Degenerate Near-planar Road Surface 3D Reconstruction and Automatic Defects Detection

Yazhe Hu

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mechanical Engineering

Tomonari Furukawa, Chair
Jia-Bin Huang
Corina Sandu
Steve C. Southward

May 12, 2020
Blacksburg, Virginia

Keywords: Road condition monitoring, 3D road reconstruction, Degenerate 3D reconstruction, Road defects detection

Degenerate Near-planar Road Surface 3D Reconstruction and Automatic Defects Detection

Yazhe Hu

(ABSTRACT)

This dissertation presents an approach to reconstruct degenerate near-planar road surface in three-dimensional (3D) while automatically detect road defects. Three techniques are developed in this dissertation to establish the proposed approach. The first technique is proposed to reconstruct the degenerate near-planar road surface into 3D from one camera. Unlike the traditional Structure from Motion (SfM) technique which has the degeneracy issue for near-planar object 3D reconstruction, the uniqueness of the proposed technique lies in the use of near-planar characteristics of surfaces in the 3D reconstruction process, which solves the degenerate road surface reconstruction problem using only two images. Following the accuracy-enhanced 3D reconstructed road surface, the second technique automatically detects and estimates road surface defects. As the 3D surface is inversely solved from two-dimensional (2D) road images, the detection is achieved by jointly identifying irregularities from the 3D road surfaces and the corresponding image information, while clustering road defects and obstacles using a mean-shift algorithm with flat kernel to estimate the depth, size, and location of the defects. To enhance the physics-driven automatic detection reliability, the third technique proposes and incorporates a self-supervised learning structure with data-driven Convolutional Neural Network (CNN). Different from supervised learning approaches which need labeled training images, the road anomaly detection network is trained by road surface images that are automatically labeled based on the reconstructed 3D surface information. In order to collect clear road surface images on the public road, a road surface monitoring system is designed and integrated for the road surface image capturing and visualization. The proposed approach is evaluated in both simulated environment and through real-world experiments. The parametric study of the proposed approach shows the small error of the 3D road surface reconstruction influenced by different variables such as the image noise, camera orientation, and the vertical movement of the camera in a controlled simulation environment. The comparison with traditional SfM technique and the numerical results of the proposed reconstruction using real-world road surface images then indicate that the proposed approach effectively reconstructs high quality near-planar road surface while automatically detects road defects with high precision, accuracy, and recall rates without the degenerate issue.

Degenerate Near-planar Road Surface 3D Reconstruction and Automatic Defects Detection

Yazhe Hu

(GENERAL AUDIENCE ABSTRACT)

Road is one of the key infrastructures for ground transportation. A good road surface condition can benefit mainly on three aspects: 1. Avoiding the potential traffic accident caused by road surface defects, such as potholes. 2. Reducing the damage to the vehicle initiated by the bad road surface condition. 3. Improving the driving and riding comfort on a healthy road surface. With all the benefits mentioned above, it is important to examine and check the road surface quality frequently and efficiently to make sure that the road surface is in a healthy condition.

In order to detect any road surface defects on public road in time, this dissertation proposes three techniques to tackle the road surface defects detection problem: First, a near-planar road surface three-dimensional (3D) reconstruction technique is proposed. Unlike traditional 3D reconstruction technique, the proposed technique solves the degenerate issue for road surface 3D reconstruction from two images. The degenerate issue appears when the object reconstructed has near-planar surfaces. Second, after getting the accuracy-enhanced 3D road surface reconstruction, this dissertation proposes an automatic defects detection technique using both the 3D reconstructed road surface and the road surface image information. Although physics-based detection using 3D reconstruction and 2D images are reliable and explainable, it needs more time to process these data. To speed up the road surface defects detection task, the third contribution is a technique that proposes a self-supervised learning structure with data-driven Convolutional Neural Networks (CNN). Different from traditional neural network-based detection techniques, the proposed combines the 3D road information with the CNN output to jointly determine the road surface defects region. All the proposed techniques are evaluated using both the simulation and real-world experiments. Results show the efficacy and efficiency of the proposed techniques in this dissertation.

Dedication

This dissertation is dedicated to my parents Zhonghua Hu and Xueling Wang. It is your love and support that encourage me all the time during this incredible journey.

Acknowledgments

First, I am truly grateful to Prof. Tomonari Furukawa for his continuous supports and advices during my PhD study. He introduced me to this excited field of road surface monitoring and taught me how to be a researcher to influence people and the society. His vision and guidance always encourage me to improve myself throughout my study.

I am so thankful to my committee members Dr. Jia-Bin Huang, Dr. Corina Sandu, and Dr. Steve Southward for their valuable feedback and comments. It is your insightful advices that make me think about my research in different perspectives.

I must also thank my colleagues in CMS Lab. Kuya, Boren, Affan, Luan, Rich, Chris, Tian, Orson, Hangxin, Murat, Mengyu, Yoon, Mickey, George, Cong, Josiah, Diya, Jihong, Urvi, Tamer, Jeeseop, Ash, Ioannis, Jonathan, Tahsin, Arthur, Junyang, and Dean. I am glad to have you studying together for all these years I spent in CMS Lab.

I would like to thank my family and all my friends for their love and support. My life becomes more colorful because of you. I want to also say thank you to my roommates Dong Nan, Sumin Shen, and Mengyu Song for their accompany. I will always remember the time we spent together in this beautiful college town.

I would like to express my deepest gratitude to my fiancée Dr. Maggie for her encouragement, understanding, and love all the time.

Lastly, I appreciate all the people who helped me during this incredible journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

During recent years, more and more efforts are focusing on intelligent monitoring for transportation infrastructures. Among various objects, road surface is one of the most important and frequent articles which vehicles interact with. A correct perception of the road surface is critical for vehicles to understand the road environment. Besides the 2D knowledge of road surface from data such as images, the 3D data of road surface can provide richer information for the cars. On the one hand, vehicles can use the 2D and 3D road data to avoid road defects or obstacles, in the meanwhile increases the ride comfort and decreases the possibility of damaging themselves. On the other hand, the road maintenance process can be facilitated based on the 3D road surface data so that hazardous road regions are being fixed on time. Thus the understanding of road surface through both 2D and 3D data becomes key to the safety and healthiness for the traffic.

1.2 Objectives

The primary objectives of this dissertation are the following:

- To propose a 3D reconstruction technique for near-planar (degenerate) road surface.

- To propose a technique which automatically detects and estimates road defects.
- To develop a system capable of capturing road surface condition information.

1.3 Principal Contributions

The technical contributions of this dissertation are summarized as follows:

- A near-planar road surface 3D reconstruction technique from one camera which solves the degenerate issue.
- Road surface defects detection and estimation technique based on 3D information.
- A self-supervised learning technique for road surface anomaly detection by combining physics-driven 3D reconstruction with data-driven CNN.
- A designed road surface monitoring system composed of camera, GPS, Field-programmable gate array (FPGA), and On-board diagnostics-II (OBD-II).

1.4 Publication

- **Hu, Yazhe** and Tomonari Furukawa. ”A high-resolution surface image capture and mapping system for public roads.” SAE International Journal of Passenger Cars-Electronic and Electrical Systems 10.2017-01-0082 (2017): 301-309.
- **Hu, Yazhe**, and Tomonari Furukawa. ”Automatic Detection and Evaluation of 3D Pavement Defects Using 2D and 3D Information at the High Speed.” International Journal of Automotive Engineering 9.4 (2018): 323-331.

- Li, Diya, **Yazhe Hu**, and Tomonari Furukawa. ”AEKF-Based 3-D Localization of Road Surface Images with Sparse Low-Accuracy GPS Data.” 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall). IEEE, 2018.

- **Hu, Yazhe**, and Tomonari Furukawa. ”Road Obstacle Detection Based on Near-planar Road 3D Reconstruction from a Front Camera.” 5th International Symposium on Future Active Safety Technology toward Zero Accidents, 2019.

- **Hu, Yazhe**, and Tomonari Furukawa. ”A Self-Supervised Learning Technique for Road Defects Detection Based on Monocular Three-Dimensional Reconstruction.” International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Vol. 59216. American Society of Mechanical Engineers, 2019.

- **Hu, Yazhe**, and Tomonari Furukawa. ”Degenerate Near-Planar 3D Reconstruction from Two Overlapped Images for Road Defects Detection.” Sensors 20.6 (2020): 1640.

1.5 Organization

This dissertation is organized as follows:

- Chapter 2 reviews past works on general object 3D reconstruction, road surface 3D reconstruction, road surface defects detection, and data-driven road defects detection techniques.

- Chapter 3 describes the proposed technique of road surface 3D reconstruction from two overlapped images. The chapter first formulates the two-image road surface 3D reconstruction problem and introduces the fundamental knowledges. Then the proposed degenerate near-planar road 3D reconstruction technique is mathematically demonstrated. Comprehensive evaluation is conducted in both simulation and real-world experiments to show the efficacy of the proposed technique.

- Chapter 4 illustrates the automatic road defects detection using near-planar road surface 3D reconstruction. The proposed technique includes preprocessing, anomaly road surface points detection, automatic road defects detection from 2D and 3D information, and defects clustering and estimation. The results of the defects detection are shown through experiments.
- Chapter 5 presents a self-supervised learning technique for the road surface defects detection based on 3D reconstruction. The self-supervised learning procedures are described for the near-planar road surface. Experimental results show the effectiveness of the proposed technique at last.
- Chapter 6 shows the design of a road surface monitoring system. Following by the design, this chapter also shows the simultaneous 2D and 3D road surface mapping technique and the corresponding visualization of the road maps as results.
- Chapter 7 summarizes the conclusions and future works of this dissertation.

Chapter 2

Literature Review

This chapter reviews past works related to the proposed original contributions in this dissertation. The general 3D object reconstruction techniques are reviewed first, followed by reviewing the road surface 3D reconstruction techniques. Past works on road surface defects detection are given next. Related works on data-driven road surface defects detection are reviewed at last.

2.1 Object Reconstruction in 3D

The goal of the 3D object reconstruction is to obtain objects' three-dimensional appearance by using sensor data. The types of 3D reconstruction sensors are classified as contact and non-contact. The contact type sensors usually use the stylus based design to measure the 3D surface according to the stylus tip's position change [1] [2]. The stylus based technique provides accurate measurement at each point. However, because the stylus based technique measure the object point-by-point, the measurement speed is low. Also, if area 3D reconstruction is required, the point-by-point measurement is inefficient.

The second type of measurement is conducted by non-contact techniques. The sensors used in non-contact techniques are categorized as active or passive. Active sensors has their own illumination or light source, where passive sensors just utilize natural light from the environment. Laser scanning 3D reconstruction technique is one of the active technique [3]

[4] [5] [6] [7]. The laser scanning techniques are based on Time-of-Flight (ToF) or phase shift to measure the distance between the point it measures and the sensor. Different from laser techniques' point-scanning or line-scanning, RGB-D techniques have area-based 3D reconstruction and obtain the RGB image at the same time [8] [9] [10] [11] [12] [13] [14] [15] [16] [17]. Besides ToF and stereo-based RGB-D sensors, structured light RGB-D techniques project known light pattern on the object to reconstruct the 3D structure. There are also passive techniques to reconstruct 3D from one camera based on SfM [18] [19] [18] [20] [21] [22]. The SfM pipeline contains feature detection from detectors such as SIFT [23], SURF [24] [25], ASIFT [26], Harris corner [27], and FAST [28]. SfM then matches the detected features [29] [30] [31] and reconstruct the 3D points from matched feature points [32]. Video-based are another passive 3D reconstruction techniques which uses similar processes from stereo vision and SfM to reconstruct 3D from key video frames [33] [34] [35] [36] [37] [38] [39].

2.2 Road Surface 3D Reconstruction

Road surface is one of the key infrastructures for ground transportation. In recent years, research efforts and industrial needs keep increasing in intelligent transportation systems and vehicles . The aims of knowing the road surface condition are twofold: increasing safety and ride comfort for vehicles; providing inspection information for timely road maintenance. The ground surface estimation problem is still an active research topic among researchers [40] [41] [42] [43] [44]. Since the performance of the vehicle is influenced by the geometry of the road surface, a knowledge on the 3D profile for the road surface is substantial for the physical vehicle-environment interaction.

Related works for road surface reconstruction can be classified into two categories. The first type of approach uses multiple sensors or a non-monocular camera to reconstruct the road surface. Jia et al. [44] used two cameras to reconstruct the road surface from a reference plane recovered by the stereo vision. Panahandel et al. [45] deployed ground plane detection technique by exploiting the rotation and translation information from Inertial Measurement Unit (IMU)-camera sensor fusion. Holz et al. [46] used the RGB-D camera to get point cloud data for surface normals reconstruction as well as the ground plane segmentation from the 3D points. While being effective, multiple-sensor methods often involve a system-level calibration. In this case, a small change of the sensor pose caused by vibration, acceleration, or collision will result in incorrect road surface reconstruction.

The second category falls into the the deployment of using one camera. Since the road surface is a near-planar object, [42, 47, 48, 49, 50] detects a plane from several views and segments them in 2D images without reconstructing a 3D ground surface. The main purposes of these papers are to segment ground surface or detect obstacles on the road surface. Kaneko et al. [51] developed a monocular depth measurement using the geometrical relationship of the Flat Surface Model and triangulation. Zhou et al. [52] proposed a robust plane-based SfM which detects and tracks the planar object first and then reconstruct planes from multiple views. Collins et al. [53] proposed a technique using Affine camera model for planar SfM. The paper showed a closed-form solution from multiple views and avoided the degenerate problem. Therefore, most existing one camera-based method either only segment planar objects in 2D images or reconstruct the 3D near-planar objects from more than two views, and the rest two-view methods do not consider the plane degenerate issue. The disadvantage of reconstruction from more than two views lies on not only the difficulty to keep tracking the same plane, but also when there are no more than two images capturing the same area.

2.3 Road Surface Defects Detection

As the road is a fundamental infrastructure in the transportation system, a healthy and intact road surface condition increases the ride comfort and vehicle safety for through traffic [54] [55]. The road surface condition inevitably downgrades and is affected by stresses from the traffic as well as the climate impacts such as humidity or temperature change. Thus frequent inspections of the road surface are vital in identifying road surface defects along with carrying out timely maintenance. Labor intensiveness, inefficiency, and subjectivity of manual inspection has resultantly necessitated automatic measurement of the road surface defects such as potholes and ruts, which are most characterized by geometry [56, 57, 58, 59, 60, 61].

Past works on automatic geometrical road defects detection can be classified by three types: the detection by acceleration, the detection by color, and the detection by geometry. The detection by acceleration technique uses accelerometers as irregular geometrical changes create vibration that can be measured by accelerometers. Yu et al. [62] analyzed acceleration and automatically detected road defects for the first time to the best of the author's knowledge. Tai et al. [63] and Eriksson et al. [64] proposed methods using a machine learning approach to detect road anomaly where Support Vector Machine (SVM) and unsupervised learning were used respectively to enhance detection accuracy. Xue et al. [65] adopted a self-learning one degree-of-freedom vibration signals to predict the pothole. Mednis et al. [66] implemented and compared several acceleration data processing algorithms for pothole detection, which resulted in a detection rate between 68% to 90%. Although detection by acceleration techniques directly and thus accurately sense geometrical road defects, they miss detection if no tire steps exactly on the road defects.

For detection by color methods, image sensors are often equipped to obtain the appearance of defects. Tedeschi et al. [67] proposed a method using Local Binary Pattern (LBP) feature-based cascade classifiers to detect road defects from images. Koch et al. [68, 69] used the histogram and four different image filters to extract road distress texture features. Banharnsakun et al. [70] deployed an Artificial Neural Network (ANN) which can categorized the distress into longitudinal crack, transversal crack, and pothole. Ryu et al. [71] separated the pothole region from the background by Histogram Shape-Based Thresholding (HST) and then used multiple filters to find the pothole features. The detection by color method provides intuitive information about road defects' position and size. However, the RGB image analysis may not capture geometry and contains unnecessary information such as shadows, oil stains and pavement markings which affect the detection.

Among detection by geometry methods, Chang et al. [72] and Yu et al. [5, 73] detected potholes by analyzing topological features obtained from 3D laser scanning data. Hou et al. [74] and El et al. [75] applied stereo-vision systems to get the 3D point cloud from road surface images and detect potholes directly from the 3D model of the road obtained from point cloud datawhile no precision rate was investigated. Ahmed et al. [76] proposed a pothole detection method by using multiple-view images to reconstruct 3D points of road surface from manually marked artificial features on the road surface. They reported around 0.1 mm accuracy in Z direction when measured statically. Antol et al. [77] and Moazzam et al. [78] implemented the road distress detection by 3D point cloud data from the RGB-D camera. The former used a movable RGB-D camera box to enable depth measurement at a low speed, while the latter used a tripod to statically measure the 3D road surface by the RGB-D camera. However, the accuracy of the 3D reconstruction by using laser sensor or stereo-vision system is influenced by vibration when the measuring sensors are moving with the vehicle. In general, the issue of the 3D reconstruction based technique is its accuracy in 3D reconstruction since the road surface is near-planar and thus provides poor vertical information.

2.4 Road Surface Defects Detection based on Data-driven Approach

The data-driven techniques use a large amount of data as the input to learn and predict defects through machine learning algorithms. The road surface defects detection using on data-driven techniques is fundamentally an image classification problem. Existing data-driven deep learning CNN such as LeNet-5 [79], AlexNet [80], VGG-16/19 [81], GoogLeNet [82], Inception v4 [83], ResNet [84], ResNeXt [85], DenseNet [86], and SENet [87]. In addition, deep learning techniques have already been used in various defects detection scenarios [88] [89] [90] [91] [92] [93] [94] [95] [96] [97].

Among data-driven road surface defects detection techniques, Li et al. [98] took 3D Laser-scanned images as the input, and passed them to a CNN model for road surface cracks detection. Park et al. [99] compared the performance of several CNN models and found the optimal one for road defects detection. Yarram et al. [100] combined both deep learning and semantic segmentation, as well as proposed a hierarchy of labeling road defects. The existing data-driven techniques still need tremendous amount of human supervision and labeling. Also, these techniques usually only take the images as the input (which is 2D) while missing the important 3D charateristics of road defects.

2.5 Summary

This chapter has presented past works on the 3D reconstruction of the road surface and defects detection techniques. First, the general 3D reconstruction techniques and the road surface techniques are reviewed, which motivates and leads to the proposed road surface 3D reconstruction technique. Second, this chapter reviews past road surface defects detection techniques which guides to the proposed road defects detection technique. Current data-driven road surface defects detection techniques are given at the end. The proposed road surface monitoring using one regular camera is developed based on these literatures to achieve the objectives of this dissertation.

Chapter 3

Degenerate Three-dimensional Reconstruction of Near-planar Surfaces from Two Images

This chapter presents a 3D reconstruction technique for the near-planar road surface using one camera. The proposed technique solves the degeneracy problem in near-planar object 3D reconstruction by obtaining a non-degenerate fundamental matrix. Instead of using multiple views which many previous works relied on, only two images are required to reconstruct the 3D profile for the road surface. By scaling the reconstructed 3D points, absolute metric scales are obtained for each point. Besides a comprehensive simulation experiment, the experiment conducted on the real-world road surface show the robustness and effectiveness of the proposed technique.

The chapter is organized as follows. The next section introduces the fundamental knowledge about the road surface 3D reconstruction from two images and the degenerate issue for road surface reconstruction. Section 3.2 then shows the proposed degenerate near-planar road surface 3D reconstruction from two images. Section 3.3 demonstrates the experimental results for the proposed technique in both simulation and real-world environment. Finally, the last section summarizes this chapter.

3.1 Road Surface 3D Reconstruction from Two Overlapped Images

3.1.1 Road Surface 3D Reconstruction Problem Formulation

Figure 3.1 shows the problem settings of 3D road surface reconstruction using one camera. The road is shown to be as a near-planar surface. In order to reconstruct the 3D road surface, a camera facing downward is mounted on a vehicle and moves along the road surface at height h where $\mathbf{X}_k^c$ is the position of the camera at time step k. Let the sequence of road surface images collected from time step 0 to time step K be $I_{0:K} \equiv \{I_k | \forall k \in \{0, ..., K\}\}$ where every two adjacent images $\{I_{k-1}, I_k\}$ capture an overlapping region on the road surface. The end objective of the 3D reconstruction is to derive a road surface point cloud, $\mathbf{X}_{0:K}^r$, which is well distributed over the road surface.

This objective can be tackled by reconstructing the 3D surface of each overlapping region from the two adjacent images independently. At time step k, the 2D feature points in the images $\{I_{k-1}, I_k\}$, $\mathbf{x}_{k-1}^r$ and $\mathbf{x}_k^r$, define the overlapping region shown by a red rectangle. The association of the corresponding feature points in the overlapping region reconstructs the 3D road surface of the overlapping region in terms of a point cloud, which is denoted by $\mathbf{X}_k^r$. It is to be noted here that the ith point of $\mathbf{X}_k^r$, $\mathbf{X}_{k,i}^r$, has a projection $\mathbf{x}_{k,i}^r$ in I_k. The entire road surface point cloud, $\mathbf{X}_{0:K}^r = \{\mathbf{X}_k^r | k = 0, 1, ..., K\}$, are obtained by combining the road surface reconstruction from every two adjacent image feature points $\{\{\mathbf{x}_{k-1}^r, \mathbf{x}_k^r\} | k = 1, 2, ..., K\}$.

Figure 3.2 illustrates the significance of the two-image problem formulation where the vehicle speed is shown with respect to different numbers of overlapped images when the camera frame rate is 60, 30 and 15 FPS. Note that these are the common frame rates in industrial cameras, and each image covers a $1m \times 1m$ road surface area. For every number

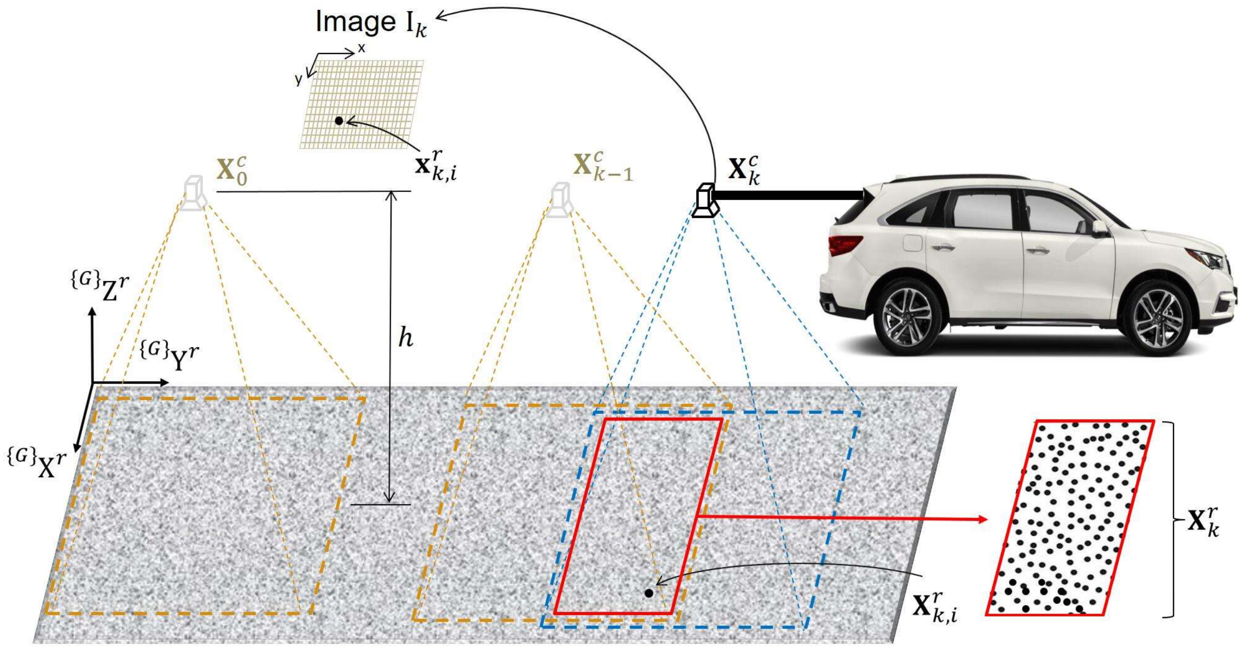


Figure 3.1: Road surface 3D reconstruction from one camera.

of images overlapped, N_o, the overlapping area between every two neighboring images is at least $(100 - \frac{100}{N_o})\%$. As the curves exhibit, every camera sees common vehicle speeds when the number of overlapped images is only two. Therefore, 3D road surface reconstruction will fail if it is not possible from two images.

3.1.2 Two-image 3D Road Surface Reconstruction

Figure 3.3 shows the notations and the operation of the general road surface 3D reconstruction from image features $\mathbf{x}_{k-1}^r$ and $\mathbf{x}_k^r$. To present the mathematical derivation of the two-image 3D reconstruction for road surfaces, a line is plotted passing the camera centers, $\mathbf{X}_{k-1}^c$ and $\mathbf{X}_k^c$. This line intersects with image I_{k-1} at point $\mathbf{e}_{k-1}$ as well as image I_k at point $\mathbf{e}_k$. $\mathbf{l}_{k-1,i}$ is a line passing through $\mathbf{e}_{k-1}$ and $\mathbf{x}_{k-1,i}^r$, a projection from road surface point $\mathbf{X}_{k-1,i}^r$ to I_{k-1}. Similarly, $\mathbf{l}_{k,i}$ is a line passing through $\mathbf{e}_k$ and $\mathbf{x}_{k,i}^r$, and this is given by:

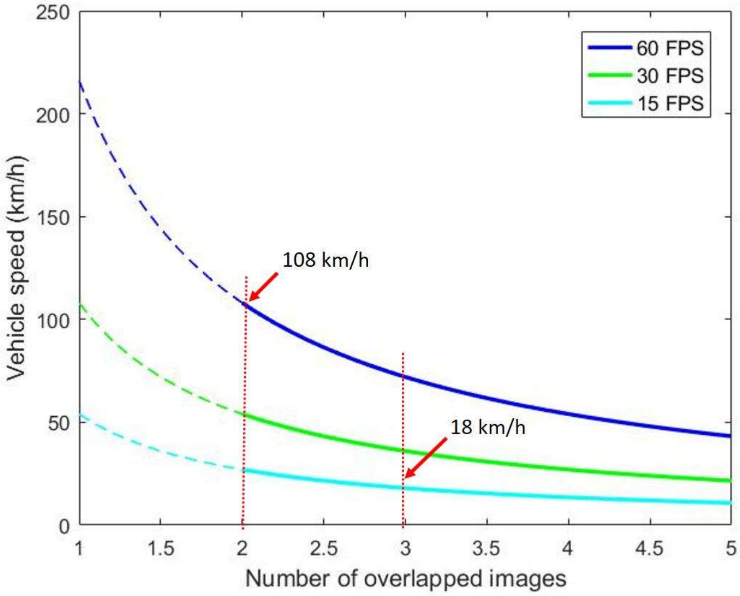


Figure 3.2: The number of images overlapped on each road surface at various vehicle speed.

