Di Guo, Jingwen Yan, Xiaobo Qu. High quality multi-focus image fusion using self-similarity and depth information, Optics Communications, DOI:10.1016/j.optcom.2014.10.031, 2014.

High quality multi-focus image fusion using self-similarity and depth information

Di Guo[1], Jingwen Yan[2], Xiaobo Qu[3*]

1. School of Computer and Information Engineering, Fujian Provincial University Key Laboratory of Internet of Things Application Technology, Xiamen University of Technology, Xiamen 361024, China
2. Guangdong Provincial Key Laboratory of Digital Signal and Image Processing Techniques, Department of Electronics Engineering, Shantou University, Shantou 515063, China
3. Department of Electronic Science, Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance Research, Xiamen University, Xiamen 361005, China

Submitted to ***Optics Communications***

***Corresponding author:**
Xiaobo Qu,
Department of Electronic Science
P. O. Box 979, Xiamen University, Xiamen 361005, China
quxiaobo@xmu.edu.cn

Abstract

Due to the limited depth of field in a camera, some imaging objects will be blurred if they are located far from the focus plane and the other objects on the plane will be clear. Multi-focus image fusion synthesizes a sharp image from multiple partially focused images. However, traditional fused images usually suffer from blurring effects and pixel distortions. In this paper, we explore two unique characteristics of multi-focus images: 1) The self-similarity of a single image and the shared similarity among multiple source images; 2) The distances from object to focal plane. The former characteristic is used to identify image structure-driven regions while the latter refine the image clarity by automatically estimating depth information of blurred images. Experimental results demonstrate that the proposed method outperforms the state-of-the-art fusion methods on image quality and objective fusion criteria.

Keywords: Image fusion, self-similarity, adaptive, depth

1. Introduction

Focusing is important for acquiring a clear image in photography. However, images will be blurred if an object lies far from the focal plane. Therefore, it will be hard to acquire a clear image for a camera with limited depth length. Multi-focus image fusion technology can provide a clearer and more reliable image by combining the information from multiple images focusing on different objects of the same or similar scene [1-3]. It has been applied in non-diffracting imaging system [4], mobile microscope processing software [5] and fluorescence intraoperative surgery [6] in biophotonics.

Existing fusion methods can be roughly divided into two types, which perform the fusion in the spatial domain [7-11] or multi-scale decomposition (MSD) domains [1, 12], using sparse transforms, e.g. wavelets [1, 13, 14], bandelets [15], contourlets [16, 17], shearlets [18], surfacelets [19], trained dictionaries [3, 20, 21]. Although many MSD methods produce nice images, they lead to pixel distortion (Figs. 1(c) and (e)) due to nonlinear operations in the MSD domain. Making use of the coefficients statics in MSD domain helps to suppress this limitation but needs to train parameters in statics models comprehensively [22]. On the contrary, much less distortions are introduced if linear fusion rules in the spatial domain, e.g. the maximum rule, directly choose the pixel values in a well-focused region (Figs. 1(d) and (f)). But the shape of a region will seriously affect fused images. For example, the commonly used isotropic square regions in the spatial domain will easily lead to blocky artifacts in blurred areas around the edge (Figs. 1(d) and (f)). How to choose image pixels in an adaptive region with high clarity remains open.

In this paper, a new multi-focus image fusion method in the spatial domain is proposed. A data-driven scheme, based on the shared similarity of source images, is proposed to generate adaptive regions. Moreover, the distance from an object to the focal plane is automatically estimated. This prior information can hopefully improve the fusion performance since the distance will seriously affect clarity of images. Our contributions are summarized as follows:

1) Adaptive regions based on shared similarity of source images are generated to measure the

clarity, thus will potentially avoid the blurring on one object and make more reliable decisions for fusion.

2) A weighting fusion rule based on adaptive regions voting is proposed to further suppress the blurring artifacts.

3) The distances between multiple objects and the focal plane are incorporated in adaptive regions to measure the clarity.

