*Article*

# Degenerate Near-Planar 3D Reconstruction from Two Overlapped Images for Road Defects Detection

**Yazhe Hu [1],\* and Tomonari Furukawa [2]**

[1]  Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA 24060, USA
[2]  Department of Mechanical and Aerospace Engineering, University of Virginia, Charlottesville, VA 22903, USA; tf4rp@virginia.edu
\*  Correspondence: yazhehu@vt.edu

check for updates

**Abstract:** This paper presents a technique to reconstruct a three-dimensional (3D) road surface from two overlapped images for road defects detection using a downward-facing camera. Since some road defects, such as potholes, are characterized by 3D geometry, the proposed technique reconstructs road surfaces from the overlapped images prior to defect detection. The uniqueness of the proposed technique lies in the use of near-planar characteristics of road surfaces' in the 3D reconstruction process, which solves the degenerate road surface reconstruction problem. The reconstructed road surfaces thus result from the richer information. Therefore, the proposed technique detects road surface defects based on the accuracy-enhanced 3D reconstruction. Parametric studies were first performed in a simulated environment to analyze the 3D reconstruction error affected by different variables and show that the reconstruction errors caused by the camera's image noise, orientation, and vertical movement are so small that they do not affect the road defects detection. Detailed accuracy analysis then shows that the mean and standard deviation of the errors are less than 0.6 mm and 1 mm through real road surface images. Finally, on-road tests demonstrate the effectiveness of the proposed technique in identifying road defects while having over 94% in precision, accuracy, and recall rate.

**Keywords:** road surface 3D reconstruction; degenerate reconstruction; road defects detection; pothole detection

## 1. Introduction

A road is one of the most fundamental infrastructures in the transportation system. A healthy and intact road surface condition increases ride comfort and vehicle safety for through traffic [1,2]. The road surface condition inevitably downgrades and is affected by stresses from traffic as well as 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 have resultantly necessitated automatic measurement of the road surface defects such as potholes and ruts, which are mostly characterized by geometry [3–8].

Past works on automatic road defects detection can be classified into three types: the acceleration-based detection, the color-based detection, and the geometry-based detection. The acceleration-based technique uses accelerometers as irregular geometrical changes create vibration that can be measured by accelerometers. Yu et al. [9] analyzed acceleration and automatically detected road defects for the first time to the best of the authors' knowledge. Vittorio et al. [10] detected the road anomalies based on the abnormal accelerometer data from the cellphone. Tai et al. [11] and Eriksson et al. [12] proposed a technique 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. [13] adopted a self-learning one degree-of-freedom vibration signal to predict potholes. Mednis et al. [14] 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 the detection if no tire steps exactly on the road defects.

For color-based techniques, image sensors are often equipped to obtain the appearance of defects. Tedeschi et al. [15] proposed a technique using Local Binary Pattern (LBP) feature-based cascade classifiers to detect road defects from images. Koch et al. [16,17] used the histogram and four different image filters to extract road distress texture features. Jo et al. [18] constrained the road defect region between two lanes through the lane detection technique to increase the precision of pothole detection. Banharnsakun et al. [19] deployed an Artificial Neural Network (ANN) which can categorize the distress into longitudinal crack, transversal crack, and pothole. Ryu et al. [20] separated the pothole region from the background by Histogram Shape-Based Thresholding (HST) and then used multiple filters to find the pothole features. The color-based technique 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 geometry-based techniques, Chang et al. [21] and Yu et al. [22,23] detected potholes by analyzing topological features obtained from 3D laser scanning data. Hou et al. [24], Fan et al. [25], and El et al. [26] applied stereo-vision systems to extract a 3D point cloud from road surface images and detect potholes directly from the 3D model of the road obtained from point cloud data while no 3D reconstruction precision was investigated. Ahmed et al. [27] proposed a pothole detection technique by Structure from Motion (SfM) taking multiple images on one road surface region to reconstruct 3D points of road surface. While accuracy in depth was reported to be in the order of 0.1 mm, the accuracy was attained by manually marking artificial features on the road surface. Antol et al. [28] and Moazzam et al. [29] implemented the road distress detection by 3D point cloud data from an RGB-D camera. The former used a movable RGB-D camera box to enable depth measurement at a low speed, while the latter mounted on a tripod to statically measure the 3D road surface by the RGB-D camera. However, the accuracy of the 3D reconstruction by using a laser sensor or stereo-vision system can be degraded if the vibration of the measuring sensors is significant. Further, 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.

This paper presents a new geometry-based technique that reconstructs road surfaces from two overlapped images captured by a downward-facing camera with little influence caused by the vibration and then detects road defects based on the 3D reconstructed road. The 3D reconstruction performed by using an improved SfM technique is extensively formulated such that the road surfaces, which are near-planar and have small vertical variations, can be reconstructed accurately. By solving the degenerate issue for near-planar road surface reconstruction, the proposed technique thus detects road defects from the accuracy-enhanced 3D reconstructed road surfaces.

This paper is organized as follows. The following section refers to the traditional SfM for the road surface and the degeneracy issue for the planar object reconstruction. Section 3 first presents the proposed 3D reconstruction technique for near-planar road surfaces and then describes the detection of road defects detection based on reconstructed 3D road. Section 4 investigates the ability of the proposed technique parametrically in simulated environments and then applies to real road surface images. Conclusions are summarized in the last section.

## 2. 3D Road Surface Reconstruction from Two Overlapped Images

### 2.1. Problem Formulation

Figure 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 consecutive images $\{I_{k-1}, I_k\}$. This problem is converted into localizing the road surface point cloud $\mathbf{X}^r_k \equiv \{\mathbf{X}^r_{k,i} | \forall i\}$ using the homogeneous two-dimensional (2D) image features $\mathbf{x}^r_{k-1} \equiv \{\mathbf{x}^r_{k-1,i} | \forall i\}$ and the corresponding $\mathbf{x}^r_k \equiv \{\mathbf{x}^r_{k,i} | \forall i\}$, which are extracted from image $I_{k-1}$ and $I_k$ respectively. It is to be noted that $\mathbf{X}^c_k$ should be derived simultaneously with $\mathbf{X}^r_k$ 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 $\mathbf{X}^{rn}_k$ and defect road surface $\mathbf{X}^{rd}_k$. In Figure 1, $\{G\}$ represents the global coordinate system while $\{L\}$ is the local coordinate for two neighboring camera positions.
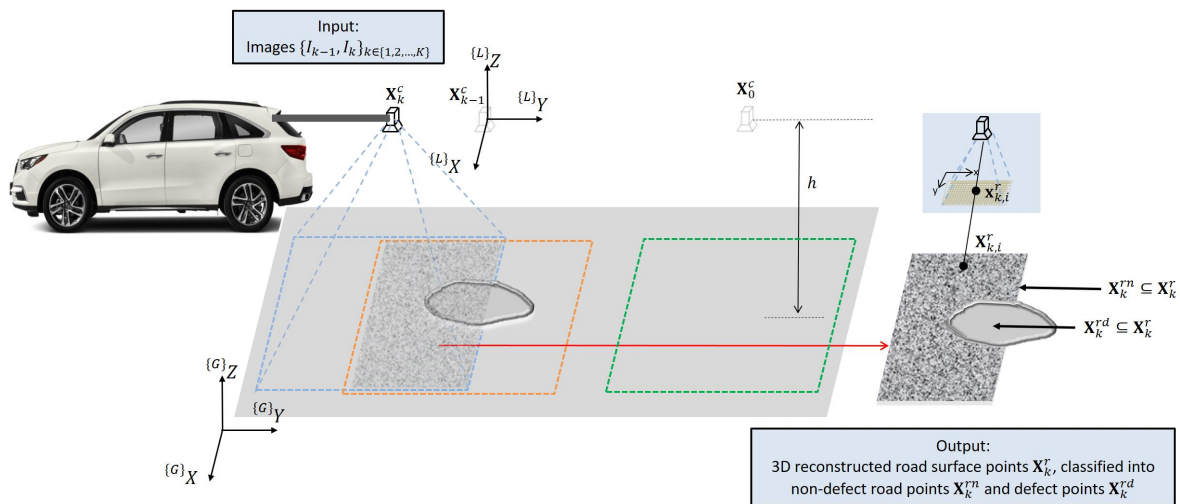


**Figure 1.** Road surface reconstruction settings for defects detection from one downward-facing camera. 3D point cloud are reconstructed from consecutive images to represent the road surface, followed by classifying the road into defective and non-defective surfaces.

Figure 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 1 m × 1 m road surface area. For every number 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.
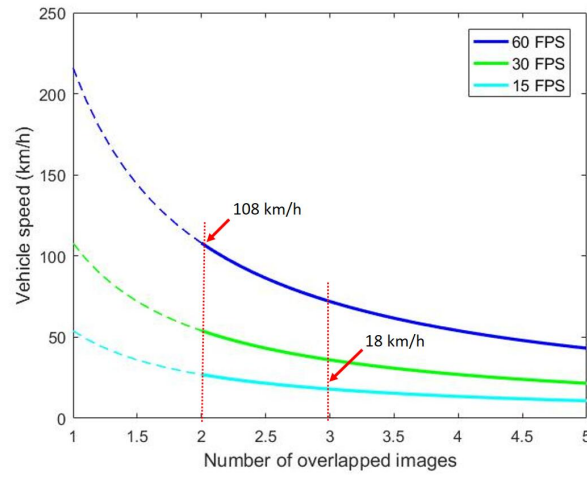
**Figure 2.** The number of images overlapped on each road surface at various vehicle speed.

## 2.2. Two-image 3D Road Surface Reconstruction

Figure 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:

$$\mathbf{l}_{k,i} = \mathbf{e}_k \times \mathbf{x}_{k,i}^r = [\mathbf{e}_k]_\times \mathbf{x}_{k,i}^r \tag{1}$$

Combining Equation (1) with $\mathbf{x}_{k,i}^{r\,T}\mathbf{l}_{k,i} = 0$ yields:

$$\mathbf{x}_{k,i}^{r\,T}[\mathbf{e}_k]_\times \mathbf{x}_{k,i}^r = 0 \tag{2}$$
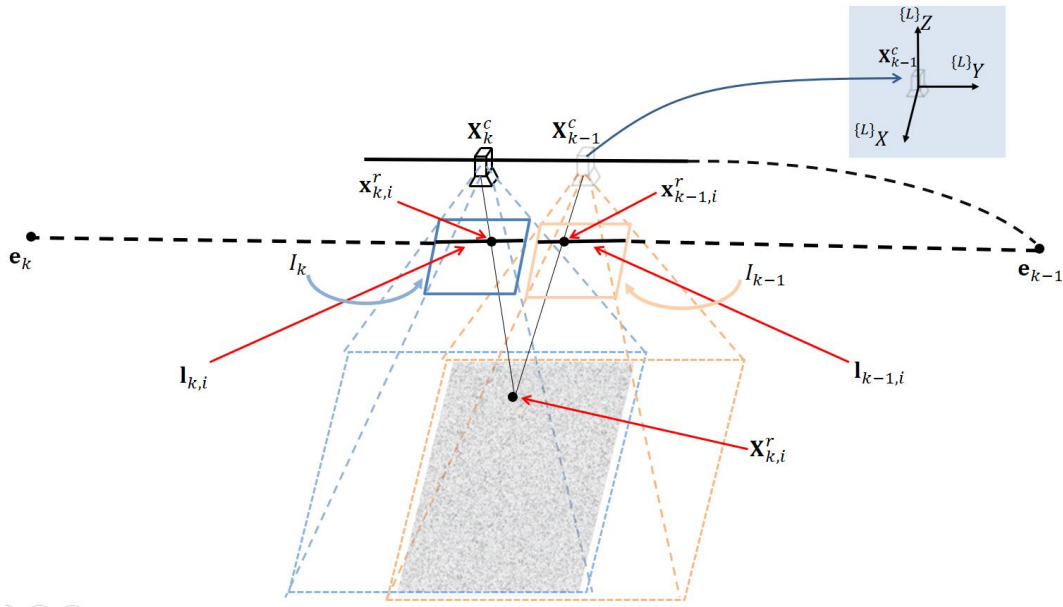


