ADHERENT RAINDROPS DETECTION AND REMOVAL IN VIDEO

動画像におけるレンズ上の雨滴の認識と除去

SHAODI YOU

尤　少迪

A MASTER DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL OF
THE UNIVERSITY OF TOKYO

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF INFORMATION SCIENCE AND TECHNOLOGY

August 2012

Thesis Supervisor: Katsushi IKEUCHI　池内 克史

ABSTRACT

Raindrops appeared on windscreens or window glass can degrade the visibility of the outside scenes. If we can detect and later remove the raindrops, many applications such as intelligent vehicle system will benefit from it.

In this thesis, we intend to focus on developing a method of automatic raindrop detection. We mathematically analyze the characteristics of raindrops adhered to a clear transparent surface, particularly the characteristics of the motion and the changes of the intensity values. The analysis is motivated by the fact that a raindrop conceives the information of the whole background scene, behaving like a minuscule catadioptric lens. In conducting the analysis, we found that motion of raindrop pixels is slower than that of non-raindrop pixels, and the change of intensity of raindrop pixels is smaller than that of non-raindrop pixels. Based on these findings, we developed our detection method. Experimental results using real videos show the effectiveness of our method.

Having detected raindrops, in this thesis, algorithms to automatically repair the video in complex outdoor environment are proposed. We categorize outdoor environments into 3 situations and propose a video repairing algorithm for each of them. First, for a fast motion, structured and textured video, a sparse matching based algorithm to estimate dense motion is described. A classification based on the dense motion is introduced and raindrop removal are proposed. Second, for a slow motion, structured and textured video, a raindrop removal method using image registration is introduced. Third, for a static and non-structured video, a modified image inpainting method is introduced for raindrop removal. Experimental results using real videos show the effectiveness of our method.

Acknowledgements

First and foremost, I would like to express my gratitude to my advisor, Prof. Katsushi Ikeuchi, for his supervision. He shares his vast knowledge and experience and insightful thoughts through the weekly discussion. While keeping me in track, he also gives freedom to follow my own interests. I have really enjoyed doing research during my master's.

I would like also express my gratitude to my mentors: Dr. Robby T. Tan at Utrecht University, The Netherlands and Dr. Rei Kawakami at The University of Tokyo. Dr. Tan uses to be very strict but patient, I have learned from him in every detailed aspects of doing research. Besides I have also learned a lot from him in finding the personal values in becoming a researcher. Dr. Kawakami also shares her vast knowledge and experience in research. She gives me suggestions in research life. As a native, she also helps me a lot in Japanese culture.

Many thanks go to the people that I have been working with, the current and former members of the Computer Vision Laboratory at the University of Tokyo. Special thanks go to Dr. Takeshi Oishi who gives me suggestion and support in career planning; Dr. Shintaro Ono, who gives me suggestions and support in Tohoku project; Boun Vinh Lu, who shares his codes and suggestions; Bjoern Rennhak and Boxin Shi in proof reading. I am also very proud of working with all the Photometry group members. Although, due to limited space, I cannot name everyone who has helped me, I am very grateful to all the people I have met in this lab.

Finally, I would like to thank my family and my friends for constant support in all my life.

August 2012
Shaodi You

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Outdoor vision system is used for various tasks such as navigation and surveillance. It can be adversely affected by bad weather conditions. In a rainy day, it is inevitable that raindrops will appear on the windscreen, camera lens, or the protecting shield. These adherent raindrops will cause large area of data to be missing. Because of this, the performances of many algorithms of outdoor vision systems (such as feature detection, tracking, stereo correspondence, etc.) will be significantly degraded.

Especially, in order to digitally archive the 2011 Japan Earthquake, our lab uses car-mounted video camera to record street views in the earthquake area. Some of the video is taken in rainy day. Raindrops adhered to camera lens cause large area of data missing. Performance of pro-processing computer vision tasks such as object detection, image registration, video stabilization and frame interpolation are significantly degraded.

1.2 Motivation and Goals

To make the video with adherent raindrops visually satisfying and to enhance the performance of computer vision tasks, it is essential to remove the raindrops.

There are 2 main steps of adherent raindrops removal.:

1. Raindrops detection.
2. Video Repairing.

For raindrops detection, algorithms to automatically detection raindrops with any shape and size are demanded. Above that, efficient method which could detect raindrops in real-time or nearly real-time is demanded.

For video repairing, algorithm which could handle complex outdoor environment is demanded. The repairing algorithm should robustly repair video with both spatially large and temporally large data missing. The algorithm should work well on fast changing video, slowly changing video and static image. The video should work well on both textured and non-textured video. Above that, computational efficiency is also required for large scale data.

1.3 Approaches to Achieve The Goal

1.3.1 Raindrops Detection in Video

For raindrops detection, while all the previous methods try to model each raindrop as a unit object, we model raindrops locally from its derivative properties. Modeling a whole raindrop needs too many parameters which are impractical when those parameters are unknown; on the contrary, modeling the derivative properties needs only few parameters.

In this thesis, we model the derivative properties between a raindrop pixel and non-raindrop pixel that are originated from the same point in the environment. We observed that the imagery of an adherent raindrop is in fact the contraction of the environment. Based on this, we theoretically found that the contraction ratio is at least $\frac{1}{8}$ in any place on the raindrop. Based on this property, we propose the raindrop detection method by using dense motion estimation (e.g. optical flow). We also found that each pixel in a raindrop represents at least 64 pixels in the environment. Because of this ratio, raindrop area changes less compared to the environment. Relying on this analysis, we propose a detection method based on the change of intensity.

1.3.2 Video Completion

For video repairing, to handle complex outdoor environment raindrop removal tasks are divided into 3 situations and solved separately.

Firstly, for fast motion, structured and textured area, theory and methodology to estimate dense motion in video with data missing are proposed. Raindrop removal sit-

uation classification based on the dense motion are introduced. Then, raindrop removal for fast motion, structured and textured area by using dense motion are proposed.

Secondly, for slow motion, structured and textured area, raindrop removal method using image registration are introduced.

Lastly, for static and non-structured area, a modified image inpainting method is introduced for raindrop removal.

1.4 Contributions

In this thesis, we proposed a complete algorithm to remove adherent raindrops in video.

For raindrops detection:

I. We propose algorithm which could detect raindrops with any size and shape.

II. Accuracy of our algorithm outperforms all existing algorithms.

III. Our proposed real-time computational efficiency which is essential for many outdoor vision tasks.

For video repairing:

I. We propose algorithm which could repair video with both spatially and temporally large missing area.

II. We propose case by case solution which could handle complex situations (complex motion and complex structure) in outdoor vision system.

III. Computational efficiency is achieved by using the proposed sparse matching based motion estimation.

1.5 Thesis Overview

Chapter 2 introduces background knowledge and related works on raindrop detection and video completion. In the first section, adherent raindrops are categorized into several situations according to there appearance; situations will be solved in this thesis is specified. In the second section, background knowledge on imagery model of adherent raindrop is described; related works on raindrop detection are introduced. In the third section, background knowledge on spatial-temporal space which is a useful tool for video analysis is introduced; related works on video completion are introduced

along with the analysis in spatial-temporal space. Lastly, adherent raindrops is analyzed spatial-temporal space. Based on the analysis, tasks on raindrops removal are specified.

Chapter 3 is the theory and implementation on raindrop detection. Firstly, we theoretically estimate the linear expansion ratio of φ. Based on it, we propose dense motion based detection method. Then, we theoretically estimate the area expansion ratio and the intensity change based detection method is thus proposed. Lastly, the methodology is described.

Chapter 4 is the theory and implementation on raindrop removal (video completion). Raindrop removal tasks are divided into 3 situations and solved separately. Firstly, for fast motion, structured and textured area, theory and methodology to estimation dense motion in video with data missing are described. Raindrop removal situation classification based on the dense motion are introduced. Then, raindrop removal for fast motion, structured and textured area by using dense motion are proposed. Secondly, for slow motion, structured and textured area, raindrop removal method using image registration are introduced. Lastly, for static and non-structured area, a modified image inpainting method is introduced for raindrop removal.

Chapter 5 is the experiments. Firstly, our raindrop detection method are tested on both experimental data set and real data set. Quantitative analysis are provided on experimental data set. To demonstrate the effectiveness, our method are compared with 2 existing raindrop detection method. Secondly, our raindrop detection method are tested on both experimental data set and real data set. Quantitative analysis are provided on experimental data set. To demonstrate the effectiveness, our method are compared with 3 existing raindrop detection method.

Chapter 6 is the conclusion.

Chapter 2

Background Knowledge, Problem Formation and Related Works

Before proposing methods to detect and remove adherent raindrops in video. In this chapter, we explicitly form the problems.

In the first section, adherent raindrops are categorized into several situations according to there appearance. Then, situations will be solved in this thesis is specified.

In the second section, background knowledge on imagery model of adherent raindrop is described. Then, several related works on raindrop detection are introduced.

In the third section, background knowledge on spatial-temporal space which is a useful tool for video analysis is introduced. Then, related works on video completion are introduced along with the analysis in spatial-temporal space. Lastly, we analyze adherent raindrops in spatial-temporal space. Based on the analysis, we specify the tasks on raindrops removal.

2.1 Categorization of Adherent Raindrops

As shown in Fig. 2.1, raindrops are totally transparent, which the appearance totally depends on the environment and in most cases varies in terms of intensity. Before proposing methods to detecting and removing raindrops in video, in this section, we first categorize the adherent raindrops according to its appearance. Based on the categorization, we form the raindrop removal problems on specific categories.

2.1.1 Categorization of Adherent Raindrops

Based on a sufficient collection, the appearance of adherent raindrops are categorized based on the following aspects:

I. Shape.

II. Distribution.

III. Scene and background.

IV. Camera condition.

Shape.

Unlike falling raindrops, whose shapes are always nearly round, the shape of adherent raindrops various according to the surface they adhered to. Exactly categorizing and modeling the shape according to the surface which raindrops adhere to is under the scope of physics. Here, for raindrop detecting and removal, we categorize the shapes according to their appearance.

According to the appearance, the shapes of raindrops are categorized as:

I. Round

II. Non-round patch

III. Streaks

IV. Mixed.

Distribution.

The density of raindrops adhere to a given surface depends on the precipitation rate and the time interval that raindrops are collected. With other parameters fixed, the heavier the rain is or the longer raindrops are collected, the denser the adherent raindrops are.

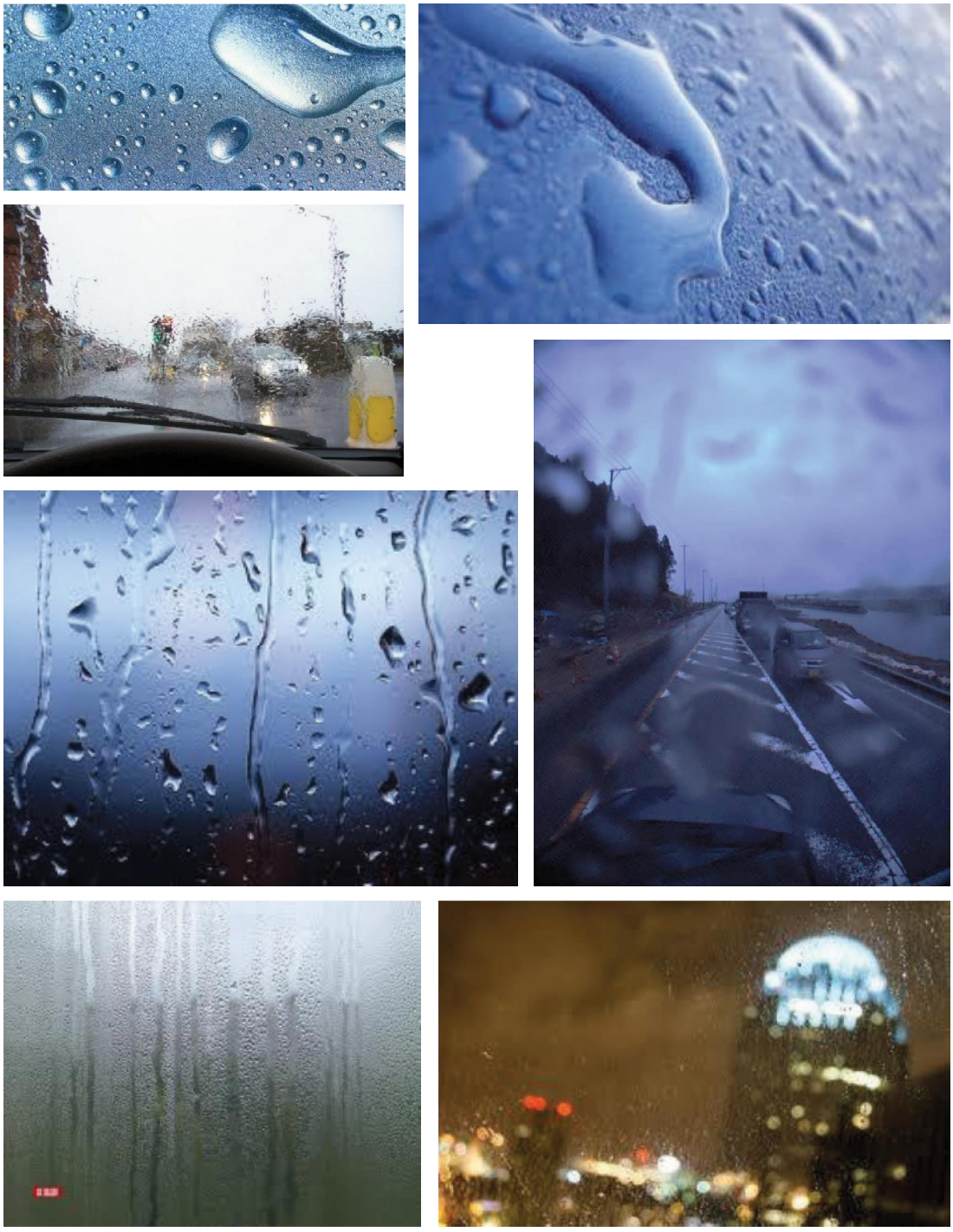


Figure 2.1: A collection of appearances of adherent raindrops.
These figures are the courtesy of Kenna Westerman, Vladimir Voronin and Nans photo challenge.

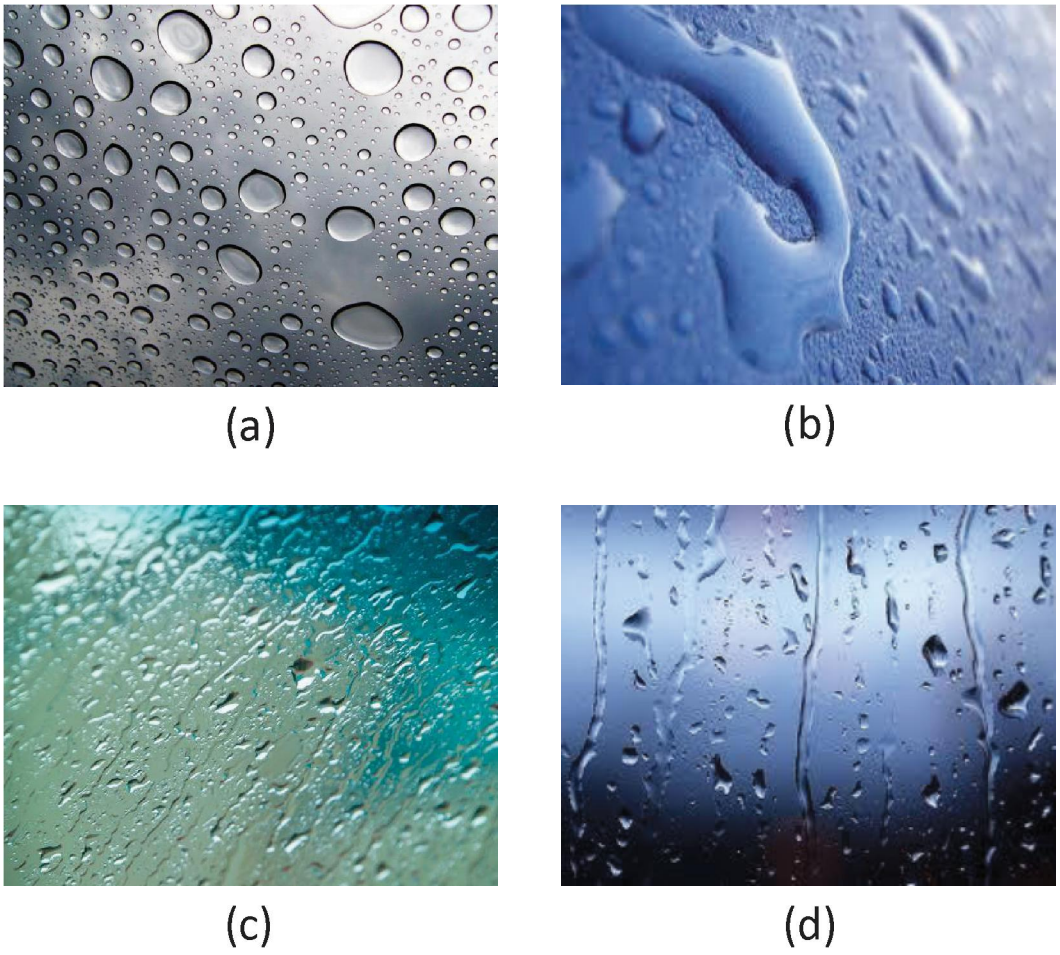


Figure 2.2: Adherent raindrops categorization according to shapes.
(a) Round. (b) Non-round patch. (c) Streaks. (d) Mixed. These figures are the courtesy of Kenna Westerman and Vladimir Voronin.

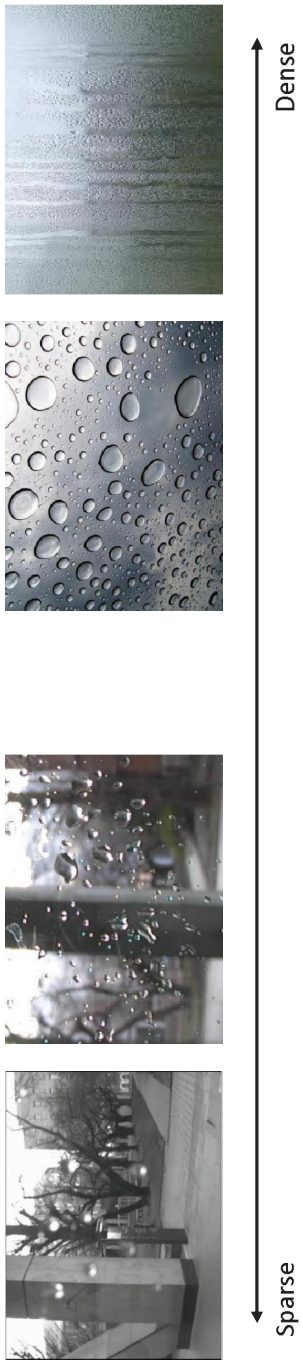


Figure 2.3: Adherent raindrops in different distributions.

As illustrated in Fig. 2.3, the density is a continuous parameter. In this thesis, we give the threshold that the distribution of adherent raindrops are considered to be sparse that:

1: There are clear intervals between raindrop area and non-raindrop area so that raindrop area is detectable.

2. Roughly less than 20% of the image are covered by raindrops so that there are sufficient information from non-raindrop area to repair the video.

Scene and background.

The appearance of raindrops is highly depend on the environment. It is impractical to numerate all the environmental situations in rainy whether. In this thesis, several situations which significantly affect the appearance are selected:

I. Light sources: high-lights will appear on raindrops if the sun is not totally covered by cloud or there are other point light sources like street lamps. Figs. 2.4 (a) and (b) are 2 examples. Usually, high lights does not appear in day-time because the sun is totally covered by thick cloud and there are no other light sources.

II. Texture richness: in outdoor environment, the background are usually texture-rich except the sky and ground are usually texture-less. Figs. 2.4 (c) and (d) are 2 examples. Most outdoor environment contains both textured and non-textured backgrounds.

Camera condition.

Besides outer parameters, the appearance of raindrops also depends on camera inner parameters. Among camera inner parameters, the appearance is most affected by lens focus. Because of the significant depth difference, except pinhole, video camera lens cannot focus on both the raindrops and the background at the same time. Thus, the appearance of raindrops can be categorized according to the camera focus. Fig. 2.5 is an illustration of the appearance of same raindrops on difference focus.

