Raindrop Removal in a Vehicle Camera Video Considering the Temporal Consistency for Driving Support

Hiroki Inoue[1], Keisuke Doman[1], Jun Adachi[2] and Yoshito Mekada[1]
[1]*Graduate School of Engineering, Chukyo University, Toyota, Aichi, Japan*
[2]*Aisin Seiki Co., Ltd., Kariya, Aichi, Japan*

Keywords: Raindrop Removal, Vehicle Camera Video, Deep Learning, Optical Flow.

Abstract: This paper proposes a recursive framework for raindrop removal in a vehicle camera video considering the temporal consistency. Raindrops attached to a vehicle camera lens may prevent a driver or a camera-based system from recognizing the traffic environment. This research aims to develop a framework for raindrop detection and removal in order to deal with such a situation. The proposed method sequentially and recursively restores a video containing no raindrops from original one that may contain raindrops. The proposed method uses an output (restored) image as one of the input frames for the next image restoration process in order to improve the restoration quality, which is the key concept of the proposed framework. In each restoration process, the proposed method first detects raindrops in each input video frame, and then restores the raindrop regions based on the optical flow. The optical flow can be calculated in the outer part of the raindrop region more accurately than the inner part due to the difficulty of finding a corresponding pixel, which is the assumption for designing the proposed method. We confirmed that the proposed framework has the potential for improving the restoration accuracy through several preliminary experiments and evaluation experiments.

1 INTRODUCTION

Camera-based Driving Safety Support Systems (DSSS) have an important role as key techniques for reducing traffic accidents. One of the systems enables a driver to clearly see the surrounding environment, for example, by adjusting the image quality of a captured video and displaying the video on the side/rearview mirror or the monitor of a navigation system. Also, such a system detects objects and white lines on a road, and provide a driver with information according to the traffic scene.

One of the serious problems on such a system is that, in a rainy day, raindrops attached to a camera lens prevent a driver from recognizing the traffic environment. Raindrops could be obstacles and cause the oversight of important objects such as pedestrians. Attaching raindrops to a camera lens also causes the unstable behavior of autonomous driving systems, which may lead to fatal traffic accidents. It is necessary to develop a raindrop removal technique for both camera-based DSSSs and autonomous driving systems.

As for the solution for raindrop removal, a windshield wiper or an air spray can physically remove raindrops. Such physical devices are, however, not only difficult to be installed as add-on parts on a vehicle, but also easy to be broken. This research focuses on vision-based raindrop removal in a vehicle camera video.

Many methods for image restoration under bad weather conditions (e.g. fog and mist (Garg and Nayar, 2007; He et al., 2011; He et al., 2016), falling rain and snow (Garg and Nayar, 2007; Barnum et al., 2010)) have been proposed. They do not deal with raindrops on the surface of a camera lens, and cannot be directly applied to the task focused on this research. Qian et al. proposed a method for raindrop removal from a single image (Qian et al., 2018). The method can output an accurately-restored image. However, it cannot restore an image perfectly in principle, because it tries to restore from a single image, and consequently, cannot use the information on objects occluded by raindrops for image restoration. Xu et al. proposed a method for video inpainting (Xu et al., 2019). The method restores each frame in an input video considering temporal information, that is, the consistency of the bidirectional optical flow between adjacent frames. Note that the method does not detect obstacles (e.g. raindrops) but just inpaint

Inoue, H., Doman, K., Adachi, J. and Mekada, Y.
Raindrop Removal in a Vehicle Camera Video Considering the Temporal Consistency for Driving Support.
DOI: 10.5220/0009106504290436
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 4: VISAPP*, pages 429-436
ISBN: 978-989-758-402-2; ISSN: 2184-4321

manually-given missing regions.

This research tries to combine the methods described above, and improve the accuracy of image restoration. Accordingly, the method first detects raindrops in each input image by using a method based on the technique (Qian et al., 2018), and then restores each image considering the temporal consistency by using a method based on the technique (Xu et al., 2019). Here, as described in Section 3, we consider to use the output of the image restoration as a part of the next inputs for the restoration process, considering the spatial distribution of the restoration confidence. The proposed concept can gradually improve the quality of the image restoration over time. We also report experimental results that the proposed concept has the potential for improving the image restoration.

2 RELATED WORK

This section summarizes the related work on raindrop detection and image restoration.

2.1 Raindrop Detection and Removal from a Single Image

Kurihata et al. have proposed a PCA-based method for raindrop detection (Kurihata et al., 2005). The method learns the various shapes of raindrops within an eigenspace method, and detect raindrops by evaluating the similarity of eigendrops.

