [14], spatial frequency (SF) [15], multi-scale weighted gradient (MWG) [16], and variance [13] aim to directly select the best pixels or regions to fuse multiple images. The main drawback of these methods is that they may produce blocking artifacts because of misaligned decision map in focused boundary and wrong selection in sub-focused regions. Fusion methods from the second class suggest to apply multiscale transforms to decompose source images, and construct an all-in-focus image in the inverse transform domain. Corresponding algorithms include discrete wavelet transform (DWT) [17], curvelet transform (CVT) [9], non-subsampled contourlet transform (NSCT) [18] among others. However, all aforementioned traditional methods are still sensitive to image misregistration, and they may suffer from undesirable artifacts.

Looking at recent approaches, sparsity and overcompleteness have been successfully used for computational image fusion [19]–[27]. The methods exploit the fact that patches of natural image can be compactly represented using an overcomplete dictionary as a linear combination of only few *atoms*. It means that the vector of weighting coefficients for the atoms is sparse. Formally, the basic *synthesis model* suggests that the signal \mathbf{x} can be described as being a linear combination of few atoms over an overcomplete dictionary \mathbf{D} , and the problem of seeking such sparse representation can be formulated as

$$\min_{\alpha} \|\alpha\|_0 \quad \text{s.t. } \mathbf{x} \approx \mathbf{D}\alpha \quad (1)$$

where α is the sparse vector of coefficients applied to *atoms* in \mathbf{D} and $\|\cdot\|_0$ denotes the operator that counts the number of non-zero entries in a vector. As problem (1) is known to be NP-hard because of its combinatorial nature, a suboptimal strategy for addressing the above problem is based on replacing the operator $\|\cdot\|_0$ with l_1 -norm and learning adaptive

dictionary from the training data. Many image processing applications have benefited remarkably by using such approach with learned overcomplete dictionary. Representative examples include the K-SVD method [28], the method of optimal directions (MOD) [29], the online dictionary learning (OLD) method [30], and their variants. “Good” dictionaries are expected to be highly adaptive to the observed images/signals and to contribute accurate sparse representations. While the *single dictionary* model has been extensively studied, there is an advanced *coupled dictionary* viewpoint to sparsity and overcompleteness that has tracked the double feature space representation problem [31]–[34]. Corresponding approaches explicitly learn a pair of dictionaries from training sets of double image patches, and formulate sparse and redundant representations for various applications and scenarios. The combination of learned coupled dictionary and sparse approximation is shown to be superior for representing double feature spaces [35]–[38].

The aforementioned image fusion methods [19]–[25], [27] directly learn and exploit a single overcomplete dictionary in a single feature space in order to describe multiple images which contain both the focused and blurred categories of image features. Hence, these methods ignore the sparse representations in blurred feature space, and set limits on sparsity of the vector of coefficients. The latter consequently leads to less accurate fusing coefficients. These disadvantages, which are associated with working in single feature spaces, motivate us to perform fusion instead in double feature spaces. In particular, instead of learning a single overcomplete dictionary from the focused features only, this paper suggests to learn two dictionaries over focused and blurred feature spaces, and then use the pair of dictionaries to perform fusing via coupled sparse representation. In this way, we exploit the existing structure shared by all available multi-focus images, correlate the representations over double feature spaces, and significantly improve the fusion performance. Our approach based on dictionaries learned from double feature spaces and fusion in the space of sparse representations proves to be more accurate than traditional methods based on single learned dictionary.

A. Contributions

In this paper, we propose to extend the coupled dictionary learning approach based on coupled sparse and redundant representations to the problem of fusing multi-focus images. Coupled overcomplete dictionaries are expected to lead to more compact representation of the focused and blurred categories of images. Based on the coupled dictionary, a simple fusion of sparse coefficients based on plain averaging can be then used while seeking for optimal fusion coefficients

needed to reconstruct an all-in-focus image. The paper presents both algorithmic developments and simulation results for multi-focus image fusion. Compared to the existing prior works, the main difference of the proposed approach is that the coupled dictionary is exploited for more flexible and accurate representation of multi-focus images. Such sparse representation then lead to higher quality results for image fusion. Towards realizing this approach, this paper makes the following main contributions.

- We formulate the multi-focus image fusion problem as a problem of obtaining an all-in-focus image representation based on sparse representation over a couple (focused and blurred) of dictionaries. Sparse representations from different feature spaces are fused by obtaining joint sparse representation. The main difficulties to be addressed in such formulation are the coupled dictionary learning and fusion of corresponding sparse representations. In our conference paper [39], some initial results for the first contribution have been presented.
- We develop a fusion procedure for finding and merging sparse vectors of coefficients, which represent focused and blurred image feature spaces defined by learned coupled dictionary. Such procedure forces to produce a higher quality all-in-focus image by using the plain averaging. This contribution differs from the previous approaches on the basis of the sparsity model used, i.e., a coupled dictionary is used for fusion. Furthermore, we introduce a local sparsity in a global Bayesian reconstruction framework of our fusion procedure.
- We develop a K-SVD-based coupled dictionary learning algorithm by extending the well-known K-SVD algorithm [28]. It explicitly enforces the requirement that sparse approximations evaluated for double feature spaces must generate an all-in-focus image using corresponding pair of dictionaries. Our optimization algorithm simultaneously learns the focused and blurred dictionaries under the joint sparse coding requirement in order to enforce correlation between the dictionaries. Such coupled dictionary learning algorithm alternates between the updates of dictionary atoms and sparse coding.

B. Organization of the Paper and Notation

The remainder of the paper is organized as follows. Section II gives the problem description and summarizes some general assumptions. Section III gives a detailed explanation of our fusion procedure. Section IV focuses on the problem of coupled dictionary learning. Specifically, a K-SVD-based coupled dictionary learning algorithm is presented. Simulation results are provided in Section V. Finally, we conclude the paper in Section VI with summary of this work.

We use bold capital letters for matrices, for example, we denote k -th multi-focus and all-in-focus images as the following matrices of pixels \mathbf{I}_k and \mathbf{I}^F , respectively. Similarly, the matrices \mathbf{D}^F and \mathbf{D}^B denote the focus and blurred dictionaries, respectively. Bold lowercase letters stand for vectors. Two-dimensional image patches ordered lexicographically as column vectors are denoted, for example, as the following vectors \mathbf{i}_k and \mathbf{i}^F for patches of k -th multi-focus and all-in-focus images, respectively. The notation $[\mathbf{X}]_{i,j}$ is used to specify an element located on the intersection of i -th row and j -th column of matrix \mathbf{X} . We use $p(\mathbf{x})$ to denote the probability density function (pdf) of a random quantity \mathbf{x} and $p(\mathbf{x}|\mathbf{y})$ to denote the conditional probability of a random quantity \mathbf{x} conditioned to another quantity \mathbf{y} . The vector norm $\|\cdot\|_p$ for $p \geq 1$ is the standard l_p -norm, and the meaning of the notation $\|\cdot\|_0$ has been already introduced above. For a matrix \mathbf{X} , we define the Frobenius norm as $\|\mathbf{X}\|_F = (\sum_{i,j}([\mathbf{X}]_{i,j})^2)^{\frac{1}{2}}$. The symbol \odot represents element-wise product of matrices, $\nabla(\cdot)$ denotes the forward finite difference operator on the vertical and horizontal directions, and $(\cdot)^T$ stands for the transpose operation.

II. PROBLEM DESCRIPTION

Consider the problem of constructing/reconstructing a high quality all-in-focus image \mathbf{I}^F from a set of multi-focus source images $\{\mathbf{I}_k\}_{k=1}^K$, which can be abstractly written in the form of the following fusing process

$$\mathbf{I}^F = \mathcal{F}\{\mathbf{I}_k\}_{k=1}^K + \mathbf{V} \quad (2)$$

where $\mathcal{F}\{\cdot\}$ stands for a fusing operator and \mathbf{V} is a zero-mean independent and identically distributed (i.i.d.) Gaussian additive noise matrix, with entries drawn at random from the normal distribution $\mathcal{N}(0, \sigma^2)$. The fusion goal is to obtain \mathbf{I}^F from $\{\mathbf{I}_k\}_{k=1}^K$. Some further assumptions as to the capture of multiple images are the following. Each multi-focus image \mathbf{I}_k is captured for the same scene and all multi-focus images are properly aligned. Note that the latter assumption is typical in the literature with a focus on image fusion algorithms design, but proper alignment of images is also an important practical issue. The problem of finding optimal fusing operator $\mathcal{F}\{\mathbf{I}_k\}_{k=1}^K$ in (2) is an ill-posed problem in general. Therefore, different types of priors and regularizations have to be enforced and used.

Patch-Sparsity Prior: Images are typically processed by patches.¹ Image patches are easy to obtain from, for example, the all-in-focus image \mathbf{I}^F by multiplying the image with matrix $\mathbf{W}_{i,j}$

¹It is because adapting a dictionary to large size images is impractical.

that extracts (i, j) -th patch from an image by means of sliding window (moving through an image staring from the left-top corner of an image and moving to the right-bottom corner) technique. Formally, by multiplying \mathbf{I}^F with $\mathbf{W}_{i,j}$, we obtain $\mathbf{I}_{i,j}^F = \mathbf{W}_{i,j} \mathbf{I}^F$. Then the image patch of the size $d \times d$ pixels $\mathbf{I}_{i,j}^F \in \mathbb{R}^{d \times d}$ is ordered lexicographically as a column vector and denoted as $\mathbf{i}_{i,j}^F \in \mathbb{R}^{d^2}$. For notation simplicity and without loss of generality, we drop hereafter the indices (i, j) marking the patch position in an image and denote patches of k -th multi-focus image \mathbf{I}_k and all-in-focus image \mathbf{I}^F just as $\mathbf{i}_k \in \mathbb{R}^{d^2}$ and $\mathbf{i}^F \in \mathbb{R}^{d^2}$, respectively.

