Multi-focus Image Fusion with Online Sparse Dictionary Learning

Jun Wang , Lu Liu , Xuan Zhu and Na Ai
School of Information Science and Technology
Northwest University
Xi'an, Shaanxi Province, 710127, China
wangjun05062403@163.com

Kun Yan
Institute of Remote Sensing and Data Transmission
China Academy of Space Technology (Xi'an)
Xi'an, Shaanxi Province, 710000, China
yankun035938@sina.com

Abstract—**This paper presents an effective multi-focus image fusion method based on the online sparse dictionary learning with double sparsity model. First, we learn the dictionaries through the source images using the online sparse dictionary learning algorithm, which enables a multi-scale analysis and train an adaptive dictionary. Then, the sparse representation coefficients of the source images may be acquired by the learned dictionary. Finally, the fused image is formed by choosing the max fusion rule and the learned dictionary. Experimental results show that the proposed method is superior to the conventional fusion methods in terms of the visual and indicator evaluation.**

Keywords-multi-focus image fusion; online sparse dictionary learning; double-sparsity model; K-SVD

I. INTRODUCTION

Multi-focus image fusion technique is capable to combine the complementary information of different focus multiple images of the same scene and have been registered into an image to achieve a clear scene description [1][2]. In contrast, the ordinary optical systems cannot even be clearly imagined for all targets in the same scene.

Conventional multi-focus image fusion methods are mainly based on the spatial domain and the transform domain. The method of image fusion based on the transform domain includes Laplacian tower decomposition [3] method and the wavelet transform [4] method. The Laplacian tower method performs the fusion processing in different spatial frequency bands, which is widely applied. However, it suffers from a correlation between the amount of the inter-layer decomposition. The wavelet transform has a good spatial and frequency domain localization, which may well preserve the high frequency information of the multi-focus image, but still lose some characteristic information of the original image to a certain degree, such as for instance the edge information.

Recently, new redundant basis functions have become a hotspot for the signal sparse approximation [5], due to presenting reasonable image fusion result. For example, Li et al. [6] applied the sparse representation in the field of image fusion, and also the sparseness of the image to propose a multi-focus image fusion method based on sparse representation, achieving fusion results superior to the conventional methods. Yu et al. [7] proposed a feature extraction and fusion method based on the joint sparse representation, explicitly extracting the common features and characteristics of the source images, and achieved good fusion effect. Wang et al. [8][9] proposed a series of interesting methods based on the double sparse fusion. These methods are further sparse under the DWT or NSCT wavelet basis and effectively extract and fuse characteristics of the source images. However, the image fusion method based on the sparse representation hardly takes into account the quality of the fused image and the computational complexity. As a result, development of the image fusion method based on the sparse representation is very limited.

This paper proposes a new image fusion method based on the online sparse dictionary learning technique. We establish base dictionaries of the source images by online sparse dictionary learning (OSDL) method, and then represent the source images on their basis. Finally, we choose the maximum rules of the coefficient activity to achieve the image fusion. The experimental results show that, compared with the conventional methods, the proposed method provide better fusion result. It also requires lower computational complexity than the fusion methods based on the sparse representation.

II. ONLINE SPARSE DICTIONARY LEARNING

Based on the sparse representation theory, any signal can be expressed as a linear combination of a few atoms in a complete dictionary. It considers a sparse characteristic between the atoms in the dictionary, where the atoms in the valid dictionary may be expressed as a linear combination of a few arbitrary atoms in the base dictionaries [10]. The double sparse model of the sparse representation theory may be expressed as:

$$\min_{A,X} \frac{1}{2}\|Y-\Phi AX\|_F^2 \quad s.t. \quad \begin{cases} \|x_i\|_0 \le p & \forall i \\ \|a_j\|_0 = k & \forall j \end{cases}. \tag{1}$$

where $\Phi \in R^{n\times L}$ is the base dictionary, $A \in R^{L\times m}$ and $X \in R^{m\times n}$ respectively represent the sparse dictionaries and coefficient matrices, and p and k are respectively the sparse coefficients and the dictionary sparsities.