Qian et al. have proposed a deep learning-based method for raindrop removal. The method calculates an attention map and removes raindrops within a Generative Adversarial Network (GAN) (Goodfellow et al., 2014). The attention map is used to guide the discriminator to focus on the features of raindrops. Qian et al. reported that the method could restore an image accurately compared with other detection and restoration method. As described in Section 1, the method uses a single image.

Iizuka et al. have proposed a deep learning-based method for image inpainting (Iizuka et al., 2017). The method uses two types of classifiers, global and local classifiers, in order to take the scene context into account. Liu et al. have also proposed a method based on deep learning, which uses a partial convolution layer to gradually complete the missing regions and achieves the high accuracy of image restoration (Liu et al., 2018). Although these methods are effective for a single image, a method considering temporal consistency is required for better image restoration accuracy.

2.2 Raindrop Detection and Removal from a Video

You et al. have proposed a method for raindrop detection and removal (You et al., 2015). The method detects raindrops based on the temporal derivatives of a video, and removes raindrops based on a blending function and a video completion technique (Wexler et al., 2004). They reported that the method performed quantitatively better compared with the original method (Wexler et al., 2004). The resultant images restored by the method was, however, blurred and not accurate enough. Thus, further improvement is required.

Xu et al. also proposed a deep learning-based method, which is designed for video inpainting (Xu et al., 2019). The method first estimates and restores optical flow maps from an image sequence containing missing regions, and then interpolates each input image based on the restored optical flow. The method achieved higher image restoration accuracy, compared with other video inpainting methods (Huang et al., 2016; Newson et al., 2014). Also, the method can generate a visually-natural image for the complex background. We thus study an accurate image restoration method based on Xu's method.

3 METHOD

The raindrop removal framework of the proposed method is shown in Fig. 1. The proposed method sequentially and recursively restores a vehicle camera video containing no raindrops from original one that may contain raindrops. The proposed method uses an output (restored) image as one of the input frames for the next image restoration process in order to improve the image restoration quality, which is the key concept of the proposed framework. In each image restoration process, the proposed method first detects raindrops in each input video frame, and then restores the raindrop regions by using a technique for deep flow-guided video inpainting (Xu et al., 2019). The technique restores an input video frame based on the optical flow, and the optical flow can be calculated in the outer part of the raindrop region more accurately than the inner part due to the difficulty of finding a corresponding pixel, which is the assumption for designing the proposed method.

The overall framework and each step of the proposed method are described below.

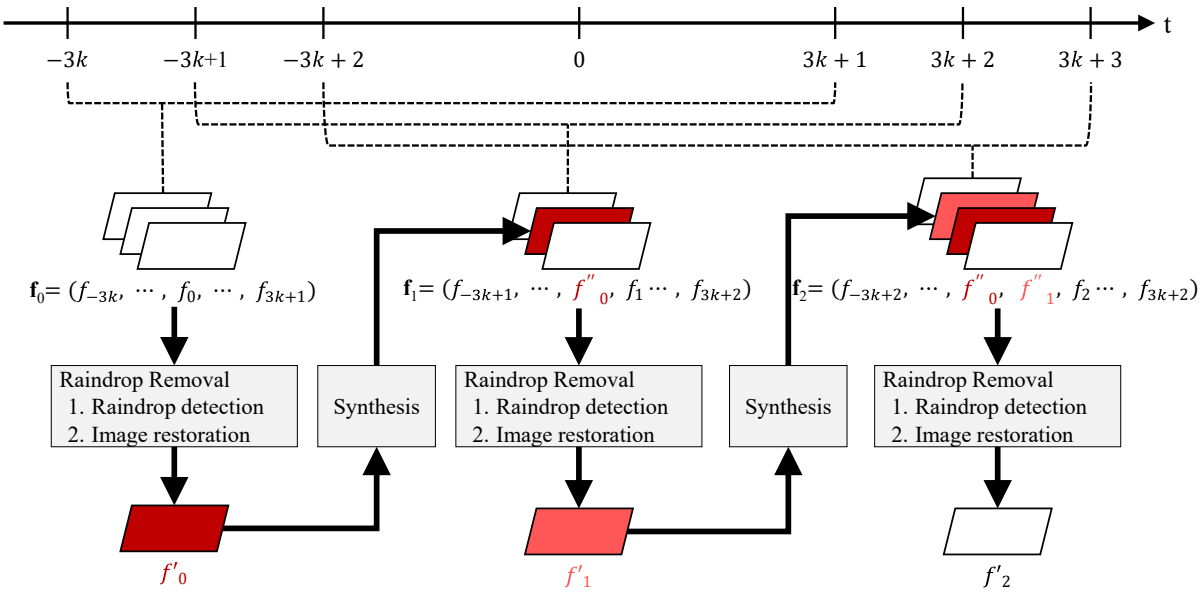


Figure 1: Proposed framework for raindrop removal based on recursive image restoration.