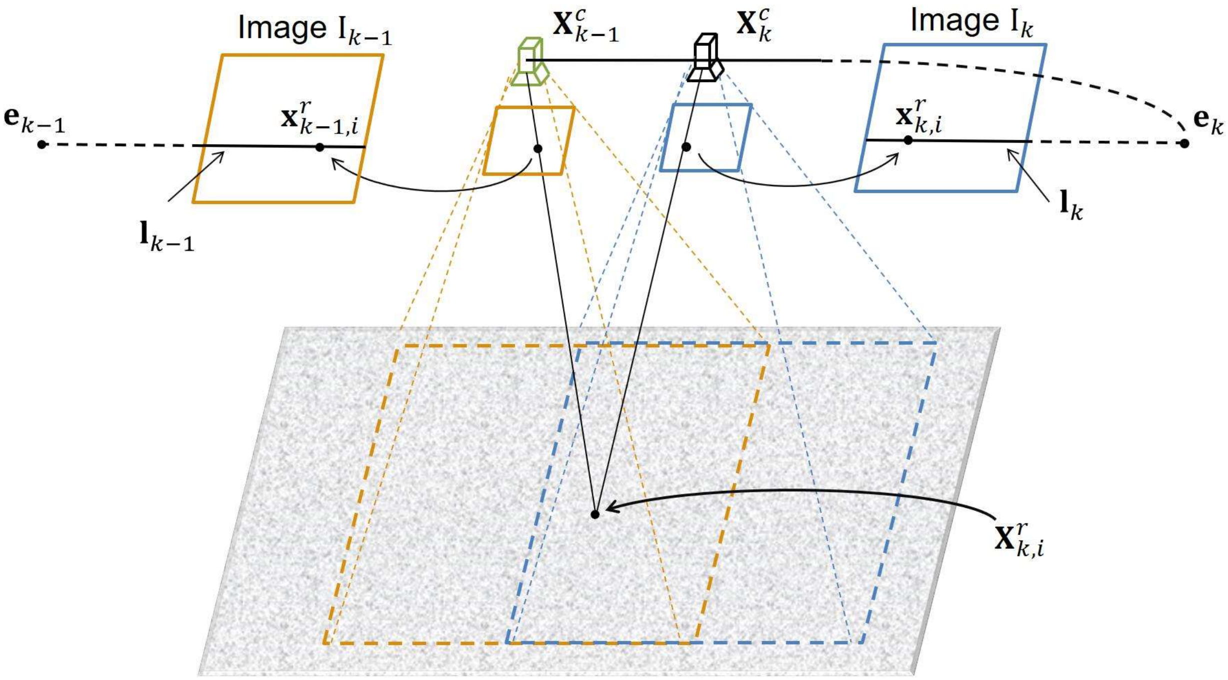


Figure 3.3: 3D road surface reconstruction from two views

$$\mathbf{l}_{k,i} = \mathbf{e}_k \times \mathbf{x}^r_{k,i} = [\mathbf{e}_k]_\times \mathbf{x}^r_{k,i} \tag{3.1}$$

Combining Eq. 3.1 with ${\mathbf{x}^r_{k,i}}^T \mathbf{l}_{k,i} = 0$ yields:

$${\mathbf{x}^r_{k,i}}^T [\mathbf{e}_k]_\times \mathbf{x}^r_{k,i} = 0 \tag{3.2}$$

If $\mathbf{X}^r_{k,i}$ is located on a road surface plane, then $\mathbf{x}^r_{k-1,i}$ and $\mathbf{x}^r_{k,i}$ are related by a homography matrix $\mathbf{H}_{ab}$:

$$\mathbf{x}^r_{k,i} \propto \mathbf{H}_{ab} \mathbf{x}^r_{k-1,i} \tag{3.3}$$

Substituting Eq. 3.2 to Eq. 3.3 results in:

$${\mathbf{x}^r_{k,i}}^T [\mathbf{e}_k]_\times \mathbf{H}_{ab} \mathbf{x}^r_{k-1,i} = {\mathbf{x}^r_{k,i}}^T \mathbf{F}_k \mathbf{x}^r_{k-1,i} = 0 \tag{3.4}$$

where $\mathbf{F}_k = [\mathbf{e}_k]_\times \mathbf{H}_{ab}$ is the fundamental matrix of the two images. Eq. 3.4 holds for all the n correspondences $\left\{ \left\{ \mathbf{x}^r_{k,i}, \mathbf{x}^r_{k-1,i} \right\} | i = 1, 2, ..., n \right\}$ [32], which means:

$$\begin{gathered} {\mathbf{x}^r_k}^T \mathbf{F}_k \mathbf{x}^r_{k-1} = \\ \begin{bmatrix} x^r_{k,1} & ... \; x^r_{k,n} \\ y^r_{k,1} & ... \; y^r_{k,n} \\ 1 & ... \; 1 \end{bmatrix}^T \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x^r_{k-1,1} & ... \; x^r_{k-1,n} \\ y^r_{k-1,1} & ... \; y^r_{k-1,n} \\ 1 & ... \; 1 \end{bmatrix} \\ = 0 \end{gathered} \tag{3.5}$$

To solve $\mathbf{F}_k$, Eq. 3.5 is rearranged to a form of $\mathbf{A}\mathbf{f}_k = 0$, where:

$$A_i = \begin{bmatrix} x^r_{k,i}x^r_{k-1,i} \\ x^r_{k,i}y^r_{k-1,i} \\ x^r_{k,i} \\ y^r_{k,i}x^r_{k-1,i} \\ y^r_{k,i}y^r_{k-1,i} \\ y^r_{k,i} \\ x^r_{k-1,i} \\ y^r_{k-1,i} \\ 1 \end{bmatrix}, \ \mathbf{A} = \begin{bmatrix} A_1^T \\ A_2^T \\ . \\ . \\ . \\ A_n^T \end{bmatrix} \tag{3.6}$$

$$\mathbf{f}_k = (f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33})^T$$

In SfM [32], $\mathbf{F}_k$ is obtained by solving the minimization problem:

$$\min_{\mathbf{f}_k} \mathbf{A}\mathbf{f}_k \tag{3.7}$$

subject to

$$\|\mathbf{f}_k\| = 1 \tag{3.8}$$

After the fundamental matrix $\mathbf{F}_k$ is calculated, by following the subsequent SfM process, the essential matrix $\mathbf{E}_k$ is calculated as:

$$\mathbf{E}_k = \mathbf{K}^T\mathbf{F}_k\mathbf{K} \tag{3.9}$$

where $\mathbf{K}$ is the intrinsic matrix of the calibrated camera. The Singular Value Decomposition (SVD) of $\mathbf{E}_k$ then contributes to the calculation of rotation matrix $\mathbf{R}_k$ and the up-to-scale translation vector $\mathbf{t}_k$ between time step $k-1$ and k:

$$\begin{aligned}
\mathbf{E}_k &= \mathbf{U}\mathbf{D}\mathbf{V}^T \\
W &= \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
\mathbf{R}_k = \mathbf{U}\mathbf{W}\mathbf{V}^T &\ \mathit{or}\ \mathbf{R}_k = \mathbf{U}\mathbf{W}^T\mathbf{V}^T \\
\mathbf{t}_k = \mathbf{U}(0,0,1)^T &\ \mathit{or}\ \mathbf{t}_k = -\mathbf{U}(0,0,1)^T
\end{aligned} \tag{3.10}$$

where there is one correct combination of $\mathbf{R}_k$ and $\mathbf{t}_k$ which can make all $\mathbf{X}_k^r$ be in front of the camera. The projection matrix $\mathbf{P}_k$ is identified by the rotation matrix $\mathbf{R}_k$ and the translation vector $\mathbf{t}_k$. The 3D reconstructed points $\mathbf{X}_k^r$ are finally obtained by the triangulation $f_t(\cdot)$:

$$\begin{aligned}
\mathbf{x}_{k-1}^r &= \mathbf{P}_{k-1}\mathbf{X}_k^r = \mathbf{K}[\mathbf{I},\mathbf{0}]\mathbf{X}_k^r \\
\mathbf{x}_k^r &= \mathbf{P}_k\mathbf{X}_k^r = \mathbf{K}[\mathbf{R}_k,\mathbf{t}_k]\mathbf{X}_k^r
\end{aligned} \tag{3.11}$$

$$\mathbf{X}_k^r = f_t(\mathbf{K},\mathbf{R}_k,\mathbf{t}_k,\mathbf{x}_k^r,\mathbf{x}_{k-1}^r) \tag{3.12}$$

3.1.3 Planar Surface Degeneracy Problem

Since the road surfaces are near-planar, it suffers from the degenerate issue which will be shown by the rest of this section. As $\mathbf{X}_k^r$ are located on the near-planar road surface, $\mathbf{x}_{k-1}^r$ and $\mathbf{x}_k^r$ can be related by a 3×3 homography matrix $\mathbf{H}_k$:

$$\mathbf{x}_k^r \propto \mathbf{H}_k\mathbf{x}_{k-1}^r \tag{3.13}$$

in which $\mathbf{x}_k^r$ is proportional to $\mathbf{H}_k\mathbf{x}_{k-1}^r$. This means that the cross product of $\mathbf{x}_k^r$ and $\mathbf{H}_k\mathbf{x}_{k-1}^r$ is $\mathbf{x}_k^r \times \mathbf{H}_k\mathbf{x}_{k-1}^r = \mathbf{0}$. Thus solving $\mathbf{H}_k$ equals to solving the equation $\mathbf{A}'\mathbf{h}_k = \mathbf{0}$ where $\mathbf{H}_k$, $\mathbf{A}'$ and $\mathbf{h}_k$ are expressed as:

$$\mathbf{H}_k = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

$$A_i' = \begin{bmatrix} -x_{k-1,i}^r & 0 \\ -y_{k-1,i}^r & 0 \\ -1 & 0 \\ 0 & -x_{k-1,i}^r \\ 0 & -y_{k-1,i}^r \\ 0 & -1 \\ x_{k,i}^r x_{k-1,i}^r & y_{k,i}^r x_{k-1,i}^r \\ x_{k,i}^r y_{k-1,i}^r & y_{k,i}^r y_{k-1,i}^r \\ x_{k,i}^r & y_{k,i}^r \end{bmatrix}, \mathbf{A}' = \begin{bmatrix} A_1^T \\ A_2^T \\ \cdot \\ \cdot \\ \cdot \\ A_n^T \end{bmatrix} \tag{3.14}$$

$$\mathbf{h}_k = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33})^T$$

To solve $\mathbf{h}_k$, the problem is equivalent to minimizing $\|\mathbf{A}'\mathbf{h}_k\|$ subject to $\|\mathbf{h}_k\| = 1$ because of image noises. Therefore, solving $\mathbf{h}_k$ is similar to solving $\mathbf{f}_k$ in the previous section.

Degeneracy is defined as the situation when fundamental matrix $\mathbf{F}_k$ obtained from the previous procedure is not unique. The planar object, which the road can be approximated as, is one of the degenerate geometries. If $\mathbf{X}_k^r$ are located on a plane surface, the correspondences in the two views $\mathbf{x}_{k-1}^r$ and $\mathbf{x}_k^r$ satisfy Eq. 3.13. Also, $\mathbf{x}_{k-1}$ and $\mathbf{x}_k$ satisfy Eq. 3.5. The substitution of Eq. 3.13 into Eq. 3.5 yields

$$\mathbf{x}_k^{rT}\mathbf{S}_k\mathbf{x}_k^r = 0 \tag{3.15}$$

where $\mathbf{S}_k = \mathbf{F}_k\mathbf{H}_k^{-1}$. To satisfy Eq. 3.15, $\mathbf{S}_k$ must be a skew-symmetric matrix given by

$$\mathbf{S}_k = \begin{bmatrix} 0 & -s_3 & s_2 \\ s_3 & 0 & -s_1 \\ -s_2 & s_1 & 0 \end{bmatrix} \tag{3.16}$$

As a result, the fundamental matrix $\mathbf{F}_k$ is:

$$\mathbf{F}_k = \mathbf{S}_k\mathbf{H}_k = \begin{bmatrix} 0 & -s_3 & s_2 \\ s_3 & 0 & -s_1 \\ -s_2 & s_1 & 0 \end{bmatrix}\mathbf{H}_k \tag{3.17}$$

Thus $\mathbf{F}_k$ has a solution with three degree-of-freedom (determined by s_1, s_2, and s_3). Since $\mathbf{F}_k$ is up-to-scale, the solution of $\mathbf{F}_k$ becomes to have two degree-of-freedom. Therefore the existing 3D reconstruction technique from I_{k-1} and I_k cannot lead to correct 3D reconstructed points for planar road surface because of the ambiguity of $\mathbf{F}_k$ introduced to reconstruction process from Eq. 3.9 to 3.12. While 3D reconstruction techniques exist, the issue of their direct application to road surface profiling is the ill-posedness of the problem due to the lack of depth information and the incorrect feature matching due to the noisy image. The next section will present the proposed technique, which solves the ambiguity issue of $\mathbf{F}_k$ for the road surface reconstruction, and leads to correct defects detection based on the 3D information.

3.2 Degenerate Near-planar 3D Reconstruction for Road Surface

3.2.1 Overview

Figure 3.4 shows a schematic diagram of the proposed degenerate road surface 3D reconstruction. Unlike the conventional SfM, the proposed technique derives a unique fundamental matrix $\mathbf{F}_k$ with no degenerate issue by not deriving it directly from the 2D feature points. After the extraction of 2D feature points in the images I_{k-1} and I_k by a technique such as SIFT [23] and the matching of the feature points, the proposed technique first identifies if the road surface is near-planar by deriving the homography $\mathbf{H}_k$ of the road surface with RANSAC and checking the outlier ratio against the threshold. If so, the initial rotation $\mathbf{R}_{k,0}$ and translation $\mathbf{t}_{k,0}$ of the camera from the constraint of vehicle's motion are computed for 3D reconstruction by the proposed technique. The acquisition of the initial rotation and translation makes the fundamental matrix $\mathbf{F}_k$ unique. The matrix $\mathbf{F}_k$ is then derived by combining the homography and the translation.

The remaining process can adopt a conventional SfM shown within the red dashed box. After getting the essential matrix $\mathbf{E}_k$ using $\mathbf{F}_k$ followed by obtaining camera's rotation and translation from SVD of $\mathbf{E}_k$, the non-degenerate fundamental matrix $\mathbf{F}_k$ is updated by iterating the camera's rotation $\mathbf{R}_{k,\xi}$ and translation $\mathbf{t}_{k,\xi}$. It is also a contribution and difference from the conventional 3D reconstruction technique that the proposed non-degenerate 3D road surface reconstruction outputs metric 3D points which have real-world units. It is noted that the proposed technique requires only two views with an overlapping area, instead of multiple views which are difficult to capture, while the vehicle moves fast.

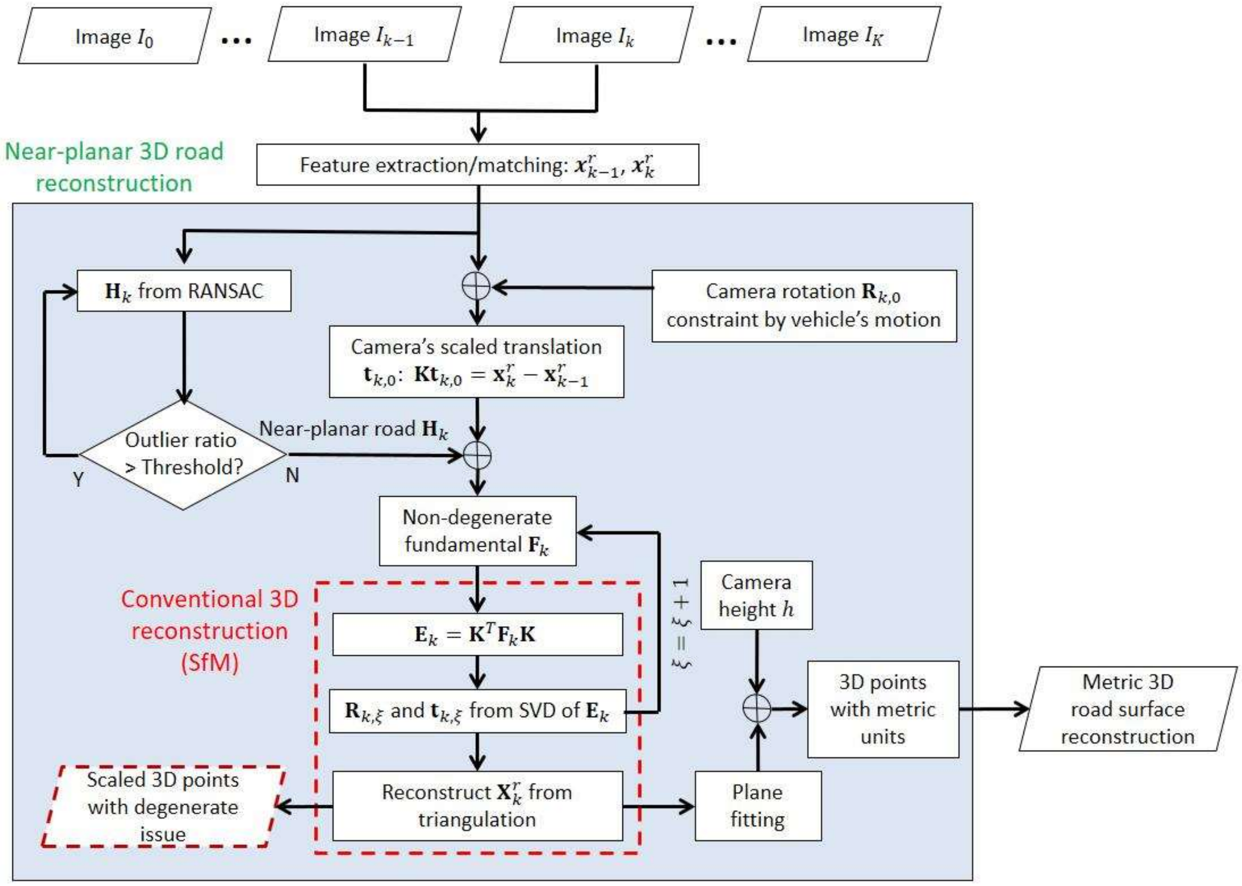


Figure 3.4: Proposed degenerate near-planar road surface 3D reconstruction technique.

3.2.2 3D Reconstruction for Near-planar Road Surface

The proposed technique solves the ambiguity issue of $\mathbf{F}_k$ by mathematically deriving a unique fundamental matrix for the near-planar road surface. In the local coordinate $\{L\}$, ${}^{\{L\}}\mathbf{X}^c_{k-1} = (0,0,0)^T$ and its projection to image I_k, $\mathbf{e}_k$, is expressed as:

$$\mathbf{e}_k = \mathbf{K}[\mathbf{R}_k, \mathbf{t}_k] \cdot \begin{bmatrix} {}^{\{L\}}\mathbf{X}^c_{k-1} \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{R}_k, \mathbf{t}_k] \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{t}_k \tag{3.18}$$

It is noted that from Eq. 3.1, all the lines $\mathbf{l}_k$ have the following for road surface images:

$$\mathbf{l}_k = \mathbf{e}_k \times \mathbf{x}^r_k \tag{3.19}$$

Meanwhile, Eq. 3.5 and ${\mathbf{x}^r_k}^T \mathbf{l}_k = 0$ relates $\mathbf{F}_k$ and $\mathbf{l}_k$ as:

$$\mathbf{F}_k \mathbf{x}^r_{k-1} = \mathbf{l}_k \tag{3.20}$$

Substitute Eq. 3.19 and 3.13 into Eq. 3.20 resulting in:

$$\mathbf{F}_k \mathbf{x}^r_{k-1} = \mathbf{e}_k \times \mathbf{x}^r_k = [\mathbf{e}_k]_\times \mathbf{H}_k \mathbf{x}^r_{k-1} \tag{3.21}$$

Combining Eq. 3.21 with Eq. 3.18, it derives $\mathbf{F}_k$ for the near-planar road surface as:

$$\mathbf{F}_k = [\mathbf{e}_k]_\times \mathbf{H}_k = [\mathbf{K}\mathbf{t}_k]_\times \mathbf{H}_k \tag{3.22}$$

where $\mathbf{H}_k$ is calculated recursively by RANSAC using $\mathbf{x}^r_{k-1}$ and $\mathbf{x}^r_k$ after mismatched points rejection.

Comparing Eq. 3.17 with Eq. 3.22, instead of representing $\mathbf{F}_k$ with any 3-vector $\mathbf{s}$, $\mathbf{F}_k$ is determined in Eq. 3.22 by $\mathbf{t}_k$ which is the up-to-scale translation between the camera positions in two views:

$$\mathbf{t}_k = \mathbf{X}^c_k - \mathbf{X}^c_{k-1} \tag{3.23}$$

Since the vehicle moving along the road has small rotation $\mathbf{R}_k$ for the camera in such a short period from time step $k-1$ to k, $\mathbf{R}_k$ is expressed as $\mathbf{R}_k \approx \mathbf{I}$. Eq. 3.25 can be obtained from Eq. 3.24:

$$\begin{aligned}
\mathbf{x}^r_{k-1} &= \mathbf{P}_{k-1}\mathbf{X}^r_k = \mathbf{K}[\mathbf{I}, \mathbf{0}]\mathbf{X}^r_k \\
\mathbf{x}^r_k &= \mathbf{P}_k\mathbf{X}^r_k = \mathbf{K}[\mathbf{R}_k, \mathbf{t}_k]\mathbf{X}^r_k \\
\mathbf{x}^r_k - \mathbf{x}^r_{k-1} &= (\mathbf{P}_k - \mathbf{P}_{k-1})\mathbf{X}^r_k = \mathbf{K}[(\mathbf{R}_k - \mathbf{I}) \mid (\mathbf{t}_k - \mathbf{0})]\mathbf{X}^r_k
\end{aligned} \tag{3.24}$$

$$\mathbf{x}^r_k - \mathbf{x}^r_{k-1} = \mathbf{K}[\mathbf{0}_{3\times 3} \mid \mathbf{t}_k] \begin{bmatrix} X^r_k \\ Y^r_k \\ Z^r_k \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{t}_k \tag{3.25}$$

The substitution of Eq. 3.25 into Eq. 3.22 determines $\mathbf{F}_k$ as:

$$\mathbf{F}_k = [\mathbf{x}^r_k - \mathbf{x}^r_{k-1}]_\times \mathbf{H}_k \tag{3.26}$$

As a result, a unique fundamental matrix $\mathbf{F}_k$ is obtained from Eq. 3.26 when the road surface is near-planar. This $\mathbf{F}_k$ leads to the correct reconstructed road surface points $\mathbf{X}_k^r$ following by identifying defects.

Because of various uncertainties in the 3D reconstruction process, errors will propagate and affect the 3D points $\mathbf{X}_k^r$. Let $\hat{\mathbf{x}}_k^r$ be the measured value of $\mathbf{x}_k^r$ where $\hat{\mathbf{x}}_k^r = \mathbf{x}_k^r + \boldsymbol{\omega}$ and $\boldsymbol{\omega} \sim N(0, \boldsymbol{\Sigma}_{\mathbf{x}_k^r})$ follows a normal distribution. Eq. 3.24 can be rewritten as:

$$\mathbf{X}_k^r = \mathbf{P}_k^+ \hat{\mathbf{x}}_k^r \tag{3.27}$$

where $\mathbf{P}_k^+ = (\mathbf{P}_k^T \mathbf{P}_k)^{-1} \mathbf{P}_k^T$ is the pseudo-inverse matrix of $\mathbf{P}_k$. Let Eq. 3.23 be written as $\mathbf{X}_k^r = f(\mathbf{x}_k^r)$. By using the first-order Taylor series expansion Eq. 3.23 becomes:

$$f \approx f_0 + \mathbf{J}_k \hat{\mathbf{x}}_k^r \tag{3.28}$$

where $\mathbf{J}_k$ represents the Jacobian matrix of $f(\cdot)$. The covariance matrix of $\mathbf{X}_k^r$ thus is approximated by

$$\boldsymbol{\Sigma}_{\mathbf{X}_k^r} \approx \mathbf{J}_k \boldsymbol{\Sigma}_{\mathbf{x}_k^r} \mathbf{J}_k^T \tag{3.29}$$

Since $\mathbf{J}_k$ in this scenario equals to $\mathbf{P}^+$ Eq. 3.25 is deduced to be

$$\boldsymbol{\Sigma}_{\mathbf{X}_k^r} \approx \mathbf{P}^+ \boldsymbol{\Sigma}_{\mathbf{x}_k^r} \mathbf{P}^{+T} \tag{3.30}$$

Therefore, although with a unique $\mathbf{F}$ for the near-planar road surface, the noises in the image inevitably cause errors for the 3D reconstructed surface points $\mathbf{X}_k^r$ due to the ill-posedness of the problem.

3.2.3 Metric Scale 3D Reconstruction

As the 3D points $\mathbf{X}_k^r$ reconstructed using the non-degenerate fundamental matrix $\mathbf{F}_k$ is up to a scale factor, the proposed technique lastly converts $\mathbf{X}_k^r$ to the metric scale 3D points $\mathbf{X}_k^{r,m}$ by using the near-planar characteristics of the road surface. First, a plane is fitted from $\mathbf{X}_k^r$ as:

$$\begin{bmatrix} X_{k,1}^r & Y_{k,1}^r & 1 \\ \vdots & \vdots & \vdots \\ X_{k,N}^r & Y_{k,N}^r & 1 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} Z_{k,1}^r \\ \vdots \\ Z_{k,N}^r \end{bmatrix} \tag{3.31}$$

where a least-squares solution of $\mathbf{p}$ is obtained from Eq. 3.36. Then, every point in $\mathbf{X}_k^r$ has a distance $d_{k,i}^r$ towards the fitted plane:

$$d_{k,i}^r = \frac{p_0 X_{k,i}^r + p_1 Y_{k,i}^r - Z_{k,i}^r + p_2}{\sqrt{p_0^2 + p_1^2 + 1}} \tag{3.32}$$

The camera to plane distance d_k^c in the same scale as the reconstructed $\mathbf{X}_k^r$ is obtained by setting $X_{k,i}^r = Y_{k,i}^r = Z_{k,i}^r = 0$:

$$d_k^c = \frac{|p_2|}{\sqrt{p_0^2 + p_1^2 + 1}} \tag{3.33}$$

Therefore, by knowing the real camera to planar object distance h a scale factor α_k is derived to get the metric scale 3D reconstruction:

$$\mathbf{X}_k^{r,m} = \alpha_k \mathbf{X}_k^r = \frac{h}{d_k^c} \mathbf{X}_k^r \tag{3.34}$$

in which α_k converts the up-to-scale $\mathbf{X}_k^r$ to the metric scale 3D points $\mathbf{X}_{k,i}^{r,m}$. To simplify the notation, let the following $\mathbf{X}_k^{r,m}$ still be written as $\mathbf{X}_k^r$ for the rest of the dissertation.

3.3 EXPERIMENTAL RESULTS

Figure 3.5 illustrates the simulated camera and the road surface in the simulation environment. On the right, the simulated camera is facing towards the simulated road surface, and has simulated properties such as intrinsic matrix and field of view. On the left, the environment creates 3D points $\mathbf{X}_k^r \equiv \left\{(X_{k,i}^r, Y_{k,i}^r, Z_{k,i}^r)^T | \forall i\right\}$ to represent the road surface. $Z_k^r = Z_m + \omega_r$, where $\omega_r \sim N(0, \delta)$ is used to change the evenness of the road in ${}^{\{L\}}Z$ direction. Z_m is the mean distance between camera and the road surface. The default unit in the simulation environment is millimeter.

The simulated images are obtained by reprojecting $\mathbf{X}_k^r$ to the simulated camera. $\hat{\mathbf{x}}_k^r$ are the measured value of $\mathbf{x}_k^r$ defined as $\hat{\mathbf{x}}_k^r = \mathbf{x}_k^r + \boldsymbol{\omega}$, where $\boldsymbol{\omega} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{x}_k^r})$ has the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}_k^r}$ and is used to model the uncertainty for matched features in image. The covariance matrix of $\boldsymbol{\Sigma}_{\boldsymbol{x}_k^r}$ is:

$$\boldsymbol{\Sigma}_{\boldsymbol{x}} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \tag{3.35}$$

As for the orientation, θ_x, θ_y, and θ_z, are the change of angles for the camera about ${}^{\{L\}}X$ axis, ${}^{\{L\}}Y$ axis, and ${}^{\{L\}}Z$ axis between two time steps. Disturbances such as the vibration of the camera cause the orientation change of the camera. Define the error for 3D reconstruction as

$$\epsilon = \frac{1}{N}\sum_{i=1}^{N} |\hat{d}_{k,i}/d_k^c - d_{k,i}/Z_m| \cdot Z_m \tag{3.36}$$

where $\hat{d}_{k,i}$ is the measured distance and Eq. 3.24 shows the ground truth distance $d_{k,i}$. Table 3.1 lists the experimental parameters analyzed in the experiment.

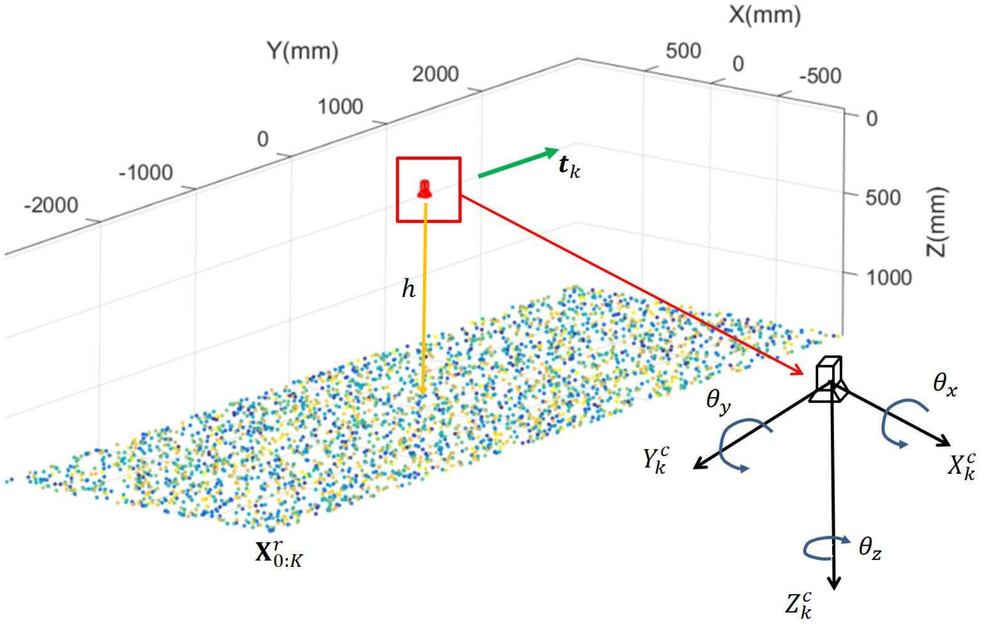


Figure 3.5: Camera and road surface in the simulation environment.

Figure 3.6 shows the comparison of 3D reconstruction error between the proposed technique and traditional SfM. The left figure shows the 3D reconstruction error when the road surface is changing from planar ($\delta = 0$) to non-planar ($\delta >> 0$). When δ is small, the reconstruction error is large for traditional SfM as the degenerate issue still exists, while the proposed technique has small reconstruction errors. The error for the proposed technique in this case is mainly from image noise σ. When the road surface is non-planar, both SfM and the proposed technique have reconstruction error $\epsilon \approx 2$ mm. The right figure shows the reconstruction error influenced by image noise σ at $\delta = 0.1$ and $\delta = 10$. For non-planar road surface which has $\delta = 10$ mm, the proposed technique and traditional SfM both have small

Table 3.1: Parameters for simulated road surface and simulated camera

Parameter	Value
Road unevenness: δ [mm]	0.1, 5, 10
Image noise: σ [pixel]	0.001, 0.002, ..., 0.1
Z_m [mm]	500, 800, 1100, 1400, 1700
θ_x [degree]	0.05, 0.10, ... , 5
θ_y [degree]	0.05, 0.10, ... , 5
θ_z [degree]	0.05, 0.10, ... , 5
Two-view translation $\mathbf{t}$ [mm, mm, mm]	$(200, 30, 0)^T$
Change of h: δ_h [mm]	0.2, 0.4, ... ,20

and similar reconstruction error. When $\delta = 0.1$ mm, i.e., road surface is near-planar, SfM has error usually between 10 to 1000 mm while the proposed technique has error usually less than 1 mm, and even for a much worse case when $\sigma = 0.2$, the error is less than 2 mm.

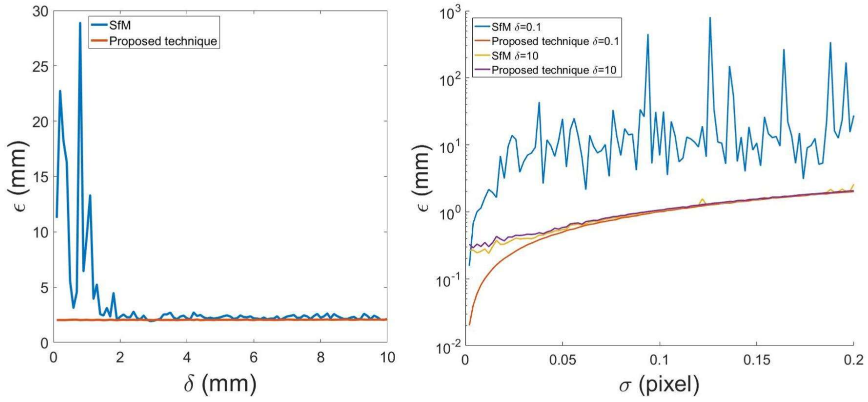


Figure 3.6: Left: 3D reconstruction error comparison between the proposed technique and traditional SfM when road unevenness δ is changing from 0 to 10 mm. Right: 3D reconstruction error comparison between the proposed technique and traditional SfM at different image noise σ while $\delta = 0.1$ or 10 mm.

