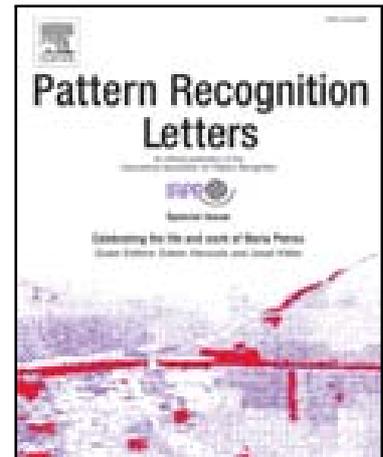


## Journal Pre-proof

Real-MFF: A large realistic multi-focus image dataset with ground truth

Juncheng Zhang, Qingmin Liao, Shaojun Liu, Haoyu Ma, Wenming Yang, Jing-Hao Xue

PII: S0167-8655(20)30300-7  
DOI: <https://doi.org/10.1016/j.patrec.2020.08.002>  
Reference: PATREC 7993



To appear in: *Pattern Recognition Letters*

Received date: 16 January 2020  
Revised date: 2 July 2020  
Accepted date: 2 August 2020

Please cite this article as: Juncheng Zhang, Qingmin Liao, Shaojun Liu, Haoyu Ma, Wenming Yang, Jing-Hao Xue, Real-MFF: A large realistic multi-focus image dataset with ground truth, *Pattern Recognition Letters* (2020), doi: <https://doi.org/10.1016/j.patrec.2020.08.002>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Research Highlights (Required)**

To create your highlights, please type the highlights against each `\item` command.

It should be short collection of bullet points that convey the core findings of the article. It should include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point.)

- A novel large and realistic multi-focus dataset with ground truth is constructed.
- The proposed dataset can serve as a test bench for multi-focus image fusion methods.
- Benefit future development of deep-learning-based methods for multi-focus image fusion.



Pattern Recognition Letters  
journal homepage: [www.elsevier.com](http://www.elsevier.com)

## Real-MFF: A large realistic multi-focus image dataset with ground truth

Juncheng Zhang<sup>a</sup>, Qingmin Liao<sup>a</sup>, Shaojun Liu<sup>b,\*\*</sup>, Haoyu Ma<sup>a</sup>, Wenming Yang<sup>a</sup>, Jing-Hao Xue<sup>c</sup>

<sup>a</sup>Shenzhen International Graduate School, Tsinghua University, China

<sup>b</sup>Department of Electric and Computer Engineering, Hong Kong University of Science and Technology, China

<sup>c</sup>Department of Statistical Science, University College London, U.K.

Article history:

### ABSTRACT

**Keywords:** Image fusion, multi-focus images, multi-focus dataset, deep learning

Multi-focus image fusion, a technique to generate an all-in-focus image from two or more partially-focused source images, can benefit many computer vision tasks. However, currently there is no large and realistic dataset to perform convincing evaluation and comparison of algorithms in multi-focus image fusion. Moreover, it is difficult to train a deep neural network for multi-focus image fusion without a suitable dataset. In this letter, we introduce a large and realistic multi-focus dataset called Real-MFF, which contains 710 pairs of source images with corresponding ground truth images. The dataset is generated by light field images, and both the source images and the ground truth images are realistic. To serve as both a well-established benchmark for existing multi-focus image fusion algorithms and an appropriate training dataset for future development of deep-learning-based methods, the dataset contains a variety of scenes, including buildings, plants, humans, shopping malls, squares and so on. We also evaluate 10 typical multi-focus algorithms on this dataset for the purpose of illustration.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

For most computer vision tasks, such as object detection and identification, it is desirable to use the in-focus images as input rather than blurred ones. However, due to the limited depth-of-field (DOF) of cameras, it is usually difficult to capture the all-in-focus images directly. Therefore, multi-focus image fusion, a technique to fuse two or more partially-focused source images into an all-in-focus image, is very important in the fields of computer vision and image processing, and has drawn considerable attention in recent years.

Existing multi-focus image fusion methods can be roughly categorized into three groups: transform-domain-based methods, spatial-domain-based methods, and deep-learning-based methods.

Transform-domain-based methods usually first decompose the source images in a transform domain, then fuse the features in the transform domain, and finally reconstruct the all-in-focus

image. Laplacian pyramid (LP) [5], ratio of low-pass pyramid (RP) [6], curvelet transform (CVT) [7], discrete wavelet transform (DWT) [8], dual-tree complex wavelet transform (DTCWT) [9], non-subsampled contourlet transform (NSCT) [10], principal component analysis (CPA) [11], and sparse representation-based methods [12–14] have been explored to build transform-domain-based methods.