2.1.2 Problem Formation

Based on the categorization in the former subsection, in this thesis, we propose the method to detect and remove the adherent raindrops in video belongs to the following categories:

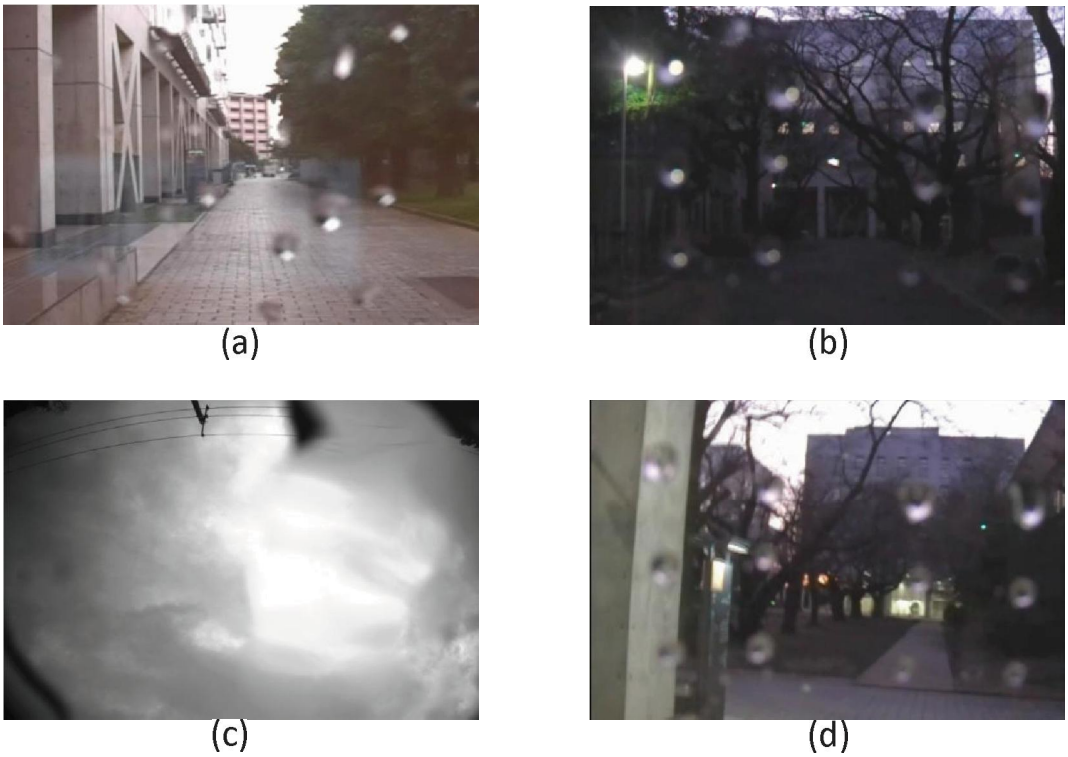


Figure 2.4: Appearance of same raindrops in different scenes.
(a) High light appear on raindrops when the sun is not totally covered by cloud. (b) High light appear on raindrops when there are point light sources light street lamps. (c) Appearance of raindrops in texture-less background (sky). (d) Appearance of raindrops in texture-rich background.

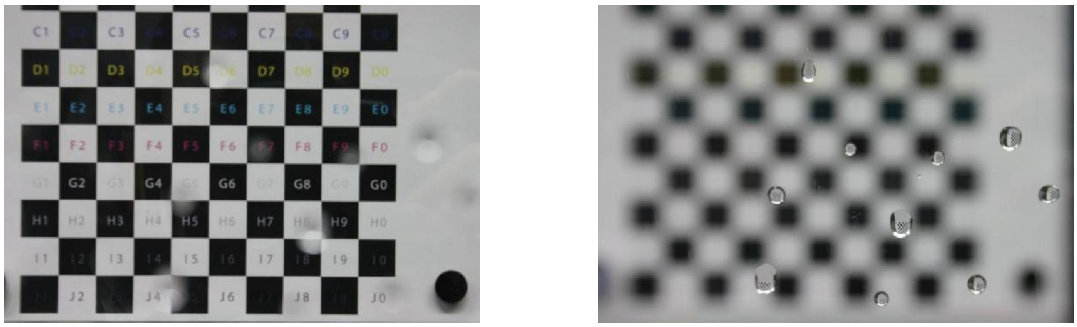


Figure 2.5: Appearance of same raindrops with difference camera focus.
(a) Camera focused to the background, clear background, blurred raindrops. (b) Camera focused to the raindrops, blurred background, clear raindrops.

I. Shape: round or non-round patches.

II. Distribution: Sparse.

III. Scene and background: no highlight, both textured and texture-less background.

IV. Camera condition: camera is focused to the background such that raindrops are blurred and background is sharp.

2.2 Raindrop Detection

Detecting raindrops is challenging. Unlike opaque objects, raindrops are totally transparent, which the appearance totally depends on the environment and in most cases varies in terms of intensity. While modeling and detecting falling raindrops in video have been exploited by Barnum et al. [2], Garg and Nayar [8, 9], Kang et al. [14], modeling and detecting adherent raindrops have been less exploited. The main reason is that, as shown in Figs. 2.6(a) and (b), it is not trivial to model the shape of adherent raindrops due to their various shapes, intensities, and distributions. To the extent of our knowledge, in physics, modeling the shapes of adherent water-drops is still in progress [11].

2.2.1 Formation of Raindrops Imagery Model

As shown in Fig. 2.7(a), the appearance of each raindrop is a contracted image of the background, as if it is taken from a catadioptric camera. Mathematically, for a given raindrop, we describe the smooth expand mapping start from raindrop area Ω_r into the environment scene Ω_e as φ:

$$\varphi : \Omega_r \rightarrow \Omega_e \tag{2.1}$$

The appearance of the raindrop and the environment share the same image plane and coordinates. In order to distinguish, in this thesis, we denote the points and coordinates in raindrop Ω_r as: $P_r = (u, v)$ and the corresponding points and coordinates in environment Ω_e as $P_e = (x, y)$. Then φ can be expressed as:

$$P_e = (x, y) = \varphi(P_r) = \varphi(u, v) = (\varphi^1(u, v), \varphi^2(u, v)) \tag{2.2}$$

As illustrated in Fig. 2.8, if A. all the camera inner parameters; B. all the geometric information of the raindrops; and C. all the depth information of the environment are determined, then φ is uniquely determined according to the refraction model.

For detection, except A, all other parameters should not be assumed known a priori. Roser et al. assumed B as a part of an ideal sphere [20] or Bezier curves [21], which only covers a small group of possible shapes.

Our task is detection, therefore, other than exhaustively solve φ, we extract differential properties of φ which are sufficient for detection. In Section 3, we theoretically estimate the linear expansion ratio of φ. Based on it, we propose dense motion based detection method. In Section 4, we theoretically estimate the area expansion ratio and the intensity change based detection method is thus proposed.

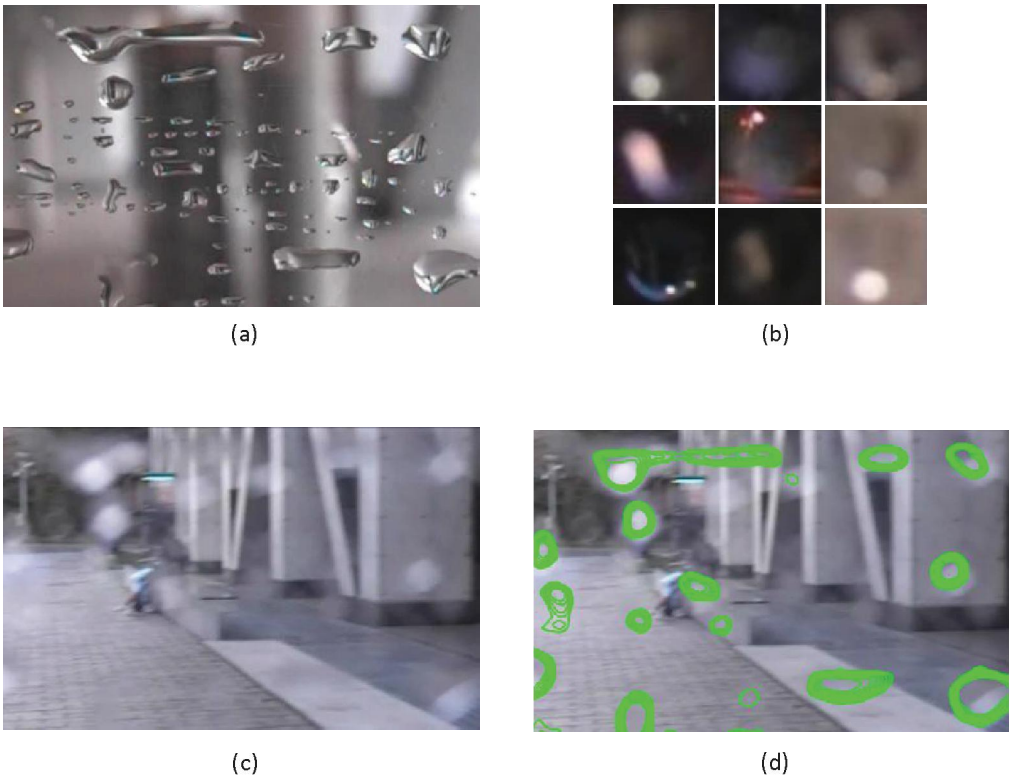
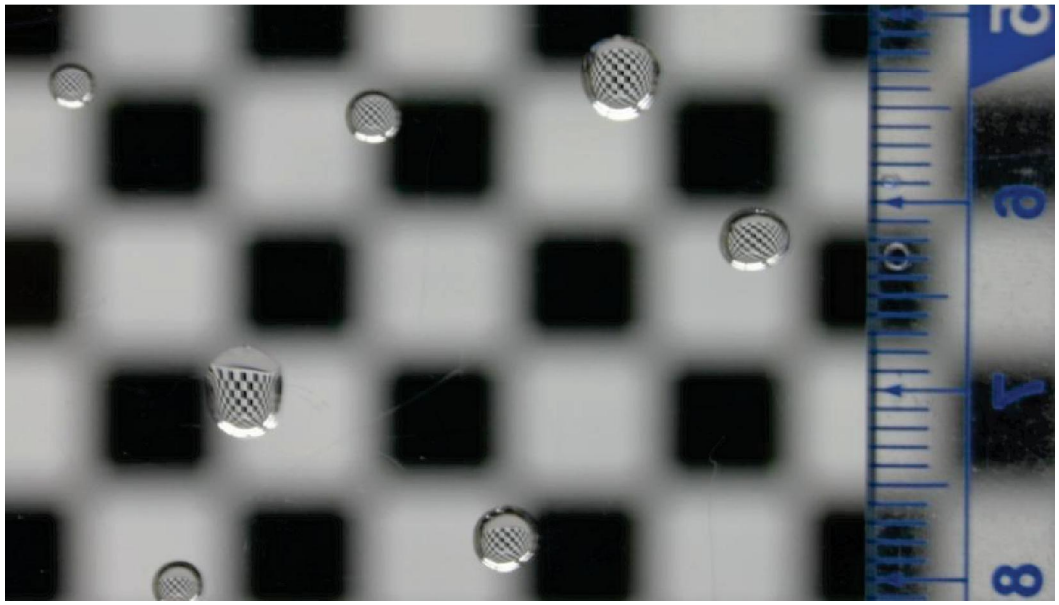
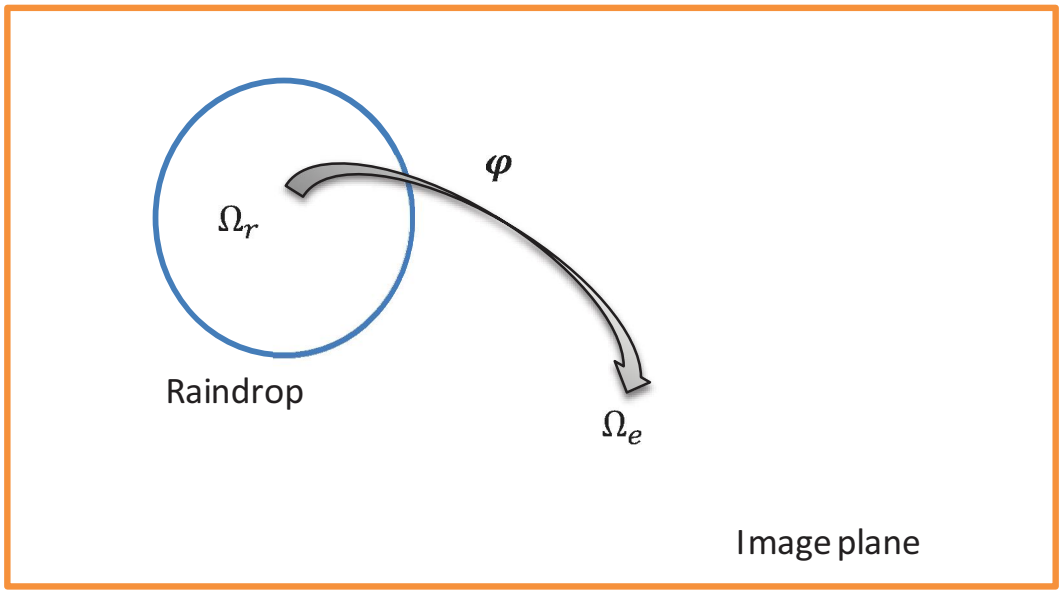


Figure 2.6: Detection of various adherent raindrops in video by our proposed method. (a) Adherent raindrops could have various shapes which are impractical to be geometrically modeled. (b) The appearance of the same raindrop various significantly under different environment. (c) The adherent raindrops in (a) are blurred when the camera is focused to the environment. (d) Detecting raindrops in (c) using our proposed method.



(a)



(b)

Figure 2.7: Raindrop imagery formation.
(a) The appearance of each raindrop is a contracted image of the background, as if it is taken from a catadioptric camera. (b) For a given raindrop area Ω_r, there is a smooth expand mapping φ start from the Ω_r into the environment scene Ω_e.

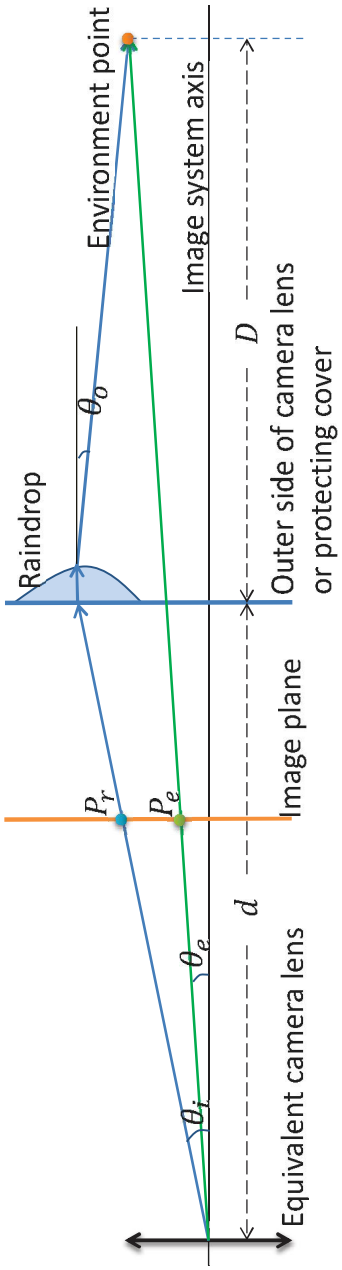


Figure 2.8: Refraction model of a pair of corresponding points on an image plane. There are two refractions on the light path through a raindrop. (The camera lens cover or protecting shield is assumed to be a thin plane and thus neglected.)

2.2.2 Related Works

In raindrop detection:

Roser et al. [20] attempt to model the adherent raindrops by assuming the raindrop shapes as a part of an ideal sphere in their first work. Later, they [21] focus on improving the model by using Bezier curves. However, the modeling is still not sufficient since it still assumes the boundary of a raindrop forms an ellipse or circle which is not always true for many raindrops as shown in Fig. 2.6(a).

Kurihata et al. [16] directly collect image templates of many raindrops and calculate their principle components . However, as shown in Figs. 2.6(a), (b) and (c) collecting and registrating training images for all various shapes, environment, illumination and blurring are considerably challenging.

Yamashita et al. [31, 33, 32] avoid modeling the raindrops by using stereo cameras. They detect raindrops by exploiting the differences in the two cameras. Their method only works with stereo cameras and assumes the two cameras share the same adherent raindrops, which is not tractable to create in real scenes.

A comparison on the applicability of these method and our method is illustrated in Table: 2.1

While all the previous methods try to model each raindrop as a unit object, we model raindrops locally from its derivative properties. Modeling a whole raindrop needs too many parameters which are impractical when those parameters are unknown; on the contrary, modeling the derivative properties needs only few parameters.

In this thesis, we model the derivative properties between a raindrop pixel and non-raindrop pixel that are originated from the same point in the environment. We observed that the imagery of an adherent raindrop is in fact the contraction of the environment. Based on this, we theoretically found that the contraction ratio is at least $\frac{1}{8}$ in any place on the raindrop. Based on this property, we propose the raindrop detection method by using dense motion estimation (e.g. optical flow). We also found that each pixel in a raindrop represents at least 64 pixels in the environment. Because of this ratio, raindrop area changes less compared to the environment. Relying on this analysis, we propose a detection method based on the change of intensity. Both methods detect raindrops on a pixel basis, making them generally applicable for raindrops with any shape and size. Fig. 2.6(d) shows one result of our proposed detection method.

Table 2.1: Comparison on the applicability of raindrop detection methods.

	Roser et al.	**Kurihata et al.**	**Yamashita et al.**	**Our method**
Source	image,video	image,video	stereo images	video
Shape	round	round	round	any
Distribution	sparse	sparse	sparse	sparse
Camera:raindrops	blurred	blurred	clear	blurred
Camera:background	clear	clear	clear	clear
High light	no	no	no	no
Background	textured	textured	textured	textured

2.3 Video Completion

After deleting the raindrop area, video completion techniques are used to repair the video. For video completion, the basic idea is that the missing information in a given frame appears elsewhere in the video. Thus, those information available elsewhere could copy to the targeting area and repair the video.

To form this idea, spatial-temporal space is introduced first. Then, by using spatial-temporal space, existing works on video completion are introduced. Lastly, our video completion task is formed by using spatial-temporal space and it is compared with the related works.

2.3.1 Spatial-Temporal Space

For video (image sequence), spatial-temporal space is a 3D space with 2 spatial dimensions and 1 temporal dimension. The 2 spatial dimensions are expanded as 2D imagery space. As illustrated in Fig. 2.9(a), for a given video (image sequence), its spatial-temporal space is formed by lining up its frames in temporal order.

As illustrated in Fig. 2.9(b), to better analyze the spatial-temporal behavior of a video, 2D slice with 1 spatial dimension and 1 temporal dimension are used. An example of 2D slice on a street view video is shown in Fig. 2.9(c).

As illustrated in Fig. 2.10(a), if the camera is moving continuously. Then any object which is visible in the camera will form a smooth trajectory in the spatial temporal space. The trajectory is better visualized in 2D slice. (Fig. 2.10(b)).

In video complete, for a given object trajectory in spatial-temporal space, if part of the trajectory is missing while the left part remains, it means the object is missing in some frames but appears in other frames. This missing object could be repaired by copy its appearance from other frames. (Fig. 2.10(c)).

In next subsection, the related works on video completion are introduced with the illustration in 2D slice spatial temporal space. For abbreviation, we call the 2D slice of spatial-temporal space as spatial-temporal space in the rest of the thesis.

2.3.2 Related Works

Before introduction video completion, some related works on image completion (image inpainting) are introduced. Then, related works on video completion are in-

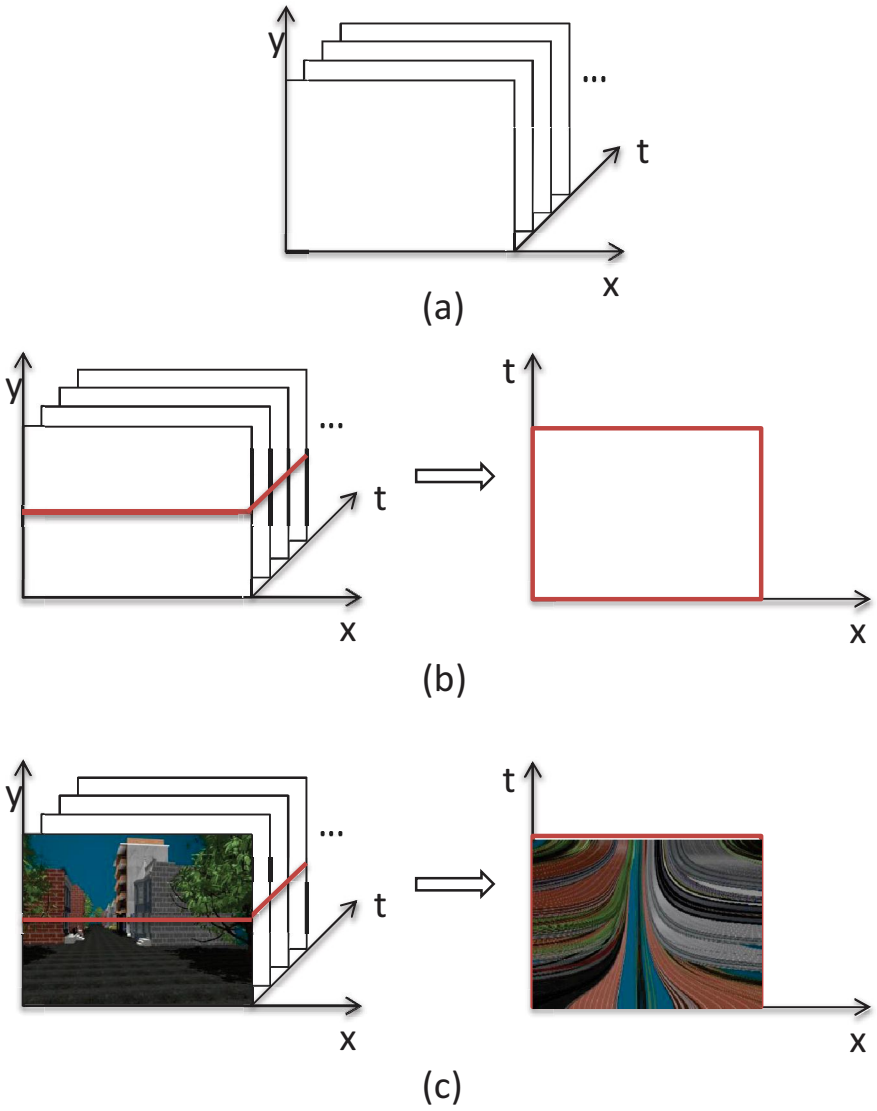


Figure 2.9: Spatial-temporal space.
(a) For a given video (image sequence), its spatial temporal space is formed by lining up its frames in temporal order. (b) to better analyze the spatial-temporal behavior of a video, 2D slice with 1 spatial dimension and 1 temporal dimension are used. (c) 2D spatial-temporal slice from a street view video. The pictures are the courtesy of Danping Zou.

