A ROBUST FUSION SCHEME FOR MULTIFOCUS IMAGES USING SPARSE FEATURES

Tao Wan[1], *Zengchang Qin*[2*], *Chenchen Zhu*[2], *Renjie Liao*[2]

[1]Boston University School of Medicine, MA 02215, USA
[2]Intelligent Computing and Machine Learning Lab,
School of ASEE, Beihang University, Beijing, 100191, China

ABSTRACT

Multifocus image fusion is an important research topic in the computer vision and image processing field. The optical lenses that are commonly used by imaging devices, such as auto-focus cameras, have a limiting focus range. Thus, only objects within the range of distances from the devices can be captured and recorded sharply while out-of-range objects become blur. In this paper, we present a novel image fusion scheme for combining two or multiple images with different focus points to generate an all-in-focus image. We formulate the problem of fusing multifocus images as choosing most significant features from a sparse matrix produced by a newly developed robust principal component analysis (RPCA) decomposition method to form a composite feature space. Thus, the salient features presented in sharp regions can be captured and integrated into a single representation. The sparse matrix is first divided into small blocks, and standard deviation is then calculated on each block as a selection criterion. To reduce blocking artifacts, a sliding window technique is utilized to smooth the transitions between blocks. The proposed fusion scheme has been demonstrated to successfully improve fusion quality in terms of visual and quantitative evaluations. The method is also able to effectively handle both grayscale and color images.

Index Terms— Multifocus image fusion, robust principal component analysis, sparse matrix.

1. INTRODUCTION

Due to the fact that commonly used optical lenses suffer from a problem of limited depth of field, images being captured are not in focus everywhere. Only objects at one particular depth will be truly in focus while out-of-focus objects remain blur, which is usually undesirable for human visual perception and often cause difficulties in image-processing tasks, such as segmentation, feature extraction, and object recognition. Image fusion provides a promising way to solve this problem by combining multiple images taken with diverse focuses into a single image in which all the objects within the image are in focus [1, 2].

Multifocus image fusion has been widely used in various fields, such as computer vision, remote sensing, digital imaging, and microscopic imaging. Currently, the exiting methods can be mainly categorized into two groups based on the different domains in which the fusion task is performed. Spatial domain based image fusion methods apply fusion rules directly to image pixels or image regions rather than transformed coefficients. For instance, Li *et al.* [3] devised a multifocus image fusion scheme by decomposing the source images into blocks and combining them based on the spatial frequency (SF). In the past few years, some more sophisticated fusion rules were proposed in [2]. These pixel or region based methods are simple to implement and fast to compute. However, they are usually subject to noise interference or blocking artifacts since the selection criteria used are computed based on single or neighboring pixels. Multiscale transforms have become popular in the image fusion field [4, 5]. These methods first decompose the source images into multiscale representations using a certain transformation. Some selection rules are then applied to the transformed images to form an integrated fusion map. Finally, a fused image is reconstructed via an inverse transformation. The transform domain based methods have showed many advantages, including improved contrast, better signal-to-noise ratio, and increased fusion quality.

Most recently, Yang and Li [6, 7] proposed a sparse representation theory based image fusion method, in which the source image can be described by a sparse linear combination of atoms from a dictionary. A set of sparsity coefficients was obtained via a simultaneous orthogonal matching pursuit. However, the redundant dictionary construction and sparse representation optimization are computationally expensive. Therefore, the method requires a longer computation time compared to spatial and transform domain based approaches.

In this paper, we propose a novel image fusion scheme which is truly different from the above approaches. Our method utilizes a newly introduced technique referred to as robust principal component analysis (RPCA), in which the data matrix is a superposition of a low-rank component and a sparse component [8]. In theory, under certain assumptions, it is feasible to recover both low-rank and sparse components

* Corresponding author's email: zcqin@buaa.edu.cn. This work is partially funded by the NCET Program of MOE, China and the SRF for ROCS.

exactly by solving the principal component pursuit. There are many important applications can be naturally modeled using this methodology, such as video surveillance, face recognition, bioinformatics, and web search [8]. We establish a multifocus image fusion framework based on the sparse features computed from the sparse matrix. The problem of fusing multifocus images is converted to the problem of selecting the most essential sparse features from the source images to form a composite feature space. The blocking effect is eliminated via a sliding window technique. The fused image is constructed through a decision scheme based on the extracted sparse features. Being implemented in this fashion, the proposed methodology is not only robust to noise interference by choosing the most significant sparse features, but also flexible to adopt different fusion rules in the sparse domain.

