Coupled Sparse Representations for Multi-Focus Image Fusion

Rui Gao, *Student Member, IEEE*, Sergiy A. Vorobyov, *Senior Member, IEEE*, and Hong Zhao

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. 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 novel multi-focus image fusion approach using coupled sparse representations. 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 is better than just selecting the sparsest one in the estimate of the original signal. By jointly learning the coupled dictionary, we enforce the similarity of sparse representations in the focused and blurred feature spaces, and then introduce a new 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 representations 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. As a demonstration of these benefits, we present high-quality image fusion results based on this new approach.

Index Terms

R. Gao is with Northeastern University, Dept. Computer application technology, Shenyang 110819, China. She is also with Aalto University, Dept. Signal Processing and Acoustics. E-mail: gaorui@research.neu.edu.cn

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

H. Zhao is with Northeastern University, Dept. Computer application technology, Shenyang 110819, China. E-mail: zhaoh@neusoft.com

Image fusion, coupled sparse representations, dictionary learning, multi-focus image.

I. INTRODUCTION

Over the last several decades, much research has been devoted to multi-focus image fusion problem [1]–[5]. Multi-focus image fusion is an effective technique for combining multiple images captured with different focal distances into an all-in-focus image, without sacrificing image quality or using specialized optic sensors [6]. The problem is of high importance in various applications, such as remote sensing, defense systems, and medical imaging [7]–[9]. One important paradigm for addressing the problem is fusion processing of multi-focus images. The majority of literature on the topicis traditionally categorized into two basic approaches [10]: the *spatial frequency-based* approaches and the *transform domain-based* approaches. In the first category, the methods such as image fusion based on Laplacian pyramid (LP) [11], spatial frequency (SF) [12], multi-scale weighted gradient (MWG) [13], and variance [10] 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 the misalignment of decision map with boundary of the focused object and wrong decision in sub-focused regions. The second category suggests to apply multiscale transforms to decompose source images, and construct an all-in-focus image in the inverse transform domain. Algorithms of this type include discrete wavelet transform (DWT) [14], curvelet transform (CVT) [7], non-subsampled contourlet transform (NSCT) [15] among others. However, all aforementioned traditional methods are still sensitive to image misregistration, and they may suffer from undesirable artifacts.

Until recently, sparsity and overcompleteness have been successfully used for image fusion [16]–[23]. The approaches exploit the fact that natural images can be compactly represented on an overcomplete dictionary as a linear combination of sparse coefficients. Formally, the basic model suggests that the signal $\boldsymbol{x} \in \mathbb{R}^d$ can be described as being a linear combination of few *atoms* over an overcomplete dictionary $\boldsymbol{D} \in \mathbb{R}^{d \times n}$. The processing can be formulated as sparse representation

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \quad \text{s.t. } \boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{\alpha} \tag{1}$$

where $\boldsymbol{\alpha}$ is the sparse vector of $\boldsymbol{x}$ applied to *atoms* in $\boldsymbol{D}$, and $\|\cdot\|_0$ is the l_0-norm which counts the non-zero entries. As the problem (1) is known to be NP-hard problem because of its combinational nature, a suboptimal strategy for the above problem is replacing l_0-norm

with l_1-norm and learning adaptive dictionary from the training data. Many image processing applications have benefited remarkably from learned dictionary. Representative algorithms on this topic include the K-SVD [24], the mothod of optimal derectons (MOD) [25], online dictionary learning (OLD) [26], and others. "Good" dictionaries are expected to be highly adaptive to the observed signals and to contribute accurate sparse representations. More recently, some studies have tracked the double feature space representation problem via learning coupled dictionary [27]–[30]. These approaches explicitly learn the pair of dictionaries from a training set of double images, and formulate sparse and redundant representations according to various applications and scenarios. The combination of learned coupled dictionary and sparse approximations proved the superior in the representations of double feature spaces [31]–[33].

Current image fusion methods based on sparse representations directly present one overcomplete dictionary in a single feature space to describe multi-focus images which contain the focused and blurred categories of features. The methods ignore the sparse representations in blurred feature space, and set the limits on the sparsity of the coefficients. That consequently leads to inaccuracy in the fusing coefficients. The difficulties on working in a single feature space motivate us to perform the fusion operations on double feature spaces. Similar to the sparsity-based methods, we will rely on the sparse coefficients from multi-focus source images. The success of such approach depends on how accurately it describes the multi-focus images and how efficiently it fuses the coefficients. Instead of directly learning one overcomplete dictionary from the focused images, this paper represents the first successful attempt to use coupled sparse representations into focused and blurred feature spaces, significantly improving the performance of image fusion method. Coupled sparse representations correlate the focused and blurred feature spaces, and bridge the fusion processes of double feature spaces.

Here, we propose a novel approach towards the multi-focus image fusion problem that utilizes coupled sparse and redundant representations over learned dictionary pair. The paper presents both the algorithmic developments and simulation results in multi-focus image fusion. Compared to prior work, the main difference of the proposed approach is that the coupled dictionary is exploited for more compact and accurate representations of different focus images. Such sparse representations will be shown to lead to higher quality results for fusion problem.

A. Contributions

Given the pair of training image sets, we seek for a couple of overcomplete dictionaries that lead to more compact representations for the focused and blurred categories of images. Based upon the coupled dictionary, we then use the plain averaging fusion criteria to find a more accurate sparse representation for reconstructing an all-in-focus image.

The main contributions of this work are threefold:

1) We cast the multi-focus image fusion problem formulation for achieving an all-in-focus image representation, based on sparsity over a couple of dictionaries. Information from different feature spaces is fused by using joint sparse coefficients. The main difficulties in the formulation are learning of coupled dictionary and fusing of corresponding sparse representations. In our conference paper [34], the first contribution has been presented.

2) We define a new fusion criteria that merging sparse coefficients over the couple of dictionaries. This criteria forces to produce a better estimated representation by plain averaging. As such, this contribution differs from previous approaches on the basis of used sparsity model. We also introduce the local sparsity in a global Bayesian reconstruction framework. The proposed criterion improves performance of multi-focus image fusion compared to some of the competing methods.

3) The K-SVD-based coupled dictionary learning algorithm is proposed such that we explicitly enforce the requirement that sparse approximations evaluated for double feature spaces can generate an all-in-focus image using their corresponding dictionaries. The proposed optimization algorithm simultaneously trains the focused and blurred dictionaries under the joint sparse coding in order to enforce collaborations between the dictionaries. A novel coupled dictionary learning algorithm alternates between the update of dictionaries atoms and sparse coding.

B. Organization of the Paper and Notation

The remainder of this paper is organized as follows. Section II gives the problem formulation and summarizes some general assumptions. Section III gives the detailed description of our proposed fusion criteria. In Section IV, we focus on jointly learning coupled dictionary. 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 a summary of this work and a description of its future extensions.

We generally use bold letters for vectors and hold capital letters for matrices. $\mathbb{R}^d$ and $\mathbb{R}^{d\times n}$ denote the set of real $d\times 1$ vectors and set of real $d\times n$ matrices, respectively. The notation $[\boldsymbol{x}]_{i,j}$ have been used to specify the element located in the i^{th} row and j^{th} column of the matrix $\boldsymbol{X}$. We use $p(\boldsymbol{x})$ to denote the probability density function (pdf) of random quantity $\boldsymbol{x}$, and use $\boldsymbol{X^T}$ to denote transpose. The vector norms $\|\cdot\|_p$ for $p \geq 1$ are the standard l_p-norms. The so-called l_0-norm, which counts the nonzero entries of its argument, is denoted by $\|\cdot\|_0$. The matrix norms $\|\cdot\|$ and $\|\cdot\|_F$ are used to denote the operator norm and the Frobenius norm, respectively. The symbol $\odot$ represents element-wise product of matrices, and the operator $\langle .,. \rangle$ means the inner product of vectors.

II. Problem Formulation

Consider the scenario with multi-focus source images $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \cdots, \boldsymbol{I}_n\}$ that acquire with different focal parameters. Our objective is to recover an all-in-focus image $\boldsymbol{I}_F$ from the observations $\{\boldsymbol{I}_k\}_{k=1}^n$ in the same scene. Assume that the observations $\{\boldsymbol{I}_k\}_{k=1}^n$ are properly aligned, and the sequence of multi-focus images and the unknown all-in-focus image are modelled as [3], [27], [35]

$$\boldsymbol{I}_k = \mathcal{F}(\boldsymbol{I}_F) \tag{2}$$

where the operator $\mathcal{F}$ is a blurring function [36], [37]. The mathematical model describes that the physical process of capturing n multi-focus images can be represented as the convolution of the all-in-focus image.

Assuming statistically independent observations, we now define the optimal solution for $\boldsymbol{I}_F$ based on maximum *a posteriori* probability (MAP)

$$\boldsymbol{I}_F = \underset{\boldsymbol{I_F}}{\operatorname{argmax}}\ p(\boldsymbol{I}_F|\{\boldsymbol{I}_k\}_{k=1}^n) \tag{3}$$

as the minimizer of a well-defined global penalty term. $p(\boldsymbol{I}_F|\{\boldsymbol{I}_k\}_{k=1}^n)$ models the observation mechanism that the estimate $\boldsymbol{I}_F$ is from the observations $\{\boldsymbol{I}_k\}_{k=1}^n$ which are statistically coupled to the transform outputs.