3.1 Overall Framework for Raindrop Removal

The proposed method restores a video frame containing no raindrops from its original adjacent frames, as shown in Fig. 1. The proposed method extracts video section between the $(-3k)$-th frame and the $(3k+1)$-th frame, and takes them as an input frame sequence. For example, in the case of k=5, 32 $(= 3\times5+1+3\times5+1)$ frames in total are input to the proposed method. The proposed method restores a video frame of interest from the frame sequence in the video section.

For the first input section $[-3k, 3k+1]$, the proposed method takes a frame sequence $\mathbf{f}_0 = (f_{-3k}, \ldots, f_0, \ldots, f_{3k+1})$ as its input, and makes its corresponding mask images $\mathbf{M}_0 = (M_{-3k}, \ldots, M_0, \ldots, M_{3k+1})$, which indicate the missing regions (raindrop regions). Then, the proposed method outputs restored image f'_0 from $\mathbf{f}_0$ with $\mathbf{M}_0$.

For the second input section $[-3k+1, 3k+2]$, the proposed method uses the refined image generated from the original f_0 and the first output f'_0 for better restoration accuracy.

Here, we assume that the quality of the image restoration is different between the outer and the inner areas of the restored region. That is, the outer the area is, the better the restoration quality is, because the image restoration should be easier in the outer part than the inner part. Thus, the proposed method makes the refined image f''_0 by synthesizing the original inner part of f_0 with the reliably-restored outer part of f'_0, and also makes its corresponding mask image M''_0, as shown in Fig. 2. The proposed method finally outputs a restored image f'_1 from $\mathbf{f}_1 = (f_{-3k+1}, \ldots, f''_0, f_1, \ldots, f_{3k+2})$ with its corresponding mask images $\mathbf{M}_1 = (M_{-3k+1}, \ldots, M''_0, M_1, \ldots, M_{3k+2})$.

In a similar manner, for the third input section $[-3k+2, 3k+3]$, the proposed method uses the first and the second outputs f''_0 and f''_1 instead of f_0 and f_1. That is, the proposed method outputs f'_2 from $\mathbf{f}_2 = (f_{-3k+2}, \ldots, f''_0, f''_1, f_2, \ldots, f_{3k+3})$ with its corresponding mask images $\mathbf{M}_2 = (M_{-3k+2}, \ldots, M''_0, M''_1, M_2, \ldots, M_{3k+3})$.

In summary, the proposed method uses the restored images with its corresponding masks instead of the corresponding original ones. This recursive framework can gradually improve the quality of image restoration over time.

3.2 Raindrop Detection

The proposed method detects raindrops in each image in the input section by using an Attentive Recurrent Network (ARN) (Qian et al., 2018). Although the network was originally proposed for generating an attention map toward raindrop removal, we consider to directly use the output of the ARN as the result of raindrop detection.

The network architecture of the ARN is shown in Fig. 3. The network has four time steps, and each time step is composed of three blocks: a five-layer ResNet (He et al., 2016), a convolutional LSTM (Xingjian et al., 2015), and a standard convolutional layer. The output of the network is in the range of [0, 1]. The