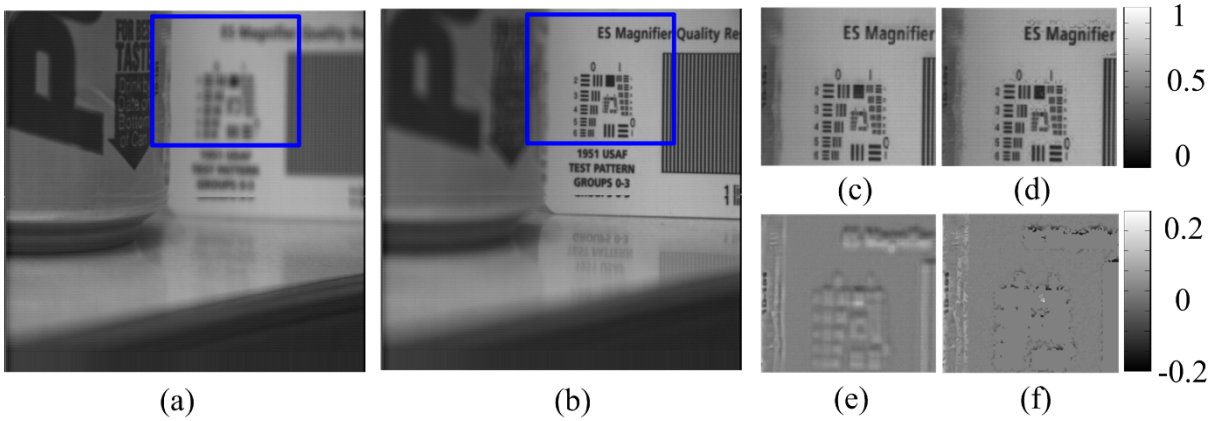

Fig. 1. Artifacts produced by the wavelet and the traditional spatial method. (a) and (b) are source images, (c) and (d) are cropped parts of fused images using wavelet and typical spatial method with isotropic regions, (e) and (f) are fusion errors compared with the sharp image regions in (b). Note: Some-of-Modified-Laplacian (SML) is applied in the spatial method.

2. Method

The outline of the proposed method is illustrated in Fig. 2. First, adaptive regions are generated using image similarity information. Next, the depth information is estimated, and combined with the image gradient to measure the clarity in adaptive regions. Finally, a pixel is fused based on the weightings from source images.

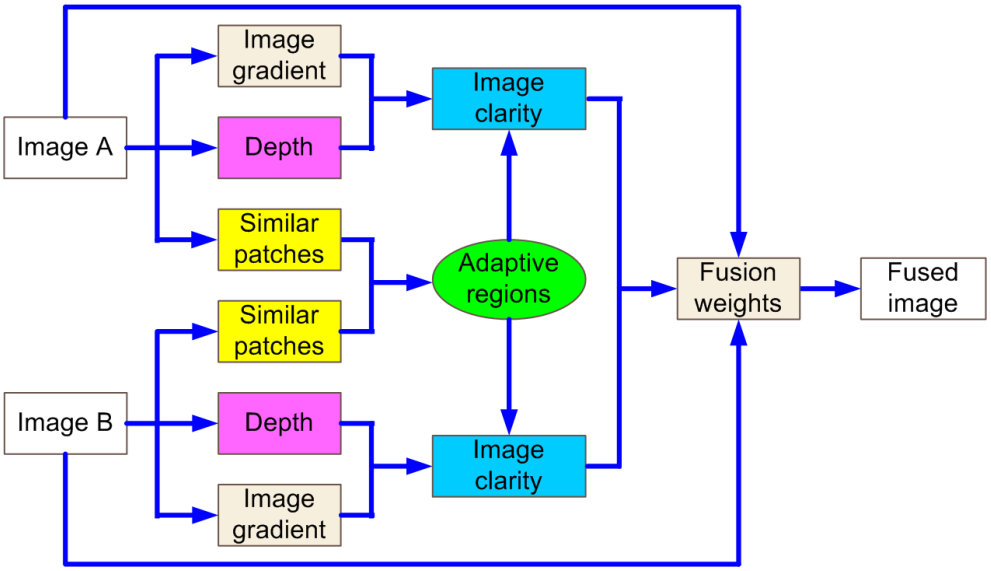

Fig.2. Flowchart of the proposed method.

2.1. Shared similarity regions

The adaptive region is generated using the shared similarity of source images. Images are first divided into multiple overlapped square patches, and similar patches are searched [23, 24].

Here, a shared self-similarity of source images is defined as follows to generate an adaptive region of pixels for fusion. Given a reference patch $P_r \in \mathbb{R}^{m\times m}$ and a region $R(r) \in \mathbb{R}^{n\times n}$ centered at pixel r, the similarity of any candidate patch $P_q \in \mathbb{R}^{m\times m}$ to the P_r is defined as

$$\eta_q = \left\| P_q - P_r \right\|_F, \tag{1}$$

where $\|A\|_F = \sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\left|a_{ij}\right|^2}$ denotes the Frobenius norm of matrix A. By sorting the η_q by the descending order for all the patches in this region, the most k similar patches to P_r are found and the collection of this patches are expressed as $L_R(r) = \left\{P_{q_1}, P_{q_2}, \cdots, P_{q_k}\right\}$. Similar patches shared by both L_{R^A} and L_{R^B} are

$$L_R^S(r) = L_{R^A}(r) \cap L_{R^B}(r) \tag{2}$$

where $L_R^S(r)$ is one of adaptive regions for fusion, $R^A(r)$ and $R^B(r)$ denote the same region of source images f^A and f^B, respectively. The locations and the number of similar patches of each adaptive region vary with the shared similarity of source images. For example, the adaptive region is composed of 4 similar patches as shown in Fig. 3.