Figure 3.7 demonstrates the comparison between the traditional SfM and proposed reconstruction technique for planar road and non-planar road 3D reconstruction with different σ. The columns from left to right illustrate the 3D reconstruction under $\delta = 0.1$ and $\delta = 5$ respectively. For each column, the top figure is the 3D reconstruction error obtained

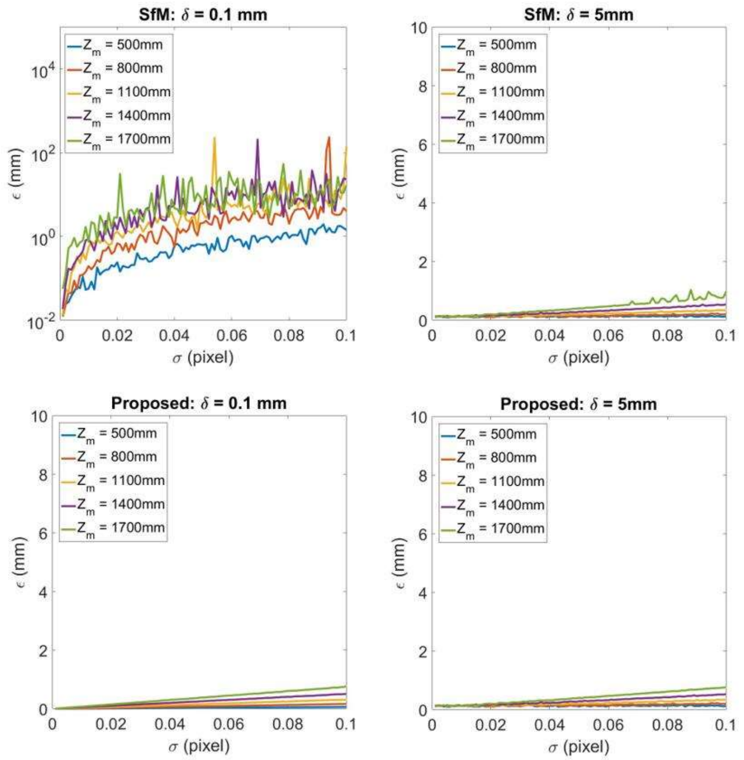


Figure 3.7: 3D reconstruction error for different image noise $\sigma = 0.001, 0.002, ..., 0.1$. From left to right each column represents the results for road unevenness $\delta = 0.1$ and $\delta = 5$ respectively. For each column, the top figure shows the 3D reconstruction by traditional SfM technique, while the bottom figure illustrates 3D reconstruction by the proposed technique.

by traditional SfM and the bottom figure is the error of 3D reconstruction by the proposed technique. The image uncertainty σ is changed from 0.001 to 0.1, while the experiment also alters the distance from camera to road surface Z_m to discover the influence to the results. It can be discovered that when δ becomes larger which means the road is not a planar surface, SfM gives close results to the proposed technique. When δ becomes smaller the error for SfM increases but for the proposed technique the error remains small.

Figure 3.8, 3.9, and 3.10 shows the 3D reconstruction error ϵ by the influence of errors in rotation matrix $\mathbf{R}$. In this simulation experiment, the rotation matrix $\mathbf{R}$ is decomposed as $\mathbf{R} = \mathbf{R}_z\mathbf{R}_y\mathbf{R}_x$ where

$$
\begin{aligned}
\mathbf{R}_x &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \\
\mathbf{R}_y &= \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \\
\mathbf{R}_z &= \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned}
\tag{3.37}
$$

$\mathbf{R}_x, \mathbf{R}_y, \mathbf{R}_z$ are the rotation matrices about the ${}^{L}X$ axis, ${}^{L}Y$ axis, and ${}^{L}Z$ axis correspondingly. The initial camera pose has $\theta_x = 0°, \theta_y = 0°$, and $\theta_z = 0°$. $\mathbf{t} = (200, 30, 0)^T$ in this simulation experiment. In Fig. 3.8, it demonstrates the 3D reconstruction error by changing θ_x. From left to right, each column represents the result under $\delta = 0.1, 5$. σ is set to be 0.2 to represent a worse (relatively large) image noise. The top figure in each column illustrates the results of using traditional SfM, while bottom figure represents the results using the proposed technique. Fig. 3.9 and Fig. 3.10 represents the same experiment by changing θ_y and θ_z.

Figure 3.8 shows the influence to 3D reconstruction error by different θ_x. For the SfM results, when $\delta = 0.1$ the error is usually more than 5% of camera-to-road distance because in this case the error is dominated by the influence of the degenerate issue. In the meantime, 3D reconstruction error is much less by using the proposed technique for the planar road surface. For $\delta = 5$ SfM has error under 2 mm. While for the proposed technique, when $\theta_x = 5°$, the error is only around 1 mm larger than the 3D reconstruction error using SfM.

Figure 3.9 identifies the influence to 3D reconstruction error by different θ_y. The error

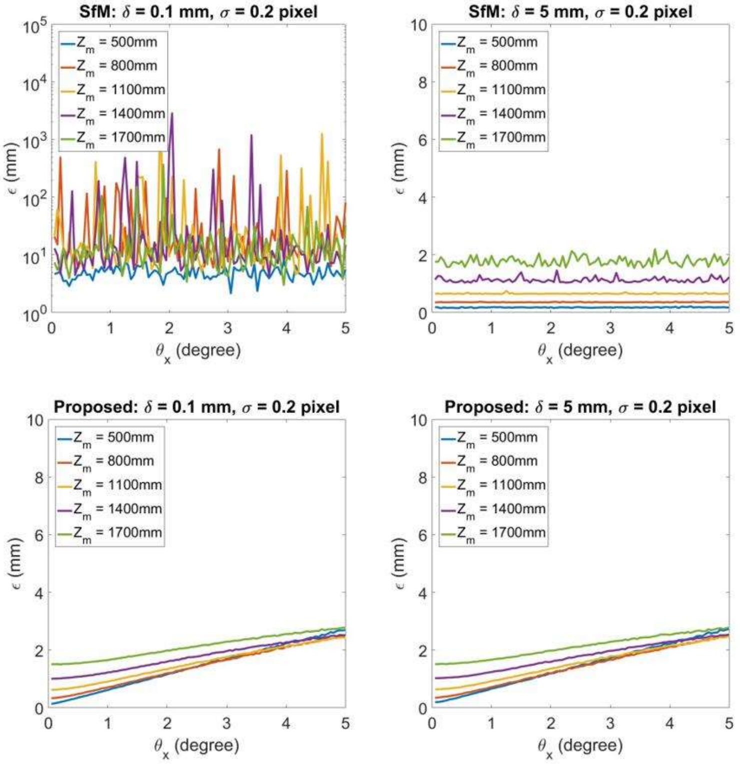


Figure 3.8: 3D reconstruction error for $\theta_x = 0.05°, 0.1°, ..., 5°$. From left to right each column represents the results for $\delta = 0.1, 5$ respectively. For each column, the top figure shows the 3D reconstruction by SfM, while the bottom figure illustrates 3D reconstruction by the proposed degenerate reconstruction technique.

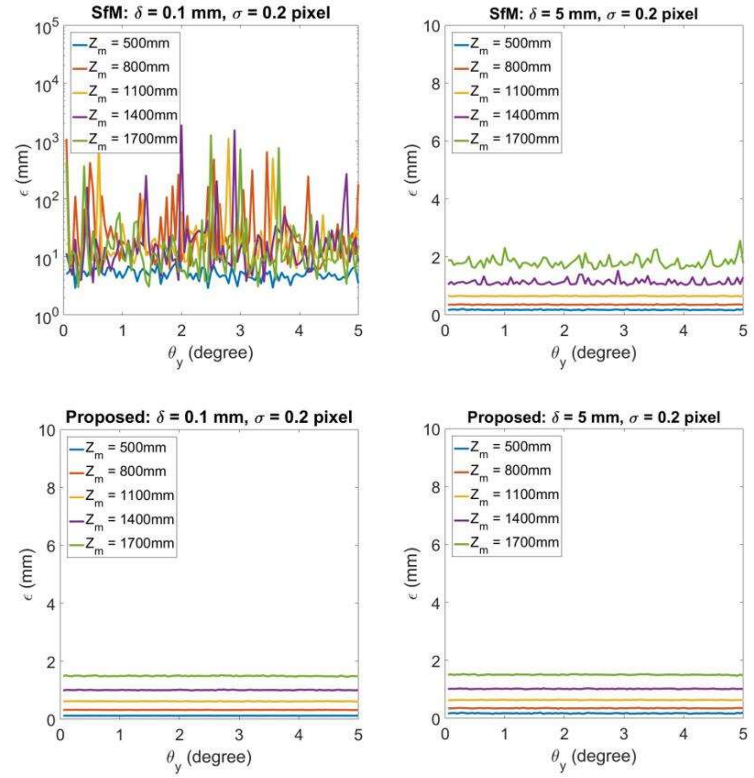


Figure 3.9: 3D reconstruction error for $\theta_y = 0.05°, 0.1°, ..., 5°$. From left to right each column represents the results for $\delta = 0.1, 5$ respectively. For each column, the top figure shows the 3D reconstruction by SfM, while the bottom figure illustrates 3D reconstruction by the proposed degenerate reconstruction technique.

is large and dominated by the influence of degenerate issue for SfM when $\delta = 0.1$, while the proposed technique constructs road with less than 2 mm error. When $\delta = 5$, SfM has comparable error with the proposed technique. For the proposed, the change of θ_y has little influence on the 3D reconstruction errors which are under 2 mm even at the worst case.

Figure 3.10 demonstrates the influence to 3D reconstruction error by different θ_z. For $\delta = 0.1$, the error is also large for traditional SfM because of the degenerate issue while the error is small for the proposed technique. When $\delta = 5$, traditional SfM has comparable error with the proposed technique. For the proposed, the change of θ_z almost has no influence to the 3D reconstruction error. The error in this case is mainly influenced by the variable Z_m.

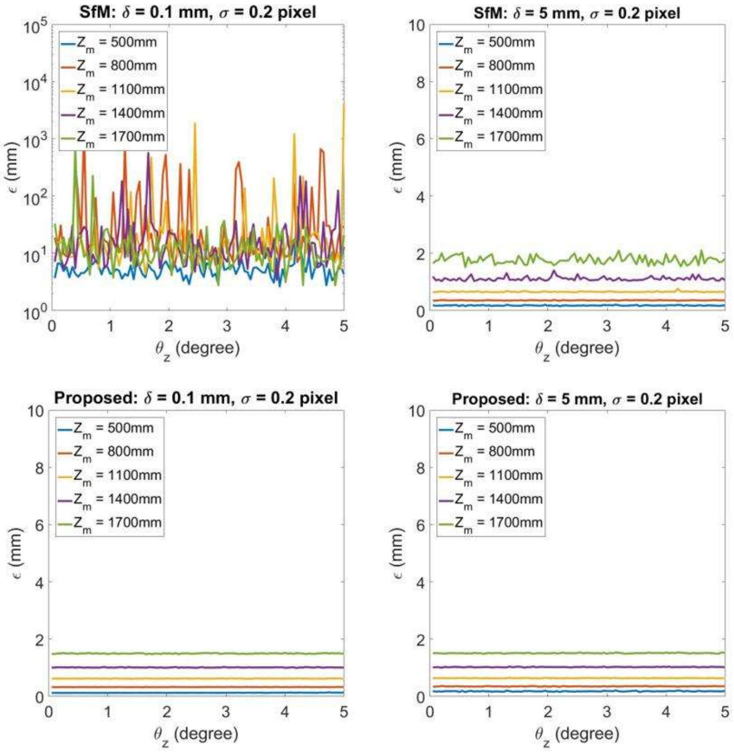


Figure 3.10: 3D reconstruction error for $\theta_z = 0.05°, 0.1°, ..., 5°$. From left to right each column represents the results for $\delta = 0.1, 5$ respectively. For each column, the top figure shows the 3D reconstruction by SfM, while the bottom figure illustrates 3D reconstruction by the proposed degenerate reconstruction technique.

The larger the Z_m, the larger the error ϵ.

Figure 3.11 shows the 3D reconstruction error when there exists a change of height δ_h caused by the vibration in camera to road surface distance h. The measured distance $\hat{h}$ is expressed as

$$\begin{aligned} \hat{h} &= h - \Delta h \\ \Delta h &\sim U(0, \delta_h) \end{aligned} \tag{3.38}$$

where Δh is simulated to having a uniform distribution from 0 to δ_h. In Fig. 3.11 from left to right each column represents the result under $\delta = 0.1, 5$ when σ is 0.2. Each top figure

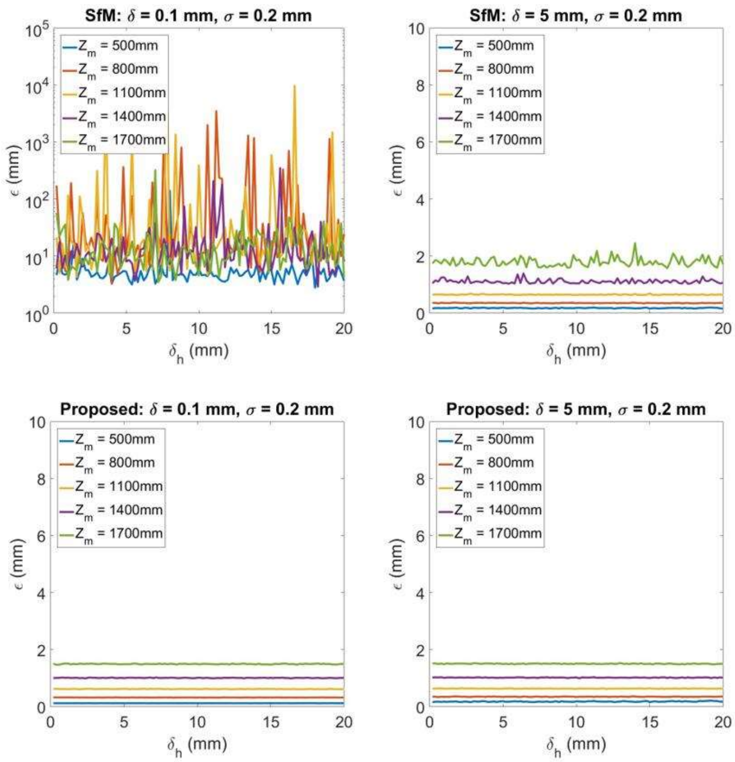


Figure 3.11: 3D reconstruction error of different error $\delta_h = 0.2, 0.4, ..., 20$ for camera to road surface distance h. From left to right each column represents the results for $\delta = 0.1, 5$ respectively. For each column, the top figure shows the 3D reconstruction by SfM, while the bottom figure illustrates 3D reconstruction by the proposed degenerate reconstruction technique.

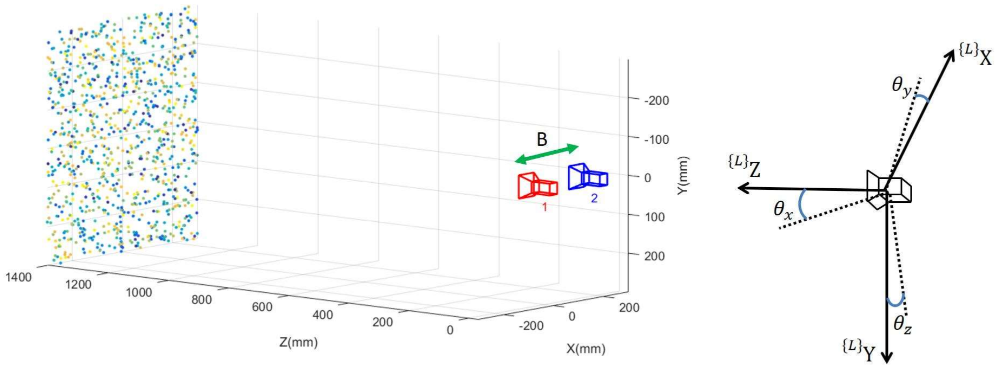
illustrates the results of using traditional SfM, while bottom figure represents the counterpart using the proposed technique. Withing each plot δ_h is changing from 0.2 to 20. The results show that when $\delta = 0.1$ the error from traditional SfM is large for the road surface. When $\delta = 5$ SfM starts to give comparable error with the proposed technique. For the proposed, it can be identified that when δ_h is changing from 0.2 to 20, the error remains almost the same for different δ_h. It means that as the change of camera to ground height h is small during vehicle driving, it has little influence to the 3D reconstruction results by using the proposed technique.

Figure 3.12 illustrates the comparison between Fan's [101] stereo vision road 3D reconstruction technique, traditional SfM, and the proposed technique. Figure 3.12a shows the simulation environment for stereo camera, where the baseline between the two cameras, B, is set to be $B = 200$ mm. Figure 3.12b compares the 3D reconstruction error ϵ from a changing θ_y, caused by the vibration of the vehicle, using stereo technique, traditional SfM, and the proposed technique on the same simulated road which has $\delta = 0.1$ mm and $\sigma = 0.2$ mm. The camera(s) has a height $h = 1400$ mm. To simplify the comparison, let θ_y be the angle for camera 2 respect to camera 1 caused by the vibration. It can be seen that the error from stereo technique exponentially increases when θ_y is larger. Even a relatively small vibration, when $\theta = 0.1$ degree, $\epsilon \approx 10$ mm which is still large for the road surface reconstruction task. Although SfM has smaller error than the stereo technique most of the time after $\theta = 0.2$ degree, it still has a mean error which is over 10 mm. This is still mainly caused by the degenerate issue of the road surface 3D reconstruction. The proposed technique, however, has less than 2 mm reconstruction error which is mainly caused by the image noise σ.

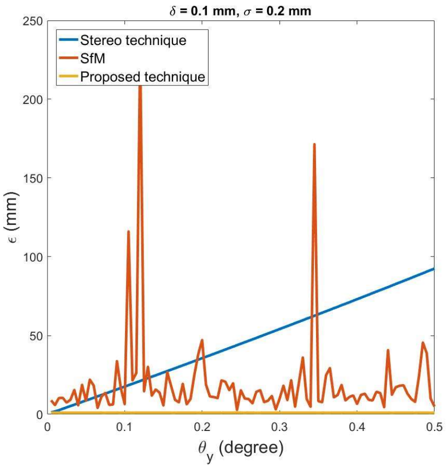
3.3.1 Real-world Field Test

The proposed technique is also analyzed by the real-world experiments. The first real-world experiment is conducted by a gantry setup for error analysis of the camera as shown in Fig. 3.13(a). Fig. 3.13(b) demonstrated a road surface image capturing system developed by the authors [102]. Real road surfaces captured by the system are reconstructed to analyze the quality of road surface reconstruction.

Figure 3.14 illustrates an example of the 3D reconstruction for a same flat plate image using SfM and the proposed technique. Traditional SfM fails in this example since the road surface in the image is a near-planar. However, the proposed one gives the correct planar-like 3D surface reconstruction as shown in Figure 3.14c.

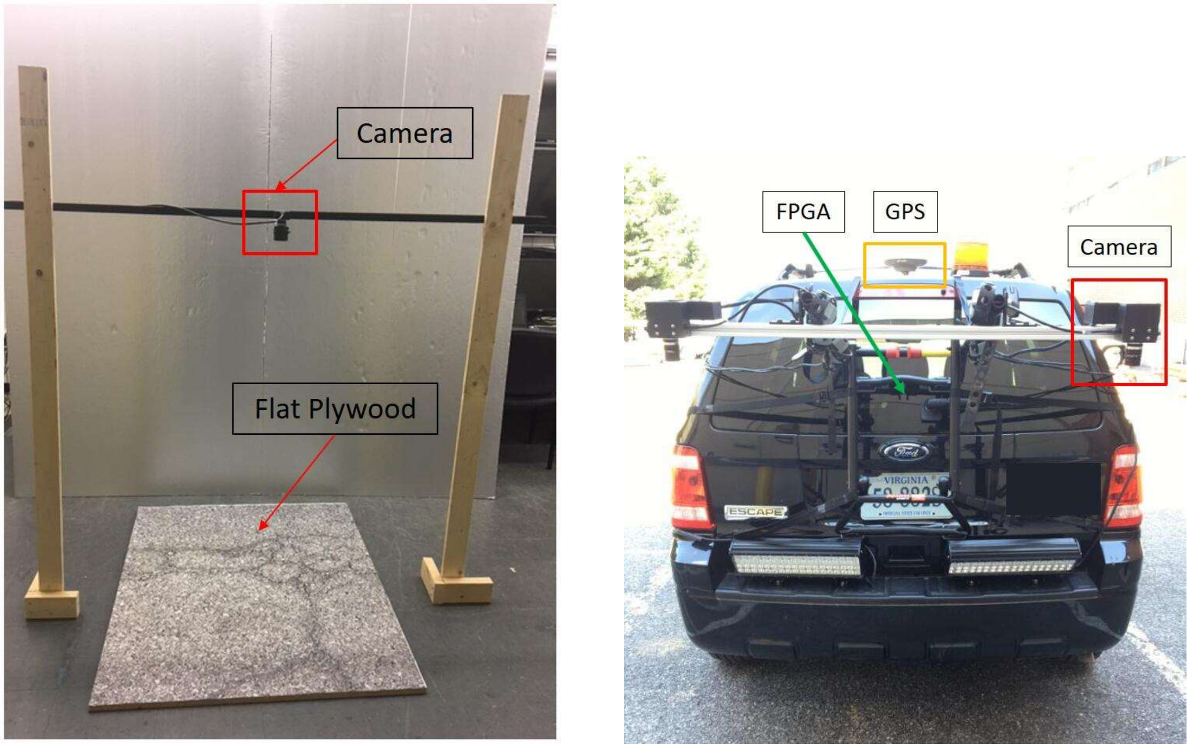


(a) Simulation environment for stereo vision-based technique. B is the baseline between the stereo cameras.



(b) Reconstruction error for stereo technique, traditional SfM, and the proposed technique under vibration which causes changes to θ_y.

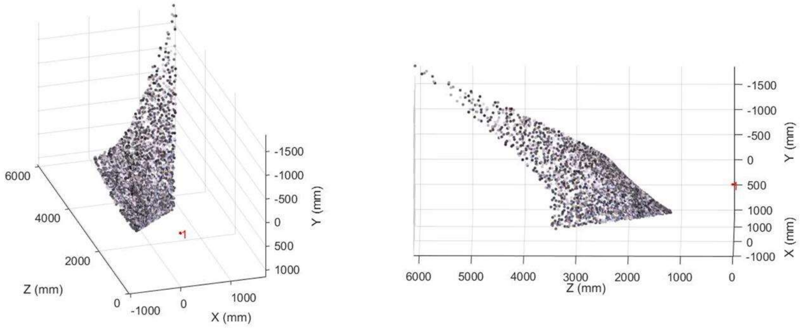
Figure 3.12: Comparison between stereo vision technique, traditional SfM, and the proposed technique under the influence of a changing θ_y which is caused by the vibration.



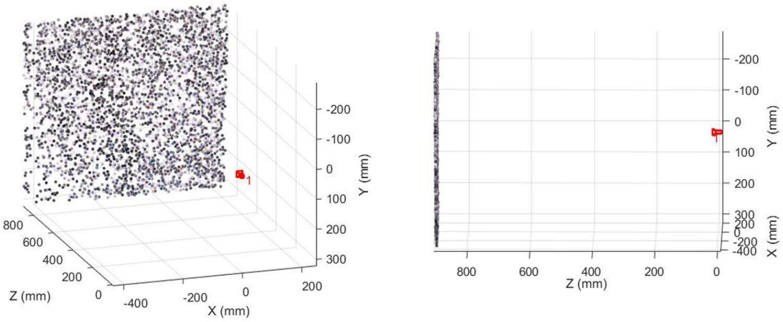
(a) (b)

Figure 3.13: Experimental setups for the near-planar object 3D reconstruction. (a): The camera to plywood distance h is set to be $h = 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600$ mm separately. A flat plywood with mimic ground pattern is placed as a planar object. (b): A road surface capturing system with a downward-facing camera.

(a) An image of the flat plate sticked with mimic road pattern.



(b) 3D reconstruction of flat surface in (a) using traditional SfM. The left image shows the front view of the reconstructed 3D points and the right image shows the left view.



(c) 3D reconstruction of flat surface in (a) using the proposed technique. The left image shows the front view of the reconstructed 3D points and the right image shows the left view.

Figure 3.14: 3D reconstruction of a flat plate using traditional (SfM) and proposed degenerate technique.

Figure 3.15 shows the 3D reconstruction error for near-planar road surface between traditional SfM and the proposed technique using the setup illustrated in Fig. 3.13(a). It can be discovered that the reconstruction by conventional technique has almost one order of magnitude larger errors than the proposed reconstruction technique. Also, the proposed technique has less variance in 3D reconstruction than SfM for near-planar surface.

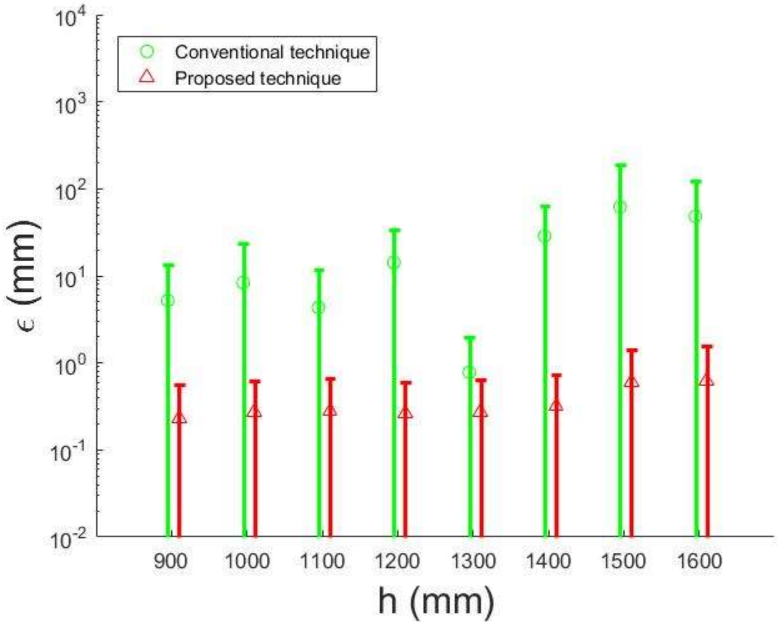


Figure 3.15: 3D reconstruction error ϵ (means and error bars) between the conventional and the proposed technique at different h using the height-adjustable gantry.

Figure 3.16 demonstrates the 3D reconstruction results by the collected road surface images. The colormap of the $Z^r_{0:K}$ from all the reconstructed road surface $\mathbf{X}^r_{0:K}$ are shown using the mesh. The first and the fourth image each represents a road section with pothole and with near-planar characteristics. Each road section is stitched by 50 images and represents about 15-meter road. Image 2, 3, 5, and 6 are the colormap plots for $Z^r_{0:K}$. It can be found that for the road section contains potholes, the performance of 3D reconstruction for the proposed and SfM are similar as shown on image 3 and 4. However, when the road is near-planar, the proposed technique still show a stable and accurate reconstruction of the

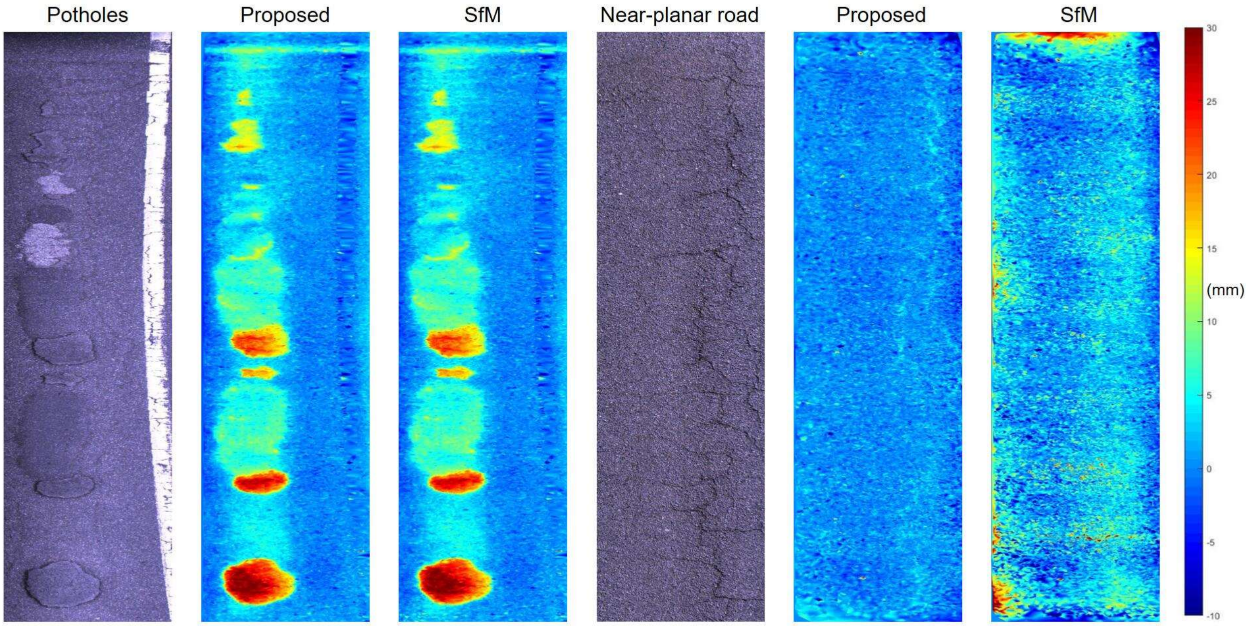


Figure 3.16: Comparison between the proposed technique with SfM on non-planar and near-planar road surfaces. The first and the fourth image each shows a section of road with potholes and near-planar surfaces separately. Image 2, 3, 5, and 6 are the colormap plot for all $Z^r_{0:K}$ with the mesh compared between the proposed technique and SfM.

road surface in image 5 while SfM reconstruct road surface with larger variance for $Z^r_{0:K}$ and fails to reconstruct correct 3D road surface at many places as shown in image 6. Image 5 also demonstrates that the proposed technique can even reconstruct and differentiate miniature road cracks. This comparison coincides with the simulation results for the proposed technique and SfM.