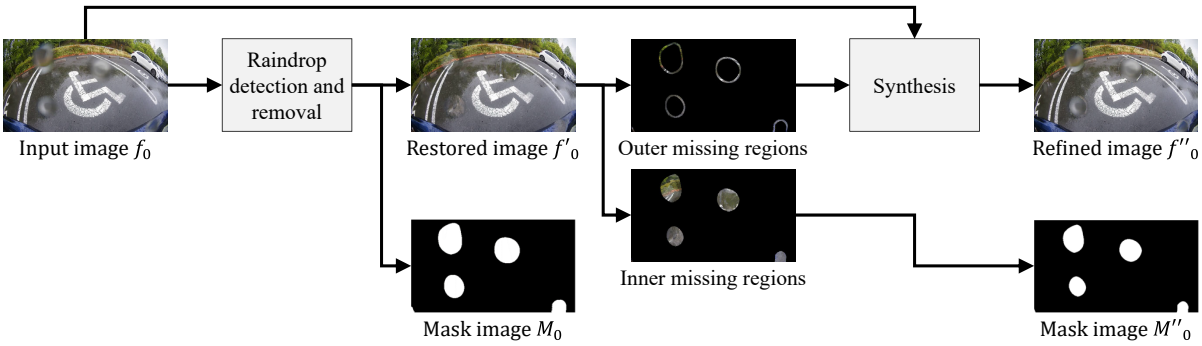


Figure 2: Process-flow of generating a refined image f''_0 by synthesizing the original inner missing parts of f_0 with the reliably-restored outer missing parts of f'_0.

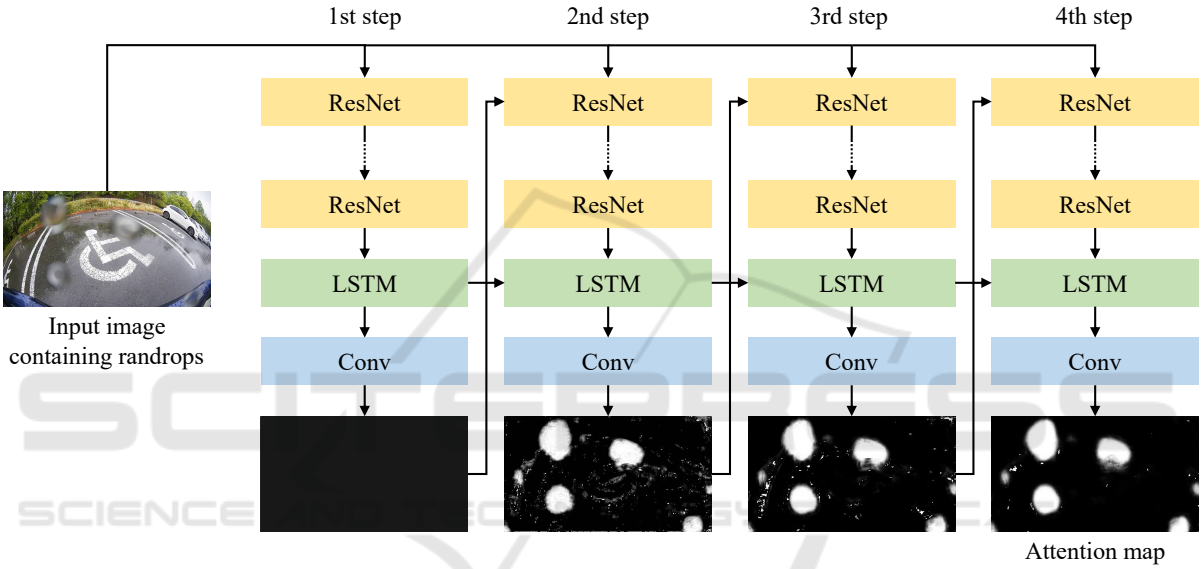


Figure 3: Architecture of the ARN (Qian et al., 2018) (In the proposed method, the output attention map is binarized to make the mask image for an input image).

higher the value is, the more attentive the region is. For each input image f_i ($i-3k \leq i \leq i+3k+1$), the proposed method makes the binary mask M_i by binarizing the output of the ARN, which indicates the missing regions (raindrop regions) to be restored.

3.3 Image Restoration

The proposed method uses a Deep Flow Completion Network (DFC-Net) (Xu et al., 2019) in order to restore a frame of interest f_i in an input video section $[i-3k, i+3k+1]$. The DFC-Net is composed of three subnetworks, and each subnetwork calculates one restored optical flow map for each sequence of $2k$+1 optical flow maps. An optical flow map $F_{i,i+1}$ is first calculated from an input frame sequence $\mathbf{f}_i = (f_{i-3k}, \ldots, f_i, \ldots, f_{i+3k+1})$ with its corresponding masks $\mathbf{M}_i = (M_{i-3k}, \ldots, M_i, \ldots, M_{i+3k+1})$. Here, missing (masked) regions in each frame of f_i are gradually completed and refined according to a coarse-to-fine manner through the subnetworks. Finally, a frame of interest f_i is restored based on the refined optical flow $F_{i,i+1}$. For more details, the refined optical flow $F_{i,i+1}$ is validated considering photometric consistency, and the pixel values in the missing regions are filled based on the flow using a inpainting technique (Yu et al., 2018).