Base on the synthesis model, the patch \mathbf{i}^F can be represented as a linear combination of few atoms over an overcomplete dictionary \mathbf{D} . Then the problem of seeking a sparse representation of \mathbf{i}^F over the dictionary \mathbf{D} , i.e., the problem of finding sparse vector of coefficients $\boldsymbol{\alpha}$ for atoms of \mathbf{D} , is the same as problem (1). It can be written in the regularized form as

$$\min_{\boldsymbol{\alpha}} \|\mathbf{D}\boldsymbol{\alpha} - \mathbf{i}^F\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \quad (3)$$

where λ is the regularization parameter and $\lambda \|\boldsymbol{\alpha}\|_1$ is a regularization term enforcing sparsity of $\boldsymbol{\alpha}$. The dictionary here has to be learned in advance by using a training set of image patches or multiple training sets of image patches for images with different focuses.

Double Sparse Representation: Image patches in the multi-focus training sets can be sparsely represented by more than one overcomplete dictionary \mathbf{D} . Specifically, for the problem of constructing an all-in-focus image from a set of available multi-focus images, it makes sense to consider a couple of overcomplete dictionaries, $\mathbf{D}^F, \mathbf{D}^B \in \mathbb{R}^{d^2 \times N}$ of N atoms ($N > d^2$), corresponding to focus and blurred image features, respectively. Then, sparse representations can be found for each dictionary using patch-sparsity prior.

The benefit of using a coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$ versus using a single focused dictionary \mathbf{D}^F only, as in the existing fusion approaches [19]–[25], [27], is the following. The use of a singly dictionary implies that the coupled features in each multi-focus patch \mathbf{i}_k are described by one indiscriminate linear combination of rows from \mathbf{D}^F . This description lacks flexibility in characterizing the blurred image structures that negatively reflects on reconstruction quality. The coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$ is expected to overcome this limitation by enabling a more accurate reconstruction in terms of directly considering the coupled features in each multi-focus patch. Merging multiple sparse vectors of coefficients for image patch representations over the dictionaries corresponding to focused and blurred feature spaces is then better than just selecting the sparsest representation over a single dictionary alone, because a combination of focused and

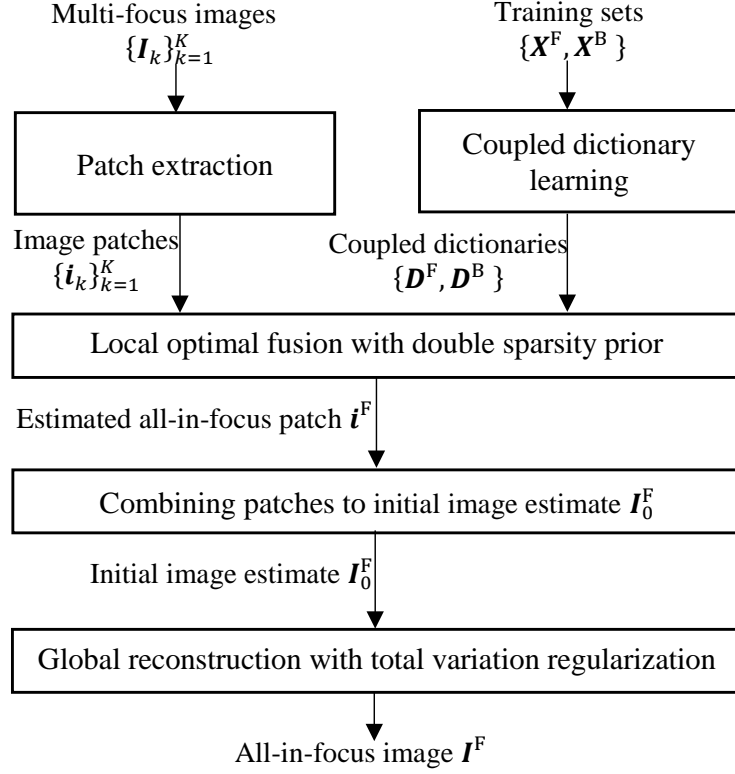


Fig. 1. Block-diagram of the procedure for reconstructing an all-in-focus image I^F from the given set of multi-focus images $\{I_k\}_{k=1}^K$ via coupled sparse representations in focus and blurred spaces and coupled dictionary learning.

blurred feature spaces gives a larger space and more degrees of freedom to the problem [41]. Of course, the problem of fusing such sparse representation of more than one dictionary then appears, and will be addressed in this paper.

Global Reconstruction: After merging multiple sparse vectors of coefficients for image patch representation over the dictionaries corresponding to focused and blurred feature spaces and finding corresponding patches of the desired all-in-focus image, the global reconstruction still has to be performed to ensure proximity between the whole reconstructed all-in-focus image and the available multi-focus images $\{I_k\}_{k=1}^K$. It is needed to make the entire reconstructed all-in-focus image consistent, and can be enforced by applying another type of regularization, which is the total variation regularization.

Summarizing, Fig. 1 shows the block-diagram of the procedure for constructing/reconstructing an all-in-focus image I^F from the given set of multi-focus images $\{I_k\}_{k=1}^K$. In this block-diagram, the block “Patch extraction” represents the above described simple process of extracting patches

applied to the available multi-focus images $\{\mathbf{I}_k\}_{k=1}^K$. The input to the block “Coupled dictionary learning” is a the set of two available subsets of M training patches, i.e., $\{\mathbf{X}^F, \mathbf{X}^B\}$. Here $\mathbf{X}^F \triangleq [\mathbf{x}_1^F, \mathbf{x}_2^F, \dots, \mathbf{x}_M^F] \in \mathbb{R}^{d^2 \times M}$ and $\mathbf{X}^B \triangleq [\mathbf{x}_1^B, \mathbf{x}_2^B, \dots, \mathbf{x}_M^B] \in \mathbb{R}^{d^2 \times M}$ are the subsets of available in-focus and blurred training patches, which can be extracted from actual images. The output of the block “Coupled dictionary learning” is the coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$ representing the in-focus and blurred image feature spaces. Section IV will be devoted to the coupled dictionary learning problem, which is otherwise decoupled from the image fusion problem addressed in the next section.

The procedure of fusing multi-focus images is presented in the block-diagram in Fig. 1 by three blocks. The block “Local optimal fusion with double sparsity prior” represents a procedure for finding and fusing vectors of sparse representations of patches of multi-focus images over the coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$. Knowing the dictionaries and the fused sparse vector of coefficients, we can reconstruct all patches of the all-in-focus image and combine them together to an initial estimate of the all-in-focus image denoted as \mathbf{I}_0^F , which corresponds in Fig. 1 to the block “Combining patches to initial image estimate \mathbf{I}_0^F ”. Finally, the block “Global reconstruction with total variation regularization” represents a procedure of global all-in-focus image reconstruction.

III. FUSION VIA COUPLED SPARSE REPRESENTATION

A. The Problem of Finding Fusing Operator

The problem of constructing/reconstructing a high quality clean all-in-focus image \mathbf{I}^F can be expressed in general as the following maximum a-posteriori probability (MAP) estimator design problem

$$\begin{aligned} \mathbf{I}^F &= \operatorname{argmax}_{\mathbf{I}^F} p\left(\mathbf{I}^F | \{\mathbf{I}_k\}_{k=1}^K\right) \\ &= \operatorname{argmax}_{\mathbf{I}^F} \exp \left\{ -\frac{1}{2\sigma^2} \left\| \mathbf{I}^F - \mathcal{F} \{ \mathbf{I}_k \}_{k=1}^K \right\|_F^2 \right\} \\ &= \operatorname{argmin}_{\mathbf{I}^F} \left\| \mathbf{I}^F - \mathcal{F} \{ \mathbf{I}_k \}_{k=1}^K \right\|_F^2. \end{aligned} \quad (4)$$

where the second row follows from model (2) and the assumption that the noise is Gaussian distributed, while the third row expresses the equivalence between the maximization of exponential function of a Frobenius norm square and minimization of the exponent, which is a Frobenius norm square, of exponential function. The MAP estimator (4) is equivalent to minimizing a well-defined global penalty and recovering/constructing \mathbf{I}^F in terms of finding the most probable set

of observations $\{\mathbf{I}_k\}_{k=1}^K$. Note that the main difficulty of solving (4) is that the fusing operator $\mathcal{F}\{\cdot\}$ is unknown and (4) is ill-posed.

Using the image patch double sparse representation discussed in the previous section for finding local optimal fusion and (4) for global reconstruction, the overall fusion via coupled sparse representation can be described in terms of the following optimization problem

$$\begin{aligned} \min_{\mathbf{I}^F, \boldsymbol{\alpha}^F} \sum_{i,j} \left\{ \|\mathbf{D}^F \boldsymbol{\alpha}^F - \mathbf{i}^F\|_2^2 + \lambda \|\boldsymbol{\alpha}^F\|_1 \right\} \\ + \left\| \mathbf{I}^F - \mathcal{F} \left\{ \begin{bmatrix} \mathbf{D}^F \\ \mathbf{D}^B \end{bmatrix}; \{\mathbf{I}_k\}_{k=1}^K \right\} \right\|_F^2 + \eta \rho(\mathbf{I}^F) \end{aligned} \quad (5)$$

where $\boldsymbol{\alpha}^F$ denotes the local sparse vector of coefficients needed for reconstructing the all-in-focus image patch \mathbf{i}^F over the focused dictionary \mathbf{D}^F , the summation $\sum_{i,j}$ means that sparse reconstruction error for all image patches is considered, $\rho(\mathbf{I}^F)$ stands for the penalty function that enforces prior knowledge about the entire all-in-focus image, and λ and η is the regularization parameters controlling the local (for image patch) sparsity penalty and the global (for the whole image) prior knowledge enforcement penalty, respectively. The notation $\mathcal{F} \left\{ \begin{bmatrix} \mathbf{D}^F \\ \mathbf{D}^B \end{bmatrix}; \{\mathbf{I}_k\}_{k=1}^K \right\}$ is used here to reflect the fact that fusion is performed for sparse representations of the images $\{\mathbf{I}_k\}_{k=1}^K$ over the coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$.

Problem (5) is separable to local optimal fusion and global reconstruction subproblems. Indeed, the first row in (5) includes image patches only, and therefore, the corresponding optimization problem is the local optimal fusion. The second row in (5) includes the whole image, and therefore, the corresponding optimization problem is the global reconstruction by which the proximity between the whole reconstructed all-in-focus image and the available multi-focus images $\{\mathbf{I}_k\}_{k=1}^K$ is ensured to make the entire all-in-focus image consistent. Then the penalty function $\rho(\mathbf{I}^F)$ takes here the form of generic regularization applied to the whole image, e.g., total variation regularization.