This paper has at the disposal two pieces of knowledge: the first one is that the patches in the training sets can be sparsely represented by a couple of overcomplete dictionaries, and the other one is that merging multiple sparse coefficients on one signal is better than just selecting the sparsest one alone [38]. Analysing the features on multi-focus images, we objectively define two

feature spaces: the focused feature space and the blurred feature space. A couple of overcomplete dictionaries $\left\{\boldsymbol{D^F}, \boldsymbol{D^B}\right\}$ for double spaces are jointly constructed by using the training sets from the focused and blurred feature spaces, with the same supports. With these analysis, we boost the fusion criteria based on the coupled sparse representations.

The solution of the MAP estimator (3) for multi-focus image fusion is built by

$$\begin{aligned}\left\{\boldsymbol{A}, \tilde{\boldsymbol{I}}_F\right\} = \underset{\boldsymbol{A}, \boldsymbol{I}_F}{\operatorname{argmin}} \|\boldsymbol{I}_F - \mathcal{F}(\{\boldsymbol{I}_k\}_{k=1}^n)\|_2^2 + \lambda\|\boldsymbol{A}\|_1 \\ + \left\|\boldsymbol{D^F} \cdot \mathcal{T}\left(\{\boldsymbol{I}_k\}_{k=1}^n ; \left\{\boldsymbol{D^F}, \boldsymbol{D^B}\right\}\right) - \boldsymbol{I}_F\right\|_2^2.\end{aligned} \tag{4}$$

where $\boldsymbol{A}$ represents the sparse coefficients of $\boldsymbol{I}_F$, $\mathcal{T}(\cdot)$ is the fusing criteria which merges sparse coefficients from $\{\boldsymbol{I}_k\}_{k=1}^n$ into $\boldsymbol{A}$, and λ is a regularization parameter. The first term in (4) is the log-likelihood global force that enforces the proximity between the estimated image $\boldsymbol{I}_F$ and its real version. The second and third terms are the image prior that means $\boldsymbol{I}_F$ has the compact sparse representations $\boldsymbol{A}$ from each multi-focus image $\boldsymbol{I}_k$, using the couple of overcomplete dictionaries $\left\{\boldsymbol{D^F}, \boldsymbol{D^B}\right\}$. The algorithm are highly efficient, due to the natural relation to learning the coupled of dictionaries. When $\boldsymbol{D^F}$ and $\boldsymbol{D^B}$ are not coupled, the fusion criteria $\mathcal{T}(\cdot)$ is unreliable.

For solving the problem (4), we divide it into two sub-problems: One problem is defining the fusion criteria $\mathcal{T}(\cdot)$ based on coupled sparse coefficients, and the second problem is learning the couple of dictionaries $\boldsymbol{D^F}$ and $\boldsymbol{D^B}$ for describing joint sparse representation $\boldsymbol{A}_k$.

In what follows, we will present the proposed fusion criteria, and learn the couple of overcomplete dictionaries $\boldsymbol{D^F}$ and $\boldsymbol{D^B}$ via jointly sparse coding.

III. Proposed Fusion Criteria

Assume that the couple of dictionaries $\boldsymbol{D^F}$ and $\boldsymbol{D^B}$ are known and fixed, the proposed penalty term in (4) has two kinds of unknown parameters: the compact sparse representation $\boldsymbol{A}$ of all-in-focus image $\boldsymbol{I}_F$, and each sparse coefficients $\boldsymbol{A}_k$ of multi-focus image $\boldsymbol{I}_k$. We start with the collection of the sparse representations $\boldsymbol{A}_k$ over the coupled dictionary $\left\{\boldsymbol{D^F}, \boldsymbol{D^B}\right\}$, and then seek the optimal $\boldsymbol{A}$ over the focused dictionary $\boldsymbol{D^F}$.

For convenience, the approach now introduces the operator $R(\{\boldsymbol{A}_k\}_{k=1}^n)$ to choose the optimal sparse coefficients, and then define the fusion criteria $\mathcal{T}$ as

$$\mathcal{T}(\{\boldsymbol{I}_k\}_{k=1}^n) \triangleq R(\{\boldsymbol{A}_k\}_{k=1}^n) \cdot \mathcal{L}(\boldsymbol{A}_k; \boldsymbol{I}_k) \tag{5}$$

where $\mathcal{L}(\boldsymbol{A}_k; \boldsymbol{I}_k)$ is a loss function which the image $\boldsymbol{I}_k$ should be sparely represented by the coefficient $\boldsymbol{A}_k$. Using one dictionary, as in the existing approach, implies that the focused and

blurred features in each image $\boldsymbol{I}_k$ are directly represented by one linear combination, and the optimal representation for blurred features is ignored. Hence, we intent to use a collection of sparse representations, with respect to the focused and blurred categories of dictionaries, in order to produce a better sparse representation.

Based on these observations, we build the averaging operator $\mathcal{C}$ on the sparse representations via the couple of dictionaries $\left\{\boldsymbol{D^F}, \boldsymbol{D^B}\right\}$ to seek the suitable sparse representation $\boldsymbol{A}_k$, and the representation is described by solving the following optimization function

$$\mathcal{L}(\boldsymbol{A}_k; \boldsymbol{I}_k) = \underset{\boldsymbol{A}_k}{\arg\min} \left\| \mathcal{C} \cdot \left[\boldsymbol{D}^F \vdots \boldsymbol{D}^B \right] \boldsymbol{A}_k - \boldsymbol{I}_k \right\|_2^2 + \lambda \| \boldsymbol{A}_k \|_1. \tag{6}$$

In doing so, we need to work on smaller patches of size $m \times m$ since adapting a dictionary to large signal is impractical. Using the sliding window technique, we divide the signal into small patches, from left-top to right-bottom. Visible artifacts may occur on block boundaries, and we also introduce overlapping patches for each $m \times m$ small patch on image $\boldsymbol{I}_k$.

Put formally, we decompose the loss function (6) into two sub-functions and operate them on each small patch. The sub-functions are described as

$$\left[\boldsymbol{a}_k^F\right]_{ij} = \underset{\left[\boldsymbol{a}_k^F\right]_{ij}}{\arg\min} \lambda \left\| \left[\boldsymbol{a}_k^F\right]_{ij} \right\|_1 + \left\| \boldsymbol{D}^F \left[\boldsymbol{a}_k^F\right]_{ij} - \mathbf{W}_{i,j} \boldsymbol{I}_k \right\|_2^2 \tag{7}$$

$$\left[\boldsymbol{a}_k^B\right]_{ij} = \underset{\left[\boldsymbol{a}_k^B\right]_{ij}}{\arg\min} \lambda \left\| \left[\boldsymbol{a}_k^B\right]_{ij} \right\|_1 + \left\| \boldsymbol{D}^B \left[\boldsymbol{a}_k^B\right]_{ij} - \mathbf{W}_{i,j} \boldsymbol{I}_k \right\|_2^2. \tag{8}$$

Here, $\mathbf{W}_{i,j}$ is a projection matrix that extracts the $(i,j)^{th}$ block from the image, and the choice for λ dictates that the errors $\left\| \boldsymbol{D}^F \left[\boldsymbol{a}_k^F\right]_{ij} - \mathbf{W}_{i,j} \boldsymbol{I}_k \right\|_2^2$ and $\left\| \boldsymbol{D}^B \left[\boldsymbol{a}_k^B\right]_{ij} - \mathbf{W}_{i,j} \boldsymbol{I}_k \right\|_2^2$ go below ϵ. We will see that the fusion quality has a close relationship with the tolerance error ϵ in Section V. Considering the function (7) and (8) respectively, we may employ the basic orthogonal matching pursuit (OMP) algorithm [39] to estimate the sparse representations with the residual energy at each iteration. In our scheme, we intend to build a bridge in relationships between $\left[\boldsymbol{a}_k^F\right]_{ij}$ and $\left[\boldsymbol{a}_k^B\right]_{ij}$, representing double kinds of feature space. Each set of sparse coefficients represents its own significant salient structures, and averaging these coefficients leads to a fusion of different features [38]. Hence, the solutions of the two functions can be jointly obtained by the randomized orthogonal matching pursuit (RandOMP) that generate two sets of competitive representations $\left[\boldsymbol{a}_k^F\right]_{ij}$ and $\left[\boldsymbol{a}_k^B\right]_{ij}$. Suppose at the j-th iteration, the t-th *atoms* $\left\| \boldsymbol{d}_t^F \left[\boldsymbol{a}_k^F\right]_{ij}(t) - \mathbf{W}_{i,j} \boldsymbol{I}_k(t) \right\|_2^2$ and $\left\| \boldsymbol{d}_t^B \left[\boldsymbol{a}_k^B\right]_{ij}(t) - \mathbf{W}_{i,j} \boldsymbol{I}_k(t) \right\|_2^2$ need to be computed. Instead of finding the smallest errors using

OMP algorithm, we randomize the choice of the t-th *atoms* with probability linearly proportional to

$$\exp\left\{\frac{c^2}{2\sigma^2}\cdot\frac{|\boldsymbol{d}_t^F\mathbf{W}_{i,j}\boldsymbol{I}_k^{j-1}|^2+|\boldsymbol{d}_t^B\mathbf{W}_{i,j}\boldsymbol{I}_k^{j-1}|^2}{\left\|\boldsymbol{d}_t^F\right\|_2^2+\left\|\boldsymbol{d}_t^B\right\|_2^2}\right\}, c^2=\frac{\sigma_k^2}{\sigma_k^2+\sigma^2} \tag{9}$$

where σ_k is the variance on the nonzero entries of the $\boldsymbol{I}_k$ representation. When the errors $\left\|\boldsymbol{D}^F\left[\boldsymbol{a}_k^F\right]_{i,j}-\mathbf{W}_{i,j}\boldsymbol{I}_k\right\|_2^2$ and $\left\|\boldsymbol{D}^B\left[\boldsymbol{a}_k^B\right]_{i,j}-\mathbf{W}_{i,j}\boldsymbol{I}_k\right\|_2^2$ go below ϵ, the randomization yields two sets of $\left\{\left[\boldsymbol{a}_k^F\right]_{i,j},\left[\boldsymbol{a}_k^B\right]_{i,j}\right\}$. Details of this processing are given in Appendix A.

