IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 5, 2020, accepted March 14, 2020, date of publication March 18, 2020, date of current version April 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2981641

Raindrop Removal With Light Field Image Using Image Inpainting

TAO YANG[1], XIAOFEI CHANG[2], HANG SU[3,4], NATHAN CROMBEZ[1], YASSINE RUICHEK[1], (Senior Member, IEEE), TOMAS KRAJNIK[5], AND ZHI YAN[1], (Member, IEEE)

[1]CIAD, University Bourgogne Franche-Comté, UTBM, 90010 Belfort, France
[2]School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China
[4]National Time Service Center, CAS, Xi'an 710600, China
[5]Artificial Intelligence Center, Czech Technical University in Prague, 121 35 Prague, Czech Republic

Corresponding author: Xiaofei Chang (changfei@nwpu.edu.cn)

This work was supported in part by the Natural Science Foundation of Shaanxi Province, China, under Grant 2019ZY-CXPT-03, in part by the PHC Barrande Programme under Grant 40682ZH (3L4AV) and the CZ MSMT Projects under Grant FR-8J18FR018, and in part by the CSF under Grant 17-27006Y and Grant 20-27034J.

ABSTRACT In this paper, we propose a method that removes raindrops with light field image using image inpainting. We first use the depth map generated from light field image to detect raindrop regions which are then expressed as a binary mask. The original image with raindrops is improved by refocusing on the far regions and filtering by a high-pass filter. With the binary mask and the enhanced image, image inpainting is then utilized to eliminate raindrops from the original image. We compare pre-trained models of several deep learning based image inpainting methods. A light field raindrop dataset is released to verify our method. Image quality analysis is performed to evaluate the proposed image restoration method. The recovered images are further applied to object detection and visual localization tasks.

INDEX TERMS Raindrop removal, light field, image inpainting.

I. INTRODUCTION

Object detection and self-localization are the fundamental tasks in autonomous driving as well as mobile robotics. In recent years, thanks to the rapid development of deep learning, vision-based perception and localization [1]–[3] have been greatly improved in both accuracy and reliability. Deep learning based approaches also show good generalization ability under the deep neural network structure and the support of a large amount of training data. Facing complex environments in practical applications, more and more work has been initiated to focus on edge cases, including adverse weather conditions.

Raindrops falling on vehicle windows (in case of built-in camera) or camera lenses (in case of external camera) on a rainy day are one of them. They typically cause vision sensors to produce blurry images which in turn interfere with high-level environmental perception tasks. To alleviate this problem, recently, Qian *et al.* [4] collected image pairs with and without raindrops and trained a Generative Adversarial Network (GAN) [5] in supervised learning paradigm to remove raindrops, that can recover test images effectively. However, due to the relatively small dataset for neural network training, we found that there is still room for improvement of model generalization ability, especially when inputting non-homologous images. Although a direct way in this matter is to increase the model capacity by feeding more training examples, it is very challenging to collect image pairs with and without raindrops in different lighting conditions, scenes, and environments.

As an emerging device, light field cameras not only retain light intensity information, but also direction information of the ray. A typical light field imaging system is a camera array. Adelson and Wang [6] pioneeringly designed a plenoptic camera that consists of a main lens, a microlens array, and an imaging sensor. Later on, Ng *et al.* [7] redesigned it in order to have a handheld light field camera based on a conventional device. Along with the development of hardware, considerable work has emerged in recent years focusing on depth estimation [8], [9] based on light field image.

The associate editor coordinating the review of this manuscript and approving it for publication was Feng Shao.

Although the handheld light field camera cannot obtain accurate depth maps over longer distances due to the short baseline, it can well distinguish the far and near areas of the scene, which makes it straightforward to detect raindrops on the glass or window in front of camera. Different from the single image raindrop removal methods, which have the limitations on generalization capabilities, the image pairs, which are used to train the image inpainting task in a supervised learning paradigm, can be easily generated. It makes deep learning based image inpainting methods [10], [11] have million-level training datasets and work well even on the non-homologous images with pre-trained models. We therefore believe that the combination of light field image and image inpainting is helpful for raindrop removal task in general scenarios.

In this paper, we propose a method that can effectively remove raindrops from the images captured by a handheld light field camera using image inpainting. Since a light field image is equivalent to an image array captured by a camera array, we use the central one as the original image. Examples with raindrops are shown on the left side of Fig. 1. We use the estimated depth map to detect raindrop regions which are subsequently expressed as a binary mask. The original image is first improved by refocusing on the far regions and high-pass filtering. With the binary mask and enhanced image, image inpainting is then applied to eliminate raindrops. Examples of recovered images are shown on the right side of Fig. 1. We compare several deep learning based image inpainting methods and directly use the pre-trained models. Image quality analysis is used to evaluate the image restoration. The recovered images are finally applied to the object detection and self-localization tasks.

FIGURE 1. Left: The central images in the image arrays captured by the light field camera. Right: The recovered images which eliminate raindrops by binary masks, enhanced images, and image inpainting.

The contributions of this paper are threefold:

- We propose a novel raindrop removal method with light field image using image inpainting.
- We release a new dataset which contains light field images with raindrops collected in different scenarios, publicly available at https://github.com/cavayangtao/light-field-raindrop-dataset.
- We comparatively evaluate our method in different scenarios using image quality analysis and apply the recovered images to object detection and self-localization tasks. The experimental results show that the use of recovered images can help improve system performance.

