Image Fusion With Convolutional Sparse Representation

Yu Liu, Xun Chen, *Member, IEEE*, Rabab K. Ward, *Fellow, IEEE*, and Z. Jane Wang, *Senior Member, IEEE*

***Abstract*—As a popular signal modeling technique, sparse representation (SR) has achieved great success in image fusion over the last few years with a number of effective algorithms being proposed. However, due to the patch-based manner applied in sparse coding, most existing SR-based fusion methods suffer from two drawbacks, namely, limited ability in detail preservation and high sensitivity to misregistration, while these two issues are of great concern in image fusion. In this letter, we introduce a recently emerged signal decomposition model known as convolutional sparse representation (CSR) into image fusion to address this problem, which is motivated by the observation that the CSR model can effectively overcome the above two drawbacks. We propose a CSR-based image fusion framework, in which each source image is decomposed into a base layer and a detail layer, for multifocus image fusion and multimodal image fusion. Experimental results demonstrate that the proposed fusion methods clearly outperform the SR-based methods in terms of both objective assessment and visual quality.**

***Index Terms*—Convolutional sparse representation (CSR), detail preservation, image fusion, shift invariance, sparse representation (SR).**

I. Introduction

THE main purpose of image fusion technique is to integrate complementary information from multiple source images of the same scene by generating a fused image [1]. The source images may originate from different types of imaging sensors or from a certain kind of sensor but with different imaging parameters. The fused image is supposed to be more comprehensive and suitable for machine or human perception than each source image individually. During recent years, image fusion has become an active topic in the image processing community, primarily owing to the increasing demand of various image-based applications in digital photography, video surveillance, medical imaging, and remote sensing.

A variety of image fusion algorithms have been developed in recent years. In general, these methods could be classified into two main categories: spatial domain methods and transform domain methods [2]. Spatial domain methods usually address the fusion issue via image blocks [3]–[6] or pixel-wise gradient information [7]–[10]. This category of methods is more applicable to fusion tasks in which the source images are obtained from the same type of sensors, such as multifocus fusion [3], [5]–[7], [10] and multiexposure fusion [4], [8], [9]. Transform domain methods merge the information contained in the source images after transforming them into another domain. The fused image is obtained by performing relevant reconstruction methods on the merged coefficients. Popular transform domains investigated in image fusion include Laplacian pyramid [11], discrete wavelet transform [12], discrete cosine transform [13], nonsubsampled contourlet transform (NSCT) [14], independent component analysis [15], sparse representation (SR) [16], etc. Since these image representation approaches are in accord with the physiological mechanism of the human visual system (HVS), transform domain methods are recognized to be very effective in multimodal image fusion [1], [17], where the source images are captured by different types of imaging sensors. Another advantage of most transform domain methods is that they can handle the fusion of noisy images by integrating the denoising process into the fusion framework [16], [18]. In this letter, we mainly concentrate on transform domain methods.

SR-based image fusion has emerged as an attractive direction in transform domain methods over the last few years. Many effective SR-based fusion methods have been proposed [16], [18]–[21]. It is generally observed that the SR-based methods tend to provide better performances than conventional multiscale transform based methods. However, it will be discussed in Section II that the SR-based fusion methods suffer from two main drawbacks, namely, *limited ability in detail preservation* and *high sensitivity to misregistration*, while these two issues are of great concern in image fusion.

In this letter, we apply a newly emerged signal decomposition model, known as *convolutional sparse representation (CSR)* [22], to the image fusion problem. We show that CSR can successfully overcome the above two drawbacks. The contributions of this work are two-fold

1) We introduce CSR into the field of image fusion to address the drawbacks of SR-based fusion methods. To the best of our knowledge, this is the first time that CSR is applied to image fusion.
2) A CSR-based image fusion framework is proposed for multifocus image fusion and multimodal image fusion. Experimental results demonstrate that the proposed methods clearly outperform the SR-based methods in terms of both objective assessment and visual quality.