Spatial-domain-based methods can be further classified into three sub-groups: block-based methods, region-based methods and pixel-based methods. The block-based methods [15] first divide images into blocks, then calculate the focus measure of each block, and finally choose the block with the highest focus measure as the corresponding block in the fusion result. Consequently, these algorithms are often affected by the granularity of block partitioning. The region-based methods [16] first segment the input images and then fuse the focused segments of these input images. Therefore, the fusion results of region-based methods highly rely on the segmentation accuracy. The pixel-based methods [17] calculate the focus measure and fuse the images at the pixel level. Usually, pixel-based methods often produce poor results near the boundary between a focused area and a defocused area.

\*\*Corresponding author.

e-mail: [liusj14@tsinghua.org.cn](mailto:liusj14@tsinghua.org.cn) (Shaojun Liu)

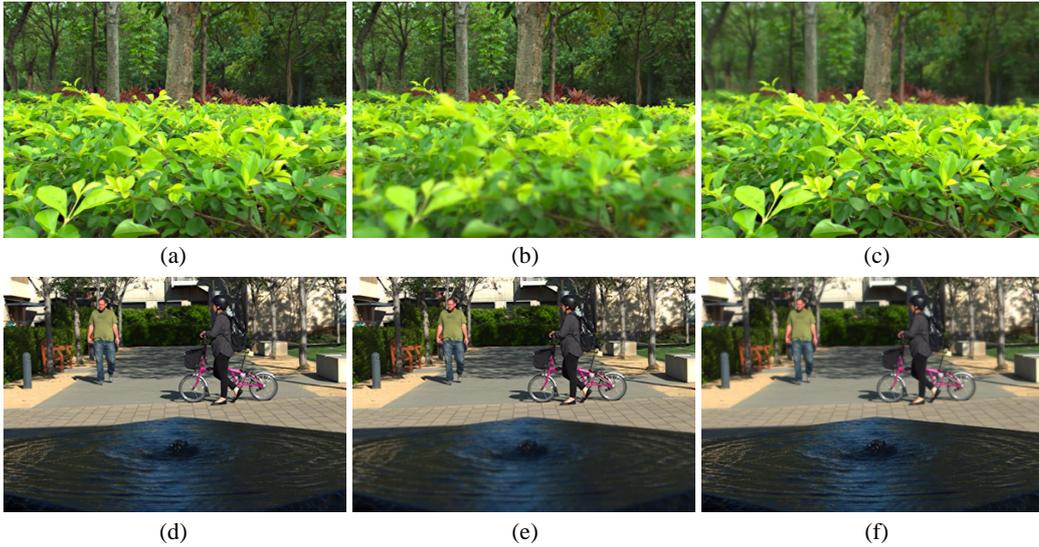


Fig. 1: Examples in our dataset. (a) and (d) are the source images focused on the foreground; (b) and (e) are the source images focused on the background; and (c) and (f) are the all-in-focus images. The upper row is for a complex example and the lower row is for a simple example.

Table 1: Characteristics of various datasets.

Dataset	Data generation method	size	realistic	ground truth
Lytro [1]	Captured by light field camera	20 pairs, 520×520	Yes	No
CNN [2]	Synthetically generated based on the ImageNet dataset	1,000,000 pairs, 16×16	No	Yes
BA-Fusion [3]	Synthetically generated based on the Matting dataset	2,268,000 pairs, 16×16	No	Yes
FuseGan [4]	Synthetically generated based on segmentation datasets	5,850 pairs, 320×480	No	Yes
Our Real-MFF	Captured by light field camera	710 pairs, 625×433	Yes	Yes

In the past several years, many deep learning methods have been proposed for multi-focus image fusion. Liu *et al.* [2] used a deep convolutional neural network (CNN) to generate a decision map just like the pixel-based methods, and then did some post-processing to produce a final decision map. FuseGan [4] used a generative adversarial network (GAN) to generate decision map. Such methods can be regarded as the network-based implementations of the spatial-domain-based methods. Different from them, Wen *et al.* [18] designed an end-to-end neural network for image fusion.