The paper is organized as follows. Section 2 describes the problem setting under the RPCA framework. The detailed methodology based on the RPCA model is presented in Section 3. The experimental results are demonstrated in Section 4. Section 5 concludes the paper.

2. PROBLEM SETTING

One core task of multifocus image fusion is to identify focused regions within each source image and eventually combine these objects in focus into a single image. In general, defocused objects appear very blur while objects located within the focus range are clearly captured. Therefore, the problem of fusing multifocus images can be treated as separating clear parts from blur parts of the images. A recently emerged RPCA technique tends to decompose the input data matrix into a low-rank principal matrix and a sparse matrix [8]. The sparse matrix represents dissimilar information from the principal components which can be useful to solve our problem. Assume we have an input data matrix $D \in \mathbb{R}^{M \times N}$ (M and N are matrix dimensions) that can be decomposed as:

$$D = A + E \tag{1}$$

where A is a principal matrix known to be low rank, and E is a sparse matrix. Although under general conditions this problem is intractable to solve, recent studies [8] have discovered that the principal component pursuit, a convex program, is able to effectively solve this problem under broad conditions. The sparse matrix E can be computed by solving the following convex optimization problem:

$$\min_{A,E} \|A\|_* + \lambda \|E\|_1 \quad \text{s.t.} \quad A + E = D \tag{2}$$

where $\|\cdot\|_*$ represents the nuclear norm of a matrix, $\|\cdot\|_1$ is the l_1 norm denoting the sum of the absolute values of matrix entries, and $\lambda > 0$ is a parameter for weighting the contribution of the sparse matrix in the optimization process. In our approach, the data matrix D with $M \times N$ dimensions

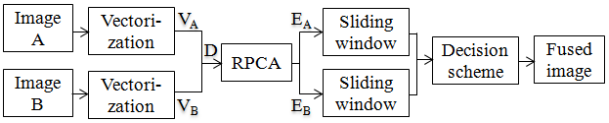

Fig. 1. The schematic diagram of the proposed fusion algorithm.

contains an $M \times 1$ matrix of N source images after vectorization. Thus, a color image can be handled as three individual images to form a single matrix. For a fast implementation, λ is set as $1/\sqrt{M}$. Fig. 2(c-d) show two examples of constructed images obtained from the sparse matrices after performing the RPCA decomposition on the source images. The figures clearly show that the sparse matrix contains salient information which reflects the edges of objects or regions in good focus. The detailed implementation is described in the following section.

3. METHODOLOGY

A schematic diagram of the proposed fusion method is shown in Fig. 1. For a simple case, we only consider the problem of fusing two source images, though it can be extended straightforwardly to process more than two images. In addition, we assume the source images are pre-registered. Therefore, image registration is not included in the entire framework. The algorithm consists of 5 steps:

Step 1: Transform the 2-dimensional source images $A, B \in \mathbb{R}^{H \times W}$ to the column vectors V_A and V_B, respectively. V_A and V_B are combined together to formulate a data matrix D:

$$D = [V_A \quad V_B] \tag{3}$$

where D is the input matrix for the RPCA model.

Step 2: Perform the RPCA decomposition on D to obtain a principal matrix A and a sparse matrix E. Reshape each column of matrix E to have two $H \times W$ matrices E_A and E_B.

Step 3: Divide the matrices E_A and E_B into K small blocks. For each pair of corresponding blocks, the standard deviations $SD_A(k)$ and $SD_B(k)$, $k = 1, ..., K$, are calculated. In a general rule, the block with a bigger standard deviation is chosen to construct the fused image F. However, This will lead to non-smooth transitions between blocks. In order to eliminate blocking artifacts, a sliding window technique is applied to the matrices E_A and E_B. Let $n_A(i,j)$ and $n_B(i,j)$ store the frequency of pixel location (i,j) being selected when a small window is moved from previous position to the current position on E_A and E_B. If a pixel location (i,j) in E_A has a higher standard deviation when the sliding window covers this position, the corresponding $n_A(i,j)$ is added one. This is also applied to E_B and $n_B(i,j)$.