B. Local Optimal Fusion

Seeking the local optimal all-in-focus patch \mathbf{i}_F , we first need to find and consider the collection of sparse vectors of coefficients $\{\boldsymbol{\alpha}_k\}_{k=1}^K$ representing the patches $\{\mathbf{i}_k\}_{k=1}^K$ of the corresponding multi-focus images $\{\mathbf{I}_k\}_{k=1}^K$ over the coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$. Due to the fact that visible artifacts may occur on patch boundaries, overlapping patches that include pixels of neighboring patches are typically used to suppress such artifacts.

Fusing operator applied to the set of image patches $\{\mathbf{i}_k\}_{k=1}^K$ can be formulated as

$$\mathbf{i}^F \triangleq \mathcal{F} \left\{ \{\mathbf{i}_k\}_{k=1}^K \right\} = \mathbf{D}^F \cdot \underbrace{\mathcal{R} \left\{ \{\alpha_k\}_{k=1}^K \right\} \cdot \mathcal{L} \{\alpha_k; \mathbf{i}_k\}}_{\text{optimal } \alpha^F} \quad (6)$$

where the operator $\mathcal{L}\{\cdot\}$ yields the collections of sparse vectors of coefficients $\{\alpha_k\}_{k=1}^K$ representing corresponding patches $\{\mathbf{i}_k\}_{k=1}^K$, and the operator $\mathcal{R} \left\{ \{\alpha_k\}_{k=1}^K \right\}$ describes selection of the optimal sparse vector, i.e., α^F , from the collection of individual sparse vectors of coefficients for patches $\{\mathbf{i}_k\}_{k=1}^K$ obtained by $\mathcal{L}\{\cdot\}$, i.e., from $\{\alpha_k\}_{k=1}^K$.

In (6), we first need to find the operator $\mathcal{L}\{\cdot\}$. This operator must generate a collection of sparse representations of image patches over the coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$, and it can be expressed then as the following l_1 -norm minimization problem

$$\begin{aligned} \mathcal{L} \{\alpha_k; \mathbf{i}_k\} = & \underset{\alpha_k}{\operatorname{argmin}} \|\alpha_k\|_1 \\ \text{s.t. } & \|\mathbf{D}^F \cdot \mathcal{C} \{\alpha_k^F, \alpha_k^B\} - \mathbf{i}_k\|_2^2 \leq \epsilon \end{aligned} \quad (7)$$

where $\mathcal{C} \{\alpha_k^F, \alpha_k^B\}$ is the operator applied to the sparse representations α_k^F and α_k^B over the corresponding focused and blurred components of the coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$ and ϵ is a tolerance parameter of user's choice that defines an acceptable representation accuracy. The operator $\mathcal{C} \{\alpha_k^F, \alpha_k^B\}$ must provide the sparse vector of coefficients α_k , and it can be, for example, the plain averaging of the sparse vectors α_k^F and α_k^B . Since the plain averaging is a separable function in \mathbf{D}^F and \mathbf{D}^B , problem (7) can be decomposed and separately solved with respect to α_k^F and α_k^B , and then followed by finding the plain average of α_k^F and α_k^B . In other words, instead of (7), we first solve the problems

$$\begin{cases} \min_{\alpha_k^F} \|\alpha_k^F\|_1 & \text{s.t. } \|\mathbf{D}^F \alpha_k^F - \mathbf{i}_k\|_2^2 \leq \epsilon \\ \min_{\alpha_k^B} \|\alpha_k^B\|_1 & \text{s.t. } \|\mathbf{D}^B \alpha_k^B - \mathbf{i}_k\|_2^2 \leq \epsilon \end{cases} \quad (8)$$

and then find the plain average of α_k^F and α_k^B .

Problems (8) can be solved efficiently by many existing greedy methods, e.g., conventional orthogonal matching pursuit (OMP) algorithm [40]. However, specifically for our fusing process, each coefficient from α_k^F and α_k^B represents its own significant salient structure, and merging these coefficients must form a more accurate fusion result, which can be achieved by randomization. In particular, the two vectors of competitive representations α_k^F and α_k^B in (8) can be found by randomized orthogonal matching pursuit (RandOMP) method [41], [42]. According

to the RandOMP method as applied to (8), instead of adding the coefficients corresponding to the atoms (columns of \mathbf{D}^F and \mathbf{D}^B), denoted hereafter as \mathbf{d}_t^F and \mathbf{d}_t^B , respectively, to the corresponding support sets $\text{Support}\{\boldsymbol{\alpha}_k^F\}$ and $\text{Support}\{\boldsymbol{\alpha}_k^B\}$ by checking the minimum errors $\|\mathbf{d}_t^F \boldsymbol{\alpha}_k^F(t) - \mathbf{i}_k(t)\|_2^2$ and $\|\mathbf{d}_t^B \boldsymbol{\alpha}_k^B(t) - \mathbf{i}_k(t)\|_2^2$, respectively, for all candidates indexed with t , the RandOMP method makes a random choice of candidates \mathbf{d}_t^F and \mathbf{d}_t^B . Here $\boldsymbol{\alpha}_k^F(t)$ and $\boldsymbol{\alpha}_k^B(t)$ stand for t -th elements of the vectors $\boldsymbol{\alpha}_k^F$ and $\boldsymbol{\alpha}_k^B$, respectively, and $\mathbf{i}_k(t)$ is the t -th pixel of \mathbf{i}_k . The method starts with assigning empty support sets to the supports of $\boldsymbol{\alpha}_k^F$ and $\boldsymbol{\alpha}_k^B$, i.e., $\text{Support}\{\boldsymbol{\alpha}_k^F\} = \text{Support}\{\boldsymbol{\alpha}_k^B\} = \emptyset$; and assigning the residuals as $\mathbf{r}_F^{(0)} = \mathbf{r}_B^{(0)} = \mathbf{i}_k$. Then at each i -th iteration, atoms \mathbf{d}_t^F and \mathbf{d}_t^B are randomly and independently selected with probabilities proportional to $\exp\{\frac{c^2}{2\sigma^2}|(\mathbf{d}_t^F)^T \mathbf{r}_F^{(l-1)}|^2 / \|\mathbf{d}_t^F\|_2^2\}$ and $\exp\{\frac{c^2}{2\sigma^2}|(\mathbf{d}_t^B)^T \mathbf{r}_B^{(l-1)}|^2 / \|\mathbf{d}_t^B\|_2^2\}$, respectively. Here $c^2 = \sigma_k^2 / (\sigma_k^2 + \sigma^2)$ and σ_k is the variance of the non-zero entries of \mathbf{i}_k representation. After running through these steps L times, we obtain randomized solutions $L^{-1} \sum_{l=1}^L \boldsymbol{\alpha}_k^{F(l)}$ and $L^{-1} \sum_{l=1}^L \boldsymbol{\alpha}_k^{B(l)}$, respectively. Then by applying the plain averaging operator, we find the desirable sparse vector of coefficients

$$\boldsymbol{\alpha}_k = \mathcal{C}\{\boldsymbol{\alpha}_k^F, \boldsymbol{\alpha}_k^B\} = \frac{1}{2L} \left(\sum_{l=1}^L \boldsymbol{\alpha}_k^{F(l)} + \sum_{l=1}^L \boldsymbol{\alpha}_k^{B(l)} \right). \quad (9)$$

After obtaining each vector of sparse representation $\boldsymbol{\alpha}_k$ for the corresponding patch \mathbf{i}_k , the operator $R\{\{\boldsymbol{\alpha}_k\}_{k=1}^K\}$ is applied in accordance with (6) to find the optimal sparse vector of coefficients $\boldsymbol{\alpha}^F$. According to the general rule [23], the sparse vector of coefficients $\boldsymbol{\alpha}^F$ is the one from the set of vectors $\{\boldsymbol{\alpha}_k\}_{k=1}^K$ that has the largest l_1 -norm, that is,

$$\begin{aligned} \boldsymbol{\alpha}^F &= \mathcal{R}\{\{\boldsymbol{\alpha}_k\}_{k=1}^K\} = \underset{k \neq m}{\operatorname{argmax}} \{\|\boldsymbol{\alpha}_k\|_1, \|\boldsymbol{\alpha}_m\|_1\} \\ 1 &\leq k, m \leq n. \end{aligned} \quad (10)$$

Finally, the reconstructed all-in-focus patch \mathbf{i}^F can be found as

$$\mathbf{i}^F = \mathbf{D}^F \boldsymbol{\alpha}^F. \quad (11)$$

Going across the whole image, all reconstructed all-in-focus image patches \mathbf{i}_F can be found separately.

By focusing on the image patches, we have enhanced so far local details of the all-in-focus image, i.e., spatial edges, local textures. The global reconstruction then has to be performed according to the second row of (5).

C. Global Reconstruction

Given all reconstructed all-in-focus image patches \mathbf{i}_F , we can concatenate them trivially by placing to the corresponding positions to form the initial estimate of the all-in-focus image, which we denote as \mathbf{I}_0^F . In order to remove possible artifacts and improve spatial smoothness, the global reconstruction problem needs to be solved for ensuring the consistency between the initial estimation \mathbf{I}_0^F and the final outcome.

With the prior commonly used in the natural image analysis on the image gradients magnitude, the global reconstruction problem can be written as

$$\mathbf{I}^F = \underset{\mathbf{I}^F}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{I}^F - \mathbf{I}_0^F\|_2^2 + \eta \sum_{i,j} \rho(\|\nabla \mathbf{I}^F\|_{i,j}) \quad (12)$$

where $\rho(\cdot)$ takes the form of total variation, $[\nabla \mathbf{I}^F]_{i,j}$ denotes the discretization of the gradient for (i, j) -th element, defined as

$$\|[\nabla \mathbf{I}^F]_{i,j}\|_2 = \sqrt{[D_h \nabla \mathbf{I}^F]_{i,j}^2 + [D_v \nabla \mathbf{I}^F]_{i,j}^2} \quad (13)$$

with linear operators D_h and D_v representing finite difference approximations of the first-order horizontal and vertical partial derivatives [43].