Armed with the analysis, we push the step further to define the function $\mathcal{C}$ as the optimal value

$$\begin{aligned}\left[\boldsymbol{a}_k\right]_{i,j}&=\mathcal{C}\left(\left[\boldsymbol{a}_k^F\right]_{i,j},\left[\boldsymbol{a}_k^B\right]_{i,j}\right)\\&=\frac{\left[\boldsymbol{a}_k^F\right]_{i,j}+\left[\boldsymbol{a}_k^B\right]_{i,j}}{2}.\end{aligned} \tag{10}$$

Given all $\left[\boldsymbol{a}_k\right]_{i,j}$, we now turn to define the function $R(\boldsymbol{A}_k)$ and seek the sparse coefficient $\boldsymbol{A}$. In a general rule, the block with a bigger defined value is chosen to construct the fused image [20]. Here, we compute

$$\mathcal{R}\left(\left[\boldsymbol{a}_k\right]_{i,j}\right)=\underset{u\neq v}{\operatorname{argmax}}\left(\left\|\left[\boldsymbol{a}_u\right]_{i,j}\right\|_1,\left\|\left[\boldsymbol{a}_v\right]_{i,j}\right\|_1\right) \tag{11}$$

where $\left\|\left[\boldsymbol{a}_u\right]_{i,j}\right\|_1$ denotes the sum of the magnitude of its coefficients, and $1\leqslant u,v\leqslant n$. Returning to (5), we can get the fusion coefficient as

$$\mathcal{T}\left(\{\boldsymbol{I}_k\}_{k=1}^n\right)=\left[\boldsymbol{a}\right]_{i,j}. \tag{12}$$

In this strategy, the compact representations are built for each patches in the all-in-focus image. Gathering all $\boldsymbol{a}_{ij}$, the all-in-focus image is reconstructed as $\boldsymbol{I}_F^0=\boldsymbol{D}^F\boldsymbol{A}$. Up to here, we have enhanced the local details (i.e.,spatially edges, local textures) observed in the all-in-focus image.

In order to remove possible artifacts and improve spatially smoothness, the global reconstruction constraint as well as the consistency between the initial estimation $\boldsymbol{I}_F^0$ and the final outcome can be applied to make a further improvement. In this paper, we utilize the gradient-descent based minimization [43] to solve this function as the replacement of the function (4), given by

$$\tilde{\boldsymbol{I}}_F=\underset{\boldsymbol{A},\boldsymbol{I}_F}{\operatorname{argmin}}\left\|\boldsymbol{I}_F-\mathcal{F}(\{\boldsymbol{I}_k\}_{k=1}^n)\right\|_2^2+\lambda\left\|\boldsymbol{I}_F^0-\boldsymbol{I}_F\right\|_2^2. \tag{13}$$

Hence, the overall multi-focus image fusion algorithm is summarized as Algorithm 1.

Algorithm 1 Image fusion from sparsity.

Input: the couple of learned dictionaries $\boldsymbol{D}^F$ and $\boldsymbol{D}^B$; multi-focus source images $\{\boldsymbol{I}_k\}_{k=1}^n$.

1: Remove the mean intensity for each source image.

2: **for** each 8×8 patch $\boldsymbol{i}_k$ from $\boldsymbol{I}_k$, starting from the upper-left corner with 1 pixel overlap,

· Solve the optimization problems (7), (8);

· Compute the averaging coefficients using (10);

· Solve the operator $\mathcal{R}(\boldsymbol{a}_{i,j})$, using (11);

· Solve the fusion criteria $\mathcal{T}(\{\boldsymbol{I}_k\}_{k=1}^n)$ using (4), (12);

· Reconstruct the initial all-in-focus image $\boldsymbol{I}_F^0$.

3: **end for**

4: Build the final image using (13).

5: Put the mean intensity into the all-in-focus image $\boldsymbol{I}_F$.

Output: the all-in-focus image $\boldsymbol{I}_F$

IV. Learning Coupled Dictionary

The entire definition on fusion criteria so far is based on the assumption that the couple of dictionaries $\boldsymbol{D^F}$ and $\boldsymbol{D^B}$ from the focused and blurred feature spaces are known and fixed. So a key issue to achieving the fusion criteria is the couple of dictionaries to learn. In this section, we focus on the coupled dictionary learning problem and more specifically, the improvement of learning efficiency.

Assuming that the sets $\boldsymbol{X}^F = \left[\boldsymbol{x}_1^F, \boldsymbol{x}_2^F, \ \cdots, \boldsymbol{x}_n^F\right] \in \mathbb{R}^{d\times n}$ and $\boldsymbol{X}^B \triangleq \left[\boldsymbol{x}_1^B, \boldsymbol{x}_2^B, \ \cdots, \boldsymbol{x}_n^B\right] \in \mathbb{R}^{d\times n}$ as the matrix of n sampled focused image vectors emerge from focused and blurred feature spaces, respectively and d is the dimension of the sampled image vectors. The aim of coupled dictionary learning is to seek the pair of dictionaries $\boldsymbol{D}^F$, $\boldsymbol{D}^B \in \mathbb{R}^{d\times N}$ adapted to two sets of focused images $\boldsymbol{X}^F$ and blurred images $\boldsymbol{X}^B$ with respect to the best possible representation $\boldsymbol{\Gamma}$.

The basic coupled dictionary learning is formulated as

$$\begin{aligned} &\min_{\boldsymbol{D}^F,\boldsymbol{D}^B,\boldsymbol{\Gamma}} \left\|\boldsymbol{X}^F - \boldsymbol{D}^F\boldsymbol{\Gamma}\right\|_2^2 + \left\|\boldsymbol{X}^B - \boldsymbol{D}^B\boldsymbol{\Gamma}\right\|_2^2 \\ &\text{s.t. } \|\boldsymbol{\Gamma}\|_0 \leqslant T_0, \left\|\boldsymbol{d}_i^F\right\|_2^2 \leqslant 1, \left\|\boldsymbol{d}_i^B\right\|_2^2 \leqslant 1, \forall 1 \leqslant i \leqslant N \end{aligned} \tag{14}$$

where $\boldsymbol{\Gamma}$ is the joint spare coding of $\boldsymbol{X}^F$ and $\boldsymbol{X}^B$, $\boldsymbol{d}_i^F$ and $\boldsymbol{d}_i^B$ are the i-th columns of $\boldsymbol{D}^F$ and $\boldsymbol{D}^B$, respectively, and T_0 is the parameter controlling the sparsity penalty. Many algorithms [27]–[30], [32] have been introduced to solve the problem (14) for learning the coupled dictionary.

The coefficients are estimated via l_1 minimization, keeping the two dictionaries fixed alternately, and the dictionaries are estimated through least squares, keeping the coefficients fixed. Though these algorithms benefit from having the reliability, their efficiency still need to be improved.

The K-SVD [24] is a well-known and frequently-employed tool that aims to handle dictionary learning problem. It updates the dictionary atoms and the corresponding coefficients alternately, and improves the atoms to better fit the signals efficiently. Particularly, the algorithm updates one atom and all the non-zero columns corresponding to the atom at a time in dictionary updating stage, and then sparse approximation updates the non-zero columns again in sparse coding stage. Modifying the redundant updating, the improved K-SVD algorithm [40] introduces a mask matrix of zeros and ones, and keeps all the non-zero columns intact. Armed with this algorithm, we propose an improved K-SVD-based coupled dictionary learning approach.

Formally, we add additional constraint on $\mathbf{\Gamma}$ for keeping all the zero atoms in $\mathbf{\Gamma}$ intact, given by

$$\mathbf{\Gamma} \odot \mathbf{M} = 0 \tag{15}$$

where the mask matrix $\mathbf{M}$ is defined as

$$\mathbf{M}_{i,j} \triangleq \begin{cases} 1, & \text{if } \mathbf{\Gamma}_{i,j} = 0 \\ 0, & \text{if } \mathbf{\Gamma}_{i,j} \neq 0 \end{cases}. \tag{16}$$

Thus, our coupled dictionary learning problem is expressed as

$$\begin{aligned} &\min_{\boldsymbol{D}^F, \boldsymbol{D}^B, \mathbf{\Gamma}} \left\| \boldsymbol{X}^F - \boldsymbol{D}^F \mathbf{\Gamma} \right\|_2^2 + \left\| \boldsymbol{X}^B - \boldsymbol{D}^B \mathbf{\Gamma} \right\|_2^2 \\ &\text{s.t. } \left\| \boldsymbol{d}_i^F \right\|_2^2 \leqslant 1, \ \left\| \boldsymbol{d}_i^B \right\|_2^2 \leqslant 1, \| \mathbf{\Gamma} \|_0 \leqslant T_0, \mathbf{\Gamma} \odot \mathbf{M} = 0. \end{aligned} \tag{17}$$

We expect the learned atoms of dictionaries $\boldsymbol{D}^F$, $\boldsymbol{D}^B$ can describe the coherent structures of each individual feature space and correlation characteristics between the focused feature space and the blurred feature space, and thus and facilitate the fusion criteria. The objective problem (17) has three parameters to optimize: the atoms of coupled dictionaries $\boldsymbol{D}^F$, $\boldsymbol{D}^B$, and joint sparse coefficient $\mathbf{\Gamma}$ that linearly relate those atoms to the learning data sets $\boldsymbol{X}^F$ and $\boldsymbol{X}^B$. Solving these parameters, the proposed approach involves two main iteration steps that (i) the atoms of the coupled dictionaries $\boldsymbol{D}^F$ and $\boldsymbol{D}^B$ are alternately updated by the sparse coefficients; and (ii) the joint sparse coefficient $\mathbf{\Gamma}$ is given by fixing the coupled dictionaries.

