Removing Reflection from In-vehicle Camera Image

Keisuke Inoue, Fumihiko Sakaue and Jun Sato
Nagoya Institute of Technology, Japan
{*inoue@cv., sakaue@, junsato@*}*nitech.ac.jp*

Keywords: In-vehicle Camera, Reflection Removal, Optical Flow.

Abstract: When taking images with an in-vehicle camera, objects in the vehicle are often reflected on the windshield due to sunlight, and they appear in the camera image. Since these reflections cause malfunction of autonomous driving systems, it is very important to remove the reflections from in-vehicle camera images. Thus, we in this paper propose a method for separating reflections and background scenes, and for generating images without reflections. Unlike the existing reflection removal methods, our method conducts the signal separation and the motion field computation simultaneously, so that we can separate images without using edge information. The efficiency of the proposed method is demonstrated by comparing with existing state-of-the-art methods.

1 INTRODUCTION

When we look at outdoor scenes through glass windows, we often have reflections from indoor objects on the glass windows, which make it difficult to see outdoor scenes. Such reflections cause serious problems for controlling autonomous vehicles using in-vehicle cameras. We often have strong reflections from in-vehicle objects at the windshield on sunny days, and they appear in images observed by in-vehicle cameras as shown in Fig. 1. If such a reflection occurs, the reflection may be misrecognized as an object in front of the vehicle, leading to a malfunction of the autonomous driving system such as an automatic braking system, which may cause a serious accident. Thus, we in this paper consider a method for removing such reflections in images.

In recent years, some methods have been proposed for separating the reflections from the background scenes in images. Xue et al. (Xue et al., 2015) used the optical flow estimation for removing reflection in images, and succeeded to separate complex reflections from background scene images. However, their method requires edge information in images for estimating optical flows, and thus if the change in intensity of reflection is not enough, it cannot separate image signals properly. More recently, some authors (Fan et al., 2017; Zhang et al., 2018) proposed reflection removal methods by using a deep learning technique. Although the efficiency of deep learning has been shown by these authors, their methods suffer from the domain shift problem, that is if the test images mismatch the training dataset, their networks no longer work properly.

Figure 1: Image captured by in-vehicle camera, which contains reflections on the windshield.

Thus, we in this paper propose a new method for separating reflections and back ground scenes accurately without using edge information and without using neural networks. The fundamental idea of our method is to conduct the image signal separation and the motion field computation simultaneously by minimizing a single cost function given a sequence of images. Although this is a very difficult problem, we make it tractable by representing the image motion field parametrically, and by using the property of in-vehicle camera. That is, the in-vehicle camera is always fixed at the same position in the passenger compartment and does not move relative to the vehicle. Thus, when taking an image with an in-vehicle camera attached to a running vehicle, the outside scene moves in the image, but the reflection of objects in the vehicle does not move. By using these properties, we separate the outside scene image and the reflected image from the inside of the vehicle efficiently. Our

Inoue, K., Sakaue, F. and Sato, J.
Removing Reflection from In-vehicle Camera Image.
DOI: 10.5220/0008941002220228
In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2020) - Volume 4: VISAPP*, pages 222-228
ISBN: 978-989-758-402-2; ISSN: 2184-4321

method does not rely on edge information in images, and does not use image prior learned from the training dataset. Thus, it works properly even if there is no strong edge information in reflection, and does not suffer from the domain shift problem.

2 RELATED WORK

In recent years, various approaches have been proposed for image signal separation. These methods fall into two classes, i.e. single image based methods and multiple image based methods.

The single image based methods are ill-posed, and hence it is necessary to combine a priori knowledge in these methods. Levin et al.(Levin et al., 2004) proposed a technique for separating reflections so that the brightness gradient of the recovered image approaches to that of the image in the natural image database. More recently, deep learning techniques are used for learning image features of reflection and separating reflections from images (Fan et al., 2017; Zhang et al., 2018; Wan et al., 2018). These methods enable us to separate reflections more accurately and faster than before. However, since these methods learn the image features of reflection based only on the training data, an enormous amount of training data is required to deal with various types of reflections in various scenes. As a result, these methods often suffer from the domain shift problem caused by the limited number of data. On the contrary, our method is based on the imaging model of reflection and does not rely on training dataset, so it does not suffer from the domain shift problem.