One of the bottlenecks of using deep neural networks to solve multi-focus fusion problems is the lack of suitable large and realistic database with ground truth for the network training. The most widely used dataset in this area is the Lytro Multi-focus dataset [1], which contains 20 pairs of multi-focus images with size 520×520 pixels. This dataset is very small and has no ground truth, therefore, it cannot be used for training a deep neural network. To break through this bottleneck, there are some valuable trials. Liu *et al.* [2] used high-quality natural images blurred by Gaussian filters with five different levels of blur to generate a dataset. The images generated by this method were not as real as naturally defocused images, because they were blurred with a spatially invariant defocus kernel and therefore lacked defocus changes. [4, 18] used segmentation datasets with manually labeled segmentation as the ground truth map. The ground truth was used as a 0-1 mask map, with 0 for foreground and 1 for background. The foreground and back-

ground were then blurred by Gaussian filters separately and finally merged together. In our previous work of BA-Fusion [3], we generated a dataset based on a matting dataset. We chose matted object as the foreground object and high-quality picture as the background. However, the images generated by these methods do not follow the real defocus model and thus need further improvement.

In this letter, we propose a new large and realistic dataset for multi-focus image fusion. The dataset, called Real-MFF, consists of various natural multi-focus images with ground truth, generated by light field images. Fig. 1 shows two pairs of partially-focused source images and their all-in-focus ground-truth images in our dataset for examples. The contributions of our work can be summarized as follows.

Firstly, we construct a new large and realistic multi-focus dataset, which contains 710 pairs of images that can be used for training deep neural networks. Each pair of images contains two partially-focused images as the source images and an all-in-focus image as the ground truth. The dataset is generated using a light field camera: Lytro illum camera. The source images are produced by choosing different focus planes.

Secondly, to the best of our knowledge, our dataset is the first large and realistic dataset that can serve as a test bench for validating multi-focus image fusion methods.

Finally, the proposed dataset is both large and realistic so that it can benefit the development of deep-learning-based multi-focus image fusion methods.

## 2. Related Work

### 2.1. Multi-focus image fusion dataset

As mentioned in Section 1, Lytro [1], currently the most widely used multi-focus image fusion dataset, only has 20 pairs of images without ground truth. Therefore, it cannot be used for training of deep neural networks. Alternatively, many deep-learning-based algorithms [2–4, 18] generated datasets artificially as the training set. But these generated datasets are not natural or realistic, especially near the focused/defocused boundary. The characteristics of these datasets are summarized in Table 1.

### 2.2. Light field image processing

Different from the conventional cameras, the light field camera records the complete light field information in one shot, and generates images afterwards. In 2005, Ng proposed a Fourier Slice Photography Theorem [19], which proved that images focused at different depths can be computed from a single light field image. Dansereau *et al.* [20] designed a 4D hyperfan shaped band-pass filter in the frequency domain, which can control the range of focused depths of images in the spatial domain.

## 3. Dataset

### 3.1. Image capture and categories

We use the Lytro illum camera to capture light field images at different places, such as square, campus, shopping mall and street. Photo categories include people, plants, buildings, objects, etc. In addition to the photos taken by us, we also selected 443 images from the Stanford database [21], which mainly include some scenes with obviously separable foreground and background, especially those with more complex boundary between foreground and background.

### 3.2. Pre-processing

We use the *LFLytroDecodeImage* function in the light field toolbox (LFtoolbox) [22] to decode light field data and obtain the raw 4D light field data as well as the 2D Lenslet images. A two-dimensional representation of 4D light field data is shown in Fig. 2(a).

Then, we use the *LFDisp* function to generate 2D images. However, the images generated by LFtoolbox are with color deviation and shape distortion effects. We fix the color difference problem through the gamma transform; and with the help of a calibration plate, the problem of shape distortion is solved. An illustrative example is shown in Fig. 3. In Fig. 3(a), the image before color correction is dim and lacks color details; in Fig. 3(b), the problem has been fixed after color correction. However, it still can be seen that the object shape in Figs. 3(a) and (b) has obvious distortions (e.g. the table is curved). After distortion correction, the table becomes straight in Fig. 3(c).