A. Dictionary update

Let us first optimize the dictionary update steps where we assume that $\mathbf{\Gamma}$ is fixed, and consider the optimization problem (17) as processing the update of two dictionaries with the coefficients summarized in $\mathbf{\Gamma}$, which are described as

$$\tilde{\boldsymbol{D}}^F = \underset{\boldsymbol{D}^F}{\operatorname{argmin}} \left\| \boldsymbol{X}^F - \sum_{i=1}^{N} \boldsymbol{d}_i^F \boldsymbol{\gamma}_i^T \right\|_2^2 \tag{18}$$

$$\tilde{\boldsymbol{D}}^B = \underset{\boldsymbol{D}^B}{\operatorname{argmin}} \left\| \boldsymbol{X}^B - \sum_{i=1}^{N} \boldsymbol{d}_i^B \boldsymbol{\gamma}_i^T \right\|_2^2 . \tag{19}$$

Based on the prior knowledge of sparse representation, it is not necessary to update the whole atoms in each atom updating step. For optimizing $\boldsymbol{d}_i^F$ and $\boldsymbol{d}_i^B$ in each i-th *atom*, we fix the remaining *atoms* $\sum_{j\neq i}\boldsymbol{d}_i^F$, $\sum_{j\neq i}\boldsymbol{d}_i^B$ and the corresponding coefficients $\sum_{j\neq i}\boldsymbol{\gamma}$, and then rewrite the functions (18) and (19) as

$$\tilde{\boldsymbol{d}}_i^{\,F} = \underset{\boldsymbol{d}_i^F}{\operatorname{argmin}} \left\| \left(\boldsymbol{X}^F - \sum_{j\neq i} \boldsymbol{d}_j^F \boldsymbol{\gamma}_j^T \right) \odot \mathbf{M}_i - \boldsymbol{d}_i^F \boldsymbol{\gamma}_i^T \right\|_2^2 \tag{20}$$

$$\tilde{\boldsymbol{d}}_i^{\,B} = \underset{\boldsymbol{d}_i^B}{\operatorname{argmin}} \left\| \left(\boldsymbol{X}^B - \sum_{j\neq i} \boldsymbol{d}_j^B \boldsymbol{\gamma}_j^T \right) \odot \mathbf{M}_i - \boldsymbol{d}_i^B \boldsymbol{\gamma}_i^T \right\|_2^2 \tag{21}$$

where $\mathbf{M}_i \triangleq \mathbf{1}_d \cdot \mathbf{m}_i^T$, and $\mathbf{1}_d$ represents the d time replications of $\mathbf{m}_i^T$. With the mask matrix $\mathbf{M}$, we ignore all the columns in $\left\{ \boldsymbol{X}^F - \sum_{j\neq i} \boldsymbol{d}_j^F \boldsymbol{\gamma}_j^T, \boldsymbol{X}^B - \sum_{j\neq i} \boldsymbol{d}_j^B \boldsymbol{\gamma}_j^T \right\}$ without using the i-th *atom*. The benefits of this update approach are two fold, not only it lowers the computational complexity, but also it avoids redundant and unnecessary updates. The error matrices $\boldsymbol{E}_i^F$, $\boldsymbol{E}_i^B$ are defined as

$$\boldsymbol{E}_i^F \triangleq \left(\boldsymbol{X}^F - \sum_{j\neq i} \boldsymbol{d}_j^F \boldsymbol{\gamma}_j^T \right) \odot \mathbf{M}_i \tag{22}$$

$$\boldsymbol{E}_i^B \triangleq \left(\boldsymbol{X}^B - \sum_{j\neq i} \boldsymbol{d}_j^B \boldsymbol{\gamma}_j^T \right) \odot \mathbf{M}_i . \tag{23}$$

Updating a new column $\boldsymbol{d}_i^F$ and its coefficients $\boldsymbol{\gamma}_i$, we minimize the mean square errors (MSE) by

$$\tilde{\boldsymbol{d}}_i^F = \underset{\boldsymbol{d}_i^F}{\operatorname{argmin}} \left\| \boldsymbol{E}_i^F - \boldsymbol{d}_i^F \boldsymbol{\gamma}_i^F \right\|_F^2 \tag{24}$$

$$\tilde{\boldsymbol{d}}_i^B = \underset{\boldsymbol{d}_i^B}{\operatorname{argmin}} \left\| \boldsymbol{E}_i^B - \boldsymbol{d}_i^B \boldsymbol{\gamma}_i^B \right\|_F^2 . \tag{25}$$

Then, we directly apply the singular value decomposition (SVD) operation on the problems (24) and (25), and get the decompositions

$$\boldsymbol{E}_i^F = \boldsymbol{U}_i^F \boldsymbol{\Delta}_i^F \boldsymbol{V}_i^F,\ \boldsymbol{E}_i^B = \boldsymbol{U}_i^B \boldsymbol{\Delta}_i^B \boldsymbol{V}_i^B. \tag{26}$$

Here, $\boldsymbol{d}_i^F$, $\boldsymbol{d}_i^B$ are respectively updated by $\boldsymbol{U}_i^B$ and $\boldsymbol{U}_i^B$, and $\boldsymbol{\gamma}_i^T$ is alternately valued by $\boldsymbol{\Delta}_i^F(1,1)\cdot \boldsymbol{V}_i^F$ and $\boldsymbol{\Delta}_i^B(1,1)\cdot \boldsymbol{V}_i^B$ on each iteration.

Such approach effectively updates the atoms in $\boldsymbol{D}^F$ and $\boldsymbol{D}^B$, without multiplying all the non-zero representations.

B. Coefficient update

In the sparse coding stage, one might be tempted to suggest using columns of coefficients in dictionary update stage directly, however, that does not imply a guaranteed improvement [24]. We proceed with updating $\boldsymbol{\Gamma}$ in the second stage, and search the sparsest representations for the signals $\boldsymbol{X}^F$ and $\boldsymbol{X}^B$ in (17).

The natural approach alternately optimizes the following sparse approximations

$$\tilde{\boldsymbol{\gamma}}_i^T = \underset{\boldsymbol{\gamma}_i^T}{\operatorname{argmin}} \left\| \boldsymbol{X}^F - \sum_{i=1}^{N} \boldsymbol{d}_i^F \boldsymbol{\gamma}_i^T \right\|_2^2 \tag{27}$$

$$\tilde{\boldsymbol{\gamma}}_i^T = \underset{\boldsymbol{\gamma}_i^T}{\operatorname{argmin}} \left\| \boldsymbol{X}^B - \sum_{i=1}^{N} \boldsymbol{d}_i^B \boldsymbol{\gamma}_i^T \right\|_2^2. \tag{28}$$

Considering computational complexity, the proposed approach ignores the formulations in (27) and (28), and try to update the coefficients using the error matrices $\boldsymbol{E}_i^F$ and $\boldsymbol{E}_i^F$. Once the dictionary update is performed, we can obtain the expressions

$$\sum_{i=1}^{N} \left\| \boldsymbol{E}_i^F - \boldsymbol{d}_i^F \boldsymbol{\gamma}_i^F \right\|_2^2 \tag{29}$$

$$\sum_{i=1}^{N} \left\| \boldsymbol{E}_i^B - \boldsymbol{d}_i^B \boldsymbol{\gamma}_i^B \right\|_2^2. \tag{30}$$

The algorithm can find the suitable sparse representations during the dictionary updating stage. Then, we combine the functions (29) and (30), forcing $\boldsymbol{d}_i^F$ and $\boldsymbol{d}_i^B$ to share the same sparse representation $\boldsymbol{\gamma}$, given by

$$\begin{aligned} &\min_{\boldsymbol{\gamma}} \frac{1}{2} \left\| \boldsymbol{E}_i^F - \boldsymbol{d}_i^F \boldsymbol{\gamma} \right\|_2^2 + \frac{1}{2} \left\| \boldsymbol{E}_i^B - \boldsymbol{d}_i^B \boldsymbol{\gamma} \right\|_2^2 + \lambda \left\| \boldsymbol{\gamma} \right\|_1 \\ &\text{s.t. } \left\| \boldsymbol{d}_i^F \right\|_2^2 \leqslant 1,\ \left\| \boldsymbol{d}_i^B \right\|_2^2 \leqslant 1,\ \forall 1 \leqslant i \leqslant N. \end{aligned} \tag{31}$$