The multiple image based methods use the property that the motion of the background scene is different from that of the reflection on glass windows. Yu and Brown (Yu and Brown, 2013) proposed a method for separating images into foreground and background by matching image features using SIFT. Xue et al. (Xue et al., 2015) recovered dense motion fields from sparse motion fields obtained by edge information, and showed that the dense motion fields enable us to separate foreground and background images more accurately. Nandoriya et al. (Nandoriya et al., 2017) also used image edge information for obtaining initial motion field and for separating foreground and background information in video frames. However, all these methods require edge information for obtaining motion fields in images, and hence they cannot separate foreground and background information accurately, when we do not have abrupt change in intensity and cannot obtain image features in input images. On the contrary, our method does not require abrupt change in images and does not need to extract image features for separating image signals.

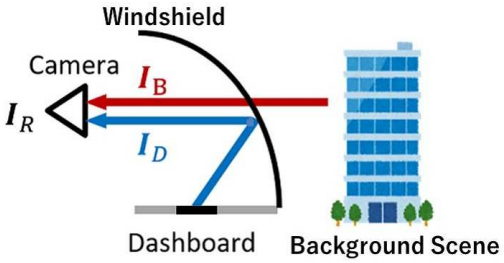


Figure 2: In-vehicle camera observes not only background scene, but also reflection of objects in vehicle.

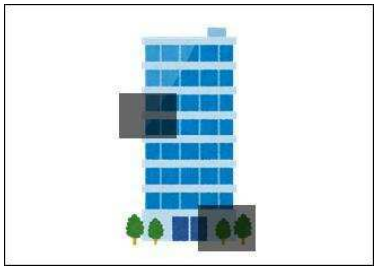


Figure 3: Observed image which contains reflection from objects in vehicle, i.e. two gray squares.

3 TRANSMISSION AND REFLECTION

We first consider an imaging model of an in-vehicle camera, in which reflected light and transmitted light are simultaneously captured in a single image.

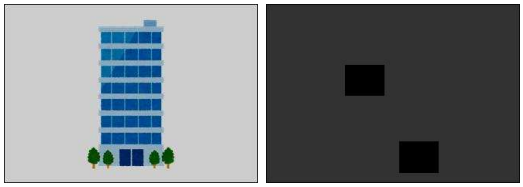
In the image of an in-vehicle camera, we often observe light reflected by the windshield of the vehicle $\mathbf{I}_B$ as well as light transmitted through the windshield $\mathbf{I}_D$. Fig. 2 shows the scene where two types of light $\mathbf{I}_B$ and $\mathbf{I}_D$ enter the in-vehicle camera, and Fig. 3 shows observed image $\mathbf{I}_R$ by the camera. The image $\mathbf{I}_R$ includes regions that are darker than the surrounding area. This is the reflection caused by objects in the vehicle.

Xue et al. formulated that the observed image is expressed by the alpha blending of the background image and the foreground image at a fixed ratio as follows:

$$\mathbf{I}_R(\mathbf{x}) = (1-\alpha(\mathbf{x}))\mathbf{I}_B(\mathbf{x}) + \alpha(\mathbf{x})\mathbf{I}_D(\mathbf{x}) \qquad (1)$$

where $\mathbf{x} = [x,y]^T$ is an image pixel, and $\alpha(\mathbf{x})$ denotes the mixing ratio at $\mathbf{x}$, which ranges from 0 to 1.

It seems that Eq. (1) is correct, but actually it is physically wrong, since in this model the light $\mathbf{I}_B$ from the outside of the vehicle is attenuated with $(1-\alpha)$, which does not happen in reality. Reflection is caused



(a) Background image (b) reflection image

Figure 4: Background image $\mathbf{I}_B$ and reflection image $\mathbf{I}_D$ derived from the observed image $\mathbf{I}_R$ in Fig. 3.

by the addition of the light $\mathbf{I}_D$ from the inside of the vehicle to the light $\mathbf{I}_B$ from the outside of the vehicle as shown in Fig. 2, and the light $\mathbf{I}_B$ from the outside of the vehicle is not attenuated. Also, for using this equation, we have to estimate not only two images, $\mathbf{I}_B$ and $\mathbf{I}_D$, but also a mixing parameter α, which is over parameterized for representing reflected light and transmitted light. Thus, we next consider an imaging model based on the reflection principle where light inside the vehicle is added to light outside the vehicle.