**Figure 3.** 3D road surface reconstruction from two views

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

$$\mathbf{x}_{k,i}^r \propto \mathbf{H}_{ab}\mathbf{x}_{k-1,i}^r \tag{3}$$

Substituting Equation (2) to Equation (3)results in:

$$\mathbf{x}_{k,i}^{r}{}^{T}[\mathbf{e}_k]_{\times}\mathbf{H}_{ab}\mathbf{x}_{k-1,i}^{r} = \mathbf{x}_{k,i}^{r}{}^{T}\mathbf{F}_k\mathbf{x}_{k-1,i}^{r} = 0 \tag{4}$$

where $\mathbf{F}_k = [\mathbf{e}_k]_{\times}\mathbf{H}_{ab}$ is the fundamental matrix of the two images. Equation (4) holds for all the $n$ correspondences $\{\{\mathbf{x}_{k,i}^{r}, \mathbf{x}_{k-1,i}^{r}\}|i = 1, 2, \ldots, n\}$ [30], which means:

$$\mathbf{x}_{k}^{r}{}^{T}\mathbf{F}_k\mathbf{x}_{k-1}^{r} =$$

$$\begin{bmatrix} x_{k,1}^{r} & \cdots & x_{k,n}^{r} \\ y_{k,1}^{r} & \cdots & y_{k,n}^{r} \\ 1 & \cdots & 1 \end{bmatrix}^{T} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_{k-1,1}^{r} & \cdots & x_{k-1,n}^{r} \\ y_{k-1,1}^{r} & \cdots & y_{k-1,n}^{r} \\ 1 & \cdots & 1 \end{bmatrix} \tag{5}$$

$$= 0$$

The solving of fundamental matrix $\mathbf{F}_k$, as well as the rotation matrix $\mathbf{R}_k$ and the translation $\mathbf{t}_k$ are given by the Appendix A. The final 3D reconstructed road surface $\mathbf{X}_k^{r}$ is given by the triangulation $f_t(\cdot)$:

$$\mathbf{X}_k^{r} = f_t(\mathbf{K}, \mathbf{R}_k, \mathbf{t}_k, \mathbf{x}_k^{r}, \mathbf{x}_{k-1}^{r}) \tag{6}$$

where $\mathbf{K}$ is the camera's intrinsic matrix.

### 2.3. Planar Surface Degeneracy Problem

Since the road surfaces are near-planar, it suffers from the degenerate issue which will be shown by the 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 $3 \times 3$ homography matrix $\mathbf{H}_k$:

$$\mathbf{x}_k^{r} \propto \mathbf{H}_k\mathbf{x}_{k-1}^{r} \tag{7}$$

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 $\mathbf{x}_k^{r} \times \mathbf{H}_k\mathbf{x}_{k-1}^{r} = 0$. Thus solving $\mathbf{H}_k$ equals to solving the equation $\mathbf{A}'\mathbf{h}_k = 0$ where $\mathbf{H}_k$, $\mathbf{A}'$ and $\mathbf{h}_k$ are expressed as:

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

$$A_i' = \begin{bmatrix} -x_{k-1,i}^{r} & 0 \\ -y_{k-1,i}^{r} & 0 \\ -1 & 0 \\ 0 & -x_{k-1,i}^{r} \\ 0 & -y_{k-1,i}^{r} \\ 0 & -1 \\ x_{k,i}^{r}x_{k-1,i}^{r} & y_{k,i}^{r}x_{k-1,i}^{r} \\ x_{k,i}^{r}y_{k-1,i}^{r} & y_{k,i}^{r}y_{k-1,i}^{r} \\ x_{k,i}^{r} & y_{k,i}^{r} \end{bmatrix}, \mathbf{A}' = \begin{bmatrix} A_1^T \\ A_2^T \\ . \\ . \\ . \\ A_n^T \end{bmatrix} \tag{8}$$

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

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 Equation (7). Also, $\mathbf{x}_{k-1}$ and $\mathbf{x}_k$ satisfy Equation (5). The substitution of Equation (7) into Equation (5) yields

$$\mathbf{x}_k^r{}^T \mathbf{S}_k \mathbf{x}_k^r = 0 \tag{9}$$

where $\mathbf{S}_k = \mathbf{F}_k \mathbf{H}_k^{-1}$. To satisfy Equation (9), $\mathbf{S}_k$ must be a skew-symmetric matrix given by

$$\mathbf{S}_k = \begin{bmatrix} 0 & -s_3 & s_2 \\ s_3 & 0 & -s_1 \\ -s_2 & s_1 & 0 \end{bmatrix} \tag{10}$$

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

$$\mathbf{F}_k = \mathbf{S}_k \mathbf{H}_k = \begin{bmatrix} 0 & -s_3 & s_2 \\ s_3 & 0 & -s_1 \\ -s_2 & s_1 & 0 \end{bmatrix} \mathbf{H}_k \tag{11}$$

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 Equations (A4) to (A7) and 6. 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. Proposed Degenerate Near-Planar 3D Reconstruction for Road Defects Detection

### 3.1. Overview

Figure 4 shows the proposed degenerate near-planar 3D reconstruction technique for road defects detection. The proposed technique consists of three parts: preprocessing, 3D reconstruction for near-planar road, and post-processing. The preprocessing rejects the mismatched feature correspondences to dramatically improve the feature matching between $I_{k-1}$ and $I_k$, which contributes to resolving the degenerate issue for near-planar road surface reconstruction. Then, a newly derived fundamental matrix $\mathbf{F}_k$ with no ambiguity improves SfM and significantly resolves the degenerate issue. In the post-processing, since the reconstructed points $\mathbf{X}_k^r$ are unitless, the proposed technique converts $\mathbf{X}_k^r$ to metric points ${}^m\mathbf{X}_k^r$. As a result, road defects can be detected reliably due to the enhanced accuracy in 3D surface reconstruction.

### 3.2. Preprocessing

The preprocess rejecting mismatched correspondences is formulated as follows. Let the difference of the $i$th corresponding feature at time step $k-1$ and $k$ be:

$$\mathbf{d}_{k,i}^f \equiv \mathbf{x}_{k,i}^r - \mathbf{x}_{k-1,i}^r \tag{12}$$

This makes the set $\mathbf{d}_k^f \equiv \{\mathbf{d}_{k,i}^f | i = 1, 2, \ldots, n\}$, 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 $k$:

$$\mathbf{d}_k^f \propto \mathbf{X}_k^c - \mathbf{X}_{k-1}^c \tag{13}$$

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

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

$$\hat{\mathbf{x}}_{k-1}^r = \mathbf{x}_{k-1}^r + \omega_{k-1}, \quad \omega_{k-1} \sim N(0, \Sigma_{k-1}) \tag{14}$$

$$\hat{\mathbf{x}}_k^r = \mathbf{x}_k^r + \omega_k, \quad \omega_k \sim N(0, \Sigma_k) \tag{15}$$

Combining Equations (14) and (15) with Equation (12), the proposed technique models $\mathbf{d}_k^f$ as a Gaussian distribution $\mathbf{d}_k^f \sim N(\overline{\mathbf{d}}_k^f, \Sigma^f)$:

$$\overline{\mathbf{d}}_k^f = \overline{\mathbf{x}}_k^r - \overline{\mathbf{x}}_{k-1}^r, \quad \Sigma^f = \Sigma_{k-1} + \Sigma_k \tag{16}$$

where $\overline{\mathbf{d}}_k^f$ is the mean value and $\Sigma^f$ is the covariance matrix of $\mathbf{d}_k^f$. As $\mathbf{d}_{k,i}^f$ 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:

$$\mathbf{d}_k^{f,c} = \{\mathbf{d}_k^f | \overline{\mathbf{d}}_k^f - \lambda \Sigma^f \mathbf{1} < \mathbf{d}_k^f < \overline{\mathbf{d}}_k^f + \lambda \Sigma^f \mathbf{1}\} \tag{17}$$

where $\lambda$ 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, The 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, Equation (17) only needs to find a reasonable $\lambda$ and operate once to keep the correct matching within a range $(\overline{\mathbf{d}}_k^f - \lambda \Sigma^f \mathbf{1}, \overline{\mathbf{d}}_k^f + \lambda \Sigma^f \mathbf{1})$. Therefore, the proposed technique obtains correct feature matchings for the following near-planar 3D reconstruction.
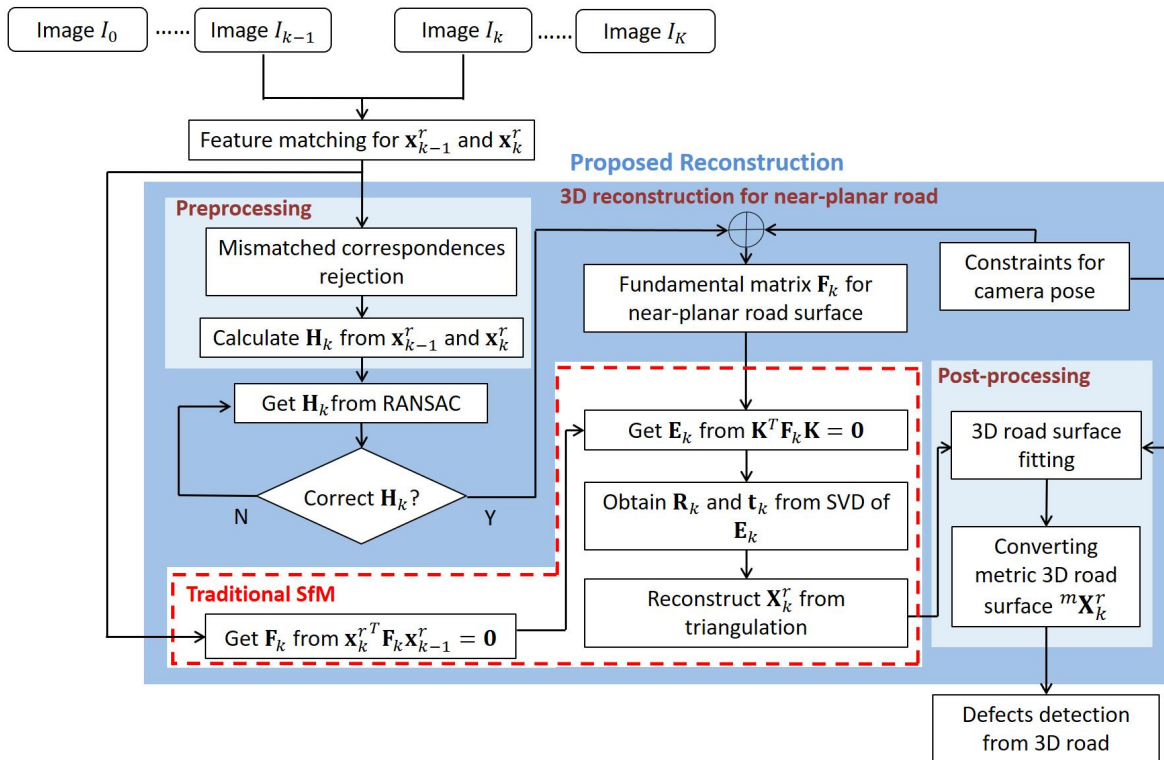