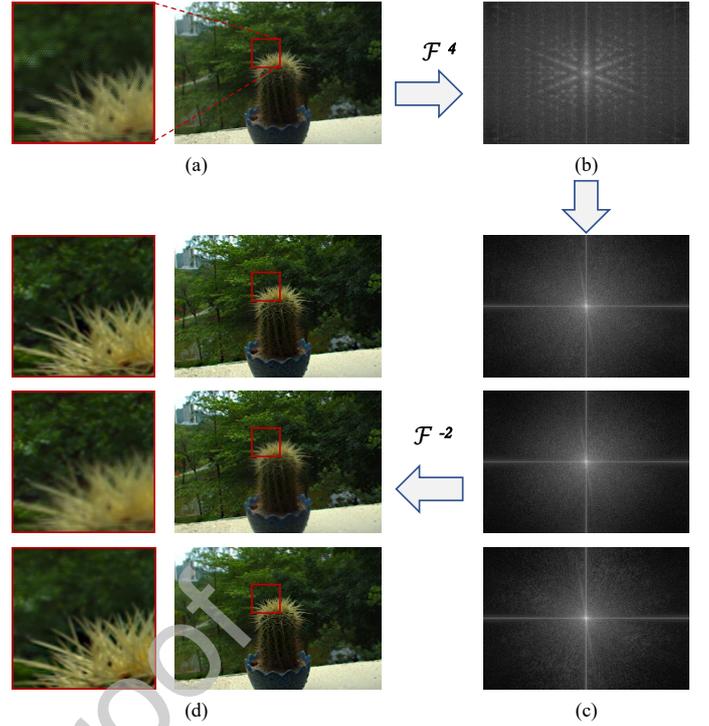


Fig. 2: Refocusing. (a) is the 4D light field captured by a Lytro camera and (b) is the corresponding 4D Fourier spectrum. (c) shows three 2D slices of the 4D Fourier spectrum, and (d) demonstrates the images generated via applying inverse 2D transforms on (c).

### 3.3. Refocusing

The main processing of the dataset is refocusing. For better visualization, we illustrate the processing with an example in Fig. 2. Firstly, a 4D light field image (Fig. 2(a)) is transformed to a 4D Fourier spectrum space (Fig. 2(b)):

$$F(\Omega_u, \Omega_s, \Omega_v, \Omega_t) = DFT(f(u, s, v, t)), \quad (1)$$

where  $f$  denotes the 4D light field image,  $DFT$  represents the discrete Fourier transform,  $F$  is 4D Fourier spectrum,  $(u, s, v, t)$  are the two-plane parameterizations of light rays, and  $(\Omega_u, \Omega_s, \Omega_v, \Omega_t)$  denotes the corresponding coordinates in the 4D Fourier spectrum space.

According to the Fourier Slice Photography Theorem [19], we then use 4D hyperfan, *LFBUILD4DFreqHyperfan* in LFtoolbox [22], as a high-dimensional passband filter to filter the 4D Fourier spectrum (Fig. 2(c)):

$$F_{depth} = H_{HF}(\Omega_u, \Omega_s, \Omega_v, \Omega_t, \theta_1, \theta_2)F(\Omega_u, \Omega_s, \Omega_v, \Omega_t), \quad (2)$$

where  $H_{HF}$  is the 4D hyperfan passband filter,  $F_{depth}$  means the 4D Fourier spectrum of 2D image focused on the specific depth, and parameters  $\theta_1$  and  $\theta_2$  are the slope parameters of the 4D hyperfan.

Finally, the 2D images (Fig. 2(d)) can be generated by the inverse Fourier transform of the processed 4D Fourier spectrum:

$$I = IDFT(F_{depth}), \quad (3)$$

where  $IDFT$  represents inverse discrete Fourier transform, and  $I$  denotes the 2D image focused on the specific depth.

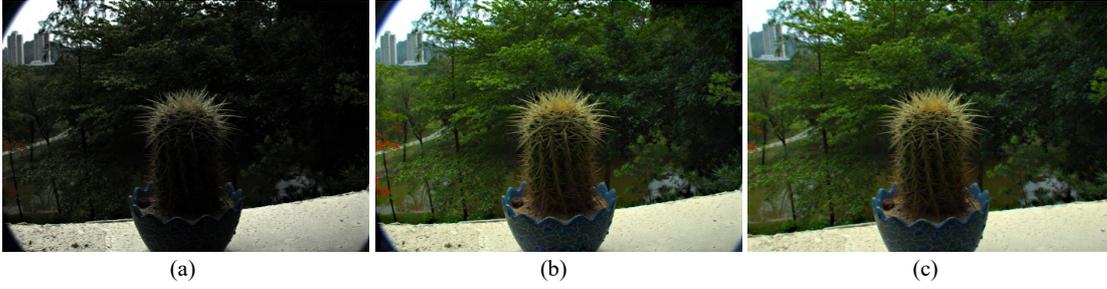


Fig. 3: Example for pre-processing. (a) is the original image before pre-processing, (b) is the intermediate image after color correction and (c) is the final image after color correction and distortion correction.