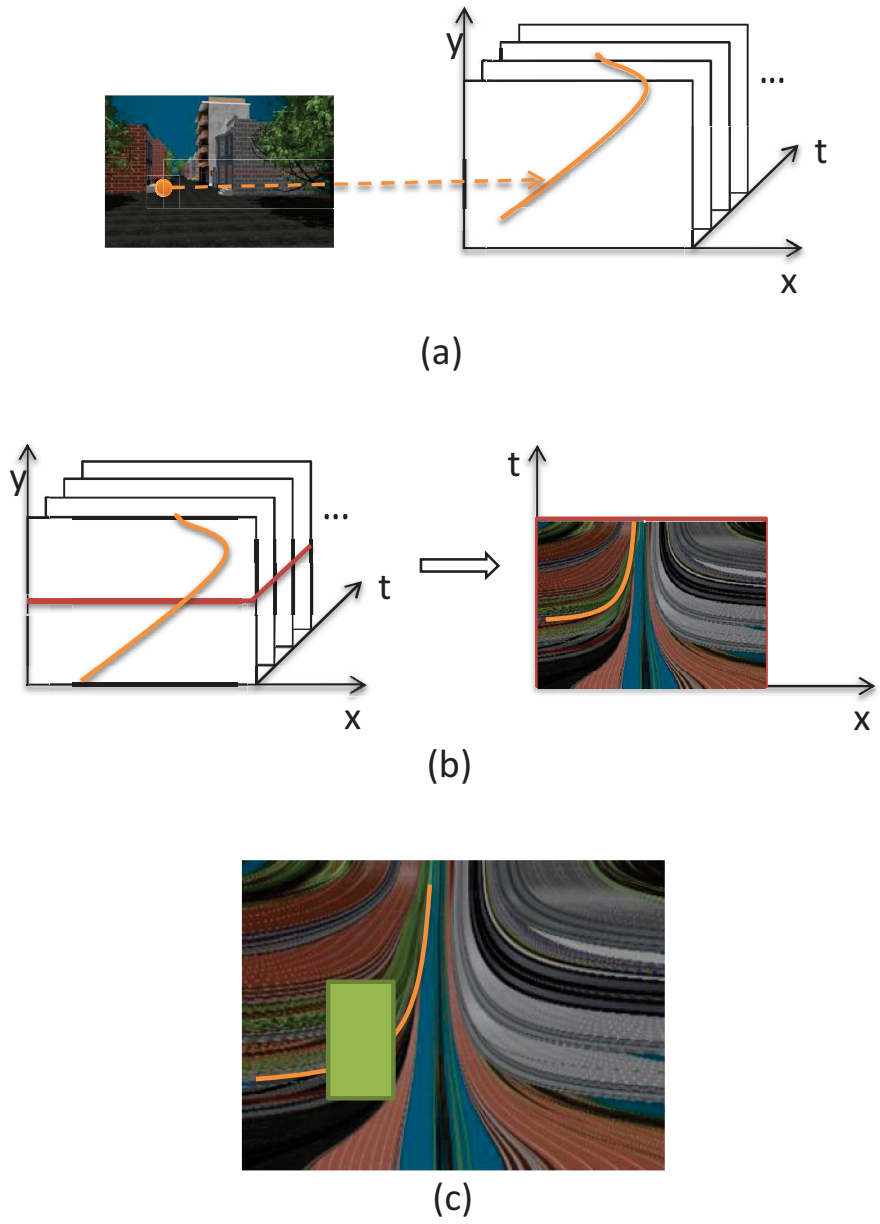


Figure 2.10: Object trajectory in spatial-temporal space.
(a) If the camera is moving continuously. Then any object which is visible in the camera will form a smooth trajectory in the spatial temporal space. (b) Visualize the trajectory in 2D spatial-temporal slice. (c) Part of the trajectory is missing (green block). The missing part could be repair by using the rest trajectory.

troduced with three categories according the scale of data missing in spatial-temporal space.

Image completion (image inpainting)

Unlike video, with only one given image, the missing area does not reappear else where. Missing area are synthesized according the texture and structure information in its neighboring area. One of the popular image completion method is proposed by Criminisi et al. [6, 1]. Their method combines exemplar-based texture synthesis and structure propagation. Exemplar-based texture synthesis contains the essential process required to replicate both texture and structure; the structure propagation, however, is highly dependent on the order in which the filling proceeds. The propose a best-first algorithm in which the confidence in the synthesized pixel values is propagated in a manner similar to the propagation of information in inpainting.

Small Spatial Missing Area

Bertalmio [3, 4], Kokaram [15] and Sahay et al. [22] propose methods on repairing small spatial-temporal missing area in videos. The missing area is illustrated in Figs. 2.11 (a) and (b). Since the missing area is spatially small, the texture and structure information is highly related to its neighboring area. Thus, a image completion techniques would work well.

Large Spatial Missing Area

As shown in Fig. 2.11 (c), large spatial missing but small temporal missing usually occurs in vintage films. Tang et al. [28] proposed method on repairing vintage films with large spatial missing. Cheung et al. [5] proposed method for interpolating frames in video. Although the missing area is spatially large, in the spatial-temporal view (Fig. 2.11 (d)), the missing area is temporally small. This means the missing area is highly related to its temporal neighboring area. Thus, copying from the temporally neighboring area with refinement on displacement and illumination will repair the video.

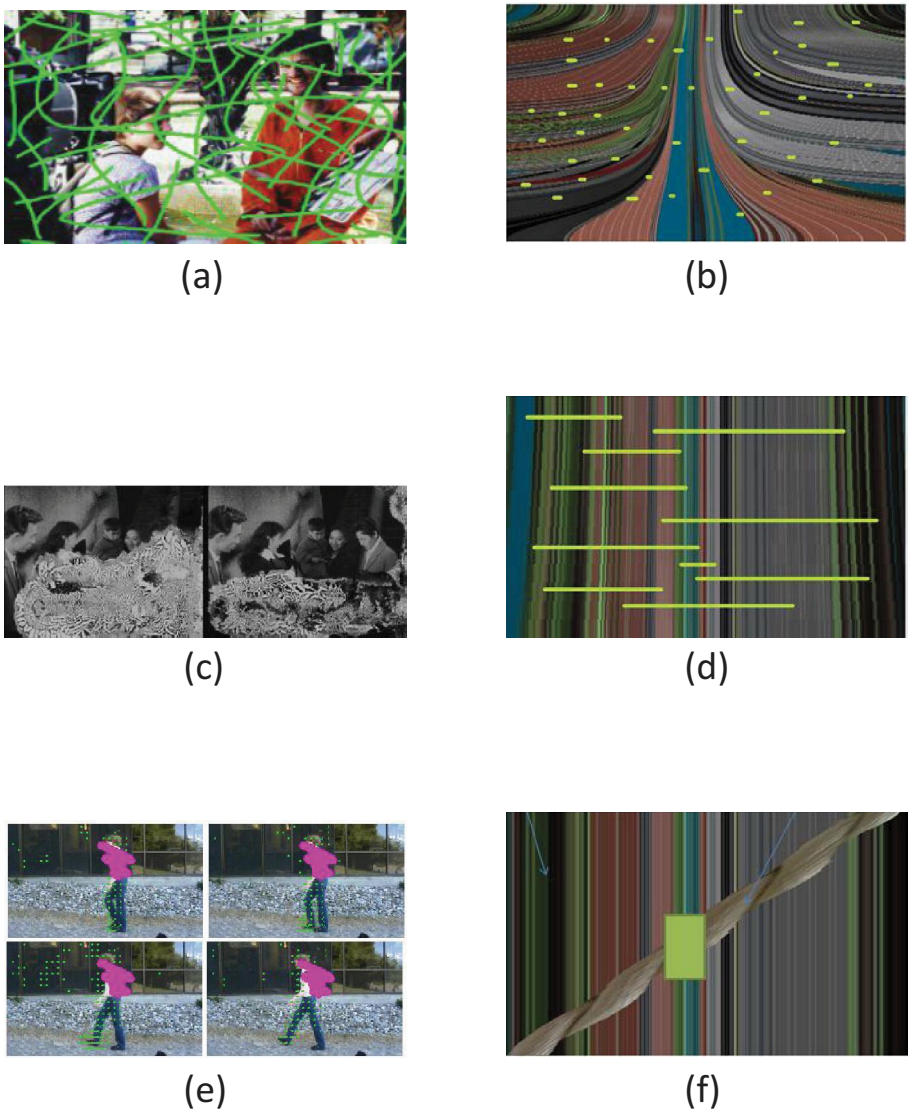


Figure 2.11: Different scale of data missing in spatial-temporal aspect. Green color denotes the missing area in spatial temporal space. (a) and (b) The missing area is both spatially small and temporally small. (c) and (d) The missing area is spatially large but temporally small. (e) and (f) The missing area is both spatially large and temporally large. (a) Cited from [3]. (c) Cited from [28]. (e) Cited from [27].

Large Spatial-Temporal Missing Area

For video completion, the most common situation is that the missing area is both the spatially and temporally large. Figs. 2.11 (e) and (f) is an illustration.

To repair the missing area, Wexler et al. [30] extend Criminisi's inpainting method. While Criminisi's image inpainting method propagate 2D image patches, Wexler et al. extended the 2D image patches to 3D voxels in the spatial-temporal space. Venkatesh et al. [29] also propose video completion techniques based on inpainting.

As illustrated in Fig. 2.10, other than directly propagate the image voxels in 3D spatial-temporal space. Videos with large missing area could be completed by first modeling the object trajectory, then image patches are copied to the targeting area according to the trajectory. Ding et al. [7], Jia et al. [13, 12], Liu et al. [18, 17], Matsushita et al. [19], Sahay et al [22], Sapiro et al [23, 24], Shih et al. [26, 25], Shiratori et al [27] and Zhang et al. [34] propose method on modeling object with cyclic trajectories, specifically, side view of walking human. For simplicity and focusing, all the above methods modeling the high repeating cyclic motion of walking human. And the backgrounds are mostly static which do not need to model or can be easily modeled.

2.3.3 Problem Formation.

This thesis is focusing on repair the missing area caused by raindrops in vehicle mounted camera video. As illustrated in Fig. 2.12, missing area caused by raindrops are both spatially large and temporally large. In spatial scale, or videos with 720*480 pixels, diagram of raindrops can be bigger than 100 pixels. In temporal scale, unless the raindrops slide down or been removed by windscreen wiper, it will always stay in the same place.

As we can see in the spatial-temporal space, repairing missing areas caused by raindrops are more complex than all the relative works. On one hand, the missing area is both spatially large and temporally large which means those repairing method for small missing area is not directly applicable. One the other hand, there are situations that the whole object trajectory is covered by raindrops, which means those areas can not be recovered by recovering the object trajectory.

In Chapter 4, Repairing missing area caused by raindrops will be divided into 3 situations.

I. Fast motion, structured and textured area: of which the object trajectory is half covered by raindrops and can be recovered. Figs. 2.12 (a) and (b) are illustrations of

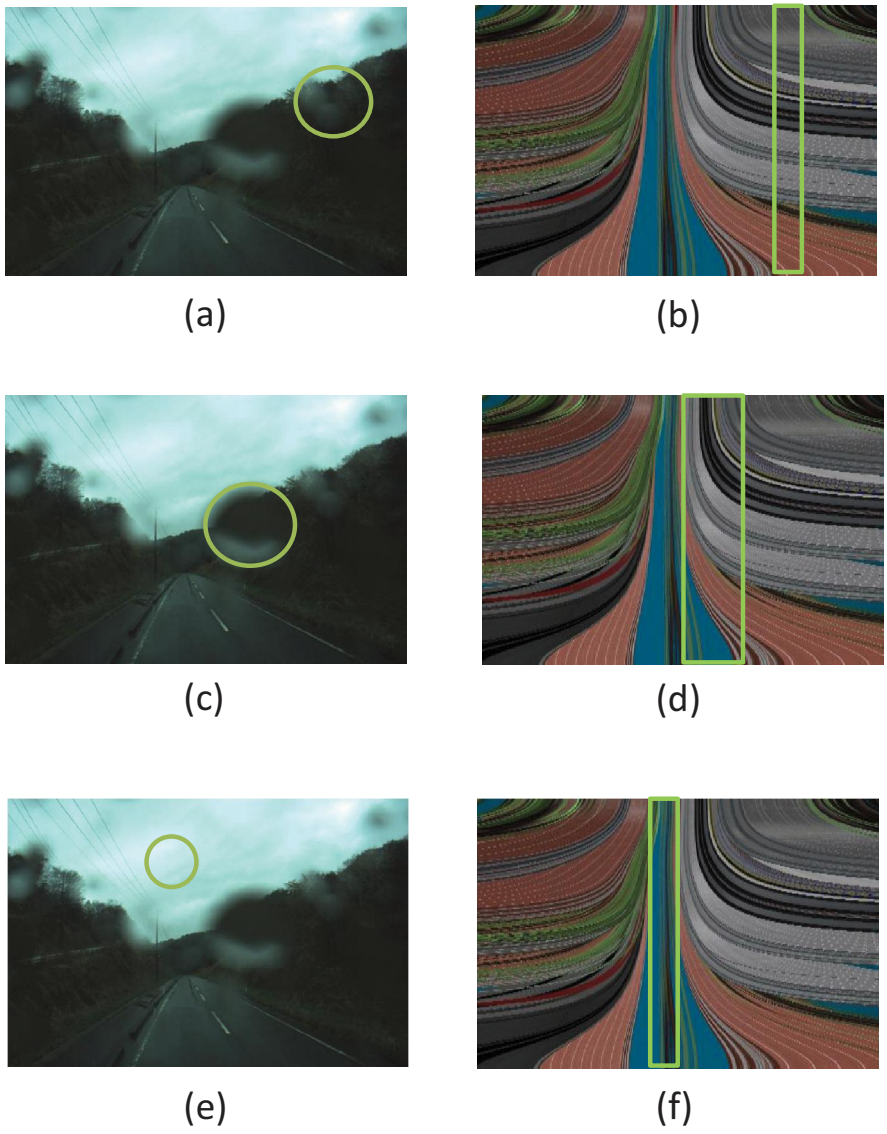


Figure 2.12: Three kind of missing cased by raindrops.
Green color denotes the missing area. (a) and (b) Fast motion, structured and textured area: of which the object trajectory is half covered by raindrops and can be recovered. (c) and (d) Slow motion, structured and textured area: of which the object trajectory is totally covered by raindrops. (e) and (f) Static and non-structured area, of which the object trajectory is totally covered by raindrops area.

this situation.

II. Slow motion, structured and textured area: of which the object trajectory is totally covered by raindrops. (Figs. 2.12 (c) and (d))

III. Static and non-structured area, of which the object trajectory is totally covered by raindrops area. In street view video, repairing the sky area belongs to this situation. (Figs. 2.12 (e) and (f)).

Methods to repair videos in each of the situations will be proposed correspondingly.

Chapter 3

Raindrops Detection in Video

To detect raindrops, other than exhaustively solve the raindrop imagery function, φ, we extract differential properties of φ which are sufficient for detection. In Subsection 3.1.1, we theoretically estimate the linear expansion ratio of φ. Based on it, we propose dense motion based detection method. In Subsection 3.1.2, we theoretically estimate the area expansion ratio and the intensity change based detection method is thus proposed.

3.1 Theory

3.1.1 Motion Analysis

In this section, we first mathematically form the derivative of φ and the expansion ratio. Then, we theoretically estimate a lower boundary of the expansion ratio according to the refraction model. Based on it, lastly, we propose a detection method based on dense motion.

Local differentials and expansion ratio

Referring to Eq.(2.2), the local differentials at $P_r = (u, v)$ is defined as:

$$J_\varphi(P_r) = J_\varphi(u, v) = \begin{pmatrix} \varphi_u^1(u,v) & \varphi_v^1(u,v) \\ \varphi_u^2(u,v) & \varphi_v^2(u,v) \end{pmatrix} \tag{3.1}$$

with: $\varphi_u^1(u,v) = \frac{\partial \varphi^1(u,v)}{\partial u}$.

The local motion at (u, v), denoted as $(\delta u, \delta v)^T$, and the local motion at (x, y), denoted as $(\delta x, \delta y)^T$, is linearly associated by $J_\varphi(u, v)$:

$$\begin{pmatrix} \delta x \\ \delta y \end{pmatrix} = J_\varphi(u, v) \begin{pmatrix} \delta u \\ \delta v \end{pmatrix} \tag{3.2}$$

Instead of modeling φ or $J_\varphi(u, v)$, we are interested in the ratio between $\| (\delta x, \delta y) \|$ and $\| (\delta u, \delta v) \|$. According to Eq.(3.2):

$$\| (\delta x, \delta y) \|^2 = (\delta x, \delta y)^T (\delta x, \delta y) = (\delta u, \delta v)^T (J_\varphi(u, v))^T J_\varphi(u, v) (\delta u, \delta v) \tag{3.3}$$

with $(J_\varphi(u, v))^T J_\varphi(u, v)$ is symmetric and positive-semidefinite, and can be diagonalized as:

$$(J_\varphi(u, v))^T J_\varphi(u, v) = E^T \begin{pmatrix} \lambda_1^2(u, v) & \\ & \lambda_2^2(u, v) \end{pmatrix} E \tag{3.4}$$

where E is an orthogonal matrix, and $0 \leq \lambda_1(u, v) < \lambda_2(u, v)$. Therefore, according to Eqs.(3.3) and (3.4), for any directional motion $(\delta u, \delta v)$ at (u, v):

$$\frac{\| (\delta x, \delta y) \|}{\| (\delta u, \delta v) \|} \geq \lambda_1(u, v) \tag{3.5}$$

We can give a lower boundary, denoted as λ_{lower}, for all $\lambda_1(u, v)$ inside the raindrop area Ω_r:

$$\lambda_{lower} \leq \min\{\lambda_1(u, v) | (u, v) \in \Omega_r\} \tag{3.6}$$

We call it the lower boundary of the expansion ratio.

Estimating the Lower Boundary of Expansion Ratio

A light ray passing through a raindrop undergoes two refractions: first, the refraction from the air to the water, and second, the refraction from the water to the air. Thus, the mapping function, φ, can be separated as two continuous mappings:

$$\varphi = \overset{a-w}{\varphi} \circ \overset{w-a}{\varphi} \tag{3.7}$$

index a stands for air, and w stands for water.

Assuming the contact surface between the camera lens cover and raindrop is flat, $\overset{a-w}{\varphi}$ should be analytically solvable.

According to Snell's law, where $n_a \sin\theta_i = n_w \sin\theta_o$, we can have:

$$\frac{P_e}{P_r} = \frac{d\tan\theta_e}{d\tan\theta_i} = \frac{n_a}{n_w}\frac{1+\frac{n_w}{n_a}\frac{d}{D}}{1+\frac{d}{D}} = constant \tag{3.8}$$

where the notation is defined in Fig. 2.8, and $\frac{n_w}{n_a}$ is approximately $\frac{4}{3} > 1$. Thus the ratio:

$$\frac{\mathrm{d}P_e}{\mathrm{d}P_r} = \frac{P_e}{P_r} = \frac{n_a}{n_w}\frac{1+\frac{n_w}{n_a}\frac{d}{D}}{1+\frac{d}{D}} > \frac{n_a}{n_w} = 0.75 \tag{3.9}$$

Hence, a lower boundary of the expansion (contraction here) ratio is:

$$\overset{a-w}{\lambda}_{lower} = 0.75 \tag{3.10}$$

Now, we estimate the expansion ratio of the second refraction $\overset{w-a}{\varphi}$. Note that, referring to Fig. 2.8, although in the first refraction the direction and position of the emergence light trace could be analytically solved, the position and angle of the incident light of the second refraction is still unsolvable. This is because we have no knowledge of the position and shape of the raindrop.