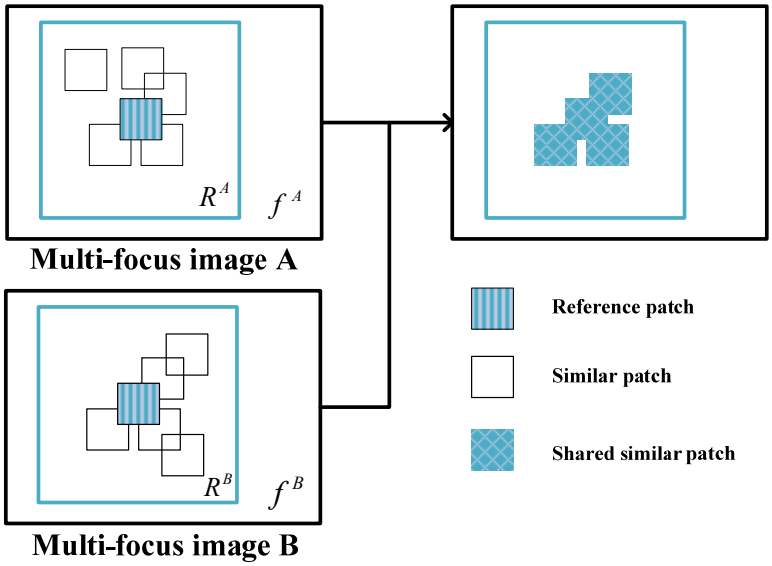

Fig. 3. An adaptive region of shared similar patches

How to make use of these adaptive regions in image fusion? Traditionally, a pixel of the fused image is chosen from a source image with higher clarity, which is called the maximum rule [1, 7], as follows:

$$f^F(r)=\begin{cases} f^A(r) & if \quad C^A(r)>C^B(r) \\ f^B(r) & \text{otherwise} \end{cases} \tag{3}$$

where C denotes the clarity metric and f^F is the fused image. In order to overcome the blocky artifacts in fusion, instead of selecting the pixel with a larger value, $C(r)$ is usually computed as a sum of pixel-wise charity measure, e.g. gradients like Some-of-Modified-Laplacian (SML) in a fixed size of region centered at r [7, 17]. SML at $r=(x_0,y_0)$ is defined as

$$SML(x_0,y_0)=\sum_{x=x_0-N}^{y=x_0+N}\sum_{x=y_0-N}^{y=y_0+N}\nabla_{ML}^2 f(x_0,y_0) \tag{4}$$

where $\nabla^2_{ML} f(x_0, y_0)$ is the Modified Laplacian (ML), whose expression is

$$\begin{aligned}\nabla^2_{ML} f(x_0, y_0) = &\left|2f(x_0, y_0) - f(x_0 - 1, y_0) - f(x_0 + 1, y_0)\right| \\ &+ \left|2f(x_0, y_0) - f(x_0, y_0 - 1) - f(x_0, y_0 + 1)\right|\end{aligned}. \tag{5}$$

In our approach, clarity is measured in each adaptive region. If the clarity in an adaptive region $L^s_R(r)$ of f^A is larger than that of f^B, one vote will be assigned to all the pixels $j \in L^s_R(r)$ in this adaptive region of f^A, meaning

$$v^A(j) = v^A(j) + 1, \quad j \in L^s_R(r). \tag{6}$$

where $v^A(j)$ denotes the counter at the spatial location j. The initial values of $v^A(j)$ and $v^B(j)$ are 0 and voting stops until all the adaptive regions are compared. Finally, one can have counter maps v^A and v^B that are in the same size as source images. In the following, $v^A(r)$ and $v^B(r)$ are used to denote the final vote for pixels at location r.

Figs. 4(e) and 4(f) show the improvement using the proposed adaptive region over the fixed size region in traditional methods. Obviously, more accurate pixels are selected from the clear images using the new approach. This is mainly because more homogeneous pixels are collected to sum pixel-wise charity measure within the proposed data-driven adaptive region.