**Figure 4.** Proposed degenerate near-planar surface reconstruction technique for road defects detection.

### 3.3. 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:

$$\mathbf{e}_k = \mathbf{K}[\mathbf{R}_k, \mathbf{t}_k] \cdot \begin{bmatrix} ^{\{L\}}\mathbf{X}^c_{k-1} \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{R}_k, \mathbf{t}_k] \cdot \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{t}_k \tag{18}$$

It is noted that from Equation (1), all the lines $\mathbf{l}_k$ have the following for road surface images:

$$\mathbf{l}_k = \mathbf{e}_k \times \mathbf{x}^r_k \tag{19}$$

Meanwhile, Equation (5)and $\mathbf{x}^{r\,T}_k \mathbf{l}_k = 0$ relates $\mathbf{F}_k$ and $\mathbf{l}_k$ as:

$$\mathbf{F}_k \mathbf{x}^r_{k-1} = \mathbf{l}_k \tag{20}$$

Substitute Equations (19) and (7) into Equation (20) resulting in:

$$\mathbf{F}_k \mathbf{x}^r_{k-1} = \mathbf{e}_k \times \mathbf{x}^r_k = [\mathbf{e}_k]_\times \mathbf{H}_k \mathbf{x}^r_{k-1} \tag{21}$$

Combining Equation (21) with Equation (18), it derives $\mathbf{F}_k$ for the near-planar road surface as:

$$\mathbf{F}_k = [\mathbf{e}_k]_\times \mathbf{H}_k = [\mathbf{K}\mathbf{t}_k]_\times \mathbf{H}_k \tag{22}$$

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 Equation (11) with Equation (22), instead of representing $\mathbf{F}_k$ with any 3-vector $\mathbf{s}$, $\mathbf{F}_k$ is determined in Equation (22) by $\mathbf{t}_k$ which is the up-to-scale translation between the camera positions in two views:

$$\mathbf{t}_k = \mathbf{X}^c_k - \mathbf{X}^c_{k-1} \tag{23}$$

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 $\mathbf{R}_k \approx \mathbf{I}$. Equation (25) can be obtained from Equation (24):

$$\begin{aligned} \mathbf{x}^r_{k-1} &= \mathbf{P}_{k-1}\mathbf{X}^r_k = \mathbf{K}[\mathbf{I}, \mathbf{0}]\mathbf{X}^r_k \\ \mathbf{x}^r_k &= \mathbf{P}_k\mathbf{X}^r_k = \mathbf{K}[\mathbf{R}_k, \mathbf{t}_k]\mathbf{X}^r_k \\ \mathbf{x}^r_k - \mathbf{x}^r_{k-1} &= (\mathbf{P}_k - \mathbf{P}_{k-1})\mathbf{X}^r_k = \mathbf{K}[(\mathbf{R}_k - \mathbf{I}) \mid (\mathbf{t}_k - \mathbf{0})]\mathbf{X}^r_k \end{aligned} \tag{24}$$

$$\mathbf{x}^r_k - \mathbf{x}^r_{k-1} = \mathbf{K}[\mathbf{0}_{3\times3} \mid \mathbf{t}_k] \begin{bmatrix} X^r_k \\ Y^r_k \\ Z^r_k \\ 1 \end{bmatrix} = \mathbf{K}\mathbf{t}_k \tag{25}$$

The substitution of Equation (25) into Equation (22) determines $\mathbf{F}_k$ as:

$$\mathbf{F}_k = [\mathbf{x}^r_k - \mathbf{x}^r_{k-1}]_\times \mathbf{H}_k \tag{26}$$

As a result, a unique fundamental matrix $\mathbf{F}_k$ is obtained from Equation (26) when the road surface is near-planar. Then by using the traditional SfM technique, this $\mathbf{F}_k$ leads to the correct reconstructed road surface points $\mathbf{X}^r_k$ 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 $\hat{\mathbf{x}}_k^r = \mathbf{x}_k^r + \omega$ and $\omega \sim N(\mathbf{0}, \Sigma_{\mathbf{x}_k^r})$ follows a normal distribution. Equation (24)can be rewritten as:

$$\mathbf{X}_k^r = \mathbf{P}_k^+ \hat{\mathbf{x}}_k^r \tag{27}$$

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 Equation (23) be written as $\mathbf{X}_k^r = f(\mathbf{x}_k^r)$. By using the first-order Taylor series expansion Equation (23) becomes:

$$f \approx f_0 + \mathbf{J}_k \hat{\mathbf{x}}_k^r \tag{28}$$

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

$$\Sigma_{\mathbf{X}_k^r} \approx \mathbf{J}_k \Sigma_{\mathbf{x}_k^r} \mathbf{J}_k^T \tag{29}$$

Since $\mathbf{J}_k$ in this scenario equals to $\mathbf{P}^+$ Equation (25) is deduced to be

$$\Sigma_{\mathbf{X}_k^r} \approx \mathbf{P}^+ \Sigma_{\mathbf{x}_k^r} \mathbf{P}^{+T} \tag{30}$$

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.

### 3.4. Post-Processing

After getting the near-planar road surface $\mathbf{F}_k$ with no ambiguity from Equation (26), $\mathbf{X}_k^r$ are reconstructed from Equations (A4) to (A7) and 6. Although, the obtained 3D road surface points $\mathbf{X}_k^r$ are unitless up to a scale factor. In order to get ${}^m\mathbf{X}_k^r$, the proposed technique fits a plane on $\mathbf{X}_k^r$ to represent the road surface:

$$\begin{bmatrix} X_k^r & Y_k^r & 1 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \end{bmatrix} = Z_k^r \tag{31}$$

Then the surface normal vector $\mathbf{n}_k$ and the up-to-scale distance from the camera to the road surface $h_u$ are obtained from $\mathbf{X}_k^r$ based on plane parameters $p_0$, $p_1$, and $p_2$:

$$\mathbf{n}_k = \frac{(p_0, p_1, -1)}{\sqrt{p_0^2 + p_1^2 + 1}} \tag{32}$$

$$h_u = \frac{|p_2|}{\sqrt{p_0^2 + p_1^2 + 1}} \tag{33}$$

The reconstructed surface and the distance $h_u$ obtained by Equations (32) and (33), 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 Equation (34) to Equation (36):

$$d_{k,i} = \frac{p_0 X_{k,i}^r + p_1 Y_{k,i}^r - Z_{k,i}^r + p_2}{\sqrt{p_0^2 + p_1^2 + 1}} \tag{34}$$

$$\mathbf{X}_{k,i}^r \in \begin{cases} \mathbf{X}_k^{rd}, if \ d_{k,i} < 0 \ and \ d_{k,i} \leq T_d \\ \mathbf{X}_k^{rn}, else \end{cases} \tag{35}$$

$$T_n = \frac{size(\mathbf{X}_k^{rn})}{size(\mathbf{X}_k^{rn} + \mathbf{X}_k^{rd})} \tag{36}$$

In Equation (34), $d_{k,i}$ is a signed value calculated as the distance of $\mathbf{X}_{k,i}^r$ to the current reconstructed road surface. The positive $d_{k,i}$ represents the point $\mathbf{X}_{k,i}^r$ located in between the camera and the current fitted road surface. The negative $d_{k,i}$ means the point $\mathbf{X}_{k,i}^r$ is at the other side of the current road surface. Equation (35) illustrates the classification of $\mathbf{X}_{k,i}^r$ into possible defect points $\mathbf{X}_k^{rd}$ and non-defect points $\mathbf{X}_k^{rn}$ by a depth threshold $T_d$. $T_n$ in Equation (36) is a threshold refers to the percentage of non-defect points among all the points $\mathbf{X}_k^r$. If it is assumed that at least $m$ percent of the points $\mathbf{X}_k^r$ 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 $\mathbf{X}_k^{rn}$ from the last iteration. The recursive process will continue until $T_n < m$ is reached.

After the recursive process, an updated camera to road up-to-scale distance $h_u$ was obtained from Equation (33). Then a metric scale factor $\alpha_k$ is calculated based on the real camera to road surface distance $h$:

$$^m\mathbf{X}_k^r = \alpha_k\mathbf{X}_k^r = \frac{h}{h_u}\mathbf{X}_k^r \tag{37}$$

where $^m\mathbf{X}_k^r$ are the metric points with units. From here, the proposed technique converts the up-to-scale points $\mathbf{X}_k^r$ into metric scale road surface points $^m\mathbf{X}_k^r$. Thus the road defects are detected by the depth ($^{\{G\}}Z$ direction) values of $^m\mathbf{X}_k^r$ based on the correct geometry. It is noted here that in order to simplify the notation, $^m\mathbf{X}_k^r$ are still written as $\mathbf{X}_k^r$ in this paper.

## 4. Experimental Results

This section provided two types of experiment to analyze the proposed technique. The first type of experiment was in a Matlab simulated environment which contained the simulated road surface, simulated camera model, and simulated camera motion. The simulation experiments analyzed the influence of different variables to the proposed road surface reconstruction. The second type of experiment was performed on the real road surfaces captured by a road surface imaging system. The real-world experiments demonstrated the accuracy of the proposed technique and its effectiveness on road defects detection.