Step 4: For a simple case where there are only two input images A and B, the decision map W can be created by:

$$W(i,j) = \begin{cases} 1 & n_A(i,j) > n_B(i,j) \\ -1 & n_A(i,j) < n_B(i,j) \\ 0 & n_A(i,j) = n_B(i,j) \end{cases} \quad (4)$$

Step 5: A 3×3 majority filter is applied to W to correct the wrong selected pixels due to the image noise. A fused image F is finally obtained after majority filtering.

$$F(i,j) = \begin{cases} A(i,j) & W(i,j) = 1 \\ B(i,j) & W(i,j) = -1 \\ (A(i,j) + B(i,j))/2 & W(i,j) = 0 \end{cases} \quad (5)$$

4. RESULTS AND DISCUSSION

The fusion method has been tested on various pairs of grayscale and color images, which are publicly available online [9]. The proposed approach has one tunable parameter of the block size S. In the experiments, S is set as 35×35 pixels for grayscale and 38×38 pixels for color images, respectively. Three reference methods are used for comparison. A simple discrete wavelet transform (DWT) based method utilizes a maximum selection rule to the high-pass coefficients and a mean operation to the low-pass coefficients. Tian and Chen [10] employed the spreading of the wavelet coefficients distribution as an image sharpness measure using a Laplacian mixture model (LMM). In addition, Li *et al.* [3] devised a multifocus image fusion method which adopted the spatial frequency as a selection criterion. For the sake of fair comparison, we use all the parameters that were reported by the authors to yield the best fusion results.

Fig. 2 shows the fusion results for two grayscale images focusing on left or right side. The original images with size of 512×512 pixels are displayed in Fig. 2(a-b), respectively. By inspecting the figures, it can be seen that the result obtained from DWT subjects to a severe ringing effect making the entire image blur. The LMM based method provides a sharp image but exhibits artifacts around edges of both clocks as indicated by the yellow rectangles. Fig. 2(g) yields a comparable result but still suffers a blocking effect. For example, vague edges can be observed on the top and bottom of the right clock. Our proposed algorithm achieves a superior result by containing all the sharp contents from the source images without introducing artifacts.

The second example combines two color images. Two individuals are standing about 30 feet apart with extended illumination as shown in Fig. 3(a-b). Both images are resized to be power of 2 to meet the requirement of the LMM method. Similarly, the DWT based method suffers a ringing effect that deteriorates the fusion quality. The fused image obtained from the LMM based method shows clear artifacts around figures and ceiling lights. The SF based method performs well but observes blocking artifacts on the right corner of the fused image (indicated by the yellow rectangle shown on Fig. 3(e)). Compared to these three methods, our result in Fig. 3(f) yields the best quality image in terms of visual