This paper is organized as follows: in Section II, we give the related work of light field imaging, image inpainting, and raindrop removal methods; Section III describes in detail our proposed method; Section IV first introduces the dataset of light field images with raindrops and then shows the experimental results and analysis; Section V summarizes the paper and prospects future work.

II. BACKGROUND

A. LIGHT FIELD IMAGING

According to the light field rendering theory proposed by Levoy and Hanrahan [12], the arbitrary ray can be represented by a radiation function. The light field is all the rays in space. The four-dimensional direction information in ray space can be parameterized by two parallel planes which are shown as coordinates of lens and sensor in Fig. 2. Based on this principle, Wilburn *et al.* [13] captured light field by an array with different numbers of cameras. Similar to the camera array, Georgiev and Intwala [14] designed a light field acquisition method that placed the lens array in front of the camera lens. Liang *et al.* [15] added a programmable liquid crystal filter in front of the lens to sample sub-aperture images by multiple exposures, which directly transformed the conventional camera into a light field one with high image resolution. However, multiple exposures require longer time and larger amount of data storage space. Ng *et al.* [7] therefore simplified the design of a plenoptic camera and made a handheld light field camera based on a conventional camera.

The device we used for dataset collection was the first generation Lytro camera. It is actually a plenoptic camera consisting of a microlens array inserted in front of the sensor

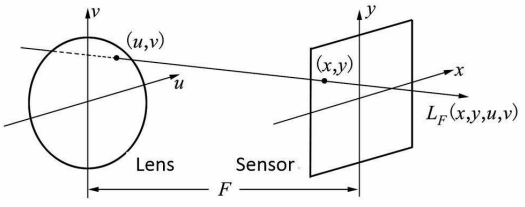


FIGURE 2. The ray going from the position (u, v) of the lens to the position (x, y) of the sensor, where F is the distance between the lens and sensor, can be represented by a four-dimensional light field function $L_F(x, y, u, v)$ [7]. The light field is all the rays in space.

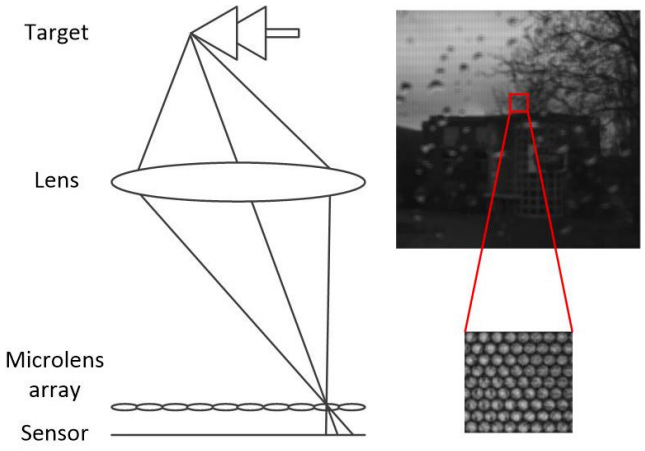


FIGURE 3. Left: Schematic diagram of a plenoptic camera. Right: The circular spots imaged by the microlens array are called macro pixels. The position of a pixel in the macro pixel corresponds to the coordinates of (u, v), while the position of a macro pixel on the sensor corresponds to the coordinates of (x, y).

(see Fig. 3). This light field camera has a refocusing function because it is capable of capturing multi-viewpoint images [7]. Unlike current refocusing methods applied to smartphones (i.e. blurring the background according to the depth map), refocusing by a light field camera is equivalent to an optical lens. When refocusing on a distant region, small objects nearby can be eliminated [16]. Considerable work has been focused on estimating the depth map from the light field image. Dansereau and Bruton [17] applied 2D gradient operators to Epipolar Plane Image (EPI) slides from the light field to estimate the depth map. Tao *et al.* [18] first estimated depth combining defocus and correspondence cues, then proposed a method [19] estimating local shape from defocus and correspondence cues to further refine the depth. With the development of deep learning technology in the field of image processing, Convolutional Neural Networks (CNN) were applied to light field images for depth estimation [8] and materials recognition [20]. Lately, Mildenhall *et al.* [21] presented a CNN-based method for light field synthesis from a set of input images captured by a handheld camera.

B. IMAGE INPAINTING

Image inpainting, also referred as image completion, aims to fill missing areas of an image. It is an important task in computer vision and image editing. The applications include damaged image restoration, object removal and more. Traditional approaches can be divided into diffusion-based method and patch-based method. The former propagates information from surrounding regions of missing areas [22], while the later fills in missing regions by getting information of the similar regions [23], [24]. However, these methods suffer from challenging situations such as urban traffic scenes, as they rely solely on texture features without semantic information.

Recently, deep learning based methods have made remarkable improvements in image inpainting by learning data distribution combining the texture and semantic information with CNN and GAN. Pathak *et al.* [25] implemented image inpainting using an encoder-decoder CNN architecture. Yang *et al.* [26] jointly optimized image content and texture constraints through a multi-scale optimization approach, which improved the output of the context-encoder at the cost of high demand of computing time and storage. GAN models the distribution of data by adversarial training [5]. Isola *et al.* [27] proposed a conditional GAN framework for image-to-image translation problems, which is widely used for image synthesis. Yeh *et al.* [28] employed GAN for image inpainting task with uncorrupted data. Iizuka *et al.* [29] addressed blurriness problems using global and local context discriminators to distinguish the generated images from real ones relying on DCGAN loss [30]. Yu *et al.* [31] used an attention mechanism to refine image inpainting results with a modified version of WGAN-GP loss [32]. Nazeri *et al.* [10] proposed a two-stage adversarial model that synthesizes edges of missing regions by an edge generator, then inputs predicted edges and incomplete color image to another generator for final inpainting. Moreover, Hong *et al.* [11] recently achieved good performance of image completion using a U-Net architecture embedded with fusion blocks to the last few decoder layers.