Note that (31) can be equivalently rewritten as the optimization function for $\boldsymbol{\gamma}$

$$\min_{\boldsymbol{\gamma}} \mathcal{F}(\boldsymbol{\gamma}) = \frac{1}{2} \underbrace{\left\| \tilde{\boldsymbol{E}}_i - \tilde{\boldsymbol{d}}_i \boldsymbol{\gamma} \right\|_2^2}_{\mathcal{F}_1(\boldsymbol{\gamma})} + \lambda \underbrace{\|\boldsymbol{\gamma}\|_1}_{\mathcal{F}_2(\boldsymbol{\gamma})} \tag{32}$$

where

$$\tilde{\boldsymbol{E}}_i = \begin{bmatrix} \boldsymbol{E}_i^F \\ \boldsymbol{E}_i^B \end{bmatrix}, \quad \tilde{\boldsymbol{d}}_i = \begin{bmatrix} \boldsymbol{d}_i^F \\ \boldsymbol{d}_i^B \end{bmatrix}. \tag{33}$$

Here, we divide the optimization function $\mathcal{F}(\boldsymbol{\gamma})$ into two parts $\mathcal{F}_1(\boldsymbol{\gamma})$ and $\mathcal{F}_2(\boldsymbol{\gamma})$ and assume $(\boldsymbol{\gamma}^t, \tilde{\boldsymbol{d}}_j)$ be the point obtained by the function $\mathcal{F}(\boldsymbol{\gamma})$ at iteration t point. Modified by [41], [42], our optimization approach is based on the repeated minimization of a differentiable surrogate function.

Define the function $\mathcal{G}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^t)$ as a surrogate function for the function $\mathcal{F}(\gamma)$ on the point $\boldsymbol{\gamma}^t$ with the following two conditions hold.

$$\begin{aligned} \mathcal{G}(\boldsymbol{\gamma}^t, \boldsymbol{\gamma}^t) &= \mathcal{F}(\gamma^t) \\ \mathcal{G}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^t) &\geq \mathcal{F}(\gamma) \end{aligned} \tag{34}$$

Then, the above optimization function can be solved by the following approximation

$$\boldsymbol{\gamma}^{t+1} = \operatorname*{argmin}_{\boldsymbol{\gamma}} \mathcal{G}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^t) \tag{35}$$

where

$$\mathcal{G}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^t) \triangleq \langle \nabla_{\boldsymbol{\gamma}} \mathcal{F}_1(\boldsymbol{\gamma}^t), \boldsymbol{\gamma} \rangle + \frac{1}{2} \left\| \boldsymbol{\gamma} - \boldsymbol{\gamma}^t \right\|_F^2 + \mathcal{F}_2(\boldsymbol{\gamma}). \tag{36}$$

More specifically, we attempt to get the solution on the variable $\boldsymbol{\gamma}$ by this rule. We describe the well-known soft-thresholding operation $\mathcal{S}_\lambda(a)$ as

$$\mathcal{S}_\lambda(a) \triangleq \begin{cases} a - \lambda, & \text{for } a > \lambda \\ 0, & \text{for } |a| \leq \lambda \\ a + \lambda, & \text{for } a < \lambda \end{cases}. \tag{37}$$

Combining (36) with (37), the updating formula for $\boldsymbol{\gamma}$ is obtained by the closed-form solution

$$\boldsymbol{\gamma}^{t+1} = \boldsymbol{\gamma}^t - \mathcal{S}_{\frac{\lambda}{\mathcal{T}^2(\tilde{\boldsymbol{d}}_j)}} \left(\boldsymbol{\gamma}^t - \frac{1}{\mathcal{T}^2(\tilde{\boldsymbol{d}}_j)} \tilde{\boldsymbol{d}}_j^T (\tilde{\boldsymbol{d}}_j \boldsymbol{\gamma}^t - \tilde{\boldsymbol{E}}_j) \right) \tag{38}$$

where $\mathcal{T}(\tilde{\boldsymbol{d}}_j)$ describe the maximum singular value of $\tilde{\boldsymbol{d}}_j$. Appendix B shows the convergence analysis.

The complete implement is given as Algorithm 2.

Algorithm 2 Learning the coupled dictionary.

Input: two signal sets $\boldsymbol{X}^F$ and $\boldsymbol{X}^B$, initial dictionaries $\boldsymbol{D}_0^F$ and $\boldsymbol{D}_0^B$, initial sparse coefficients Γ.

1: **Initialization**: Set $\boldsymbol{D}^F := \boldsymbol{D}_0^F$, $\boldsymbol{D}^B := \boldsymbol{D}_0^B, \boldsymbol{\Gamma} := \boldsymbol{\Gamma}_0$
2: **for** $j = 1 \cdots n$ **do**
3: compute the matrices $\boldsymbol{E}_i^F$ and $\boldsymbol{E}_i^B$, using (22), (23);
4: update the coupled atoms $\boldsymbol{d}_i^F$and $\boldsymbol{d}_i^B$ using (24), (25);
5: **for** $i = 1 \cdots t$ **do**
6: update sparse coefficients $\boldsymbol{\gamma}$ using (37), (38);
7: **end for**
8: **end for**

Output: the couple of dictionaries $\boldsymbol{D}^F$ and $\boldsymbol{D}^B$.

V. SIMULATION AND RESULTS

This section demonstrates experiential results of our approach on simulated data. We firstly evaluate the proposed approach by visual comparisons, and then discuss the quantitative assessments of the proposed implementation. At last, we describe various the influence of factors to the proposed approach including patch size, tolerance error, and number of overlapping pixels. The objective evaluation is based on two state-of-the-art fusion performance metrics: (i) Q_{MI} [44], which measures how well the mutual information from the source images is preserved in the fused image; (ii) $Q^{AB/F}$ [45], which evaluates how well the success of edge information transfers from the source images to the fused image. All the experiments are performed on a PC running a Inter(R) Xeon(R) 3.40GHz CPU.

A. Experimental Setup

The proposed method is compared to the following state-of-art multi-focus image fusion algorithms:

- LP: aplacian pyramid-based image fusion approach [11];
- MWG: multi-scale weighted gradient-based spatial image fusion approach [13];
- DWT: discrete wavelet transform-based image fusion approach [14];
- NSCT: non-subsampled contourlet transform-based image fusion approach [15];

- PCA: robust principal component analysis-based image fusion approach [1];
- SR-CM: sparse representation "choose-max"-based image fusion approach [20];
- SR-KSVD: sparse representation K-SVD-based image fusion approach.

The parameter settings of these methods are as follows. For the LP and DWT methods, the source images are decomposed to 3 levels. The wavelet basis "db1" is applied to the DWT algorithm. For the NSCT approach, the direction numbers of the 4 decomposition levels from coarse to fine are selected as 4, 8, 8, and 16. For fair comparison, all algorithms based on sparsity model are implemented using the same dictionary size.

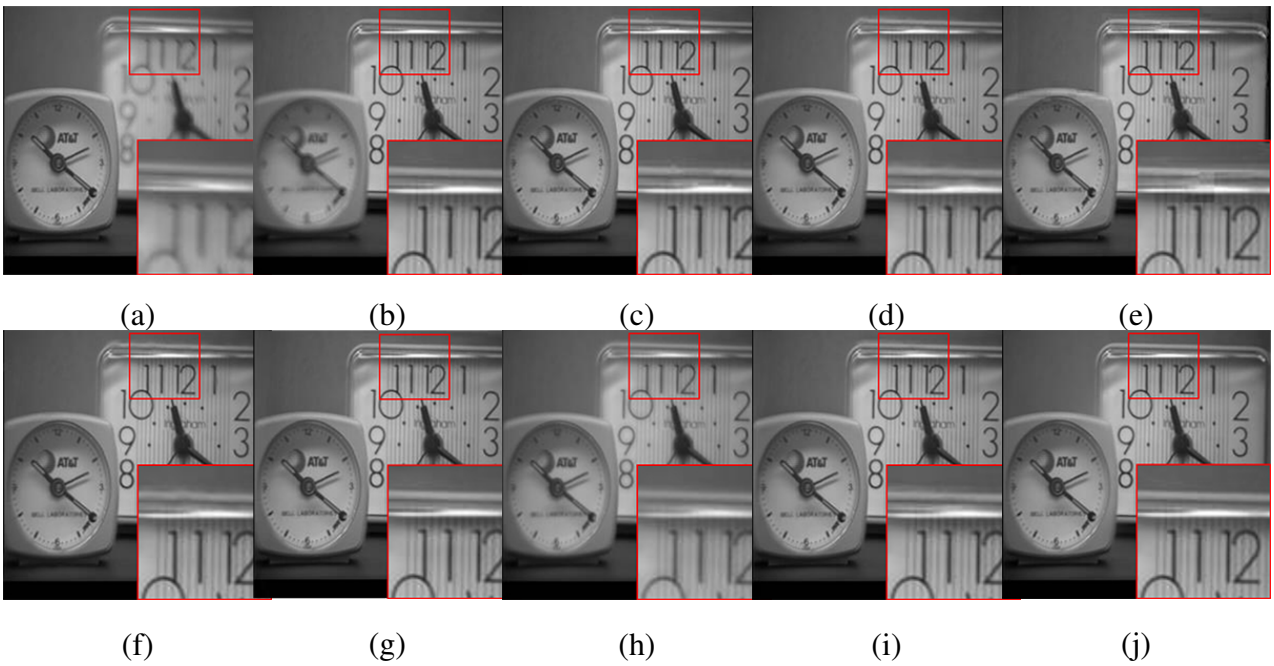
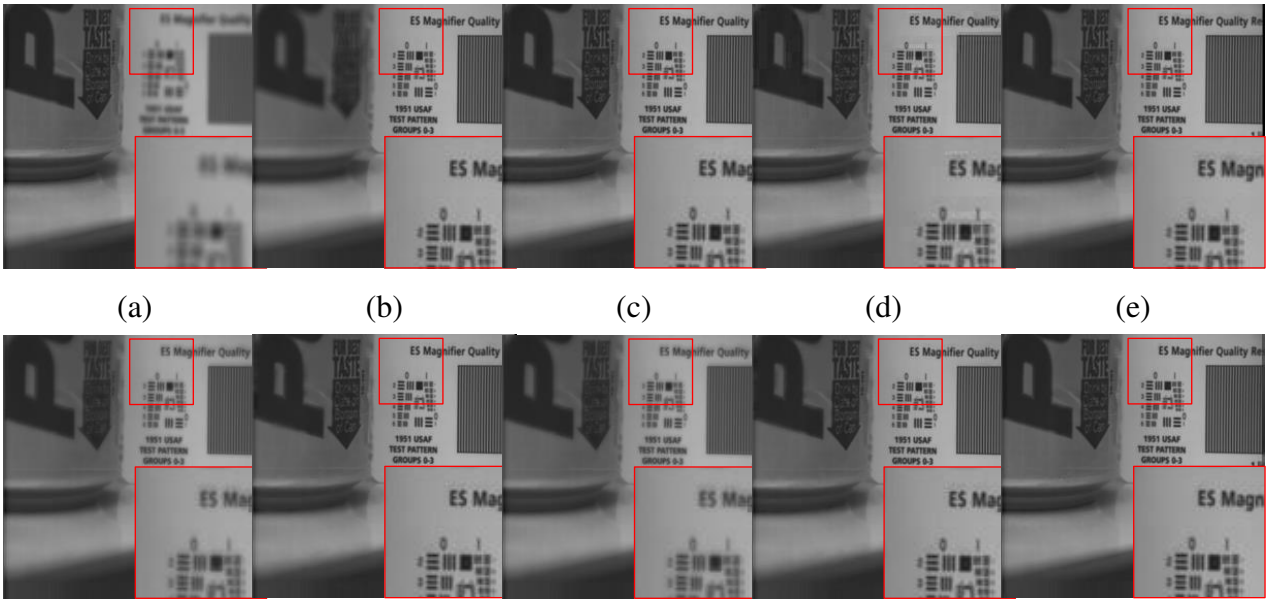