To estimate the expansion ratio of the second refraction, we start from the differential geometry on the outer surface of the raindrop. For a given position (u, v) on the surface of the raindrop, its up to second order differential geometry values are illustrated as in Fig. 3.1 [35]. The upper principle curvature vector, k_1, points to the direction where the raindrop surface bends most. And the lower principle curvature vector, k_2, points to the direction where the surface bends least. The curvature vector of any other direction, k, is the linear combination of k_1 and k_2. The values of any curvature vector k is bounded by k_1 and k_2:

$$k_2 \leq k \leq k_1 \tag{3.11}$$

The reciprocal of curvature is called curvature radius: $R = \frac{1}{k}$. In any direction, it is bounded by two principle curvature radius: $R_1 \leq R \leq R_2$.

As illustrated in Fig. 2.8, we now consider the second refraction locally at given point (u, v). Mention that there is no knowledge about how this local coordinates is aligned to the global coordination. First, we try to estimate the angular ratio $\frac{\mathrm{d}\theta_o}{\mathrm{d}\theta_i}$.

According to Snell's law, we have:

$$\frac{\mathrm{d}\theta_o}{\mathrm{d}\theta_i} = \frac{\frac{n_w}{n_a}\cos\theta_i}{(1-(\frac{n_w}{n_a})^2\sin^2\theta_i)^{\frac{1}{2}}} \tag{3.12}$$

where we know that $\frac{n_w}{n_a} > 1$, thus Eq.(3.12) gets its minimum when θ_i=0:

$$\min\left(\frac{\mathrm{d}\theta_o}{\mathrm{d}\theta_i}\right) = \left.\frac{\mathrm{d}\theta_o}{\mathrm{d}\theta_i}\right|_{\theta_i=0} \tag{3.13}$$

This is in accordance with the observation of real raindrop image shown in Fig. 2.7(a).

As illustrated in Fig. 3.2(b), according to Eq.(3.13), we may put the normal of the raindrop surface considerably close to the image system axis. Assuming every angle is significantly small:

$$\theta_i \ll 1, \theta_o \ll 1, \theta_e \ll 1, \theta_r \ll 1 \tag{3.14}$$

According to Eq.(3.14), we can use the following approximation: $\sin\theta = \tan\theta = \theta$, $\frac{\mathrm{d}P_e}{\mathrm{d}P_r} = \frac{P_e}{P_r}$. The expansion ratio is estimated as:

$$\frac{\mathrm{d}P_e}{\mathrm{d}P_r} = \frac{P_e}{P_r} = \frac{\theta_e}{\theta_r} = 1 - \frac{n_w}{n_a}\frac{d}{R}\frac{D}{d+D} \tag{3.15}$$

To estimate Eq.(3.15), we only need to give an estimation of the upper boundary of the curvature radius R for any raindrop at any given position. Since, the camera lens cover is vertical to the ground, big raindrops will slide down, and according to our observation of real data, almost all raindrops has a diameter smaller than $5mm$: $2R < 5mm$. Then $R < 5mm$ is a very safe assumption. Assuming $d > 100mm$ and $D > 1m$, we have:

$$\frac{P_e}{P_r} < -11 \tag{3.16}$$

The negative sign means the image on the raindrop is inverted, this is in accordance with our observation on real data.

The expansion ratio of the second refraction is then estimated as:

$$\overset{w-a}{\lambda}_{lower} > 11 \tag{3.17}$$

Substituting Eqs.(3.10) and (3.17) into Eq.(3.7), we have the overall expansion ratio estimation:

$$\lambda_{lower} > 8 \tag{3.18}$$

This means the motion in the environment $\| (\delta x, \delta y) \|$ is at least 8 times as great as the corresponding motion $\| (\delta u, \delta v) \|$ on the raindrop. Fig. 3.3 is an observation of real data, which demonstrates our estimation.

Pixel-wised raindrop detection using dense motion

We have demonstrated that the motion on the raindrop should be significantly smaller than the corresponding motion in the environment. Thus, we could detect raindrop area pixel-wisely by using optical flow. (Specifically, we use SIFT flow [17] in this thesis). Fig. 3.4 is an example.

3.1.2 Intensity Change Analysis

Based on the linear expansion ratio which is given in previous section, in this section, we estimate the area expansion ratio of φ. Then we propose the raindrop detection by intensity change.

Estimate area expansion ratio

Observing Eqs.(3.4) and (3.18), both eigenvalues of $(J_\varphi(u,v))^T J_\varphi(u,v)$ is greater than 0 which means $J_\varphi(u,v)$ is reversible at (u,v):

$$\begin{pmatrix} \delta u \\ \delta v \end{pmatrix} = J_\varphi^{-1}(x,y)\begin{pmatrix} \delta x \\ \delta y \end{pmatrix} \tag{3.19}$$

And at the neighborhood of (u,v), denoted as $V(u,v)$, φ is a one to one mapping which maps $V(u,v)$ to $V(x,y) = \varphi(V(u,v))$.

If we select a pixel in $V(x,y)$, its area can be defined as:

$$|\Delta x||\Delta y|\sin\theta = |\Delta x||\Delta y| \tag{3.20}$$

where $\Delta x = (\Delta x, 0)^T$ and $\Delta y = (0, \Delta y)^T$ denotes the directional width and height of the pixel, θ is the angle between Δx and Δy. For a square pixel, Δx is orthogonal to Δy, thus $\sin\theta = 1$.

For the given pixel area, its corresponding area on the raindrop can be calculated using Eq.(3.19):

$$|J_\varphi^{-1}(x,y)\Delta x||J_\varphi^{-1}(x,y)\Delta y|\sin\theta' \tag{3.21}$$

where θ' is the angle between $J_\varphi^{-1}(x,y)\Delta x$ and $J_\varphi^{-1}(x,y)\Delta y$. Using the lower bound of linear expansion ratio in Eq.(3.18), we have:

$$\begin{aligned} |J_\varphi^{-1}(x,y)\Delta x||J_\varphi^{-1}(x,y)\Delta y|\sin\theta' &\le |J_\varphi(x,y)^{-1}\Delta x||J_\varphi(x,y)^{-1}\Delta y| \\ &\le |\lambda_{lower}^{-1}\Delta x||\lambda_{lower}^{-1}\Delta y| < (\tfrac{1}{8})^2|\Delta x||\Delta y| < \tfrac{1}{64}|\Delta x||\Delta y| \end{aligned} \tag{3.22}$$

Comparing Eqs.(3.22) and (3.20), we have, any pixel in the environment can cover at most the area of $\frac{1}{64}$ pixel in the raindrop. Or inversely, a pixel on the raindrop corresponds to the area of at lest 64 pixels in the environment.

Raindrop detection using intensity change

For digitized data, only one set of value is sampled at a pixel. Thus, each pixel value $p(u, v, t)$ in the raindrop area is a weighted average of at least 64 pixels in the environment:

$$p(u, v, t) = \sum_{(x,y)\in\varphi(V(u,v))} W(x, y)p(x, y, t) \tag{3.23}$$

where $V(u, v)$ denotes the neighborhood of a pixel which covers point (u, v). $W(x, y) > 0$ is the weight at (x, y), and $\sum_{(x,y)\in\varphi(V(u,v))} W(x, y) = 1$.

According to Eq.(3.23), for different time t_1 and t_2 ($t_1 \neq t_2$), the change of intensity at the raindrop area is smaller than its corresponding environment area:

$$\begin{aligned} &|p(u, v, t_1) - p(u, v, t_2)| \\ &= |\textstyle\sum_{(x,y)\in\varphi(V(u,v))} W(x, y)(p(x, y, t_1) - p(x, y, t_2))| \\ &\leq \textstyle\sum_{(x,y)\in\varphi(V(u,v))} W(x, y)|p(x, y, t_1) - p(x, y, t_2)| \end{aligned} \tag{3.24}$$

For raindrop detection, we do not have previous knowledge on how the intensity changes at different places. If we assume at each pixel, $|p(x, y, t_1) - p(x, y, t_2)|$ obeys a $\sigma = 1$ independent normal distribution. Then $|p(u, v, t_1) - p(u, v, t_2)|$ obeys a $\sigma = \frac{1}{8}$ normal distribution. This means the change of intensity on raindrop area is significantly smaller than the environment. Therefore, change of intensity along time can be used to detect raindrop area. Fig. 3.5 is an example.

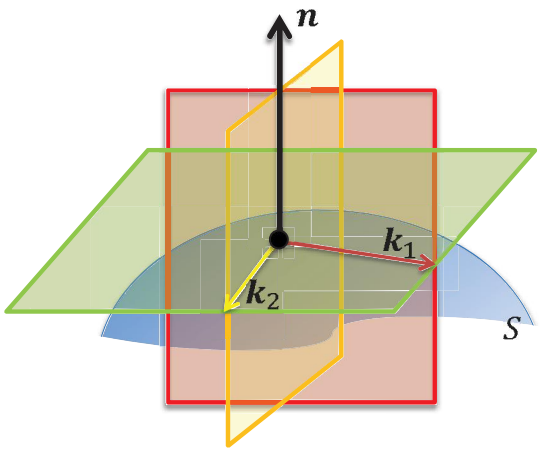


Figure 3.1: Local linear space

Given a point on raindrop surface S, its two principle curvatures vectors k_1, k_2 and the normal n are orthogonal to each other.

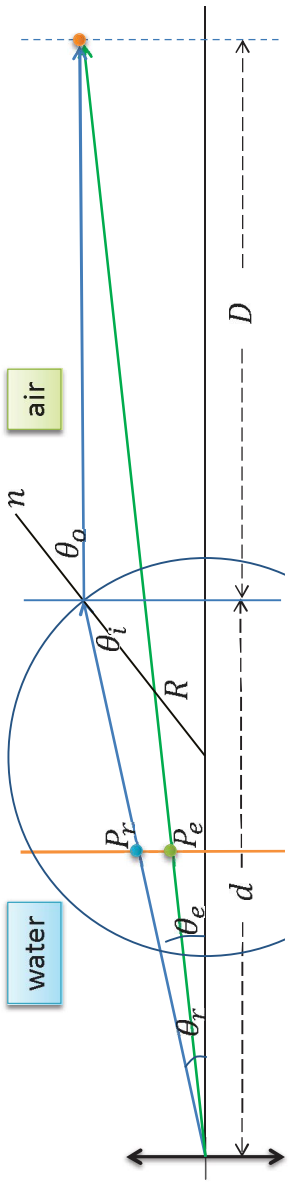


Figure 3.2: Simplified refraction model of the second refraction using principle curvature.

Refraction model of the second refraction when assuming the normal of refraction is very close to the image system axis. The notation are same with Fig. 2.8, R is the curvature radius at the place and direction where the refraction happens.

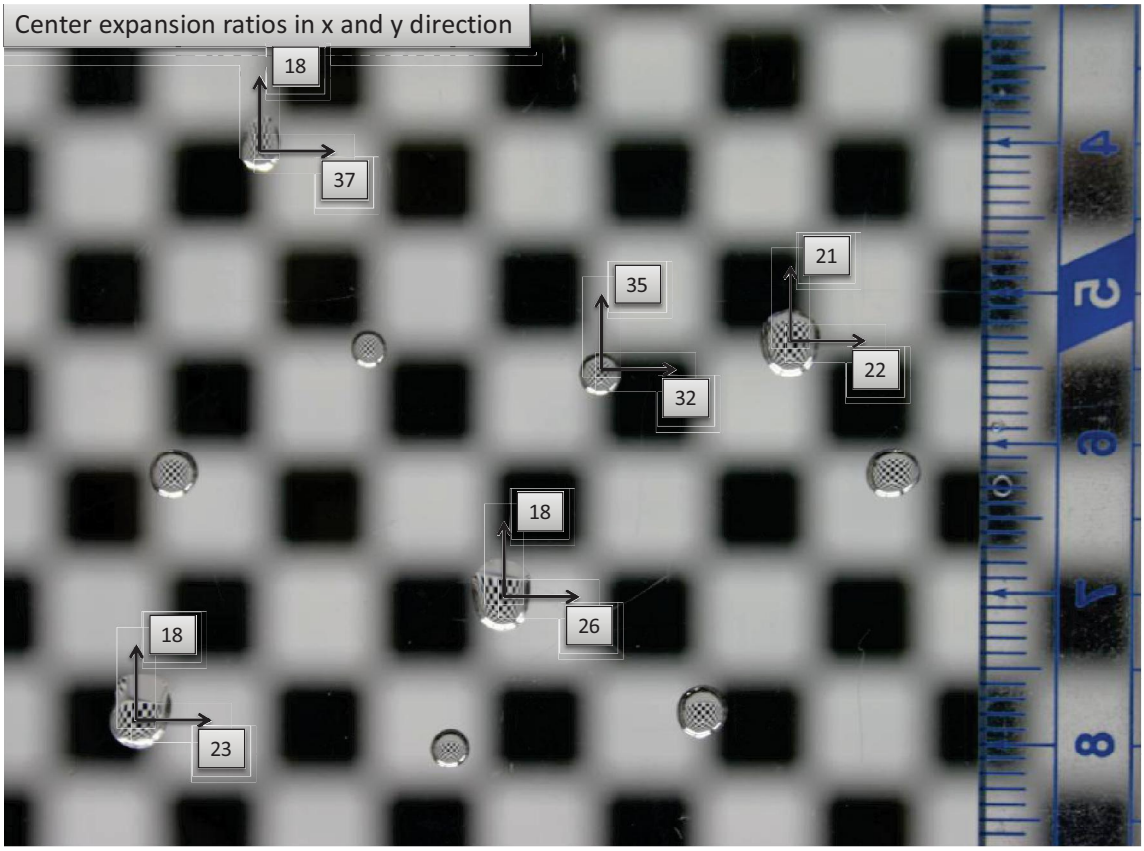


Figure 3.3: Observing the expansion ratio in x and y direction on real data.

Figure 3.4: Raindrop area detection using optical flow.
(a) Input video with blurred raindrops. (b) Inter frame optical flows. (c) Accumulated intensity of optical flow on 200 frames.

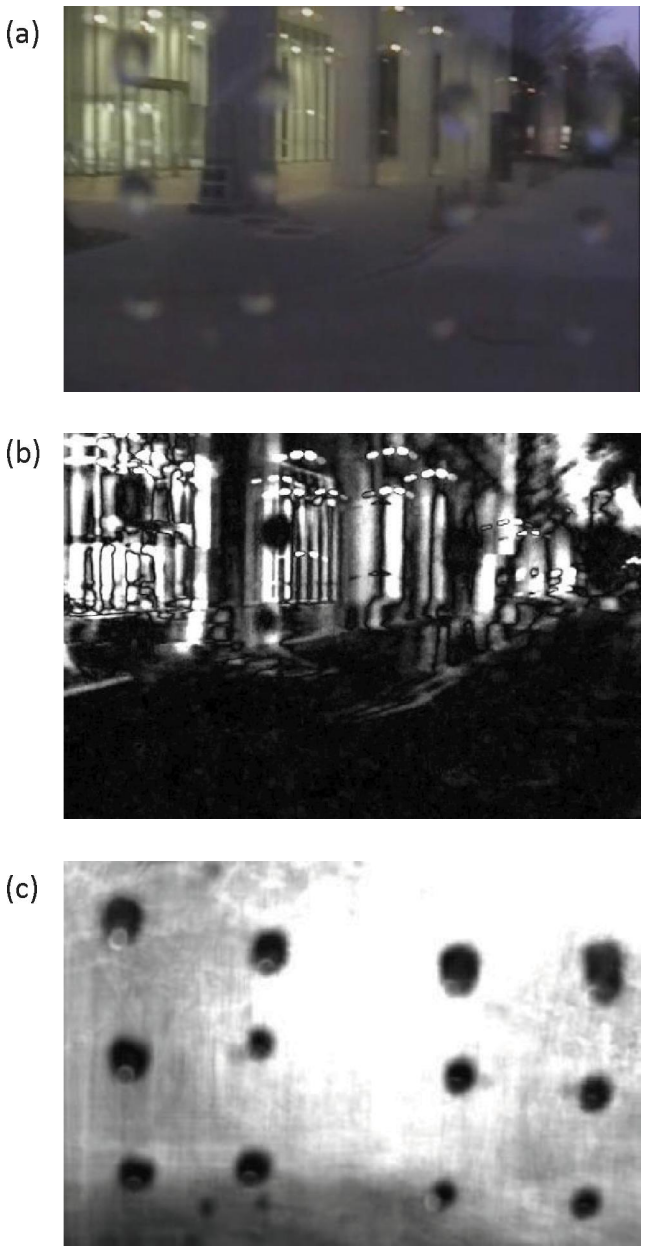


Figure 3.5: Pixel-wised intensity change.
(a) Input video with blurred raindrops. (b) Inter frame intensity change. (c) Accumulated intensity change over 200 frames.

3.2 Implementation

This section discusses the algorithm and implementation of raindrop detection based on the theoretical analysis explained in the previous sections.

3.2.1 Feature extraction

Real time feature extraction using sliding window

Our features for raindrop detection (motion and change of intensity) will be more robust to obtain if they are accumulated over a certain period of time. As illustrated in Fig. 3.6(a), we use a sliding window to accumulate features in real time.

Choose temporal length

Statistically, the longer the time period is, the more distinctive the features are. Fig. 3.7(a)-(c) is an example of accumulated intensity change over difference time period.

However, a long temporal window will increase the latency of detection. If the window length is set as T seconds, then a raindrop is expected to be detected within T seconds after it falls onto the lens' cover. Hence, to balance the latency and accuracy, we usually accumulate feature over 200 frames, which is about 8 seconds if the frame rate is 24 per second.

3.2.2 Raindrop area detection

Real time detection

For real time detection, we can simply set a threshold for the accumulated change of intensity. Pixels that have the accumulated change of intensity lower than a threshold is considered to be inside the raindrop area. An example is shown in Fig. 3.8(b).

Refined detection using level sets

If time is not a concern, level sets is used to achieve better accuracy. First, convolution of Gaussian with $\sigma = 2$ pixels is used to reduce noise. Then, as shown in Fig. 3.8(c), level set is calculated. We use the following criteria to select raindrop areas:

1. Threshold: as proved previously, raindrop area should have smaller optical flow and intensity change.

2. Closure: the level set around a raindrop area should be close.

3. Circumference. The diameter of a raindrop area is usually smaller than 5mm. For our camera setting, almost all raindrop diameter are smaller than 80 pixels. And thus the threshold for raindrop circumference is set less than 250 pixels.

4. Roundness. Although the contour of real raindrops can vary significantly, approximately the contour is close to round. We define the roundness of a contour as following:

$$\frac{Area}{(Circumference)^2} \tag{3.25}$$

A rounder shape would have a bigger roundness value and a circle has the maximum roundness value: $\frac{\pi r^2}{(2\pi r)^2} = \frac{1}{2\pi} = 0.16$. Specifically, square has the roundness value as $\frac{1}{4^2}$=0.0625. For our data, the threshold for roundness is set greater than 0.12.

Fig. 3.8(d) is an example of refined detection result by using the criterion above.

3.2.3 Overall work flow

The overall flowchart of our raindrop detection method is illustrated in Fig. 3.10. Firstly, the features, either optical flow or intensity change, are accumulated continuously by using a sliding window. Among which, accumulated intensity change could be extracted in real time. Secondly, a real time detection could be performed by using pixel-wisely compare the feature to a given threshold. At last, a refined detection is performed by using level set: after noise reduction, level set is calculated and raindrop areas are chosen based on the four proposed criterion.

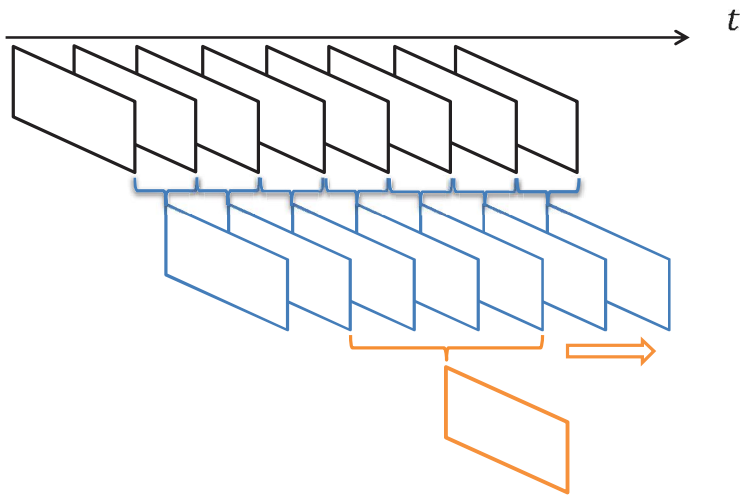


Figure 3.6: Accumulated feature in real time by using sliding window. Orange arrow: sliding window. Black frames: input frames; Blue frames: inter frame feature; Orange frames: accumulated feature.