The training dictionary learning under the double sparse model represents a key problem. The learning process uses the batch processing instead of the original small image block processing. This yields reducing number of the alternate transformations when the sparse coding and the sparse dictionary are updated, which may further reduce the computational complexity. First, we initialize a sparse dictionary, and then solve the sparse representation

978-1-5386-0497-7/17/$31.00 ©2017 IEEE

coefficients under a sparse dictionary for the training samples, and finally update the sparse dictionary. The sparse dictionary learning may be expressed as:

$$\min_{A} \frac{1}{2}\|Y-\Phi AX\|_F^2 \qquad s.t. \qquad \|a_j\|_0 = k \qquad \forall j. \tag{2}$$

with matrix $A \in R^{L\times m}$ being a sparse dictionary, where each column contains k nonzero elements. At present, many dictionary learning methods are to update the atoms in the dictionary column by column. As a result, the m-minimization problem may be changed to:

$$\min_{A} \underbrace{\frac{1}{2}\|E_j-\Phi a_j x_j^T\|_F^2}_{f(a_j)} \qquad s.t. \qquad \|a_j\|_0 = k. \tag{3}$$

where the residual $E_j = Y - \sum_{i\neq j}\Phi a_j x_j^T$, and x_j^T denotes the j^{th} row of x, for j = 1, 2, ..., m, to update each column of A. In the OSDL, each atom in the dictionary is updated by the random gradient descent method. Since the global update of the dictionary is faster than the dictionary atom at a particular step size, the update of the atoms in the sparse dictionary can be expressed as follows:

$$A^{t+1} = P_k\left[A^t - \eta^t \nabla f(A^t)\right]. \tag{4}$$

where P_k represents the threshold operator and selects the largest value in each column. Moreover, in (4), η^t represents the step size, and the global step in OSDL is expressed as:

$$\eta^* = \frac{\|\nabla f(A_s)\|_F}{\|\Phi\nabla f(A_s)X\|_F}. \tag{5}$$

The desired dictionary may be achieved by combining the learned sparse and the base dictionaries. The OSDL algorithm improves the convergence speed in the learning process, greatly reduces the computational complexity, effectively avoids the local minimum, and provides the structural dictionary with better performance, faster computation and adaptability.

III. PROPOSED MULTI-FOCUS IMAGE FUSION METHOD

A. Source Image sparse representation

Since the pixel-level image fusion is used to process the image local information, we first divide the source images into small patches. Considering Y as the input image, and y^i as the i^{th} patch, with the size of $\sqrt{n}\times\sqrt{n}$, extracted from the input image Y, as a n-dimension-ality column vector. The image patch y^i may be sparsely represented with the dictionary D, so the sparse representation based decomposition model of the image patch y^i may be represented as:

$$\min_{\alpha^i}\|\alpha^i\|_0 \qquad s.t. \qquad \|y^i - D\,\alpha^i\|_2^2 \leq \varepsilon. \tag{6}$$

with α^i denoting the sparse coefficients, and the parameter ε as the decomposition error.

In the above model, selection of the dictionary is vital, and OSDL method is used to learn the dictionary. Moreover, we use the OMP algorithm to solve the above sparse decomposition model, because of its various advantages, e.g. fast computation and low complexity.

B. Fusion step

This paper proposes a multi-focus image fusion method called *online sparse dictionary learning image fusion method*, dubbed as OSDLF. Figure 1 shows the flowchart of the OSDLF

Step 1: the source images, I_1 and I_2, should be divided into patches with size of $\sqrt{n}\times\sqrt{n}$, and all patches are lexicographically ordered as n -dimensional column vectors, represented by Y_1 and Y_2.

Step 2: the training set Y are considered as the concatenation of Y_1 and Y_2, and then the train Y by using OSDL algorithm according to the formulas (2) - (5) to achieve the sparse dictionary A. We may achieve the online sparse learning dictionary D by forming the base dictionary and A linearly.

Step 3: According to (6), the OMP algorithm is used for the sparse coding of Y1 and Y2 on the dictionary D .We then achieve the sparse coefficients α_1 and α_2.