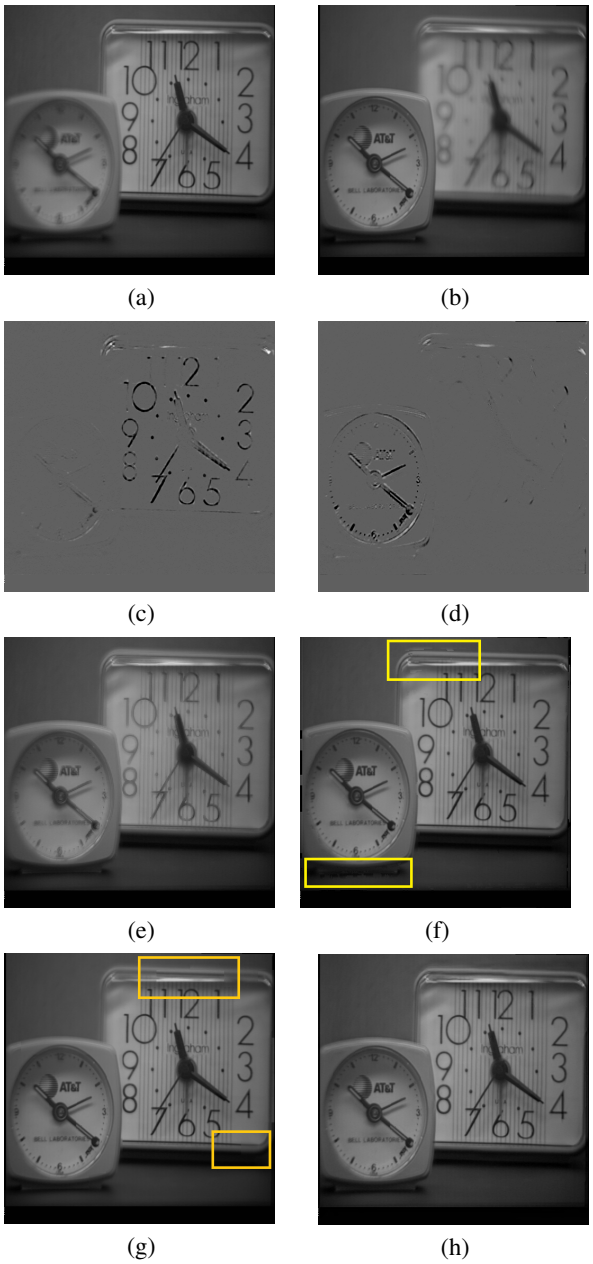

Fig. 2. Fusion Results using "Clock" grayscale images. (a-b) Original images. (c-d) Images constructed from the sparse matrix after RPCA decomposition. (e) DWT. (f) LMM [10]. (g) SF [3]. (h) RPCA. The yellow rectangles indicate the artifacts.

perception. For example, the lights on both right and left corners appear sharp, and lines on the ceiling are well connected. Further, the boundaries of both figures are smooth and clear.

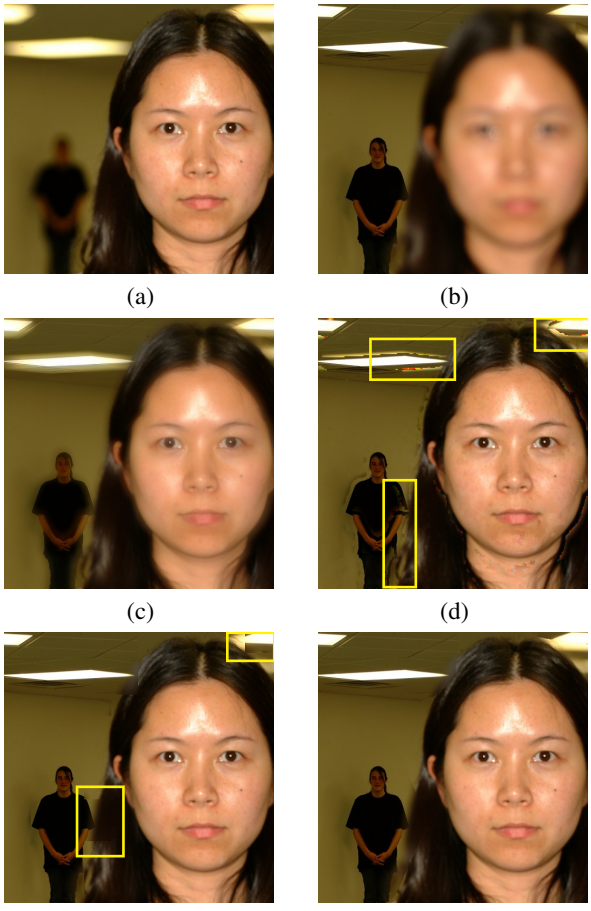

Fig. 3. Fusion results using "Human" color images. (a-b) Original images. (c) DWT. (d) LMM [10]. (e) SF [3]. (f) RPCA. The yellow rectangles indicate the artifacts.