Similar to the approaches in [43]–[45], optimization problem (12) can be efficiently solved by the alternating directions method of multipliers (ADMM) [46], [47], which decomposes a large global problem into a series of smaller local subproblems. The result \mathbf{I}_F from the above optimization is then taken as the final estimate of the all-in-focus image.

D. Summary of the Fusion Via Sparse Representation Algorithm

In summary, when the underlying dictionaries \mathbf{D}^F and \mathbf{D}^B are known, the optimal fusion via local sparse representations is first calculated. After that, the global reconstruction is employed for improving spatial smoothness of the reconstructed all-in-focus image. The overall multi-focus image fusion via sparse representation algorithm is summarized as Algorithm 1. Note that before reconstructing the all-in-focus image \mathbf{I}^F from the available multi-focus source images $\{\mathbf{I}_k\}_{k=1}^K$ using given/learned dictionaries \mathbf{D}^F and \mathbf{D}^B , we need to ensure that all images $\{\mathbf{I}_k\}_{k=1}^K$ are of the same intensity, which can be achieved by removing the mean intensity for source images as reflected in step 1 of Algorithm 1. At the end the mean intensity needs to be added to the reconstructed all-in-focus image as reflected in step 10 of Algorithm 1.

Algorithm 1 Image fusion via sparse representation.

Input: Multi-focus source images $\{\mathbf{I}_k\}_{k=1}^K$ and learned dictionaries \mathbf{D}^F and \mathbf{D}^B .

- 1: Remove the mean intensity for each source image.
- 2: **for** each d^2 (e.g., $d = 8$) image patch \mathbf{i}_k , taken starting from the left-top corner with p (e.g., $p = 1$) pixel overlap between neighboring patches,
- 3: Solve (8) to find α_k^F and α_k^B ;
- 4: Compute α_k using (9);
- 5: Apply the operator $\mathcal{R}\{\{\alpha_k\}_{k=1}^n\}$ using (10);
- 6: Find each patch \mathbf{i}_F using (11);
- 7: **end for**
- 8: Form the initial estimate of the all-in-focus image \mathbf{I}_0^F .
- 9: Perform the global reconstruction using (12).
- 10: Put the mean intensity into the reconstructed \mathbf{I}^F .

Output: The all-in-focus image \mathbf{I}^F .

IV. COUPLED DICTIONARY LEARNING

The description of the fusing process in the previous section has been based on the assumption that the coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$ is known. However, the dictionary also has to be learned as shown in the block-diagram in Fig. 1. In this section, the problem of the coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$ learning from a set of available in-focus and blurred image patches is addressed with the objective of designing a computationally efficient algorithm.

A. Coupled Dictionary Learning Problem Formulation

We aim to find a coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$ best representing a given available sets of training image patches for focused \mathbf{X}^F and blurred \mathbf{X}^B feature spaces. Specifically, we aim to find the best possible representation Γ for \mathbf{X}^F and \mathbf{X}^B in the coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$. Then the problem can be formulated as the following optimization problem

$$\begin{aligned}
 & \min_{\mathbf{D}^F, \mathbf{D}^B, \Gamma} \|\mathbf{X}^F - \mathbf{D}^F \Gamma\|_F^2 + \|\mathbf{X}^B - \mathbf{D}^B \Gamma\|_F^2 \\
 & \text{s.t. } \|\gamma_t^T\|_1 \leq T_0, \|\mathbf{d}_t^F\|_2^2 \leq 1, \|\mathbf{d}_t^B\|_2^2 \leq 1, \forall t
 \end{aligned} \tag{14}$$

where the matrix Γ is the joint sparse coding matrix for both \mathbf{X}^F and \mathbf{X}^B , \mathbf{d}_t^F and \mathbf{d}_t^B are as before t -th columns of \mathbf{D}^F and \mathbf{D}^B , respectively, γ_t^T is t -th row of Γ , and T_0 is the parameter controlling the sparsity.

Many numerical algorithms [31]–[34] have been developed to address problems of type (14) in alternating manner. Specifically for problem (14), the optimization variables can be split into two subsets of variables, where one subset includes the coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$, and the other consists of the sparse coding matrix Γ . By addressing (14) in alternating manner, we first initialize/update and fix $\{\mathbf{D}^F, \mathbf{D}^B\}$ and optimize (14) over Γ . Then Γ is fixed and (14) is optimized over $\{\mathbf{D}^F, \mathbf{D}^B\}$ until a stopping criterion is satisfied. Although the algorithms in [31]–[34] can be used for addressing (14) after some necessary modifications, their computational efficiency is not satisfactory.

The K-SVD algorithm [28] is another well-known alternating approach to overcomplete dictionary learning. It alternates between the so-called *sparse coding* of the training set (when the coupled dictionary is fixed) and *dictionary update* (when the matrix of sparse coefficients is fixed) steps. The K-SVD algorithm is computationally more efficient than the above mentioned methods, however, its computational efficiency can be further improved by using some simplifying structures present in (14). Specifically, in the dictionary update step, while updating each atom, the standard K-SVD algorithm has to update all non-zero coefficients of a corresponding column of Γ . Moreover, Γ is updated again at the sparse coding step. To avoid some redundant updates, we can introduce a so-called mask matrix [48] that consists of zeros and ones, and aims to keep all the non-zero column elements intact. It can be done by adding the following constraint

$$\Gamma \odot \mathbf{M} = 0 \quad (15)$$

to problem (14). Here, (i, j) -th element of the mask matrix \mathbf{M} is defined as

$$m_{i,j} \triangleq \begin{cases} 1, & \text{if } [\Gamma]_{i,j} = 0 \\ 0, & \text{if } [\Gamma]_{i,j} \neq 0 \end{cases}. \quad (16)$$

The dictionaries \mathbf{D}^F and \mathbf{D}^B capture the coherent structures of each corresponding individual feature space, and at the same time, describe correlation characteristics between the focused and blurred feature spaces. It facilitates the fusing process as explained in the previous section.

Our algorithm for addressing problem (14) with additional constraint (15) in alternating matter will contain two stages: (i) coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$ update (atom by atom updates) for fixed Γ ; and (ii) sparse coding for optimizing the matrix of joint sparse coefficients Γ for fixed coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$.

B. Dictionary Update

For fixed Γ , we solve (14) for finding the coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$. Since the objective function of (14) is separable with respect to the dictionaries \mathbf{D}^F and \mathbf{D}^B , and different sets of constraints are applied to the atoms of \mathbf{D}^F and \mathbf{D}^B , (14) can be split into two subproblems of finding updates to the dictionaries \mathbf{D}^F and \mathbf{D}^B separately. The corresponding optimization problems are given as

$$\mathbf{D}^F = \underset{\mathbf{D}^F}{\operatorname{argmin}} \left\| \mathbf{X}^F - \sum_{t=1}^N \mathbf{d}_t^F \gamma_t^T \right\|_F^2 \quad (17)$$

$$\mathbf{D}^B = \underset{\mathbf{D}^B}{\operatorname{argmin}} \left\| \mathbf{X}^B - \sum_{t=1}^N \mathbf{d}_t^B \gamma_t^T \right\|_F^2 \quad (18)$$

subject to the constraints in (14) applicable to corresponding atoms.

In (17) and (18), we rewrite the products $\mathbf{D}^F \Gamma$ and $\mathbf{D}^B \Gamma$ as the sums of vector outer products. After such modification, it appears to be not necessary to update all atoms simultaneously. Instead each atom can be updated disjoint from the others. Thus, to update the atoms \mathbf{d}_t^F and \mathbf{d}_t^B , we fix the remaining atoms, and rewrite optimization problems (17) and (18) as

$$\mathbf{d}_t^F = \underset{\mathbf{d}_t^F}{\operatorname{argmin}} \left\| \left(\mathbf{X}^F - \sum_{s \neq t} \mathbf{d}_s^F \gamma_s^T \right) \odot \mathbf{M}_t - \mathbf{d}_t^F \gamma_t^T \right\|_F^2 \quad (19)$$

$$\mathbf{d}_t^B = \underset{\mathbf{d}_t^B}{\operatorname{argmin}} \left\| \left(\mathbf{X}^B - \sum_{s \neq t} \mathbf{d}_s^B \gamma_s^T \right) \odot \mathbf{M}_t - \mathbf{d}_t^B \gamma_t^T \right\|_F^2 \quad (20)$$

where \mathbf{M}_t is built by repeating d times t -th row of the mask matrix \mathbf{M} , i.e., the row \mathbf{m}_t^T .

Using the mask matrix, all the columns of Γ , except t -th one, are removed, that is, $\mathbf{X}^F - \sum_{s \neq t} \mathbf{d}_s^F \gamma_s^T$ and $\mathbf{X}^B - \sum_{s \neq t} \mathbf{d}_s^B \gamma_s^T$, and only the atoms of \mathbf{D}^B and \mathbf{D}^F involved in the sparse representation picked by \mathbf{M} are updated. Since redundant and unnecessary updates of irrelevant atoms is avoided, the computational complexity can be significantly reduced. Then optimization

problems (19) and (20) can be further rewritten as the following simple rank-1 approximation problems

$$\mathbf{d}_t^F = \underset{\mathbf{d}_t^F}{\operatorname{argmin}} \left\| \mathbf{E}_t^F - \mathbf{d}_t^F \boldsymbol{\gamma}_t^T \right\|_F^2 \quad (21)$$

$$\mathbf{d}_t^B = \underset{\mathbf{d}_t^B}{\operatorname{argmin}} \left\| \mathbf{E}_t^B - \mathbf{d}_t^B \boldsymbol{\gamma}_t^T \right\|_F^2 \quad (22)$$

where $\mathbf{E}_t^F \triangleq \left(\mathbf{X}^F - \sum_{s \neq t} \mathbf{d}_s^F \boldsymbol{\gamma}_s^T \right) \odot \mathbf{M}_t$ and $\mathbf{E}_t^B \triangleq \left(\mathbf{X}^B - \sum_{s \neq t} \mathbf{d}_s^B \boldsymbol{\gamma}_s^T \right) \odot \mathbf{M}_t$ are the error matrices which do not include t -th atoms.