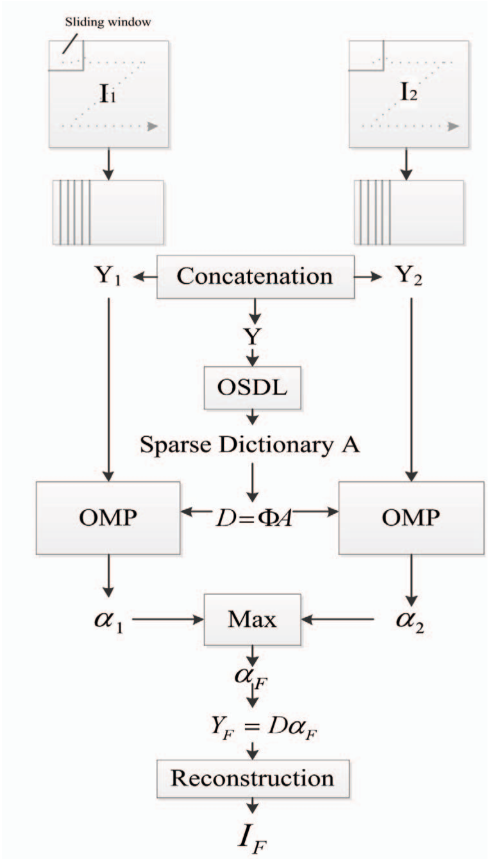

Figure 1. Flower chart of the OSDLF

Step 4: Choose the largest activity coefficient as the fused image in the fusion image coefficient α_F^j , where j th row of α_F reads:

$$\alpha_F^j = \begin{cases} \alpha_1^j, & \|\alpha_1^j\|_1 > \|\alpha_2^j\|_1 \\ \alpha_2^j, & \|\alpha_1^j\|_1 \le \|\alpha_2^j\|_1 \end{cases} \tag{7}$$

Step 5: The fused vector Y_F is given by:

$$Y_F = Da_{\mathrm{F}} . \tag{8}$$

Step 6: Reshape each vector Y_{F_j} in Y_F into a block with size of $\sqrt{n} \times \sqrt{n}$, then add the block to I_F at its responding position., and apply the average processing to each location to obtain the final fused image I_F .

IV. Experimental Analysis

This section compares the performance of the proposed method with five different methods. The five methods are as follows. Three classical image fusion algorithms, i.e. the Discrete Wavelet Transform (DWT), the Laplacian Pyramid (LP) and the Principal Components Analysis (PCA). The NSCT_K-SVD method [8] is an advanced image fusion method based on the sparse representation, which performs the Nonsubsampled Contourlet (NSCT) on the image, and uses the K-SVD algorithm to train the dictionary. Finally, we consider the NSCT_OSDLF method to analysis performance of the proposed method under the NSCT domain.

The experiment includes two types of images, i.e. the reference image and the non-reference image. The objective evaluation is necessary for the fused images. The reference fused image is available in the experiment of the reference image. Therefore, we use two objective fusion metrics, i.e. the peak signal to noise ratio (PSNR) and the root mean square error (RMSE). The RMSE values the smaller the better and the PSNR values the bigger the better. In the experiment of the non-reference image, the reference fused image is unavailable, while the space frequency (SF) and the $Q^{AB/F}$ are used to evaluate the fusion image. Besides, larger value for the two aforementioned evaluations implies better fusion result. In addition, the experiments are implemented in Matlab 2014a.

A. Experimental results of the Reference image

Figure 2 shows the experiment, where a pair of multi-focus images with reference image is used to test the proposed method. Figures 2(a) and (b) show the source images, where Fig. 2(a) focuses on the wolf which is in front of the image, and Fig. 2(b) focuses on the mountain behind the wolf. Figure 2(c) is the reference fused image and Figs. 2(d)-(i) depict the fusion results achieved by different methods. It may be seen that Fig. 2(i) presents visually better result than Figs.2 (d)-(i). Besides of the subjective visual analysis, we see the objective evaluation results.