The rest of this letter is organized as follows. In Section II, related work and the motivation of this work are presented.

Manuscript received August 23, 2016; revised October 10, 2016; accepted October 13, 2016. Date of publication October 19, 2016; date of current version November 30, 2016. This work was supported by the Qatar National Research Fund, a member of Qatar Foundation, under National Priorities Research Program Grant 7-684-1-127. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Marco Felipe Duarte. *(Corresponding author: X. Chen.)*

Y. Liu and X. Chen are with the Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China (e-mail: yuliu@hfut.edu.cn; xun.chen@hfut.edu.cn).

R. K. Ward and Z. J. Wang are with Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: rababw@ece.ubc.ca; zjanew@ece.ubc.ca).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org.

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2016.2618776

Section III describes the proposed fusion methods in detail. The experimental results and discussions are provided in Section IV. Finally, Section V concludes the letter.

II. Related Work and Motivation

A. SR-Based Image Fusion

SR addresses the sparsity of natural signals, which is consistent with the mechanism of HVS [23]. Given a signal $\mathbf{s} \in \mathrm{R}^n$ and an over-complete dictionary $\mathbf{D} \in \mathrm{R}^{n \times m} (n < m)$, the goal of SR is to estimate a sparse vector $\mathbf{x} \in \mathrm{R}^m$ with only a few nonzero entries such that $\mathbf{s} \approx \mathbf{D}\mathbf{x}$. This problem can be formulated as [24]

$$\arg\min_{\mathbf{x}} ||\mathbf{x}||_0 \quad s.t. \quad ||\mathbf{D}\mathbf{x} - \mathbf{s}||_2 < \varepsilon \tag{1}$$

where $\|\cdot\|_0$ indicates the l_0-"norm" that counts the number of nonzero entries and ε is the error tolerance.

SR has been widely used in various image processing and computer vision applications [25], [26]. In image fusion research, SR was first introduced by Yang and Li [16]. In their method, the sliding-window technique is applied to divide the source images into a large number of overlapping patches, and sparse decomposition is performed independently on each image patch using orthogonal matching pursuit (OMP) algorithm. The l_1-norm of the sparse vector is employed as the activity level measurement for fusion. Finally, the fused image is obtained by aggregating all the reconstructed patches into a whole image with the pixel-wise "overlap-averaging" strategy. In [19], this method was further modified by adding a constraint to ensure that the patches of different source images at the same location are decomposed into the same subset of dictionary atoms. The related algorithm is called simultaneous OMP (SOMP). Since then, many improved SR-based image fusion methods have been proposed [18], [20], [21]. However, in these methods, as in most SR-based image processing algorithms, to reduce the modeling burden and computational cost, sparse coding has always been performed on local image patches rather than on the entire image. This patch-based manner tend to cause the following two drawbacks of SR-based image fusion methods.

1) Limited Ability in Detail Preservation: In SR-based methods, the resulting representation at each pixel is multivalued since adjacent patches are overlapped. Ideally, the multiple values of one pixel should be exactly the same to keep the consistency among overlapped patches [27]. However, since sparse coding is independently performed on each patch or patch set, this consistency constraint is ignored. In [22], Wohlberg pointed out that this representation is not optimized with respect to the entire image. The aggregation and averaging strategies are applied to obtain the final value for each pixel, which will inevitably change the original local structures in the source image. In particular, to avoid blocking artifacts, the degree of overlap within adjacent patches in SR-based methods is usually very high. As a result, some details are smoothed or even lost in the fused image.

2) High Sensitivity to Misregistration: It is well known that shift invariance is a crucial issue in image fusion. An image representation approach is shift invariant if it outputs the same representation (feature measure) for a target feature even if the feature is shifted in spatial domain. Based on a shift invariant representation approach, such as NSCT, it is usually easy to develop fusion methods that are insensitive to misregistration, e.g., by designing some simple fusion rules like the window-based averaging. Otherwise, if the representation approach is not shift invariant, the output measures of different source images will not represent the same feature when they are misregistered, leading to meaningless comparison. As a consequence, the pseudo-Gibbs phenomenon tends to be observed around the misregistered regions in the fused image.