The initial inputs for the first subnetwork are forward and backward optical flow maps $\mathbf{F}^{(0)}_{\mathrm{fi}} = (\mathrm{F}^{(0)}_{\mathrm{i-3k,i-3k+1}}, \ldots, \mathrm{F}^{(0)}_{\mathrm{i,i+1}}, \ldots, \mathrm{F}^{(0)}_{\mathrm{i+3k,i+3k+1}})$ and $\mathbf{F}^{(0)}_{bi} = (F^{(0)}_{i-3k+1,i-3k}, \ldots, F^{(0)}_{i+1,i}, \ldots, F^{(0)}_{i+3k+1,i+3k})$ calculated by using FlowNet 2.0 (Ilg et al., 2017) in addition to mask images $\mathbf{M}_i$. The first subnetwork then outputs refined forward and backward optical flow maps $\mathbf{F}^{(1)}_{fi} = (F^{(1)}_{i-2k,i-2k+1}, \ldots, F^{(1)}_{i,i+1}, \ldots, F^{(1)}_{i+2k,i+2k+1})$ and $\mathbf{F}^{(1)}_{bi} = (F^{(1)}_{i-2k+1,i-2k}, \ldots, F^{(1)}_{i+1,i}, \ldots, F^{(1)}_{i+2k+1,i+2k})$. The second subnetwork takes the outputs of the

first subnetwork, $\mathbf{F}_{fi}^{(1)}$ and $\mathbf{F}_{bi}^{(1)}$, and $\mathbf{M}_i$ as its inputs, and outputs more refined optical flow maps $\mathbf{F}_{fi}^{(2)} = (F_{i-k,i-k+1}^{(2)}, \ldots, F_{i,i+1}^{(2)}, \ldots, F_{i+k,i+k+1}^{(2)})$ and $\mathbf{F}_{bi}^{(2)} = (F_{i-k+1,i-k}^{(2)}, \ldots, F_{i+1,i}^{(2)}, \ldots, F_{i+k+1,i+k}^{(2)})$. In a similar manner, the third subnetwork refines the outputs of the second subnetwork, $\mathbf{F}_{fi}^{(2)}$ and $\mathbf{F}_{bi}^{(2)}$ with $\mathbf{M}_i$, and finally outputs the optical flow map $F_{i,i+1}$ corresponding to the input frame f_i.

4 EXPERIMENTS

We conducted an evaluation experiment following two kinds of preliminary experiments. The first preliminary experiment was to investigate the effectiveness of the raindrop detection and removal method without introducing the concept of the recursive image restoration described in Section 3. The second preliminary experiment was to confirm the validity of the assumption of the proposed concept. Finally, we evaluated the image restoration accuracy of the proposed framework described in Section 3.1 quantitatively and qualitatively.

In all the experiments, we used a vehicle camera video captured in a parking scenario in which the vehicle moved backward and stopped moving at a parking space between white lines. The camera was attached by the rear license plate, and its angle of view was 151 degrees. The image resolution was 1,920×1,080 pixels, and the frame rate was 6 fps. The details of each experiment are described below.

4.1 Preliminary Experiment 1: Evaluation on the Effectiveness of the Raindrop Detection and Removal Method

In the first preliminary experiment, we investigated the effectiveness of the raindrop detection and removal method without introducing the concept of the recursive image restoration described in Section 3.

4.1.1 Method

As for the module for the raindrop detection, the ARN was trained with 1,105 images containing raindrops and annotated with their regions. Here, the optimization function was Adam, and the loss function was the Mean Squared Error (MSE). The iteration of the ARN training was 500 epochs. In the test step, the ARN output (the attention map) was binarized with the threshold of 0.5 to generate the mask image for raindrops.

As for the module for the raindrop removal, the DFC-Net was fine-tuned with 10 parking-scene videos containing no raindrops and 10 mask images for simulating the regions missed by raindrops, based on the pre-trained model provided by Xu et al. [1]. The target optical flow in the training was calculated by FlowNet 2.0 from the 10 parking-scene videos without applying the mask images. Here, the optimization function was SGD, and the loss function was the Mean Absolute Error (MAE). The iteration of the DFC-NET training was 500 epochs

4.1.2 Results

Figure 4 shows the results of the raindrop detection and removal method without introducing the concept of the recursive image restoration described in Section 3. The method could accurately detect raindrops throughout the video. We can see, however, the method could not perfectly remove the raindrops. The experimental results showed both the effectiveness and the problem of the method without introducing the concept of the recursive image restoration toward raindrop detection and removal.