Let us consider the case where light from outside of the vehicle $\mathbf{I}_B$ and light from inside of the vehicle $\mathbf{I}_D$ are incident as shown in Fig. 3. If the light $\mathbf{I}_B$ and the light $\mathbf{I}_D$ are added to form an observed image, the intensity $\mathbf{I}_R$ of the observed image can be expressed as follows:

$$\mathbf{I}_R(\mathbf{x}) = \mathbf{I}_B(\mathbf{x}) + \mathbf{I}_D(\mathbf{x}) \tag{2}$$

Fig. 4 shows the background image $\mathbf{I}_B$ and the reflection image $\mathbf{I}_D$ when considered based on this model. The observed image in Fig. 3 is the sum of these two images according to Eq.(2). As is clear from the image in Fig.4 (b), the two gray squares in the observed image in Fig. 3 is generated not because the reflection of gray square objects exists, but because the intensity of the square objects is lower than the surrounding area. The reflection model in Eq. (2) accurately models this actual reflection process.

We next consider the difference in the characteristics of the external image $\mathbf{I}_B$ and in-vehicle image $\mathbf{I}_D$ over time. In the outside image $\mathbf{I}_B$, the position of 3D object in the image changes with the motion of the vehicle, so the brightness $\mathbf{I}_B(\mathbf{x})$ at a pixel $\mathbf{x}$ in the image changes with time. On the other hand, since the object in the vehicle is stationary relative to the in-vehicle camera, the intensity $\mathbf{I}_D(\mathbf{x})$ of the in-vehicle image at pixel $\mathbf{x}$ can be considered to be constant in a short period of time.

In the following sections, we propose a method for separating the outsize image $\mathbf{I}_B$ and the inside image $\mathbf{I}_D$ from the observed image $\mathbf{I}_R$ assuming that the inside image $\mathbf{I}_D$ is constant in a short period of time.

4 SEPARATING OUTSIDE AND INSIDE IMAGES

Suppose we have an observed image $\mathbf{I}_B^t$ at time t. Then, assuming that the optical flow does not change in a short period of time, the observed image $\hat{\mathbf{I}}_B^s$ at time s $(s \neq t)$ can be described by using the observed image $\mathbf{I}_B^t$ at time t as follows:

$$\hat{\mathbf{I}}_B^s(\mathbf{I}_B^t, \Delta\mathbf{x}) = \mathbf{I}_B^t(\mathbf{x} - (s-t)\cdot\Delta\mathbf{x}) \tag{3}$$

where $\Delta\mathbf{x}$ denotes the optical flow at $\mathbf{x}$.

Furthermore, we represent the optical flow of background scene motion parametrically by using affine transformation, that is *affine flow* (Sabater et al., 2012; Ju et al., 1996). In the affine flow, the motion vector $\Delta\mathbf{x} = [\Delta\mathbf{x}, \Delta\mathbf{y}]^\top$ at $\mathbf{x} = [x, y]^\top$ can be described as follows:

$$\begin{bmatrix} \Delta\mathbf{x} \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} \tag{4}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

where $\mathbf{A}$ denotes affine transformation matrix with six parameters, and we represent the whole optical flow in the image just by these six parameters. It seems that six parameters are not enough for representing whole scene flow. However, since the affine flow can represent divergence as well as rotation and translation, it can describe scene flow of in-vehicle camera images efficiently.

Now, let us consider three consecutive images, $\mathbf{I}_R^0(\mathbf{x})$, $\mathbf{I}_R^1(\mathbf{x})$ and $\mathbf{I}_R^2(\mathbf{x})$. From Eq. (3), these three images can be described by using reflection image $\mathbf{I}_D$, affine parameter $\mathbf{A}$ and background image $\mathbf{I}_B^0$ at time 0 as follows:

$$\mathbf{I}_R^0(\mathbf{x}) = \hat{\mathbf{I}}_B^0(\mathbf{I}_B^0, \Delta\mathbf{x}(\mathbf{A})) + \mathbf{I}_D(\mathbf{x}) \tag{6}$$