Problems (21) and (22) can be then easily solved by singular value decomposition (SVD) [28] of the corresponding error matrices. In particular, we first perform SVD for \mathbf{E}_t^F and \mathbf{E}_t^B , that is,

$$\mathbf{E}_t^F = \mathbf{U}_t^F \boldsymbol{\Delta}_t^F \mathbf{V}_t^F \quad (23)$$

$$\mathbf{E}_t^B = \mathbf{U}_t^B \boldsymbol{\Delta}_t^B \mathbf{V}_t^B. \quad (24)$$

Then the updated atoms \mathbf{d}_t^F and \mathbf{d}_t^B are given by the principal left singular vectors (first columns) of \mathbf{U}_t^B and \mathbf{U}_t^F , respectively. Also the product of the first column of \mathbf{V}_t^F with the largest/first singular value in $\boldsymbol{\Delta}_t^F$ and the product of the first column of \mathbf{V}_t^B with the first singular value in $\boldsymbol{\Delta}_t^B$ can be viewed as two non-sparse versions of $\boldsymbol{\gamma}_t^T$.

C. Sparse Coding

We yet need to find optimal sparse representation matrix $\boldsymbol{\Gamma}$. The problems for finding optimal $\boldsymbol{\gamma}_t^T$ that represents the image patches from \mathbf{X}^F and \mathbf{X}^B over the dictionaries \mathbf{D}^F and \mathbf{D}^B are, respectively, given as

$$\tilde{\boldsymbol{\gamma}}_t^T = \underset{\boldsymbol{\gamma}_t^T}{\operatorname{argmin}} \left\| \mathbf{X}^F - \sum_{t=1}^N \mathbf{d}_t^F \boldsymbol{\gamma}_t^T \right\|_F^2 + \lambda \left\| \boldsymbol{\gamma}_t^T \right\|_1 \quad (25)$$

$$\tilde{\boldsymbol{\gamma}}_t^T = \underset{\boldsymbol{\gamma}_t^T}{\operatorname{argmin}} \left\| \mathbf{X}^B - \sum_{t=1}^N \mathbf{d}_t^B \boldsymbol{\gamma}_t^T \right\|_F^2 + \lambda \left\| \boldsymbol{\gamma}_t^T \right\|_1. \quad (26)$$

The objective functions in problems (25) and (26) can be rewritten in terms of the corresponding error matrices \mathbf{E}_t^F and \mathbf{E}_t^B as $\left\| \mathbf{E}_t^F - \mathbf{d}_t^F \boldsymbol{\gamma}_t^T \right\|_F^2 + \lambda \left\| \boldsymbol{\gamma}_t^T \right\|_1$ and $\left\| \mathbf{E}_t^B - \mathbf{d}_t^B \boldsymbol{\gamma}_t^T \right\|_F^2 + \lambda \left\| \boldsymbol{\gamma}_t^T \right\|_1$, respectively. Moreover, both atoms from the focused and blurred dictionaries, i.e., \mathbf{d}_t^F and \mathbf{d}_t^B ,

must share the same sparse representation, i.e., the same γ_t^T , which means that problems (25) and (26) have to be solved jointly. Then the resulting optimization problem is

$$\min_{\gamma_t} \frac{1}{2} \|\mathbf{E}_t^F - \mathbf{d}_t^F \gamma_t^T\|_F^2 + \frac{1}{2} \|\mathbf{E}_t^B - \mathbf{d}_t^B \gamma_t^T\|_F^2 + \lambda \|\gamma_t^T\|_1. \quad (27)$$

Let us denote the whole objective function of (27) as $f(\gamma_t)$,² and split it into the following two parts $f_1(\gamma_t) \triangleq 0.5 \left(\|\mathbf{E}_t^F - \mathbf{d}_t^F \gamma_t^T\|_F^2 + \|\mathbf{E}_t^B - \mathbf{d}_t^B \gamma_t^T\|_F^2 \right) = \|\mathbf{E}_t - \mathbf{d}_t \gamma_t^T\|_F^2 / 2$ and $f_2(\gamma_t) \triangleq \lambda \|\gamma_t^T\|_1$, where $\mathbf{E}_t \triangleq \mathbf{E}_t^F + \mathbf{E}_t^B$ and $\mathbf{d}_t \triangleq \mathbf{d}_t^F + \mathbf{d}_t^B$. Then (27) can be efficiently addressed using block successive minimization method [49], which is a general framework for many large-scale optimization methods. Specifically, (27) is solved iteratively by finding a sequence $\{\gamma_t^{(q)}\}$ where the superscript q stands for the iteration index. At q -th iteration, the following optimization problem is solved

$$\gamma_t^{(q+1)} = \operatorname{argmin}_{\gamma_t} \mathcal{G}(\gamma_t, \gamma_t^{(q)}) \quad (28)$$

where $\mathcal{G}(\gamma_t, \gamma_t^{(q)})$ is a surrogate function for the function $f(\gamma_t)$ at the point $\gamma_t^{(q)}$, for which the conditions $\mathcal{G}(\gamma_t^{(q)}, \gamma_t^{(q)}) = f(\gamma_t^{(q)})$ and $\mathcal{G}(\gamma_t, \gamma_t^{(q)}) \geq f(\gamma_t)$ are satisfied.

Selecting $\mathcal{G}(\gamma_t, \gamma_t^{(q)})$ as

$$\mathcal{G}(\gamma_t, \gamma_t^{(q)}) \triangleq \nabla_{\gamma_t}^T f_1(\gamma_t^{(q)}) \gamma_t + \frac{1}{2} \|\gamma_t - \gamma_t^{(q)}\|_2^2 + f_2(\gamma_t) \quad (29)$$

the update rule for γ_t can be found to be

$$\gamma_t^{(q+1)} = \gamma_t^{(q)} - \mathcal{S}_{\frac{\lambda}{\mathcal{T}^2(\mathbf{d}_t)}} \left\{ \gamma_t^{(q)} - \frac{1}{\mathcal{T}^2(\mathbf{d}_t)} \mathbf{d}_t^T \left(\mathbf{d}_t (\gamma_t^{(q)})^T - \mathbf{E}_t \right) \right\}. \quad (30)$$

where the operator $\mathcal{S}_{\frac{\lambda}{\mathcal{T}^2(\mathbf{d}_t)}} \{\cdot\}$ is the soft thresholding operator with the threshold $\lambda/\mathcal{T}^2(\mathbf{d}_t)$ applied on the entries of its argument [50], [51] and $\mathcal{T}(\mathbf{d}_t) = \|\mathbf{d}_t\|$ to guarantee the convergence (see details in [49]). Since problem (27) is convex, the convergence to global optimum is guaranteed. Also note that with respect to $f_1(\gamma_t)$ only, the selection of $\mathcal{G}(\gamma_t, \gamma_t^{(q)})$ as (29) is the same as in the well-known mirror descent method. Therefore, (30) can be interpreted as a modified mirror descent for a composite objective function containing convex smooth, i.e., $f_1(\gamma_t)$, and convex non-smooth, i.e., $f_2(\gamma_t)$, parts, where we have the proximal soft thresholding operator due to the sparsity enforcing part $f_2(\gamma_t)$.

²We drop for simplicity the transpose of γ_t , but always mean that we work with rows of $\mathbf{\Gamma}$.

D. Summary of the Coupled Dictionary Learning Algorithm

Alternating between the dictionary update and sparse coding steps, the overall algorithm for coupled dictionary learning can be summarized as in Algorithm 2. The final update rules for both the dictionary update and sparse coding are obtained by using the most computationally efficient approaches from the state-of-the art approaches. Also the fast convergence for updates of type (30) is well documented fact (see the summary in [49]).

Algorithm 2 Coupled Dictionary Learning.

Input: Training sets of vectorized focus and blurred image patches \mathbf{X}^F and \mathbf{X}^B , the initial dictionaries \mathbf{D}_0^F and \mathbf{D}_0^B , and the initial sparse coding matrix Γ_0 .

- 1: **Initialization:** Set $\mathbf{D}^F := \mathbf{D}_0^F$, $\mathbf{D}^B := \mathbf{D}_0^B$, $\Gamma := \Gamma_0$.
- 2: **for** $t = 1 \cdots N$ **do**
- 3: Compute the error matrices \mathbf{E}_t^F and \mathbf{E}_t^B for atoms \mathbf{d}_t^F and \mathbf{d}_t^B , respectively;
- 4: Update the atoms \mathbf{d}_t^F and \mathbf{d}_t^B using (21) and (22) for all t ;
- 5: Define the surrogate function as given in (29);
- 6: **for** $q = 1$: maximum iterations **do**
- 7: Update rows of the sparse coding matrix Γ , i.e., γ_t^T , using (30) for all t ;
- 8: **end for**
- 9: **end for**

Output: The coupled dictionary $\{\mathbf{D}^F, \mathbf{D}^B\}$.

V. EXPERIMENTAL RESULTS

A. Experimental Setup

In this section, we first evaluate the proposed approach and compare it to existing state-of-the-art approaches in terms of visual comparison, and then discuss also quantitative assessments. We then discuss various factors that influence the performance including patch size, tolerance error, and number of overlapping pixels between neighboring patches. In addition, we test different approaches in terms of robustness to noise for different noise levels.

The quantitative assessments are based on two state-of-the-art fusion performance metrics.

- The measure of how well the mutual information from the source images is preserved in the fused image, denoted as Q_{MI} [52];

- The measure of how well the success of edge information transfers from the source images to the fused image, denoted as, $Q^{AB/F}$ [53].

The proposed approach is compared to the following existing state-of-the-art multi-focus image fusion algorithms.

- LP: Laplacian pyramid-based image fusion approach [14];
- MWG: multi-scale weighted gradient-based spatial image fusion approach [16];
- DWT: discrete wavelet transform-based image fusion approach [17];
- NSCT: non-subsampled contourlet transform-based image fusion approach [18];
- RPCA: robust principal component analysis-based image fusion approach [1];
- SR-CM: sparsely representation “choose-max”-based image fusion approach [23];
- SR-KSVD: sparsely representation K-SVD-based image fusion approach.