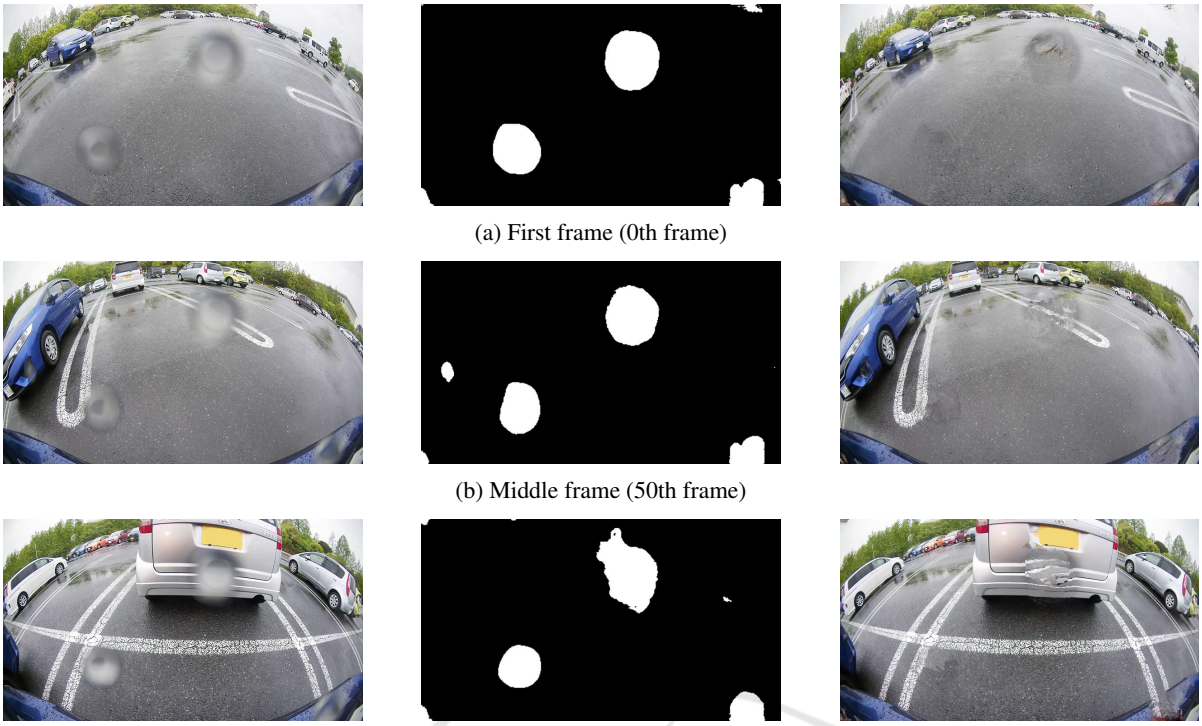
4.2 Preliminary Experiment 2: Investigation on the Validity of the Assumption of the Proposed Concept

In the second preliminary experiment, we investigated the validity of the assumption of the proposed concept, that is, the assumption that the accuracy of optical flow restoration in the outer part of the missing region is higher than the inner part.

4.2.1 Method

We calculated the optical flow maps using FlowNet 2.0 from nine vehicle videos with or without masking for simulating a missing region, and then restored each map using the DFC-Net. The mask here was a circle whose radius was 200 pixels. Its center circle area whose radius was 141 pixels was defined as the inner part, whereas the remaining part was defined as the outer part. Note that here the inner and the outer parts were the same area. Finally, we calculated the cosine similarity between the two maps in order to investigate the restoration confidence. The higher the similarity is, the higher the flow restoration accuracy

[1] https://github.com/nbei/Deep-Flow-Guided-Video-Inpainting



(a) First frame (0th frame)

(b) Middle frame (50th frame)

(c) Last frame (107th frame)

Figure 4: Examples of the results of raindrop detection and removal (Left: input image, Center: mask image (detected raindrop regions), Right: output image).

is. If the similarity in the outer part of a missing region is higher than that in the inner part, the assumption of the proposed concept can be regarded as valid.

4.2.2 Results

Table 1 shows the calculated cosine similarity for each of the inner and the outer parts. We can see that the similarities in the outer parts were generally higher than those in the inner parts. These results indicated that the optical flow calculated in the outer part was more confident, and consequently, the image restoration accuracy of the outer part should be higher than that of the inner part. We thus confirmed that the assumption of the proposed concept was valid.

