Automatic Detection and Evaluation of 3D Pavement Defects Using 2D and 3D Information at the High Speed

Yazhe Hu 1)  Tomonari Furukawa 1)

1) Department of Mechanical Engineering, Virginia Tech Blacksburg, VA, USA
(E-mail: yazhehu@vt.edu) (E-mail: tomonari@vt.edu)

Received on October 12, 2017

ABSTRACT: This paper presents a technique of automatic 3D pavement defects detection using both the two-dimensional (2D) and three-dimensional (3D) information from the images captured at high speed. Scaled 3D points reconstructed from Structure from Motion algorithm are first used to detect the defect regions based on the 3D information. A mismatched points rejection method then eliminates incorrectly matched points and reconstructs a final 3D road surface. The proposed technique uses both 3D and 2D information for the defects detection where the characteristics of defect regions in 2D images together with the 3D information from the 3D reconstruction detect the defect region. The capability of the proposed technique was first investigated through parameters studies. Quantitative analyses have shown the accuracy, precision, and recall rate of the proposed technique are all above 90%. The result demonstrates the potential of the proposed technique for the automatic detection of 3D pavement defects.

KEY WORDS: information, communication, and control, automatic road defects detection, pavement surface 3D reconstruction, structure from motion 

1. Introduction

The road is the most fundamental and important infrastructure for motor vehicles to travel. An intact road condition can reduce the damage to the vehicle as well as the potential traffic accidents. Roads however keep developing defects inevitably by the heavy usage and the environmental inputs, and a tremendous amount of the pavement surface in the United States and in the European Union have been reported to be in poor condition accordingly. Since the currently dominating human inspection is subject to inefficiency, subjectivity, and high cost, automatic detection of pavement defects is becoming a key breakthrough for the effective road inspection.

Past works on the automatic pavement defects detection can be classified into three types: the vibration-based method, the two-dimensional (2D) image-based method, and the three-dimensional (3D) reconstruction-based method. The vibration-based methods apply an accelerometer on the vehicle and detect pavement defects using the vibration information. Eriksson et al. proposed the road defects detection by using the machine learning, Support Vector Machine (SVM), and unsupervised learning respectively on accelerometer data collected by the taxi and motorcycle. Mednis et al. utilized the accelerometer in the smart-phone with Android OS, and used several signal processing algorithms to find the pattern of road defects such as the pothole. Xue et al. proposed a method also using the accelerometer in smart-phone, but adopted a one degree-of-freedom model to estimate the depth and size of the pothole on the road. Although the vibration-based methods require simple and cheap accelerometer sensors attached to the vehicle, they can only detect road defects when the vehicle (tire) passes above the defect regions. Meanwhile, different kinds of noise to the accelerometer will also affect the detection accuracy.

Methods based on using 2D image information are applied on automatic pavement defects detection. Koch et al. at first segmented the road images into defect and non-defect regions, and then detected the defects based on the statistical measurements in both defect and non-defect regions. Later on they extended the previous method from single frames to continuous frames detection, and keep updating the texture information of defect and non-defect regions while tracking the defects over frames. Jo et al. utilized a regular camera to continuously capture the road surface when the vehicle was moving. Then noises were removed by the lane detection which confined the road surface areas, and the defects were detected based on the color and size. Doycheva et al. implemented the wavelet transform to detect pavement distress. By using a graphical processing unit, a real-time road distress detection was achieved. Azhar et al. adopted the histograms of oriented gradients (HOG) to describe the feature in images, and used Naïve Bayes classifier to classify pothole and non-pothole images. They also implemented graph cut algorithm to segment pothole locations in the road surface images. The 2D image-based road defects detection methods use image information to detect defects on the road, and mounting the cameras to the vehicle is not difficult. However, the accuracy of detecting pavement defects by using only the image processing can be affected by the shadows, oil stains, and patches on the pavement surface.

Road defects can also be detected based on 3D reconstruction-based methods. Chang et al. used a 3D laser scanner to detect road defects based on the topological information given by the 3D scanning data. Hou et al. and El Gendy et al. obtained 3D data of road surface based on stereo vision and the road defects are detected based on the 3D reconstruction of the road surface. Ahmed et al. applied 3D reconstruction based on multiple-view images and the reconstruction of the pavement surface was done by also using artificial features marked on the pavement. Antol et al. and Moazzam et al. reconstructed road surface with Kinect RGB-D sensor which was a 3D area scanner. Then the road defects could be detected based on the depth and size information. The laser scanner method and the stereo vision method are influenced by the vibration when the vehicle they were mounted was moving at a high speed. Meanwhile, the previous work of multiple-view images 3D reconstruction method was a static method. Also, the RGB-D cameras worked well in the indoor at static but could not work at the outdoor environment and...
when the RGB-D cameras were moving. These methods are not suitable for defects detection at a high speed since the fast moving traffic on the road requires a fast moving speed of the road defects measurement system.