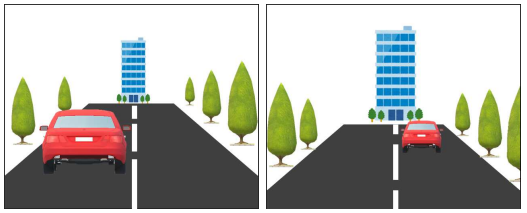
$$\mathbf{I}_R^1(\mathbf{x}) = \hat{\mathbf{I}}_B^1(\mathbf{I}_B^0, \Delta\mathbf{x}(\mathbf{A})) + \mathbf{I}_D(\mathbf{x}) \tag{7}$$

$$\mathbf{I}_R^2(\mathbf{x}) = \hat{\mathbf{I}}_B^2(\mathbf{I}_B^0, \Delta\mathbf{x}(\mathbf{A})) + \mathbf{I}_D(\mathbf{x}) \tag{8}$$

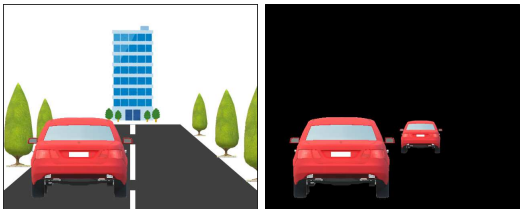
where $\Delta\mathbf{x}(\mathbf{A})$ denotes motion vector at $\mathbf{x}$ in the affine flow represented by an affine transformation matrix $\mathbf{A}$. Assuming that $\Delta\mathbf{x}(\mathbf{A})$ does not change even if the background images change in a short time, we can obtain constraints on optical flow in the background image. As a result, we can describe a series of observed images $\mathbf{I}_R^t$ $(t = 0, \cdots, 2)$using a background image $\mathbf{I}_B^0$, a reflection image $\mathbf{I}_D$ and an affine transformation $\mathbf{A}$.

Since the observed images $\mathbf{I}_R^t$ at three consecutive time are represented as shown in Eq. (6), Eq. (7) and Eq. (8), we can estimate $\mathbf{I}_B^0$, $\mathbf{I}_D$ and $\mathbf{A}$ by minimizing the following cost function E_c:

$$E_c(\mathbf{I}_B^0, \mathbf{I}_D, \mathbf{A}) = \sum_{t=0}^{2} ||\mathbf{I}_R^t - \hat{\mathbf{I}}_B^t(\mathbf{I}_B^0, \mathbf{A}) - \mathbf{I}_D||^2 \tag{9}$$



(a) Background image $\mathbf{I}_B^0$ (b) Background image $\mathbf{I}_B^1$



(c) Synthesized image $\hat{\mathbf{I}}_B^1$ (d) $||\mathbf{I}_B^1 - \hat{\mathbf{I}}_B^1||$

Figure 5: Separating images under the existence of independent motions.

In Eq.(9), we use three observations $\mathbf{I}_R^t$ $(t = 1, \cdots, 2)$, since we estimate two images, $\mathbf{I}_B^0$ and $\mathbf{I}_D$, and an affine transformation.

5 SEPARATING IMAGES UNDER INDEPENDENT MOTIONS

In section 4, we estimated outside image $\mathbf{I}_B$ and inside image $\mathbf{I}_D$ by assuming that the entire optical flow in the image can be represented by a single affine transformation. Although this assumption is valid when we observe a static scene from a moving vehicle, it is no loner valid if we have some independently moving objects in the scene. Thus, in this section, we extend the method described in section 4, and propose a method for extracting outside image and inside image from the observed image under the existence of multiple independent motions.