4.3 Evaluation Experiment: Effectiveness of the Proposed Framework

We evaluated the image restoration accuracy of the proposed framework quantitatively and qualitatively with three vehicle videos containing no raindrops.

4.3.1 Method

We manually set mask images simulating missing regions by raindrops, and gradually reduced the missing regions over time by replacing with the pixel values of the original images. In this setting, we aimed to investigate the effectiveness (the improvement limit) of the proposed framework in the case of no raindrop detection error and no restoration error in the outer missing part. The mask reduction was performed by erosion with a 5×5 morphological kernel until the mask region disappeared completely. We evaluated the image restoration accuracy based on the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity (SSIM).

4.3.2 Results

Table 2 shows the restoration accuracy of the proposed method. As a reference, we also investigated the restoration accuracy of the conventional method (Xu et al., 2019). The examples of the restored images are shown in Fig 5.

The proposed method could improve the restoration accuracy compared with the conventional one, although this was strictly not a fair comparison because the proposed method used the manually-restored results in the outer part of the missing regions whereas the conventional one did not. Such a case would be realistic, considering the preliminary experimental re-

Table 1: Accuracy of optical flow restoration for each part of missing regions.

Video	Missing Region Position	Camera Movement	Cosine Similarity	
			Inner Part	Outer Part
1	Top left	Turn right	0.9969	**0.9992**
2	Top middle	Turn left	0.9939	**0.9999**
3	Top right	Turn right	0.9028	**0.9994**
4	Middle left	Turn right	0.9927	**0.9996**
5	Center	Turn left	0.9997	**0.9999**
6	Middle right	Turn right	0.9990	**0.9999**
7	Bottom left	Turn right	**0.9997**	0.9905
8	Bottom middle	Turn right	0.9994	**0.9999**
9	Bottom right	Turn left	0.9396	**0.9956**
Average			0.9804	**0.9982**

Table 2: Image restoration accuracy of the proposed method (with recursive restoration) and the conventional method (without recursive restoration) (Xu et al., 2019).

Video	Camera Movement	PSNR		SSIM	
		Proposed	Conventional	Proposed	Conventional
1	Move forward	**44.36**	29.15	**0.9867**	0.9464
2	Turn left	**42.67**	33.63	**0.9953**	0.9836
3	Turn right	**41.34**	28.62	**0.9866**	0.9533
Average		**42.79**	30.47	**0.9895**	0.9611

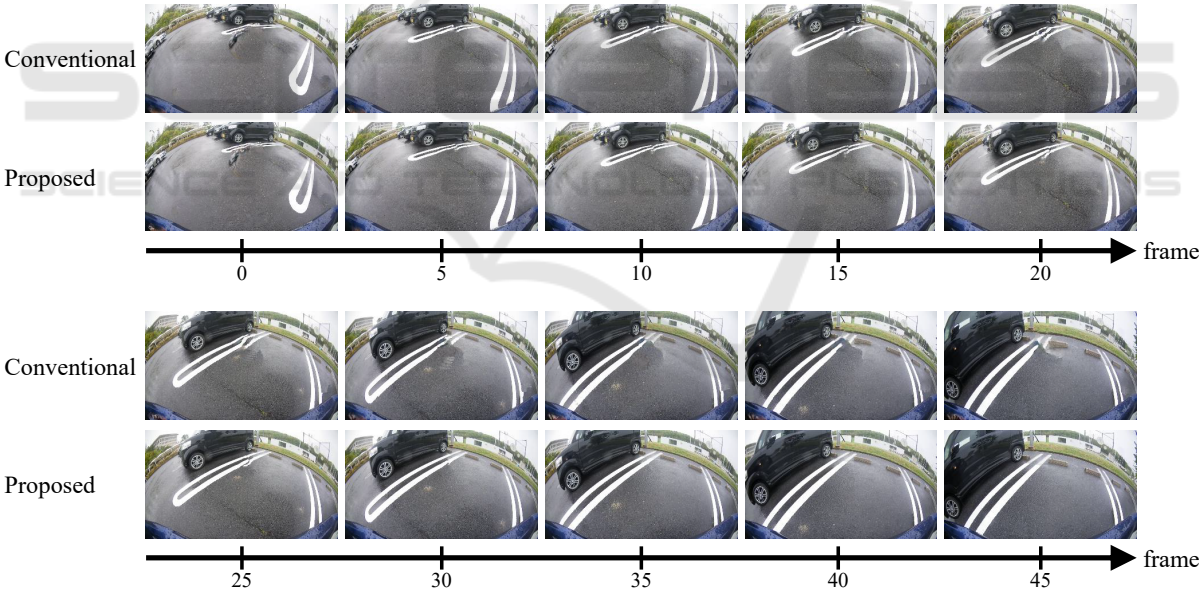


Figure 5: Comparison of the raindrop removal results: Proposed method (with recursive restoration) vs. Conventional method (without recursive restoration).

sults (Section 4.2) that the optical flow in the outer part of a missing region was relatively easy to be estimated. Therefore, the proposed framework has the potential for improving the restoration accuracy.