The SR of an image is only shift invariant when the stride of patches is one pixel in both horizontal and vertical directions. However, the computational efficiency in this situation is usually very low. Even in this case, most patch-wise SR-based fusion methods [16], [18]–[20] are not robust to misregistration. This is because in these methods, one patch from a source image is just associated with the patches at the same location from other source images, while the information contained in their neighboring patches are not involved. Clearly, the sparse coefficients of a patch will not remain the same when a certain feature (e.g., an edge) within it undergoes a translation, so the corresponding comparison in these methods may be meaningless. When it comes to the entire image, the aggregation and averaging strategies may slightly alleviate this problem, but will not completely address the concern as the block dividing approach is independent of image contents.

B. Convolutional Sparse Representation

CSR can be viewed as an alternative representation to SR using the convolutional form, which aims to achieve the SR of an entire image rather than a local image patch. The basic idea of CSR is to model an image $\mathbf{s}$ as the sum over a set of convolutions between sparse coefficient maps $\mathbf{x}_m$ and dictionary filters $\mathbf{d}_m$ [22]:

$$\arg\min_{\{\mathbf{x}_m\}} \frac{1}{2} \left\| \sum_m \mathbf{d}_m * \mathbf{x}_m - \mathbf{s} \right\|_2^2 + \lambda \sum_m \|\mathbf{x}_m\|_1 \tag{2}$$

where $*$ denotes the convolution operator. In the past few years, several efficient algorithms [22], [28], [29] within the alternating direction method of multipliers [30] framework have been proposed to solve this problem in the frequency domain. This has improved the potential of this model for practical usage to a large extent.

According to a recent comprehensive survey by Wohlberg [22], CSR has been studied under two main streams. In the machine learning community, the CSR model was introduced by Zeiler *et al.* [31] in their deconvolutional networks for feature learning. While in the signal processing community, CSR is also known as translation/shift-invariant SR [32], [33]. The convolutional form is derived starting from the shift-invariance principle.

In this letter, we demonstrate that CSR is an effective model for image fusion as it can overcome the above two drawbacks of SR-based fusion methods, primarily based on the following two motivations:

1) Since the result of CSR is single-valued and optimized over the entire image [22], the issue of detail preservation should be well solved;

2) Since CSR is essentially a shift-invariant image representation model, the fusion quality in misregistered regions should be significantly improved.

III. Proposed Fusion Scheme

Suppose that there are K preregistered source images denoted as $\mathbf{I}_k, k \in \{1, \ldots, K\}$, and suppose a set of dictionary filters $\mathbf{d}_m, m \in \{1, \ldots, M\}$ have been learned by the dictionary learning method in [22]. The proposed CSR-based image fusion framework consists of the following four steps.

A. Two-Scale Image Decomposition

Each source image $\mathbf{I}_k$ is first decomposed into a base layer $\mathbf{I}_k^b$ and a detail layer $\mathbf{I}_k^d$. The base layer is obtained by solving the following optimization problem:

$$\arg\min_{\mathbf{I}_k^b} ||\mathbf{I}_k - \mathbf{I}_k^b||_F^2 + \eta(||\mathbf{g}_x * \mathbf{I}_k^b||_F^2 + ||\mathbf{g}_y * \mathbf{I}_k^b||_F^2) \quad (3)$$

where $\mathbf{g}_x = [-1\ 1]$ and $\mathbf{g}_y = [-1\ 1]^T$ are the horizontal and vertical gradient operators, respectively. The regularization parameter η is set to 5 in this letter. This is a Tikhonov regularization problem and can be efficiently solved by the fast Fourier transform. Having $\mathbf{I}_k^b$, the detail layer can be obtained by subtraction:

$$\mathbf{I}_k^d = \mathbf{I}_k - \mathbf{I}_k^b. \quad (4)$$

This kind of two-scale decomposition approach has been employed in many image fusion methods such as [34].