Throughout all experiments, the parameter used in the methods tested are set as follows. For the LP and DWT methods, the source images are decomposed to 3 levels. The wavelet basis “db1” is applied in the DWT algorithm. For the NSCT approach, the direction numbers at 4 decomposition levels from coarse to fine are selected as 4, 8, 8, and 16 for levels 1 to 4, respectively. For fair comparison, all algorithms based on sparsity model are implemented using the same dictionary size. For the proposed method, the tolerance error is set to $\epsilon = 0.1$, the patch size is equal to 8^2 , and the overlapping length between neighboring patches is 1. For the coupled dictionary learning, the training sets of focused image patches of size 8^2 consisting of 50,000 patches taken from the USC-SIPI image database [55] of 40 natural images, and 50,000 blurred patches of the same size, which are created from focused patches using Gaussian blur function, are used. The coupled dictionaries is then learned by Algorithm 2. Randomly selected patches from the training sets of focused and blurred sets of patches are used as initialization. The coupled dictionary size is selected to be 64×256 as a reasonable tradeoff between fusion quality and computation, while in general larger size dictionary results in heavier computation, but better fusion quality. Also for coupled dictionary learning, we execute 6 multiple dictionary update cycles (DUC) and 30 iterations per DUC. All our experiments are performed on a PC running a Inter(R) Xeon(R) 3.40GHz CPU.

B. Visual Comparison

We conduct our fusion experiments over the standard multi-focus dataset [54], and show some representative fusion results in Figs. 2 and 3. All the figures also include the magnified details of

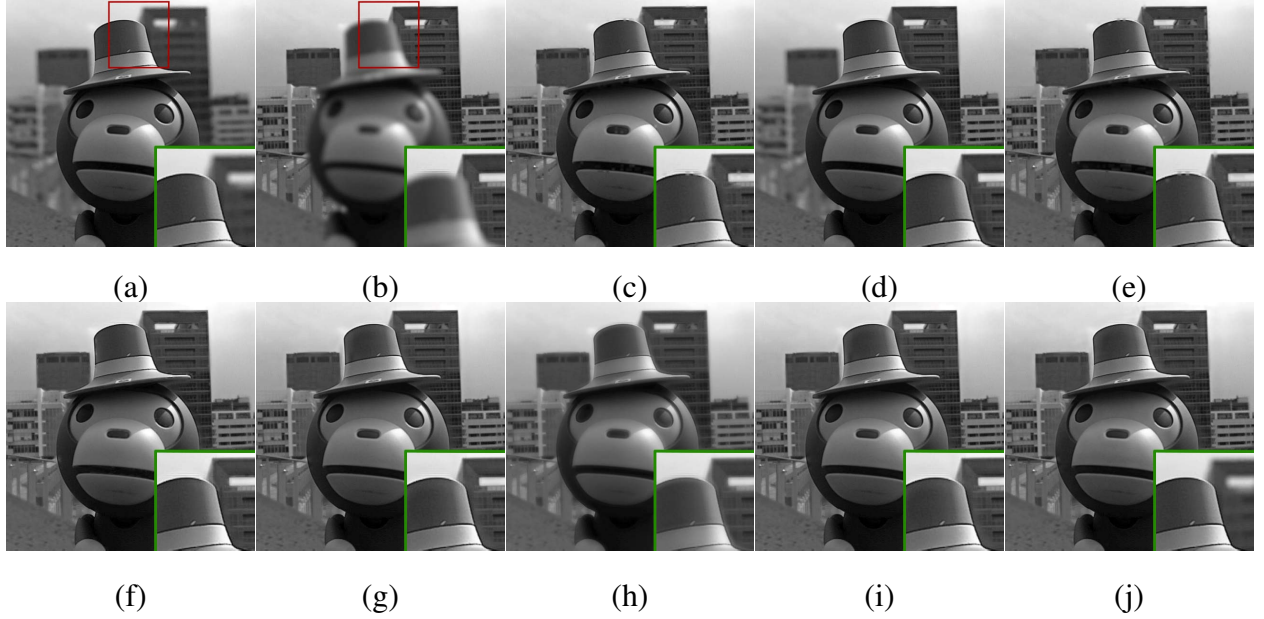


Fig. 2. Source images “Monkey” and fusion results. (a) First source image with focus on the background. (b) Second source image with focus on the front. Fused images obtained by LP (c), MWG (d), DWT (e), NSCT (f), RPCA (g), SR-CM (h), SR-KSVD (i), and the proposed method (j).

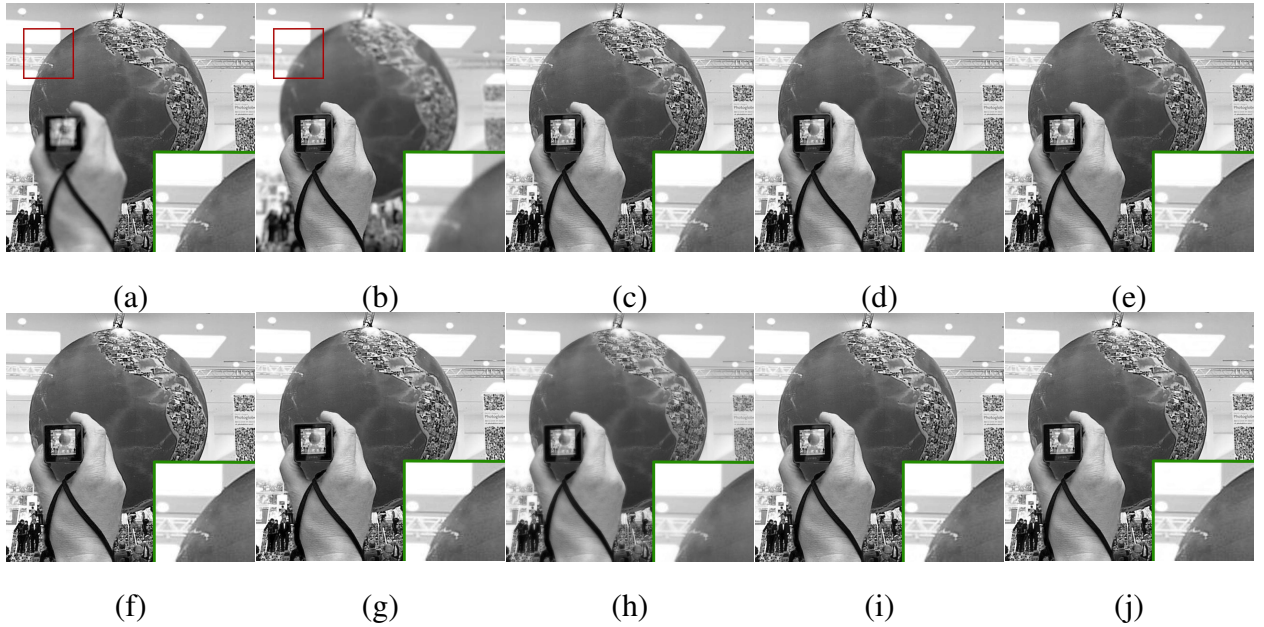


Fig. 3. Source images “Earth” and fusion results. Same order as in Fig. 2.

the images in the lower right corners. Figs. 2(a) and (b) show the source images, and the fused images obtained by different fusion methods tested are presented in Figs. 2(c)-(j). The same order of images and reconstructions obtained by different methods is also used in the Fig. 3. By inspecting the results visually, we can see that the LP method does not produce continuous

TABLE I
OBJECTIVE EVALUATION OF IMAGE FUSION PERFORMANCE

Methods	Measures in Fig. 2			Measures in Fig. 3		
	Q_{MI}	$Q^{AB/F}$	$Time(s)$	Q_{MI}	$Q^{AB/F}$	$Time(s)$
LP	0.9083	0.6879	0.0042	0.8998	0.6864	0.0053
MWG	0.9045	0.7243	0.8383	0.8992	0.7202	0.8424
DWT	0.8991	0.7013	0.1542	0.8876	0.6991	0.1784
NSCT	0.9245	0.7185	1.0748	0.9188	0.7101	1.1092
RPCA	0.9381	0.6620	0.0074	0.9311	0.6601	0.0102
SR-CM	0.9276	0.7214	3.1894	0.9241	0.7201	3.2785
SR-KSVD	0.9386	0.7326	3.1992	0.9381	0.7314	3.2198
Proposed	0.9432	0.7451	3.2123	0.9428	0.7439	3.4009

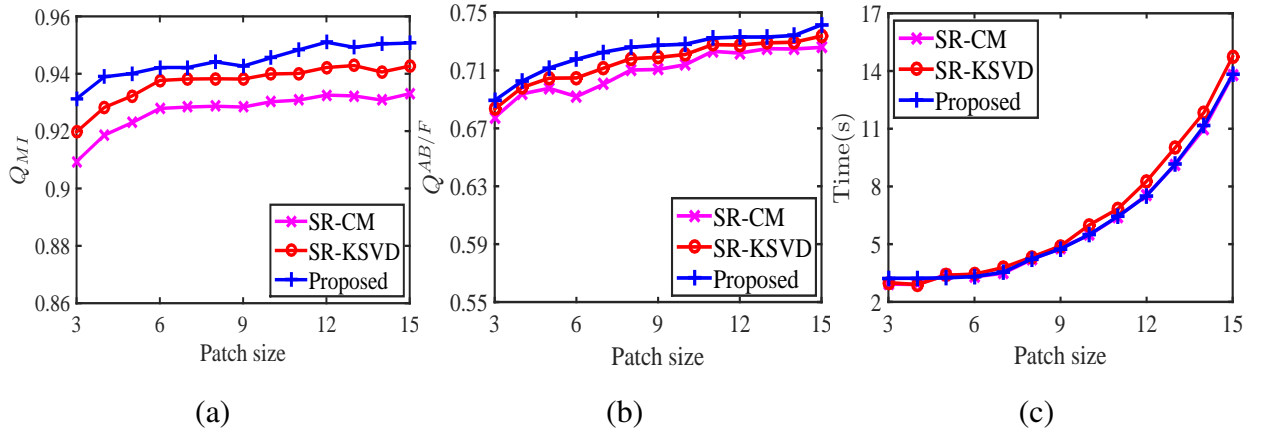


Fig. 4. Fusion performance vs. patch size for $\epsilon = 0.1$, $p = 1$.