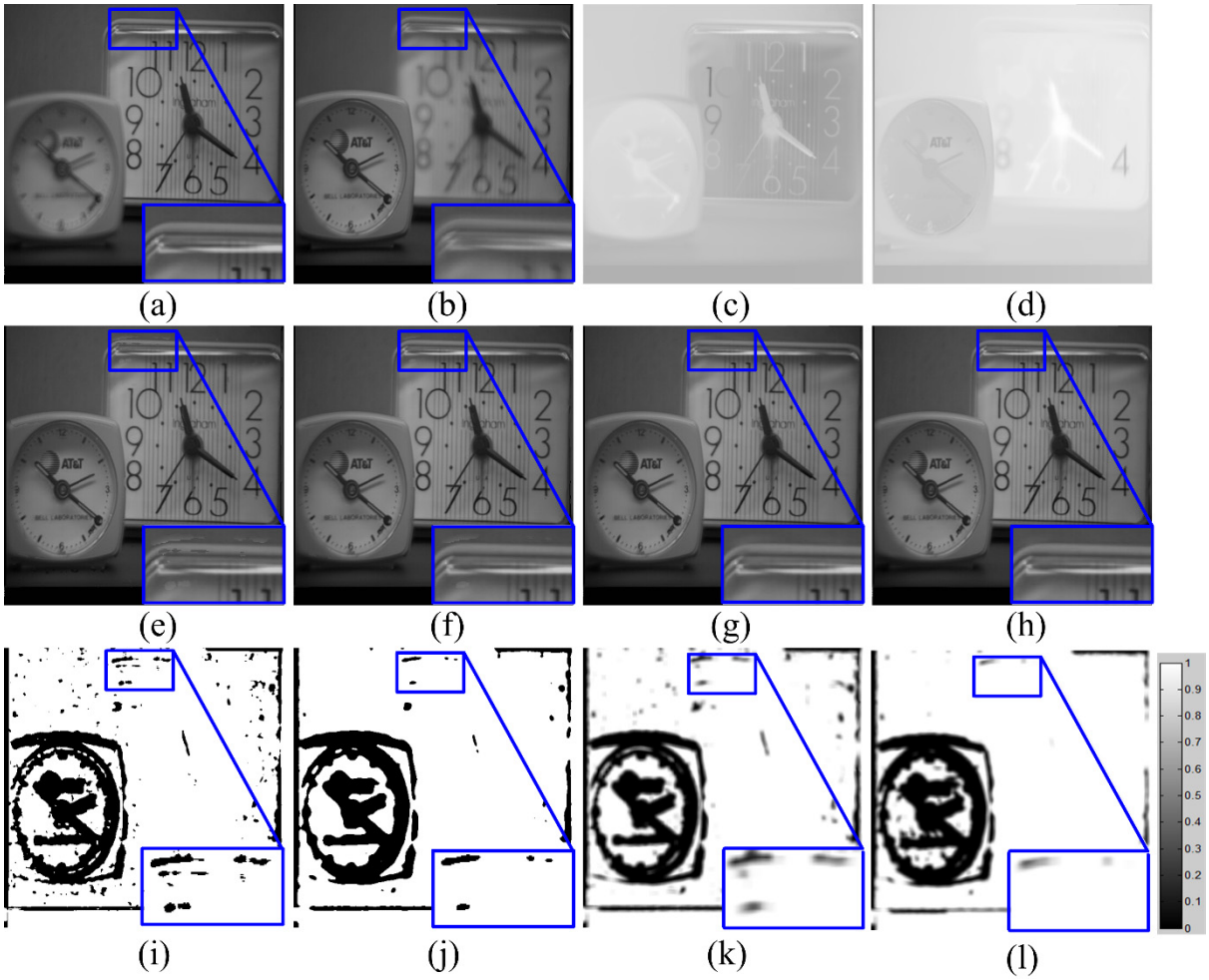

Fig. 4. Illustration of the proposed method. (a) and (b) are multi-focus source images, (c) and (d) are estimated distances to focal plane of (a) and (b) where darker gray value implies closer distance, (e) and (f) are fused images using fixed and the proposed adaptive regions with a maximum rule, (g) and (h) are the fused image using adaptive regions with the weighting rule but without and with depth information, (i)-(l) are the weight for source image (b) used in the fusion of (e)-(h), respectively.

However, one may observe some blurred pixels around edges in Figs. 4(f). One possible solution is to use weightings to compose the pixel according to

$$f^F(r) = \frac{v^A(r)}{v^A(r)+v^B(r)} f^A(r) + \frac{v^B(r)}{v^A(r)+v^B(r)} f^B(r). \tag{7}$$

Since $v^A(r)$ or $v^B(r)$ are usually nonzero, the wrongly selected pixels will be smoothened. For example, a binary decision is made by the traditional maximum rule as shown in Fig. 4(j), where a white (or black) pixel indicates choosing a pixel from Fig. 4(a) (or Fig. 4(b)). But the proposed rule

generates a weighting between 0 and 1 as shown in Fig. 4(k) to extract gray values from both Figs. 4(a) and (b). Therefore, this weighting fusion rule can inhibit the defect on image edges comparing Figs. 4(f) and 4(g), although some pixels are still wrongly selected around edges in Fig. 4(g). This problem will be further resolved using the following estimated depth information.