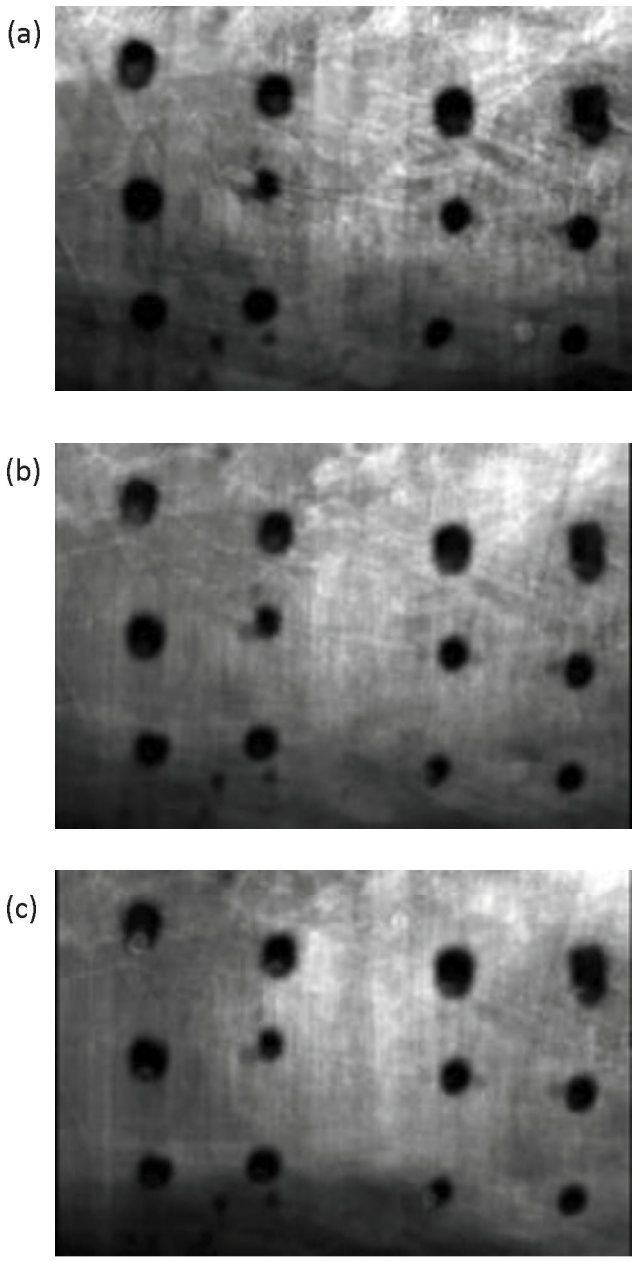


Figure 3.7: Accumulate feature over different time periods.
(a) 100 frames, (b) 200 frames and (c) 400 frames.

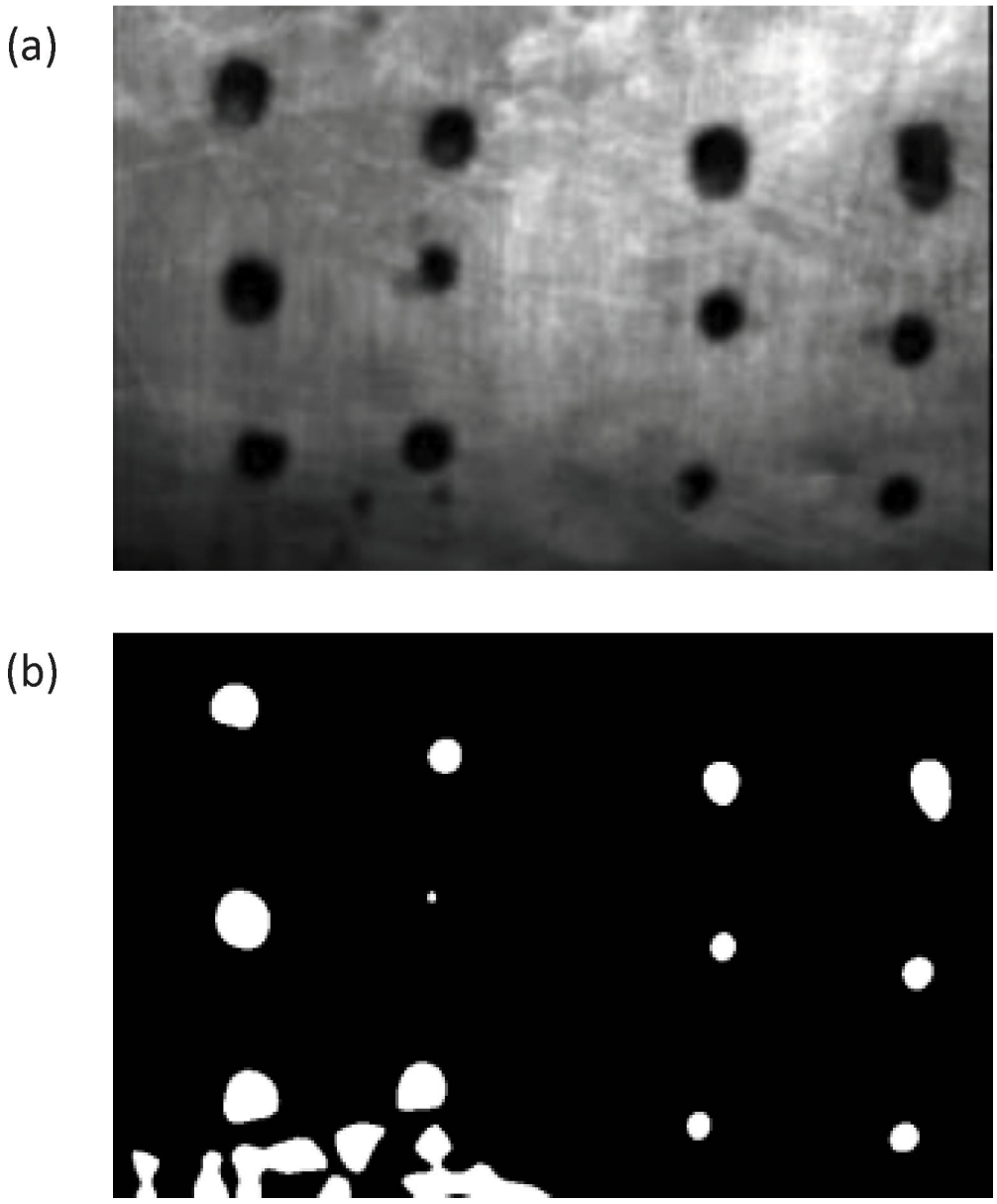


Figure 3.8: Real time detection.
(a) Accumulated features. (b) Real time detection using threshold.

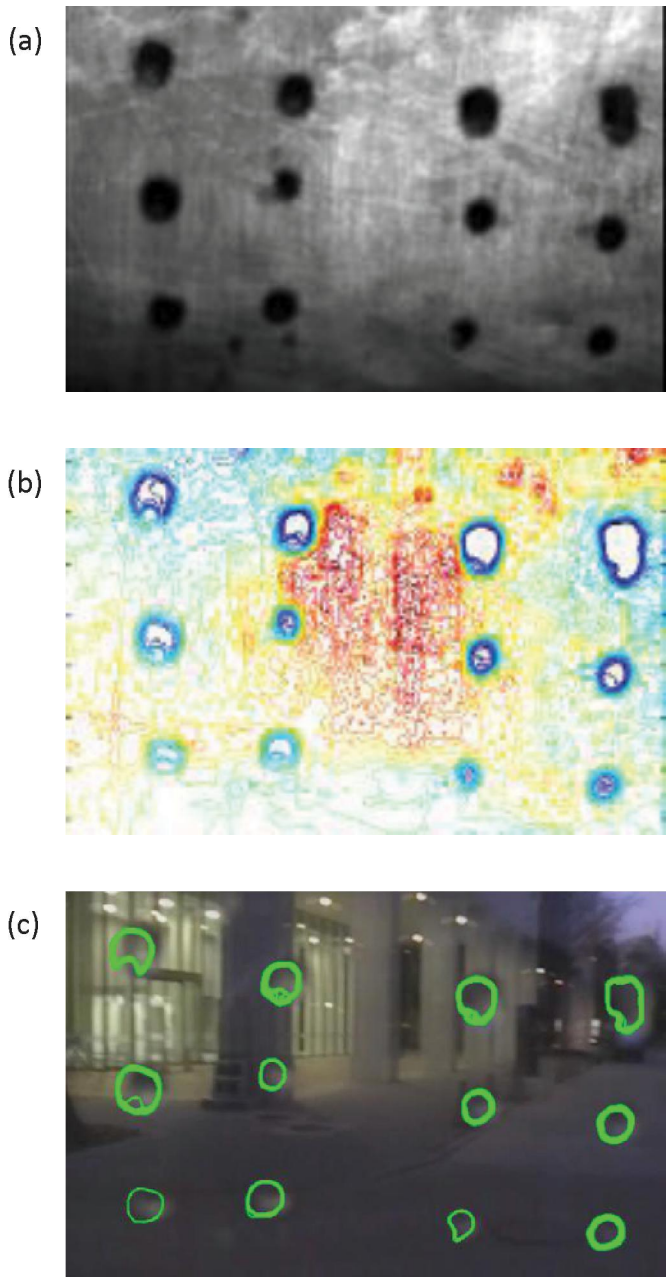


Figure 3.9: Refined detection.
(a) Accumulated features. (b) Level sets of accumulated feature. (c) Selected level sets using 4 proposed criterion.

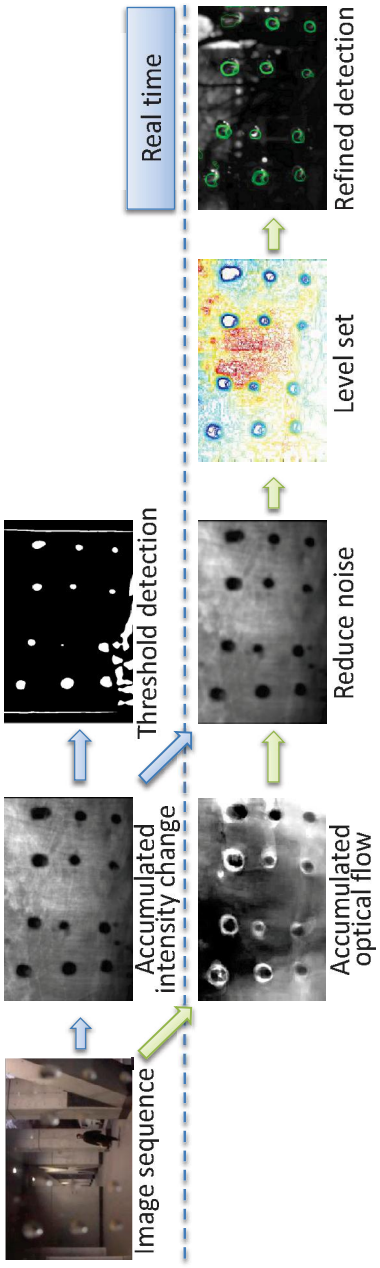


Figure 3.10: The overall work flow of our proposed raindrop detection method.

Chapter 4

Video Completion

In this chapter, methods to repair those area covered by raindrops are proposed.

Repairing missing area caused by raindrops will be divided into 3 situations.

I. Fast motion, structured and textured area: of which the object trajectory is half covered by raindrops and can be recovered. Figs. 2.12 (a) and (b) are illustrations of this situation.

II. Slow motion, structured and textured area: of which the object trajectory is totally covered by raindrops. (Figs. 2.12 (c) and (d))

III. Static and non-structured area, of which the object trajectory is totally covered by raindrops area. In street view video, repairing the sky area belongs to this situation. (Figs. 2.12 (e) and (f))

Methods to repair videos in each of the case are proposed correspondingly in the first 3 sections of this chapter. The overall work flow of the repairing method is introduced in the last section.

4.1 Video Completion in Fast Motion Area

4.1.1 Motion Trajectory Modeling using Interpolation

As illustrated in Figs. 4.1(a) and (b), if a given object is moving fast and if it is covered by a raindrop in certain frame, after several frames, it will move out of the raindrop area and appear again. If we can model its motion trajectory, then, the raindrop area can be repairing by copying the appearance of object according to the motion trajectory.

In video taken by vehicle, the object motion trajectory is smooth. And, as illustrated in Fig. 4.1 (c), in a small time period, it can be well approximated by using linear interpolation. Thus, the problem left to us is to model the motion in the neighborhood of raindrop area, either the motion from Point A to Point B, or the motion from Point D to Point C.

4.1.2 The Limitation of Optical Flow

To model the inter frame motion, one possible method is to use optical flow. However, as discussed in raindrop detection, those area covered by raindrop has very small optical flow in comparing with its neighborhood. Optical flow estimation in the neighborhood or raindrops is not robust. This is because smooth constraints are used to improve the robustness of optical flow estimation, such that the estimation in neighborhood of raindrops will be significantly affected by raindrops. Fig. 4.4(a) is an example of motion estimation using SIFT-flow [17], the estimation in the neighborhood of raindrops are erroneous. Thus, traditional optical flow method cannot be used to estimate motion in the neighborhood of raindrops directly.

4.1.3 Motion Modeling using Sparse Matching

To overcome the effect of raindrops, in this thesis, sparse matching based method is proposed. As illustrated in Figs. 4.2 (a) and (b), given two temporally neighboring frames, there sparse correspondences can be find by using SIFT matching.

To make the sparse matching evenly distributed in the whole frame, as illustrated in Fig. 4.2 (c). Each frame is divided into several small patches. SIFT matching are performed in each of these corresponding patches. By doing so, it is ensured that there exists SIFT matchings in each of the patch. In this thesis, the patch size is set as 120*120

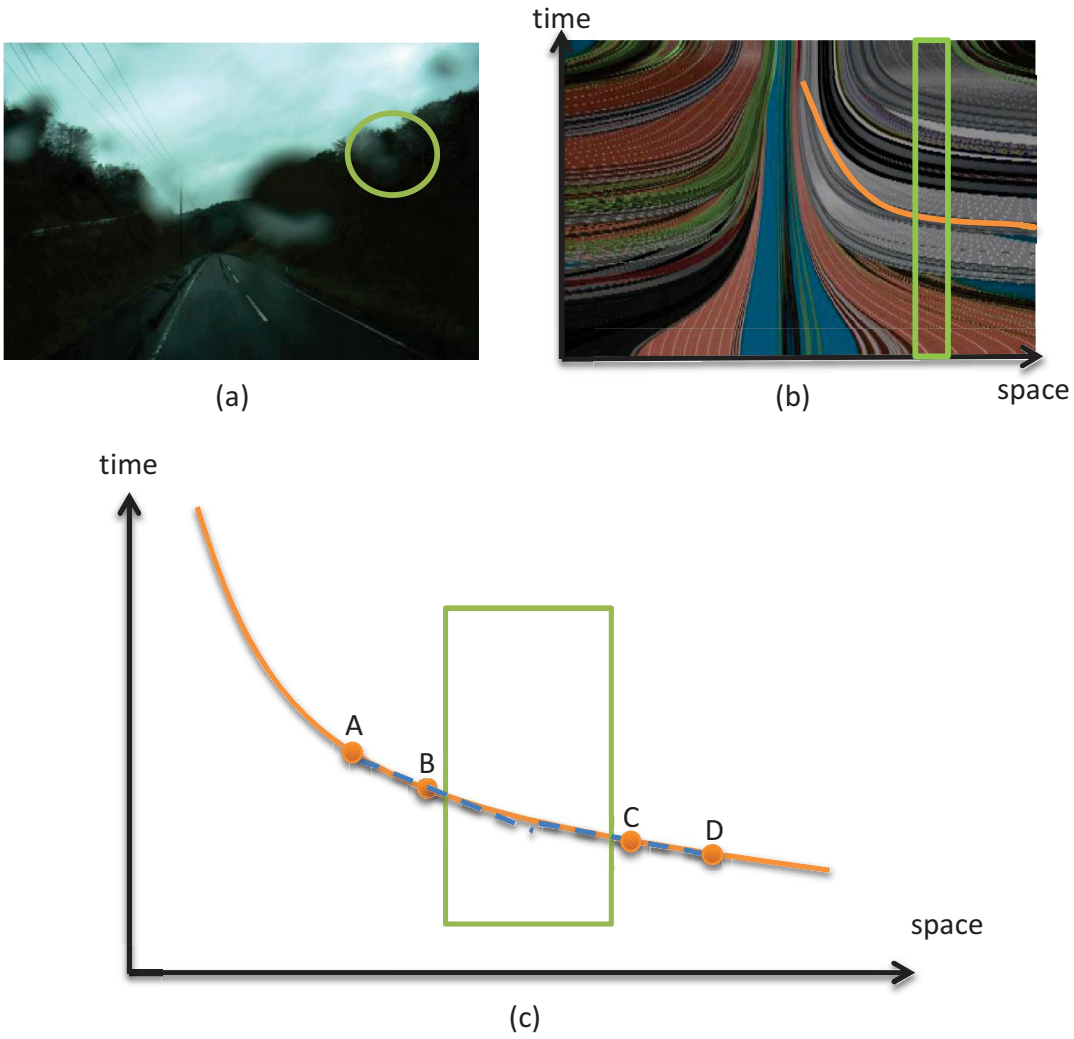


Figure 4.1: Recover the motion trajectory by linearly interpolation.
Green area denotes those cover by raindrops. (a) Fast moving area covered by a raindrop. (b) Fast moving object trajectory in spatial-temporal view. (c) Object trajectory covered by raindrop would be well approximated by linearly interpolate the motion from its neighboring area.

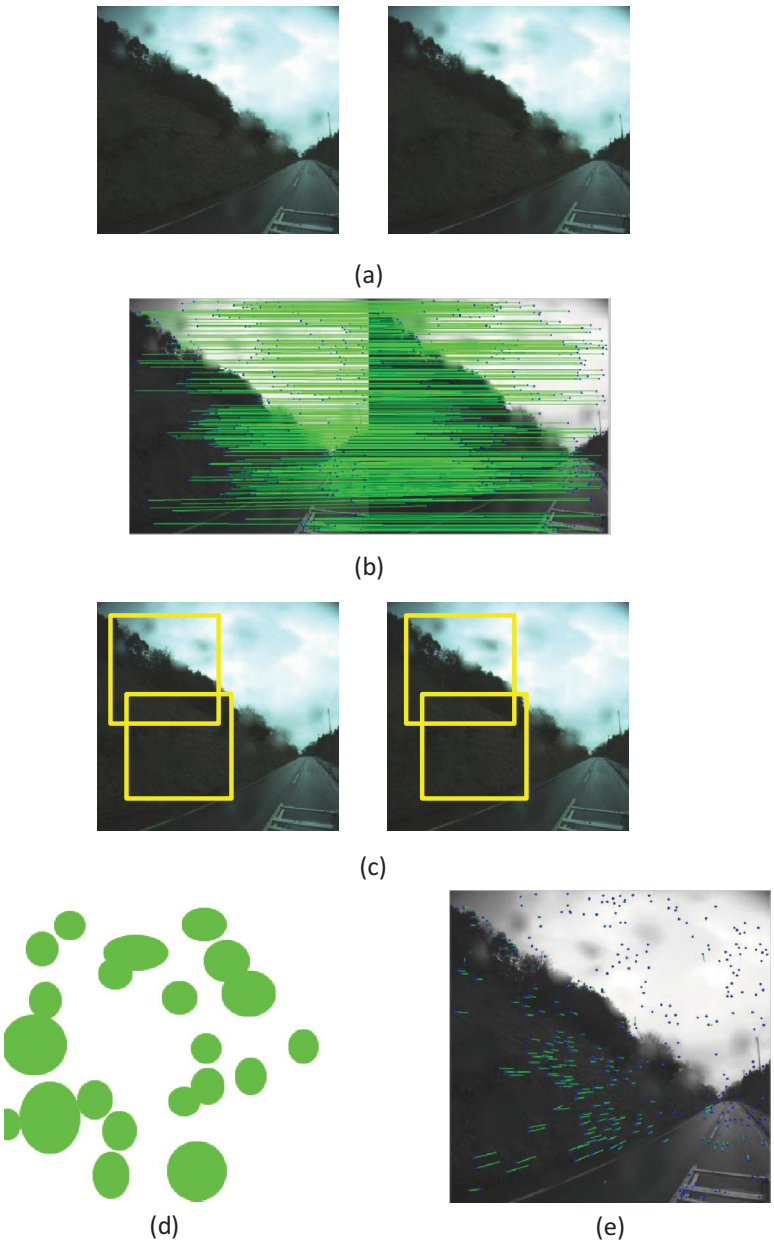


Figure 4.2: Sparse inter-frame motion from sparse SIFT matching.
(a) Two succeeding frames. (b) Sparse SIFT matching between two frames.(c) Patch by patch SIFT matching to make the match points uniformly distributed. (d) Matches on the raindrops are not useful, and thus removed. (e) Sparse motion by overlapping the two frames.

pixels as default. Every patch overlaps its neighboring patch in order to not to missing the matchings on the boundary of each patch.

Then, as illustrated in Fig. 4.2 (d). those matches in the raindrop area are not useful, and are thus removed according to the raindrop area detection result. For sparse matching, every correspondence is independence so that these motion correspondence are not affected by the raindrops.

Finally, by overlapping two frames together, we get the sparse inter-frame motion which are well distributed and not affected by raindrops.

4.1.4 Refined Sparse Matching Using Focus Point Constraints

To improve the robustness of sparse matching, besides RANSAC, directional constraints are used.

For cameras mounted on a vehicle, the motion of camera is limited. While a totally free camera has 6 motion freedoms, a vehicle mounted camera has very limited motion freedom. The motion of the camera is assumed to be the same of the vehicle. In a very short time period, say the time interval between taking 2 frames, the motion of the vehicle can be considered as 1 dimensional linear motion.