### 4.1. Experiments in Simulation Environment

Figure 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 $\mathbf{X}_k^r \equiv \{(X_{k,i}^r, Y_{k,i}^r, Z_{k,i}^r)^T | \forall i\}$ 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 $\hat{\mathbf{x}}_k^r = \mathbf{x}_k^r + \omega$, where $\omega \sim N(0, \Sigma_{\mathbf{x}_k^r})$ has the covariance matrix $\Sigma_{\mathbf{x}_k^r}$ and is used to model the uncertainty for matched features in image. The covariance matrix of $\Sigma_{\mathbf{x}_k^r}$ is:

$$\Sigma_x = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \tag{38}$$

As for the orientation, $\theta_x, \theta_y$, and $\theta_z$, 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

$$\epsilon = \frac{1}{N} \sum_{i=1}^{N} |\hat{d}_{k,i}/d_k^c - d_{k,i}/Z_m| \cdot Z_m \tag{39}$$

where $\hat{d}_{k,i}$ is the measured distance and Equation (34) shows the ground truth distance $d_{k,i}$. Table 1 lists the experimental parameters analyzed in the experiment.
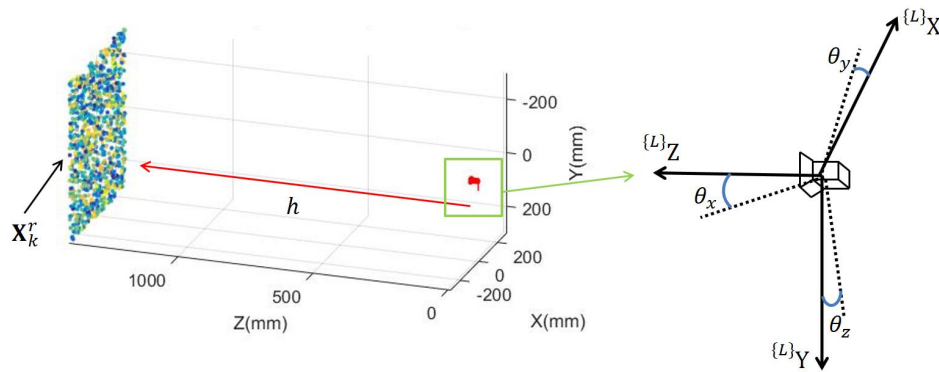


**Figure 5.** Camera and road surface in the simulation environment.

**Table 1.** Parameters for simulated road surface and simulated camera.

| Parameter | Value |
|---|---|
| Road unevenness: $\delta$ [mm] | 0.1, 5, 10 |
| Image noise: $\sigma$ [pixel] | 0.001, 0.002, . . . , 0.1 |
| $Z_m$ [mm] | 500, 800, 1100, 1400, 1700 |
| $\theta_x$ [degree] | 0.05, 0.10, . . . , 5 |
| $\theta_y$ [degree] | 0.05, 0.10, . . . , 5 |
| $\theta_z$ [degree] | 0.05, 0.10, . . . , 5 |
| Two-view translation | |
| **t** [mm, mm, mm] | $(200, 30, 0)^T$ |
| Change of h: $\delta_h$ [mm] | 0.2, 0.4, . . . ,20 |

Figure 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 ($\delta \gg 0$). When $\delta$ 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 $\sigma$. 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 $\sigma$ 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 and similar reconstruction error. When $\delta = 0.1$ mm, i.e., road surface is near-planar, SfM has error usually between 10 and 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.
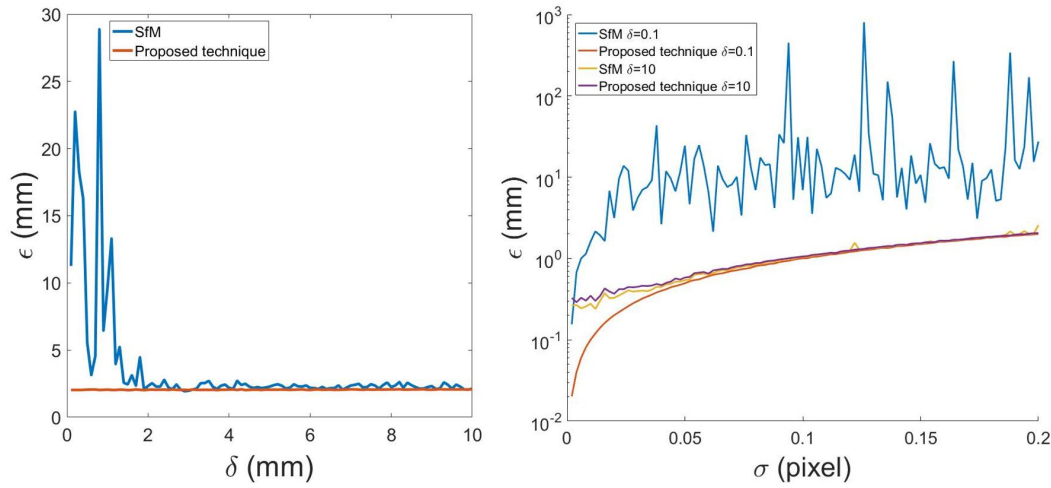
**Figure 6.** Left: 3D reconstruction error comparison between the proposed technique and traditional SfM when road unevenness $\delta$ is changing from 0 to 10 mm. Right: 3D reconstruction error comparison between the proposed technique and traditional SfM at different image noise $\sigma$ while $\delta = 0.1$ or 10 mm.

Figure 7 demonstrates the comparison between the traditional SfM and proposed reconstruction technique for planar road and non-planar road 3D reconstruction with different $\sigma$. 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 by traditional SfM and the bottom figure is the error of 3D reconstruction by the proposed technique. The image uncertainty $\sigma$ 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 $\delta$ becomes larger which means the road is not a planar surface, SfM gives close results to the proposed technique. When $\delta$ becomes smaller the error for SfM increases but for the proposed technique the error remains small.

Figures 8–10 shows the 3D reconstruction error $\epsilon$ by the influence of errors in rotation matrix **R**. In this simulation experiment, the rotation matrix **R** is decomposed as $\mathbf{R} = \mathbf{R}_z \mathbf{R}_y \mathbf{R}_x$ where

$$
\begin{aligned}
\mathbf{R}_x &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \\
\mathbf{R}_y &= \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \\
\mathbf{R}_z &= \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned}
\tag{40}
$$

$\mathbf{R}_x, \mathbf{R}_y, \mathbf{R}_z$ 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 Figure 8, it demonstrates the 3D reconstruction error by changing $\theta_x$. From left to right, each column represents the result under $\delta = 0.1, 5$. $\sigma$ 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. Figures 9 and 10 represents the same experiment by changing $\theta_y$ and $\theta_z$.
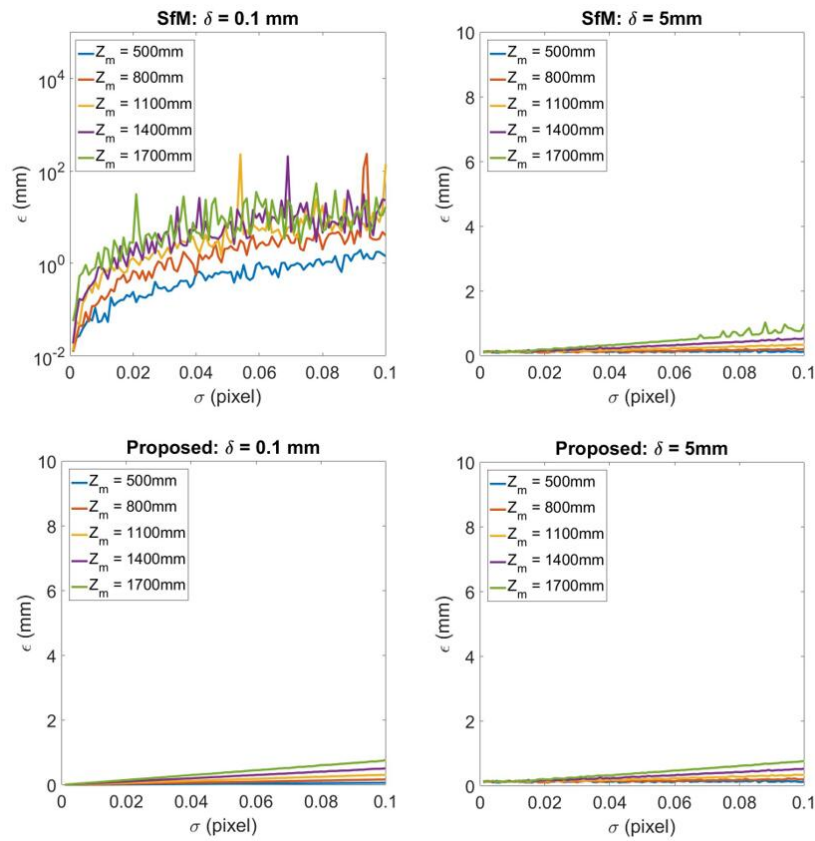
**Figure 7.** 3D reconstruction error for different image noise $\sigma = 0.001, 0.002, \ldots, 0.1$. From left to right each column represents the results for road unevenness $\delta = 0.1$ and $\delta = 5$ respectively. For each column, the top figure shows the 3D reconstruction by traditional SfM technique, while the bottom figure illustrates 3D reconstruction by the proposed technique.

Figure 8 shows the influence to 3D reconstruction error by different $\theta_x$. 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 9 identifies the influence to 3D reconstruction error by different $\theta_y$. The error 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 $\theta_y$ has little influence on the 3D reconstruction errors which are under 2 mm even at the worst case.

Figure 10 demonstrates the influence to 3D reconstruction error by different $\theta_z$. 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 $\theta_z$ almost has no influence to the 3D reconstruction error. The error in this case is mainly influenced by the variable $Z_m$. The larger the $Z_m$, the larger the error $\epsilon$.
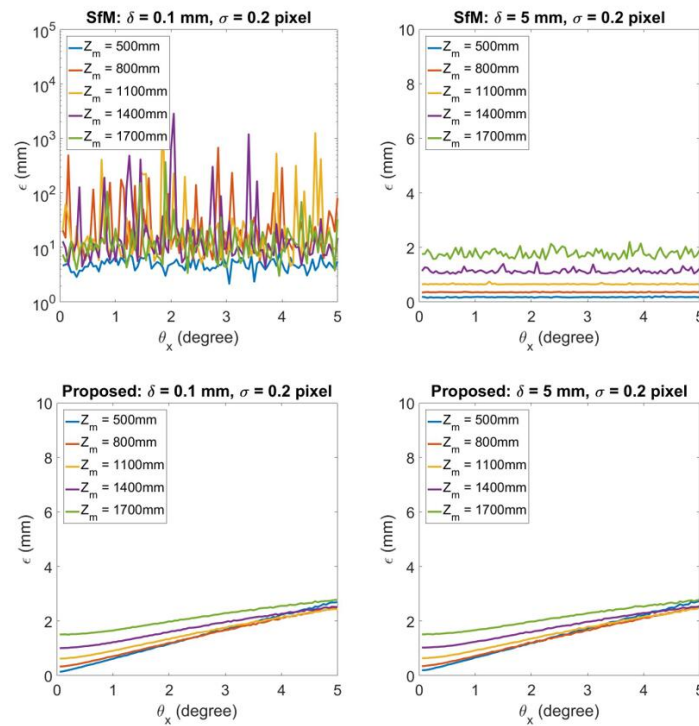
**Figure 8.** 3D reconstruction error for $\theta_x = 0.05°, 0.1°, \ldots, 5°$. From left to right each column represents the results for $\delta = 0.1, 5$ respectively. For each column, the top figure shows the 3D reconstruction by SfM, while the bottom figure illustrates 3D reconstruction by the proposed degenerate reconstruction technique.
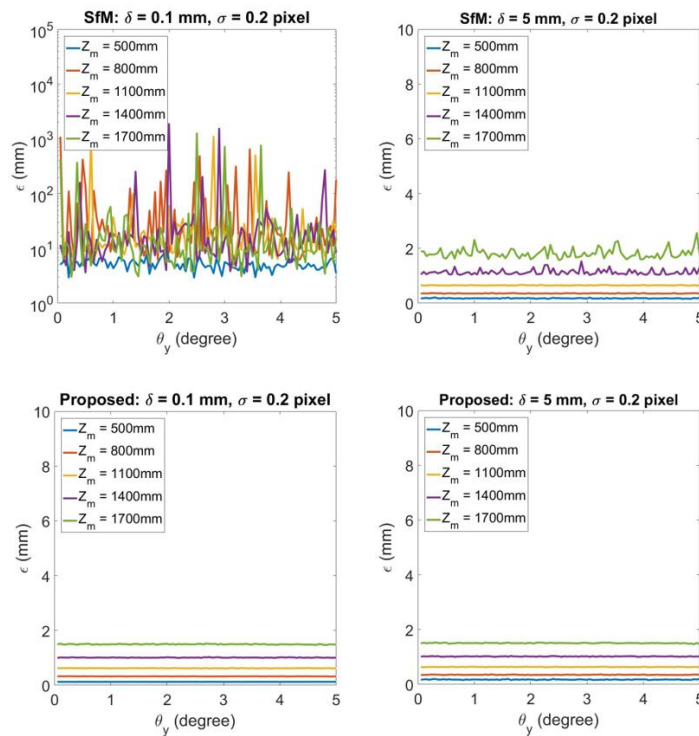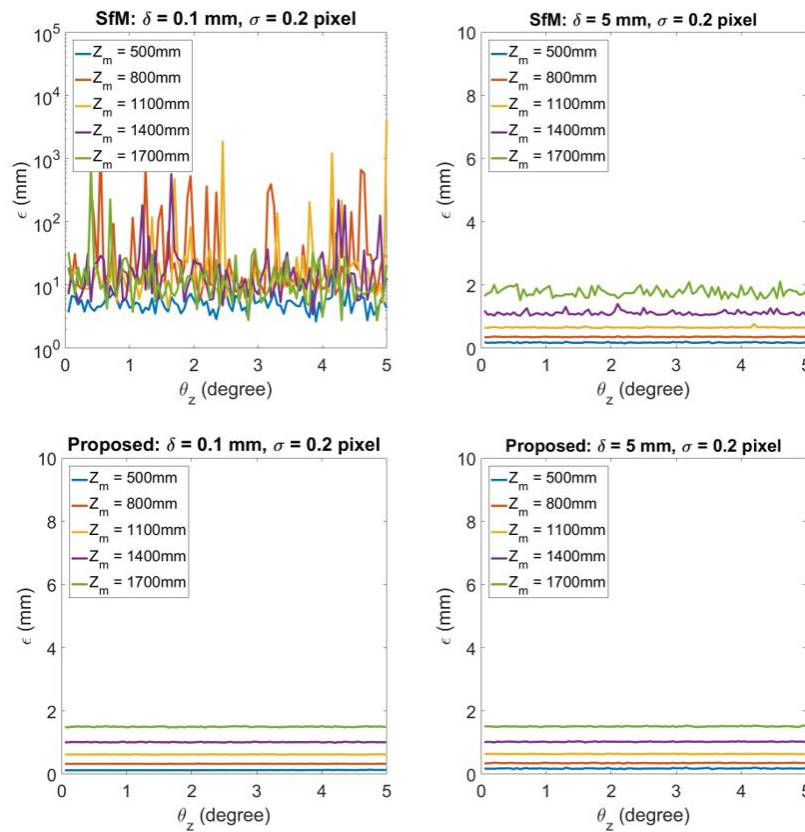