2.2. Incorporate the distance from the object to the focal plane

Here explains how to incorporate the depth information to remove the blurring around edges.

Let $d(r)$ be the distance between the focal plane and the object whose location corresponds to the image pixel r. When an object is placed at the focal plane d_F, all the rays from a point of the object will converge to a single sensor point and the image will appear sharp. Rays from a point of another object at $d(r)$ will reach multiple sensor points and result in a blurred image. Let $s(r)$ denote the diameter of the circle of confusion, and it can be written as

$$s(r) = \frac{|d(r) - d_F|}{d(r)} \frac{F_0^2}{N(d_F - F_0)} \tag{8}$$

where F_0 and N are the focal length and the stop number of the camera, respectively.

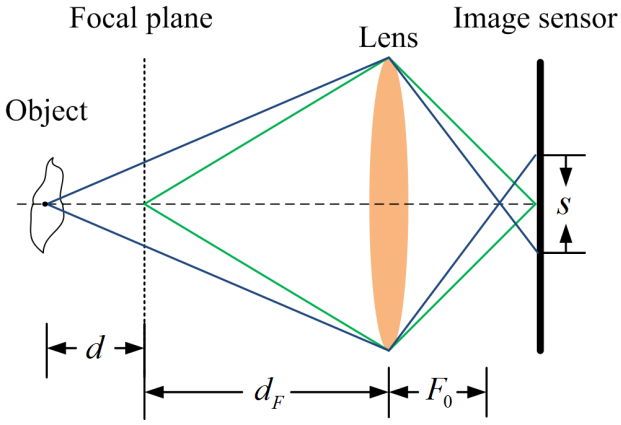

Fig. 5. Focus and defocus for thin lens model.

In this paper, a linear response camera $d(r) \propto s(r)$ is assumed to automatically estimate the depth using a defocus map estimation method proposed in [25]. The whole estimation process is shown in Fig. 6 and a brief review will be given below.

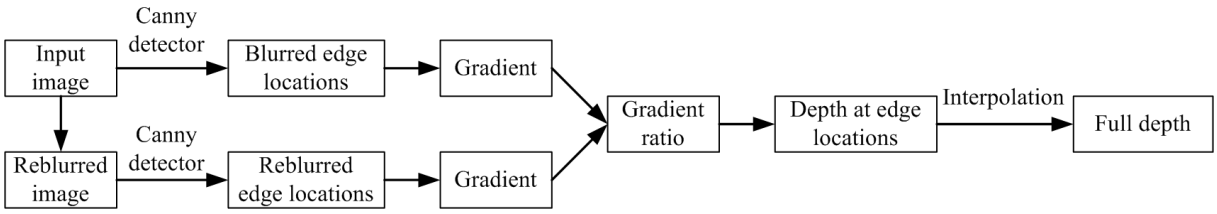

Fig. 6. The process of depth estimation.

First, the $d(r)$ at edge locations $r \in L$ is calculated by

$$d(r) \propto \frac{1}{\sqrt{\left[R(r)\right]^2 - 1}} \tag{9}$$

where $R(r)$ is a ratio between the gradients of input out-of-focus image and its re-blurred image [25]. Then, the depth estimation is propagated the from edge locations to the entire image, and a full depth map is obtained finally [25]. In this paper, the maximum $d(r)$ is normalized to be 1 for all images.

If the distance from an object to the focal plane is close, the pixel belongs to well-focused region as shown in Figs. 4(c) and (d). Thus, the depth information can be used as a refinement for the clarity metric on images

$$C(r) = M(r) \times \left[d(r)\right]^{-2} \tag{10}$$

where $M(r)$ is the SML, and $d(r)$ is the distance between the object at pixel *r* and the focal plane. By incorporating this depth information into fusion, more accurate weightings are performed on source images as shown in Fig. 4(l). Therefore, the blurring artifacts are significantly removed as shown in Fig. 4(h).

3. Results

In simulations, the proposed method is compared with the typical image fusion methods using the pulse coupled neural networks (PCNN) [16], sharpness statistics (SS) [26], and the state-of-the-arts guided filtering (GF) [27]. The code of the three methods can be found at

http://www.quxiaobo.org, http://www.mathworks.com/matlabcentral/fileexchange/downloads/16579 and http://xudongkang.weebly.com. Default parameters in the shared source codes are used. For the proposed method, the search region size is 16×16 and the number of similar patches is 16 for all images.