C. RAINDROP REMOVAL

Unfortunately, there are few works on the detection and removal of raindrops in images. Roser and Geiger [33] detected raindrop in a single image based on a photometric raindrop model and fused multiple frames into a single frame to restore the image. To further, Wu *et al.* [34] analyzed the color, texture, and shape features to propose raindrop candidates, and then reduced false detections through a learning-based algorithm. The regions occupied by raindrops are finally filled with image inpainting. Kang *et al.* [35] removed raindrops using morphological component analysis based image decomposition. Besides using a monocular camera, Yamashita *et al.* detected and removed raindrops with a stereo system [36] and image sequences [37], respectively. However, since these methods do not benefit from the latest progress in deep learning, the image restoration is not satisfactory. Eigen *et al.* [38] first used CNN to restore the raindrop image by end-to-end training. Qian *et al.* [4] improved the raindrop removal and image recovery quality by GAN and attention mechanism. However, limited to the size of the training set, the generalization ability of neural network methods needs to be further improved. In summary, to the best of our knowledge, removing raindrops in light field image by deep learning based image inpainting and the light field raindrop dataset are proposed for the first time in this paper.

III. PROPOSED METHOD

The overall flowchart of our method is shown in Fig. 4. We first estimate the depth map D from the light field image I, and a binary mask M showing the raindrop regions is obtained

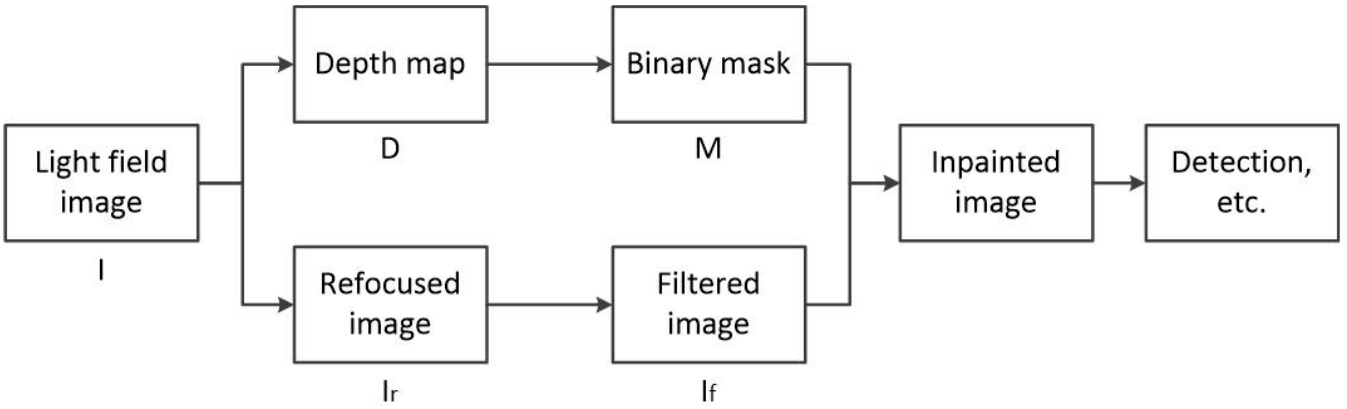


FIGURE 4. The overview flowchart of the proposed method. A binary mask showing the raindrop regions is generated using the depth map. In parallel, the image refocusing on the far regions is enhanced by a high-pass filter. Finally, a deep learning based image inpainting method is used to restore the masked raindrop image.

using the depth map. Then, the image refocusing on the far regions (I_r) is enhanced by a high-pass filter, while the filtered image is expressed as I_f. Finally, we utilize deep learning based image inpainting algorithm to restore the masked raindrop image. The recovered image can be further used for tasks such as object detection and self-localization.

A. MASKING REFOCUSED IMAGE

For the light field function $L_F(x, y, u, v)$ representing the ray in space, by fixing the parameters of (u, v) as 2D slices of the 4D light field function, we can then get a sub-aperture image, which is equivalent to the one obtained from the viewpoint (u, v) as expressed by:

$$I^{(u,v)}(x, y) = L_F^{(u,v)}(x, y) \tag{1}$$

Fig. 5 shows the sub-aperture image array obtained by rearranging the original light field image, in which the coordinates of (u, v) represent angular resolution, while the coordinates of (x, y) expresses spatial resolution.

The disparity between each sub-aperture can be used to calculate the depth map D. A common method to do so is to estimate the slopes of lines in EPI slices from the light field image [17]. To analyze the orientation of patterns in EPI, Johannsen *et al.* [39] built a dictionary with atoms of fixed disparity and used sparse coding representation to find which elements in the dictionary can best describe the patch. Jeon *et al.* [40] estimated correspondences from the light field based on EPI with sub-pixel accuracy using the cost volume. Tao *et al.* [41] developed a method for local shape estimation based on defocus and correspondence cues, and then used shading to further refine the depth. However, the Lytro software (associated with the camera) we used can directly provide competitive depth output, which can effectively lightweight our system on the one hand, and also conducive to system reuse on the other. As shown on the left side of Fig. 6, the raindrops are well distinguished in the depth map generated by the software. Furthermore, we convert the depth map D to a binary image M for subsequent image