Fig. 1. *Source images "Clock" and the fusion result comparisons. (a) The first source image with focus on the left. (b) The second source image with focus on the right. Fused images obtained by LP (c), MWG (d), DWT (e), NSCT (f), PCA (g), SR-CM (h), SR-KSVD (i) and the proposed method (j).*

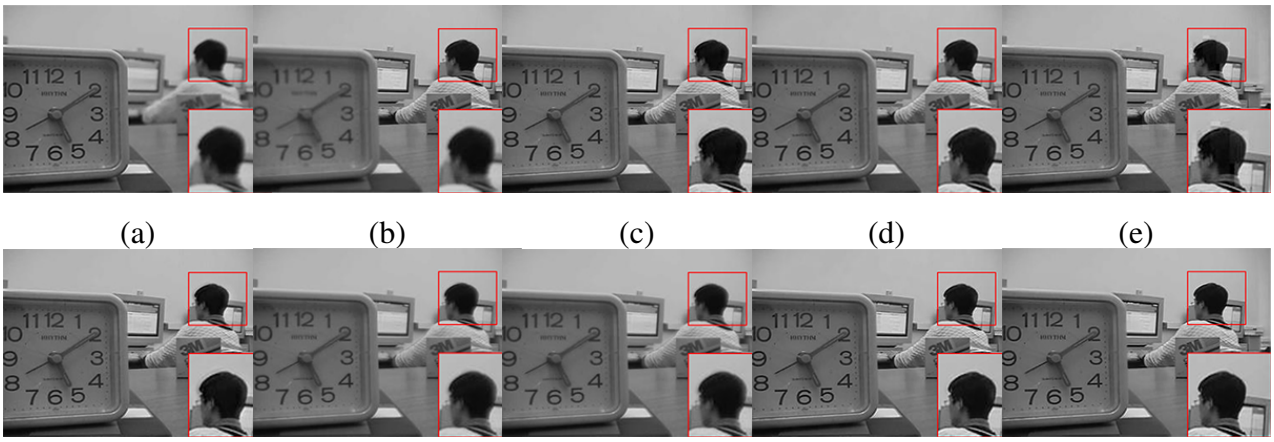
Throughout all experience, we use the following parameter setting for our method. we set the tolerance error $\epsilon = 0.1$, the pitch size $m = 8$, and the overlapping size is 1. During the coupled dictionary learning, the training data consist of 50,000 8×8 focused patches which are randomly sampled from the database of 40 natural images and 50,000 8×8 blurred patches created by focused patches using Gaussian blur function. The two dictionaries are produced by improved K-SVD algorithm, and initialized with samples from the training data. Larger dictionaries result in heavier computation and better quality. Considering the tradeoff between fusion quality and computation, we fix the dictionary size as 64×256, execute 6 multiple dictionary update cycles



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig. 2. *Source images "Pepsi" and the fusion result comparisons. Same order in the Fig. 1.*



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig. 3. *Source images "Lab" and the fusion result comparisons. Same order in the Fig. 1.*

(DUC) and 30 iterations. Experiments converge very quickly, and approximately achieve a $\times 2$ speedup compared to the standard coupled dictionary learning [27].

B. Simulated Data Experiments

To show the effectiveness of the proposed algorithm, we compare it with existing algorithms, including LP, MWG, DWT, NSCT, PCA, SR-CM, and SR-KSVD methods. The fusion results of the images "Clock", "Pepsi", and "Lab" are shown in Fig. 1, Fig. 2, and Fig. 3, including the

TABLE I

OBJECTIVE EVALUATION OF THE IMAGE FUSION

Methods	Measures in Fig. 1			Measures in Fig. 2			Measures in Fig. 3		
	Q_{MI}	$Q^{AB/F}$	$Time(s)$	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**	0.9091	0.6901	**0.0039**
MWG	0.9045	0.7243	0.8383	0.8992	0.7202	0.8424	0.9083	0.7264	0.8280
DWT	0.8991	0.7013	0.1542	0.8876	0.6991	0.1784	0.9074	0.7083	0.1501
NSCT	0.9245	0.7185	1.0748	0.9188	0.7101	1.1092	0.9246	0.7189	1.0054
PCA	0.9381	0.6620	0.0074	0.9311	0.6601	0.0102	0.9383	0.6609	0.0070
SR-CM	0.9276	0.7214	3.1894	0.9241	0.7201	3.2785	0.9289	0.7215	3.1002
SR-KSVD	0.9386	0.7326	3.1992	0.9381	0.7314	3.2198	0.9391	0.7326	3.0784
Ours	**0.9432**	**0.7451**	3.2123	**0.9428**	**0.7439**	3.4009	**0.9445**	**0.7451**	3.1459

magnified details in the lower right corners. Figs. 1(a) and (b) show the source images, and the fused images obtained by different fusion methods based on the LP, MWG, DWT, NSCT, PCA, SR-CM, SR-KSVD methods, and the proposed method are presented in Figs. 1(c)-(j). The same order is in the Fig. 2 and Fig. 3.

By carefully observing the fusion results, we see clearly that the LP method does not produce continuous edges (see Fig. 1(c)) and the DWT method results in blocking artifacts (see Fig. 1(e)). Moreover, the same results can be easily observed in the magnified images. The fusion methods based on MWG (see Fig. 1(d)) and NSCT (see Fig. 1(f)) show circle blurring effect around strong boundaries. Fig. 1(f) is visually better than Fig. 1(d). In Figs. 1(g) and (i) can be seen some artificial distortions, and Fig. 1(h) obviously lacks edge information. Our approach provides the best visual appearance (see Fig. 1(j)). Similar subjective results are obtained in these experiments as shown in Fig. 2 and Fig. 3. Here again the proposed approach yields better image quality than that of the other compared methods.

Table I summarizes the evaluations on grayscale multi-focus images. The values of Q_{MI} and $Q^{AB/F}$ range from 0 to 1, with 1 representing the ideal fusion. The bold values are the best results in the corresponding columns. As can be seen from Table I, 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 the values for the LP, MWG, DWT, NSCT, and SR-CM methods, and it is similar to the methods based on PCA and SR-KSVD. That means that PCA, SR-KSVD and the proposed approach well preserve the mutual information from different source images. The values of $Q^{AB/F}$ demonstrate that our approach reduces the blocking artifacts and artificial distortions, and combines the significant edge information into the fused image. For the running time, we optimize the sparse coding using CoefROMP [40] which is initialized with the largest $k/3$ coefficients from the previous pursuit stage unlike previous pursuit algorithm [39], so the proposed approach takes less time than the SR-KSVD approach.