edges (see Fig. 2(c)) and the DWT method results in blocking artifacts (see Fig. 2(e)). The same results for these two methods can be more clearly seen in the magnified sections of the images. The fusion methods based on MWG (see Fig. 2(d)) and NSCT (see Fig. 2(f)) show circle blurring effect around strong boundaries, while Fig. 2(f) appears visually better than Fig. 2(d). In Figs. 2(g) and (i), some artificial distortions can be seen, while Fig. 2(h) obviously lacks edge information. Our approach provides the best visual appearance (see Fig. 2(j)). Note that

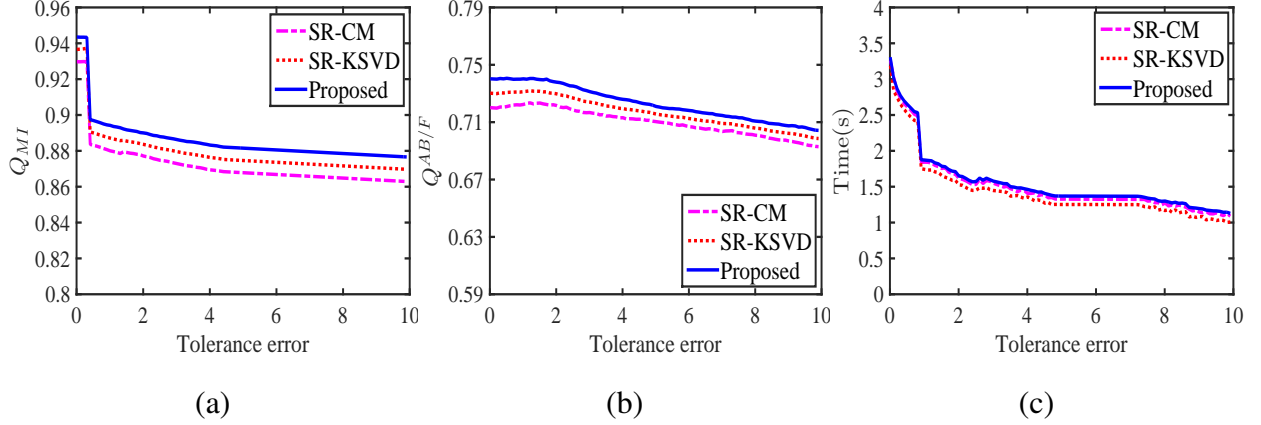


Fig. 5. Fusion performance vs. tolerance error for $d = 8$, $p = 1$.

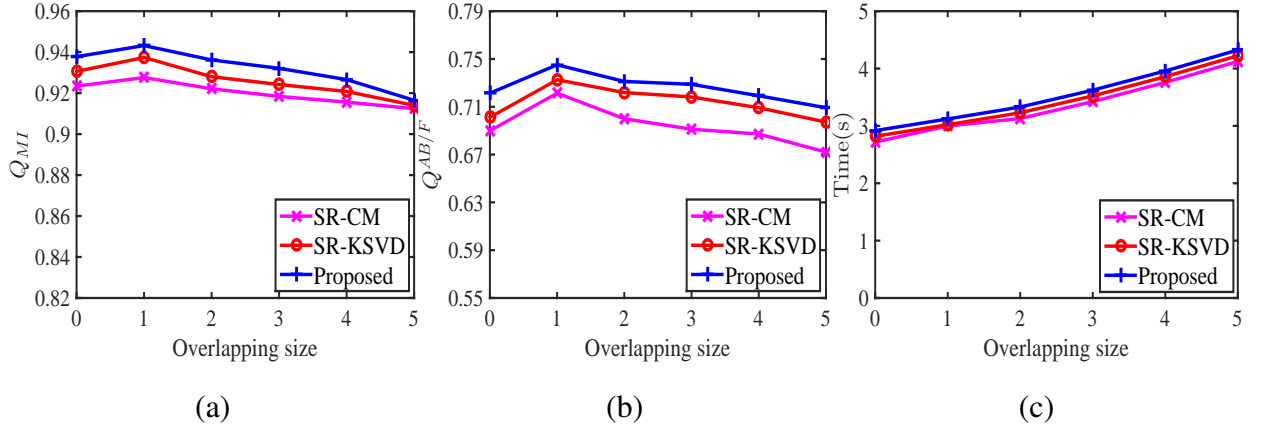


Fig. 6. Fusion performance vs. overlapping size for $\epsilon = 0.1$, $m = 8$.

similar visual subjective results are consistently obtained in our experiments with all the images tested including the other representative images shown in Fig. 3. Therefore, it can be concluded that the proposed approach yields better all-in-focus image quality than other tested methods, including state-of-the-art methods.

In addition to the visual comparisons, Table I summarizes the quantitative evaluations of different methods tested on the above grayscale multi-focus images in terms of Q_{MI} , $Q_{AB/F}$, and the running time (in seconds). The performance measures Q_{MI} and $Q_{AB/F}$ range from 0 to 1, with 1 representing the ideal fusion. The one bold value per each column of the table depicts the best result for a specific performance measure by a corresponding method. It can be seen from Table I that the proposed approach generally produces better quantitative results in terms of Q_{MI} and $Q_{AB/F}$. The value of Q_{MI} for the proposed approach is always larger than for the LP, MWG, DWT, NSCT, and SR-CM methods, and it is also better than for the

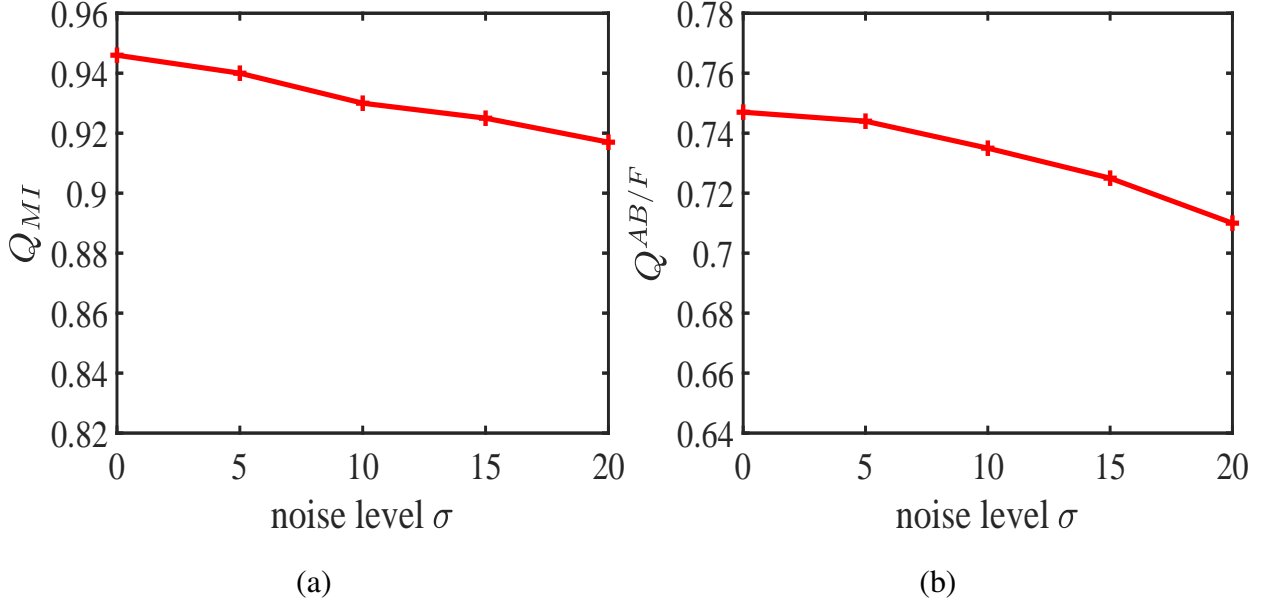


Fig. 7. Fusion performance vs. noise power σ .



Fig. 8. Source images “Friends” and fusion results for $\sigma = 15$. (a) First source image with focus on the right. (b) Second source image with focus on the left. Fused images obtained by SR-CM (c), SR-KSVD (d), and the proposed method (e).

methods based on RPCA and SR-KSVD. Therefore, the proposed approach as well as RPCA and SR-KSVD preserve well the mutual information from different source images. The values of $Q^{AB/F}$ demonstrate that the proposed approach reduces the blocking artifacts and artificial distortions, and combines the significant edge information into the fused image. Comparing the running time, it can be seen that the SR-CM, SR-KSVD, and proposed methods require more time than the other methods tested.

C. Effects of Main Parameters

The following three main parameters influence the fusion performance: patch size d^2 , tolerance error ϵ , and overlapping length between neighboring patches p . We compare here the best

performing methods, which are the SR-CM, SR-KSVD and the proposed methods, on the grayscale multi-focus image dataset which contains 10 pairs of grayscale images and 30 pairs of artificial source images with 256 level grey scales. Figs. 4– 6 show respectively the average values of Q_{MI} , $Q^{AB/F}$, and the running time for different values of pitch size, tolerance error, and overlapping length. We also give the corresponding parameter values for all the methods in the caption.

It can be seen in Figs. 4(a) and (b) that both Q_{MI} and $Q^{AB/F}$ are slightly better when the patch size is larger. As the patch size increases, the running time also increases (see Fig. 4(c)) for the proposed as well as for the SR-CM and SR-KSVD methods at about the same rate, while the SR-KSVD shows the worst time. Therefore, by trying to balance the computation time and fusion quality, we have set the patch size to 8^2 in our above reported experiments. It can be also seen in Fig. 5 that as the tolerance error ϵ increases, all the three methods tested here show degradation, as expected, in Q_{MI} and $Q^{AB/F}$, while the running time slightly reduces. Yet, the proposed method shows better results than the other two methods tested here. The results in Fig. 6 show that with the increase of the overlapping length between neighboring patches, Q_{MI} decreases, while the running time increases for all three methods tested here. The proposed method still show the best results. Therefore, we have set the overlapping size as $p = 1$ in our above experiments.