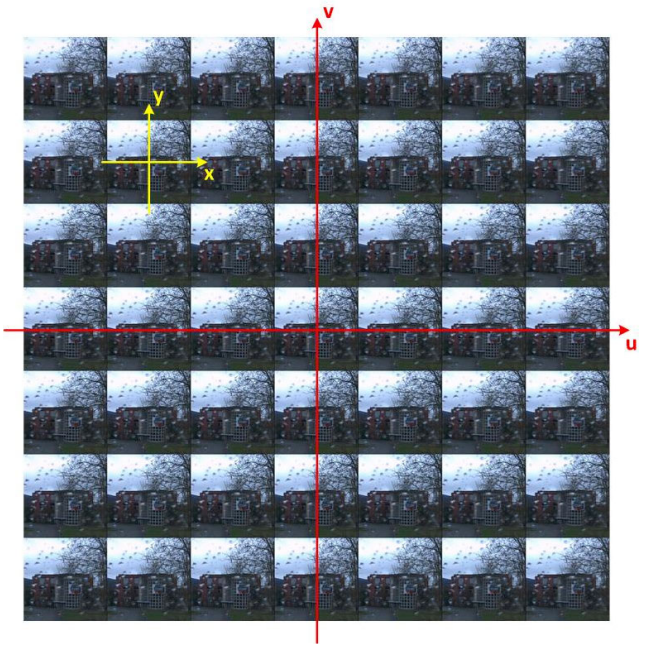


FIGURE 5. Sub-aperture image array obtained by rearranging the original light field image.

completion. In order to have M that completely covers the raindrops, we first use the close and open operators to eliminate the noise on the original binary image, and then use the erosion operator to enlarge the raindrop ranges. The final binary image is shown in right part of Fig. 6.

The refocusing of the light field image is achieved by assuming a virtual imaging plane. For a light field imaging system with a distance F between the lens and the imaging plane (c.f. Fig. 2), we assume that the imaging plane moves to a new distance F'. The ray, which passes through the positions (u, v) and (x, y), has a new position (x', y') on the virtual imaging plane. With simple triangular geometry [7], we can use the light field function of the original imaging plane to obtain a new light field function for the same ray as

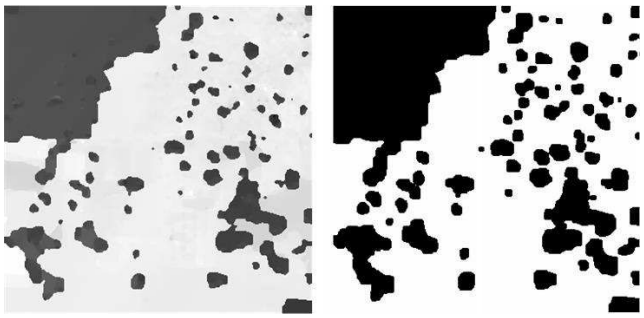


FIGURE 6. Left: The depth map generated by the Lytro light field software. Right: The binary mask processed by image morphology.

expressed in:

$$\begin{aligned} &L_{F'}\left(x', y', u, v\right) \\ &\quad = L_F\left(x, y, u, v\right) \\ &\quad = L_F\left(u\left(1-\frac{1}{\alpha}\right)+\frac{x'}{\alpha}, v\left(1-\frac{1}{\alpha}\right)+\frac{y'}{\alpha}, u, v\right) \end{aligned} \quad (2)$$

where $\alpha = F'/F$. Then the final refocused image on the virtual imaging plane can be represented as the following integral:

$$I_{\alpha F}\left(x', y'\right) = k \iint L_F\left(p, q, u, v\right) dudv \quad (3)$$

where $p = u\left(1-\frac{1}{\alpha}\right)+\frac{x'}{\alpha}$ and $q = v\left(1-\frac{1}{\alpha}\right)+\frac{y'}{\alpha}$; k is a normalization coefficient.

To highlight the background and weaken the influence of raindrops, we refocus on the far regions (I_r). Then, a high-pass filter H is further used to enhance image details as:

$$H = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 1 \end{bmatrix} \quad (4)$$

The binary mask M, which indicates the regions of raindrops, is applied to the filtered image I_f for further image restoration. Fig. 7 shows the original sub-aperture image, enhanced image, and masked image, respectively.

FIGURE 7. Left: the original sub-aperture image. Center: the refocused image processed by a high-pass filter. Right: the masked image.

B. RAINDROP REMOVAL

To eliminate raindrops, image inpainting is used to get a clear image with the enhanced image and binary mask. We have tried several state-of-the-art methods, including:

FIGURE 8. Raindrop removal with three different deep learning based methods. Left: Contextual Attention. Center: Edge Connect. Right: DFNet.

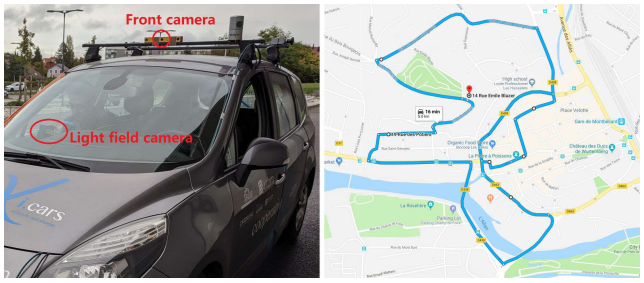


FIGURE 9. The last 50 images of our dataset was collected with the UTBM robocar. The Lytro camera was held in the car by the author, and the corresponding GPS information was also recorded. As shown on the right side in the figure, the route of data collection is the same as the EU long-term dataset [43].