Since the depth range usually changes for different scenes, we need to first manually select the specific depth range of focus in the Lytro software, then record the depth of focus for each image, and finally generate the desired image focused on different depth ranges. Therefore,  $\theta_1$  and  $\theta_2$  usually change for different images. Typically, both  $\theta_1$  and  $\theta_2$  are in the range of  $[-2, 2]$ . The pseudocode of the algorithm is shown in Algorithm 1.

---

#### Algorithm 1 Refocusing

---

```

for each light field image do
  Decode the light field data:  $f$ ;
  Transform  $f$  into the frequency domain by FFT:
   $F = DFT(f)$ ;
   $j$  = the depth at foreground;
   $k$  = the depth at background;
   $l = (j + k)/2$ ;
  # Generate the foreground focused image
  Build the 4D hyperfan with slopes  $j$  and  $l$ :
   $H_{FG} = H_{HF}(j, l)$ ;
   $F_{FG} = H_{FG}F$ ;
   $I_{FG} = IDFT(F_{FG})$ ;
  # Generate the background focused image
  Build the 4D hyperfan with slopes  $k$  and  $l$ :
   $H_{BG} = H_{HF}(k, l)$ ;
   $F_{BG} = H_{BG}F$ ;
   $I_{BG} = IDFT(F_{BG})$ ;
  # Generate the all-in-focus image
  Build the 4D hyperfan with slopes  $j$  and  $k$ :
   $H_{AF} = H_{HF}(j, k)$ ;
   $F_{AF} = H_{AF}F$ ;
   $I_{AF} = IDFT(F_{AF})$ ;
end for

```

---

## 4. Experiments

### 4.1. Dataset

Our Real-MFF dataset contains 710 pairs of images, where each pair includes two partially-focused source images focused on the foreground or background respectively, and an all-in-focus image as the ground truth. We hope that the ground truth is similar to the composition of two source images: the foreground region should be similar to the source image focused on

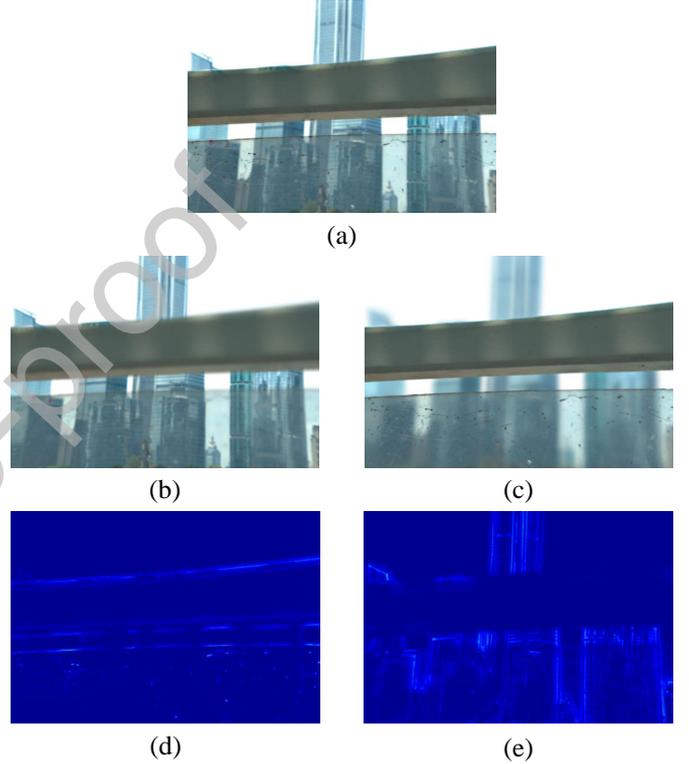


Fig. 4: Example of the generated images. (a) is the all-in-focus image; (b) is the image focused on the background; (c) is the image focused on the foreground; (d) is the difference map between (a) and (b); and (e) is the difference map between (a) and (c).

the foreground, and the background region should be similar to the source image focused on the background. Fig. 4 shows an example for the difference maps between the all-in-focus image and the two source images. The difference shown in Fig. 4(d) indicates that the ground truth (Fig. 4(a)) matches the background of the image focused on the background (Fig. 4(b)); and the difference shown in Fig. 4(e) indicates that the ground truth (Fig. 4(a)) matches the foreground of the image focused on the foreground (Fig. 4(c)).