3.4 Chapter Summary

This chapter has presented a 3D reconstruction technique which reconstructs near-planar road surface from two images. The degenerate issue for the road surface 3D reconstruction was tackled by deriving a non-degenerate fundamental matrix from the estimation of near-planar homography, camera's translation, and camera's rotation. Only two views captured with overlapped area on the road surface are required to reconstruct 3D road surface points rather than multiple views which are not guaranteed while the vehicle moves fast. Absolute metric scales are obtained by a conversion procedure for 3D points using the near-planar characteristics.

Experiments were conducted in both the simulation environment and the field test for the proposed 3D road surface reconstruction technique. The parameters in the reconstruction processes were studied in the simulation environment. By changing to different unevenness δ, image noises σ, $\Delta\theta_x$, and $\Delta\theta_y$, the proposed technique outperformed the conventional technique for road surface 3D reconstruction in the simulation test. During the field test, a height-adjustable camera gantry setup demonstrated that the accuracy of proposed technique for near-planar objects 3D reconstruction was at least one order of magnitude better in accuracy than the conventional technique. Test on the road surface capturing system showed the quality and accuracy of reconstructed near-planar objects. Both the simulation and field test demonstrate the effectiveness and reliability of the proposed non-degenerate road surface 3D reconstruction technique.

Chapter 4

Physics-based Automatic Defects Detection for Near-planar Surfaces

This chapter presents a technique that automatically detects road defects using the 3D geometry and 2D image. The image and geometry are registered on the near-planar road as the 3D surface geometry is reconstructed exactly from image features. The chapter is organized as follows. The first section introduces the formulation of the road surface detection problem. Section 4.2 illustrates the proposed automatic road surface defects detection from geometry and the corresponding images. Section 4.3 shows the experimental validation of the proposed technique. Conclusions are summarized in the last section.

4.1 Road Defects Detection Problem Formulation

Figure 4.1 shows general settings and problem formulation of road surface reconstruction using a downward-facing camera for road defects detection. The road surface, shown as a near-planar object, contains a pothole representing a defect road. A camera, facing downward to the road surface at a height h, is mounted on a vehicle. While the vehicle is moving, the camera captures images $I_{0:K}$ at positions $\mathbf{X}^c_{0:K}$ from time step 0 to time step K. Since images are captured by a camera of various frame rates at various vehicle speeds, minimally and most fundamentally required is the reconstruction of a 3D road surface overlapped by two

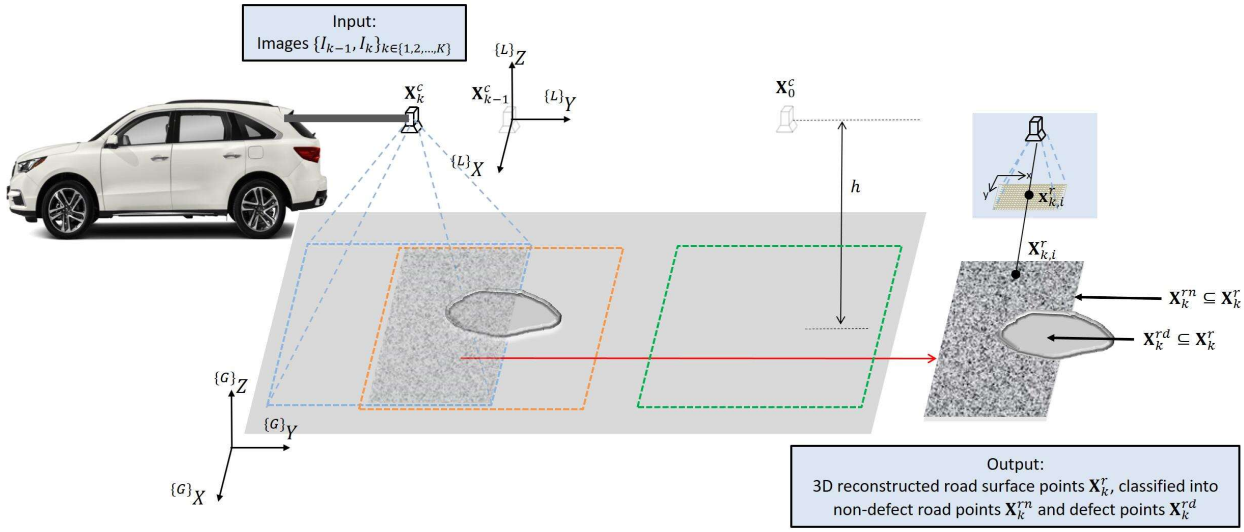


Figure 4.1: Road surface reconstruction settings for defects detection from one downward-facing camera. 3D point cloud are reconstructed from consecutive images to represent the road surface, followed by classifying road into defect and non-defect surfaces.

consecutive images $\{I_{k-1}, I_k\}$. This problem is converted into localizing the road surface point cloud $\mathbf{X}_k^r \equiv \left\{\mathbf{X}_{k,i}^r | \forall i\right\}$ using the homogeneous 2D image features $\mathbf{x}_{k-1}^r \equiv \left\{\mathbf{x}_{k-1,i}^r | \forall i\right\}$ and the corresponding $\mathbf{x}_k^r \equiv \left\{\mathbf{x}_{k,i}^r | \forall i\right\}$, which are extracted from image I_{k-1} and I_k respectively. It is to be noted that $\mathbf{X}_k^c$ should be derived simultaneously with $\mathbf{X}_k^r$ since the camera position is not precisely known due to the vehicle vibration. Once the reconstruction has been completed, road surface points are classified as normal flat road surface $\mathbf{X}_k^{rn}$ and defect road surface $\mathbf{X}_k^{rd}$. In Figure 4.1, $\{G\}$ represents the global coordinate system while $\{L\}$ is the local coordinate for two neighboring camera positions.

4.2 Automatic Road Surface Defects Detection from Degenerate Near-planar Reconstruction

Figure 4.2 shows the proposed automatic defects detection and estimation technique based on degenerate near-planar 3D reconstruction. First, the preprocessing keeps correct feature matching between I_{k-1} and I_k while removing wrong feature correspondences using the camera's motion information. Based on the accuracy-enhanced near-planar 3D reconstruction which solves the degenerate issue, road defect points are classified by fitting the reference road surface thus anomaly off-surface 3D points are detected. Since the 3D points reconstructed from the proposed technique are inversely obtained from images, the proposed technique automatically detect road defects by combining the 3D defect points and the corresponding brightness information from 2D images.

4.2.1 Preprocessing

The preprocess rejecting mismatched correspondences is formulated as follows. Right before the preprocessing, image features $\mathbf{x}^r_{k-1}$ and $\mathbf{x}^r_k$ in I_{k-1} and I_k are found by feature descriptors such as SIFT [23] which is used in this paper. For SIFT features $\mathbf{x}^r_{k-1}$, each feature is represented by a 128-vector. For the ith feature in $\mathbf{x}^r_{k-1}$, $\mathbf{x}^r_{k-1,i}$, the Euclidean distance d_e between $\mathbf{x}^r_{k-1,i}$ and the jth feature in $\mathbf{x}^r_k$ is defined as:

$$d_e = \|\mathbf{x}^r_{k-1,i} - \mathbf{x}^r_{k,j}\| \tag{4.1}$$

and the matched feature for $\mathbf{x}^r_{k-1,i}$ is the feature $\mathbf{x}^r_{k,j}$ which results in the minimum d_e.

After the feature matching process, let the difference of the ith corresponding feature

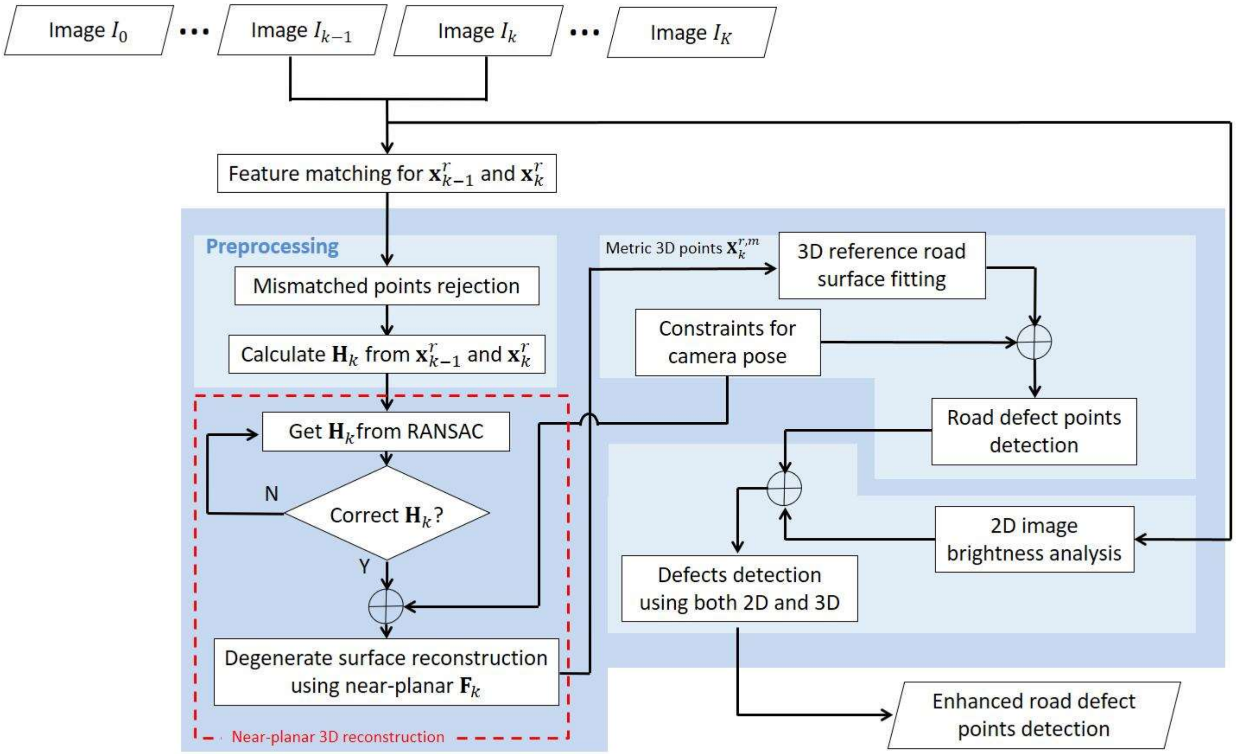


Figure 4.2: Automatic 3D reconstruction-based defects detection and estimation for near-planar road surface.

at time step $k-1$ and k be:

$$\mathbf{d}^f_{k,i} \equiv \mathbf{x}^r_{k,i} - \mathbf{x}^r_{k-1,i} \tag{4.2}$$

This makes the set $\mathbf{d}^f_k \equiv \left\{\mathbf{d}^f_{k,i} | i = 1, 2, ..., n\right\}$, which includes all the n correspondences of the images I_{k-1} and I_k. As the vehicle is moving along the road following a smooth path, it is valid to assume that the rotation of the camera is small and the camera's motion is linear in a short period between two neighboring time steps $k-1$ and k:

$$\mathbf{d}^f_k \propto \mathbf{X}^c_k - \mathbf{X}^c_{k-1} \tag{4.3}$$

which means $\mathbf{d}^f_k$ are also linear and proportional to the camera's motion.

Since I_{k-1} and I_k has Gaussian noises for $\mathbf{x}^r_{k-1}$ and $\mathbf{x}^r_k$ and n is large with the difference distributed smoothly, the measured image corresponding features $\hat{\mathbf{x}}^r_{k-1}$ and $\hat{\mathbf{x}}^r_k$ are:

$$\hat{\mathbf{x}}^r_{k-1} = \mathbf{x}^r_{k-1} + \boldsymbol{\omega}_{k-1}, \quad \boldsymbol{\omega}_{k-1} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{k-1}) \tag{4.4}$$

$$\hat{\mathbf{x}}^r_k = \mathbf{x}^r_k + \boldsymbol{\omega}_k, \quad \boldsymbol{\omega}_k \sim N(\mathbf{0}, \boldsymbol{\Sigma}_k) \tag{4.5}$$

Combining Eq. 4.4 and 4.5 with Eq. 4.2, while also based on the Kolmogorov–Smirnov test [103] for $\mathbf{d}^f_k$ to test the normality of the distribution, the proposed technique models $\mathbf{d}^f_k$ as a Gaussian distribution $\mathbf{d}^f_k \sim N(\overline{\mathbf{d}}^f_k, \boldsymbol{\Sigma}^f)$:

$$\overline{\mathbf{d}}^f_k = \overline{\mathbf{x}}^r_k - \overline{\mathbf{x}}^r_{k-1}, \quad \boldsymbol{\Sigma}^f = \boldsymbol{\Sigma}_{k-1} + \boldsymbol{\Sigma}_k \tag{4.6}$$

where $\overline{\mathbf{d}}^f_k$ is the mean value and $\boldsymbol{\Sigma}^f$ is the covariance matrix of $\mathbf{d}^f_k$. As $\mathbf{d}^f_{k,i}$ of correct matches are closer to $\overline{\mathbf{d}}_k^f$ than those of the mismatched features, mismatched correspondences can be rejected by defining correct matching as:

$$\mathbf{d}_k^{f,c} = \{\mathbf{d}_k^f \,|\overline{\mathbf{d}}_k^f - \lambda\mathbf{\Sigma}^f\mathbf{1} < \mathbf{d}_k^f < \overline{\mathbf{d}}_k^f + \lambda\mathbf{\Sigma}^f\mathbf{1}\} \tag{4.7}$$

where λ is a threshold and $\mathbf{1}$ is an all-ones vector. As the exact distance that the camera moves between time step $k-1$ and k is unknown, RANSAC technique is difficult to determine the threshold and number of iterations to filter correct feature matchings. However, the proposed technique uses the camera's linear motion as a prior knowledge, which means correct matchings have similar values in $\mathbf{d}_{k,i}^f$. Unlike RANSAC, Eq. 4.7 only need to find a reasonable λ and operate once to keep the correct matching within a range $(\overline{\mathbf{d}}_k^f - \lambda\mathbf{\Sigma}^f\mathbf{1}, \overline{\mathbf{d}}_k^f + \lambda\mathbf{\Sigma}^f\mathbf{1})$. Therefore, the proposed technique obtains correct feature matchings for the following near-planar 3D reconstruction.

4.2.2 Anomaly Road Surface Points Detection from Near-planar Road Reconstruction

After getting the 3D road surface points $\mathbf{X}_k^r$ from Section 3.2.2, the proposed technique classifies $\mathbf{X}_k^r$ into normal road points and anomaly road points. The reconstructed surface and distance d_k^r obtained by Eq. 3.32, however, may not be the final reconstruction. Because the road surface may have anomalies such as potholes, the first-time road surface reconstruction will be distorted if such anomaly exists. Thus a recursive surface fitting process is proposed to reconstruct the road surface through Eq. 4.8 to 4.10:

$$d_{k,i} = \frac{p_0 X^r_{k,i} + p_1 Y^r_{k,i} - Z^r_{k,i} + p_2}{\sqrt{p_0^2 + p_1^2 + 1}} \tag{4.8}$$

$$\mathbf{X}^r_{k,i} \in \begin{cases} \mathbf{X}^{rd}_k, if\ d_{k,i} < 0\ and\ d_{k,i} \leq T_d \\ \mathbf{X}^{rn}_k, else \end{cases} \tag{4.9}$$

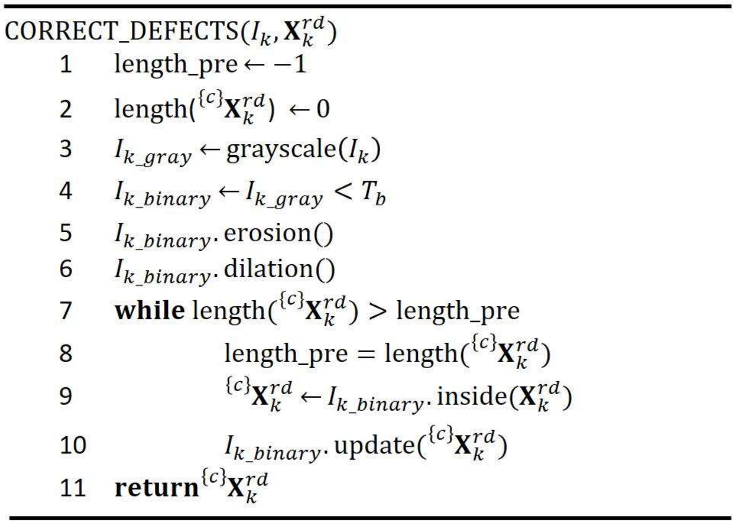
$$T_n = \frac{size(\mathbf{X}^{rn}_k)}{size(\mathbf{X}^{rn}_k + \mathbf{X}^{rd}_k)} \tag{4.10}$$

In Eq. 4.8, $d_{k,i}$ is a signed value calculated as the distance of $\mathbf{X}^r_{k,i}$ to the current reconstructed road surface. The positive $d_{k,i}$ represents the point $\mathbf{X}^r_{k,i}$ located in between the camera and the current fitted road surface. The negative $d_{k,i}$ means the point $\mathbf{X}^r_{k,i}$ is at the other side of the current road surface. Eq. 4.9 illustrates the classification of $\mathbf{X}^r_{k,i}$ into possible defect points $\mathbf{X}^{rd}_k$ and non-defect points $\mathbf{X}^{rn}_k$ by a depth threshold T_d. T_n in Eq. 4.10 is a threshold refers to the percentage of non-defect points among all the points $\mathbf{X}^r_k$. If it is assumed that at least m percent of the points $\mathbf{X}^r_k$ are actually representing non-defect road surface, then a $T_n > m$ will continue the recursive process to fit a new road surface based on all the $\mathbf{X}^{rn}_k$ from the last iteration. The recursive process will continue until $T_n < m$ is reached.

4.2.3 Automatic Road Defects Detection from both 2D and 3D Information

After getting the 3D reconstructed road surface $\mathbf{X}^r_k$ which are classified into $\mathbf{X}^{rd}_k$ and $\mathbf{X}^{rn}_k$, the proposed technique further improves the classification of road surface points. As the inevitable noises in image feature matching propagates to the 3D reconstruction based on Equations 3.27 to 3.30, current $\mathbf{X}^{rd}_k$ obtained from the latest classification process still have false classified points. In order to solve this issue, the proposed technique combines 2D image color information with 3D reconstructed points to filter out false classified points since defect road region has different image intensity than normal road region.

Figure 4.3 shows the proposed algorithm to improve the classification of road surface defect points using both the 2D color image information and the 3D road surface reconstruction. The function CORRECT_DEFECTS() takes two inputs: the kth image I_k and the defect road surface points $\mathbf{X}_k^{rd}$ classified from the last section. At line 1 (L1), a variable length_pre is initialized to -1. At L2, ${}^{\{c\}}\mathbf{X}_k^{rd}$ represents the correct road surface defect points classified in $\mathbf{X}_k^{rd}$ and length(${}^{\{c\}}\mathbf{X}_k^{rd}$) is a function which returns the number of points in ${}^{\{c\}}\mathbf{X}_k^{rd}$. L3 converts I_k to the grayscale image $I_{k_{gray}}$. L4 conducts the image binarization using a threshold T_b, which means for every pixel if its intensity is less than T_b its value becomes 1, and otherwise 0. From L5 to L6 the erosion() and dilation() are common operation to erode the image first and then dilate it to remove the noise. Since inside the 3D defect region of the road surface it has the image intensity difference comparing with the normal road surface under natural lighting condition, L5 to L6 are the processes to find the low intensity region in the image which potentially appears together with defects. Since the shadow region (where $I_{k\_binary} == 1$) and its vicinity may contain defect, the dilation() will enlarge the area of $I_{k\_binary}(I_{k\_binary} == 1)$. L7 to L10 keeps finding the correct defect points based on the updated 2D defect region. Although $\mathbf{X}_k^{rd}$ are 3D points, according to the road surface 3D reconstruction process, they corresponds to their 2D corresponding points in the image. Thus L9 checks the corresponding 2D positions of $\mathbf{X}_k^{rd}$ and returns the correct defect points which are located in the region of $I_{k\_binary}(I_{k\_binary} == 1)$, which is called the potential defects region (PDR). In L10, the I_{k_binary}.update(${}^{\{c\}}\mathbf{X}_k^{rd}$) function updates the PDR using the new defect points ${}^{\{c\}}\mathbf{X}_k^{rd}$ by the same dilation procedure used in L6. The dilation process uses a square window with a width of W, which changes the I_{k_binary} values to 1 at all the pixels within this square window if the center pixel value is originally 1. Finally, if no more new ${}^{\{c\}}\mathbf{X}_k^{rd}$ are found through this process, the current ${}^{\{c\}}\mathbf{X}_k^{rd}$ are returned as the correct defect road surface points.



```
CORRECT_DEFECTS($I_k, \mathbf{X}_k^{rd}$)
length_pre ← −1
length(${}^{\{c\}}\mathbf{X}_k^{rd}$) ← 0
$I_{k\_gray}$ ← grayscale($I_k$)
$I_{k\_binary}$ ← $I_{k\_gray} < T_b$
$I_{k\_binary}$. erosion()
$I_{k\_binary}$. dilation()
while length(${}^{\{c\}}\mathbf{X}_k^{rd}$) > length_pre
        length_pre = length(${}^{\{c\}}\mathbf{X}_k^{rd}$)
        ${}^{\{c\}}\mathbf{X}_k^{rd}$ ← $I_{k\_binary}$. inside($\mathbf{X}_k^{rd}$)
        $I_{k\_binary}$. update(${}^{\{c\}}\mathbf{X}_k^{rd}$)
return ${}^{\{c\}}\mathbf{X}_k^{rd}$
```

Figure 4.3: Defect road surface points classification algorithm using both 2D image and 3D reconstruction information.

4.3 Experimental Results

4.3.1 Experiments on Real Road Surface

The proposed mismatched points rejection was evaluated based on two consecutive images. From Eq. 4.7, the detected correct matched points ${}^{\{c\}}\mathbf{X}_k^{rd}$ are kept by using the camera's motion pattern and a parameter λ.

Figure 4.4 shows the mismatched points rejection process and its performance by analyzing two consecutive images. Fig. 4.4(a) and 4.4(b) illustrate the kth image and the $(k+1)$th image captured by the camera. Fig. 4.4(c) demonstrates the feature points matching between the kth image and the $(k+1)$th image. It can be found that there are some obvious mismatched points among the matching. Fig. 4.4(d) shows the detected mismatched points based on Eq. 4.7 where = 3. The red rectangular areas contain the correct matched points although they are detected as mismatched points, which shows the side effect of removing also some correct matching during mismatched rejection. Fig. 4.4(e) then evaluates the performance of mismatched points rejection by using different λ. During the experiment, the total number of all the matched feature points is 4362 and the number of true mismatched points in the image is 32. From $\lambda = 1$ to 4.5, all the 32 true mismatched points are detected and the number of matched points among the detected mismatched points $\mathbf{X}_k^{rd}$ is decreasing. When $\lambda = 5$, the number of true mismatched points becomes 30 which means 2 of them are not detected by Eq. 4.7. Based on the experiment, a smaller λ can detect not only most of the mismatched points but also keep more matched points into, which increases the number of 3D reconstruction points as these 3D reconstruction are coming from matched features. Meanwhile, a larger λ can reduce the number of matched points in $\mathbf{X}_k^{rd}$ while the detection for true mismatched points may be missed, which will lead to the failure of the 3D reconstruction caused by mismatched features. Since starting from $\lambda = 1.5$ the ratio between the matched points among the detected mismatched points and the total matched feature points is less than 10%, the density of the feature matching points is still high since it still keeps more than 90% of the original points. Therefore from Fig. 4.4(e) the value of λ can be between 1.5 to 2.5 in order to keep more points meanwhile remove all the mismatched points.

Figure 4.5 shows the experimental setup of the error analysis for the proposed technique using real images. The camera is facing towards the ground with its principle axis vertical to the ground surface as shown in Fig. 4.5a. The ground surface is made by a flat plate to mimic a degenerate planar road surface as illustrated in Fig. 4.5b. An image of road surface is printed and stuck to the flat plate to provide road surface patterns for the image feature

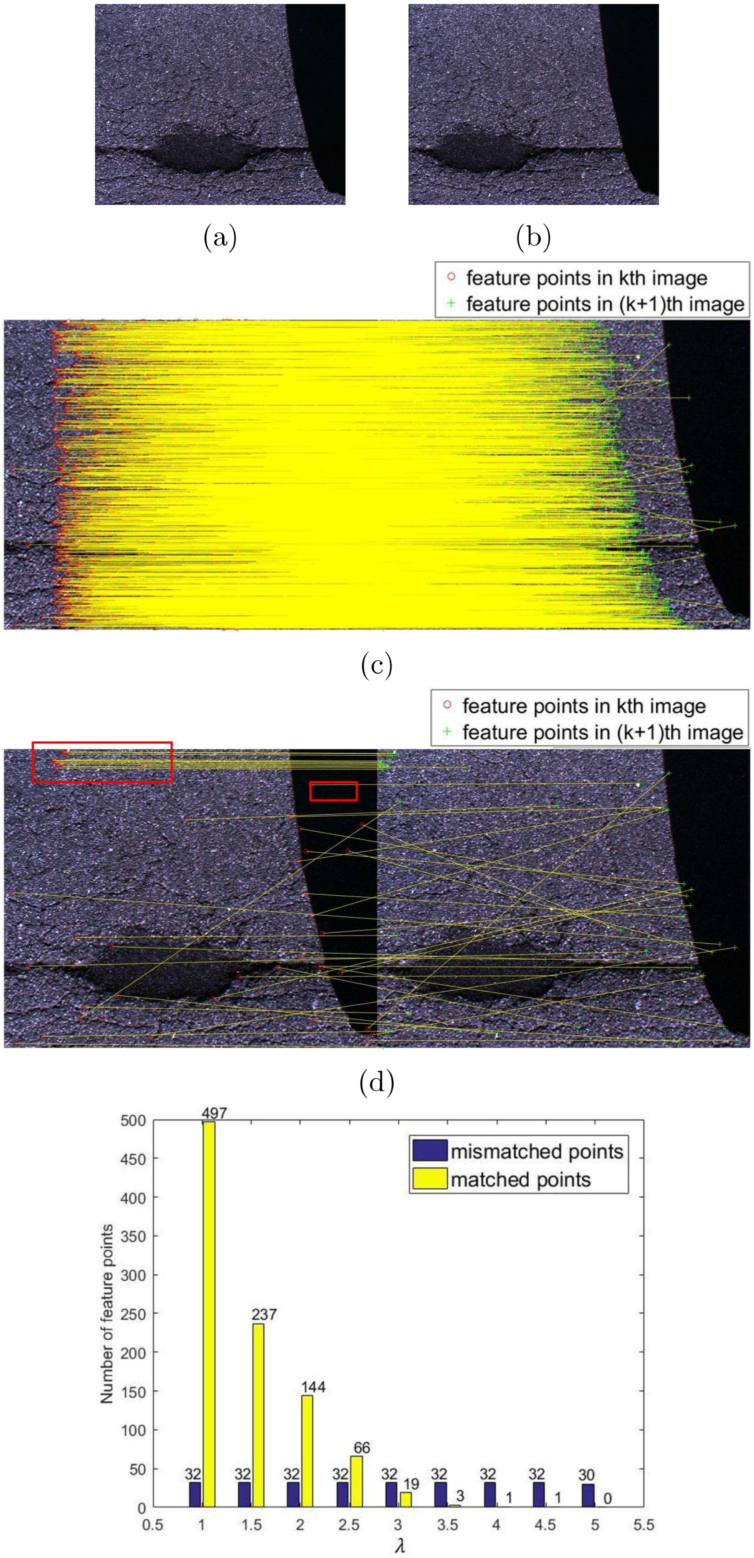
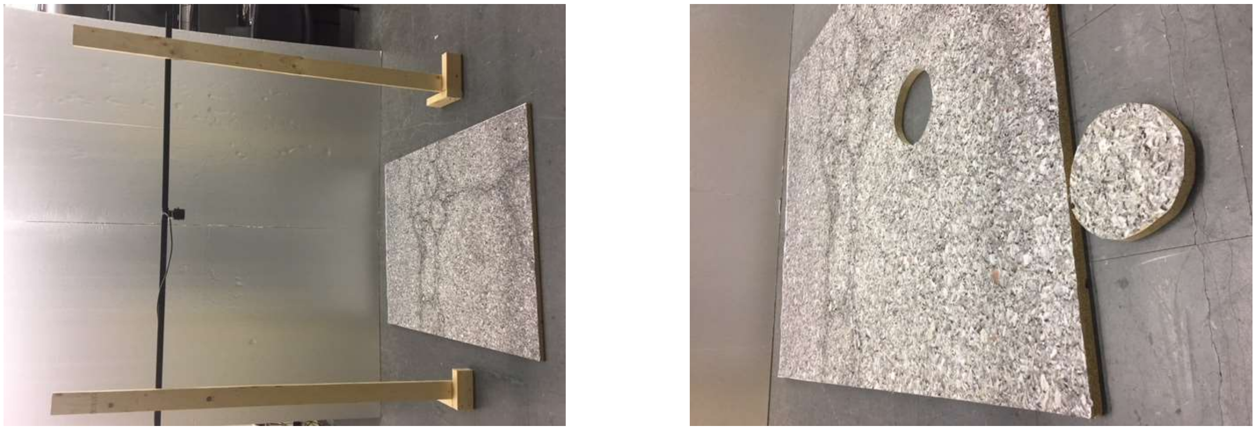


Figure 4.4: (a) (b) Two consecutive images captured by the camera. (c) Feature points matching between two images in (a) and (b). (d) Detected mismatched feature points between two images. (e) Performance of mismatched feature points detection at different λ.



(a) A height-adjustable gantry for the camera.

(b) A flat plate sticked with a mimic road pattern image.