- **Contextual Attention** [31] is a two-stage coarse-to-fine network architecture. The first stage is a dilated convolutional network with reconstruction loss to get rough completion contents. The second stage uses the rough image inpainting result as input with two Wasserstein GAN losses, where one looks at the global image and the other looks at the missing regions. A contextual attention mechanism is integrated into the second stage to borrow feature information from the known background to generate missing patches.
- **Edge Connect** [10] is a two-stage processing method. Different from Contextual Attention, the first stage is an edge generator, while the second stage is an image completion network which uses the masked image and edges generated from the first stage as input. Each stage consists of a generator/discriminator pair. The two-stage generators are first trained separately using Canny edges then fine-tuned end-to-end. This approach is inspired by how artists work and is able to recover high and low frequency information of the missing regions.
- **DFNet** [11] is a U-Net architecture embedded with fusion blocks to several layers of decoder. The fusion block extracts raw completion from feature maps and predicts a composition mask, which is the missing region but has smoother weights on the boundary regions. The final output is generated combining the information of the raw completion, composition map, as well as scaled input image. Instead of using GAN loss, a total loss combining structural loss and texture loss is used for training.

FIGURE 10. From left to right in each row: original sub-aperture image, enhanced image, masked image, recovered image with Edge Connect, recovered image with DFNet, recovered image with DeRaindrop [4].

As the Contextual Attention approach tries to find similar content from the background to fill the missing pixels, but neglects the structural information, it does not show competitive performance. In contrast, Edge Connect and DFNet show interesting results, since they consider both structural and texture information as important features, and are therefore integrated into our method. We use the models pre-trained on Places2 Dataset [42], which contains 10 million scene photographs, for our raindrop removal task. Fig. 8 shows examples of the results obtained.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. LIGHT FIELD RAINDROP DATASET

We evaluated the proposed method on a new dataset collected with the first generation Lytro camera. The dataset contains 90 light field raindrop images that cover different camera angles and scenes. For each image, we provide the light field source file, reference image, and reference depth map given by the Lytro software. The light field source file can be easily decoded by the light field toolbox.[1] We collected raindrop images with glass plate, window of train, and window of car. The camera is placed 10 to 15 cm in front of the glass plate and windows. The first 40 images in our dataset cover different situations evenly, which are used for image quality analysis and object detection task. Using the EU long-term dataset [43] as reference, the last 50 images are applied to self-localization task. For the latter, we drove on a rainy day to collect data following the same path as the previous dataset (see Fig. 9). Table 1 gives a summary of our dataset.

[1] https://github.com/doda42/LFToolbox

TABLE 1. An overview of the light field raindrop dataset.

Image number	Glass angle	Obstacle	GPS
1-10	Parallel	Raindrops	No
11-28	Tilted	Raindrops	No
29-40	Tilted	Snowflakes	No
41-90	Tilted	Raindrops	Yes

TABLE 2. Average image quality scores.

Image types	Average IQS
I	4.62
R1 - A	4.87
R1 - B	4.80
R1 - C	4.39
R1	4.95
R2	5.02

B. IMAGE QUALITY ANALYSIS

Image quality assessment measures image degradations such as noise, blur, etc. It can be categorized into full-reference and no-reference approaches. The former needs an ideal reference image to calculate image quality metrics, while the latter predicts image quality using statistical model. We use the no-reference CNN based image quality predictor that relies on the state-of-the-art deep object recognition networks, which is end-to-end trained on the dataset of image quality assessment [44]. We measured the Image Quality Scores (IQS) of the original sub-aperture image, raindrop-removed image using Edge Connect, and raindrop-removed image using DFNet, respectively. In addition, we also give the results when disabling different components in our method. The average IQS of the first 40 images of each type are calculated as shown in Table 2, where "I" is the original image; "R1" is the recovered image with Edge Connect; "R2" is the recovered image with DFNet; "R1 - A" is the output without image inpainting, which is equivalent to the enhanced image I_f; "R1 - B" is the recovered image with Edge Connect but disabling the refocusing module; "R1 - C" is the recovered image with Edge Connect but disabling the high-pass filter.

It can be seen that the quality of the enhanced images is improved compared to the central sub-aperture images for two reasons. First, the background is strengthened by refocusing and high-pass filter, while the raindrops that interfere with the image quality are weakened. Second, the refocused image is the integration of light field, thus it has a higher signal-to-noise ratio than a single sub-aperture image. It is intuitive that the images with raindrops removal have the highest average quality scores. Moreover, DFNet performs best, as it considers both the structure and texture information of the image to produce more accurate image details, rather than two steps like Edge Connect.

When the refocusing module is disabled, the IQS is decreased because the signal-to-noise ratio of the original image is relatively low, and only using a high-pass filter to strength the high-frequency information also increases noise. When the high-pass filter module is disabled, the image quality is greatly reduced, even lower than the original image. The reason is that although refocusing weakens the raindrops in the foreground, the background becomes more blurred than the original image, as the sub-aperture has a deeper depth of field. Meanwhile, the weakening of high-frequency information is not conducive to the extraction of image structure information in image inpainting.