**Figure 9.** 3D reconstruction error for $\theta_y = 0.05°, 0.1°, \ldots, 5°$. From left to right each column represents the results for $\delta = 0.1, 5$ respectively. For each column, the top figure shows the 3D reconstruction by SfM, while the bottom figure illustrates 3D reconstruction by the proposed degenerate reconstruction technique.

**Figure 10.** 3D reconstruction error for $\theta_z = 0.05°, 0.1°, \ldots, 5°$. 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 11 shows the 3D reconstruction error when there exists a change of height $\delta_h$ caused by the vibration in camera to road surface distance $h$. The measured distance $\hat{h}$ is expressed as

$$\hat{h} = h - \Delta h$$
$$\Delta h \sim U(0, \delta_h)$$

(41)

where $\Delta h$ is simulated to having a uniform distribution from 0 to $\delta_h$. In Figure 11 from left to right each column represents the result under $\delta = 0.1, 5$ when $\sigma$ is 0.2. Each top figure illustrates the results of using traditional SfM, while bottom figure represents the counterpart using the proposed technique. Withing each plot $\delta_h$ 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 $\delta_h$ is changing from 0.2 to 20, the error remains almost the same for different $\delta_h$. 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.
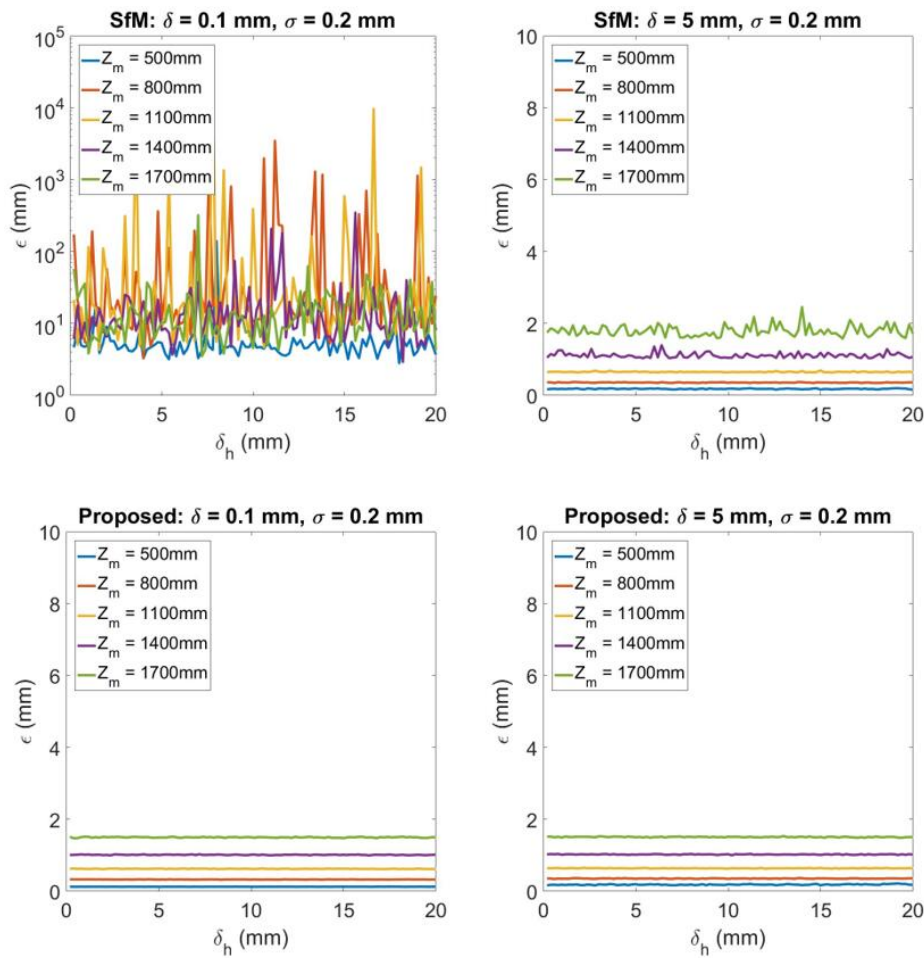
**Figure 11.** 3D reconstruction error of different error $\delta_h = 0.2, 0.4, \ldots, 20$ for camera to road surface distance h. From left to right each column represents the results for $\delta = 0.1, 5$ respectively. For each column, the top figure shows the 3D reconstruction by SfM, while the bottom figure illustrates 3D reconstruction by the proposed degenerate reconstruction technique.

Figure 12 illustrates the comparison between Fan's [25] stereo vision road 3D reconstruction technique, traditional SfM, and the proposed technique. Figure 12a shows the simulation environment for stereo camera, where the baseline between the two cameras, *B*, is set to be $B = 200$ mm. Figure 12b compares the 3D reconstruction error $\epsilon$ from a changing $\theta_y$, 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 $\theta_y$ 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 $\theta_y$ 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 $\sigma$.
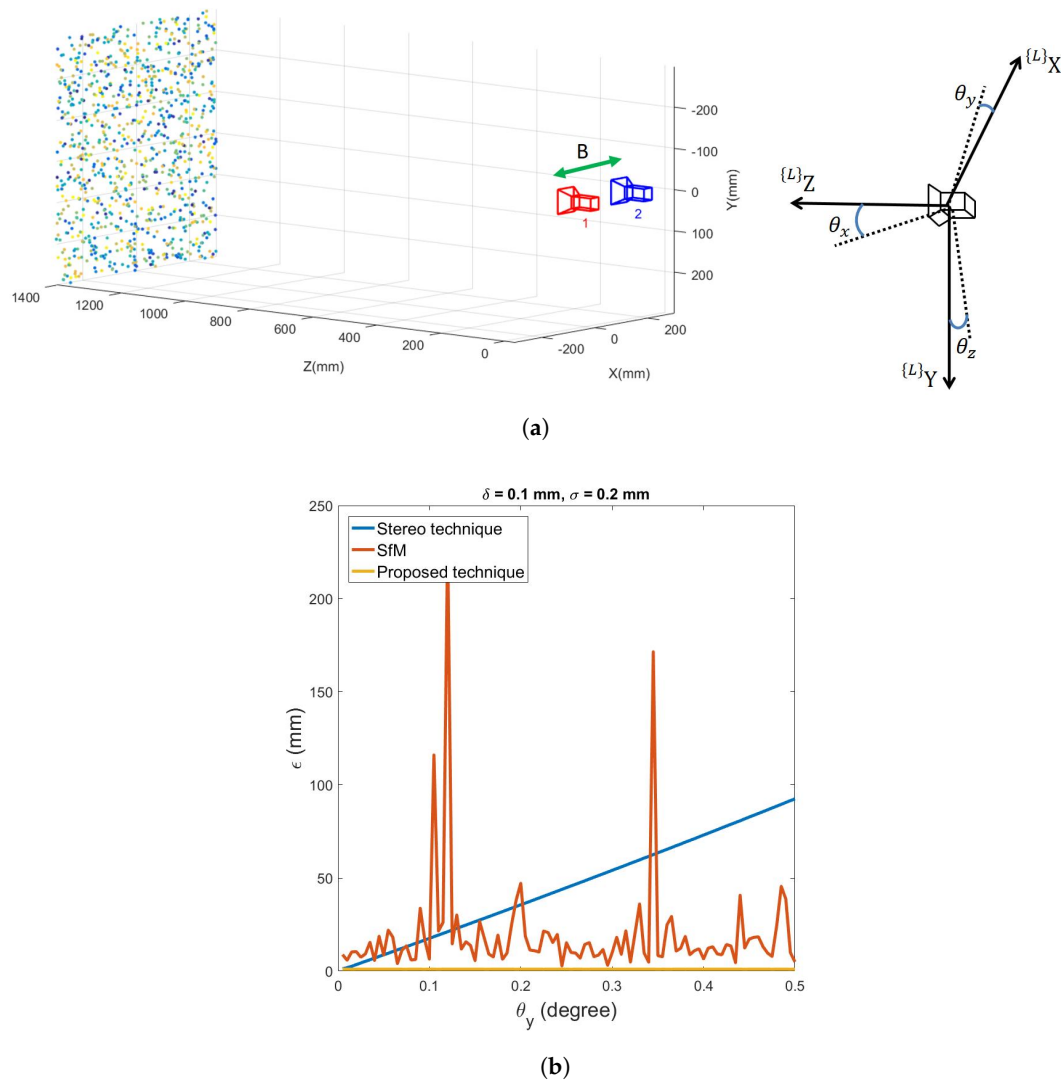
(**a**)



(**b**)

**Figure 12.** Comparison between stereo vision technique, traditional SfM, and the proposed technique under the influence of a changing $\theta_y$ which is caused by the vibration. (**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 $\theta_y$.