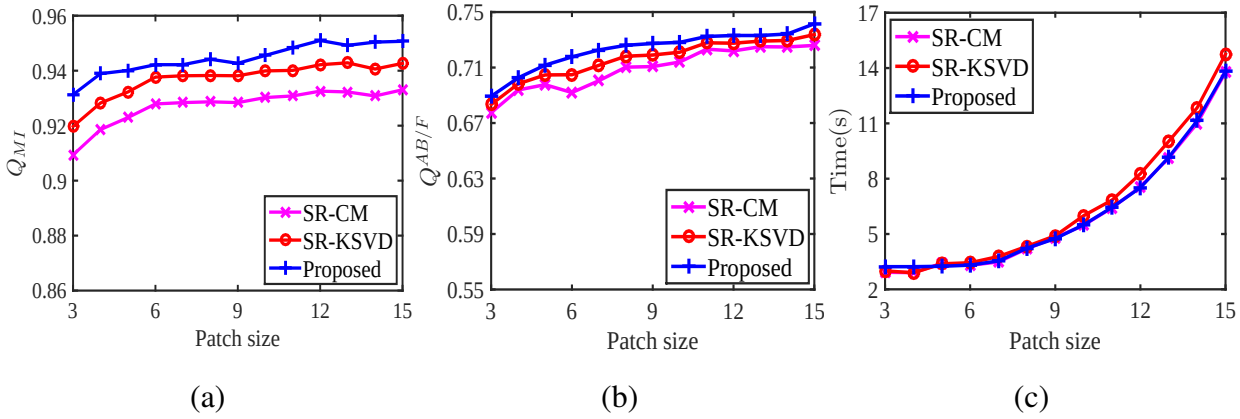


Fig. 4. *Fusion performances with different patch size.*

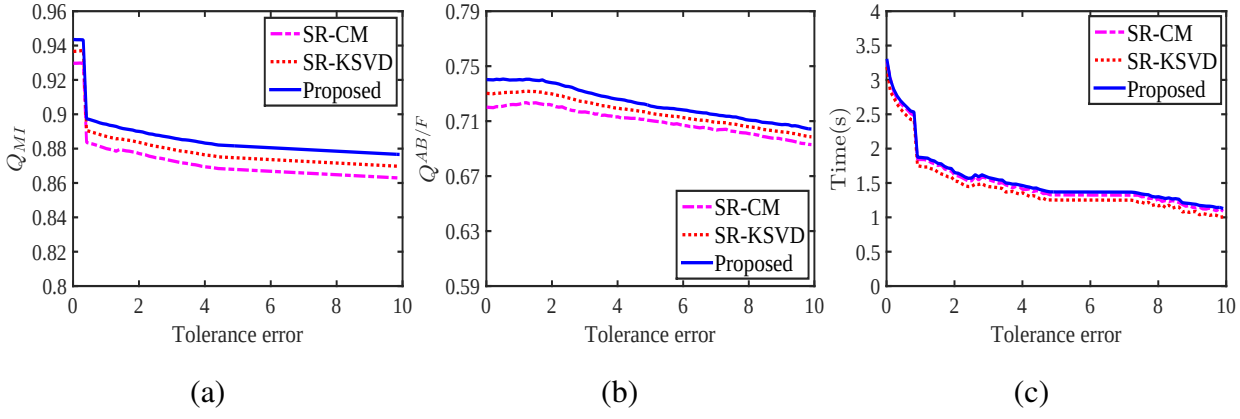
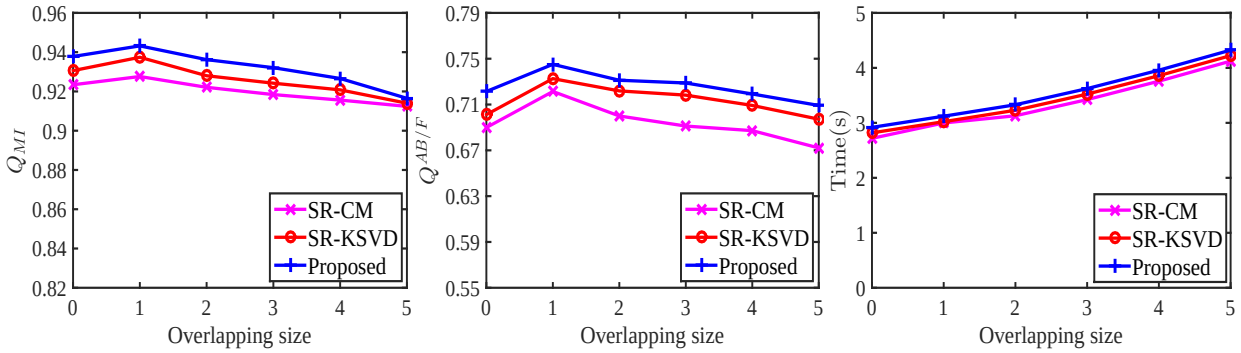


Fig. 5. *Fusion performances with different tolerance error.*



(a) (b) (c)

Fig. 6. *Fusion performances with different overlapping size.*

C. Effects of Main Parameters

In our experiments, we discuss three main parameters for the proposed approach to evaluate the fusion performance: patch size, tolerance error and number of overlapping pixels. We evaluate the proposed approach on the grayscale multifocus dataset which contain 10 pairs of grayscale images and 30 pairs of artifical source images with 256 level grey scales. Fig. 4, Fig. 5 and Fig. 6 respectively show the average values of Q_{MI}, $Q^{AB/F}$ and the running time with the the parameters pitch size, tolerance error and overlapping size.

As can be seen in Figs. 4(a) and (b), both Q_{MI} and $Q^{AB/F}$ slightly benefit from the increasing patch size. When the patch size increases to 9, the running time begins to increase sharply (see Fig. 4(a)). Balancing the computation time and fusion quality, we set the patch size to 8×8 in our experiments. Fig. 5 also illustrates the significant improvements in our approach compared to SR-CM and SR-KSVD. The changes of Q_{MI}, $Q^{AB/F}$, and the running time are shown versus the tolerance error ϵ. Fig. 5(a) and (b) show that the tolerance error slightly impacts on Q_{MI} and $Q^{AB/F}$. When ϵ is less than 2, the tested approaches have an acceptable fusion quality (see Fig. 5(a)). When ϵ is larger than 2, Q_{MI} is drastically decreasing. Additional experiments in Fig. 6 in show that for increased number of overlapping pixels, Q_{MI} drastically decreases and the running time sharply increases. Hence, our method has fixed overlapping size to 1.

VI. Conclusion

We have presented a fusion algorithm for combing multiple images with different focal settings into an all-in-focus image. The algorithm first formulated the physical process on multi-focus image fusion, and then described the basic model based on coupled sparse representations. we

introduced the K-SVD-based coupled dictionary learning algorithm that enforced the sparse approximations for double feature spaces. Using the couple of dictionaries from the focused and blurred feature spaces, we explored the an effective and accurate fusion criteria via plain averaging. We demonstrated using simulated data that the proposed approach well preserved the edge and structural information of source images, drastically reduced the blocking artifacts, circle blurring and artifical distortions.

There are two possible extensions to the algorithm. First, while we achieved the fusion criteria in the noiseless setting, it may be possible to fuse multi-focus images in the presence of noise. This may allow us to derive theoretical analysis for fusion criteria. Second, as it was demonstrated that even a RandOMP-based averaging weighting scheme improved the performance of the algorithm, designing more advanced weighting schemes may also result in further improvement.

APPENDIX A
FUSING SPARSE COEFFICIENTS

In this appendix, we describe in details the fusion on sparse representations in Section III.

Fusing sparse coefficients aims to approximate the compact spares representations $\boldsymbol{A}$ for unknown image $\boldsymbol{I}_F$. Given a pair of dictionaries $\left\{\boldsymbol{D}^F, \boldsymbol{D}^B\right\}$, the sparsity-based sparse coding problems are described as

$$\begin{aligned}\left[\boldsymbol{a}_k^F\right]_{ij} =& \arg\min_{\left[\boldsymbol{a}_k^F\right]_{ij}} \left\|\left[\boldsymbol{a}_k^F\right]_{ij}\right\|_1 \\ & \text{s.t.} \left\|\boldsymbol{d}^F\left[\boldsymbol{a}_k^F\right]_{ij} - \mathbf{W}_{i,j}\boldsymbol{I}_k\right\|_2^2 \leq \epsilon\end{aligned} \tag{39}$$

$$\begin{aligned}\left[\boldsymbol{a}_k^B\right]_{ij} =& \arg\min_{\left[\boldsymbol{a}_k^B\right]_{ij}} \left\|\left[\boldsymbol{a}_k^B\right]_{ij}\right\|_1 \\ & \text{s.t.} \left\|\boldsymbol{d}^B\left[\boldsymbol{a}_k^B\right]_{ij} - \mathbf{W}_{i,j}\boldsymbol{I}_k\right\|_2^2 \leq \epsilon\end{aligned} \tag{40}$$

Assume that the columns of $\boldsymbol{D}^F$ and $\boldsymbol{D}^B$ are normalized. The RandOMP [38] algorithm selects at each step the atoms of $\left[\boldsymbol{a}_k^F\right]_{ij}$ and $\left[\boldsymbol{a}_k^B\right]_{ij}$ with probability proportional to (9). When each atom is selected, the signal $\mathbf{W}_{i,j}\boldsymbol{I}_k$ is orthogonally projected to the span of the elected atoms, and $\left[\boldsymbol{a}_k^F\right]_{ij}$ and $\left[\boldsymbol{a}_k^B\right]_{ij}$ at the j-th iteration are computed by

$$\begin{aligned}\left[\boldsymbol{a}_k^F\right]_{ij} &= (\boldsymbol{d}^F)^{+}\mathbf{W}_{i,j}\boldsymbol{I}_k \\ &= ((\boldsymbol{d}^F)^T\boldsymbol{d}^F)^{-1}(\boldsymbol{d}^F)^T\mathbf{W}_{i,j}\boldsymbol{I}_k\end{aligned} \tag{41}$$

$$\begin{aligned}\left[\boldsymbol{a}_k^B\right]_{ij} &= (\boldsymbol{d}^B)^{+}\mathbf{W}_{i,j}\boldsymbol{I}_k \\ &= ((\boldsymbol{d}^B)^T\boldsymbol{d}^B)^{-1}(\boldsymbol{d}^F)^T\mathbf{W}_{i,j}\boldsymbol{I}_k.\end{aligned} \tag{42}$$

When the residuals $\boldsymbol{r}_j^F$ and $\boldsymbol{r}_j^B$ are below ϵ, the stopping rules are satisfied. Here, we present in Algorithm 3 the sparse coefficients fusion method.