D. Robustness to Noise

To test how robust is the proposed method to noise, we add Gaussian noise to source images, and conduct fusion experiments for different noise levels $\sigma = \{0, 5, 10, 15, 20\}$. The results are shown in Fig. 7. It can be seen that with the increase of the noise level, Q_{MI} and $Q^{AB/F}$ gradually decrease.

For comparison, we also show the results of using the SR-CM and SR-KSVD methods for noisy images with $\sigma = 15$. As it can be seen in the fused images shown in Fig. 8, there are noticeable differences in the pores on the man's face. In particular, Figs. 8(c) and (d) show over-smoothing effect. Moreover, some details also disappear. Our approach performs the best as it visually appears in Fig. 8(e). Therefore, the proposed approach is capable of providing fusion and restoration simultaneously.

VI. CONCLUSION

We have presented a fusion algorithm for combining multiple images with different focal settings into one all-in-focus image. We first have formalized the physical process of capturing multi-focus images, and then developed a basic model based on coupled sparse representation of all-in-focus image. We have introduced a K-SVD-based coupled dictionary learning algorithm that enforces the sparse approximations for double feature spaces (focused and blurred). Using the coupled dictionary from the focused and blurred feature spaces, we have developed an efficient and accurate fusing approach, and have demonstrated that the proposed approach well preserves the edge and structural information of source images; drastically reduces the blocking artifacts, circle blurring, and artificial distortions; and shows in general better results than the existing fusion methods including state-of-the-art methods.

REFERENCES

- [1] T. Wan, C. Zhu, and Z. Qin, "Multifocus image fusion based on robust principal component analysis," *Pattern Recognit. Lett.*, vol. 34, no. 9, pp. 1001–1008, Jul. 2013.
- [2] Y. Liu, S. Liu, and Z. Wang, "Multi-focus image fusion with dense SIFT," *Inf. Fusion*, vol. 23, pp. 139–155, May 2015.
- [3] S. Pertuz, D. Puig, M. A. Garcia, and A. Fusiello, "Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1242–1251, Mar. 2013.
- [4] J. Tian, and L. Chen, "Multi-focus image fusion using wavelet-domain statistics," in *Proc. IEEE Int. Conf. Image Process.*, Hong Kong, 2010, pp. 1205–1208.
- [5] J. Tian, L. Chen, L. Ma, and W. Yu, "Multi-focus image fusion using a bilateral gradient-based sharpness criterion," *Opt. Commun.*, vol. 284, no. 1, pp. 80–87, Jan. 2011.
- [6] M. Subbarao, T. Choi, and A. Nikzad, "Focusing techniques," *Opt. Eng.*, vol. 32, pp. 2824–2836, Mar. 1993.
- [7] M. Born and E. Wolf, *Principles of Optics*. Cambridge Univ. Press., 1999.
- [8] Q. Zhang, and B. L. Guo, "Multifocus image fusion using the nonsubsampling contourlet transform," *Signal Process.*, vol. 89, pp. 1334–1346, Jul. 2009.
- [9] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inf. Fusion*, vol. 8, no. 2, pp. 143–156, Apr. 2007.
- [10] G. Pajares and J. Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.*, vol. 37, no. 9, pp. 1855–1872, Sep. 2004.
- [11] O. Rockinger, "Image sequence fusion using a shift-invariant wavelet transform," in *Proc. IEEE Int. Conf. Image Process.*, Santa Barbara, CA, 1997, pp. 288–291.
- [12] V. D. Calhoun and T. Adali, "Feature-based fusion of medical imaging data," *IEEE Trans. Inf. Technol. Biomedicine*, vol. 13, no. 5, pp. 711–720, Sep. 2009.
- [13] V. Aslantas and R. Kurban, "Fusion of multi-focus images using differential evolution algorithm," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8861–8870, Dec. 2010.
- [14] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.

- [15] S. Li, B. Yang, “Multifocus image fusion using region segmentation and spatial frequency,” *Inf. Fusion*, vol. 26, no. 7, pp. 971–979, Jul. 2008.
- [16] Z. Zhou, S. Li, and B. Wang, “Multi-scale weighted gradient-based fusion for multi-focus images,” *Image Vision Comput.*, vol. 20, pp. 60–72, Nov. 2014.
- [17] J. Tian and L. Chen, “Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure,” *Signal Process.*, vol. 92, no. 9, pp. 2137–2146, Sep. 2012.
- [18] Q. Zhang and B. Guo, “Multifocus image fusion using the nonsubsampling contourlet transform,” *Signal Process.*, vol. 89, no. 7, pp. 1334–1346, Jul. 2009.
- [19] T. Wan, N. Canagarajah, and A. Achim, “Compressive image fusion,” in *Proc. IEEE Int. Conf. Image Process.*, San Diego, CA, 2008, pp. 1308–1311.
- [20] S. Ambat, S. Chatterjee, and K. Hari, “Fusion of algorithms for compressed sensing,” *IEEE Trans. Signal Process.*, vol. 61, no. 14, pp. 3699–3704, May. 2010.
- [21] T. Wan, Z. Qin, C. Zhu, and R. Liao, “A robust fusion scheme for multifocus images using sparse features,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, British Columbia, Canada, 2013, pp. 1957–1961.
- [22] H. Li, L. Li, and J. Zhang, “Multi-focus image fusion based on sparse feature matrix decomposition and morphological filtering,” *Opt. Commun.*, vol. 342, pp. 1–11, May. 2015.
- [23] B. Yang and S. Li, “Multifocus image fusion and restoration with sparse representation,” *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 884–892, Apr. 2010.
- [24] M. Nejati, S. Samavi, and S. Hirani, “Multi-focus image fusion using dictionary-based sparse representation,” *Inf. Fusion*, vol. 25, pp. 72–84, Sep. 2015.
- [25] Q. Zhang, and M. D. Levine, “Robust multi-focus image fusion using multi-task sparse representation and spatial context,” *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2045–2058, Mar. 2016.
- [26] R. Gao, S. A. Vorobyov, and H. Zhao, “Image fusion with cosparsity analysis operator,” in *IEEE Signal Process. Lett.*, vol. 24, no. 7, pp. 943–947, July 2017.
- [27] L. Cao, L. Jin, H. Tao, G. Li, Z. Zhuang and Y. Zhang, “Multi-focus image fusion based on spatial frequency in discrete cosine transform domain,” *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 220–224, Sep. 2015.
- [28] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [29] K. Engan, S. O. Aase, and J. H. Husoy, “Method of optimal directions for frame design,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, Phoenix, AZ, USA, 1999, pp. 2443–2446.
- [30] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online dictionary learning for sparse coding,” in *Proc. ACM Int. Conf. Mach. Learn.*, Montreal, QC, Canada, 2009, pp. 689–696.
- [31] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, May. 2010.
- [32] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, “Coupled dictionary training for image super-resolution,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, Aug. 2012.
- [33] J. Sadasivan, S. Mukherjee, and C. S. Seelamantula, “Joint dictionary training for bandwidth extension of speech signals,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, Shanghai, China, 2016, pp. 5925–5929.
- [34] S. Wang, L. Zhang, Y. Liang, and Q. Pan, “Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis,” in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Rhode Island, USA, 2012, pp. 2216–2223.
- [35] T. Peleg, and M. Elad, “A statistical prediction model based on sparse representations for single image super-resolution,” *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2569–2582, Jun. 2014.

- [36] T. Kohler, X. Huang, F. Schebesch, A. Aichert, A. Maier, and J. Hornegger, "Robust multiframe super-resolution employing iteratively re-weighted minimization," *IEEE Trans. Comput. Imag.*, vol. 2, no. 1, pp. 42–58, Mar. 2016.
- [37] A. Mesaros, T. Heittola, O. Dikmen, and T. Virtanen, "Sound event detection in real life recordings using coupled matrix factorization of spectral representations and class activity annotations," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, South Brisbane, Australia, 2015, pp. 151–155.
- [38] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1553–1564, Mar. 2010.
- [39] R. Gao, S. A. Vorobyov, and H. Zhao, "Multi-focus image fusion via coupled dictionary training," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process.*, Shanghai, China, 2016, pp. 1666–1670.
- [40] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit," *CS Technion*, vol. 40, no. 8, pp. 1–15, Apr. 2008.
- [41] M. Elad and I. Yavneh, "A plurality of sparse representations is better than the sparsest one alone," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4701–4714, Oct. 2009.
- [42] T. Peleg, Y. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2286–2303, Feb. 2012.
- [43] A. Lanza, S. Morigi, F. Sgallari, "Convex image denoising via non-convex regularization," *Scale Sp. Var. Methods Comput. Vis.*, vol. 9087, Springer, pp. 666–677, 2015.
- [44] M. Nikolova, S. Esedoglu, and T. F. Chan, "Algorithms for finding global minimizers of image segmentation and denoising models," *SIAM J. Appl. Math.*, vol. 66, no. 5, pp. 1632–1648, Jun. 2006.
- [45] A. Parekh, and I. W. Selesnick, "Convex denoising using non-convex tight frame regularization," *IEEE Signal Process. Lett.*, vol. 22, no. 10, pp. 1786–1790, Oct. 2015.
- [46] J. Eckstein, and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 3, pp. 293–318, Nov. 1992.
- [47] S. Xie, and S. Rahardja, "Alternating direction method for balanced image restoration," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4557–4567, Nov. 2012.
- [48] L. N. Smith and M. Elad, "Improving dictionary learning: multiple dictionary updates and coefficient reuse," *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 79–82, Jan. 2013.
- [49] M. Razaviyayn, M. Hong and Z. Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, Jan. 2013.
- [50] D. Donoho, and J. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [51] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [52] M. Hossny, S. Nahavandi, and D. Creighton, "Comments on information measure for performance of image fusion," *Electron. Lett.*, vol. 44, no. 18, pp. 1066–1067, Aug. 2008.
- [53] C. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, Feb. 2000.
- [54] M. Nejati, S. Samavi, and S. Shirani, "Multi-focus image fusion using dictionary-based sparse representation," *Inf. Fusion*, vol. 25, pp. 72–84, Sep. 2015.
- [55] A. Weber. (1981). *The USC-SIPI Image Database* [Online]. Available: <http://sipi.usc.edu/database/>