*4.2. Experiments on Real Road Surface*

Figure 13 shows the experimental setup of the error analysis for the proposed technique using real images. The camera is facing downward to the road surface with its principle axis vertical to the ground surface as shown in Figure 13a. The ground surface is made by a flat plate to mimic a planar road surface as illustrated in Figure 13b. An image of road surface is printed and stuck to the flat plate to provide road surface patterns for the image feature searching and matching. A circular part of the plate can be removed from the plate to mimic the road pothole.

Figure 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 14c.
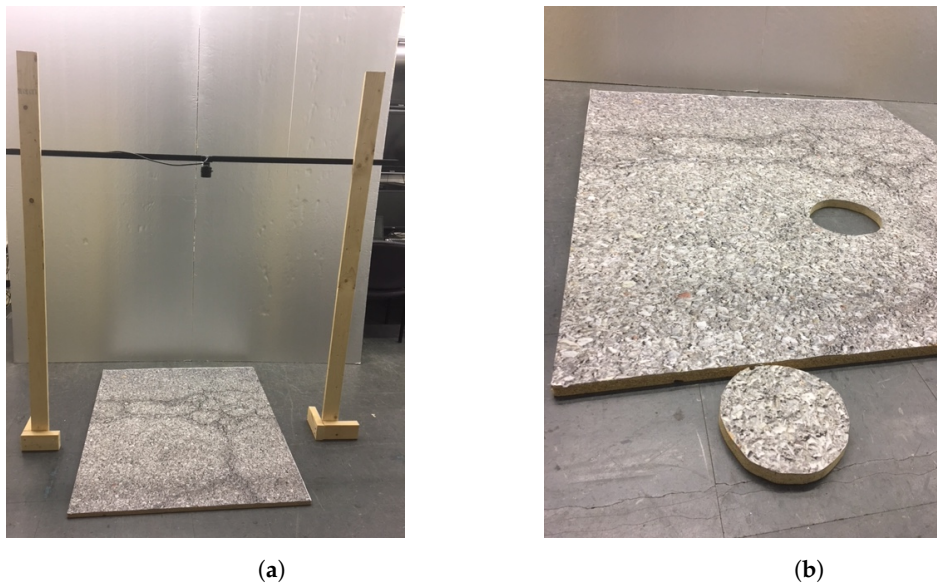
(**a**)　　　　　　　　　　　　　　　　(**b**)

**Figure 13.** Experimental setup for accuracy analysis of the proposed 3D reconstruction technique. 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 planar road. A hole on the plate can be used to simulate the road defect. (**a**) A height-adjustable gantry for the camera. (**b**) A flat plate sticked with a mimic road pattern image.
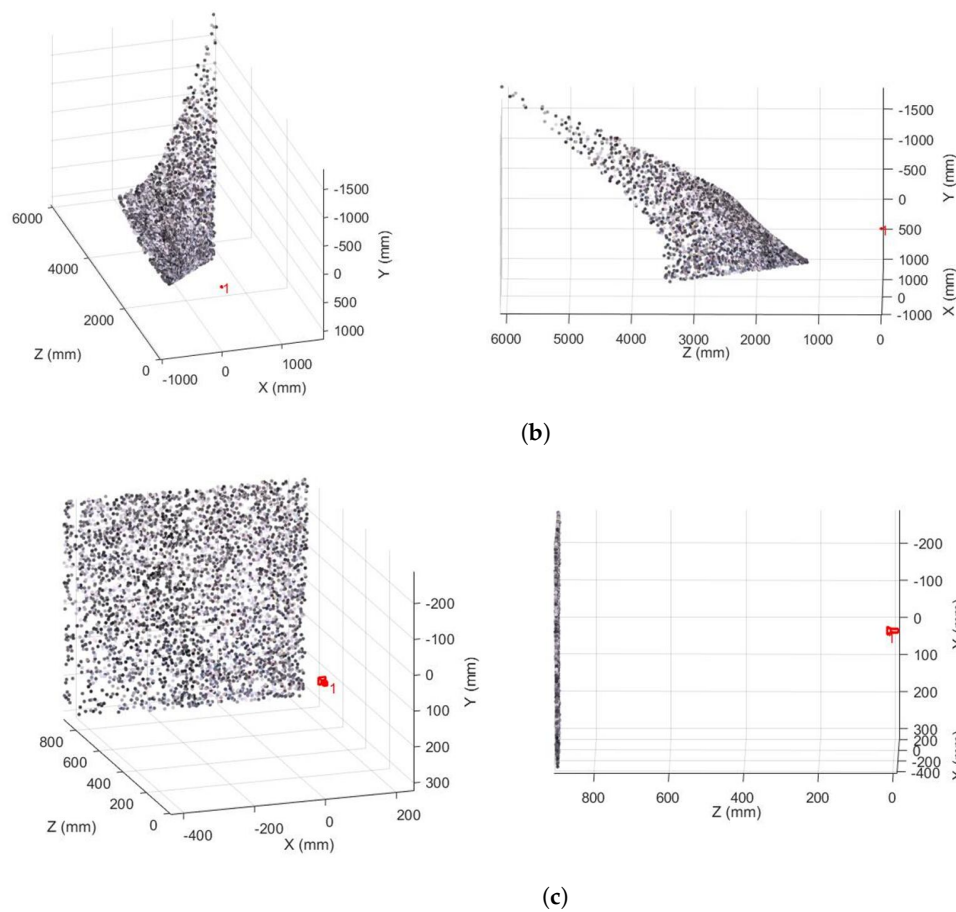


(**a**)

**Figure 14.** *Cont.*

(b)



(c)

**Figure 14.** 3D reconstruction of a flat plate using non-degenerate technique (SfM) and proposed degenerate technique. (**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 15 shows the error analysis for 3D reconstruction using real images. Table 2 lists the parameters analyzed in the experiments using real images. It is noted that the mismatched feature rejection constant is found to be robust to keep correct matchings at $\lambda = 1.5$. Figure 15a demonstrates the error of 3D reconstruction from traditional SfM. Figure 15b represents the 3D reconstruction error by using the proposed technique. The errors are compared between two techniques by changing the height of the camera $h$ from 900 to 1600 mm. The mean errors are plotted and the error bar represents the standard deviation of 10 runs of image capturing for each height. It can be identified from Figure 15 that traditional SfM gives large mean error and standard deviation for this planar plate, while the proposed technique has mean error less than 0.6 mm and standard deviation close to 1 mm.

**Table 2.** Parameters for experiments using real images.

| Parameter | Value |
|---|---|
| Camera Field of View | $56° \times 44°$ |
| Road unevenness: $\delta$ [mm] | $< 0.5$ |
| Camera to road distance: h [mm] | $900, 1000, \ldots, 1600$ |
| Image noise: $\sigma$ [pixel] | $< 0.2$ |
| Mismatched feature rejection constant: $\lambda$ | 1.5 |
| Two-view camera translation: | |
| **t** [mm, mm, mm] | $(100, 0, 0)^T$ |
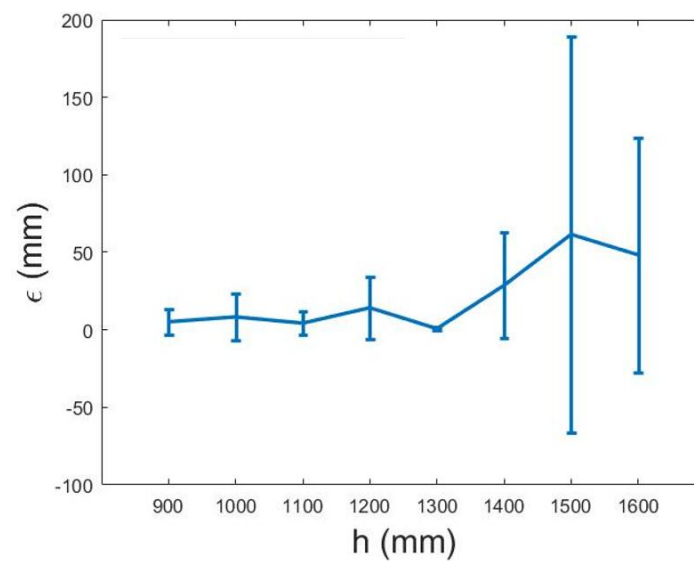
(**a**) 3D reconstruction error for traditional SfM technique.
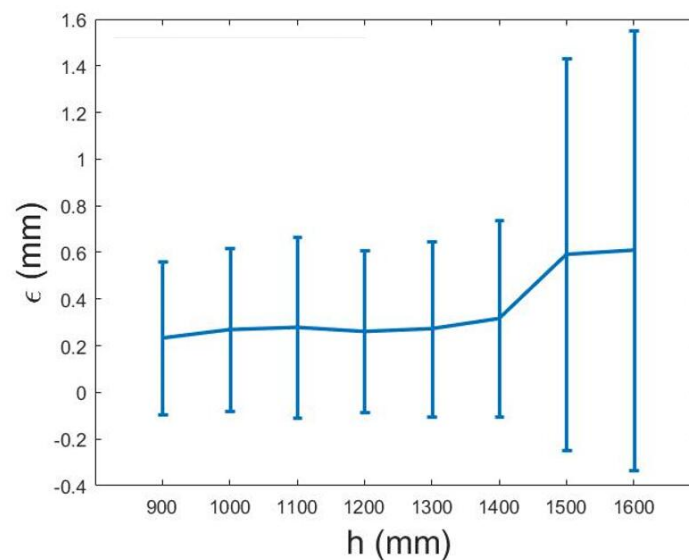


(**b**) 3D reconstruction error from the proposed technique.

**Figure 15.** 3D reconstruction error for traditional SfM technique and proposed technique at $h = 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600$ mm.

Figure 16 shows a system which captures road surface images. The authors' previous work [31] built this system which captures $1024 \times 1280$ resolution road surface images at a driving speed up to 100 km/hour. There are two cameras on this system. Although the proposed 3D reconstruction technique is based on a monocular camera, two cameras can work separately to increase the area of road surface region covered by images. This system is controlled by field-programmable gate array (FPGA) so that the frame rate of the camera is adaptive based on the vehicle speed. On-board diagnostics (OBD) port on the vehicle passes the vehicle's velocity to FPGA which will set higher frame rate for the camera when the vehicle is moving fast and lower frame rate when the vehicle is slow. The system set the frame rate so that there is at least an 50% overlapping area between two consecutive images.