Figure 4.5: Experimental setup for accuracy analysis of the proposed degenerate 3D reconstruction method. The camera to road distance h is set to be $h = 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600$ mm. A flat plate with mimic road surface pattern is placed as a degenerate plane. A hole on the plate can be used to simulate the road defect.

searching and matching. A circular part of the plate can be removed from the plate to mimic the road pothole.

Table 4.1: Parameters for experiments using real images.

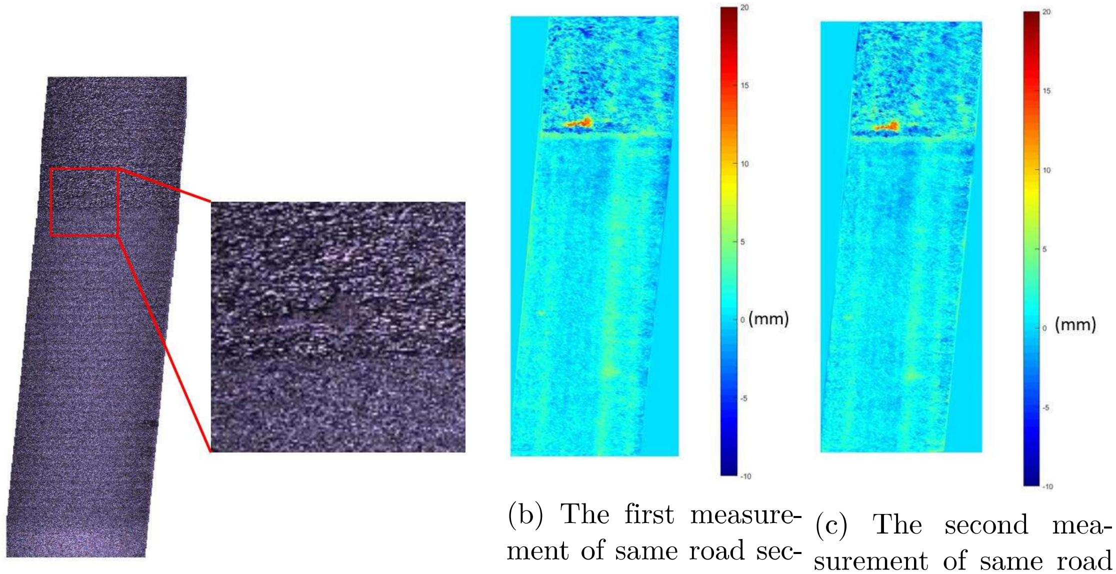
Parameter	Value
Camera Field of View	$56° \times 44°$
Road unevenness: δ [mm]	< 0.5
Camera to road distance: h [mm]	900, 1000, ... ,1600
Image noise: σ [pixel]	< 0.2
Two-view camera translation: $\mathbf{t}$ [mm, mm, mm]	$(100, 0, 0)^T$

Figure 4.6 demonstrates the repeatability experiment for the proposed technique. Figure 4.6a represents a section of the road which contains a pothole. The images of road surface are obtained by the system shown in Fig. **??** and then stitched together. In Figure 4.6b the $Z^r - h$ values of reconstructed 3D road surface points are plotted as a colormap.

In Fig. 4.6c, the proposed technique measures the same road section as Fig. 4.6b. The two measurements are then compared to validate the repeatability of the proposed technique. In Figure 4.6d, Z_1 are the Z^r values of the reconstructed road surface points from the first measurement, while Z_2 are the ones from the second measurement. The histogram shows the count of points with different $Z_2 - Z_1$ values. The mean value of $Z_2 - Z_1$ is -0.1079 mm and the standard deviation of $Z_2 - Z_1$ is 1.3515 mm. The statistics results of $Z_2 - Z_1$ reflects the high repeatability of the proposed technique.

Fig. 4.7 shows the automatic 3D defects detection process using both 2D and 3D information on five typical scenarios. Fig. 4.7(a) shows five typical situations. The first image contains both the hole-like and the crack-like defect. The second one has a defect with a small area. The third one represents a pothole region. The fourth image contains a large defect area at the bottom of the image. The last one is the flat road surface and has no defect in it. Then Fig. 4.7(b) illustrates the 3D points reconstructed from the original 2D images and their next frames. The red dots in the figures represent the defect position based on the 3D information from 3D reconstruction. Next, in Fig. 4.7(c), all the detected defects $\mathbf{X}_k^{rd}$ are mapped to the corresponding position in the image. It can be found that although most of the $\mathbf{X}_k^{rd}$ are located in the defect region, there are still several $\mathbf{X}_k^{rd}$ which are located not at the defect region. Therefore in Fig. 4.7(d), the binary 2D image is also used together with the information from 3D points. The function CORRECT_DEFECTS() defined in Fig. 4.3 is applied to refine the defects detection by using both 2D and 3D information of the road surface. Finally, Fig. 4.7(e) demonstrates the refined defects detection results. It shows the removal of the incorrect $\mathbf{X}_k^{rd}$ and keeps the true defect points in ${}^{\{c\}}\mathbf{X}_k^{rd}$.

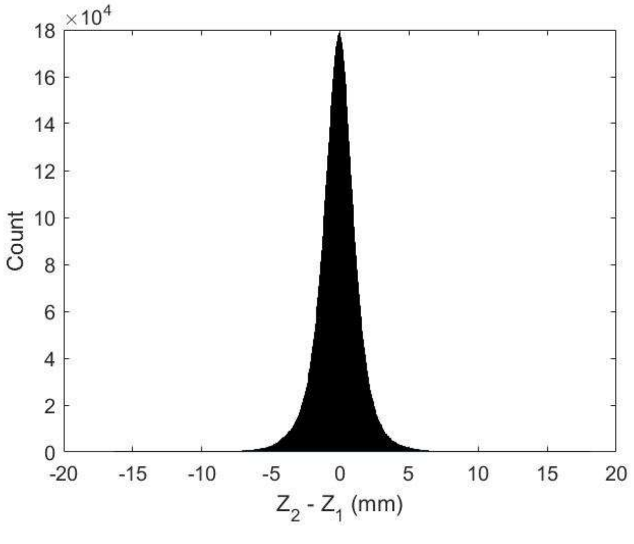
Table 5.1 compares SfM with the proposed technique on defects detection using road surface images. The comparison is based on 6300 road surface images which are collected at rural, urban, and highway road for weather conditions such as sunny, cloudy, and partly



(a) Stitched images of a road section which has a geometrical defect.

(b) The first measurement of same road section. $Z^r - h$ values of road surface point cloud data are represented by a colormap.

(c) The second measurement of same road section. $Z^r - h$ values of road surface point cloud data are represented by a colormap.



(d) Repeatability quantitative results. Z_1 are the Z^r values from the first measurement while Z_2 are the ones from the second measurement.

Figure 4.6: Repeatability test for the proposed technique.

cloudy around Blackburg, Virginia area. The real road surface images are captured at both highway driving speed (100 km/h) and local road driving speed (40 km/h). Some images capture potholes while other images capture flat road surface. From true positive (TP), false positive (FP), true negative (TN), and false negative (FN), the accuracy is expressed as $(TP+TN)/(TP+TN+FP+FN)$, precision as $TP/(TP+FP)$, while the recall illustrated by $TP/(TP+FN)$. From Table 5.1 although SfM gives higher recall rate between proposed degenerate method and the non-degenerate method, it has only 34.34% precision rate. It means that although SfM rarely misses the detection of potholes (less FN), this method also generates wrong detection of potholes (more FP). The proposed technique on the other hand, results in 98.95% accuracy, 94.33% precision and 95.76% recall rate. All the three criteria are above 94%.

Table 4.2: Performance of road surface defects detection for different methods.

	Proposed	SfM
TP	632	658
TN	5602	4382
FP	38	1258
FN	28	2
Accuracy	98.95%	80%
Precision	94.33%	34.34%
Recall	95.76%	99.70%

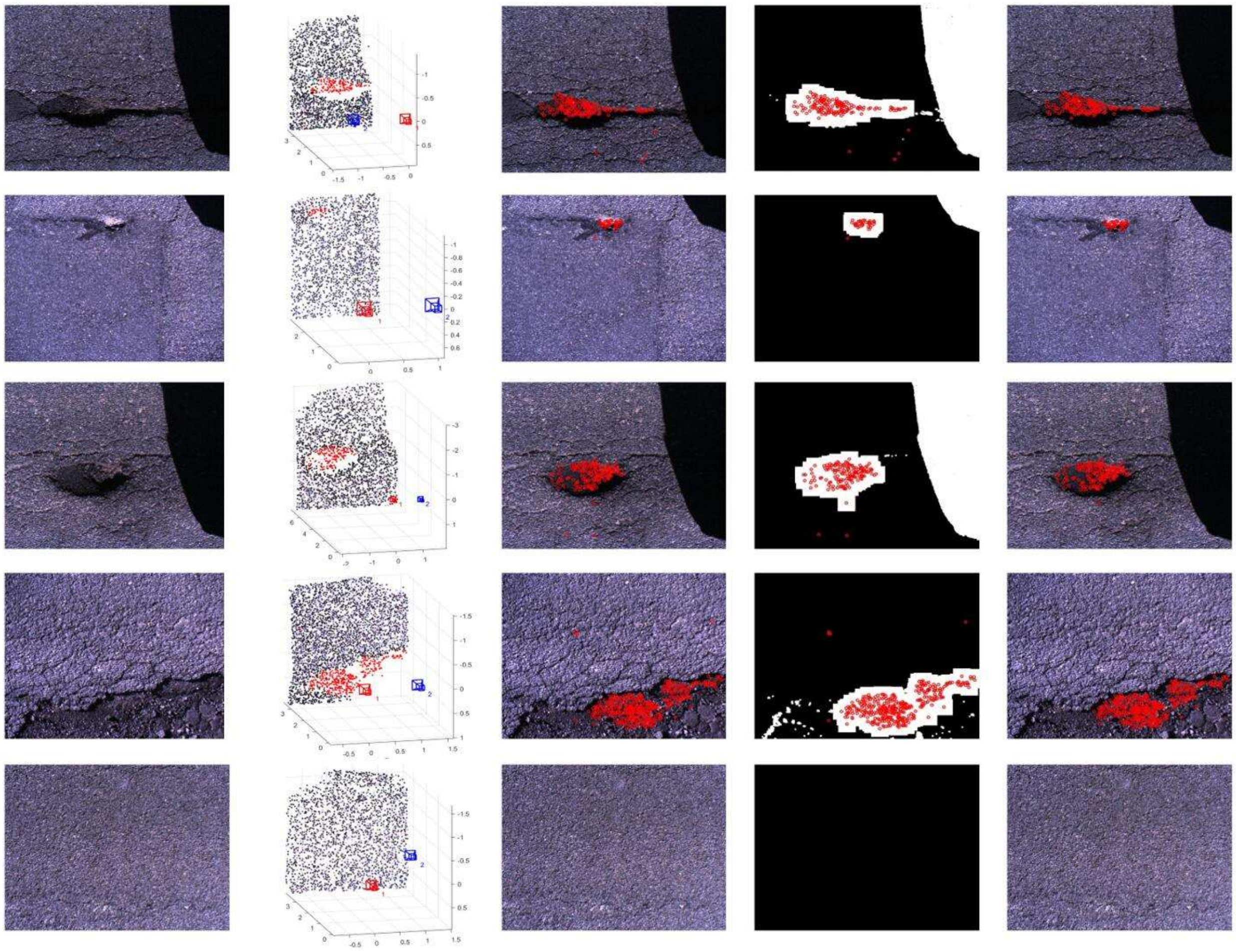
4.4 Chapter Summary

A technique of automatic road defects detection based on the 3D reconstruction for the near-planar road surface is presented in this chapter. 3D points with units are reconstructed for road surface by using the combination of mismatched points rejection and the metric 3D reconstruction. 3D profiles for both non-degenerate and degenerate road surface are reconstructed by the proposed technique. 3D road surface defects are detected automatically using the reconstructed 3D road surface points and their corresponding 2D image.

Two types of experiment were conducted to evaluate the proposed degenerate road surface 3D reconstruction and defects detection technique. In the simulation environment, the first experiment compared SfM and the proposed technique under different road unevenness represented by δ and the noise σ in images. Results showed that the changing of δ does not affect the reconstruction error ϵ using the proposed technique but increases ϵ dramatically for SfM when δ is close to 0 (planar surface). The second experiment compared SfM and the proposed technique under the different rotation angles $\theta_x, \theta_y, \theta_z$ for the camera. Results showed that by changing $theta_x, theta_y$, and $theta_z$ the error ϵ is less than 3 mm even at the worst case. The third experiment showed the change of camera to road distance δ_h almost does not change the ϵ when $0 < \delta_h < 20\ mm$. For experiments using real images, the first experiment showed the 3D reconstruction error ϵ using both SfM and the proposed technique for the reconstruction of a flat surface under laboratory environment. The results showed that the error for SfM is much higher than the proposed degenerate method, and the proposed technique have mean error within 1 mm and standard deviation within 1 mm for h from 900 mm to 1600 mm. Lastly 6300 real road surface images were captured by the presented system on both local road and highway road surface. The proposed technique increased the accuracy from 80% to 98.95% and precision from 34.34% to 94.33% for the road defects detection.

This chapter is focused on reconstruct 3D structure and automatically detect defects for road surface using physics-based technique. Although the automatic defects detection is usually accurate based on the physics model, there are mainly two drawbacks: First, the processing time of reconstructing the road surface and then detect defect is long; Second, if no/few points are reconstructed in the road defect region, it is difficult to detect the defects.

The next chapter will address these issues by incorporating a data-driven defects detection into the current framework.



(a) (b) (c) (d) (e)

Figure 4.7: (a) Original images of the road surface. (b) Reconstruction of 3D points from consecutive images with overlapping region. Red and blue simulated cameras represent two camera positions. (c) Detected defects (red circles) $\mathbf{X}_k^{rd}$ based only on 3D information. (d) The binary 2D image marked with detected defects. (e) Refined defects ${}^{\{c\}}\mathbf{X}_k^{rd}$ detection after combining 2D and 3D information together.

Chapter 5

Automatic Defects Detection and Characterization Using 3D Reconstruction-based Self-supervised Learning

This chapter presents a self-supervised learning technique for automatic road surface defects detection and characterization based on 3D reconstruction from one camera. While only unlabeled images are required as the input, the proposed self-supervised learning technique gets 3D information from a physics-based near-planar road 3D reconstruction which conducts self-labeling for the input images during training. For testing, the proposed technique combines the data-driven CNN prediction with the physics-driven 3D reconstruction information for a refined detection of road defects while characterizing the location, depth, and area of the defects.

This chapter is organized as follows. The next section shows the problem formulation for the geometrical 3D road surface defects detection using one camera. Section 5.2 presents the proposed road surface anomaly detection using self-supervised learning technique. Section 5.3 demonstrates the efficacy of the proposed technique through real-world experiment. Finally, conclusions are drawn at the last section.

5.1 Geometrical 3D Road Surface Reconstruction and Defects Detection Using One Camera

Figure 5.1 shows the 3D reconstruction of road surface by an ordinary RGB camera. The camera is facing downwards with its optical axis vertical towards the road surface. The problem is formulated as given the images $I = \{\{I_k, I_{k+1}\} | \forall k \in (0, ..., N)\}$, reconstructing the road surface points $\mathbf{X}^r = \{\mathbf{X}^r_k | \forall k \in (0, ..., N)\}$. For each reconstructed road surface section $\mathbf{X}^r_k$, $\mathbf{X}^{rd}_k \in \mathbf{X}^r_k$ represent road surface points located in the defect area. ${}^{i,j}\mathbf{X}^r_k = (X^r, Y^r, Z^r)^T$ is one point on the road surface which projects to image I_k as ${}^{i,j}\mathbf{x}^r_k = (i, j)^T$ in the image coordinate. The camera's postion at time step k is $\mathbf{X}^c_k$, while the camera to the planar road surface distance is h.

5.2 Self-supervised Learning on Road Surface Anomaly

5.2.1 Overview

Figure 5.2 illustrates the proposed self-supervised CNN structure for road surface defects detection. Unlike the supervised learning technique, self-supervised learning does not require the human labeling for input data. In the case of road surface images, each image captures the road surface only within several meters long. This makes human labeling tremendously difficult by considering processing tens of millions of road images. Thus the self-supervised learning technique is suitable for the road surface defects detection task. In the proposed self-supervised learning technique, only images $\{I_k | \forall k \in (0, ..., N)\}$ without human labeling are required as the input. First, the images are passed to the near-planar road surface 3D reconstruction module which reconstructs the 3D point cloud data $\mathbf{X}^r = \{\mathbf{X}^r_k | \forall k \in (0, ..., N)\}$

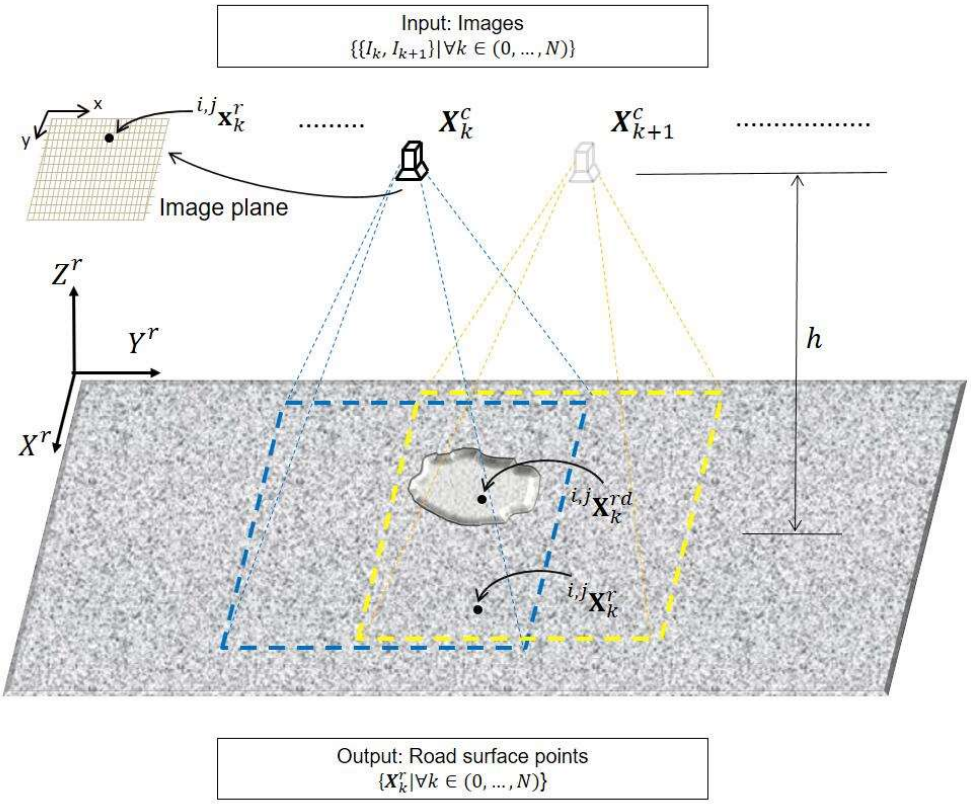


Figure 5.1: 3D reconstruction of road surface from one camera. I_k is the image captured by camera at time step k. $\mathbf{X}_k^r$ are reconstructed road surface points at time step k. $^{i,j}\mathbf{x}_k^r = (i, j)^T$ is the image point projected by the point $^{i,j}\mathbf{X}_k^r$. h is the height from the camera to the road surface. $\mathbf{X}_k^c$ represents the camera's position at time step k.

from $\{I_k | \forall k \in (0, ..., N)\}$. The road surface depth Z^r from $\mathbf{X}_r = (X^r, Y^r, Z^r)$ will then be used to determine the label l_k for image I_k as "defect" or "non-defect":

$$l_k = \begin{cases} 1, & I_k \ has \ defect \\ 0, & I_k \ has \ no \ defect \end{cases} \tag{5.1}$$

The self-labeled images are then passed to the convolutional neural network inspired by the AlexNet [80] and the output of the CNN has two probabilities corresponding to the "defect" and "non-defect" classes. The last module of the proposed technique determines whether the output probabilities are strong enough to conclude the road to be defect or non-defect category. If the belief is not strong, additional information from 3D reconstruction is included to provide a final decision on the road classifications.

The input to this network is $\{I_k | \forall k \in (0, ..., N)\}$. The input images have no labels for training process at this stage. First, the images are passed to the near-planar road surface 3D reconstruction module which calculates the 3D point cloud data to represent the road surface.

After previous 3D reconstruction and defects decision process, the images are self-labeled using their 3D geometrical information. This is one of the main difference between the proposed technique and the conventional CNN techniques for image data which require the labels for training. Also for conventional CNN techniques with RGB-D data, the input should be not only RGB image but also the depth (D) information either from Lidar or RGB-D camera. However, the proposed self-supervised learning structure only requires consecutive images as inputs. It labels the defect/non-defect image by the 3D point cloud data generated directly from image inputs. By doing this, there is no need for labeling the road surface images by human beings, which is important since the amount of the road

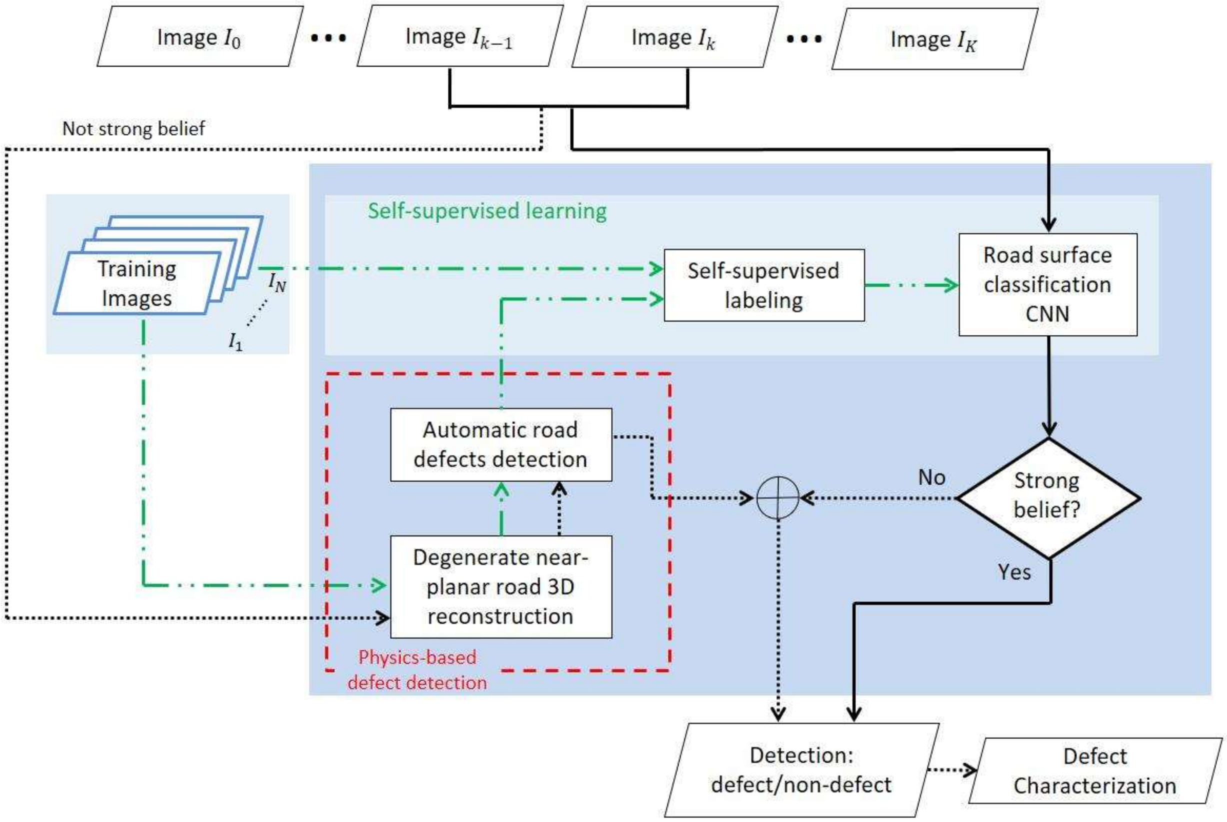


Figure 5.2: Self-supervised learning for road surface defects detection. The training process only needs unlabeled images as input. 3D depth information is obtained by a near-planar road 3D reconstruction process to conduct self-labeling of images. For testing, the image is passed to the network which gives prediction of defect or non-defect. The 3D depth information helps to refine the classification if the original predict is not strong enough.

surface images is too huge to inspect. Also, human beings judge the defects only based on the color information which causes subjectivity on defect road surface images labeling, while the proposed technique detects defect according to both the RGB image and the depth information.

5.2.2 Near-planar Road 3D Reconstruction and Self-labeling

Epipolar constraint describes the geometrical relationship between 3D object and its 2D image projection. Since road surface's near-planar characteristics, $\mathbf{x}_k^r$ and $\mathbf{x}_{k+1}^r$ satisfies Eq. 3.13. Thus given the epiploar line $\mathbf{l}_{k+1}$, epipolar constraint is illustrated as follows:

$$\begin{gathered}\mathbf{l}_{k+1} = \mathbf{e}_{k+1} \times \mathbf{x}_{k+1}^r = \mathbf{e}_{k+1} \times \mathbf{H}_k \mathbf{x}_k^r \\ \mathbf{l}_{k+1} = \mathbf{F}_k \mathbf{x}_k^r\end{gathered} \tag{5.2}$$

where $\mathbf{e}_{k+1}$ is the epipole which represents the projection of camera's center position at time step k onto I_{k+1}. As a result, the epipole $\mathbf{e}_{k+1}$ can also be expressed as:

$$\mathbf{e}_{k+1} = \mathbf{K}[\mathbf{R}_{k+1}, \mathbf{t}_{k+1}] \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{t}_{k+1} \tag{5.3}$$

where the homogeneous coordinate of the camera's center position is $(0, 0, 0, 1)^T$. $\mathbf{t}_{k+1} = \mathbf{X}_{k+1}^c - \mathbf{X}_k^c$ is the camera's translation between two time steps. From Eq. 3.11, it can be derived that

$$\mathbf{x}_{k+1}^r - \mathbf{x}_k^r = (\mathbf{P}_{k+1} - \mathbf{P}_k)\mathbf{X}_k^r = \mathbf{K}[(\mathbf{R}_{k+1} - \mathbf{I}) \mid (\mathbf{t}_{k+1} - \mathbf{0})]\mathbf{X}_k^r \tag{5.4}$$

which means if the orientation of the camera does not change much between time step k and time step $k + 1$, the rotation matrix $\mathbf{R}_{k+1}$ will be $\mathbf{R}_{k+1} \approx \mathbf{I}$ which leads to:

$$\mathbf{x}_{k+1} - \mathbf{x}_k = \mathbf{K}[(\mathbf{I} - \mathbf{I}) \mid (\mathbf{t}_{k+1} - \mathbf{0})]\mathbf{X}_k^r = \mathbf{K}[\mathbf{0}_{3\times 3} \mid \mathbf{t}_{k+1}]\mathbf{X}_k^r \tag{5.5}$$

Here $\mathbf{X}_k^r$ is abused to represent its homogeneous coordinate, which is $\mathbf{X}_k^r = (X_k^r, Y_k^r, Z_k^r, 1)$. Because of $\mathbf{0}_{3\times 3}$, Eq. 5.5 is evolved after substituting $\mathbf{X}_k^r$:

$$\mathbf{x}_{k+1}^r - \mathbf{x}_k^r = \mathbf{K}\mathbf{t}_{k+1} \tag{5.6}$$

Finally, combining Eq. 5.2, 5.3, and 5.6, the fundamental matrix $\mathbf{F}_k$ is derived to be unique as:

$$\mathbf{F}_k = [\mathbf{x}_{k+1}^r - \mathbf{x}_k^r]_\times \mathbf{H}_k \tag{5.7}$$

Thus correct 3D road surface $\mathbf{X}^r$ can be obtained by combining Eq. 5.7 and 3.12.

The advantage of getting 3D point cloud data from proposed 3D near-planar road reconstruction is that the point cloud data do not need to have another registration process onto the image. The point cloud data naturally have perfect match with image pixels since they are reconstructed from the image feature matching. The corresponding 3D points then give the image clue of depth differences within the image, which can find the defects area by examining $\mathbf{X}_k^r$. The defects area points $\mathbf{X}_k^{rd} = (X_k^{rd}, Y_k^{rd}, Z_k^{rd})^T$ have smaller Z_k^{rd} compare with Z_k^r which locate on the flat road surface. A normalization procedure is performed before labeling. $\mathbf{X}_k^r$ are normalized through the distance from camera to ground surface h, which results in normalized road surface points $\bar{\mathbf{X}}_k^r$:

$$\bar{\mathbf{X}}_k^r = \frac{\mathbf{X}_k^r}{h} \tag{5.8}$$

then a threshold α is used to decide the points $\mathbf{X}_k^{rd}$ which are located in the defects area:

$$\mathbf{X}_k^{rd} = \{\mathbf{X}_k^r | \forall \mathbf{X}_k^r < \alpha\} \tag{5.9}$$

The value of α is decided by the determination criteria of defects. If the defect is defined by its $Z_k^r < Z^d$, α is then defined as $\alpha = \frac{Z^d}{h}$. The self-labeling thus is done by examing depth Z_k^r for the road captured by I_k. If $\mathbf{X}_k^r$ have some road surface points where $Z_k^r < Z^d$, these points represents the road defect points $\mathbf{X}_k^{rd}$ and I_k is automatically self-labeled as $l_k = 1$. Otherwise, I_k represents non-defect road and is self-labeled as $l_k = 0$.

5.2.3 Network Structure and Classification using 3D Road Data

In Figure 5.2, the convolutional neural network has five convolutional layers and three fully connected layers. As the proposed technique is solving a classification problem with two classes $l_k = 1$ and $l_k = 0$, the last fully connected layer outputs 2 numbers as the scores for two classes. The score for class $l_k = 1$ and $l_k = 0$ are represented as s_1 and s_0. Besides convolutional layers (Conv), the network also has Rectified linear unit (Relu), Max pooling layer, Dropout layer, and Fully connected layer (FC) [80]. The eight layers are deployed as: (Conv1 + Relu) - (Max pooling + Conv2 + Relu) - (Max pooling + Conv3 + Relu) - (Conv4 + Relu) - (Conv5 + Relu) - (Max pooling + Dropout + FC6 + Relu) - (Dropout + FC7 + Relu) - (FC8 + Relu). The data size after each convolutional layer is shown on Figure as 5.2 as $H \times W \times C$ where H is the height, W is the width, and C is the channel of the data output after each layer. The scores s_1 and s_0 represent the unnormalized probability of each class. In order to get the normalized probability, the softmax function is used to get the new probability for each class as:

$$p_{l_k} = \frac{e^{s_{l_k}}}{\Sigma_{i=1}^{2} e^{s_i}} \tag{5.10}$$

where p_{l_k} is between 0 to 1. Therefore if one of the classes show strong belief (p_{l_k} close to 1) that I_k belongs to itself, the classification on the road for I_k can be determined. On the other hand, if p_{l_k} is far away from 1, the classification determination process needs the help from 3D reconstructed road information.