As shown in Fig. 7, PCNN leads to obvious blurred edges. SS generates very sharp edge but introduces artifacts around boundaries. GF produces nice results but still encounters blurring around boundaries. The proposed approach achieves minimal artifacts around the object boundaries and produces the sharpest images.

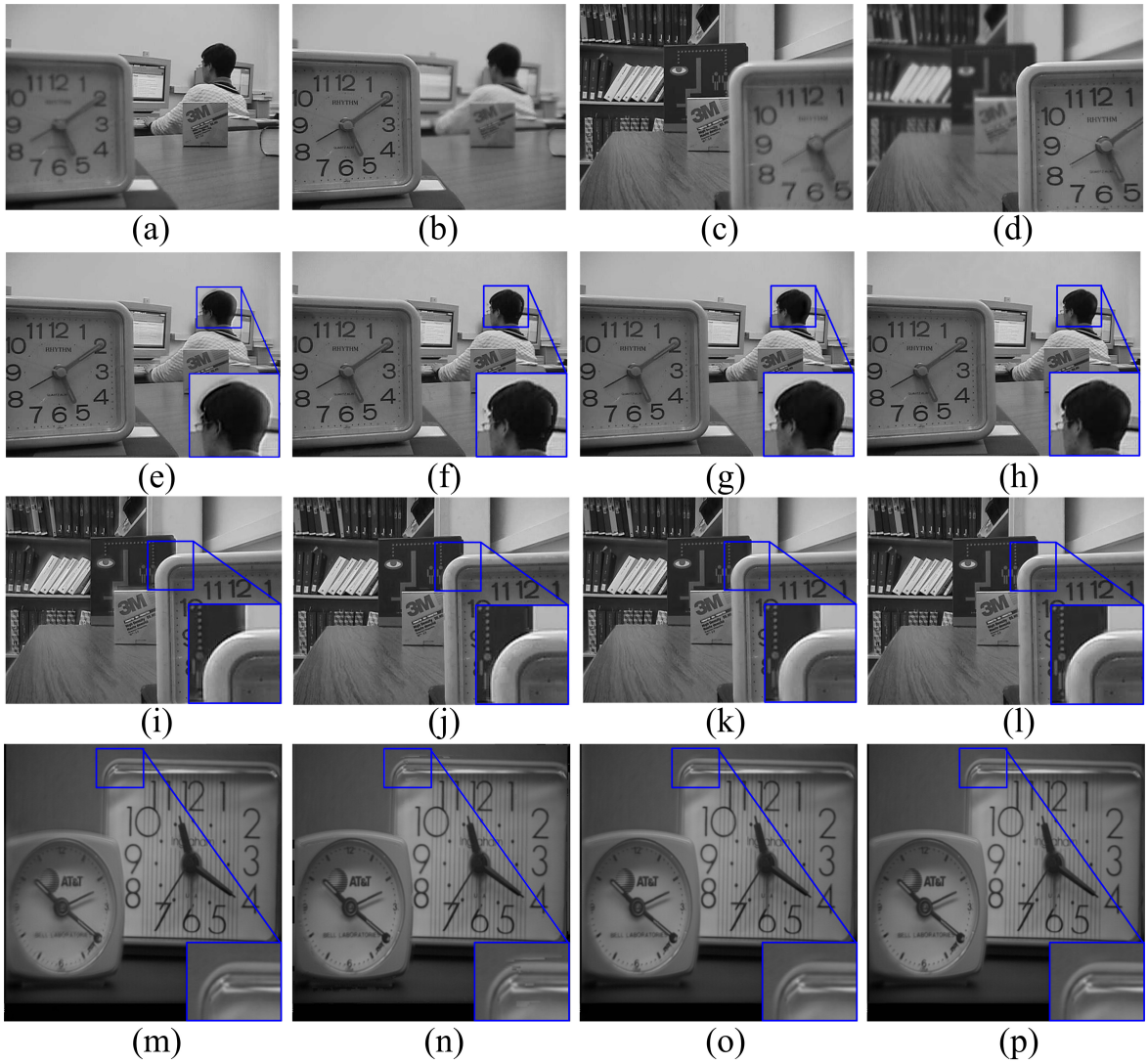

Fig. 7. Fused images using different methods. (a)-(d) are source images Disk and Lab, (e)-(h), (i)-(l) and (m)-(p) are fused images using pulsed coupled neural networks, sharpness statics, guided filter and the proposed method, respectively.

Two fusion criteria, mutual information (MI) [13] and $Q^{AB/F}$ [14] are used to evaluate the fusion performance [16, 26, 27]. MI measures the information transferred from source images to the fused image, and $Q^{AB/F}$ indicates how much edge information is preserved in the fused image. Evaluation criteria listed in Table 1 demonstrates that the proposed method provides the most consistent fused images to the source images and preserves the image edges best.