B. Fusion of Detail Layers

For each detail layer $\mathbf{I}_k^d$, its sparse coefficient maps $\mathbf{C}_{k,m}, m \in \{1, \ldots, M\}$ are obtained by solving the CSR model with the method in [22]:

$$\arg\min_{\{\mathbf{C}_{k,m}\}} \frac{1}{2} \left\| \sum_{m=1}^{M} \mathbf{d}_m * \mathbf{C}_{k,m} - \mathbf{I}_k^d \right\|_2^2 + \lambda \sum_{m=1}^{M} \|\mathbf{C}_{k,m}\|_1. \quad (5)$$

Let $\mathbf{C}_{k,1:M}(x,y)$ denote the contents of $\mathbf{C}_{k,m}$ at the position (x,y) in spatial domain. Clearly, $\mathbf{C}_{k,1:M}(x,y)$ is an M-dimensional vector. In accord with the approach used in the SR-based fusion methods [16], [19], the l_1-norm of $\mathbf{C}_{k,1:M}(x,y)$ is adopted as the activity level measure of source images. Thus, the activity level map $\mathbf{A}_k(x,y)$ is obtained by

$$\mathbf{A}_k(x,y) = \|\mathbf{C}_{k,1:M}(x,y)\|_1. \quad (6)$$

To make the method insensitive to misregistration, a window-based averaging strategy is performed on $\mathbf{A}_k(x,y)$ to obtain the final activity level map:

$$\bar{\mathbf{A}}_k(x,y) = \frac{\sum_{p=-r}^{r} \sum_{q=-r}^{r} \mathbf{A}_k(x+p, y+q)}{(2r+1)^2} \quad (7)$$

where r determines the window size. The method will be more robust to misregistration with a larger r, but some tiny details may be lost at the same time. In multifocus image fusion, object edges in multiple source images have different sharpness so that the locations are not exactly the same, thus, a relatively larger r is preferred. In multimodal image fusion, as small-scale details often exist, it is more appropriate to adopt a smaller r. In this letter, r is fixed to 9 and 3 for multifocus and multimodal image fusion, respectively.

TABLE I
Objective Assessment of Different Fusion Methods

Images	Metrics	SR-OMP	SR-SOMP	CSR-16	CSR-32	CSR-64	CSR-128
Multifocus	EN	7.2947	7.2955	7.3011	7.3062	7.3071	**7.3076**
	$Q^{\mathrm{AB/F}}$	0.7618	0.7635	0.7739	0.7782	0.7787	**0.7789**
	Q_{HVS}	0.7722	0.7729	0.7826	0.7852	**0.7865**	0.7862
	VIFF	0.9293	0.9299	0.9340	0.9387	0.9397	**0.9401**
Visible-infrared	EN	6.3029	6.3035	6.3088	6.3125	**6.3141**	6.3140
	$Q^{\mathrm{AB/F}}$	0.5615	0.5664	0.5871	0.5901	**0.5922**	0.5918
	Q_{HVS}	0.4803	0.4821	0.4964	0.4996	0.5008	**0.5014**
	VIFF	0.3437	0.3448	0.3600	0.3627	0.3635	**0.3637**
Medical	EN	5.4524	5.4537	5.4815	5.4901	5.4924	**5.4928**
	$Q^{\mathrm{AB/F}}$	0.6022	0.6056	0.6104	0.6152	0.6166	**0.6169**
	Q_{HVS}	0.5620	0.5623	0.5798	0.5823	0.5833	**0.5835**
	VIFF	0.5169	0.5174	0.5212	0.5236	0.5249	**0.5255**