5.2.4 Defects Clustering and Estimation

Figure 5.3 shows the clustering and property estimation of road defects. After road defect points $\mathbf{X}_k^{rd}$ being detected, a recursive mean shift algorithm is used to cluster defects. At iteration 0, multiple squire windows are assigned to cover the whole road surface. $\mathbf{ms}_{0,i}$ represents the center position of the ith window at iteration 0. $w \times w$ is the size of the window. $\mathbf{X}_{k,j}^{rd}$ means the jth road surface point in the defect region. At iteration 1, the windows are shifting based on the their positions towards defect points $\mathbf{X}_k^{rd}$:

$$\mathbf{ms}_{1,i} = \frac{\sum_{j=1}^{Num(\mathbf{X}_k^{rd})} K(\mathbf{ms}_{0,i} - \mathbf{X}_{k,j}^{rd})\mathbf{X}_{k,j}^{rd}}{\sum_{j=1}^{Num(\mathbf{X}_k^{rd}} K(\mathbf{ms}_{0,i} - \mathbf{X}_{k,j}^{rd})} \tag{5.11}$$

where $Num(\mathbf{X}_k^{rd})$ is the number of points in $\mathbf{X}_k^{rd}$, and $K(\mathbf{x})$ is a flat kernel function for $\mathbf{x} = (x_1, x_2, ..., x_j)$ given by:

$$K(\mathbf{x}) = \begin{cases} 1, |x_j| \leq \frac{w}{2} \; \forall j \in \mathbb{R} \\ 0, else \end{cases} \tag{5.12}$$

Thus the windows are moving to $\mathbf{X}_k^{rd}$ after each iteration. The iteration will be terminated when the window positions are converged. As a result, the windows which are closer can be clustered using the same label. The number of different labels represents the number of defects on the road $\mathbf{X}_k^r$. The area, location, and depth of each defect can then be estimated from the clustered 3D points $\mathbf{X}_k^{rd}$.

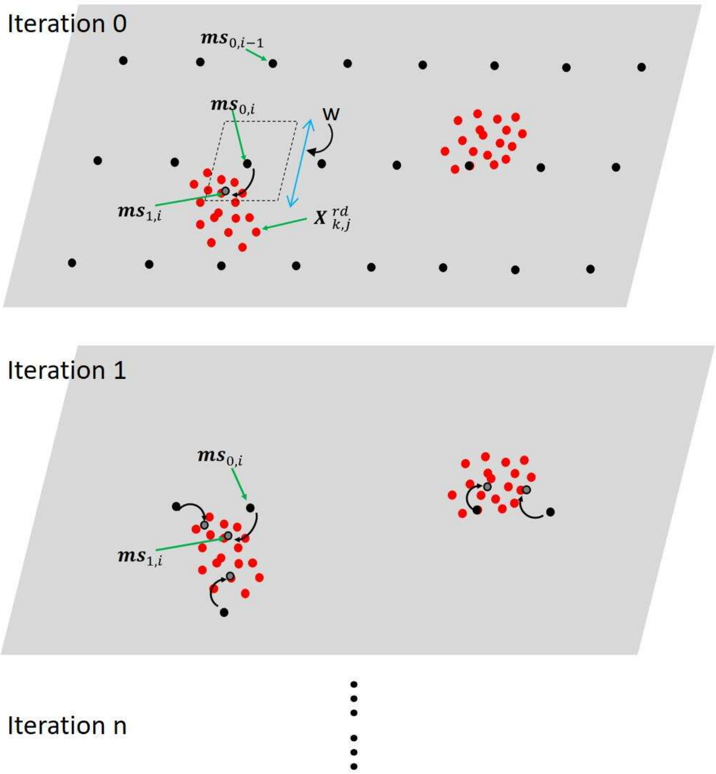


Figure 5.3: Defects clustering and estimation using mean shift algorithm.

5.3 Experimental Results

5.3.1 Data Collection

Table 5.1 shows all the conditions under which the road surface images are captured.

Table 5.1: Variables for road surface data collection

Variable	Name
Road category:	highway, urban, rural
Period of time:	morning, noon, afternoon
Weather:	sunny, partly cloudy, cloudy
Road surface material:	concrete, asphalt

5.3.2 Imbalanced Data Augmentation

Since the amount of road defects are far less than the non-defect road areas, the road surface images collected are highly imbalanced. Among nearly one million images collected from the road, over 99 % of the images are non-defect. The imbalanced data can cause strong bias to the traning of neural network. In the case of road surface images, if the amount of non-defect images are dominant within the training data, the whole network will tend to predict the road as non-defect at the test time since during training the model learns that by predicting the images as non-defect it is correct over 99 % of the time. In addition, less defect road images means that the model learns less about the road defects during training. Thus augmenting the number of defect road images will help to improve the prediction of the model.

Figure 5.4 shows the augmentation of the defect road images. After an image is self-labeled as the defect road surface, the road defect is augmented to several images by translating, rescaling, rotating, and flipping. For the translation and rescaling process, the original image is translated to several images so that each generated image contains the defect area but at different locations of the image. The newly generated images usually allocate the defect to the left, right, up, down, and center of the image. For the image rotation, new images are obtained by rotating the original image at different angles such as 90°, 180°, 270° and so on. The flipping increases the defect images by the flip of original image along their

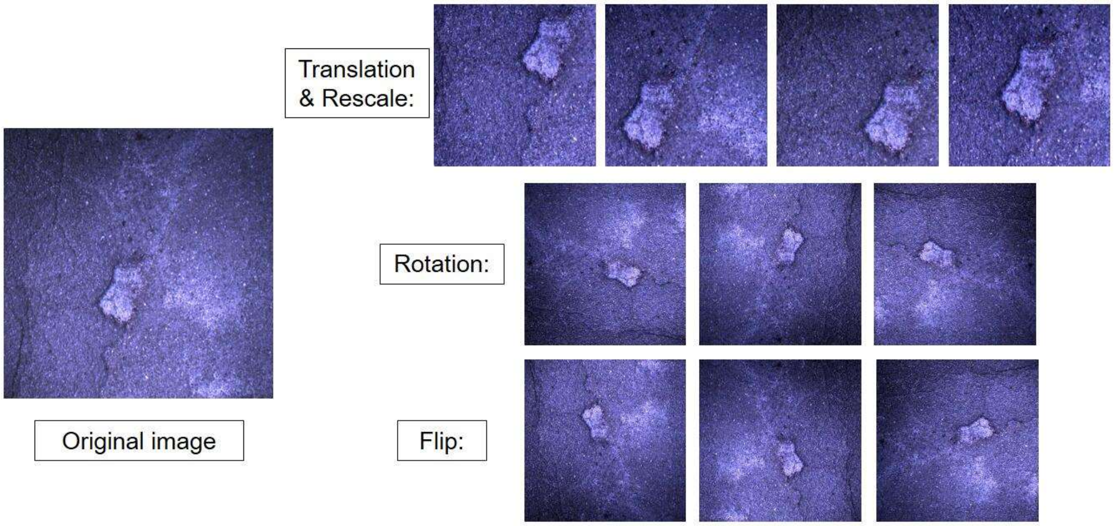


Figure 5.4: Defect image data augmentation. The imbalanced data between defect and non-defect images are solved by data augmenting through translation, rescaling, rotation, and flipping of defect images.

horizontal or vertical axes.

5.3.3 Network Performance Evaluation

The proposed self-supervised road defects learning structure was first trained by the collected road surface images. Table 5.1 lists all the conditions under which the images are collected. Among the captured images, 3155 non-defect images are randomly chosen as the training data. Also, after the image augmentation, 2700 defect images become the rest of the training data which reflects the road defect. The weights of the network are initialized according to Krizhevsky et al. [80]. During the training, the selected 5855 images are divided so that 80 % become the training data and 20 % are the validation data. The hyperparameters such as learning rate, learning decay, number of epochs, and batch size are fine-tuned to get the optimal result. The best combination of hyperparameters are: 0.00005 learning rate, 0.95 learning decay rate, 800 epochs, and a batch size of 64.

Table 5.2 demonstrates the performance of the proposed technique comparing with the traditional image only CNN defects detection using the testing data. The testing data are composed by 220 non-defect images and 204 defect images. The comparisons in precision, accuracy, and recall rates are conducted among Alexnet [80], Vgg-16 [81], Densenet [86], and the proposed (3D + Backbone: Alexnet). It can be found that the proposed Image + 3D learning network improves the performance of road defects detection in accuracy and recall comparing with the other three CNN which detects defects based on only the image information.

Table 5.2: Performance for testing data on traditional image only technique and the proposed technique.

	Accuracy	Precision	Recall
Alexnet	79.0%	99.1%	56.4%
Vgg-16	78.8%	96.7%	57.8%
Densenet	89.9%	94.4%	83.8%
Proposed (3D + Backbone: Alexnet)	95.9%	95.2%	87.7%

Figure 5.5 shows the average processing time for each image on the road surface defects classification. It can be seen that for the three CNN models, the time spent on detecting defects for one image is less than 10^{-1} second, while the time for the proposed data-driven detection is around 10^{0} second and around 10^{1} for the proposed physics-based detection. Thus, the average processing time reduced to 1/10 for the proposed data-driven technique.

Figure 5.6 shows the quantitative prediction results for several representative images. From top to bottom, each row in the figure means: the original image, ground truth for the image, class prediction by image only (without 3D information), decision making on whether the belief strong enough, 3D depth image, and the final class prediction using the image and 3D depth information. Column 1 and 2 show two cases where the image only prediction matches the ground truth. Furthermore, as the belief of predictions are strong

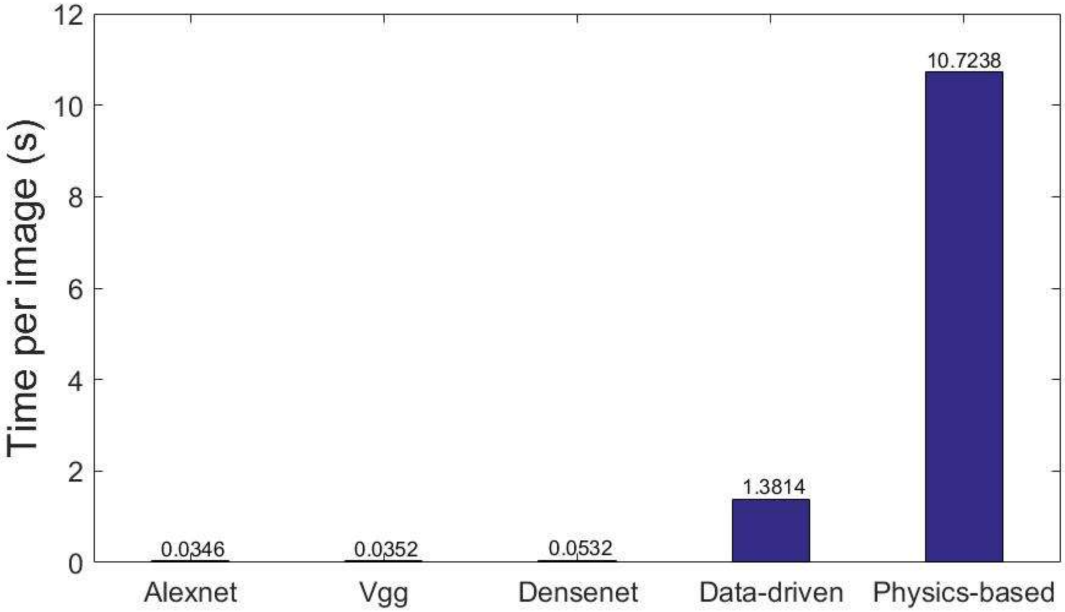
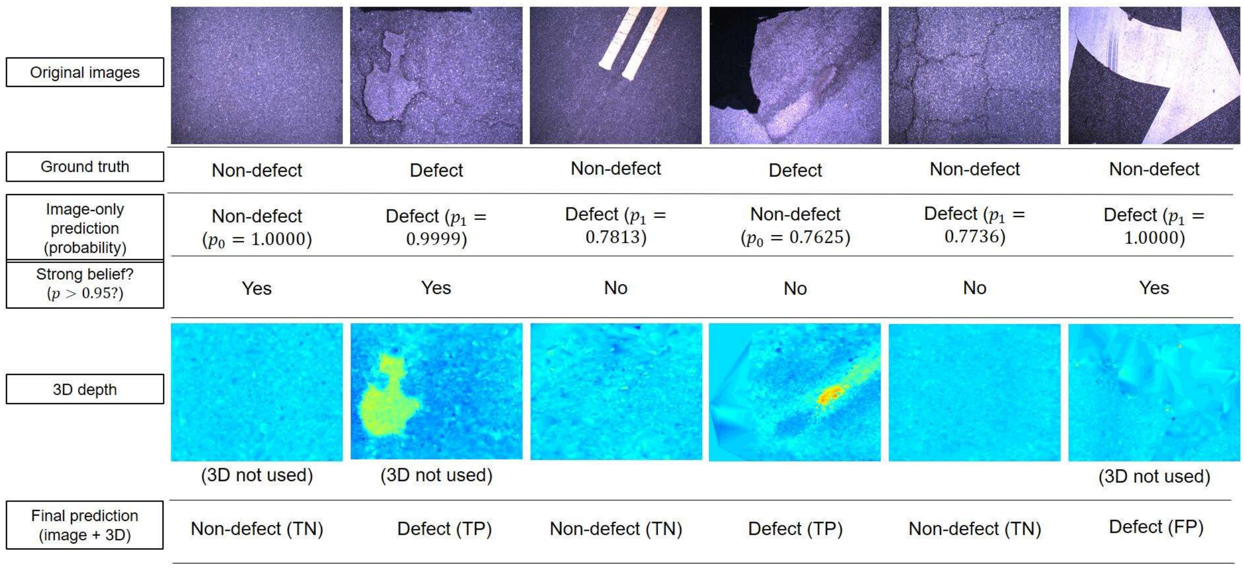


Figure 5.5: Average time spent on classifying road surface image into defect or non-defect.

enough (empirically set by $p_{l_k} > 0.95$), these two cases make the final prediction without using the 3D depth. Column 3, 4, and 5 illustrate cases that first have wrong predictions but finally make correct predictions after combining the 3D information. The original image of column 3 contains the double solid lines which predict the image to be ”Defect” with a weak belief ($p_1 = 0.7813$). With this not strong belief, a near-planar road 3D reconstruction is conducted to help the prediction. Since the depth image strongly infer that there is no defect in the input image, the final decision predicts the image as ”Non-defect” which is consistent with the ground truth. Column 4 has a small pothole in the image which is difficult to recognize. The image only classification predicts it as a ”Non-defect” road. After following the same process of using 3D information, the final prediction is corrected to predict it as ”Defect” as the 3D data for pothole is obvious in the depth image. Column 5 has some cracks in the image and is wrongly predicted as ”Defect” at the beginning. It is also corrected to be the same as ground truth after using 3D data. Column 6 is different with other cases since



Original images						
Ground truth	Non-defect	Defect	Non-defect	Defect	Non-defect	Non-defect
Image-only prediction (probability)	Non-defect ($p_0 = 1.0000$)	Defect ($p_1 = 0.9999$)	Defect ($p_1 = 0.7813$)	Non-defect ($p_0 = 0.7625$)	Defect ($p_1 = 0.7736$)	Defect ($p_1 = 1.0000$)
Strong belief? ($p > 0.95$?)	Yes	Yes	No	No	No	Yes
3D depth	(3D not used)	(3D not used)				(3D not used)
Final prediction (image + 3D)	Non-defect (TN)	Defect (TP)	Non-defect (TN)	Defect (TP)	Non-defect (TN)	Defect (FP)

Figure 5.6: Quantitative results for the proposed technique in several representative situations. Column 1 and 2 data are predicted correctly at the image-only prediction stage. Column 3, 4, and 5 data have wrong predict at the image-only prediction stage, but are corrected at the final prediction by 3D road depth. Column 6 shows a false positive case because its wrong belief is so strong that the network directly output the wrong prediction as the final classification.

it reflects a False Positive prediction at the end. The reason for this error is that the image only prediction gives strong belief of this road marking image to be a "defect" road with $p_1 = 1.0000$. The proposed technique then believes this strong prediction and does not use 3D information to facilitate the prediction. Thus the prediction is kept as "Defect" which results in a false positive case.

Figure 5.7 demonstrates the clustering of road defects. The depth of reconstructed road $\mathbf{X}^r_{0:K}$ is shown as a colormap. It can be discovered that there exists multiple defects in a section of the road. Then the proposed mean shift defects clustering algorithm initialize windows to cover the whole road. The example is converged after 5 iterations by then the defects are clustered to be four groups. Thus the size, position, and depth can be easily determined using clustered 3D points $\mathbf{X}^{rd}_k$.

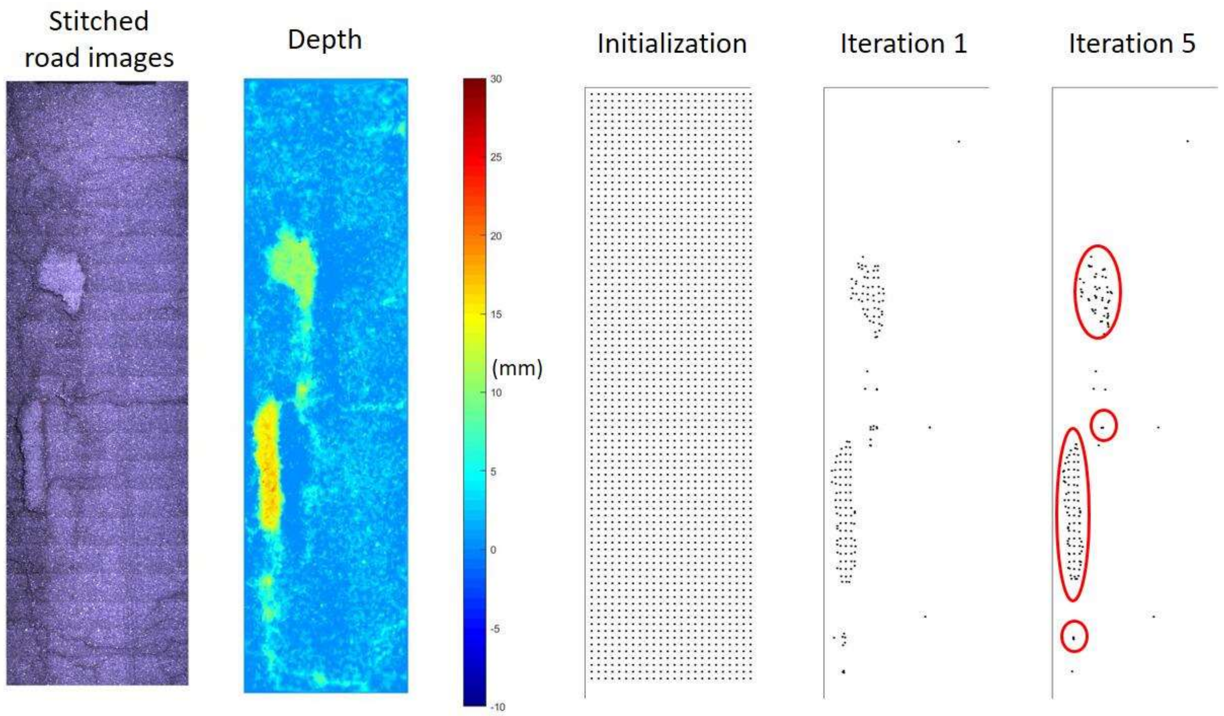


Figure 5.7: Road defects clustering and estimation. From left to right: A section of road surface images stitched together; The depth colormap of the corresponding road; Initialization (iteration 0) of proposed mean shift windows; Windows' position after iteration 1; Windows' position after iteration 5. Number of defects are recognized by the clustering. Defects' size and depth are estimated based on clustering.

5.4 Chapter Summary

This chapter has presented a self-supervised road surface defects detection learning technique based on 3D reconstruction from one camera. Relying on the near-planar road 3D reconstruction process, the proposed technique labels the input training images without the human labeling. In addition to self-supervised learning from 3D road surface, the geometrical depth information is also combined with the CNN classification output to result in a more accurate prediction. Also, the clustering for detected road defects further estimates the quantity, depth, location, and size of the defects.

Over one million road surface images are collected with the road image capturing system. The images are captured at various conditions such as different road categories, period of time, weather, and road surface materials. The experimental evaluation of the proposed technique is analyzed based on the collected road surface images. Due to the imbalanced nature of defect and non-defect road surface images, the defect image data augmentation is conducted by using translation, rescaling, rotation, and flipping of the defect images. The training of proposed network is based on road images that are random-selected and balanced. The testing data demonstrate the increase for all the Accuracy, Precision, and Recall to be 95.9%, 95.2%, and 87.7%. Several case studies show that the proposed technique effectively improve the prediction accuracy of the convolutional neural network by using Image + 3D for images containing irregular potholes, cracks, road lines and other markings. One of the future works is fine-tuning the hyperparameters and layers layout of the CNN for more accurate prediction results. Another future work is that by using a camera looking forward at road surfaces in front of the vehicle, the revised deep learning network have the potential to achieve online road defects detection in front of the car.

Chapter 6

Road Surface Monitoring System

This chapter presents a system design and the data visualization for road surface monitoring. The first section formulates the problem and lists the design specifications for the road surface monitoring system. Section 6.2 shows the system design for both the hardware installation and the electronics/sensors integration. Section 6.3 illustrates the proposed simultaneous road surface 2D and 3D mapping by fusing the sensor data from the designed road surface monitoring system. Chapter 6.4 demonstrates the performance for the proposed 2D and 3D mapping by using the designed system.

6.1 Problem Formulation

Figure 6.1 is a diagram to show the road surface monitoring problem. The proposed road surface monitoring mainly focuses on two tasks: capturing the road surface images, which are represented by $I_0, I_1, ..., I_K$, and reconstruct 3D road surface $\mathbf{X}_0^r, \mathbf{X}_1^r, ..., \mathbf{X}_K^r$ in the global coordinate $\{G\}$. Furthermore, the 3D defects such as potholes are automatically detected through the 3D road surface, and other road surface information such as cracks and road surface markings are monitored through the 2D images. Since the road surface monitoring tasks are usually performed on the public road, the system mounted on a vehicle requires to drive at common speed from 20 km/h at local road up to 100 km/h at highway. Thus, the road monitoring system needs to capture clear road surface image and reconstruct 3D road

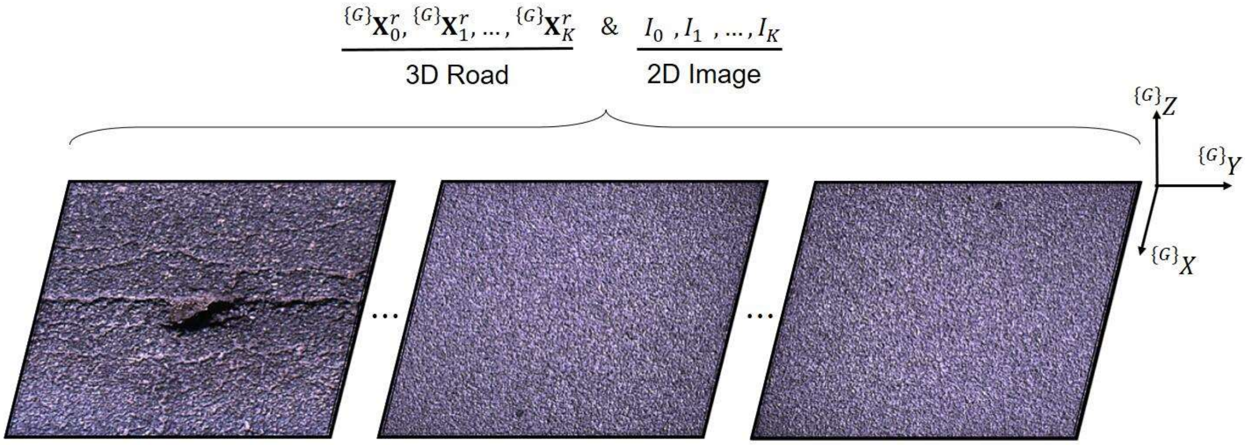


Figure 6.1: Road surface monitoring problem.

surface while drives at different public road speed.

6.2 Road Surface Monitoring System

Figure 6.2 shows the proposed road surface monitoring system. Only one camera is necessary to capture and reconstruct road surface, but two cameras are used individually to expand the system's field of view to around the lane width. The GPS records the global position of the vehicle while the system is capturing the road surface images. Two high-illumination LED lights increase the brightness of the image at low ambient light condition. The odometry data from the OBD-II port is read by the system, while a FPGA is programmed to synchronize the OBD-II odometry data with the camera and the GPS by sending the synchronized control signals. The maximum frame rate of the camera is 100 Hz and the shutter speed is 0.04 ms. Since the camera's frame rate and shutter speed are both high, the camera captures road surface images with no motion blur and more than 50% overlapping region between two consecutive images at up to 100 km/h driving speed. Therefore, the proposed design of the road surface monitoring system captures and reconstructs the road surface at various

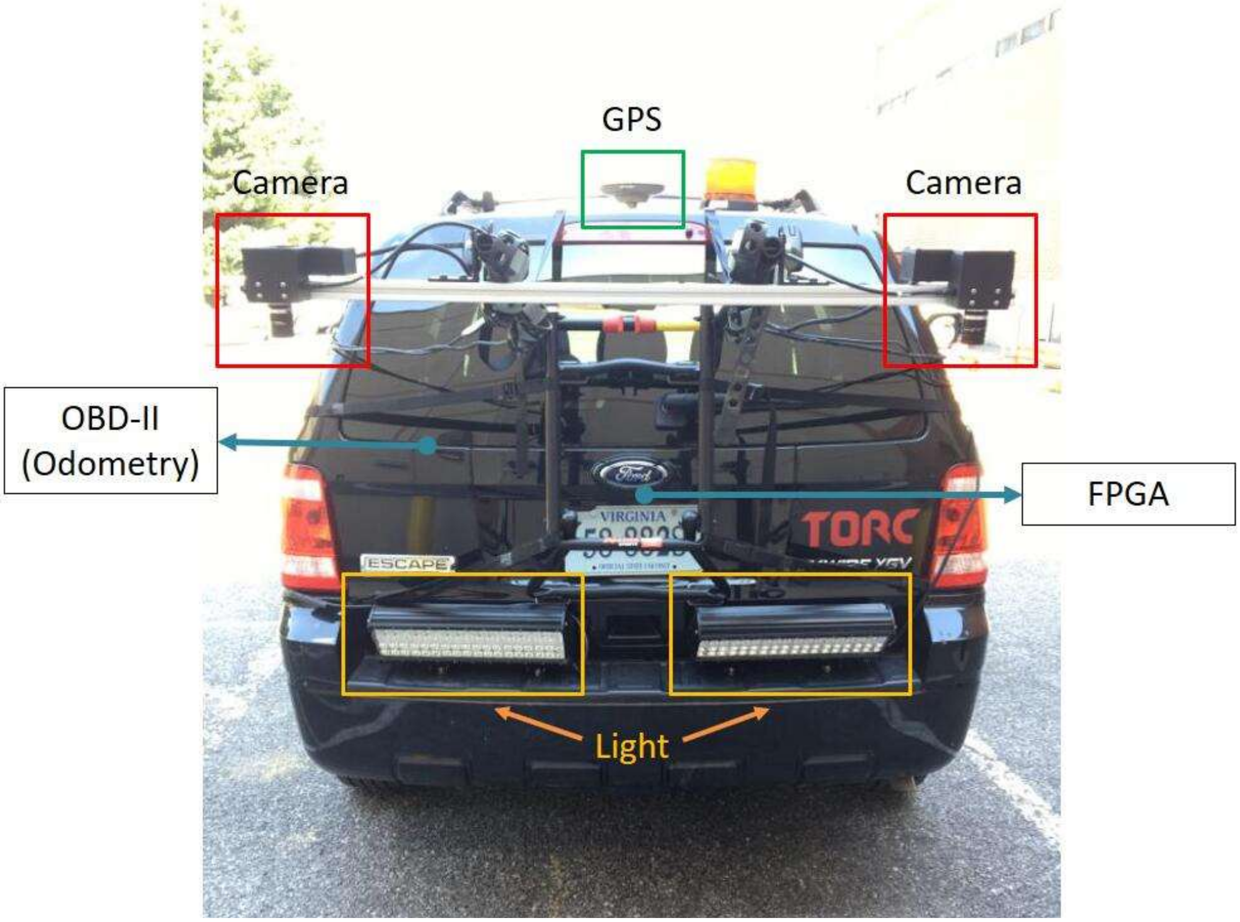
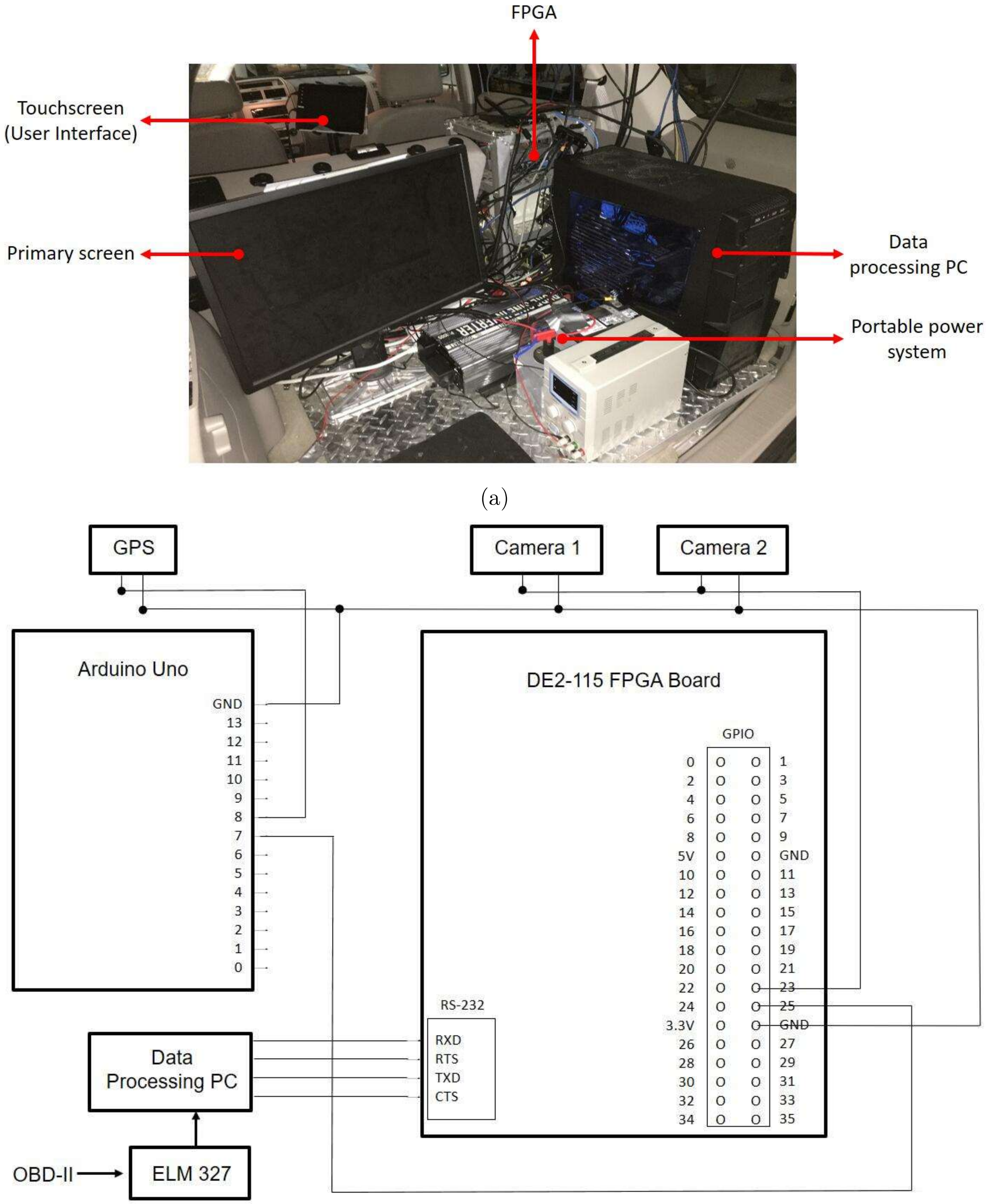


Figure 6.2: Road surface monitoring system for various driving speed.

speeds up to 100 km/h which is the highway speed.