Table 1. Objective fusion criteria for different methods

Images	**PCNN**		**SS**		**GF**		**Proposed**	
	MI	$Q^{AB/F}$	MI	$Q^{AB/F}$	MI	$Q^{AB/F}$	MI	$Q^{AB/F}$
Disk	6.15	0.656	7.55	0.715	7.06	0.724	**7.81**	**0.733**
Lab	7.59	0.700	8.04	0.740	7.91	0.751	**8.38**	**0.757**
Clock	7.47	0.681	8.03	0.693	7.89	0.716	**8.57**	**0.724**

4. Discussions

4.1 Tests on synthetic images

Three popular grayscale images are adopted as ground truth images. Different regions of the ground truth images are then blurred to synthetize the testing images (Fig. 8). The measured peak signal-to-noise-ratio (PSNR), shown in Table 2, implies that the proposed method consistently outperforms the compared methods.

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

Fig. 8. Synthetic images for fusion. (a) and (b), (c) and (d), (e) and (f) are generated on the commonly used images *Barbara*, *Boat*, and *Mandrill*, respectively.

Table 2. PSNR performance on the synthetic images (unit: dB).

Images	PCNN	SS	GF	Proposed
Barbara	33.4	42.1	46.8	**50.6**
Boat	32.7	36.6	46.6	**53.0**
Mandrill	33.3	43.9	46.7	**51.9**

4.2 Parameters of the proposed method

Parameters of the proposed method include the region size, the number of similar patches and the patch size. We analyze the effect of these parameters on mutual information (MI) in Fig. 9. Another evaluation criterion, $Q^{AB/F}$, is not shown here because of its consistent performance to that of MI. Default values of the search region size, the number of similar patches and the patch size are 16, 16, and 8, respectively. When one parameter is discussed, the other two parameters are fixed with default values.

The MI performance is stable for different region sizes and numbers of similar patches. However, the computational time goes up as the two parameters increase. Therefore, region size is set as 16 and the number of similar patches is set as 16. The patch size will affect the MI performance a little more. The MI performance always increases as the patch size becomes larger. However, the computational time increases significantly when the patch size is increased to 14. To maintain a stable computational time, the patch size is suggested to be half of the region size.

Default parameters of the depth estimation are used and one can download the code at http://sjzhuo.net/defocusEstimation/index.html.

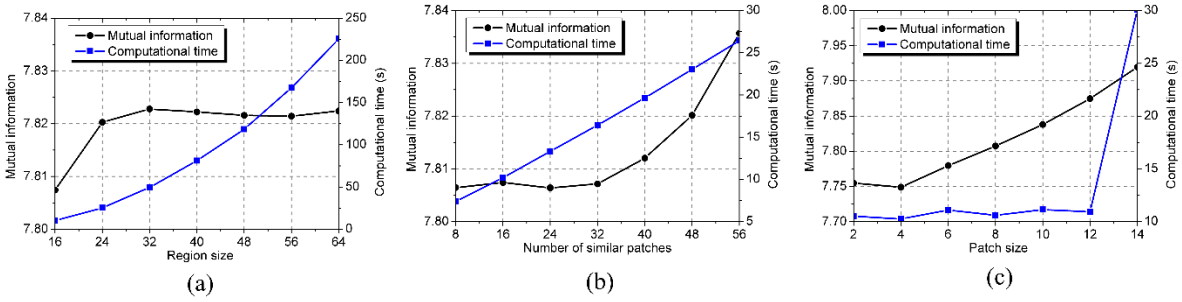

Fig. 9. The effect of the parameters in the proposed method. (a)-(c) are curves for the region size, the number of similar patches and the patch size. Experiments are conducted on fusing *Disk* images and computational time of depth estimation is not included.

4.3 Computational time

The computational time for all methods is reported in Table 3. PCNN is the slowest and GF is the fastest. The proposed method is faster than SS. In the simulation, 80% computation time of the proposed method comes from estimating the depth information. Besides, the computation spent on patch-based processing of the proposed method could be accelerated with parallel computing [28].

Table 3. Computational time for different methods (unit: second)

Images	PCNN	SS	GF	Proposed
Disk (640×480)	489.63	153.34	0.63	66.90
Lab (640×480)	495.13	163.75	0.66	70.53
Clock (512×512)	419.35	117.00	0.53	59.13

Note: Experiments run on a laptop computer with 2 Cores 2.6 GHz CPU and 12 GB RAM.