Suppose we have background images, $\mathbf{I}_B^0$ and $\mathbf{I}_B^1$ and affine flow $\mathbf{A}$ which are obtained by using the proposed method, as shown in Fig. 5 (a) and (b). Then the background image $\hat{\mathbf{I}}_B^1$ can be synthesized by using $\mathbf{I}_B^0$ and $\mathbf{A}$ according to Eq. (3) as shown in Fig. 5 (c). Since the affine flow $\mathbf{A}$ represents the background scene motion, the static objects, such as buildings and trees, in $\mathbf{I}_B^1$ and $\hat{\mathbf{I}}_B^1$ coincide with each other. However, if we have independently moving objects, such as vehicles, they do not match in $\mathbf{I}_B^1$ and $\hat{\mathbf{I}}_B^1$ as shown in Fig. 5 (b) and (c). Thus, we can extract independently moving regions as shown in Fig. 5 (d) by extracting image pixels which hold the following inequality:

$$||\mathbf{I}_B^1 - \hat{\mathbf{I}}_B^1(\mathbf{I}_B^0, \mathbf{A})|| \geq th \tag{10}$$

where, th is a threshold value. Then, by minimizing the cost function E_c on this region, the outside image $\mathbf{I}_B$ and the inside image $\mathbf{I}_D$ as well as the affine flow $\mathbf{A}$ of the independently moving region can be obtained. By applying these procedures iteratatively, we can extract outside images and inside images from captured images. Hence, we extract outside image $\mathbf{I}_B$, inside image $\mathbf{I}_D$ and affine flows of N regions $\mathbf{A}_R = \{\mathbf{A}_1, \cdots, \mathbf{A}_N\}$ by solving the following minimization problem:

$$\{\hat{\mathbf{I}}_B^0, \hat{\mathbf{I}}_D, \hat{\mathbf{A}}_R\} = \underset{\mathbf{I}_B^0, \mathbf{I}_D, \mathbf{A}_R}{\arg\min} \sum_{i=1}^{N} E_c(\mathbf{I}_{Bi}^0, \mathbf{I}_{Di}, \mathbf{A}_i) + \alpha||\mathcal{L}(\mathbf{I}_B^0)||^2 + \beta||\mathcal{L}(\mathbf{I}_D)||^2 \tag{11}$$

where, $\mathcal{L}(\cdot)$ denotes the Laplacian for smoothness constraints, and α and β is its weight.

6 SEGMENTATION IMAGE

For separating outside and inside images more accurately, we further introduce image segmentation into our method.

In Sec 2.2, we assumed that the intensity of the outside image changes with time, while that of the inside image does not change. However in the real scene image, the image motions of some outside objects, such as road, are not observable since they do not have enough texture on their surface. As a result, these objects are classified into inside objects and appear in the inside image. For solving this problem, we utilize image segmentation, and set additional constraints to be classified into inside intensity when the image point belongs to texture-less objects.

More specifically, we first classify image pixels in an image into four categories, “Buildings”, “Road”, “Sky”, and “Others” by using an image segmentation network, and “Buildings”, “Road” and “Sky” are marked as texture-less area. We next define a maximum intensity $\mathbf{I}_D^{max}(\mathbf{x})$ for inside image $\mathbf{I}_D$ at an image pixel $\mathbf{x}$ as follows:

$$\mathbf{I}_D^{max}(\mathbf{x}) = \lambda \mathbf{I}_R(\mathbf{x}) \tag{12}$$

where, λ is a scalar in the range of $[0,1]$. If an image pixel $\mathbf{x}$ is in a texture-less area, then inside image intensity $\mathbf{I}_D(\mathbf{x})$ at $\mathbf{x}$ is estimated in the range of $[0, \mathbf{I}_D^{max}(\mathbf{x})]$, and if $\mathbf{x}$ is not in a texture-less area, $\mathbf{I}_D(\mathbf{x})$ is estimated in the range of $[0,1]$ by using Eq.(11). In this way, we can avoid outside objects, such as road and sky, being classified as inside objects and appearing in the inside image. Fig. 6 shows an overview of this algorithm.

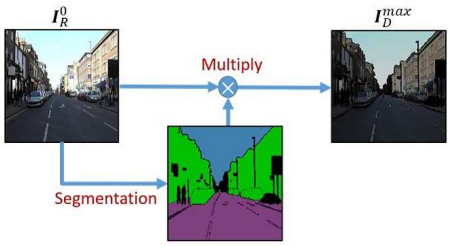


Figure 6: Overview of generating $\mathbf{I}_D^{max}(\mathbf{x})$ for a texture-less area.

Figure 7: Experimental setup of in-vehicle camera and reflection on windshield.

7 EXPERIMENTS

We next show the results of generating reflection removal images using our proposed method in various scenes.