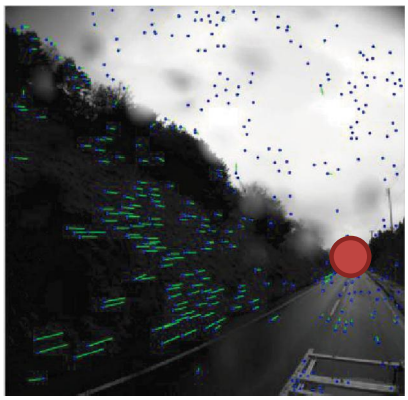
According to this constraint, as illustrated in Fig. 4.3(a), all the object motion between 2 neighboring frames should point to the a focus point. Usually, if the car is moving along the street, the focus point is located at the vanishing point of the street. If the car is turning its direction, the focus point will deviate from the vanishing point.

To estimated the focus point, similar as the Hough transform [10], as illustrated in Fig. 4.3 (b), we let each sparse matching vote for the position such that minimizes the angle between the sparse matching and the focus point: θ_i. Then, the the focus point is determined as the position minimizes the overall angle:

$$\Theta = \sum_i \theta_i \tag{4.1}$$

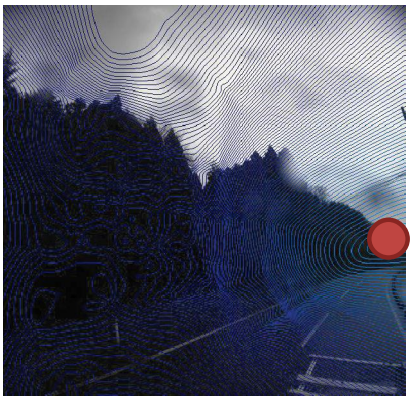
Fig. 4.3 (c) is an example of the distribution of the overall angle Θ on the whole frame.

After estimated the focus point, those matchings pointing the direct significantly deviate to the focus point are considered as wrong matching and thus need to be removed. Usually, the threshold is set as 15^o.



(a)

(b)



(c)

Figure 4.3: Refined Sparse Matching Using Focus Point Constraints.
(a) Because the motion of the vehicle mounted camera are limited, the motion of all the objects in the video should point to a focus. (b) To estimate the focus point, each matching vote for the position that minimized the angle between the motion and focus. (c) Example of the voting error from all matchings on a frame.

4.1.5 Dense Motion from Interpolating The Sparse Matching

The dense motion between two succeeding frames could be estimated by interpolating the sparse matching. Specifically, in this thesis, 2D polynomial interpolation is used. Fig. 4.4 is a comparison of motion estimation between SIFT-flow and our proposed sparse matching method. While SIFT-flow is affected by raindrops, our sparse matching based method successfully estimates the inter-frame motion.

4.1.6 Separating Video Completion Situations According to Motion Estimation

Previously, we have categorized the video completion tasks into 3 situations based on their motion behavior: fast motion, slow motion and static. By far, how to classify each raindrop area into these situations has not been introduced yet.

Intuitively, the classification is performed according to the motion surrounding the raindrop. As discussed previously, optical-flow based motion estimation is not robust in the surrounding area of raindrops and thus are not directly useful for the classification task. Thus, the proposed sparse matching and interpolation based motion estimation method would be used to perform the classification.

In this thesis, usually, the threshold for fast motion area is set that the norm of the inter-frame motion is greater than 5 pixels:

$$||(\mathrm{d}x, \mathrm{d}y)|| \geq 5 pixels \tag{4.2}$$

Given that in a 720*480 pixels image, the diagram of raindrops is usually smaller than 50 pixels. Eq. 4.2 means that, in fast motion area, if a object moves into a raindrop area, it will move out of the raindrop area with 10 frames.

As to classify the slow motion area from static area. Chromatic clues as used to improve the accuracy. This is because very small motion will be affected by noise. For street view video, usually, the sky area has the chromatic obviously differ from the street area which could be used as an robust clue for the classification.

After classify raindrop areas intro 3 situations, in this section, we propose the method to repair the fast motion situation. Method to repair other 2 situations are introduced in the next 2 sections correspondingly.

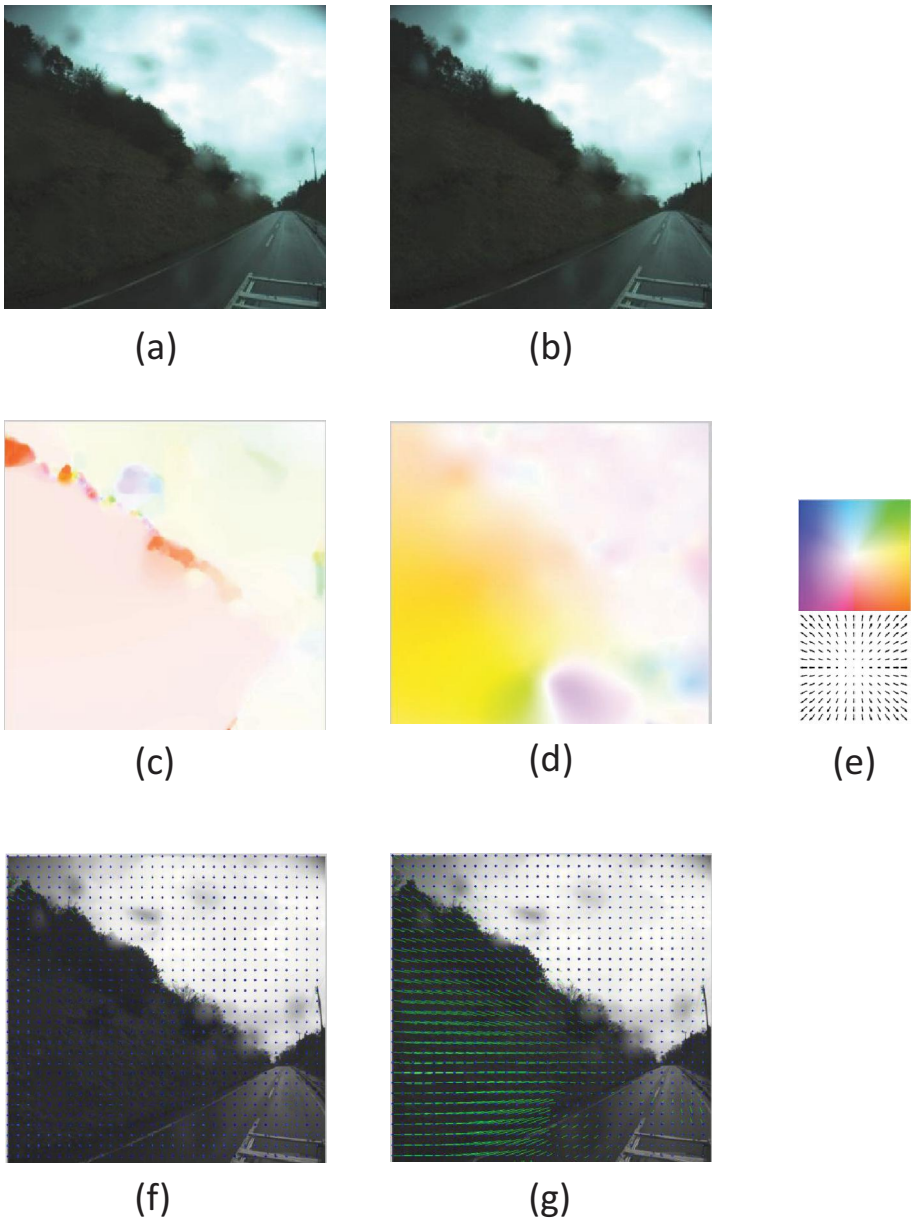


Figure 4.4: Comparison of inter-frame motion estimation.
(a) and (b) Two succeeding frames. (c)-(e) Colormap representation of the motion estimation results. (f)-(g) Motion needle representation of the motion estimation results. (c) and (f) Motion estimation using SIFT-flow. The result is erroneous because of the influence from raindrop area. (d) and (g) Motion estimation by interpolating sparse SIFT matching.

4.1.7 Video Completion According to Motion Trajectory

After densely modeling the motion in the raindrop area, the raindrop area can be repaired by copying the appearance from other frames according to the motion. Since the motion estimation is dense, this copying could be done pixel-wisely. Fig. 4.5 is an example of the repairing result.

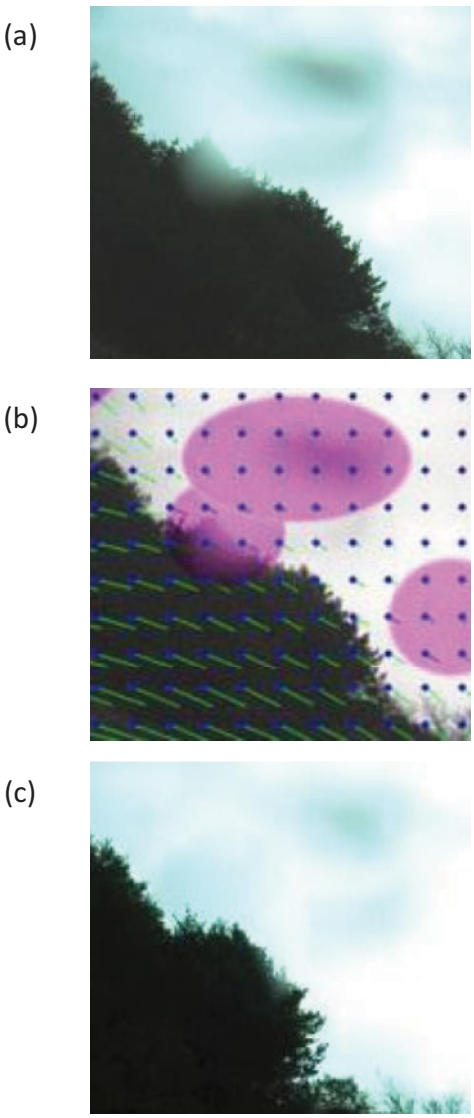


Figure 4.5: Video repairing using based on motion estimation.
(a) Area covered by raindrop. (b) Estimated motion in the raindrop area. (c) Repaired image by copying pixels from temporal-neighboring frame according to the motion.

4.2 Video Completion in Slow Motion Area

In this section, method to repair the slow motion but texture area are introduced.

4.2.1 Motion Simplification

As illustrated in Fig. 4.6, in the slow motion area, although the motion trajectory is not totally covered by a raindrop, modeling the motion trajectory using linear interpolation is not accurate.

Fortunately, in slow motion area, we can neglect the object perspective change. There are two kind to situation that the object motion is not fast changing:

I. Distanced from the camera. Since the camera motion is limited, the farther the object is away from the camera, the slower the inter-frame motion is. For very distanced object, its depth changing in its surrounding area is negligible in comparing with the depth. Thus, the perspective change is also negligible. And, thus, the appearance change of this object can be simplified as scaling and translation.

II. Close to the vanishing point. As discussed in the previous section, all object motion are pointing to the focus point. And, the closer the object to the the focus point, the slower the motion is. Usually, the focus point is very close to the vanishing point. Object appearance at the vanishing point has not perspective changes. Thus, in the area close to the vanishing point, the appearance change of the object can be simplified as scaling and translation.

4.2.2 Video Completion Using Image Registration

According to the discussion above, the appearance change in the slow motion area can be simplified as scaling and translation. So that slow motion area can be repaired using image registration.

Registration of images could be done with out much effort. By using SIFT matching, correspondence between 2 frames has already been acquired. As introduced in the previous section, the focus of each frame has already been estimated. To register 2 frames, translation are performed to set the focuses of two frames at the same position. Then, the scaling coefficient is calculated according to the distance from the matching point to the focus point. Fig. 4.7, is the result of registering frames with different temporal distance.

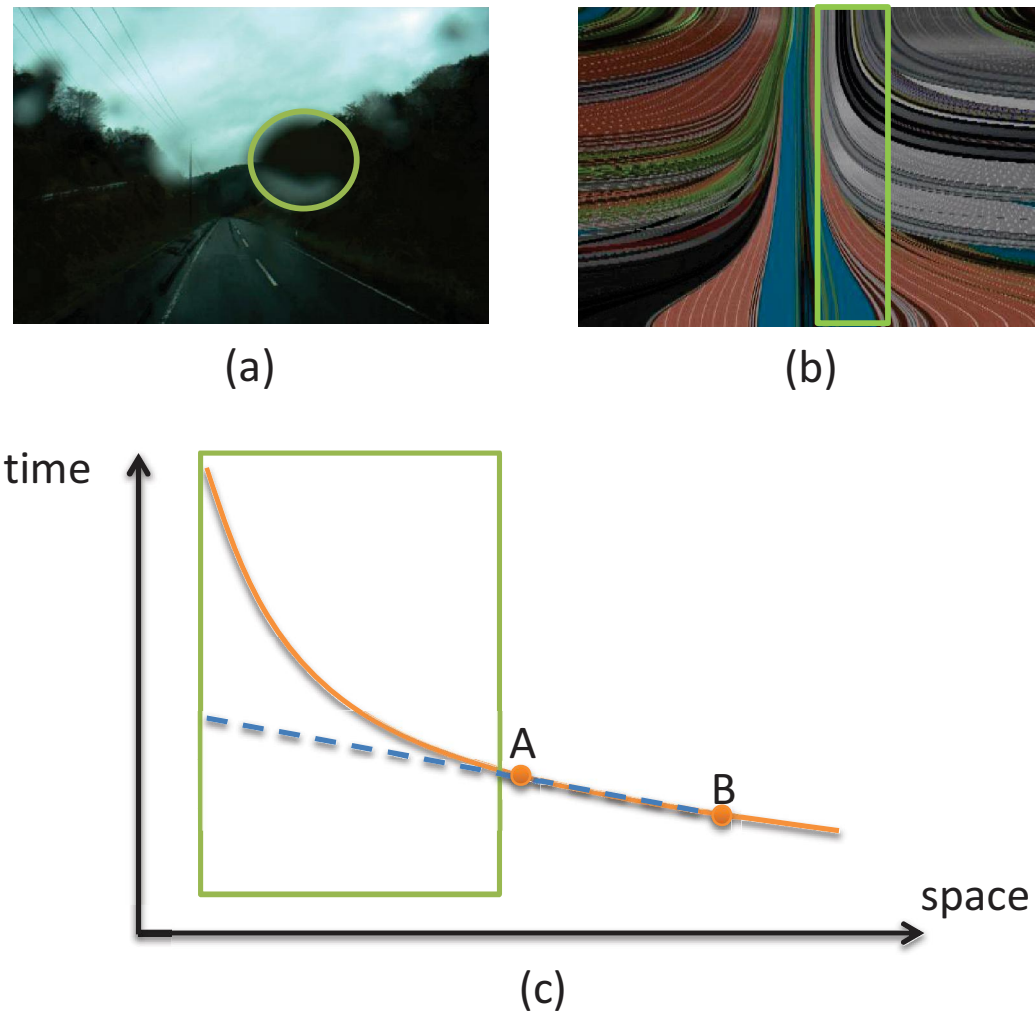


Figure 4.6: Motion interpolation does not work in slow motion area. Green area denotes those cover by raindrops. (a) Slow moving area covered by a raindrop. (b) Fast moving object trajectory in spatial-temporal view. (c) Object trajectory covered by raindrop can not be well approximated using interpolation.

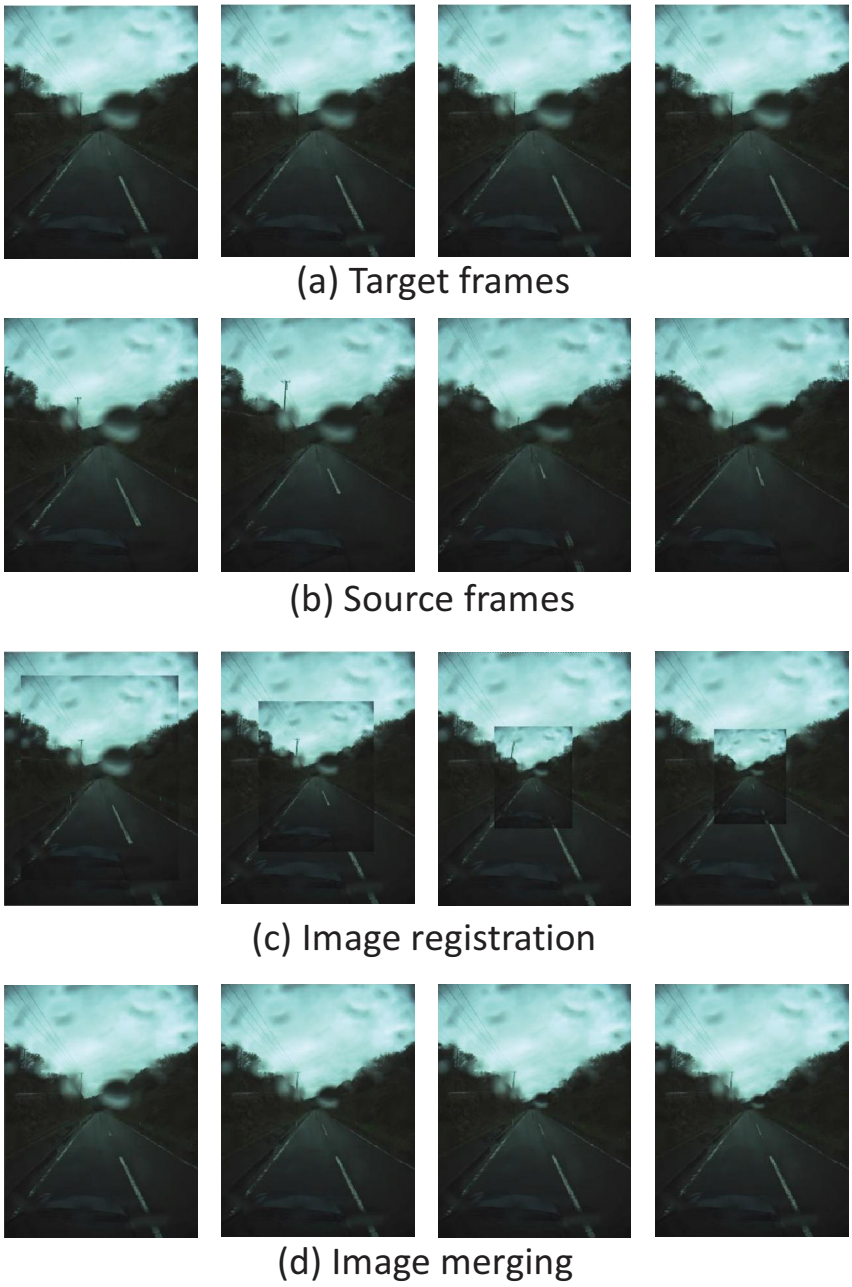


Figure 4.7: Video Completion Using Image Registration.
(a) Targeting frames, four frames are identical (b) Source frames, from left to right, the temporal distance from source frame to target frame are: 10 frames, 20 frames, 30 frames and 40 frames. . (c) Register the source frame to the target frame using translating and scaling. (d) Merge source frame to target frame in the slow motion area covered by raindrop. The scale of raindrop has been significantly decreased.

By image registration and merging, one can see that the size of the raindrop has been reduced in the slow motion area. By gradually increasing the temporal distance between the frames to register, the scale of the raindrop will gradually decrease.

However, the raindrop will not be totally removed by only changing the scale. Fortunately, if a raindrop is small enough, say smaller than 10 pixels, the structured information can be neglected. And, when the scale of raindrop is small enough, one can see that the motion in the surrounding area is almost static. Therefore, now the area covered by the small raindrop is considered as: static and non-structure. Hence, the remaining work of repairing this small raindrop area can be classified to the third situation.

4.3 Video Completion in Static and Non-structured Area

In this section, the method to repair the video on static and non-structured area is introduced.

4.3.1 Task Simplification

As introduced previously, in a street view video, the sky area and the close-to-focus area are the static and non-structured area. As illustrated in Fig. 4.8, in static area, the object trajectory will be totally covered by raindrops. Therefore, the video cannot be repaired by copying appearance from other frames. This kind of situation can be repaired only be image repairing methods.

For image inpainting, one challenge is to repair the structure information [6, 1]. Fortunately, as has been pointed out, both the sky area and the close to focus area are considered as non-structured, thus the task is simplified as texture synthesis.

4.3.2 Modified Image Inpainting

Our texture synthesis method is proposed based on Criminisi's image inpainting method. [6, 1]. It has four steps:

I. As illustrated in Fig. 4.9(a) the image has the texture source area and the blank target area.

II. As illustrated in Fig. 4.9(b), the inpainting task starts from choosing the target patch (yellow square) to be inpainted. The center of the patch is always located on the boundary between the source area and target area. For each patch with the center located on the boundary, the patch confidence is calculated as:

$$confidence = \frac{area\ in\ source\ region}{area\ of\ the\ patch} \tag{4.3}$$

After updating the confidence value along the boundary, the patch with the max confidence is chosen as the patch to inpaint. As default, the size of the target patch is set as 8*8 pixels in this thesis.

III. As illustrated in Fig. 4.9(c), by searching along the whole source region, the source patch which has the minimum intensity difference with the target patch is chosen as the optimal source patch. In this thesis, to optimize the computational efficiency, the searching area is limited as a 50*50 patch centered at the center of the target patch.