Furthermore, in order to evaluate the effect of difficulty level on an algorithm, especially for simple and complex boundaries, we manually divide the dataset into a simple set (260 pairs) and a complex set (450 pairs) according to the complexity of focused/defocused boundary. The division is conducted via a three-person majority voting to reduce the bias caused by sub-



Fig. 5: Visual comparison. (a) and (b) are the two partially-focused source images; (c) is the ground truth all-in-focus image; and (d)-(m) are the fusion results of the 10 methods.

PSNR	Dsift [17]	CNN [2]	SR [12]	GF [23]	LP [5]	RP [6]	DTCWT [9]	CVT [7]	NSCT [10]	NSCT-SR [13]	SSIM	Dsift [17]	CNN [2]	SR [12]	GF [23]	LP [5]	RP [6]	DTCWT [9]	CVT [7]	NSCT [10]	NSCT-SR [13]	EPE	Dsift [17]	CNN [2]	SR [12]	GF [23]	LP [5]	RP [6]	DTCWT [9]	CVT [7]	NSCT [10]	NSCT-SR [13]	ALL	Dsift [17]	CNN [2]	SR [12]	GF [23]	LP [5]	RP [6]	DTCWT [9]	CVT [7]	NSCT [10]	NSCT-SR [13]	
Dsift [17]	0	0	0	0	0	0	0	0	0	0	Dsift [17]	0	0	0	0	0	0	0	0	0	0	0	Dsift [17]	0	0	0	0	0	0	0	0	0	0	Dsift [17]	0	0	0	0	0	0	0	0	0	0
CNN [2]	1	0	1	0	0	0	0	0	0	0	CNN [2]	1	0	1	0	0	0	0	0	0	0	0	CNN [2]	1	0	1	0	0	1	0	0	0	0	CNN [2]	1	0	1	0	0	0	0	0	0	0
SR [12]	1	0	0	0	0	0	0	0	0	0	SR [12]	1	0	0	0	0	0	0	0	0	0	0	SR [12]	1	0	0	0	0	0	0	0	0	0	SR [12]	1	0	0	0	0	0	0	0	0	0
GF [23]	1	1	1	0	0	1	0	0	0	0	GF [23]	1	1	1	0	0	0	0	0	0	0	0	GF [23]	1	1	1	0	0	1	0	1	0	0	GF [23]	1	1	1	0	0	0	0	0	0	0
LP [5]	1	1	1	1	0	1	1	1	1	0	LP [5]	1	1	1	1	0	1	1	1	1	0	0	LP [5]	1	1	1	1	0	1	1	1	1	0	LP [5]	1	1	1	1	0	1	1	1	1	0
RP [6]	1	0	1	0	0	0	0	0	0	0	RP [6]	1	1	1	1	0	0	0	0	0	0	0	RP [6]	1	0	1	0	0	0	0	0	0	0	RP [6]	1	0	1	0	0	0	0	0	0	0
DTCWT [9]	1	1	1	1	0	1	0	0	0	0	DTCWT [9]	1	1	1	1	0	1	0	1	0	0	0	DTCWT [9]	1	1	1	0	0	1	0	1	0	0	DTCWT [9]	1	1	1	0	0	1	0	0	0	0
CVT [7]	1	1	1	1	0	1	1	0	0	0	CVT [7]	1	1	1	1	0	1	0	0	0	0	0	CVT [7]	1	1	1	0	0	1	0	0	0	0	CVT [7]	1	1	1	0	0	1	0	0	0	0
NSCT [10]	1	1	1	1	0	1	1	1	0	0	NSCT [10]	1	1	1	1	1	1	1	1	0	0	0	NSCT [10]	1	1	1	1	1	1	1	1	0	0	NSCT [10]	1	1	1	1	1	1	1	1	0	0
NSCT-SR [13]	1	1	1	1	0	1	1	1	1	0	NSCT-SR [13]	1	1	1	1	1	1	1	1	1	0	0	NSCT-SR [13]	1	1	1	1	1	1	1	1	1	0	NSCT-SR [13]	1	1	1	1	1	1	1	1	1	0

(a) PSNR

(b) SSIM

(c) EPE

(d) ALL

Fig. 6: Significance testing results for the performance of different methods in terms of (a) PSNR, (b) SSIM, (c) EPE and (d) the intersection (denoted by ‘ALL’) of (a)-(c), where ‘1’ indicates that the algorithm of the row is significantly better than the algorithm of the column, and ‘0’ otherwise, at the significance level of 0.05.