Figure 6.3 shows the equipment and control circuit of the proposed road surface monitoring system. Figure 6.3a illustrates the control units, which are located at the trunk of the car, for the road surface monitoring system. The Portable power system is composed of a 12 V DC battery and a DC-AC converter, which provides power to the data processing PC, two screens, and the FPGA. The data processing PC reads the odometry data from the OBD-II port and convert it to data which adaptively change the frame rate and frequency for the camera and GPS. Also, the data processing PC saves all the images and GPS data. The FPGA gets the output from the data processing PC and sends the synchronized con-



(a)

(b)

Figure 6.3: Design and control circuit for road surface monitoring system.

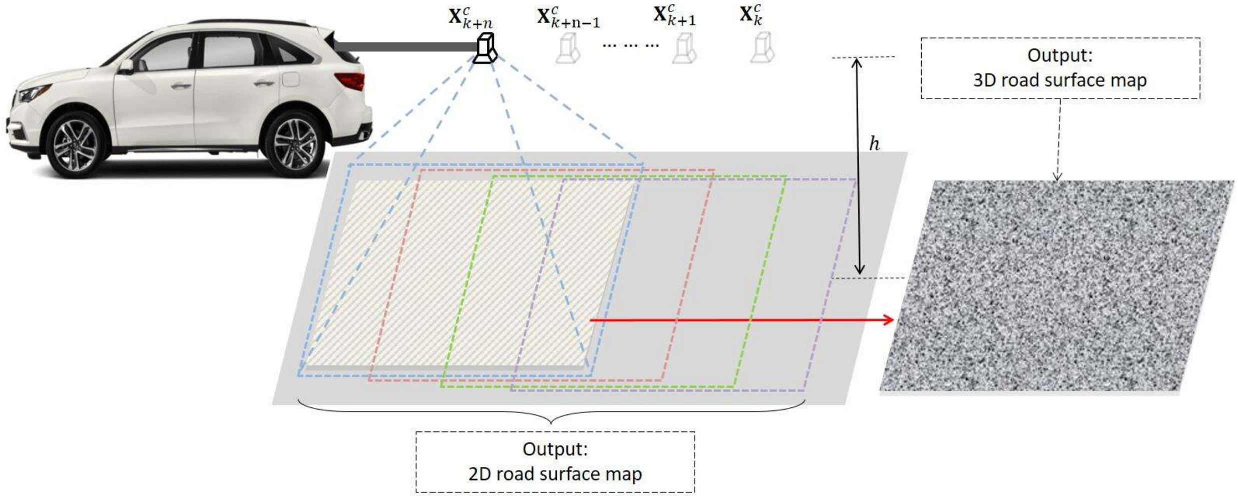


Figure 6.4: Simultaneous road surface 2D and 3D mapping.

trolling signals to both the camera and the GPS. The primary screen is used mainly in the developing and debugging period while the touchscreen acts as the main user interface to operate the system.

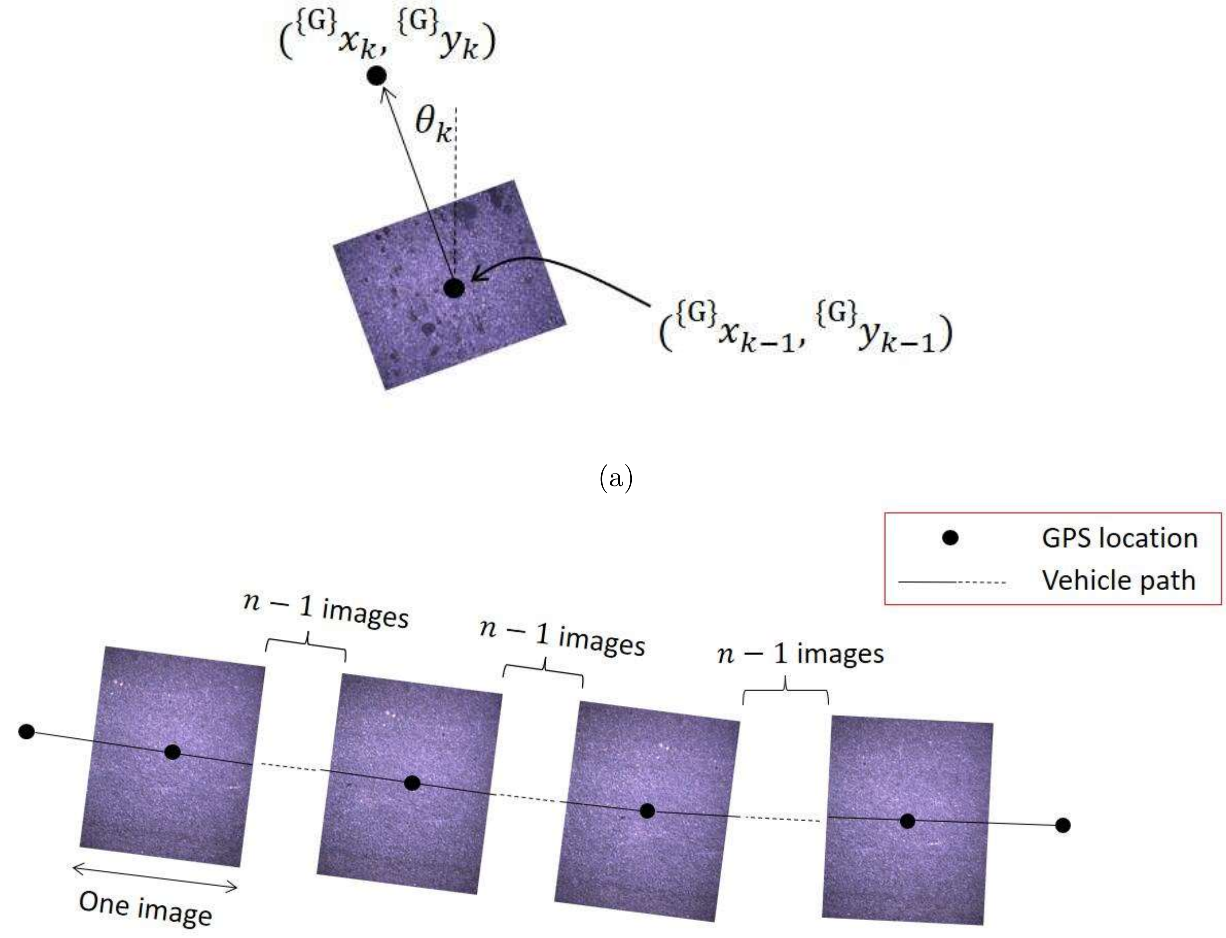
Figure 6.3b is the schematic diagram of the control circuit. First, the ELM 327 OBD-II reader is used to obtain the odometry data. ELM 327 converts the serial bus signal to USB communication protocol which the data processing PC can read. A program is developed on the data processing PC to understand the odometry information sent by ELM 327 through USB protocol. The program handles both the ISO 15765-4 11-bit and 29-bit CAN protocol, which are used by most of the cars nowadays. Then, the data processing PC sends the current vehicle speed through a serial signal to the FPGA. The FPGA receives the speed signals from the RS-232 port, converts them to controlling signal, and sens this synchronized signal to change the frame rate of the camera and GPS.

6.3 Simultaneous Road Surface 2D and 3D Mapping

Figure 6.4 shows the simultaneous 2D and 3D road surface mapping from the proposed road surface monitoring system. As the road surface monitoring system drives from time step k to $k+n-1$, it captures n road surface images which have overlapping between every neighboring two images. Therefore, from the previous chapters, a continuous 3D reconstruction for the road surface between every two images builds the 3D road surface map between time step k to $k+n-1$. Meanwhile, the 2D road surface map is obtained using the images captured between time step k and $k+n-1$.

In order to get the 2D and 3D road surface maps, several problems need to be solved based on the designed road surface system. First, the maximum frequency of the GPS is usually less than the image frame rate, thus the synchronization issue of the GPS data with images need to be addressed. Second, since road surface images are discretely captured, it is required that a technique can determine the pose and show the continuous road surface to better reflect the real road surface condition for a long section. Third, the obtained 2D and 3D map needs to be visualized properly for the road surface monitoring purpose.

Figure 6.5 shows the synchronization of the GPS data to images and the determination of road surface image pose. Figure 6.5a illustrates a road surface image captured at time step $k-1$. When the road surface monitoring system recorded this image, the GPS on the system also obtain the global position in ${}^{\{G\}}X$ and ${}^{\{G\}}Y$ direction as $({}^{\{G\}}x_{k-1}, {}^{\{G\}}y_{k-1})$. Besides the position of the road surface image, the pose also contains the orientation θ_k of the current image. However, all the sensors on the proposed system do not measure θ_k for the image. Therefore, this pose determination problem is solved by using the vehicle's motion information. When the vehicle moves to the next position $({}^{\{G\}}x_k, {}^{\{G\}}y_k)$ at time step k, the vehicle's moving direction is determined the vector $({}^{\{G\}}x_k - {}^{\{G\}}x_{k-1}, {}^{\{G\}}y_k - {}^{\{G\}}y_{k-1})$.



(a)

(b)

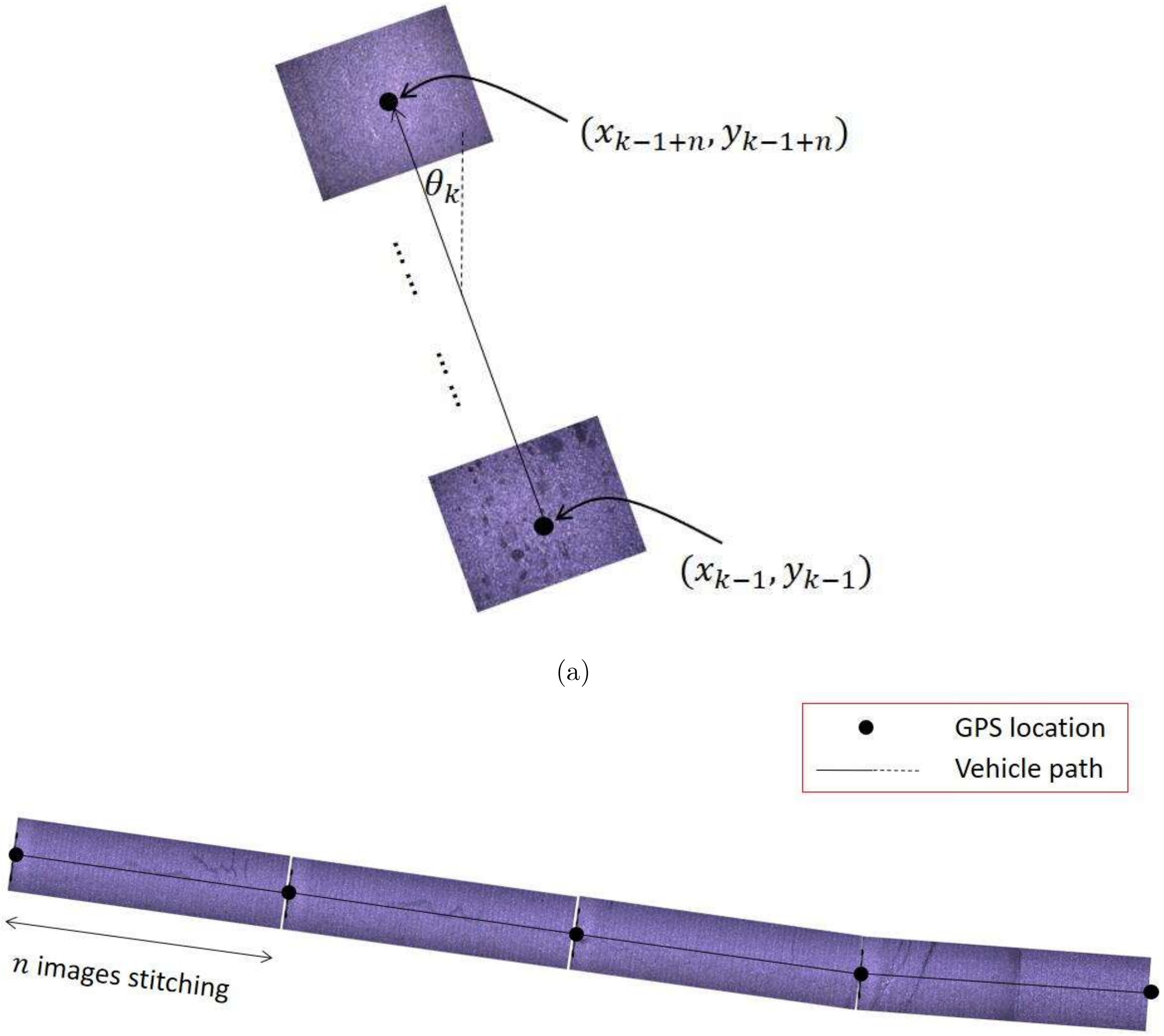
Figure 6.5: High frame rate images pose determination and synchronization to the GPS data.

Since the time between time step $k-1$ and k is short and the camera is rigidly attached to the vehicle, the vehicle's moving direction is parallel to the image's y direction. Thus the orientation of the road surface image in global coordinate is:

$$\theta_k = \arctan(-\frac{^{\{G\}}x_k - {^{\{G\}}x_{k-1}}}{^{\{G\}}y_k - {^{\{G\}}y_{k-1}}}) \tag{6.1}$$

Figure 6.5b then shows the continuous GPS data and image registration. As the frequency of the GPS is usually less than the image frame rate, shown in the Figure 6.5b, the GPS-corresponding images will have some unregistered $n-1$ images in between. The pose of these $n-1$ images are thus determined from the neighboring images.

In order to solve the unregistered images between two neighboring GPS recordings, Figure 6.6 demonstrates the proposed solution for determining the pose of images which miss GPS data. Figure 6.6a shows the situation when the GPS frequency is less than the camera frame rate. In this case, the two neighboring GPS data are recorded at $({^{\{G\}}x_{k-1}}, {^{\{G\}}y_{k-1}})$ and $({^{\{G\}}x_{k-1+n}}, {^{\{G\}}y_{k-1+n}})$ respectively. There are n images in between which are not registered with the GPS. If using interpolation to determine missing poses for unregistered images, the pose of these images will not be accurate. However, as the feature matching between two neighboring image are already obtained for the 3D reconstruction process, the proposed technique uses matched features to stitch road surface images especially for those which are unregistered with GPS data. Figure 6.6b illustrates the proposed road surface mapping. Image stitching is performed for the n images which only the first image has corresponding GPS data. Therefore, because image stitching uses matched image features to posit the images one by one, it is more accurate than pure interpolating image poses. Also, by stitching n images, the road surface map is continuously displaying a period of road, which is a more straight-forward visualization for the road surface monitoring.



(a)

(b)

Figure 6.6: road surface image stitching for unregistered GPS-missing images.

While getting the 2D road surface map, the road surfaces are reconstructed by the proposed 3D reconstruction technique. As the 3D reconstruction also relies on the features matched between two neighboring images, the transformation calculated in the 2D road surface stitching is the same for 3D reconstructed points transformation. By stitching the 3D reconstructed points from the same n images, the proposed technique and system can reconstruct the 2D image and 3D road surface simultaneously.

6.4 Experimental Validation

The designed road surface monitoring system and the proposed simultaneous 2D and 3D road surface mapping are validated in the real world environment. The experimental tests are conducted at Blacksburg Virginia and the nearby area. The road surface images are captured at different weather conditions such as sunny, cloudy, and partly cloudy. The road surface types include concrete and asphalt. The road measured contains rural road, urban local road, and highway, which requires the system vehicle to drive from 30 km/h to 100 km/h.

Figure 6.7 shows the 2D road surface mapping on Google Map and the GPS-image synchronization using the proposed system and techniques. The figure on top displays some road sections measured by the road surface monitoring. The green line shows the GPS data to draw the path of the vehicle. Each red marker represents a section of 2D road surface mapping synchronized with the GPS position. The bottom figure then is an enlarged view of the section marked by a blue rectangular box. The yellow rectangular box locates a section of 2D road surface mapping stitched by several images. It can be seen that the GPS data is well synchronized with the road surface and the road surface stitching is correct for the current location.

Figure 6.8 shows the test for the resolution of the collected GPS data. The left figure demonstrates a test at the local road. The green line represents the vehicle's path based on the GPS data. The purpose of the experiment is to show the qualitative result of the GPS path. The tested road has four lanes, and the road surface monitoring system drives on each lane one-by-one. The right figure illustrates the zoomed-in view of the yellow rectangular box area. It can be seen that the four lanes are clearly separated by the GPS data which means the proposed system can measure each lane of the road surface.

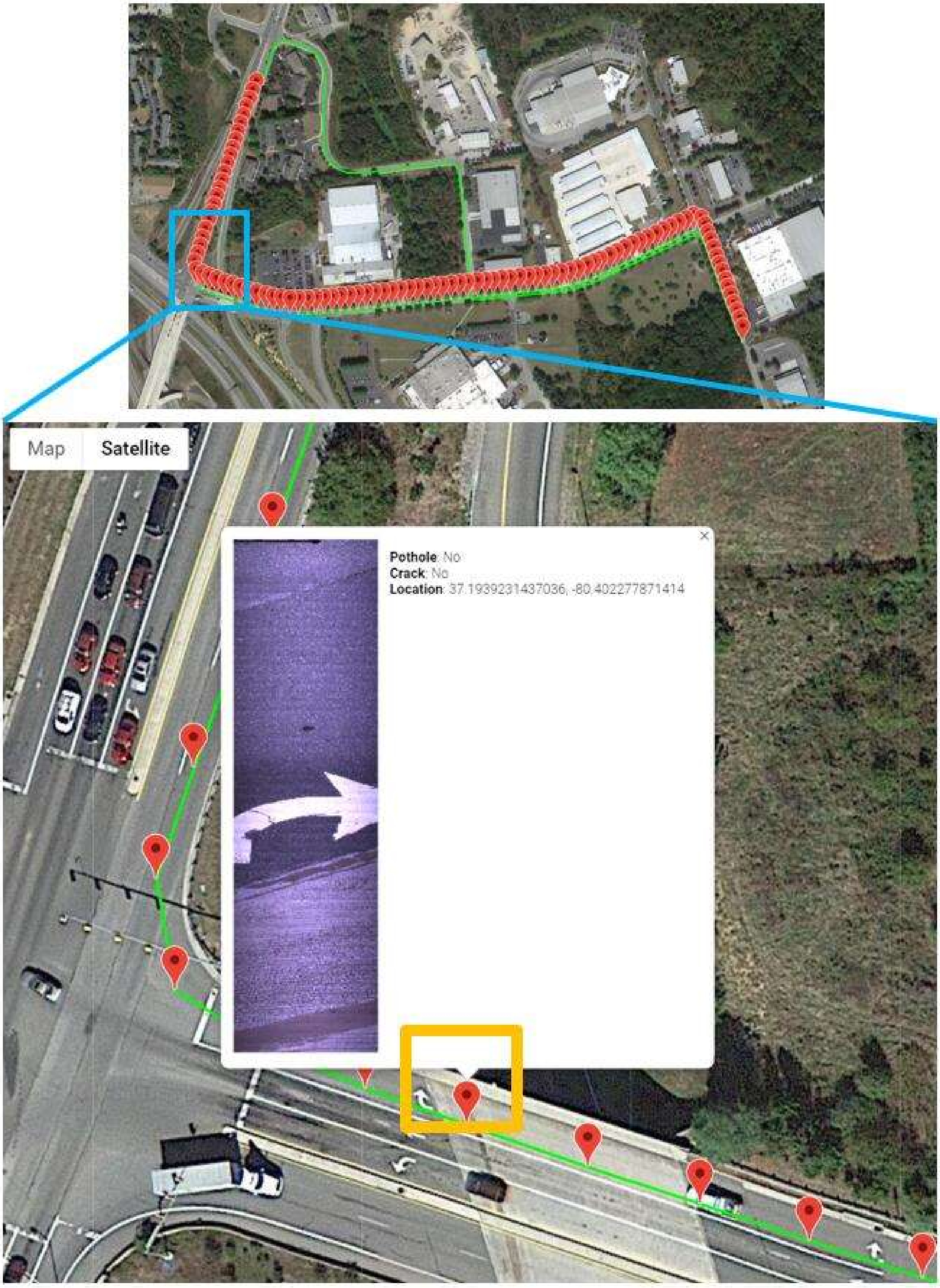
Figure 6.9 demonstrates the visualization of the proposed simultaneous 2D and 3D road surface mapping on Google Map. Figure 6.9a shows the pasting of the 2D stitched road surface map and Figure 6.9b visualizes the 3D road surface map. It can be seen that the position and orientation for both the 2D and 3D map are aligned correctly to the Google Map. This visualization of road surface map on online maps such as Google Map provides an intuitive and efficient way for the road surface monitoring task.

Figure 6.10 demonstrates the automatic long-distance 2D and 3D road surface mapping. The blue line indicates a long-distance travel of the road surface monitoring system in the urban environment. Figure 6.10 visualizes part of the long-distance 3D road surface mapping from Figure 6.10. The result shows that the proposed simultaneous 2D and 3D road surface mapping technique can automatically measure the road surface by using the designed road surface monitoring system. The system helps the automation of the traditional surveying work by human being and efficiently provides richer information on both the 2D and 3D road surface information for long-distance road measuring.

6.5 Chapter Summary

This chapter has presented a road surface monitoring system design and the simultaneous 2D amd 3D road mapping technique. The designed road surface monitoring system integrates and synchronizes FPGA, camera, OBD-II, GPS, and onboard PC together for both the sensor hardware and software. Different sensor data are then fused to generate 2D road surface image map and 3D road surface map simultaneously.

The overall performance of this designed road surface monitoring system is evaluated by the experiment. The data collecting tests are done in various speed limits such as highway, urban road, and rural road. Also, the data are collected on different weathers which are sunny day, cloudy day, and partly cloudy day. The first and second experiment are designed to evaluate the synchronization of GPS data with images and the resolution of the GPS data. Experiments show that the road surface images stitched are well synchronized with GPS data geographically and the resolution of GPS data clearly separates lanes of the road. The second experiment demonstrates that this system can simultaneously make 2D and 3D map for the road surface. The last experiment is conducted for a long-distance road surface measuring. The results show that the proposed technique automatically reconstruct the 2D and 3D map for long-distance measurement and efficiently visualizes the maps on online map server such as Google Map.



Pothole: No
Crack: No
Location: 37.1939231437036, -80.402277871414

Figure 6.7: GPS-image synchronization and 2D road surface mapping.

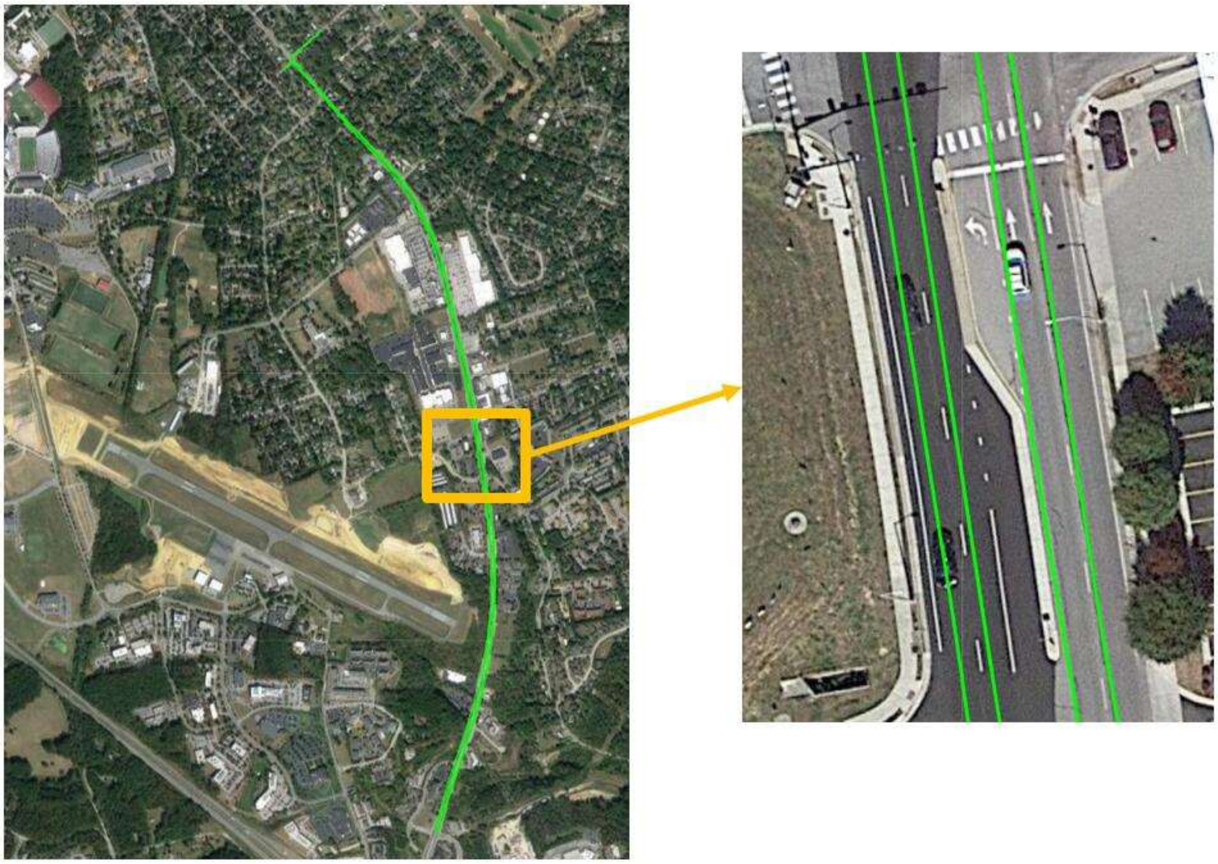
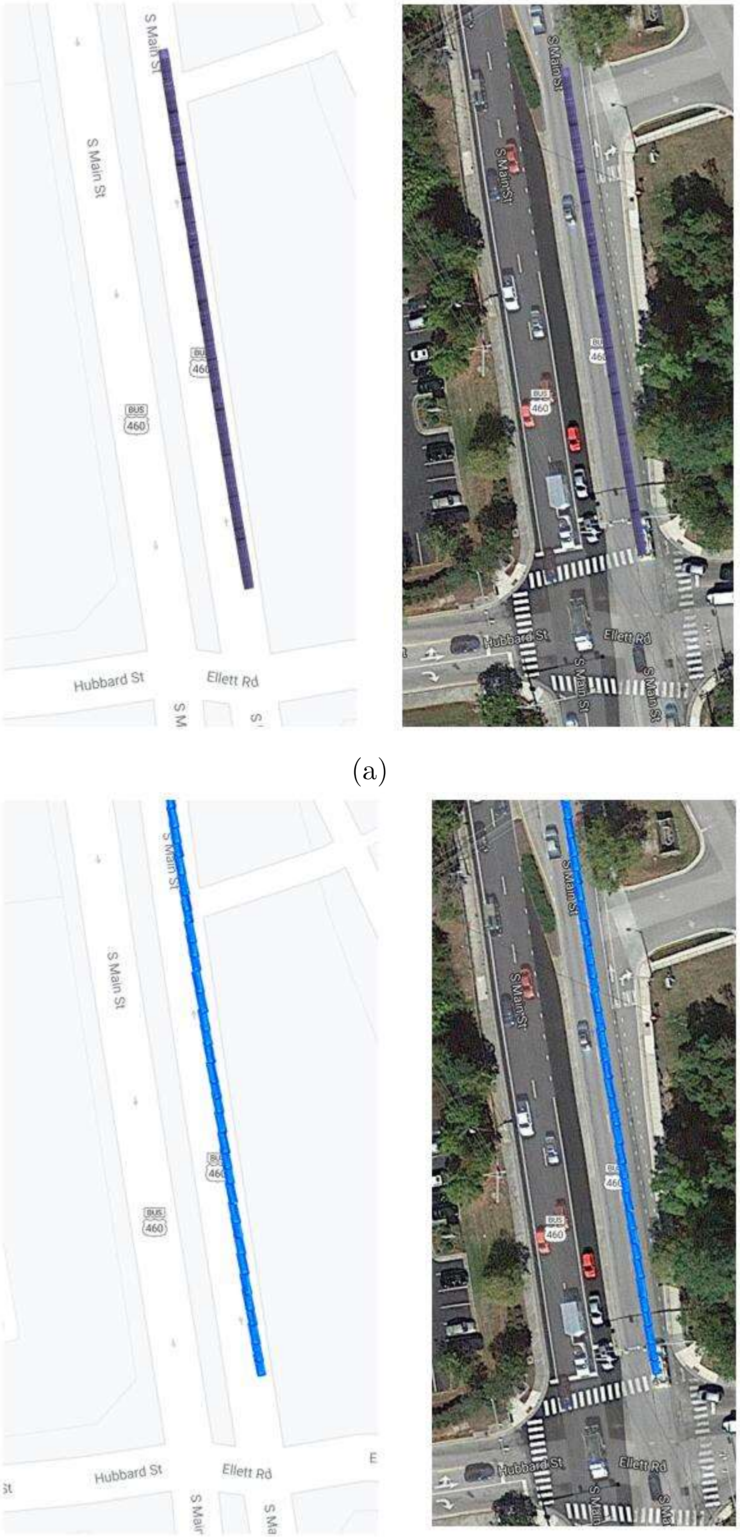


Figure 6.8: Lane-resolution GPS test.