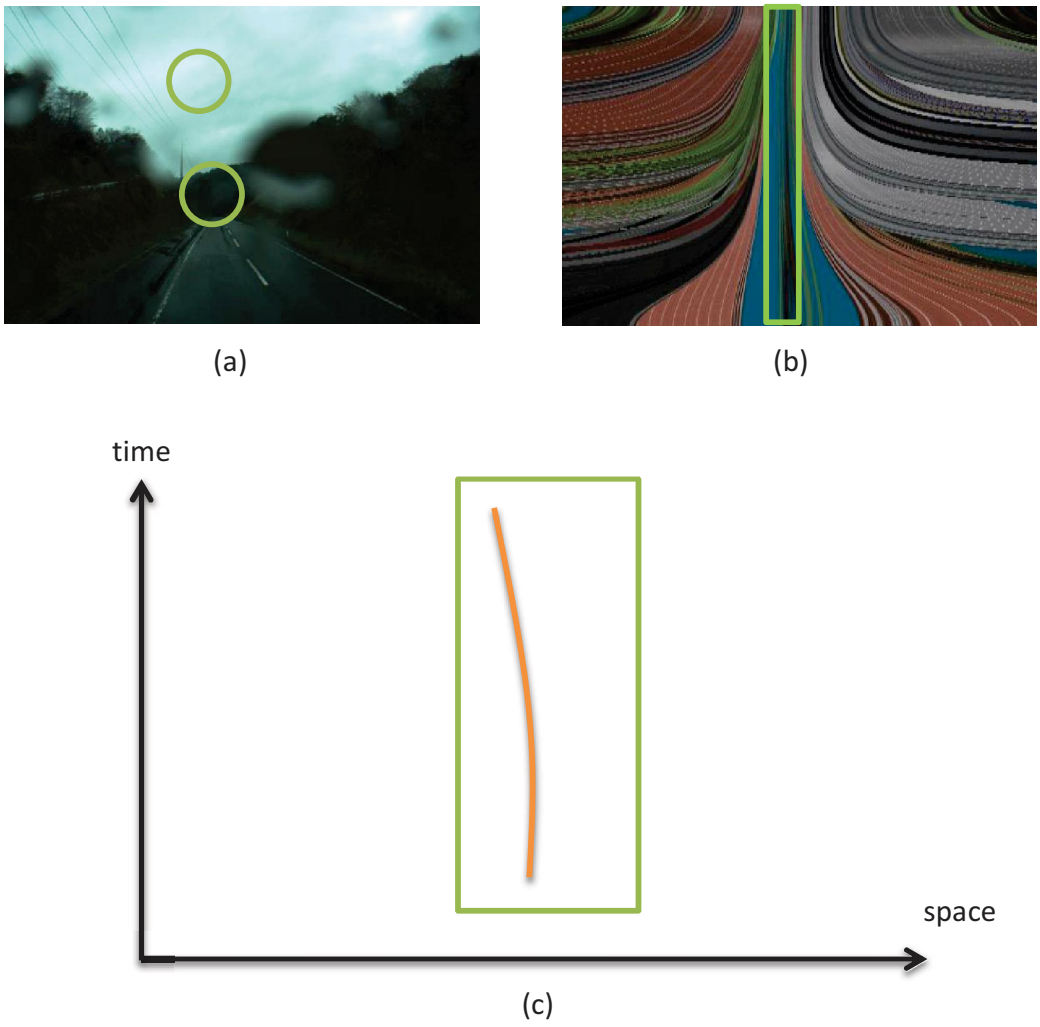


Figure 4.8: No motion information available in static area.
Green area denotes those cover by raindrops. (a) Static area covered by a raindrop. (b) Static object trajectory in spatial-temporal view. (c) The whole object trajectory is covered by raindrops. No motion information is availble.

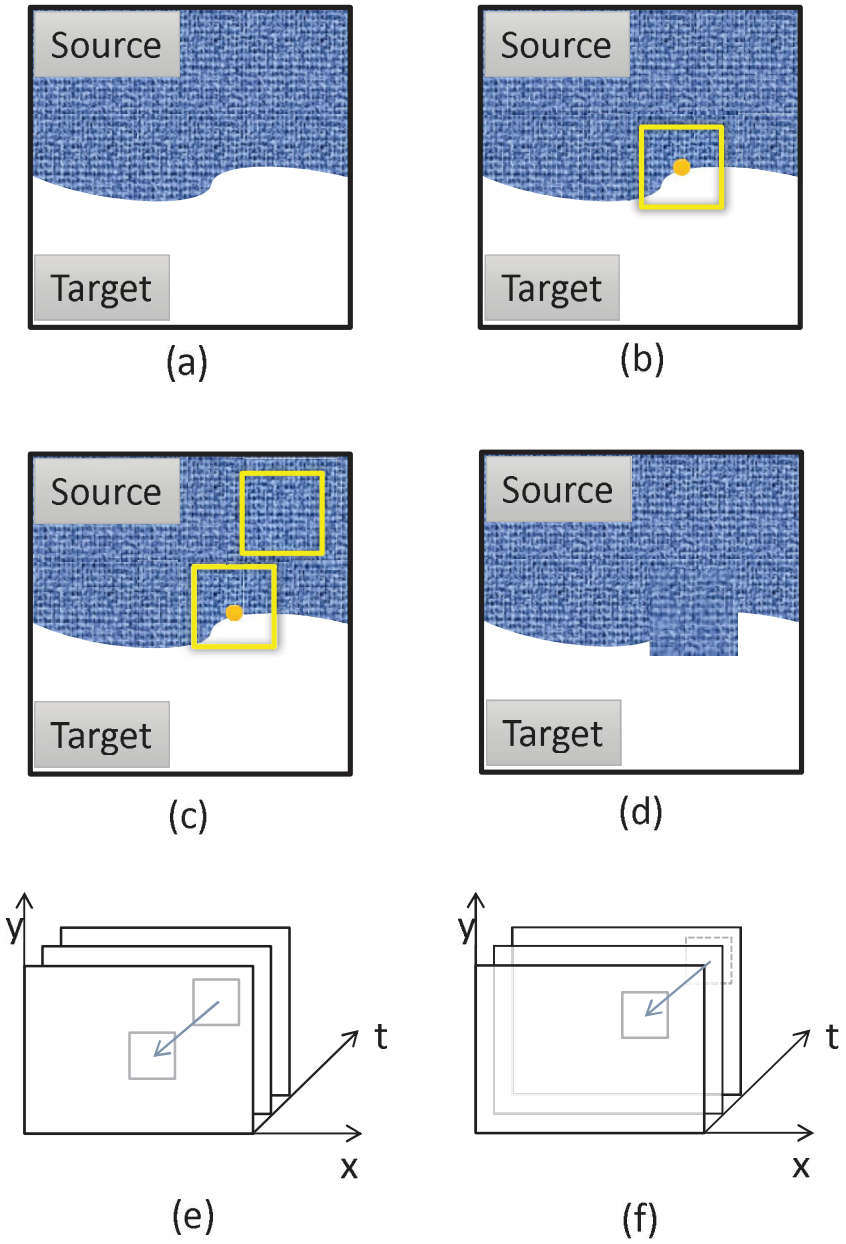


Figure 4.9: The proposed image inpainting method.
(a) Target region and source region. (b) The target patch is centered as the point with max confidence value on the boundary. (c) The source patch has the minimum intensity difference with the target patch. (d) Target patch is inpainted by copying source patch. (e) Source region is in the same frame for image inpainting. (f) For video inpainting, source region can be extended to neighboring frames.

IV. As illustrated in Fig. 4.9(d), the target patch is inpainted by copying the source patch to the target patch.

Then, the source region and target region are updated and the inpainting progress is continued by repeating Steip I to Step IV until the whole target region has been inpainted.

Although there is scarly no motion change in the static area, there might exists illumination changes between frames. To improve Step III, when inpainting videos, the source region can be extended from the same frame to several temporally neighboring frames. Fig. 4.9(e), (f) illustrated this extension.

Fig. 4.10 are several results of video repairing on static area using our modified image inpainting method.

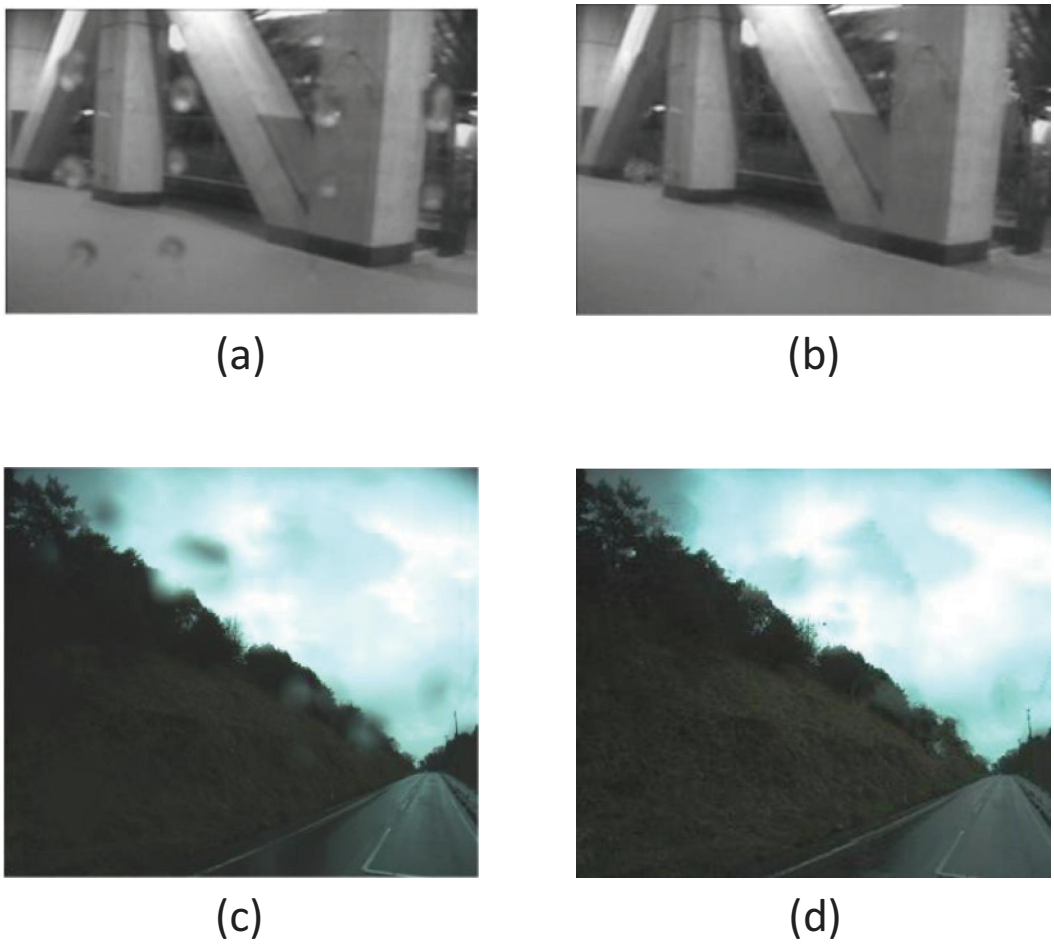


Figure 4.10: Image inpainting using the proposed method.
(a) and (c) Input image. (b) and (d) Inpainted image.

4.4 Overall Workflow

The overall workflow for video completion is illustrated as in Fig. 4.11. It is mainly consisted with 3 steps:

Step1: Dense motion estimation using our proposed sparse matching and interpolation based method.

Step2: Situations classification according to the estimated dense motion and chromatic clues.

Step3: Case by case video repairing.

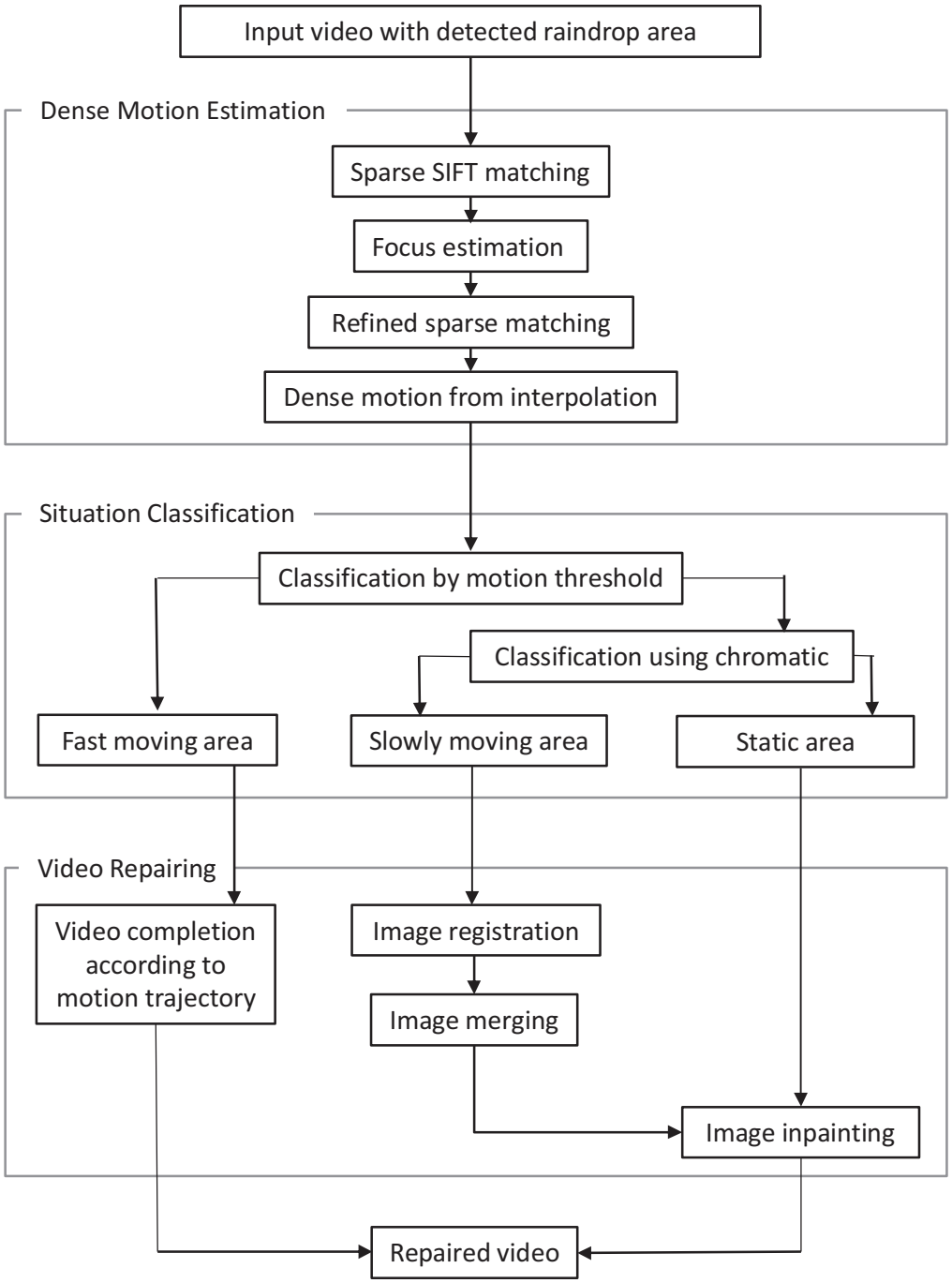


Figure 4.11: The overall workflow for video completion

Chapter 5

Experimentation Results

In this chapter, the effectiveness of the proposed raindrop detection method and video completion method are demonstrated with plenty of experiments.

Quantitative experiments for raindrop detection is described in the first section. And quantitative experiments for video completion is described in the section section.

We mention that different data set are used for quantitative experiments on raindrop detection and video completion. This is because:

1. For raindrop detection, the data with adherent raindrop are taken in real world, there is no ground truth of the appearance information covered by raindrops. Therefore, the data taken for raindrop detection has no ground truth for video completion.

2. For video completion, the raindrops area are intentionally added, so that there exists the ground truth of the appearance information for video completion. However, as previously mentioned, quantitatively modeling and generating the appearance of raindrops is not applicable. This means the artificial raindrops cannot be used to benchmark the effectiveness of raindrop detection methods.

In the third session, several experiments on raindrops removal in real video with raindrops are provided without quantitative analysis.

In all the above sections, our proposed raindrop detection and video completion methods are compared with existing raindrop detecting and video completion methods correspondingly.

5.1 Raindrops Detection

5.1.1 Equipments and settings

All our experiment data were taken in the real world. We used Victor GZ-MZ555-S video camera. The lens focus is $6.3mm$ and the aperture is $1 : 3.5$. To protect our camera, we put the raindrops on a protecting shield which is close to the video camera. The distance from the camera to the protecting shield is $d = 100mm$. The tilt θ of the shield is adjustable; usually it is set vertical to the ground: $\theta = 0$. The video data were taken in a short time after sunset to mimic the rainy day environment.

5.1.2 Detection of raindrops in various situations.

To demonstrate the generality of our method, the experiments data were taken by varying the following aspects: shape, size, blurring and tilt. Each of the variations is summarized in Table 1.

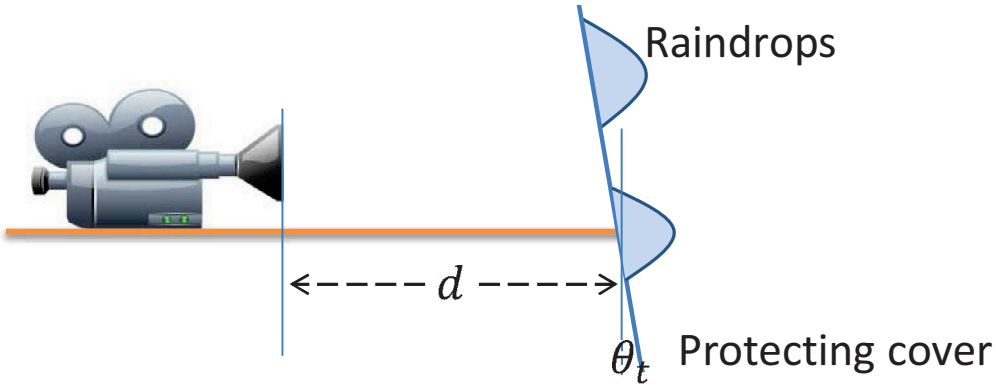


Figure 5.1: Our equipment to take experimental data.

5.1.3 Comparison with existing methods

To demonstrate the efficiency of our method, we compared it with Roser's [20] and Kurihata's [16] method. The results are shown in Figs. 5.1.2-5.1.5.

5.1.4 Quantitative analysis

We use the precision-recall curve to quantitatively analyze the performances. The precision is defined as the number of correct detections divided by the number of all the detections. Recall is defined as the correct detections divided by the number of detectable raindrops. The results for each experiment are shown in Fig. 5.6.

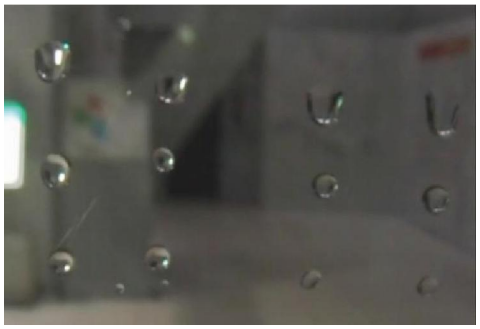
According to the results, both of our proposed method outperform the existing methods. And detection using intensity change has the best performance. While existing method only detect the center and size of raindrops. Our proposed method can detect raindrops with a large variety of shapes, and even water streaks in the second data.

5.1.5 Tohoku data

As shown in the last column of Figs. 5.1.5, we detected raindrops from videos taken by a vehicle-mounted camera. The camera is mounted on top of the car. There is no protecting shield, the raindrops is directly adherent to the lens, and thus significantly blurred. Our proposed method has successfully detected all the raindrops visible and lower false alarm in comparing with the existing methods.