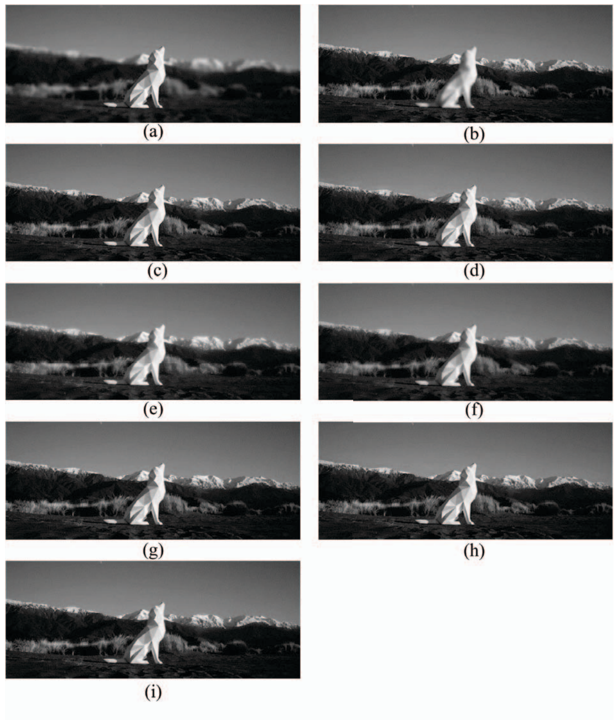

Figure 2. Multi-focus image fusion results by several methods (a) Source image A (b) Source image B (c) Reference image (d) DWT (e) LP (f) PCA (g) NSCT_K-SVD (h) NSCT_OSDLF (i) OSDLF

TABLE I. Objective evaluation of several fusion methods of reference image fusion

Methods	PSNR	RMSE	Time
DWT	24.5026	15.1848	1.782s
LP	23.0179	18.0155	1.067s
PCA	24.9195	14.4732	0.795s
NSCT_K-SVD	24.8328	14.6184	430.190s
NSCT_OSDLF	24.9542	14.4154	61.803s
OSDLF	**25.8377**	**13.0213**	243.443s

Table Ⅰ lists the objective evaluation indexes of the six fusion algorithms, where the bold indicate the best results. In comparison with the three classical methods, i.e. DWT, LP, PCA, the proposed method requires more time. However, the results of the PSNR and RMSE demonstrate that the proposed method may provide high-quality fused images. As it may be seen in Table Ⅰ， the processing time of the OSDLF is shorter than the one of the NSCT_K-SVD. In addition, better processing time is achieved by the OSDLF on the basis of NSCT.

Evaluating the results of the subjective visual and objective indicators, we may see that the fusion image given by the OSDLF has least difference from the reference image. In general, the OSDLF algorithm is superior to the other five methods.

B. Experimental results of Non-reference image

Figure 3 shows the experiment, where a pair of the multi-focus image without reference image are used to test the proposed method. Figures 3(a)(b) depict the source images.

Figure 3(a) presents the left side of the image with a clock while Fig. 3(b) shows the right side of the image with a student. Figures 3(c)-(h) present the fusion results achieved by various methods, while Fig. 3(h) represents the clearest image.

Table Ⅱ lists the objective evaluation indexes of the six fusion algorithms, and the best results have been bolded. Although value of the SF is little smaller than NSCT_K- SVD, value of the $Q^{AB/F}$ is much better. Besides, compared with the classical DWT, LP and PCA methods, the proposed method requires more computation time and gives better result. However, due to the ‘‘sliding window” scheme, the fusion method based on the sparse representation methods are time-consuming[6][7][8], even if it takes about 30 minutes to fuse the images. Nevertheless, the processing time of the proposed method is shorter than the one of the NSCT_K-SVD method, proving that the OSDLF may greatly improve the fusion efficiency. We can also see that the OSDLF on the basis of NSCT requires much shorter processing time.