Table 2: Quantitative comparison of multi-focus image fusion methods on our dataset.

	Dsift [17]	CNN [2]	SR [12]	GF [23]	LP [5]	RP [6]	DTCWT [9]	CVT [7]	NSCT [10]	NSCT-SR [13]
on the whole dataset										
PSNR	36.42	40.46	39.68	40.74	41.32	40.37	40.86	40.90	41.18	41.32
SSIM	0.9744	0.9880	0.9857	0.9890	0.9904	0.9891	0.9896	0.9892	0.9905	0.9906
EPE	2.738	1.901	2.060	1.865	1.790	1.939	1.876	1.886	1.762	1.736
on the simple dataset										
PSNR	38.48	43.41	42.13	43.45	43.93	42.88	43.46	43.52	43.72	43.87
SSIM	0.9797	0.9920	0.9888	0.9921	0.9924	0.9914	0.9922	0.9921	0.9926	0.9927
EPE	2.078	1.375	1.526	1.357	1.294	1.408	1.349	1.349	1.268	1.244
on the complex dataset										
PSNR	35.23	39.76	38.26	39.18	39.81	38.92	39.35	39.38	39.72	39.85
SSIM	0.9714	0.9857	0.9838	0.9873	0.9892	0.9878	0.9881	0.9876	0.9894	0.9894
EPE	3.119	2.205	2.368	2.159	2.076	2.245	2.181	2.196	2.048	2.020

jective selection. Two examples are shown in Fig. 1, where Fig. 1(a) is a complex case and Fig. 1(d) is a simple case. For the complex case in Fig. 1(a), we can find it difficult to draw a curve as the boundary of the foreground and the background, whereas for the simple case in Fig. 1(d), we can easily take the border of the cistern as the boundary.

#### 4.2. Test Bench

We evaluate 10 typical multi-focus image fusion methods, including some transform-domain-based methods (LP [5], RP [6], DWT [8], CVT [7], DTCWT [9], NSCT [10], SR [12], NSCT-SR [13]), some spatial-domain-based methods (Dsift [17], GF [23]), and a deep-learning-based methods (CNN [2]), on our dataset.

We employ the peak signal to noise ratio (PSNR), structure similarity (SSIM) and end point error (EPE) as the evaluation measures, and in terms of these measures, the performances of the 10 multi-focus image fusion methods on the whole dataset, the simple dataset and the complex dataset are listed in Table 2. Unsurprisingly, the results of these algorithms on the simple dataset are generally better than those on the complex dataset. The results also show that, among these algorithms, NSCT and NSCT-SR are the top two on our dataset.

For illustration, we choose an example with complex foreground borders from our dataset, and show the visual results of

various methods in Fig. 5. We can see the difference among these 10 methods in the regions of the falling leaves and the people in the background. It is hard for the spatial-domain-based methods to generate a precise decision map, which means that the resultant images from these methods have some quite blurry areas. Consequently, their PSNRs are usually lower than those of other methods. For the transform-domain-based methods, complex boundaries will decrease the performance of the algorithms, but fortunately this bad effect will not result in complete failure of the algorithms. Moreover, we can see that CNN has blurred results for the falling leaves, and Dsift has blurred results for the person, while NSCT-SR has better results in both areas.

In order to further analyze whether the performances of these methods are statistically significantly different from each other, we perform the Wilcoxon signed rank tests [24] on the experimental results in Table 2. The significance testing results in PSNR, SSIM and EPE are shown in Fig. 6, as well as the intersection of these three results (denoted by ‘ALL’), in all of which the number ‘1’ indicates that the algorithm of the row is significantly better than the algorithm of the column at the significance level of 0.05, and the number ‘0’ otherwise. Based on the results, NSCT-SR performs the best on our dataset.

## 5. Conclusions

In this letter, we propose a large and realistic dataset for multi-focus image fusion tasks. To the best of our knowledge, our Real-MFF dataset is the first large and realistic dataset for this purpose. The dataset is generated by light field images captured with a light field camera. We use the 4D hyperfan in the frequency domain as a band-pass filter to produce images focused on different depths manually. The proposed dataset can serve as a test bench for evaluating multi-focus image fusion methods; furthermore, it can benefit the development of deep-learning-based methods for multi-focus image fusion in the future. It is also our future work to continuously increase the size, diversity and quality of this dataset.