Algorithm 3 Fusing sparse coefficients.

Input: the couple of learned dictionaries $\left\{\boldsymbol{D}^F, \boldsymbol{D}^B\right\}$; multi-focus source images $\{\boldsymbol{I}_k\}_{k=1}^n$; the tolerance error ϵ.

1: **Initialization**: Set

· $\left[\boldsymbol{a}_k^F\right]_{ij}^0 := 0$, $\left[\boldsymbol{a}_k^B\right]_{ij}^0 := 0$;

· $\boldsymbol{r}_j^F := \mathbf{W}_{i,j}\boldsymbol{I}_k$, $\boldsymbol{r}_j^B := \mathbf{W}_{i,j}\boldsymbol{I}_k$;

· Support $S_F^0 = Support\left\{\left[\boldsymbol{a}_k^F\right]_{ij}^0\right\} = \varnothing$,

Support $S_B^0 = Support\left\{\left[\boldsymbol{a}_k^B\right]_{ij}^0\right\} = \varnothing$;

2: **for** j=1 : max-iter

· Seek the random j_F^r-th atom and random j_B^r-th atom with probability proportional to (9);

· Merge the supports $S_F^j = S_F^{j-1}\bigcup\{j_F^r\}$, and $S_F^j = S_B^{j-1}\bigcup\{j_B^r\}$;

· Compute $\left[\boldsymbol{a}_k^F\right]_{ij}$ and $\left[\boldsymbol{a}_k^B\right]_{ij}$, using (41) and (42);

· Compute $\boldsymbol{r}_j^F = \mathbf{W}_{i,j}\boldsymbol{I}_k - \boldsymbol{d}^F\left[\boldsymbol{a}_k^F\right]_{ij}$, and

$\boldsymbol{r}_j^B = \mathbf{W}_{i,j}\boldsymbol{I}_k - \boldsymbol{d}^F\left[\boldsymbol{a}_k^B\right]_{ij}$;

· When $\boldsymbol{r}_j^F < \epsilon$ and $\boldsymbol{r}_j^B < \epsilon$, quit;

3: **end for**

4: Compute the averaging coefficients $[\boldsymbol{a}_k]_{ij}$, using (10).

5: Choose the compact sparse coefficients $[\boldsymbol{a}]_{ij}$, using (11).

Output: the fusing coefficients $[\boldsymbol{a}]_{ij}$.

APPENDIX B
CONVERGENCE ANALYSIS ON LEARNING THE COUPLED DICTIONARY

In this section, we discuss the convergence analysis on learning the coupled dictionary. The material presented here is the review of the ideas [41], [42]. The main optimization problem in (17) is not guaranteed to seek the global optimum, and we intend to discuss the local minimum. Each dictionary updating step for $\boldsymbol{d}^F$, $\boldsymbol{d}^B$ with a fixed sparse representation $\boldsymbol{\gamma}$, we optimize the

error matrices $\boldsymbol{E}_j^F$ and $\boldsymbol{E}_j^B$ in (22) and (23). Empirically, the two dictionaries converge the local optimal with executing limited updating operations. In this case, the success of local convergence mainly depends on the sparse coding stage. Thus, we emphasize on discussing the minimum of joint sparse coding.

In the following convergence analysis, we assume a stationary point $(\boldsymbol{\gamma}^*, \boldsymbol{d})$ generated by limited iterates during sparse coding stage. Following the definitions of $\mathcal{F}(\cdot)$, the following series of inequalities are described by

$$\mathcal{F}(\boldsymbol{\gamma}^1) \geqslant \mathcal{F}(\boldsymbol{\gamma}^2) \geqslant \cdots \geqslant \mathcal{F}(\boldsymbol{\gamma}^t) \geqslant \cdots \tag{43}$$

Here exists a subsequence $(\boldsymbol{\gamma}^t, \boldsymbol{d})$ converging to the limit point $(\boldsymbol{\gamma}^*, \boldsymbol{d})$. $\mathcal{G}(\cdot,\cdot)$ represents the upper bound of $\mathcal{F}(\cdot)$, then we obtain that

$$\mathcal{G}(\boldsymbol{\gamma}^t,\boldsymbol{\gamma}^t) = \mathcal{F}(\boldsymbol{\gamma}^t) \leqslant \mathcal{F}(\boldsymbol{\gamma}^{t+1}) \leqslant \mathcal{G}(\boldsymbol{\gamma}^{t+1},\boldsymbol{\gamma}^t) \leqslant \mathcal{G}(\boldsymbol{\gamma},\boldsymbol{\gamma}^t). \tag{44}$$

Note that when $t \to \infty$, the inequalities (44) imply

$$\mathcal{G}(\boldsymbol{\gamma}^*,\boldsymbol{\gamma}^*) \leqslant \mathcal{G}(\boldsymbol{\gamma},\boldsymbol{\gamma}^*). \tag{45}$$

Thus, we conclude $\mathcal{G}'(\boldsymbol{\gamma},\boldsymbol{\gamma}^*) \geqslant 0$, that means the update on $\boldsymbol{\gamma}$ exists the local optimum $(\boldsymbol{\gamma}^*, \boldsymbol{d})$.

ACKNOWLEDGMENT

The authors would like to thank...

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," in *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," in *Opt. Commun.*, vol. 284, no. 1, pp. 80–87, Jan. 2011.

[6] Q. Zhang, and B. L. Guo, "Multifocus image fusion using the nonsubsampled contourlet transform, "*Signal Process.*, vol. 89, pp. 1334–1346, Jul. 2009.

[7] 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.

[8] G. Pajares and J. Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.*, vol. 37, no. 9, pp. 1855–1872, Sep. 2004.

[9] 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.

[10] 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.

[11] 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.

[12] S. Li, B. Yang, "Multifocus image fusion using region segmentation and spatial frequency," *Inf. Fusion*, vol. 26, no. 7, pp. 971–979, Jul. 2008.

[13] 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.

[14] 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.

[15] Q. Zhang and B. Guo, "Multifocus image fusion using the nonsubsampled contourlet transform," *Signal Process.*, vol. 89, no. 7, pp. 1334–1346, Jul. 2009.

[16] T. Wan, N. Canagarajah, and A. Achim, "Compressive image fusion," in *Proc. IEEE 15th Int. Conf. Image Process.*, San Diego, CA, 2008, pp. 1308–1311.

[17] 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.

[18] T. Wan, Z. Qin, C. Zhu, and R. Liao, "A robust fusion scheme for multifocus images using sparse features," in *Proc. IEEE 38th Int. Conf. Acoustics, Speech and Signal Process.*, British Columbia, Canada, 2013, pp. 1957–1961.

[19] 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.

[20] 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.

[21] 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.

[22] Q. Zhang, and M. D. Levine, "Robust Multi-Focus Image Fusion Using Multi-Task Sparse Representation and Spatial Context," in *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2045–2058, Mar. 2016.

[23] 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," in *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 220–224, Sep. 2015

[24] 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.

[25] 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.

[26] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th ACM Int. Conf. Mach. Learn.*, Montreal, QC, Canada, 2009, pp. 689–696.

[27] 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.

[28] 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.

[29] J. Sadasivan, S. Mukherjee, and C. S. Seelamantula, "Joint dictionary training for bandwidth extension of speech signals," in *Proc. IEEE 41st Int. Conf. Acoustics, Speech and Signal Process.*, Shanghai, China, 2016, pp. 5925–5929.

[30] 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.

[31] 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.

[32] 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.

[33] 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 40st Int. Conf. Acoustics, Speech and Signal Process.*, South Brisbane, QLD, 2015, pp. 151–155.

[34] R. Gao, S. A. Vorobyov, and H. Zhao, "Multi-focus image fusion via coupled dictionary training," in *Proc. IEEE 41st Int. Conf. Acoustics, Speech and Signal Process.*, Shanghai, China, 2016, pp. 1666–1670.

[35] T. Khler, 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.

[36] M. Subbarao, T. Choi, and A. Nikzad, "Focusing techniques," *Opt. Eng.*, vol. 32, pp. 2824–2836, Mar. 1993.

[37] M. Born, and E. Wolf, "Principles of Optics," in *Cambridge Univ. Press.*, 1999.

[38] 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.

[39] 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.

[40] 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.

[41] M. Razaviyayn, H. W. Tseng and Z. Q. Luo, "Computational Intractability of Dictionary Learning for Sparse Representation," *arXiv preprint arXiv:1511.01776*, 2015.

[42] 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.

[43] 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.

[44] 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.

[45] C. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, Feb. 2000.