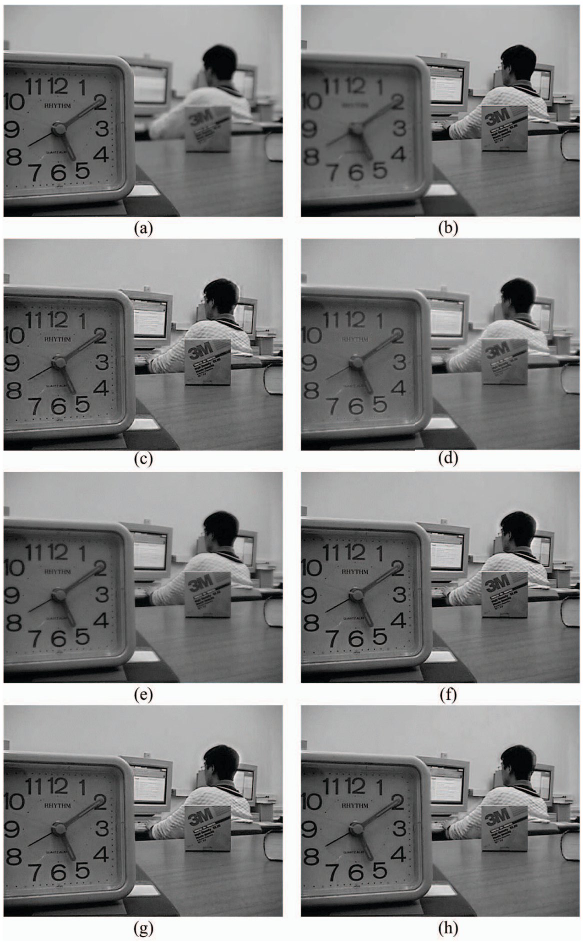

Figure 3. Multi-focus image fusion results by several methods (a) Source image A (b) Source image B (c) DWT (d) LP (e) PCA (f) NSCT_K-SVD (g) NSCT_OSDLF (h) OSDLF

In conclusion, the proposed fusion approach is superior to other methods through subjective visual analysis and objective quantitative evaluation.

TABLE II. OBJECTIVE EVALUATION OF SEVERAL FUSION METHODS OF NON-REFERENCE IMAGE FUSION

Methods	SF	$Q^{AB/F}$	Time
DWT	12.9267	0.6424	3.181s
LP	8.3403	0.5093	1.402s
PCA	7.7657	0.5611	0.978s
NSCT_K-SVD	**12.9752**	0.6757	636.716s
NSCT_OSDLF	12.9475	0.6959	85.667s
OSDLF	12.9419	**0.7002**	566.033s

V. CONCLUSION

We presented an effective multi-focus image fusion method based on online sparse dictionary learning. The major contribution of this study is to use the double sparse model to learn the dictionary and to effectively extract the source image while making the method less time consuming. The simulation data and the actual data verify that both the NSCT_OSDLF method under the NSCT double sparse model and the spatial domain OSDLF method provide better visual effects and objective indexes than the conventional fusion methods. Moreover, the calculation rate is much higher than other previously reported methods operating based on the sparse representation techniques.

ACKNOWLEDGMENT

This work is supported by Special Research Project of Shaanxi Education Department (16JK1774); the National Nature Science Foundation of China under Grant (No. 61402368, and No. 61272285).

REFERENCES

[1]Z Liu, Y Chai, H Yin, J Zhou, and Z Zhu , “A novel multi-focus image fusion approach based on image decomposition,” Information Fusion, 2017,35:102-116.

[2]D Xiao, L Wang, T Xiang, and Y Wang, “Multi-focus image fusion and robust encryption algorithm based on compressive sensing,” Optics and Laser Technology,2017,91:212-225.

[3]M Cai, J Yang, and G Cai, “Multi-focus image fusion algorithm using LP transformation and PCNN,” IEEE International Conference on Software Engineering and Service Science, 2015:237-241.

[4]Q Guo and S Liu, “Performance analysis of multi-spectral and panchromatic image fusion techniques based on two wavelet discrete approaches,” Optik, 122(9), 811-819(2011).

[5]Y Yang, Y Que, S Huang, and P Lin, “Multiple Visual Features Measurement With Gradient Domain Guided Filtering for Multisensor Image Fusion,” IEEE Transactions on Instrumentation & Measurement, 2016, PP(99):1-13.

[6]B Yang, and S Li, “Multifocus Image Fusion and Restoration With Sparse Representation,” IEEE Transactions on Instrumentation & Measurement, 2010, 59(4):884-892.

[7]N Yu, T Qiu, F Bi, and A Wang, “Image Features Extraction and Fusion Based on Joint Sparse Representation,” IEEE Journal of Selected Topics in Signal Processing, 2011, 5(5):1074-1082.

[8]J Wang, J Peng, X Feng, G He,J Wu, and K Yan, “Image fusion with nonsubsampled contourlet transform and sparse representation,”Journal of Electronic Imaging, 2013, 22(4):043019.

[9]J Wang, J Peng, G He, X Feng, and K Yan, “Fusion Method for Visible and Infrared Images Based on Non-subsampled Contourlet Transform and Sparse Representation, ”Acta Armamentarii,2013,34(7):815-820.

[10]J Sulam , B Ophir, M Zibulevsky, and M Elad, “Trainlets: Dictionary Learning in High Dimensions,” IEEE Transactions on Signal Processing, 2016, 64(12):3180-3193.