5. Conclusion

In this paper, the image self-similarity and the distance from the object to the focal plane are

explored for multi-focus image fusion in the spatial domain. The former leads to a data-driven region in arbitrary shapes, while the latter significantly remove blurring artifacts around object boundaries. Experiments demonstrate that the proposed approach outperforms the state-of-the-art fusion methods on both the image quality and objective fusion criteria. In the future, a faster depth estimation method and a parallel computing of the proposed method are expected to speed up the fusion process.

Acknowledgements

The authors are grateful to Dr. Shaojie Zhuo for sharing the depth estimation code, Drs. Jing Tian, Xudong Kang, Jianwen Hu for sharing the image fusion codes and Dr. Rick Blum for sharing multi-focus source images used in this paper. This work was partially supported by the NNSF of China (61302174, 61201045 and 40971206), Fundamental Research Funds for the Central Universities (2013SH002), Scientific Research Foundation for the Introduction of Talent at Xiamen University of Technology (YKJ12021R), Open Fund from Key Lab of Digital Signal and Image Processing of Guangdong Province (2013GDDSIPL-07) and Academic Innovation Team Building Project of Shantou University (No. ITC12002).

Reference

[1] Z. Zhong, R. S. Blum, Proceedings of the IEEE 87 (8) (1999) 1315.
[2] L. Guo, M. Dai, M. Zhu, Optics Express 20 (17) (2012) 18846.
[3] L. Chen, J. Li, C. L. P. Chen, Optics Express 21 (4) (2013) 5182.
[4] Z. Zhai, P. Zhang, Q. Lv, X. Wang, X. Zhong, Optics Communications 295 (15) (2013) 74.
[5] Z. J. Smith, K. Chu, A. R. Espenson, M. Rahimzadeh, A. Gryshuk, M. Molinaro, D. M. Dwyre, S. Lane, D. Matthews, S. Wachsmann-Hogiu, PLoS ONE 6 (3) (2011) e17150.
[6] P. F. Feruglio, C. Vinegoni, L. Fexon, G. Thurber, A. Sbarbati, R. Weissleder, Journal of Biophotonics 6 (4) (2013) 363.
[7] W. Huang, Z. Jing, Pattern Recognition Letters 28 (4) (2007) 493.
[8] V. Aslantas, A. N. Toprak, Optics Communications 332 (2014) 350.
[9] V. Aslantas, R. Kurban, Optics Communications 282 (16) (2009) 3231.
[10] J. Tian, L. Chen, L. Ma, W. Yu, Optics Communications 284 (1) (2011) 80.
[11] K.-L. Hua, H.-C. Wang, A. H. Rusdi, S.-Y. Jiang, Journal of Visual Communication and Image Representation, 25 (5) (2014) 951.
[12] H. Zhao, Z. Shang, Y. Y. Tang, B. Fang, Pattern Recognition, 46 (3) (2013) 1002.
[13] Y. Chai, H. Li, Z. Li, Optics Communications 284 (19) (2011) 4376.

[14] Y. Chai, H. F. Li, J. F. Qu, Optics Communications, 283 (19) (2010) 3591.
[15] X. Qu, J. Yan, G. Xie, Z. Zhu, B. Chen, Chinese Optics Letters, 5 (10) (2007) 569.
[16] X. Qu, J. Yan, H. Xiao, Z. Zhu, Acta Automatica Sinica, 34 (12) (2008) 1508.
[17] X. Qu, J. Yan, G. Yang, Optics and Precision Engineering, 17 (5) (2009) 1203.
[18] Q. G. Miao, C. Shi, P. F. Xu, M. Yang, Y. B. Shi, 284 (6) (2011) 1540.
[19] Q. Zhang, L. Wang, Z. Ma, H. Li, 285 (13-14) (2012) 3032.
[20] S. Gao, Y. Cheng, Y. Zhao, Optics Letters, 38 (11) (2013) 1981.
[21] B. Yang, S. Li, IEEE Transactions on Instrumentation and Measurement, 59 (4) (2010) 884.
[22] W. Wu, X. Yang, Y. Pang, J. Peng, G. Jeon, Optics Communications, 287 (2013) 63.
[23] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, IEEE Transactions on Image Processing, 16 (8) (2007) 2080.
[24] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, Z. Chen, Medical Image Analysis, 18 (6) (2014) 843.
[25] S. Zhuo, T. Sim, Pattern Recognition, 44 (9) (2011) 1852.
[26] J. Tian, L. Chen, Signal Processing, 92 (9) (2012) 2137.
[27] S. Li, X. Kang, J. Hu, IEEE Transactions on Image Processing, 22 (7) (2013) 2864.
[28] Q. Li, X. Qu, Y. Liu, D. Guo, J. Ye, Z. Zhan, Z. Chen, Computational and Mathematical Methods in Medicine, 2014 (2014) Article ID 257435.