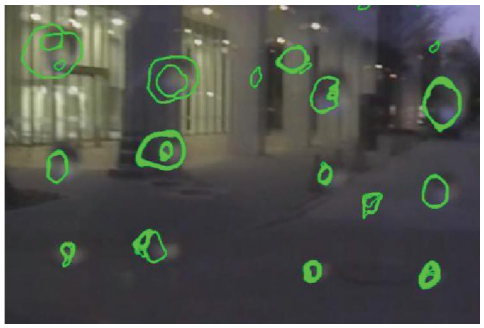
Table 5.1: Parameters and setting of the experiment data

	Raindrop Shape	**Video size** (pixel)	**Size** (pixel, diameter)	**Blurring radius** (pixel)	**Tilt** (degree)	**No. of visible raindrops** (after blurring)
Experiment 1	mostly round	720×480	30 - 80	7	0	12
Experiment 2	various	720×480	15 - 320	20	0	20
Experiment 3	various	720×480	9 - 85	16	0	18
Experiment 4	mostly round	720×480	8 - 60	7	45	20
Car-mounted	mostly round	600×400	N/A	N/A	0	6

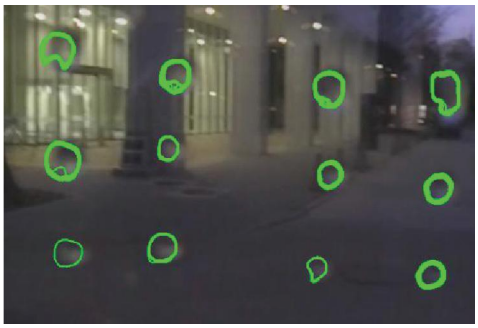


Ground truth

Input

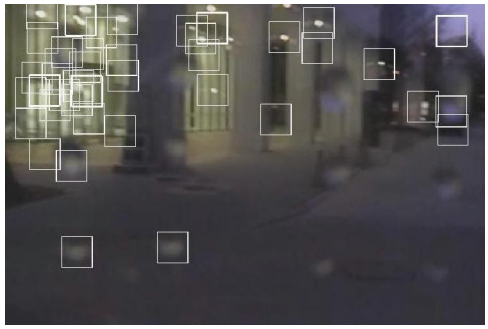


Optical flow



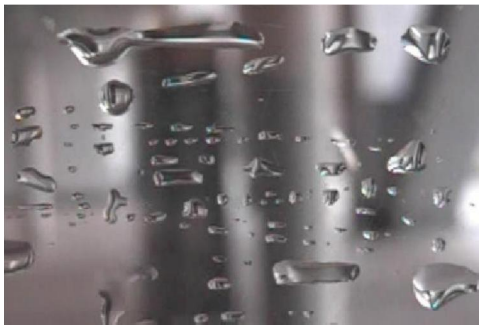
Intensity change

Kurihata 2005



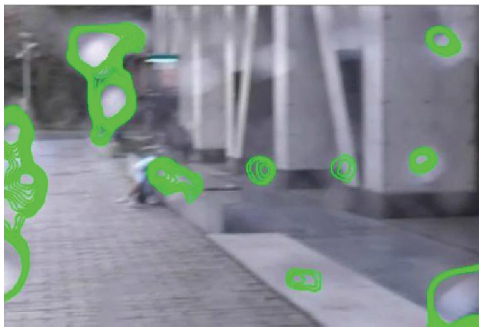
Roser 2009

Figure 5.2: Detection experiment using our proposed methods and existing methods (1).



Ground truth

Input



Optical flow

Intensity change

Kurihata 2005

Roser 2009

Figure 5.3: Detection experiment using our proposed methods and existing methods (2).

Ground truth

Input

Optical flow

Intensity change

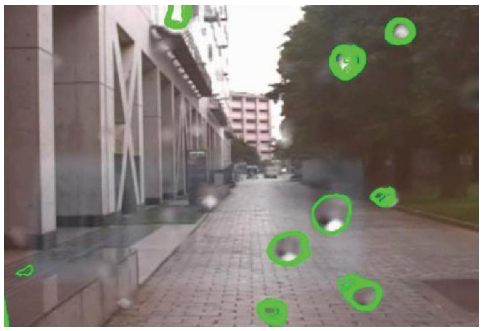
Kurihata 2005

Roser 2009

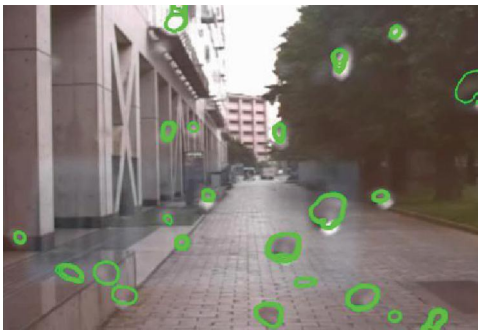
Figure 5.4: Detection experiment using our proposed methods and existing methods (3).

Ground truth

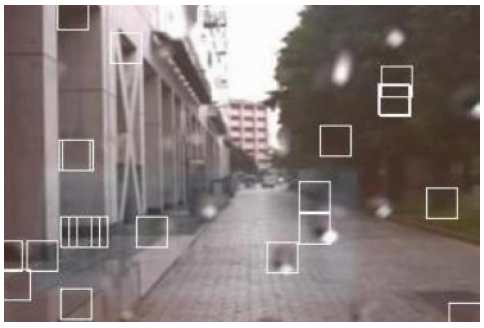
Input



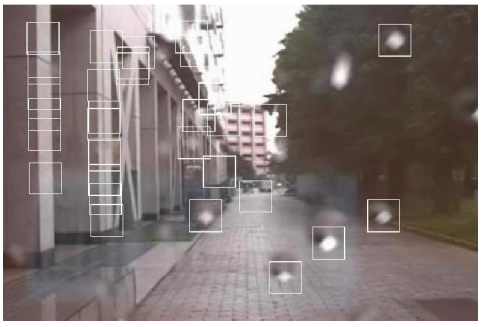
Optical flow



Intensity change



Kurihata 2005



Roser 2009

Figure 5.5: Detection experiment using our proposed methods and existing methods (4).

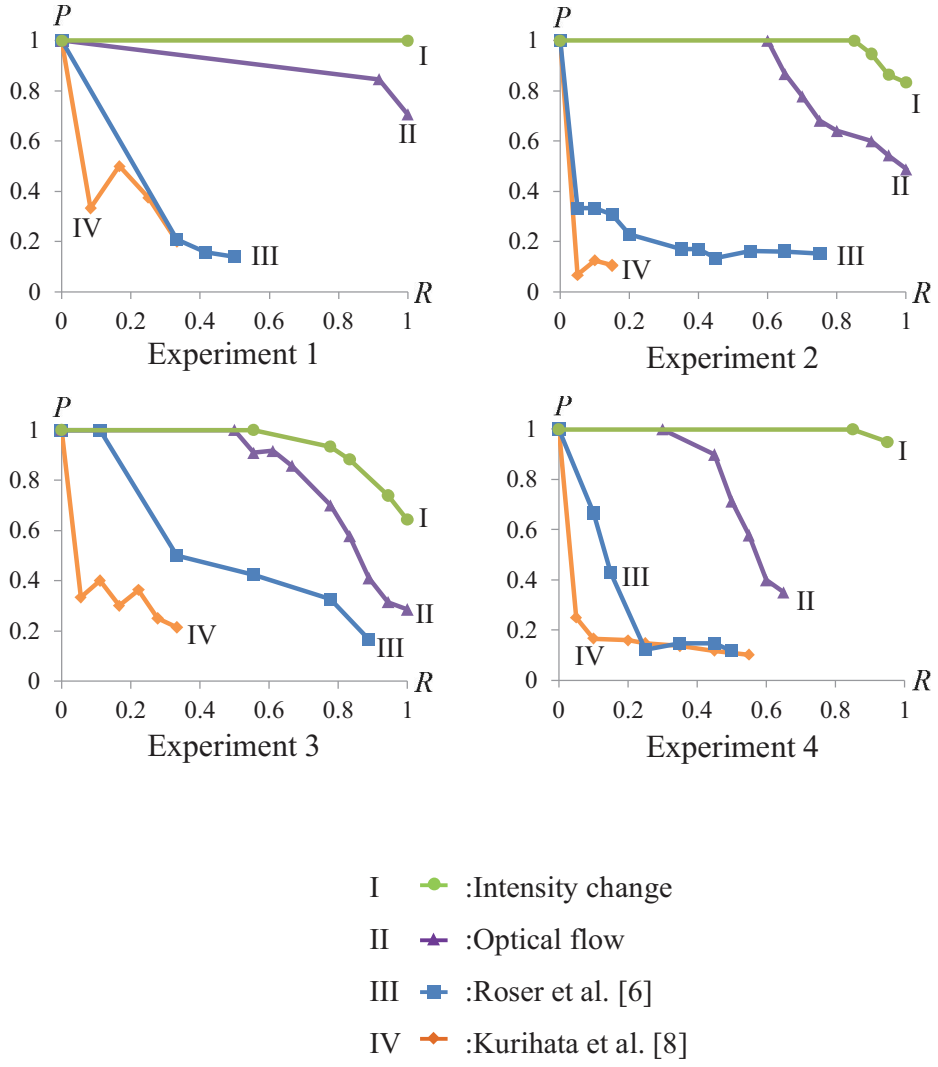


Figure 5.6: Precision-recall curves of our propose methods and two existing methods. Precision (P) is in y-axis and recall (R) is in x axis.

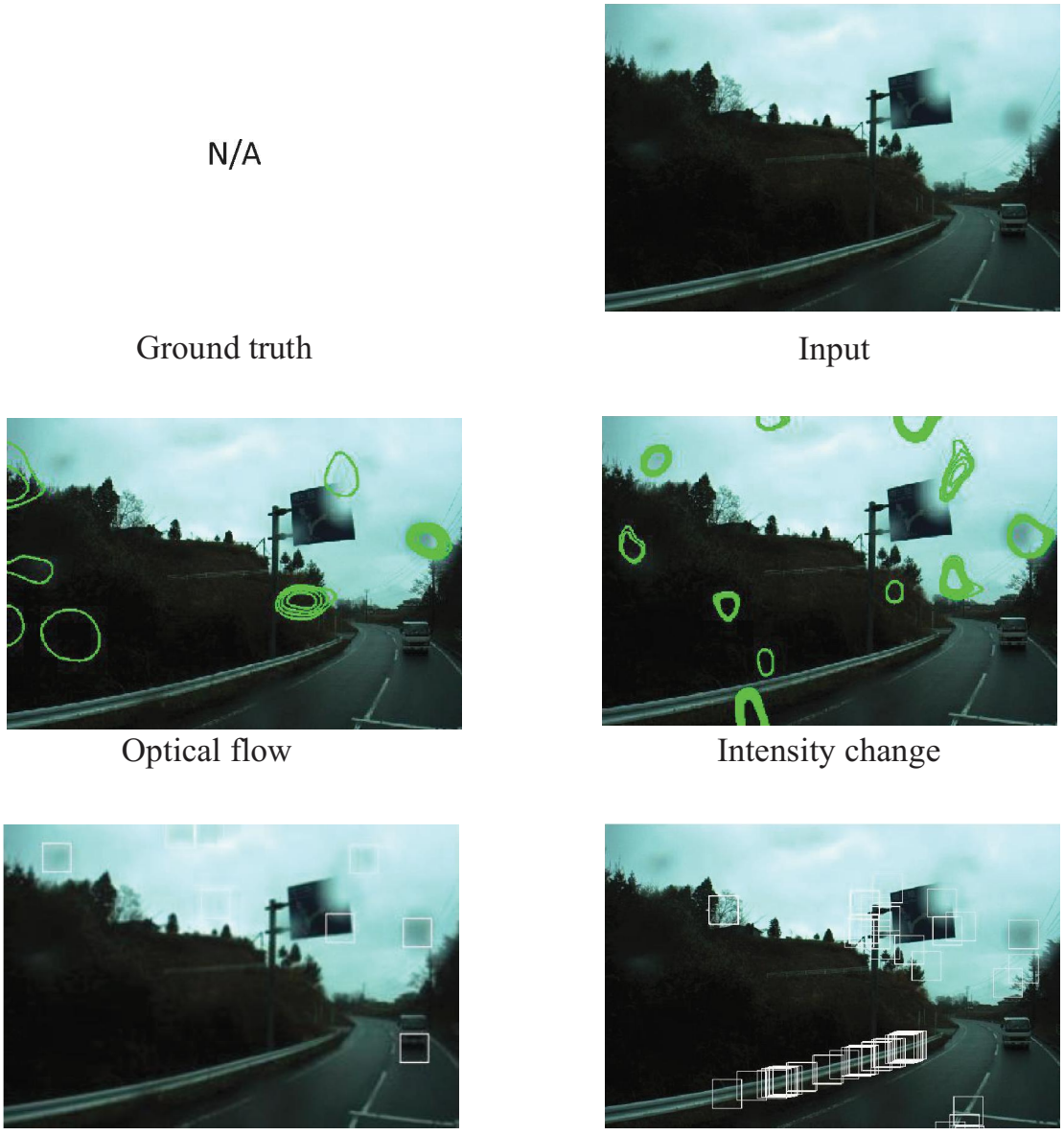


Figure 5.7: Detection experiment on Tohoku data using our proposed methods and existing methods.

5.2 Video Completion

In this section, experiments to quantitatively analyze the effectiveness of the proposed video completion methods are provided.

5.2.1 Experiments and settings

To evaluate the effectiveness of the video completion method, we use data taken by vehicle mounted camera in cloudy days. The cloudy day environment mimics the rainy day environment but do not have adherent raindrops. Raindrop areas to be removed are manually label. All the situations: fast moving area, slowly moving area and static area are tested.

5.2.2 Comparison with existing methods.

To demonstrate the effectiveness, our method are compared with existing methods. As introduced in related works, the following 3 methods are chosen:

1. Image inpainting. (Criminisi et al. [6, 1]).
2. Space-time completion. (Wexler et al. [30])
3. Motion based completion. (Shiratori et al. [27])

5.2.3 Quantitative Evaluation

For quantitative analysis, we calculate the average intensity difference between the repaired image and the ground truth.

For each pixel in the repairing area, it is considered to be a RGB pixel:

$$I = (R, G, B) \tag{5.1}$$

And each pixel is 8-bit from 0 to 255.

For each repaired pixel I_r, its error from the groundtruth I_o is calculated as:

$$E = ||I_r - I_o||_0 = |R_r - R_0| + |G_r - G_0| + |B_r - B_0| \tag{5.2}$$

For each experiment and each repairing method, the repairing error is calculated as the average error of all repaired pixels. The average error of 4 methods in 3 experiments are listed in Table 3.

(a) Ground truth

(b) Input video

(c) Image inpaiting

(d) Space-time video completion

(e) Dense motion based video completion

(f) Our proposed method

Figure 5.8: Video completion in fast moving area using our proposed methods and existing methods.

(a) Ground truth

(b) Input video

(c) Image inpaiting

(d) Space-time video completion

(e) Dense motion based video completion

(f) Our proposed method

Figure 5.9: Video completion in slowly moving area using our proposed methods and existing methods.

(a) Ground truth

(b) Input video

(c) Image inpaiting

(d) Space-time video completion

(e) Dense motion based video completion

(f) Our proposed method

Figure 5.10: Video completion in static area using our proposed methods and existing methods.

Table 5.2: Comparison on average repairing error.

	Fast motion area	Slow motion area	Static area
Image inpainting	13.5	24.4	90.0
Space-time video completion	10.0	19.6	1.7
Dense motion based video completion	80.9	77.1	190.7
Our method	13.2	20.7	3.5

5.2.4 Tohoku Data

Lastly, we show the results of video completion on real data with adherent raindrop taken in Japanese northeastern area. Our experiments are also compared with three existing methods. The results are shown in Figs. 5.2.4 and 5.2.4.

(a) Input video

(b) Image inpaiting

(c) Space-time video completion

(d) Dense motion based video completion

(e) Our proposed method

Figure 5.11: Raindrop removal on Tohoku data using our proposed methods and existing methods (I).

(a) Input video

(b) Image inpaiting

(c) Space-time video completion

(d) Dense motion based video completion

(e) Our proposed method

Figure 5.12: Raindrop removal on Tohoku data using our proposed methods and existing methods (II).

Chapter 6

Conclusion

In this thesis, we propose method to remove raindrops in video.

For raindrop detection, we have introduced a novel method of detecting adherent raindrops in video. We mathematically analyzed the relation between the appearance of raindrop and non-raindrop pixels. Our key idea of detecting raindrops is based on our theoretical findings that the motion of raindrop pixels is slower than that of non-raindrop pixels, and the change of intensity of raindrop pixels is smaller than that of non-raindrop pixels. We believe, our automatic raindrop detection method is novel and can benefit many applications that suffer from adherent raindrops.

With detected raindrops, algorithm to automatically repair the video in complex outdoor environment Has been proposed. We categorized outdoor environment into 3 situations and proposed video repairing algorithms correspondingly. Firstly, for fast motion, structured and textured video, sparse matching based algorithm to estimate dense motion in video with data missing are described. Raindrop removal situation classification based on the dense motion are introduced. Then, raindrop removal for fast motion, structured and textured area by using dense motion are proposed. Secondly, for slow motion, structured and textured area, raindrop removal method using image registration are introduced. Lastly, for static and non-structured area, a modified image inpainting method is introduced.

As future work, we consider to further exploit the video completion algorithm in aiming of handling more complex situations, better robustness and better accuracy.

References

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

[2] M. Bertalmio. Navier-stokes, fluid dynamics, and image and video inpainting. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[3] M. Bertalmio. Strong-continuation, contrast-invariant inpainting with a third-order optimal pde. *IEEE Transactions on Image Processing*, 15(7):1934–1938, 2006.

[4] V. Cheung, B. J. Frey, and N. Jojic. Video epitomes. *International Journal of Computer Vision*, 76(2):141–152, 2008.

[5] A. Criminisi, P. Perez, and K. Toyama. Object removal by exemplar-based inpainting. *IEEE International Conference on Computer Vision*, 2003.

[6] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9):1200–1212, 2004.

[7] T. Ding, M. Sznaier, and O. I. Camps. A rank minimization approach to video inpainting. *IEEE International Conference on Computer Vision*, 2007.

[8] K. Garg and S. Nayar. Photometric model of a raindrop. *CMU Technical Report*, 2003.

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

[10] J. Illingworth and J. Kittler. A survey of the hough transform. *Computer Vision, Graphics, and Image Processing*, 44(1), 1988.

[11] E. Jerison, Y. Xu, L. Wilen, and E. Dufresne. The deformation of an elastic substrate by a three-phase contact line. *Physical Review Letters*, 106(18), 2011.

[12] J. Jia, Y. Tai, T. Wu, and C. Tang. Video repairing under variable illumination using cyclic motions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):832–839, 2006.

[13] J. Jia, T. Wu, Y. Tai, and C. Tang. Video repairing: Inference of foreground and background under severe occlusion. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[14] L. Kang, C. Lin, and Y. Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Transactions on Image Processing*, 21(4):1742–1755, 2012.

[15] A. C. Kokaram. On missing data treatment for degraded video and film archives: A survey and a new bayesian approach. *IEEE Transactions on Image Processing*, 13(3):397–415, 2004.

[16] H. Kurihata, T. Takahashi, I. Ide, Y. Mekada, Hiroshi Murase, Yukimasa Tamatsu, and Takayuki Miyahara. Rainy weather recognition from in-vehicle camera images for driver assistance. *IEEE Intelligent Vehicles Symposium*, 2005.

[17] C. Liu, J. Yuen, and A. Torralba. Sift flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):978–994, 2006.

[18] M. Liu, S. Chen, J. Liu, and X. Tang. Video completion via motion guided spatial-temporal global optimization. *ACM international conference on Multimedia*, 2009.

[19] Y. Matsushita, E. Ofek, W. Ge, and X. Tang. Full-frame video stabilization with motion inpainting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1150–1163, 2006.

[20] M. Roser and A. Geiger. Video-based raindrop detection for improved image registration. *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, 2009.

[21] M. Roser, J. Kurz, and A. Geiger. Realistic modeling of water droplets for monocular adherent raindrop recognition using bezier curves. *Asian Conference on Computer Vision*, 2010.

[22] R. R. Sahay and A. N. Rajagopalan. Inpainting in shape from focus: Taking a cue from motion parallax. *British machine vision conference*, 2009.

[23] G. Sapiro and M. Bertalmio. Video inpainting of occluding and occluded objects. *IEEE International Conference on Image Processing*, 2005.

[24] G. Sapiro and M. Bertalmio. Video inpainting under constrained camera motion. *IEEE Transactions on Image Processing*, 16(2):545–553, 2007.

[25] T. K. Shih, N. C. Tan, J. C. Tsai, and H. Zhong. Video falsifying by motion interpolation and inpainting. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008.

[26] T. K. Shih, N. C. Tang, W. Yeh, and T. Chen. Video inpainting and implant via diversified temporal continuations. *ACM international conference on Multimedia*, 2006.

[27] T. Shiratori, Y. Matsushita, S. B. Kang, and X. Tang. Video completion by motion field transfer. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.

[28] N. C. Tang, C. Hsu, C. Su, and T. Shih. Video inpainting on digitized vintage films via maintaining spatiotemporal continuity. *IEEE Transactions on Multimedia*, 13(4):602–614, 2011.

[29] M. V. Venkatesh, S. S. Cheung, and J. Zhao. Efficient object-based video inpainting. *Pattern Recognition Letters*, 30(2):168–179, 2009.

[30] Y. Wexler, E. Shechtman, and M. Irani. Space-time video completion. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004.

[31] A. Yamashita, T. Harada, T. Kaneko, and K. T. Miura. Removal of adherent noises from images of dynamic scenes by using a pan-tilt camera. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004.

[32] A. Yamashita, S. Kato, and T. Kaneko. Robust sensing against bubble noises in aquatic environments with a stereo vision system. *IEEE International Conference on Robotics and Automation*, 2006.

[33] A. Yamashita, Y. Tanaka, and T. Kaneko. Removal of adherent water-drops from images acquired with stereo camera. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005.

[34] Y. Zhang, J. Xiao, and M. Shah. Motion layer based object removal in videos. *IEEE Workshops on Application of Computer Vision*, 2005.

[35] V. Zorich and R. Cooke. Mathematical analysis. *Springer*, 2004.