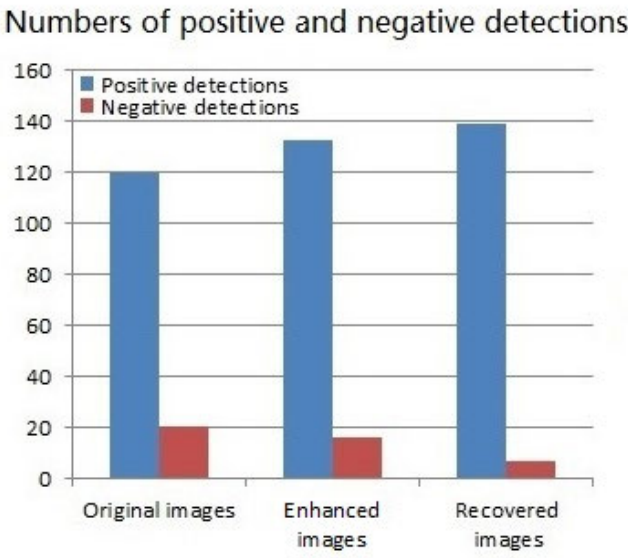


FIGURE 11. The original sub-aperture images, enhanced images, and raindrop-removed images have positive detections of 120, 133, 139 and negative detections of 20, 16, 7, respectively.

Fig. 10 shows some examples of image recovery using our method. We can see that if the raindrops detection is correct, image inpainting methods are able to remove the raindrops and restore the clear image well. Although DFNet can produce images with higher quality scores, there is no significant difference between the results and those of Edge Connect from the perspective of human eyes.

In Fig. 10, We also show the results of directly applying the single image raindrop removal method, i.e. DeRaindrop [4], on the original sub-aperture images. There is a slight effect on the first image. On the second one, only raindrops in the upper half are detected and removed. It has little effect on other images, and we find this method doesn't work well in most of the images we have collected.

C. OBJECT DETECTION

In order to evaluate the usefulness of our method for object detection, we input the original images, enhanced images, and the images after raindrop removal, respectively into Google Vision API[2] for comparative analysis. As Edge Connect keeps the original image resolution at 360 × 360, while DFNet forces to output an image with a magnified resolution

[2]http:/cloud.google.com/vision/

FIGURE 12. The object detection results of the original sub-aperture images, enhanced images, and recovered images, from left to right of each scenario.

of 512×512, we use therefore the raindrop-removed images output from Edge Connect for testing for fair comparison. The experimental results are shown in Fig. 11. We count the numbers of positive detections and negative detections, and it can be seen that the recovered images outperform those without raindrop removal.

Fig. 12 gives some intuitive examples of the object detection results. It can be seen that negative detections may appear under the interference of raindrops. For example, in the first row of Fig. 12, the background mountain is detected as a building, and the animals are erroneously detected. Compared with the original sub-aperture images, the enhanced images bring improvements to the object detection task, and the recovered images without raindrops provide the best performance. As evidence, on the right side of the second row and the left side of the third row of Fig. 12, only the images without raindrops output the bounding boxes of the buildings successfully.

D. SELF-LOCALIZATION

Finally, we perform self-localization tasks using light field images from number 41 to 90. The reference images are extracted from the data "2018-07-20" of the EU long-term dataset [43] by one-tenth of the original sampling rate of the front central camera. The localization is implemented by image retrieval, where we calculate the similarities between a processed light field image and all the reference images, then use the reference image with the maximal similarity as our localization result. The similarity score is calculated by DEep Local Feature (DEFL) [3], which is based on CNN trained on a large landmark image dataset to identify semantic local features for image retrieval. Instead of using extracted features directly, an attention mechanism is designed in DELF for keypoint selection. At the end, the number of inlier matching keypoints selected by the geometric verification is used as the similarity score.

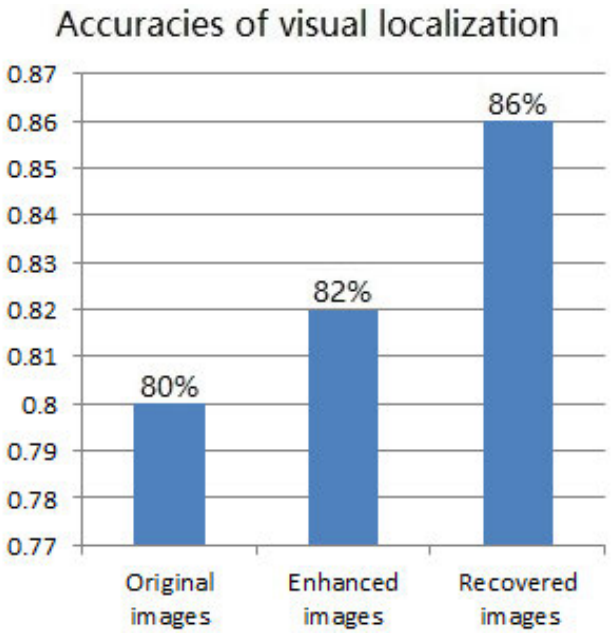


FIGURE 13. The accuracies of visual localization of the original sub-aperture images, enhanced images, and raindrop-removed images.

Fig. 13 shows the accuracies of visual localization of the original sub-aperture images, enhanced images, and raindrop

FIGURE 14. From left to right are the examples of visual localization results of sub-aperture images, enhanced images, and recovered images, respectively. Green color means the positive retrieval and red color indicates the negative retrieval.

removed images, which are 80%, 82%, 86% respectively. The results are consistent with image quality analysis and object detection results. The raindrop-removed images achieve the best result, followed by the enhanced images.

Fig. 14 gives some examples of visual results. Although deep learning features have strong robustness, raindrops still causes interference to some pictures like the first and second rows in Fig. 14. For very similar scenarios such as the third row, it leads to wrong matching for all the images. As consequence, improving image quality by refocusing and removing raindrops results in better visual localization performance.