**Figure 16.** FPGA controlled road surface capturing system with adaptive camera frame rate.

Figure 17 demonstrates the qualitative result of reconstructing the road surface using the proposed technique. The top figure shows a road surface image stitching for 20 images to visualize a section of road. The bottom one shows a colormap, which represents the depth ($^{\{G\}}Z$ direction) values of $\mathbf{X}^r_{0:20}$ reconstructed by the proposed technique. It can be seen that two major defects together with several small defects are standing out from the 3D road surface.
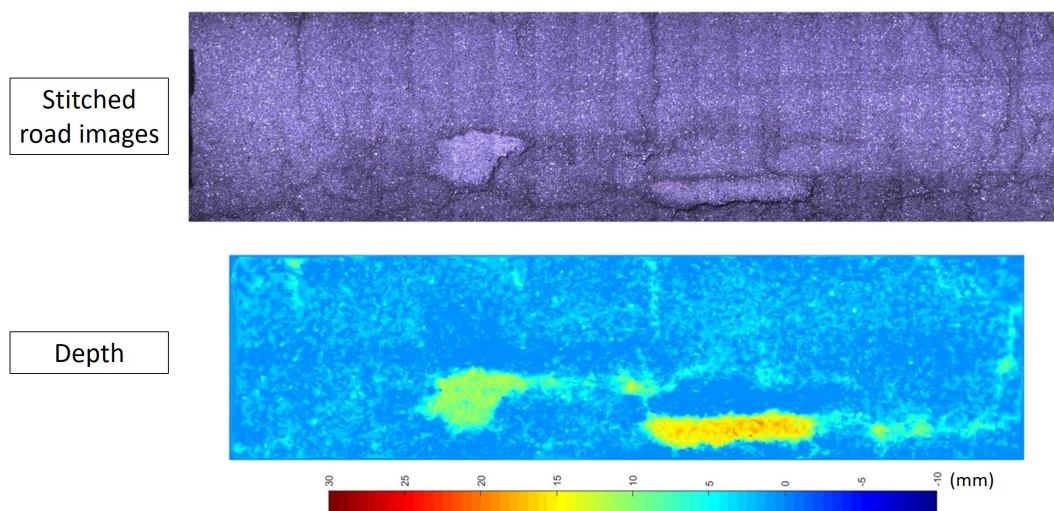


**Figure 17.** The qualitative 3D reconstruction result for a section of road using the proposed technique. Top: a road surface image stitched by 20 consecutive images captured. Bottom: A colormap image for the depth ($^{\{G\}}Z$ direction) values of reconstructed $\mathbf{X}^r_{0:20}$ using the proposed techinque.

Figure 18 compares the proposed technique with traditional SfM in the quality of near-planar road surface reconstruction. The top figure is a near-planar road surface image stitched by 20 consecutive images. The middle one shows the colormap of the $^{\{G\}}Z$ values of the reconstructed road $\mathbf{X}^r_{0:20}$ for the same road using the proposed technique, while the bottom one is the reconstructed colormap of the same road obtained by using traditional SfM. It can be found that the proposed technique has much less outliers and noises in reconstructing the near-planar road surface. The proposed technique even differentiates small cracks by showing the different color at the crack areas. On the other hand, the same road surface reconstructed by the traditional SfM technique shows large deviated depth values at many places which are obviously not correct for a near-planar road surface.
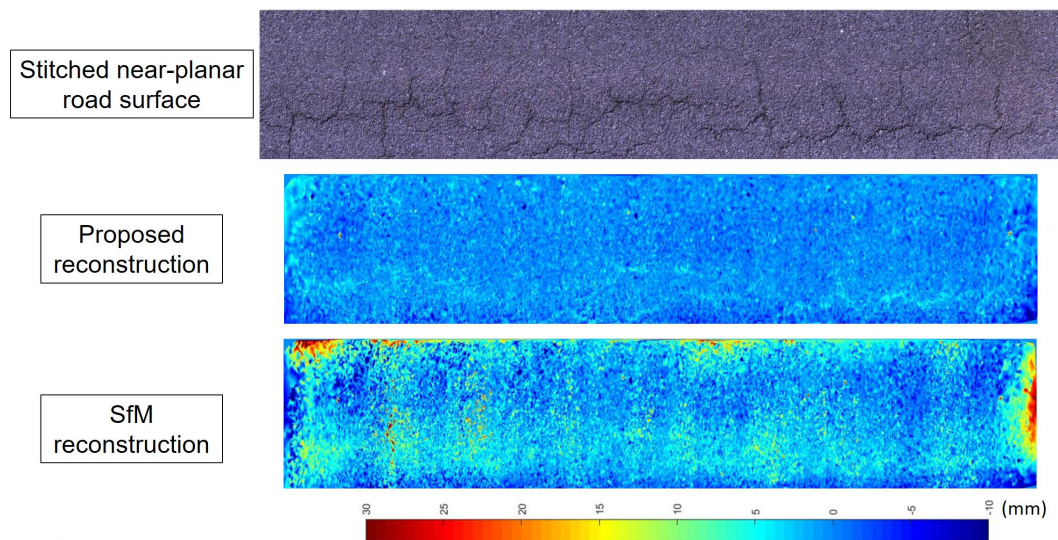
**Figure 18.** The comparison between the proposed and the traditional SfM technique for reconstructing a section of road surface. Top: a road surface image stitched by 20 consecutive images captured. Middle: A colormap image for the depth ($^{\{G\}}Z$ direction) values of reconstructed $\mathbf{X}^r_{0:20}$ from the proposed techinque. Bottom: A colormap image for the depth ($^{\{G\}}Z$ direction) values of reconstructed $\mathbf{X}^r_{0:20}$ using traditional SfM.

Figure 19 demonstrates the repeatability experiment for the proposed technique. Figure 19a represents a section of the road which contains a pothole. This section of road surface are obtained by stitching 50 images which are captured using the system shown in Figure 16. In Figure 19b the $^{\{G\}}Z^r - h$ values of reconstructed 3D road surface points are plotted as a colormap. In Figure 19c, the proposed technique measures the same road section which is reconstructed in Figure 19b. The two measurements are then compared to validate the repeatability of the proposed technique. In Figure 19d, $Z_1$ are the $^{\{G\}}Z^r$ 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 $Z_2 - Z_1$ 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.
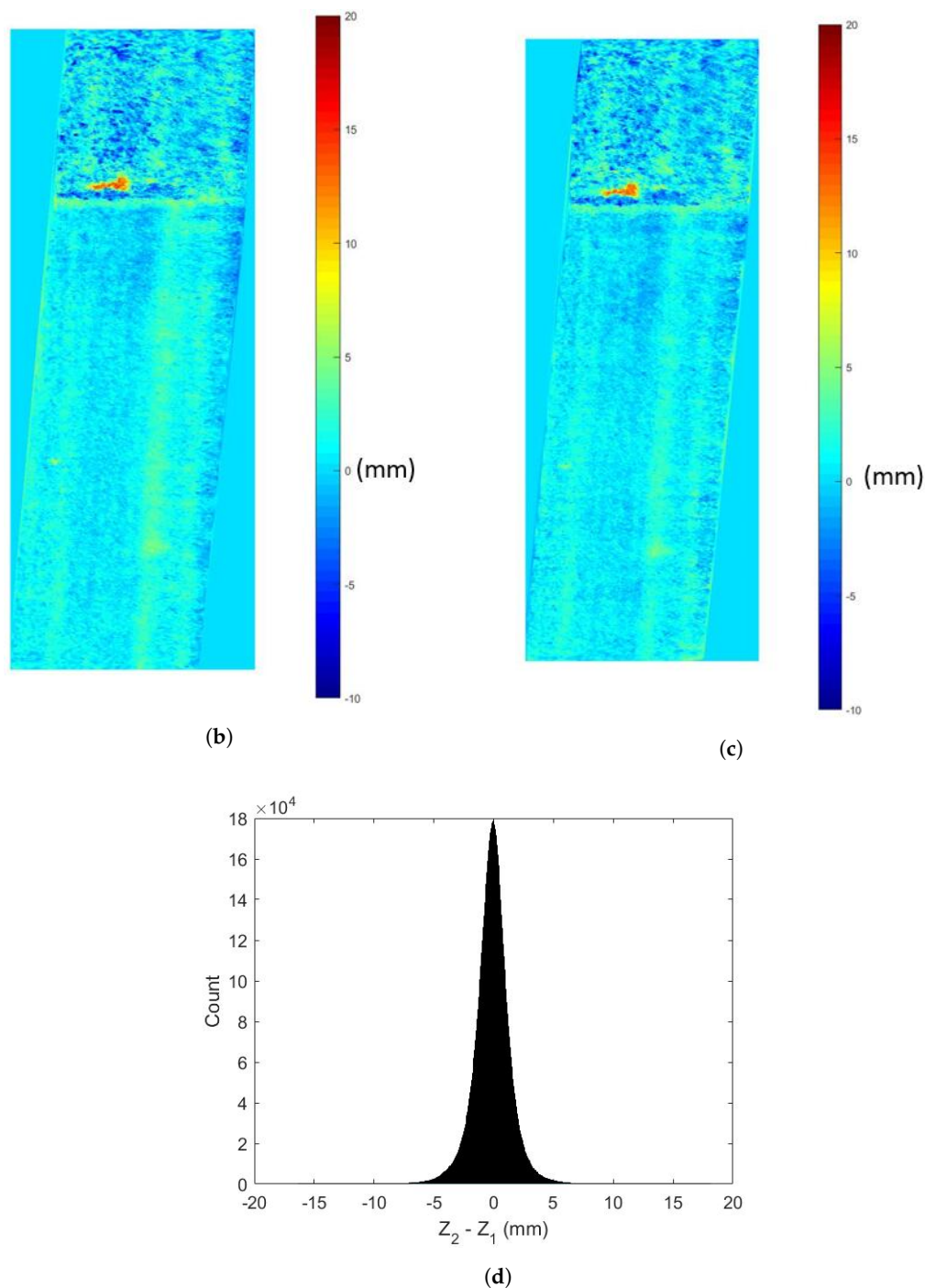


(**a**)

**Figure 19.** *Cont.*

(b)

(c)



(d)
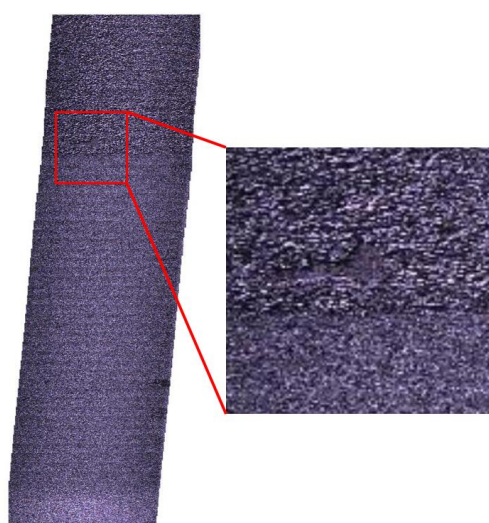
**Figure 19.** Repeatability test for the proposed technique. (**a**) Stitched images of a road section which has a geometrical defect. (**b**) The first measurement of one road section. $^{\{G\}}Z^r - h$ values of road surface point cloud data are represented by a colormap. (**c**) The second measurement of one road section. $^{\{G\}}Z^r - h$ values of road surface point cloud data are represented by a colormap. (**d**) Repeatability quantitative results. $Z_1$ are the $^{\{G\}}Z^r$ values from the first measurement while $Z_2$ are the ones from the second measurement.

Table 3 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 roads for weather conditions such as sunny, cloudy, and partly 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 3 although traditional SfM gives higher recall rate between the proposed technique and traditional SfM, it has only 34.34% precision rate. It means that although traditional SfM rarely misses the detection of potholes (less FN), it generates more 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 3.** Performance of road surface defects detection for different techniques.

|  | 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% |

## 5. Conclusions

A geometry-based technique of reconstructing degenerate near-planar road surfaces from two images for road defects detection is presented in this paper. The proposed technique mathematically formulates the near-planar road surface reconstruction problem, and improves traditional SfM for the 3D road reconstruction process. Since the degenerate issue of the near-planar road surface reconstruction is solved by the proposed technique, road surface defects are thus detected from the accuracy-enhanced 3D road surfaces.