Then, the "choose-max" strategy is applied to achieve the fused coefficient maps:

$$\mathbf{C}_{f,1:M}(x,y) = \mathbf{C}_{k^*,1:M}(x,y),\ k^* = \arg\max_k(\bar{\mathbf{A}}_k(x,y)). \quad (8)$$

Finally, the fusion result of detail layers is reconstructed by

$$\mathbf{I}_f^d = \sum_{m=1}^{M} \mathbf{d}_m * \mathbf{C}_{f,m}. \quad (9)$$

C. Fusion of Base Layers

Different fusion schemes for base layers are applied to different types of fusion tasks. For multifocus image fusion, the most important issue is detail extraction. Since some details still remain in the base layers, the "choose-max" strategy based on the activity level maps obtained from detail layer fusion is adopted. The multifocus fusion result of base layers is

$$\mathbf{I}_f^b(x,y) = \mathbf{I}_{k^*}^b(x,y),\ \ k^* = \arg\max_k(\bar{\mathbf{A}}_k(x,y)). \quad (10)$$

However, this selection strategy may cause visual inconsistency when source images are obtained with different modalities, as the grayscale values of the same position may vary significantly. Thus, the averaging strategy is applied to multimodal image fusion. The multimodal fusion result of base layers isy

$$\mathbf{I}_f^b(x,y) = \frac{1}{K} \sum_{k=1}^{K} \mathbf{I}_k^b(x,y). \quad (11)$$

D. Two-Scale Image Reconstruction

Having $\mathbf{I}_f^b$ and $\mathbf{I}_f^d$, the fused image $\mathbf{I}_f$ is reconstructed by

$$\mathbf{I}_f = \mathbf{I}_f^d + \mathbf{I}_f^b. \quad (12)$$

IV. Experiments

In our experiments, ten pairs of multifocus images, ten pairs of visible-infrared images, and ten pairs of multimodal medical images are used as test images. Four popular objective metrics, namely, the entropy EN, the gradient-based metric $Q^{\mathrm{AB/F}}$ [35],

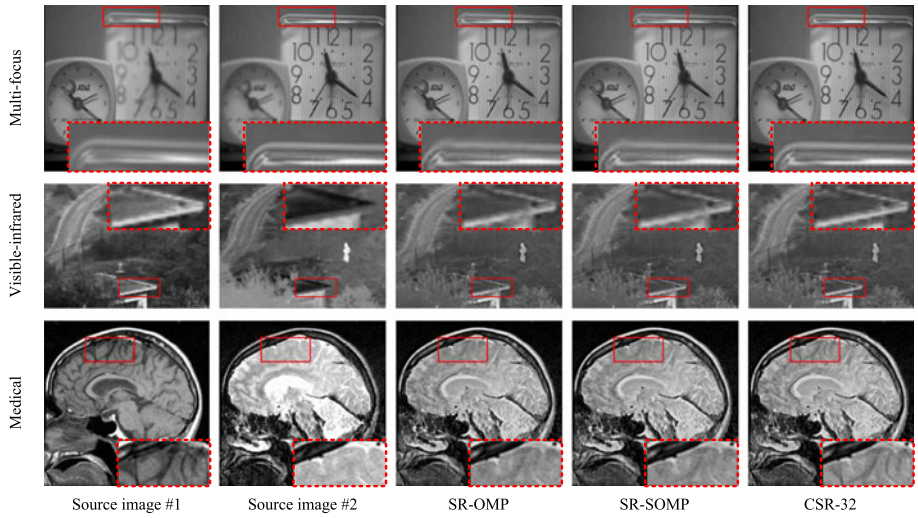

Fig. 1. Three fusion examples in our experiments.

the HVS-based metric Q_{HVS} [36] and the visual information fidelity fusion metric VIFF [37], are adopted to evaluate fusion performance.