E. DISCUSSION

By image quality analysis, it can be seen that the images without raindrops have higher IQS, which is intuitive. Meanwhile, every component in our framework contributes to the final image quality improvement. Compared to the existing single image raindrop removal method, our method is effective for more general scenarios and also able to remove the snowflakes on the window in front of camera. These are mainly due to the geometric information based depth map construction from light field image and the great generalization ability of deep learning based image inpainting.

When applying the recovered images to high-level perception tasks such as object detection and self-localization, there is a significant performance improvement over images with raindrops. Unlike the camera array, the handheld light field camera doesn't require a complex signal synchronization system. Since the depth map is important to guide the raindrop removal, our method is more suitable for light or moderate raining conditions. In summary, it's important to benefit from both the rich ray information of light field image and current progress in deep learning based image processing for the perception tasks under adverse weather conditions.

V. CONCLUSION

In this paper, we proposed a method that removes raindrops with light field image using image inpainting. The depth map generated from the light field image was used to detect raindrop regions, which were then expressed as a binary mask. In parallel, the original image was improved

by refocusing on the far regions and filtering by a high-pass filter. Image inpainting was finally utilized to eliminate raindrops with the binary mask and enhanced image. A new publicly available dataset of light field raindrop images covering different camera angles and scenes was built to evaluate our method. Image quality analysis, object detection, and vision-based self-localization were performed to prove the raindrop removal enhancement with light field images. It should be noted that our method still has many steps which makes it difficult to run in real-time. It's interesting to make raindrop removal with light field image an end-to-end paradigm in the future. Moreover, combining use of our light field dataset with the EU long-term dataset for long-term autonomy of vehicles is also a point of our future attention.

REFERENCES

[1] T. Yang, C. Cappelle, Y. Ruichek, and M. El Bagdouri, "Multi-object tracking with discriminant correlation filter based deep learning tracker," *Integr. Comput.-Aided Eng.*, vol. 26, no. 3, pp. 273–284, Apr. 2019.

[2] Y. Qiao, C. Cappelle, Y. Ruichek, and T. Yang, "ConvNet and LSH-based visual localization using localized sequence matching," *Sensors*, vol. 19, no. 11, p. 2439, 2019.

[3] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han, "Large-scale image retrieval with attentive deep local features," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3456–3465.

[4] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2482–2491.

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014, *arXiv:1406.2661*. [Online]. Available: http://arxiv.org/abs/1406.2661

[6] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 99–106, 1992.

[7] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Comput. Sci. Tech. Rep.*, vol. 2, no. 11, pp. 1–11, 2005.

[8] X. Sun, Z. Xu, N. Meng, E. Y. Lam, and H. K.-H. So, "Data-driven light field depth estimation using deep convolutional neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 367–374.

[9] J. Tian, Z. Murez, T. Cui, Z. Zhang, D. Kriegman, and R. Ramamoorthi, "Depth and image restoration from light field in a scattering medium," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2401–2410.

[10] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, "EdgeConnect: Generative image inpainting with adversarial edge learning," 2019, *arXiv:1901.00212*. [Online]. Available: http://arxiv.org/abs/1901.00212

[11] X. Hong, P. Xiong, R. Ji, and H. Fan, "Deep fusion network for image completion," 2019, *arXiv:1904.08060*. [Online]. Available: http://arxiv.org/abs/1904.08060

[12] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn.* New York, NY, USA: ACM, 1996, pp. 31–42.

[13] B. S. Wilburn, M. Smulski, H.-H. K. Lee, and M. A. Horowitz, "Light field video camera," *Proc. SPIE*, vol. 4674, Dec. 2001, pp. 29–36.

[14] T. Georgiev and C. Intwala, "Light field camera design for integral view photography," Adobe, San Jose, CA, USA, Tech. Rep., 2006.

[15] C.-K. Liang, G. Liu, and H. H. Chen, "Light field acquisition using programmable aperture camera," in *Proc. IEEE Int. Conf. Image Process.*, vol. 5, Sep. 2007, pp. 233–236.

[16] S. McCloskey, "Masking light fields to remove partial occlusion," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 2053–2058.

[17] D. Dansereau and L. Bruton, "Gradient-based depth estimation from 4D light fields," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 3, May 2004, p. III-549.

[18] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 673–680.

[19] M. W. Tao, P. P. Srinivasan, S. Hadap, S. Rusinkiewicz, J. Malik, and R. Ramamoorthi, "Shape estimation from shading, defocus, and correspondence using light-field angular coherence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 546–560, Mar. 2017.

[20] T.-C. Wang, J.-Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, and R. Ramamoorthi, "A 4D light-field dataset and CNN architectures for material recognition," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 121–138.

[21] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. K. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 29:1–29:14, Jul. 2019.

[22] S. Esedoglu and J. Shen, "Digital inpainting based on the Mumford–Shah–Euler image model," *Eur. J. Appl. Math.*, vol. 13, no. 4, pp. 353–370, Aug. 2002.

[23] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, p. 24, 2009.

[24] J.-B. Huang, S. B. Kang, N. Ahuja, and J. Kopf, "Image completion using planar structure guidance," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–10, Jul. 2014.

[25] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2536–2544.

[26] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-resolution image inpainting using multi-scale neural patch synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6721–6729.

[27] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[28] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6882–6890.

[29] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," *ACM Trans. Graph.*, vol. 36, no. 4, Jul. 2017, Art. no. 107.

[30] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434

[31] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5505–5514.

[32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2017, pp. 214–223.

[33] M. Roser and A. Geiger, "Video-based raindrop detection for improved image registration," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops, ICCV Workshops*, Sep. 2009, pp. 570–577.