## Acknowledgement

We thank the reviewers and editors for their constructive comments that improved our manuscript. This work was supported in part by the National Natural Science Foundation of China (No.61771276).

## References

- [1] M. Nejadi, S. Samavi, S. Shirani, Multi-focus image fusion using dictionary-based sparse representation, *Information Fusion* 25 (2015) 72–84.
- [2] Y. Liu, X. Chen, H. Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, *Information Fusion* 36 (2017) 191–207.
- [3] H. Ma, J. Zhang, S. Liu, Q. Liao, Boundary aware multi-focus image fusion using deep neural network, in: *IEEE International Conference on Multimedia and Expo (ICME)*, 2019, pp. 1150–1155.
- [4] X. Guo, R. Nie, J. Cao, D. Zhou, L. Mei, K. He, FuseGAN: Learning to fuse multi-focus image via conditional generative adversarial network, *IEEE Transactions on Multimedia* 21 (8) (2019) 1982–1996.
- [5] P. J. Burt, E. H. Adelson, The Laplacian pyramid as a compact image code, *IEEE Transactions on Communications* 31 (4) (1983) 532–540.
- [6] A. Toet, Image fusion by a ratio of low-pass pyramid, *Pattern Recognition Letters* 9 (4) (1989) 245–253.
- [7] F. Nencini, A. Garzelli, S. Baronti, L. Alparone, Remote sensing image fusion using the curvelet transform, *Information Fusion* 8 (2) (2007) 143–156.
- [8] H. Li, B. Manjunath, S. K. Mitra, Multisensor image fusion using the wavelet transform, *Graphical Models and Image Processing* 57 (3) (1995) 235–245.
- [9] J. J. Lewis, R. J. O’Callaghan, S. G. Nikolov, D. R. Bull, N. Canagarajah, Pixel- and region-based image fusion with complex wavelets, *Information Fusion* 8 (2) (2007) 119–130.
- [10] Q. Zhang, B.-I. Guo, Multifocus image fusion using the nonsubsampling contourlet transform, *Signal Processing* 89 (7) (2009) 1334–1346.
- [11] T. Wan, C. Zhu, Z. Qin, Multifocus image fusion based on robust principal component analysis, *Pattern Recognition Letters* 34 (9) (2013) 1001–1008.
- [12] B. Yang, S. Li, Multifocus image fusion and restoration with sparse representation, *IEEE Transactions on Instrumentation and Measurement* 59 (4) (2010) 884–892.
- [13] Y. Liu, S. Liu, Z. Wang, A general framework for image fusion based on multi-scale transform and sparse representation, *Information Fusion* 24 (2015) 147–164.
- [14] Q. Zhang, T. Shi, F. Wang, R. S. Blum, J. Han, Robust sparse representation based multi-focus image fusion with dictionary construction and local spatial consistency, *Pattern Recognition* 83 (2018) 299–313.
- [15] I. De, B. Chanda, Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure, *Information Fusion* 14 (2) (2013) 136–146.
- [16] M. Li, W. Cai, Z. Tan, A region-based multi-sensor image fusion scheme using pulse-coupled neural network, *Pattern Recognition Letters* 27 (16) (2006) 1948–1956.
- [17] Y. Liu, S. Liu, Z. Wang, Multi-focus image fusion with dense SIFT, *Information Fusion* 23 (2015) 139–155.
- [18] W. Zhao, D. Wang, H. Lu, Multi-focus image fusion with a natural enhancement via a joint multi-level deeply supervised convolutional neural network, *IEEE Transactions on Circuits and Systems for Video Technology* 29 (4) (2019) 1102–1115.
- [19] R. Ng, Fourier slice photography, *ACM Transactions on Graphics* 24 (3) (2005) 735–744.
- [20] D. G. Dansereau, O. Pizarro, S. B. Williams, Linear volumetric focus for light field cameras, *ACM Transactions on Graphics* 34 (2) (2015) 1–20.
- [21] D. G. Dansereau, B. Girod, G. Wetzstein, LiFF: Light field features in scale and depth, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8042–8051.
- [22] D. G. Dansereau, O. Pizarro, S. B. Williams, Decoding, calibration and rectification for lenselet-based plenoptic cameras, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1027–1034.
- [23] S. Li, X. Kang, J. Hu, Image fusion with guided filtering, *IEEE Transactions on Image Processing* 22 (7) (2013) 2864–2875.
- [24] F. Wilcoxon, Individual comparisons by ranking methods, *Biometrics Bulletin* 1 (6) (1945) 80–83.

## Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.