Since the design of activity level measure and fusion scheme in our methods generally follow the SR-based methods [16], [19], to make a fair and clear comparison between SR and CSR for image fusion (also considering the limitation of paper length), the proposed CSR-based methods are primarily compared with these two methods in this letter to verify the advantage of CSR. These two methods are denoted as SR-OMP [16] and SR-SOMP [19]. All the parameters are set to the recommended values as reported in [16] and [19]. The dictionary used in these two methods has 256 atoms and is learned by the K-SVD method [38] from natural image patches. For the proposed methods, the spatial size of each dictionary filter is set to 8×8, which is in accord with the size of dictionary atom in SR-OMP and SR-SOMP. The learning method presented in [22] is employed to learn the dictionary filters from 50 high-quality natural images of size 256×256. In our experiments, the number of dictionary filters is set to 16, 32, 64, and 128, respectively, to study its impact on fusion performance. The corresponding methods are denoted as CSR-16, CSR-32, CSR-64, and CSR-128, respectively. Please refer to the supplementary material for more details about dictionary learning. According to the study (for image reconstruction applications) in [22], the parameter λ in Eq. (5) is fixed at 0.01 in all the experiments.

Table I reports the objective assessment results of different fusion methods. The average scores over all ten test examples are calculated and the highest value at each row shown in bold indicates the best score among the different methods. It can be seen that the proposed CSR-based methods have considerable advantages over the SR-based methods for the three types of fusion tasks on each metric. For the CSR-based methods, the fusion performance generally improves when increasing the number of dictionary filters, but the differences are relatively small, especially when there are more than 32 filters.

One fusion example is selected from each type of image fusion task, as shown in Fig. 1 The fused images of methods SR-OMP, SR-SOMP, and CSR-32 are exhibited. A close-up is provided in each example for better comparisons. In the multifocus example, the CSR-based method obtains better performance around misregistered edge regions than the SR-based methods (due to the different focal settings applied, the locations of object edges in different source images are not precisely the same for their different sharpness). In two multimodal examples, the CSR-based method clearly extracts more details from source images than the SR-based methods. More fusion examples could be found in the supplementary material.

TABLE II
CPU RUNNING TIME IN SECONDS OF DIFFERENT METHODS FOR MERGING TWO SOURCE IMAGES OF SIZE 256×256 PIXELS

Methods	SR-OMP	SR-SOMP	CSR-16	CSR-32	CSR-64	CSR-128
Time	836	1565	14.4	39.5	101	259

Table II lists the CPU running time of different fusion methods for merging two source images of size 256×256 pixels. All the methods are implemented in pure MATLAB with a 3.6 GHz CPU and 8 GB RAM. It can be seen that the CSR-based methods are much more time-efficient when compared with the SR-based methods.

V. CONCLUSION

In this letter, we introduce CSR into image fusion to address the drawbacks of SR-based fusion methods. A simple-yet-effective fusion framework based on CSR is proposed. Experimental results on multifocus and multimodal image fusion demonstrate the advantages of the proposed CSR-based methods over SR-based methods. For future work, the impacts of some free parameters in our algorithm will be studied, such as the regularization parameter in the CSR model and the spatial size of dictionary filters. In addition, similar to the great improvements achieved in SR-based image fusion, more effective fusion schemes based on CSR could be further developed to pursue better fusion performance. We believe that CSR has great potential to become a popular model for image fusion in the near future.

REFERENCES

[1] S. Li, X. Kang, L. Fang, J. Hu, and H. Yin, "Pixel-level image fusion: A survey of the state of the art," *Inf. Fusion*, vol. 33, pp. 100–112, 2017.

[2] T. Stathaki, *Image Fusion: Algorithms and Applications*. New York, NY, USA: Academic, 2008.

[3] S. Li, J. Kwok, and Y. Wang, "Combination of images with diverse focuses using the spatial frequency," *Inf. Fusion*, vol. 2, no. 3, pp. 169–176, 2001.