In our experiments, road scene images were taken by an in-vehicle camera. As shown in Fig. 7, the in-vehicle camera was attached on the windshield so that the reflection from the object on the dashboard appears in the image. We took three consecutive images every 0.1 seconds in the driving scene, and cropped these images with the size of 200×200. To obtain the segmentation images, we used a standard image segmentation network based on pix2pix (Isola et al., 2016), which was trained on Cityscapes dataset (Cordts et al., 2016).

Fig. 8 (a), (b) and (c) show three sequential images of four difference scenes observed by the in-vehicle camera. As shown in these images, inside objects were reflected by the windshield and appear in the observed images with outside scenes. These sequential images were used for extracting outside image and inside image. The outside and inside images obtained from our method are shown in Fig. 8 (d) and (e) respectively. As shown in these images, outside scenes and inside objects are separated appropriately in these images by using our method. Fig. 8 (f) shows optical flow estimated by our method. As we can see in these images, the estimated optical flow represents the movement of the buildings, road and trees appropriately.

We next compare our method with two state-of-the art reflection removal methods. The first one is Xue's method (Xue et al., 2015) which uses edge information for computing optical flows and separating signals. This method can separate signals clearly when we have sharp reflection in images, but it degrades when the reflected objects are vague in images. The second one is Zhang's method (Zhang et al., 2018) which is based on a trained neural network.

Fig. 9 compares the results of our method with those of Xue's method and Zhang's method. As shown in this figure, Xue's method cannot remove the reflection properly, since the reflected objects do not have sharp edges in the observed images, and the motion field cannot be obtained accurately in their method. Zhang's method succeeded to remove reflection partially, but not all of the reflection was removed accurately. Since the efficiency of learning based methods heavily depends on the dataset used in the learning stage, Zhang's method cannot remove reflection accurately, when the input images are taken under different domains.

Finally, we show results from images which contain independent motions. Fig. 10 (a) show sequential images at $t = 0$ and $t = 2$. As shown in these images, the vehicle is moving forward while the bicycle and the pedestrian are moving independently. Fig. 10 (b) shows reflection removal images obtained from the proposed method, and Fig. 10 (c) shows optical flow estimated at the same time. As shown in Fig. 10 (b), the reflection at the center of the image was eliminated accurately even if we have independent motions in images. Fig. 10 (c) shows that both the background scene motions and the independent motions are estimated accurately in the proposed method.

From these results, we find that the proposed method works efficiently under various situations.

8 CONCLUSIONS

In this paper, we proposed a novel method for separating background scenes and reflections from images observed by an in-vehicle camera.

Our method estimates background scene images, reflection images and scene flows simultaneously, representing the scene flows parametrically by using affine transformations. The method does not reply on edge information in images, and hence it works efficiently even if the reflections do not have sharp image features unlike the existing methods.