(a)

(b)

Figure 6.9: 2D and 3D road surface mapping and visualization on Google Map.

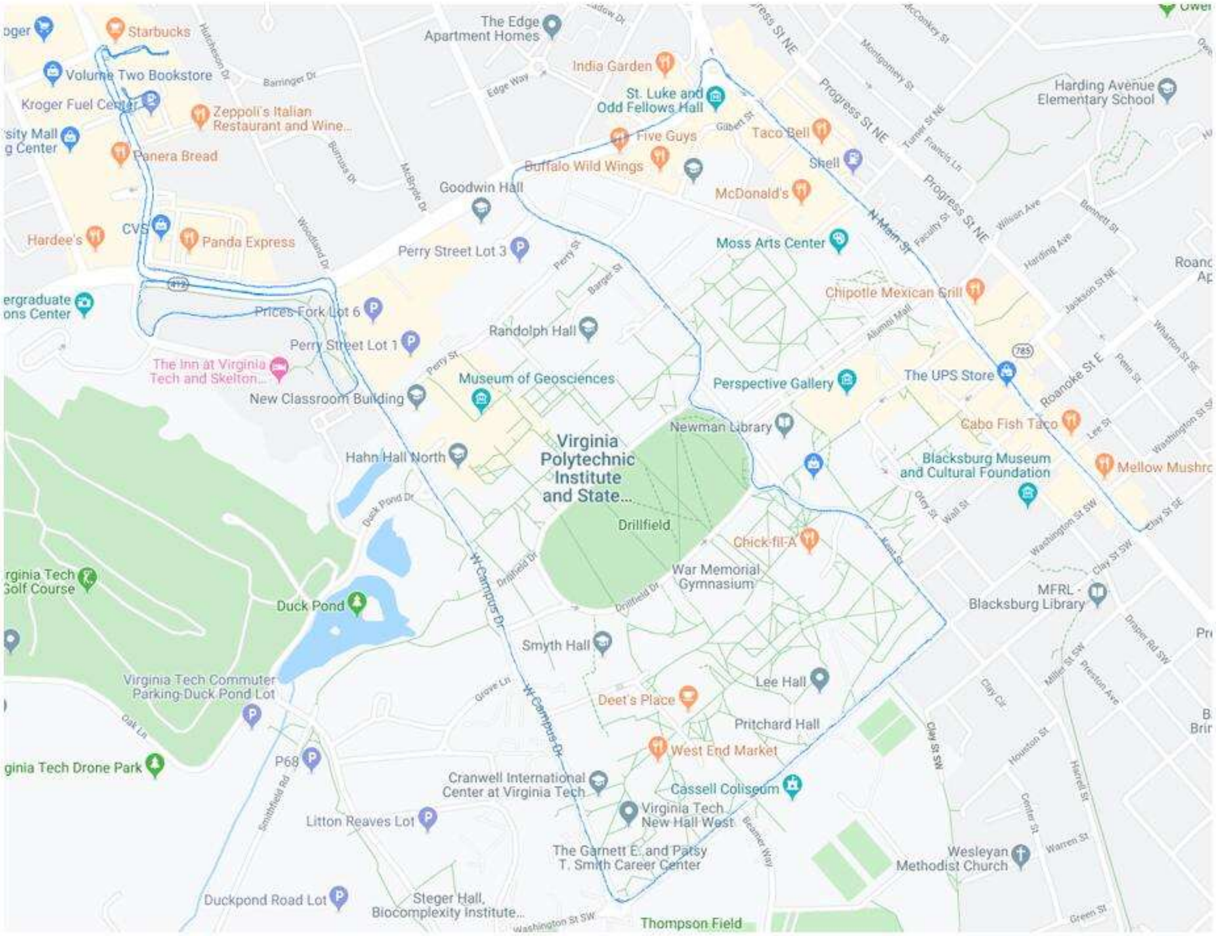


Figure 6.10: Experiment for automatic long distance 2D and 3D road surface map.

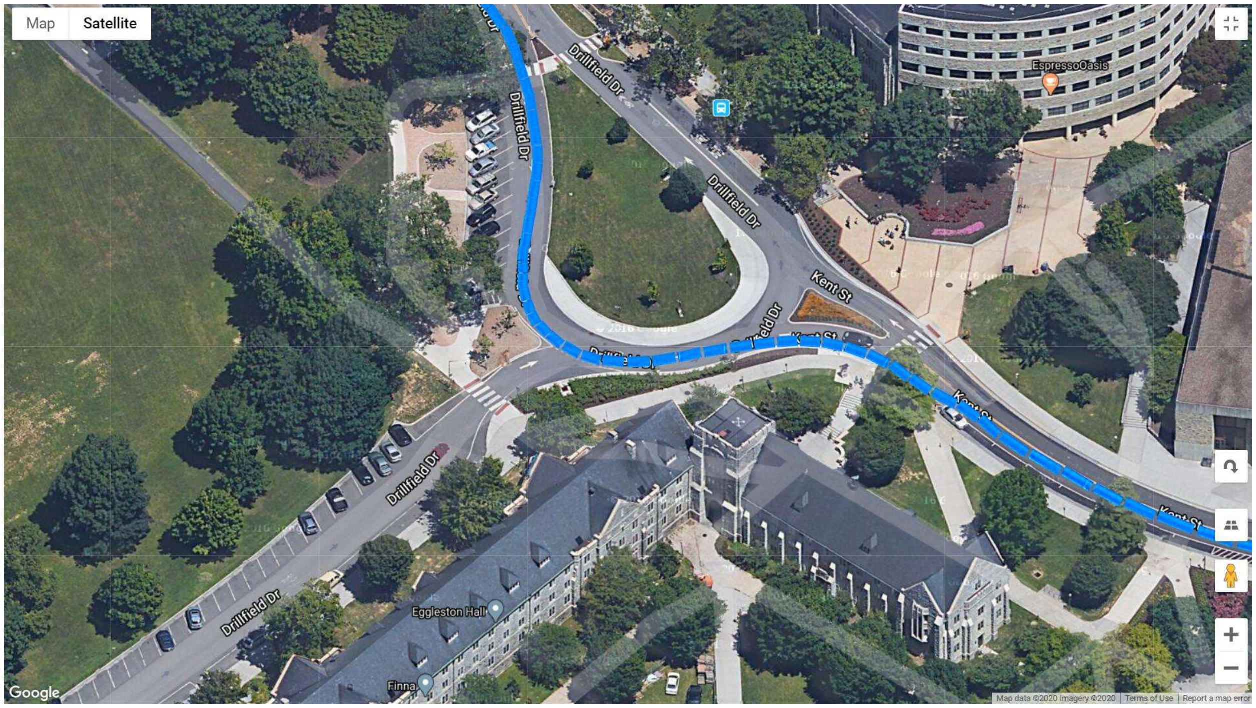


Figure 6.11: Detailed view for part of the automatic long distance 3D road surface map.

Chapter 7

Conclusions and Future Works

7.1 Conclusions

This dissertation has presented four principal contributions: a degenerate near-planar road surface 3D reconstruction from one camera, an automatic road surface defects detection and estimation technique using 2D and 3D information, a self-supervised learning technique for the road surface defects detection based on 3D physics-based data and data-driven CNN, and a designed road surface monitoring system that performs the simultaneous 2D and 3D road surface mapping.

The first proposed technique reconstructs the degenerate near-planar road surface into 3D by using one camera. The proposed technique tackles the degenerate issue for near-planar road surface by deriving a non-degenerate fundamental matrix based on the camera's translation, rotation, and the near-planar homography. Comparing with traditional SfM technique which also uses one camera to reconstruct 3D object, the proposed technique only need two images to reconstruct the 3D road surface without the degenerate issue while the traditional SfM needs three to avoid the degeneracy. The absolute metric-scaled 3D road surfaces are obtained from a conversion procedure which uses the near-planar characteristics. The experiments in both simulation environment and the real-world environment evaluate the proposed technique. The results show that by changing to different unevenness δ and image noises σ, the proposed reconstruction has less than 3 mm reconstruction error even at the worst case during the experiment under different rotation angle θ_x, θ_y, and θ_z caused by the potential vibration. The test on real road surface demonstrates the quality of the reconstructed 3D road surface and the mean reconstruction error of less than 1 mm while the standard deviation also less than 1 mm.

The second technique proposed in the dissertation is an automatic road surface defects detection based on 2D and 3D information. After getting the near-planar road surface 3D reconstruction from the previous chapter, a preprocessing process uses the camera's linear motion knowledge to remove the outliers in feature matching. Then the 3D anomaly road surface points are identified from their distances to the reference road plane fitted from all the reconstructed 3D points. Knowing the potential anomaly points from 3D points, the proposed technique automatically detect the defect region by combining the 2D image's brightness information with the 3D points classification. Finally, defects are clustered together based on a mean-shift based filter on the 3D defect points. The experiment on the real road surface first shows the effectiveness of the proposed technique in removing mismatched feature correspondences. The road surface defects detection experiment on public road surface image data demonstrates a 98.95% accuracy, 94.33% precision, and 95.76% recall rate.

The third contribution has presented a self-supervised learning technique for road surface defects detection using near-planar 3D reconstruction. Without human labeling, the proposed technique automatically labels the road surface images into defect or non-defect based on the 3D road information for training data. During testing, while the trained CNN outputs the classification probability, the outputs are combined with the 3D depth data to give a more accurate prediction for the tested road surface images. During the experiment, the image data augmentation is effective in relieving the imbalanced-data problem. Compared with traditional learning-only detection technique, the proposed technique improves the defect-accuracy, non-defect-accuracy, and accuracy of defects detection each from 79.4% to 94.6%, 94.7% to 97.5%, and 92.0% to 97.1%.

The fourth contribution is the design of a road surface monitoring system from which a simultaneous 2D and 3D road surface mapping is proposed. The road surface monitoring system is designed to synchronize different sensors and equipment together such as FPGA, OBD-II, camera, GPS, and onboard PC. By solving the pose of road surface images, the proposed technique generates 2D and 3D road surface map simultaneously. The experiments are implemented at different weathers such as sunny day, cloudy day, and partly cloudy day to capture road surface images on highway, urban road, and rural road. The results show that the GPS data and images are well synchronized and have a resolution to separate each lane of the road. Also, the stitched 2D and 3D road surface maps have correct pose in the geographical environment and can be automatically generated and visualized simultaneously on online map server.

7.2 Future Works

This dissertation has worked on the degenerate near-planar road surface reconstruction and automatic defects detection problem. However, there are still several directions which can be further investigated or explored:

- More experimental evaluations of the defects clustering and characterization.
- Evaluating more on neural networks parameters for the proposed data-driven detection.
- Road surface texture features for vehicle's localization.
- Road surface defects physics-based modeling using 3D surface measurement.

These continuing works will lead to more valuable outcomes from the information extracted from road surfaces.

Bibliography

[1] E. Morrison, “A prototype scanning stylus profilometer for rapid measurement of small surface areas,” *International Journal of Machine Tools and Manufacture*, vol. 35, no. 2, pp. 325–331, 1995.

[2] P. Pawlus and M. Śmieszek, “The influence of stylus flight on change of surface topography parameters,” *Precision Engineering*, vol. 29, no. 3, pp. 272–280, 2005.

[3] P. Dias, V. Sequeira, J. G. Gonçalves, and F. Vaz, “Automatic registration of laser reflectance and colour intensity images for 3d reconstruction,” *Robotics and Autonomous Systems*, vol. 39, no. 3-4, pp. 157–168, 2002.

[4] X. Xiong, A. Adan, B. Akinci, and D. Huber, “Automatic creation of semantically rich 3d building models from laser scanner data,” *Automation in construction*, vol. 31, pp. 325–337, 2013.

[5] S.-J. Yu, S. R. Sukumar, A. F. Koschan, D. L. Page, and M. A. Abidi, “3d reconstruction of road surfaces using an integrated multi-sensory approach,” *Optics and lasers in engineering*, vol. 45, no. 7, pp. 808–818, 2007.

[6] R. Usamentiaga, J. Molleda, and D. F. García, “Fast and robust laser stripe extraction for 3d reconstruction in industrial environments,” *Machine Vision and Applications*, vol. 23, no. 1, pp. 179–196, 2012.

[7] C. Brenner, “Building reconstruction from images and laser scanning,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 6, no. 3-4, pp. 187–198, 2005.

[8] E. Bylow, J. Sturm, C. Kerl, F. Kahl, and D. Cremers, "Real-time camera tracking and 3d reconstruction using signed distance functions.," in *Robotics: Science and Systems*, vol. 2, p. 2, 2013.

[9] R. Maier, J. Sturm, and D. Cremers, "Submap-based bundle adjustment for 3d reconstruction from rgb-d data," in *German Conference on Pattern Recognition*, pp. 54–65, Springer, 2014.

[10] K. Wang, G. Zhang, and H. Bao, "Robust 3d reconstruction with an rgb-d camera," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4893–4906, 2014.

[11] A. Dai, M. Nießner, M. Zollhöfer, S. Izadi, and C. Theobalt, "Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration," *ACM Transactions on Graphics (ToG)*, vol. 36, no. 4, p. 1, 2017.

[12] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, "3-d mapping with an rgb-d camera," *IEEE transactions on robotics*, vol. 30, no. 1, pp. 177–187, 2013.

[13] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments," *The International Journal of Robotics Research*, vol. 31, no. 5, pp. 647–663, 2012.

[14] H. Pan, T. Guan, Y. Luo, L. Duan, Y. Tian, L. Yi, Y. Zhao, and J. Yu, "Dense 3d reconstruction combining depth and rgb information," *Neurocomputing*, vol. 175, pp. 644–651, 2016.

[15] A. Adan and D. Huber, "3d reconstruction of interior wall surfaces under occlusion and clutter," in *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pp. 275–281, IEEE, 2011.

[16] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pp. 127–136, IEEE, 2011.

[17] Y. Fu, Q. Yan, L. Yang, J. Liao, and C. Xiao, "Texture mapping for 3d reconstruction with rgb-d sensor," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4645–4653, 2018.

[18] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4104–4113, 2016.

[19] J. J. Koenderink and A. J. Van Doorn, "Affine structure from motion," *JOSA A*, vol. 8, no. 2, pp. 377–385, 1991.

[20] O. D. Faugeras and F. Lustman, "Motion and structure from motion in a piecewise planar environment," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 2, no. 03, pp. 485–508, 1988.

[21] C. Wu, "Towards linear-time incremental structure from motion," in *2013 International Conference on 3D Vision-3DV 2013*, pp. 127–134, IEEE, 2013.

[22] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski, "Building rome in a day," *Communications of the ACM*, vol. 54, no. 10, pp. 105–112, 2011.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[24] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” in *European conference on computer vision*, pp. 404–417, Springer, 2006.

[25] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (surf),” *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[26] J.-M. Morel and G. Yu, “Asift: A new framework for fully affine invariant image comparison,” *SIAM journal on imaging sciences*, vol. 2, no. 2, pp. 438–469, 2009.

[27] C. G. Harris, M. Stephens, *et al.*, “A combined corner and edge detector.,” in *Alvey vision conference*, vol. 15, pp. 10–5244, Citeseer, 1988.

[28] E. Rosten and T. Drummond, “Machine learning for high-speed corner detection,” in *European conference on computer vision*, pp. 430–443, Springer, 2006.

[29] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[30] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, “An optimal algorithm for approximate nearest neighbor searching fixed dimensions,” *Journal of the ACM (JACM)*, vol. 45, no. 6, pp. 891–923, 1998.

[31] M. Muja and D. G. Lowe, “Fast approximate nearest neighbors with automatic algorithm configuration.,” *VISAPP (1)*, vol. 2, no. 331-340, p. 2, 2009.

[32] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision.* Cambridge university press, 2003.

[33] I. Brilakis, H. Fathi, and A. Rashidi, “Progressive 3d reconstruction of infrastructure with videogrammetry,” *Automation in construction*, vol. 20, no. 7, pp. 884–895, 2011.

[34] M. Golparvar-Fard, V. Balali, and J. M. de la Garza, “Segmentation and recognition of highway assets using image-based 3d point clouds and semantic texton forests,” *Journal of Computing in Civil Engineering*, vol. 29, no. 1, p. 04014023, 2015.

[35] M. Pollefeys, D. Nistér, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, P. Merrell, *et al.*, “Detailed real-time urban 3d reconstruction from video,” *International Journal of Computer Vision*, vol. 78, no. 2-3, pp. 143–167, 2008.

[36] A. Akbarzadeh, J.-M. Frahm, P. Mordohai, B. Clipp, C. Engels, D. Gallup, P. Merrell, M. Phelps, S. Sinha, B. Talton, *et al.*, “Towards urban 3d reconstruction from video,” in *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT’06)*, pp. 1–8, IEEE, 2006.

[37] C. Russell, R. Yu, and L. Agapito, “Video pop-up: Monocular 3d reconstruction of dynamic scenes,” in *European conference on computer vision*, pp. 583–598, Springer, 2014.

[38] T. Schöps, T. Sattler, C. Häne, and M. Pollefeys, “Large-scale outdoor 3d reconstruction on a mobile device,” *Computer Vision and Image Understanding*, vol. 157, pp. 151–166, 2017.

[39] H. Kim, S. Leutenegger, and A. J. Davison, “Real-time 3d reconstruction and 6-dof tracking with an event camera,” in *European Conference on Computer Vision*, pp. 349–364, Springer, 2016.

[40] S. Guo and C. Meng, “Monocular visual odometry and obstacle detection system based on ground constraints,” in *International Conference on Social Robotics*, pp. 516–525, Springer, 2012.

[41] G. Panahandeh and M. Jansson, "Vision-aided inertial navigation based on ground plane feature detection," *IEEE/ASME transactions on mechatronics*, vol. 19, no. 4, pp. 1206–1215, 2014.

[42] T.-J. Lee, D.-H. Yi, D.-I. Cho, *et al.*, "A monocular vision sensor-based obstacle detection algorithm for autonomous robots," *Sensors*, vol. 16, no. 3, p. 311, 2016.

[43] A. Asvadi, C. Premebida, P. Peixoto, and U. Nunes, "3d lidar-based static and moving obstacle detection in driving environments: An approach based on voxels and multi-region ground planes," *Robotics and Autonomous Systems*, vol. 83, pp. 299–311, 2016.

[44] B. Jia, J. Chen, and K. Zhang, "Recursive drivable road detection with shadows based on two-camera systems," *Machine Vision and Applications*, vol. 28, no. 5-6, pp. 509–523, 2017.

[45] G. Panahandeh, N. Mohammadiha, and M. Jansson, "Ground plane feature detection in mobile vision-aided inertial navigation," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pp. 3607–3611, IEEE, 2012.

[46] D. Holz, S. Holzer, R. B. Rusu, and S. Behnke, "Real-time plane segmentation using rgb-d cameras," in *Robot Soccer World Cup*, pp. 306–317, Springer, 2011.

[47] C.-H. Lin, S.-Y. Jiang, Y.-J. Pu, and K.-T. Song, "Robust ground plane detection for obstacle avoidance of mobile robots using a monocular camera," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pp. 3706–3711, IEEE, 2010.

[48] J.-M. Frahm and M. Pollefeys, "Ransac for (quasi-) degenerate data (qdegsac)," in *Computer vision and pattern recognition, 2006 IEEE computer society conference on*, vol. 1, pp. 453–460, IEEE, 2006.

[49] J. Zhou and B. Li, “Homography-based ground detection for a mobile robot platform using a single camera,” in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pp. 4100–4105, IEEE, 2006.

[50] P. R. Giordano, R. Spica, and F. Chaumette, “An active strategy for plane detection and estimation with a monocular camera,” in *IEEE Int. Conf. on Robotics and Automation, ICRA'14*, 2014.

[51] A. M. Kaneko and K. Yamamoto, “Monocular depth estimation by two-frame triangulation using flat surface constraints,” in *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*, pp. 574–581, IEEE, 2017.

[52] Z. Zhou, H. Jin, and Y. Ma, “Robust plane-based structure from motion,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 1482–1489, IEEE, 2012.

[53] T. Collins and A. Bartoli, “Planar structure-from-motion with affine camera models: closed-form solutions, ambiguities and degeneracy analysis,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1237–1255, 2017.

[54] I. O. for Standardization, *Mechanical vibration and shock-Evaluation of human exposure to whole-body vibration-Part 1: General requirements.* The Organization, 1997.

[55] S. Tighe, N. Li, L. Falls, and R. Haas, “Incorporating road safety into pavement management,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 1699, pp. 1–10, 2000.

[56] Y. Huang and B. Xu, “Automatic inspection of pavement cracking distress,” *Journal of Electronic Imaging*, vol. 15, no. 1, p. 013017, 2006.

[57] M. Herold and D. Roberts, "Spectral characteristics of asphalt road aging and deterioration: implications for remote-sensing applications," *Applied Optics*, vol. 44, no. 20, pp. 4327–4334, 2005.

[58] T. Saarenketo and T. Scullion, "Road evaluation with ground penetrating radar," *Journal of applied geophysics*, vol. 43, no. 2-4, pp. 119–138, 2000.

[59] L. Bursanescu and F. Blais, "Automated pavement distress data collection and analysis: a 3-d approach," in *3-D Digital Imaging and Modeling, 1997. Proceedings., International Conference on Recent Advances in*, pp. 311–317, IEEE, 1997.

[60] D. H. Kil and F. B. Shin, "Automatic road-distress classification and identification using a combination of hierarchical classifiers and expert systems-subimage and object processing," in *Image Processing, 1997. Proceedings., International Conference on*, vol. 2, pp. 414–417, IEEE, 1997.

[61] T. Fukuhara, K. Terada, M. Nagao, A. Kasahara, and S. Ichihashi, "Automatic pavement-distress-survey system," *Journal of Transportation Engineering*, vol. 116, no. 3, pp. 280–286, 1990.

[62] B. X. Yu and X. Yu, "Vibration-based system for pavement condition evaluation," in *Applications of Advanced Technology in Transportation*, pp. 183–189, 2006.

[63] Y.-c. Tai, C.-w. Chan, and J. Y.-j. Hsu, "Automatic road anomaly detection using smart mobile device," in *conference on technologies and applications of artificial intelligence, Hsinchu, Taiwan*, 2010.

[64] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: using a mobile sensor network for road surface monitoring," in *Proceed-*

ings of the 6th international conference on Mobile systems, applications, and services, pp. 29–39, ACM, 2008.

[65] G. Xue, H. Zhu, Z. Hu, J. Yu, Y. Zhu, and Y. Luo, "Pothole in the dark: Perceiving pothole profiles with participatory urban vehicles," *IEEE Transactions on Mobile Computing*, vol. 16, no. 5, pp. 1408–1419, 2017.

[66] A. Mednis, G. Strazdins, R. Zviedris, G. Kanonirs, and L. Selavo, "Real time pothole detection using android smartphones with accelerometers," in *Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on*, pp. 1–6, IEEE, 2011.

[67] A. Tedeschi and F. Benedetto, "A real-time automatic pavement crack and pothole recognition system for mobile android-based devices," *Advanced Engineering Informatics*, vol. 32, pp. 11–25, 2017.

[68] C. Koch and I. Brilakis, "Pothole detection in asphalt pavement images," *Advanced Engineering Informatics*, vol. 25, no. 3, pp. 507–515, 2011.

[69] C. Koch, G. M. Jog, and I. Brilakis, "Automated pothole distress assessment using asphalt pavement video data," *Journal of Computing in Civil Engineering*, vol. 27, no. 4, pp. 370–378, 2012.

[70] A. Banharnsakun, "Hybrid abc-ann for pavement surface distress detection and classification," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 2, pp. 699–710, 2017.

[71] S.-K. Ryu, T. Kim, and Y.-R. Kim, "Image-based pothole detection system for its service and road management system," *Mathematical Problems in Engineering*, vol. 2015, 2015.

[72] K. Chang, J. Chang, and J. Liu, “Detection of pavement distresses using 3d laser scanning technology,” in *Computing in Civil Engineering (2005)*, pp. 1–11, 2005.

[73] X. Yu and E. Salari, “Pavement pothole detection and severity measurement using laser imaging,” in *Electro/Information Technology (EIT), 2011 IEEE International Conference on*, pp. 1–5, IEEE, 2011.

[74] Z. Hou, K. C. Wang, and W. Gong, “Experimentation of 3d pavement imaging through stereovision,” in *International Conference on Transportation Engineering 2007*, pp. 376–381, 2007.

[75] A. El Gendy, A. Shalaby, M. Saleh, and G. W. Flintsch, “Stereo-vision applications to reconstruct the 3d texture of pavement surface,” *International Journal of Pavement Engineering*, vol. 12, no. 03, pp. 263–273, 2011.

[76] M. Ahmed, C. Haas, and R. Haas, “Toward low-cost 3d automatic pavement distress surveying: the close range photogrammetry approach,” *Canadian Journal of Civil Engineering*, vol. 38, no. 12, pp. 1301–1313, 2011.

[77] S. Antol, K. Ryu, and T. Furukawa, “A new approach for measuring terrain profiles,” in *ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. V02AT02A026–V02AT02A026, American Society of Mechanical Engineers, 2013.

[78] I. Moazzam, K. Kamal, S. Mathavan, S. Usman, and M. Rahman, “Metrology and visualization of potholes using the microsoft kinect sensor,” in *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*, pp. 1284–1291, IEEE, 2013.

[79] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[80] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[81] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[82] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[83] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[84] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[85] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500, 2017.

[86] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[87] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.

[88] D. Soukup and R. Huber-Mörk, “Convolutional neural networks for steel surface defect detection from photometric stereo images,” in *International Symposium on Visual Computing*, pp. 668–677, Springer, 2014.

[89] Y. Sun and H.-r. Long, “Detection of weft knitting fabric defects based on windowed texture information and threshold segmentation by cnn,” in *2009 International Conference on Digital Image Processing*, pp. 292–296, IEEE, 2009.

[90] Y.-J. Cha, W. Choi, and O. Büyüköztürk, “Deep learning-based crack damage detection using convolutional neural networks,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.

[91] L. Ri-Xian, Y. Ming-Hai, and W. Xian-Bao, “Defects detection based on deep learning and transfer learning.,” *Metallurgical & Mining Industry*, no. 7, 2015.

[92] G. Psuj, “Multi-sensor data integration using deep learning for characterization of defects in steel elements,” *Sensors*, vol. 18, no. 1, p. 292, 2018.

[93] E. Protopapadakis and N. Doulamis, “Image based approaches for tunnels’ defects recognition via robotic inspectors,” in *International Symposium on Visual Computing*, pp. 706–716, Springer, 2015.

[94] F.-C. Chen and M. R. Jahanshahi, “Nb-cnn: Deep learning-based crack detection using convolutional neural network and naïve bayes data fusion,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4392–4400, 2017.

[95] S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, and B. De Schutter, “Deep con-

volutional neural networks for detection of rail surface defects," in *2016 International joint conference on neural networks (IJCNN)*, pp. 2584–2589, IEEE, 2016.

[96] R. Ren, T. Hung, and K. C. Tan, "A generic deep-learning-based approach for automated surface inspection," *IEEE transactions on cybernetics*, vol. 48, no. 3, pp. 929–940, 2017.

[97] H.-w. Huang, Q.-t. Li, and D.-m. Zhang, "Deep learning based image recognition for crack and leakage defects of metro shield tunnel," *Tunnelling and Underground Space Technology*, vol. 77, pp. 166–176, 2018.

[98] B. Li, K. C. Wang, A. Zhang, E. Yang, and G. Wang, "Automatic classification of pavement crack using deep convolutional neural network," *International Journal of Pavement Engineering*, pp. 1–7, 2018.

[99] J.-K. Park, B.-K. Kwon, J.-H. Park, and D.-J. Kang, "Machine learning-based imaging system for surface defect inspection," *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 3, no. 3, pp. 303–310, 2016.

[100] S. Yarram, G. Varma, and C. Jawahar, "City-scale road audit system using deep learning," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 635–640, IEEE, 2018.

[101] R. Fan, U. Ozgunalp, B. Hosking, M. Liu, and I. Pitas, "Pothole detection based on disparity transformation and road surface modeling," *IEEE Transactions on Image Processing*, vol. 29, pp. 897–908, 2019.

[102] Y. Hu and T. Furukawa, "A high-resolution surface image capture and mapping system for public roads," *SAE International Journal of Passenger Cars-Electronic and Electrical Systems*, vol. 10, no. 2017-01-0082, pp. 301–309, 2017.

[103] F. J. Massey Jr, “The kolmogorov-smirnov test for goodness of fit,” *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.