Two types of experiment were conducted to evaluate the proposed road surface 3D reconstruction for the defects detection technique. In the simulation environment, the first experiment compared SfM and the proposed technique under different road unevenness $\delta$ and the noise $\sigma$ in images. Results showed that the changing of $\delta$ does not affect the reconstruction error $\epsilon$ using the proposed technique but increases $\epsilon$ dramatically for traditional SfM when $\delta$ is close to 0. The second experiment compared traditional 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 $\epsilon$ is less than 3 mm even at the worst case. The third experiment showed the change of camera to road distance $\delta_h$ almost does not change the $\epsilon$ when $0 < \delta_h < 20$ mm. The comparison of the stereo vision technique, traditional SfM, and the proposed technique demonstrated the robustness of the proposed technique for road surface reconstruction under the influence of vibration. For experiments using real images, the first experiment showed the 3D reconstruction error $\epsilon$ using both traditional SfM and the proposed technique for the reconstruction of a flat surface under laboratory environment. The results showed that the error for traditional SfM is much higher than the proposed technique, and the proposed technique has a mean error within 1 mm and standard deviation within 1 mm for h from 900 to 1600 mm. Lastly, 6300 real road surface images were captured by the presented system on both local road and highway road surfaces. The proposed technique increased the accuracy from 80% to 98.95% and precision from 34.34% to 94.33% for road defects detection.

This paper focused on reconstructing a 3D structure for road defects using a downward-facing camera. Future works include: 1. Making the camera facing forward to capture the images in front of the vehicle, and then detect defects and objects on the road surface to help vehicles avoid obstacles. 2. Using deep neural networks on both the images and 3D reconstructed points to improve the accuracy of road surface defects detection.

## Appendix A

To solve $\mathbf{F}_k$, Equation (5) is rearranged to a form of $\mathbf{A}\mathbf{f}_k = 0$, where:

$$A_i = \begin{bmatrix} x^r_{k,i} x^r_{k-1,i} \\ x^r_{k,i} y^r_{k-1,i} \\ x^r_{k,i} \\ y^r_{k,i} x^r_{k-1,i} \\ y^r_{k,i} y^r_{k-1,i} \\ y^r_{k,i} \\ x^r_{k-1,i} \\ y^r_{k-1,i} \\ 1 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} A_1^T \\ A_2^T \\ . \\ . \\ . \\ A_n^T \end{bmatrix} \tag{A1}$$

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

In SfM [30], $\mathbf{F}_k$ is obtained by solving the minimization problem:

$$\min_{\mathbf{f}_k} \mathbf{A}\mathbf{f}_k \tag{A2}$$

subject to

$$\|\mathbf{f}_k\| = 1 \tag{A3}$$

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

$$\mathbf{E}_k = \mathbf{K}^T \mathbf{F}_k \mathbf{K} \tag{A4}$$

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 $k$:

$$\mathbf{E}_k = \mathbf{U}\mathbf{D}\mathbf{V}^T$$
$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A5}$$
$$\mathbf{R}_k = \mathbf{U}\mathbf{W}\mathbf{V}^T \text{ or } \mathbf{R}_k = \mathbf{U}\mathbf{W}^T\mathbf{V}^T$$
$$\mathbf{t}_k = \mathbf{U}(0,0,1)^T \text{ or } \mathbf{t}_k = -\mathbf{U}(0,0,1)^T$$

where there is one correct combination of $\mathbf{R}_k$ and $\mathbf{t}_k$ which can make all $\mathbf{X}^r_k$ 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 $\mathbf{X}^r_k$ are finally obtained by the triangulation $f_t(\cdot)$:

$$\mathbf{x}_{k-1}^{r} = \mathbf{P}_{k-1}\mathbf{X}_{k}^{r} = \mathbf{K}[\mathbf{I}, \mathbf{0}]\mathbf{X}_{k}^{r}$$
$$\mathbf{x}_{k}^{r} = \mathbf{P}_{k}\mathbf{X}_{k}^{r} = \mathbf{K}[\mathbf{R}_{k}, \mathbf{t}_{k}]\mathbf{X}_{k}^{r} \tag{A6}$$

$$\mathbf{X}_{k}^{r} = f_{t}(\mathbf{K}, \mathbf{R}_{k}, \mathbf{t}_{k}, \mathbf{x}_{k}^{r}, \mathbf{x}_{k-1}^{r}) \tag{A7}$$

## References

1. Mechanical Vibration and Shock-Evaluation of Human Exposure to Whole-Body Vibration-Part 1: General Requirements. Available online: https://www.iso.org/standard/76369.html (accessed on 30 April 1997).
2. Tighe, S.; Li, N.; Falls, L.; Haas, R. Incorporating road safety into pavement management. *Transp. Res. Rec. J. Transp. Res. Board* **2000**, *1699*, 1–10. [CrossRef]
3. Huang, Y.; Xu, B. Automatic inspection of pavement cracking distress. *J. Electron. Imaging* **2006**, *15*, 013017. [CrossRef]
4. Herold, M.; Roberts, D. Spectral characteristics of asphalt road aging and deterioration: Implications for remote-sensing applications. *Appl. Opt.* **2005**, *44*, 4327–4334. [CrossRef] [PubMed]
5. Saarenketo, T.; Scullion, T. Road evaluation with ground penetrating radar. *J. Appl. Geophys.* **2000**, *43*, 119–138. [CrossRef]
6. Bursanescu, L.; Blais, F. Automated pavement distress data collection and analysis: A 3-D approach. In Proceedings of International Conference on Recent Advances in 3-D Digital Imaging and Modeling (Cat. No.97TB100134), Ottawa, ON, Canada, 12–15 May 1997; pp. 311–317.
7. Kil, D.H.; Shin, F.B. Automatic road-distress classification and identification using a combination of hierarchical classifiers and expert systems-subimage and object processing. In Proceedings of International Conference on Image Processing, Santa Barbara, CA, USA, 26–29 October 1997; pp. 414–417.
8. Fukuhara, T.; Terada, K.; Nagao, M.; Kasahara, A.; Ichihashi, S. Automatic pavement-distress-survey system. *J. Transp. Eng.* **1990**, *116*, 280–286. [CrossRef]
9. Yu, B.X.; Yu, X. Vibration-based system for pavement condition evaluation. *Appl. Adv.Tech. Trans.* **2006**, 183–189. [CrossRef]
10. Vittorio, A.; Rosolino, V.; Teresa, I.; Vittoria, C.M.; Vincenzo, P.G.; Francesco, D.M. Automated sensing system for monitoring of road surface quality by mobile devices. *Procedia Soc. Behav. Sci.* **2014**, *111*, 242–251. [CrossRef]
11. Tai, Y.C.; Chan, C.W.; Hsu, J.Y.j. Automatic road anomaly detection using smart mobile device. In Proceedings of the conference on technologies and applications of artificial intelligence, Hsinchu, Taiwan, 25–27 January 2010.
12. Eriksson, J.; Girod, L.; Hull, B.; Newton, R.; Madden, S.; Balakrishnan, H. The pothole patrol: Using a mobile sensor network for road surface monitoring. In Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services (ACM), Breckenridge, CO, USA, 17–20 June 2008; pp. 29–39. [CrossRef]
13. Xue, G.; Zhu, H.; Hu, Z.; Yu, J.; Zhu, Y.; Luo, Y. Pothole in the dark: Perceiving pothole profiles with participatory urban vehicles. *IEEE Trans. Mob. Comput.* **2017**, *16*, 1408–1419. [CrossRef]
14. Mednis, A.; Strazdins, G.; Zviedris, R.; Kanonirs, G.; Selavo, L. Real time pothole detection using android smartphones with accelerometers. In Proceedings of the 2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS), Barcelona, Spain, 27–29 June 2011; pp. 1–6.
15. Tedeschi, A.; Benedetto, F. A real-time automatic pavement crack and pothole recognition system for mobile Android-based devices. *Adv. Eng. Inf.* **2017**, *32*, 11–25. [CrossRef]
16. Koch, C.; Brilakis, I. Pothole detection in asphalt pavement images. *Adv. Eng. Inf.* **2011**, *25*, 507–515. [CrossRef]
17. Koch, C.; Jog, G.M.; Brilakis, I. Automated pothole distress assessment using asphalt pavement video data. *J. Comput. Civil Eng.* **2012**, *27*, 370–378. [CrossRef]
18. Jo, Y.; Ryu, S. Pothole detection system using a black-box camera. *Sensors* **2015**, *15*, 29316–29331. [CrossRef] [PubMed]
19. Banharnsakun, A. Hybrid ABC-ANN for pavement surface distress detection and classification. *Inter. J. Mach. Learn. Cybern.* **2017**, *8*, 699–710. [CrossRef]

20. Ryu, S.K.; Kim, T.; Kim, Y.R. Image-based pothole detection system for ITS service and road management system. *Math. Prob. Eng.* **2015**, *2015*. [CrossRef]

21. Chang, K.; Chang, J.; Liu, J. Detection of pavement distresses using 3D laser scanning technology. *J. Comput. Civ. Eng.* **2005**, 1–11. [CrossRef]

22. Yu, S.J.; Sukumar, S.R.; Koschan, A.F.; Page, D.L.; Abidi, M.A. 3D reconstruction of road surfaces using an integrated multi-sensory approach. *Opt. Lasers Eng.* **2007**, *45*, 808–818. [CrossRef]

23. Yu, X.; Salari, E. Pavement pothole detection and severity measurement using laser imaging. In Proceedings of the 2011 IEEE International Conference on Electro/Information Technology, Mankato, MN, USA, 15–17 May 2011; pp. 1–5.

24. Hou, Z.; Wang, K.C.; Gong, W. Experimentation of 3D pavement imaging through stereovision. In Proceedings of the International Conference on Transportation Engineering 2007, Chengdu, China, July 22–24 2007; pp. 376–381. [CrossRef]

25. Fan, R.; Ozgunalp, U.; Hosking, B.; Liu, M.; Pitas, I. Pothole detection based on disparity transformation and road surface modeling. *IEEE Trans. Image Process.* **2019**, *29*, 897–908. [CrossRef] [PubMed]

26. El Gendy, A.; Shalaby, A.; Saleh, M.; Flintsch, G.W. Stereo-vision applications to reconstruct the 3D texture of pavement surface. *Int. J. Pavement Eng.* **2011**, *12*, 263–273. [CrossRef]

27. Ahmed, M.; Haas, C.; Haas, R. Toward low-cost 3D automatic pavement distress surveying: The close range photogrammetry approach. *Can. J. Civ. Eng.* **2011**, *38*, 1301–1313.

28. Antol, S.; Ryu, K.; Furukawa, T. A New Approach for Measuring Terrain Profiles. In Proceedings of the ASME 2013 International Design Engineering Technical Conferences And Computers and Information in Engineering Conference, Portland, OR, USA, 4–7 August 2013. [CrossRef]

29. Moazzam, I.; Kamal, K.; Mathavan, S.; Usman, S.; Rahman, M. Metrology and visualization of potholes using the microsoft kinect sensor. In Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, 6–9 October 2013; pp. 1284–1291.

30. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Maynila, Philippines, 2003.

31. Hu, Y.; Furukawa, T. A High-Resolution Surface Image Capture and Mapping System for Public Roads. *SAE Int. J. Passenger Cars Electron. Electr. Syst.* **2017**, *10*, 301–309. [CrossRef]