[4] A. Goshtasby, "Fusion of multi-exposure images," *Image Vis. Comput.*, vol. 23, no. 6, pp. 611–618, 2005.

[5] V. Aslantas and R. Kurban, "Fusion of multi-focus images using differential evolution algorithm," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8861–8870, 2010.

[6] X. Bai, Y. Zhang, F. Zhou, and B. Xue, "Quadtree-based multi-focus image fusion using a weighted focus-measure," *Inf. Fusion*, vol. 22, no. 1, pp. 105–118, 2015.

[7] S. Li, X. Kang, J. Hu, and B. Y, "Image matting for fusion of multi-focus images in dynamic scenes," *Inf. Fusion*, vol. 14, no. 2, pp. 147–162, 2013.

[8] W. Zhang and W.-K. Cham, "Gradient-directed multiexposure composition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2318–2323, Apr. 2012.

[9] B. Gu, W. Li, J. Wong, M. Zhu, and M. Wang, "Gradient field multi-exposure images fusion for high dynamic range image visualization," *J. Vis. Commun. Image Represent.*, vol. 23, no. 4, pp. 604–610, 2012.

[10] Y. Liu, S. Liu, and Z. Wang, "Multi-focus image fusion with dense sift," *Inf. Fusion*, vol. 23, no. 1, pp. 139–155, 2015.

[11] P. Burt and R. Kolczynski, "Enhanced image capture through fusion," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1993, pp. 173–182.

[12] H. Li, B. Manjunath, and S. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57, no. 3, pp. 235–245, 1995.

[13] 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, Feb. 2015.

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

[15] N. Mitianoudis and T. Stathaki, "Pixel-based and region-based image fusion schemes using ICA bases," *Inf. Fusion*, vol. 8, no. 2, pp. 131–142, 2007.

[16] 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.

[17] A. Goshtasby and S. Nikolov, "Image fusion: Advances in the state of the art," *Inf. Fusion*, vol. 8, no. 2, pp. 114–118, 2007.

[18] Y. Liu and Z. Wang, "Simultaneous image fusion and denoising with adaptive sparse representation," *IET Image Process.*, vol. 9, no. 5, pp. 347–357, 2015.

[19] B. Yang and S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Inf. Fusion*, vol. 13, no. 1, pp. 10–19, 2012.

[20] N. Yu, T. Qiu, F. Bi, and A. Wang, "Image features extraction and fusion based on joint sparse representation," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 1074–1082, Sep. 2011.

[21] Q. Zhang and M. 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, May 2016.

[22] B. Wohlberg, "Efficient algorithms for convolutional sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 301–315, Jan. 2016.

[23] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[24] A. Bruckstein, D. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.

[25] J. Yang, T. H. J. Wright, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[26] M. Wang, X. Liu, and X. Wu, "Visual classification by l1-hypergraph modeling," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 9, pp. 2564–2574, Sep. 2015.

[27] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1823–1831.

[28] H. Bristow, A. Eriksson, and S. Lucey, "Fast convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 391–398.

[29] F. Heide, W. Heidrich, and G. Wetzstein, "Fast and flexible convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5135–5143.

[30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[31] M. Zeiler, D. Krishnan, G. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2528–2535.

[32] M. Lewicki and T. Sejnowski, "Coding time-varying signals using sparse, shift-invariant representations," in *Proc. Adv. Neural Inf. Process. Syst.*, 1999, pp. 730–736.

[33] M. Morup and M. Schmidt, "Transformation invariant sparse coding," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, 2011, pp. 1–6.

[34] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

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

[36] Y. Chen and R. Blum, "A new automated quality assessment algorithm for image fusion," *Image Vis. Comput.*, vol. 27, no. 10, pp. 1421–1432, 2009.

[37] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Inf. Fusion*, vol. 14, pp. 127–135, 2013.

[38] 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.