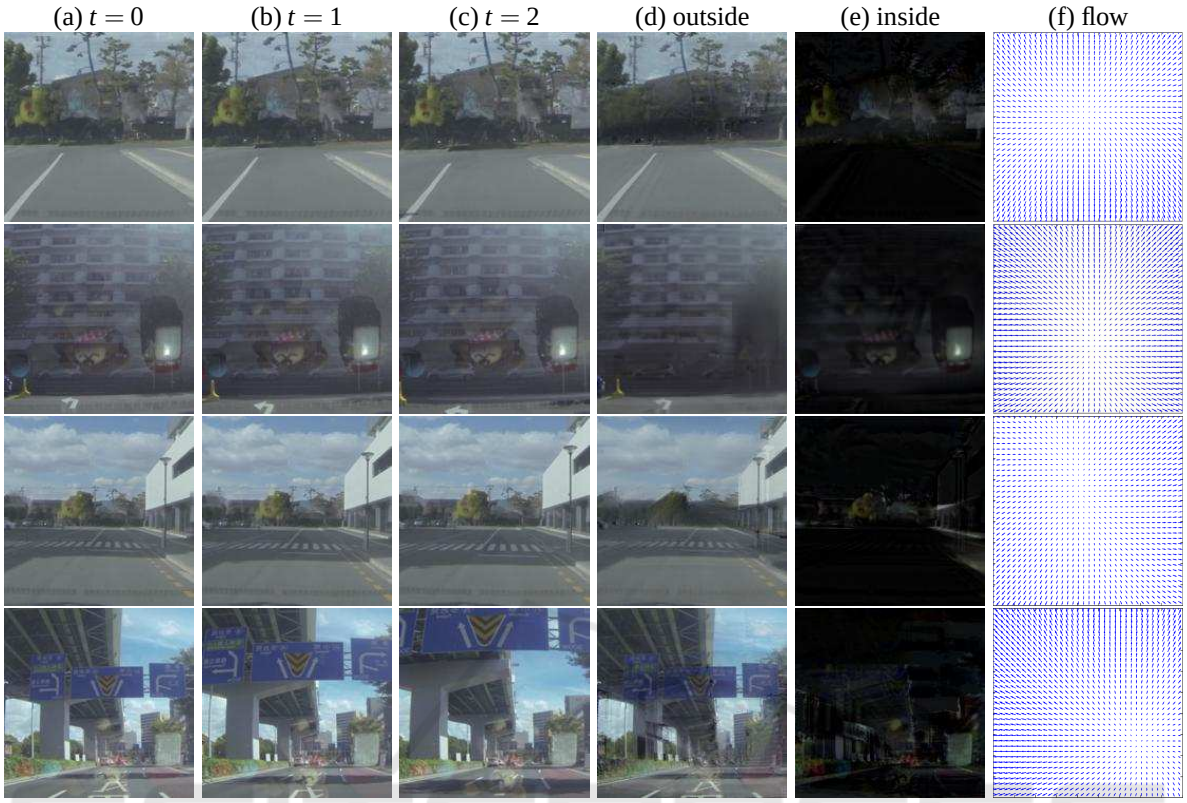


Figure 8: Result of signal separation from in-vehicle image into outside image and inside image. (a), (b) and (c) show three sequential images obtained by in-vehicle camera. (d) and (e) show outside image and inside image obtained by the proposed method from images in (a), (b) and (c). The estimated optical flow is shown in (f).

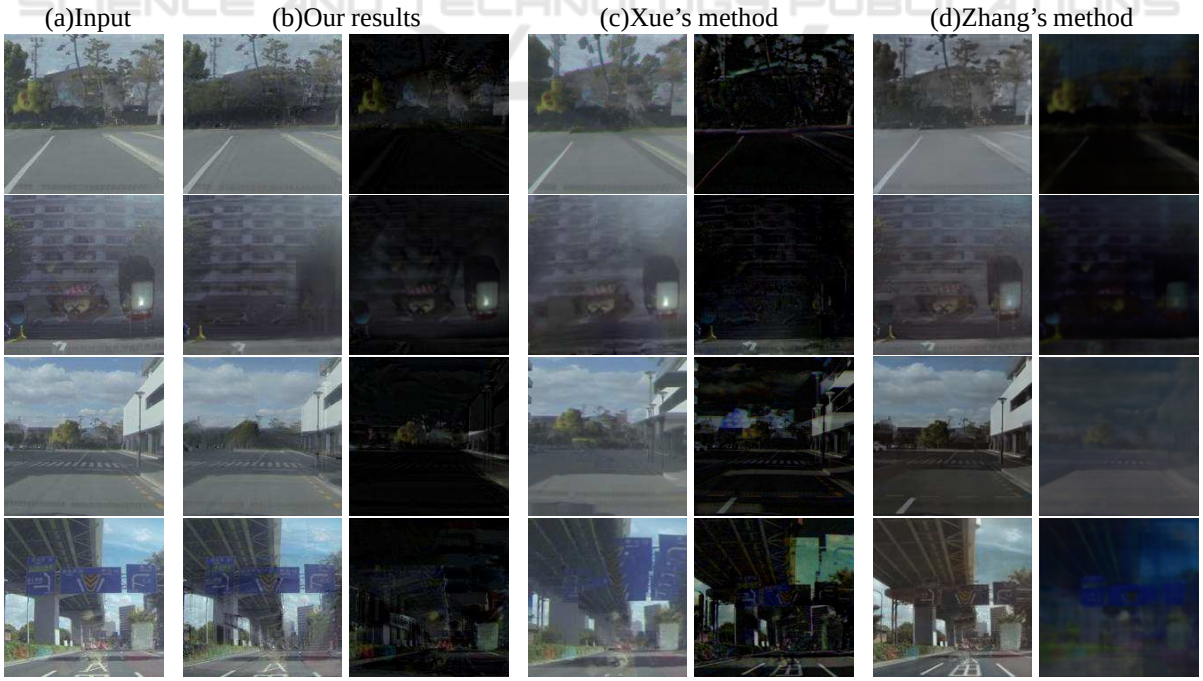


Figure 9: Comparison of reflection removal images obtained by our proposed method, Xue's method (Xue et al., 2015) and Zhang's method (Zhang et al., 2018).