Moreover, three image quality criteria are performed to provide objective evaluations. These three metrics are: (i) mutual information (MI) [11], (ii) Petrovic's metric $Q^{AB/F}$ [12] which measures the edge as well as the orientation information in both source images (denoted as A and B) and fused image (denoted as F), (iii) structural similarity index (SSIM) [13], which quantifies salient information that has been transferred into the fused image, where larger metric values imply better image quality. The quantitative results are tabulated in Table 1. Our method gains the highest MI, $Q^{AB/F}$, and SSIM values compared to other methods. In fact, due to the actual definitions of these three metrics, a difference of 0.01 is significant for the quality improvement. The computational complexity of these four methods is evaluated using the Matlab code on an Intel Core2 2.4GHz machines with a 4GB RAM. The running times are presented in Table 1, where one can see that the proposed approach yields higher computational cost than the other two methods, due to the matrix decomposition method requires a longer computational time.

Table 1. The objective evaluation and run-time performance

Image	Method	DWT	LMM	SF	PRCA
Clock	MI	7.47	8.07	7.99	**8.57**
	$Q^{AB/F}$	0.59	0.78	0.77	**0.80**
	SSIM	0.86	**0.91**	0.89	**0.91**
	run-time(s)	0.94	22.64	2.50	20.37
Human	MI	6.09	8.48	8.84	**9.29**
	$Q^{AB/F}$	0.57	0.77	0.73	**0.81**
	SSIM	0.88	0.89	0.86	**0.90**
	run-time(s)	1.67	62.49	4.72	45.26

5. CONCLUDING REMARKS

A novel image fusion scheme has been presented to combine multiple images acquired with different focus points. The RPCA technique is used to decompose the source images into principal and sparse matrices. The salient information from multifocus images can be discovered via sparse features computed based on the sparse matrix. The experiments showed that the RPCA-based approach yielded consistently superior fusion results compared to a number of state-of-the-art fusion methods in terms of both subjective and objective evaluations. Future work will involve extending the developed method to be applied to the noisy images.

6. REFERENCES

[1] C. Ludusan, O. Lavialle, Multifocus Image Fusion and Denoising: A Variational Approach, *Pattern Recognition Letters,* 33, 1388-96, 2012.

[2] S. Li, B. Yang, Multifocus Image Fusion Using Region Segmentation and Spatial Frequency, *Image and Vision Computing,* 26, 971-79, 2008.

[3] S. Li, J. Kwok, Y. Wang, Combination of Images with Diverse Focuses Using the Spatial Frequency, *Information Fusion,* 2, 169-76, 2001.

[4] T. Wan, N. Canagarajah, A. Achim, A Novel Region-based Image Fusion Framework Using Alpha-Stable Distributions in the Complex Wavelet Domain, *IEEE Tran. on Multimedia,* 11(4), 624-33, 2009.

[5] S. Li, B. Yang, J. Hu, Performance Comparision of Different Multi-resolution Transforms for Image Fusion, *Information Fusion,* 12, 74-84, 2011.

[6] B. Yang, S. Li, Multifocus Image Fusion and Restoration with Sparse Representation, *IEEE Tran. on Instrumentation and Measurement,* 59, 884-892, 2010.

[7] B. Yang, S. Li, Pixel-level Image Fusion with Simultaneous Orthogonal Matching Pursuit, *Information Fusion,* 13, 10-19, 2012.

[8] E. Candès, X.Li, Y. Ma, J.Wright, Robust Principal Component Analysis?, availabe at: *http://arxiv.org/abs/0912.3599v1,* 2009.

[9] The Online Resource for Research in Image Fusion, available at *http://www.imagefusion.org/*, 2009.

[10] J. Tian, L. Chen, Adaptive Multi-focus Image Fusion Using a Wavelet-based Statistical Sharpness Measure, *Signal Processing,* 92, 2137-46, 2012.

[11] D. MacKay, Information Theory, Inference, and Learning Algorithms, *Cambridge University Press,* 2003.

[12] C. S.Xydeas, V. Petrovic, Objective Image Fusion Performance Measure, *Electronics Letters,* 36, 308-09, 2000.

[13] M. Gaubatz, MeTriX MuX Visual Quality Assessment Package, available at *http://foulard.ece.cornell.edu/gaubatz/metrix_mux/*, 2011.