4.3.3 Discussion

We can see the improvement of the image restoration over time from Fig 5. This would be because the proposed framework recursively used the restored images. However, in a practical situation, the proposed method may not always output a perfectly-restored image, which should cause the decrease of the restoration accuracy due to the error propagation. We should also analyze the best way of reducing missing regions,

that is, how large can the missing regions be reduced. This parameter should be one of the factors affecting the accuracy improvement.

5 CONCLUSION

This paper proposed a recursive framework for raindrop removal in a vehicle video camera. The method first detects raindrops in each of an input image sequence by using a method based on the technique (Qian et al., 2018), and then restored each image considering the temporal consistency by using a method based on the technique (Xu et al., 2019). The results of the first preliminary experiment showed the effectiveness and the problem of the method without introducing the concept of the recursive image restoration toward raindrop detection and removal. The second preliminary experiment showed the validity of the assumption of the proposed concept, that is, the assumption that the accuracy of optical flow restoration in the outer part of the missing region is higher than the inner part. The results of the main evaluation experiments showed the proposed recursive framework has the potential for improving the restoration accuracy.

The future work includes the study on 1) how to deal with the error propagation and 2) how to reduce missing regions over time in the proposed recursive restoration. In addition, we will study a way for taking various possible situations into account, such as small vehicle motion and many raindrops attached to the lens of a camera, which may be the factors to decrease the accuracy of raindrop removal. Furthermore, the proposed method restores the middle frame of input frames. We will also investigate the restoration accuracy with the last frame of input ones in order to remove raindrops without delay.

REFERENCES

Barnum, P. C., Narasimhan, S., and Kanade, T. (2010). Analysis of rain and snow in frequency space. *International Journal of Computer Vision*, 86(2-3):256–274.

Garg, K. and Nayar, S. K. (2007). Vision and rain. *International Journal of Computer Vision*, 75(1):3–27.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

He, K., Sun, J., and Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.

Huang, J.-B., Kang, S. B., Ahuja, N., and Kopf, J. (2016). Temporally coherent completion of dynamic video. *ACM Transactions on Graphics*, 35(6):1–11.

Iizuka, S., Simo-Serra, E., and Ishikawa, H. (2017). Globally and locally consistent image completion. *ACM Transactions on Graphics*, 36(4):1–14.

Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., and Brox, T. (2017). FlowNet 2.0: Evolution of optical flow estimation with deep networks. In *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2462–2470.

Kurihata, H., Takahashi, T., Ide, I., Mekada, Y., Murase, H., Tamatsu, Y., and Miyahara, T. (2005). Rainy weather recognition from in-vehicle camera images for driver assistance. In *Proceedings of 2005 IEEE Intelligent Vehicles Symposium*, pages 205–210.

Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., and Catanzaro, B. (2018). Image inpainting for irregular holes using partial convolutions. In *Proceedings of 2018 European Conference on Computer Vision*, pages 85–100.

Newson, A., Almansa, A., Fradet, M., Gousseau, Y., and Pérez, P. (2014). Video inpainting of complex scenes. *SIAM Journal on Imaging Sciences*, 7(4):1993–2019.

Qian, R., Tan, R. T., Yang, W., Su, J., and Liu, J. (2018). Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2482–2491.

Wexler, Y., Shechtman, E., and Irani, M. (2004). Space-time video completion. In *Proceedings of 2004 IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 120–127.

Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c. (2015). Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems 28*, pages 802–810.

Xu, R., Li, X., Zhou, B., and Loy, C. C. (2019). Deep flow-guided video inpainting. In *Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3723–3732.

You, S., Tan, R. T., Kawakami, R., Mukaigawa, Y., and Ikeuchi, K. (2015). Adherent raindrop modeling, detection and removal in video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(9):1721–1733.

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., and Huang, T. S. (2018). Generative image inpainting with contextual attention. In *Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition*, pages 5506–5514.