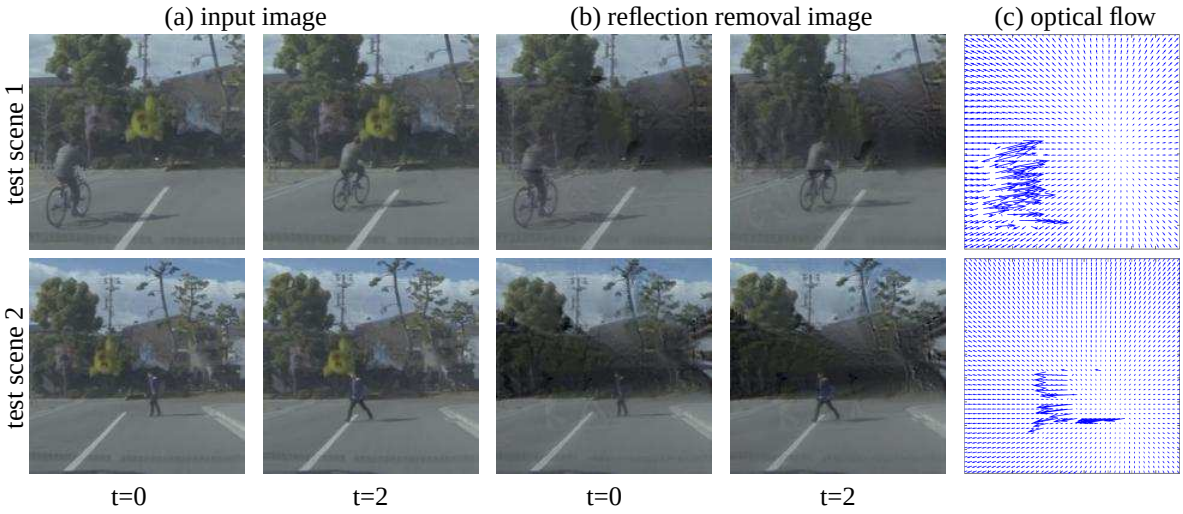


Figure 10: Results of generating reflection removal image and computing optical flow under the existence of independent motions.

The experimental results show that the proposed method outperforms the existing state-of-the-art methods.

REFERENCES

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Fan, Q., Yang, J., Hua, G., Chen, B., and Wipf, D. (2017). A generic deep architecture for single image reflection removal and image smoothing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3238–3247.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2016). Image-to-image translation with conditional adversarial networks. *arxiv*.

Ju, S. X., Black, M. J., and Jepson, A. D. (1996). Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*, pages 307–314. IEEE.

Levin, A., Zomet, A., and Weiss, Y. (2004). Separating reflections from a single image using local features. In *ECCV*.

Nandoriya, A., Elgharib, M., Kim, C., Hefeeda, M., and Matusik, W. (2017). Video reflection removal through spatio-temporal optimization. In *The IEEE International Conference on Computer Vision (ICCV)*.

Sabater, N., Leprince, S., and Avouac, J.-P. (2012). Contract invariant and affine sub-pixel optical flow. *International Conference on Image Processing*.

Wan, R., Shi, B., Duan, L.-Y., Tan, A.-H., and Kot, A. C. (2018). Crrn: Multi-scale guided concurrent reflection removal network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4777–4785.

Xue, T., Rubinstein, M., Liu, C., and Freeman, W. T. (2015). A computational approach for obstruction-free photography. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 34(4).

Yu, L. and Brown, M. (2013). Exploiting reflection change for automatic reflection removal. pages 2432–2439.

Zhang, X., Ng, R., and Chen, Q. (2018). Single image reflection separation with perceptual losses. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4786–4794.