[34] Q. Wu, W. Zhang, and B. V. K. Vijaya Kumar, "Raindrop detection and removal using salient visual features," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 941–944.

[35] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic single-image-based rain streaks removal via image decomposition," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1742–1755, Apr. 2012.

[36] A. Yamashita, Y. Tanaka, and T. Kaneko, "Removal of adherent waterdrops from images acquired with stereo camera," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Aug. 2005, pp. 400–405.

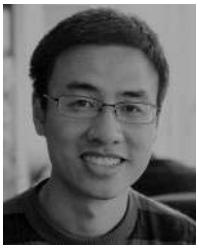
[37] A. Yamashita, I. Fukuchi, T. Kaneko, and K. T. Miura, "Removal of adherent noises from image sequences by spatio-temporal image processing," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2008, pp. 2386–2391.

[38] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 633–640.

[39] O. Johannsen, A. Sulc, and B. Goldluecke, "What sparse light field coding reveals about scene structure," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3262–3270.

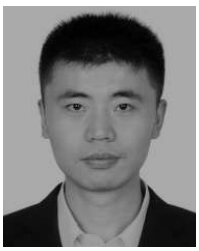
[40] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1547–1555.

[41] M. W. Tao, P. P. Srinivasan, J. Malik, S. Rusinkiewicz, and R. Ramamoorthi, "Depth from shading, defocus, and correspondence using light-field angular coherence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1940–1948.
[42] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1452–1464, Jun. 2018.
[43] Z. Yan, L. Sun, T. Krajnik, and Y. Ruichek, "EU long-term dataset with multiple sensors for autonomous driving," 2019, *arXiv:1909.03330*. [Online]. Available: http://arxiv.org/abs/1909.03330
[44] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018.



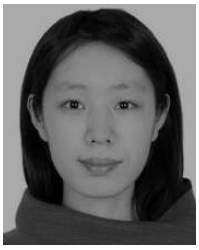
TAO YANG received the B.S. degree in detection, guidance, and control technology and the M.S. degree in control engineering from Northwestern Polytechnical University (NPU), Xi'an, China, in 2011 and 2014, respectively, and the Ph.D. degree in automation from the University of Technology of Belfort-Montbéliard (UTBM), France, in 2019.

Since 2019, he has been a Postdoctoral Research Fellow with the Distributed Artificial Intelligence and Knowledge Laboratory (CIAD), UTBM, France. His research interests include computer vision, autonomous driving, and chronorobotics.



XIAOFEI CHANG received the Ph.D. degree in navigation, guidance, and control technology from Northwestern Polytechnical University (NPU), Xi'an, China, in 2010.

He is currently an Associate Professor with the School of Astronautics, Flight Control Institute, NPU, China. His research interest includes aircraft navigation and control.



HANG SU received the B.S. degree in electronic information engineering from Northwestern Polytechnical University (NPU), Xi'an, China, in 2011, and the M.S. degree in communication and information system from the University of Chinese Academic of Sciences (CAS), Beijing, China, in 2014, where she is currently pursuing the Ph.D. degree in astrometry and celestial mechanics. From 2017 to 2019, she was an Academic Guest with ETH Zurich, Switzerland.

Since 2014, she has been a Research Associate with the Orbit and Time Transfer Department, National Time Service Center, CAS, Xi'an. Her research interest includes GNSS data-based navigation and localization.

NATHAN CROMBEZ received the Ph.D. degree in computer vision for robotics from the University of Picardie Jules Verne, Amiens, France, in 2015.

He conducted his Postdoctoral Research at the MIS Laboratory. Since 2018, he has been an Assistant Professor (Maître de conférences) with the Distributed Artificial Intelligence and Knowledge Laboratory (CIAD), University of Technology of Belfort-Montbéliard (UTBM), France. His research interests are mainly focused on the perception and the navigation of robots and autonomous vehicles based on unconventional vision systems.

YASSINE RUICHEK (Senior Member, IEEE) received the Ph.D. degree in control and computer engineering and the Habilitation à Diriger des Recherches degree in physic science from the University of Lille, France, in 1997 and 2005, respectively.

Since 2007, he has been a Full Professor with the University of Technology of Belfort-Montbéliard (UTBM). His research interests are concerned with multisensory data-based perception and localization, including computer vision, pattern recognition and classification, machine learning, and data fusion, with applications to intelligent transportation systems and video surveillance.

TOMAS KRAJNIK received the Ph.D. degree in artificial intelligence and biocybernetics from Czech Technical University (CTU), Prague, Czech Republic, in 2012.

From 2013 to 2017, he was a Research Fellow with the Lincoln Center of Autonomous Systems (L-CAS), University of Lincoln, U.K. He is currently an Associate Professor with CTU. His research interests include mobile robotics, cybernetics, and chronorobotics.

ZHI YAN (Member, IEEE) received the Ph.D. degree from the University of Paris 8, France, in 2012.

From 2013 to 2015, he was a Postdoctoral Research Fellow with the CAR Team, IMT Lille Douai, France. From 2016 to 2017, he was also a Postdoctoral Research Fellow with the Lincoln Centre for Autonomous Systems (L-CAS), University of Lincoln, U.K., working on the Horizon 2020 Project FLOBOT. Since 2017, he has been an Assistant Professor (Maître de conférences) with the Distributed Artificial Intelligence and Knowledge Laboratory (CIAD), University of Technology of Belfort-Montbéliard (UTBM). His research interests are in autonomous driving, mobile robotics, and chronorobotics.