This paper presents a technique of automatic 3D road defects detection using both 2D and 3D information with images captured at the high speed. A CMOS camera with the high shutter speed captures the road surface images without motion blur at high moving speed. Two consecutive images are captured with an overlapping region. 3D scaled points are reconstructed in the overlapping regions using structure from motion, and the road surface is represented by 3D scaled points. The error in 3D points reconstruction caused by the feature mismatching is removed by using camera’s motion model. Then, the potential road defect regions are classified based on the 3D information. Since the proposed technique obtains both the 2D image and 3D structure of the pavement surface, the road defect regions are detected and evaluated by using the characteristics of road defects in both 2D images and 3D points.

This paper is organized as follows. The following section introduces the camera motion and road surface models as well as reviews structure from motion algorithm. Section 3 describes the proposed technique which uses both 3D and 2D information to automatically detect pavement defects. Section 4 demonstrates the efficacy by the experimental results. Conclusions are summarized in the last section.

2. 3D Points Reconstruction from One Camera

2.1. Camera motion and road surface models

The models for a single-camera motion and the road surface are shown in Fig. 1. $x_k^c \in \mathcal{X}$ is the position of the camera when the kth image is captured. The camera is set to be vertically facing towards the road surface at a certain height, and it continuously moves from position k to the next position k+1 while captures an image at each position. $[u_k, v_k]^T \in \mathcal{M}$ represents the image pixel coordinate of a pixel in the kth image. The image $k$ and image $k+1$ have an overlapping region demonstrated in Fig. 1 as the region with shadow. Based on the structure from motion algorithm the three-dimensional point cloud can be reconstructed in the overlapping region:

$$x_k^r = f(u_k, v_k, K)$$ (1)

where $x_k^r \in \mathcal{X}$ is the ith 3D point of the road reconstructed by the image k and the image k+1, and $K$ is the camera’s intrinsic matrix obtained based on the camera calibration. The $x_k^r$ contains the 3D information of the points within the overlapping region and will be used to automatically detect pavement defects. The structure from motion algorithm which is used to reconstruct point cloud $x_k^r$ will be presented in the following subsection.

The defects in the pavement surface can be defined as two types: the 2D defect and the 3D defect. The 2D defect only has pattern in the x and y direction, such as thin cracks on the road surface which do not differ from the flat surface region in the z direction. However, the 3D defect has not only the pattern in the x and y direction but also in the z direction such as potholes. In this paper, the automatic defects detection technique focuses on the detection of the 3D defects.

2.2. Structure from motion

From Hartley and Zisserman\cite{18}, structure-from-motion algorithm can reconstruct scaled 3D points from two views with one camera. Camera intrinsic parameter matrix $K$ is first obtained by the camera calibration:

$$K = \begin{bmatrix} a_u & s & u_0 \\ 0 & a_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$ (2)

where $a_u$ and $a_v$ are the number of pixels along u and v direction, $s$ is the parameter of skew, and $[u_0, v_0]$ is the principle point of the camera. The 3x3 essential matrix $E$ is then given by:

$$\hat{x}^T E \hat{x}' = 0$$ (3)

$$\hat{x} \sim K^{-1} \hat{u}$$ (4)

where $\hat{x} = [\hat{x}, \hat{y}, 1]^T$ is the homogeneous 2D point in camera’s image coordinate converted from the camera’s pixel coordinate of the kth image, and $\hat{x}'$ is in the homogeneous 2D point in the $(k+1)$th image, $\hat{u} = [u, v, 1]^T$ is the homogeneous 2D point in the camera’s pixel coordinate. Equations 3 and 4 can also be rewritten as:

$$\hat{u}' (K^{-1} E K^{-1}) \hat{u}' = 0$$ (5)

$$F \sim K^{-1} E K^{-1}$$ (6)

where $F$ is called fundamental matrix. From equation 5 $F$ can be calculated given 8 or more corresponding points $(\hat{u}, \hat{u}')$, and essential matrix can be recovered as:

$$E \sim K^T F K$$ (7)

then the rotation $R$ and transformation $T$ can be obtained after computing the singular value decomposition of $E$:

$$E \sim U A V^T$$ (8)

$$[T]_x = U \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} U^T$$ (9)

$$R = U \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} V^T$$ (10)

where $A$ is the diag($\sigma_1, \sigma_2, \sigma_3$) and $U$ and $V$ are orthogonal matrix, $[T]_x$ is the cross product matrix. Projection matrix $P$ can transform 3D homogeneous points into 2D homogeneous points, and it can be obtained after getting $R$ and $T$:

$$P = K [R \mid T]$$ (11)

at last step, triangulation can reconstruct 2D corresponding points between two overlapping images to 3D points. Ideally the line passing through the first camera center and a feature point should intersect with the other line passing through the second camera center and the corresponding feature point. However, in reality
usually they will not intersect. So a minimization objective function is applied to find the best 3D reconstruction:

$$\mathbf{x}_k^r = \arg \min_{\mathbf{x}} \Sigma \| \mathbf{u}_k - \mathbf{u}_k(P_0, \mathbf{x}_k^r) \|^2$$  \hspace{1cm} (12)

where \( \mathbf{u} \) is the projected pixel coordinate from \( \mathbf{x}_k^r \). The \( \mathbf{x}_k^r \) reconstructed by structure from motion algorithm is the up-to-scale points. Scaled 3D point means the distance between points and camera poses are based on relative positions, which are not the real distance in reality. The reason for the points to be scaled is that the distance between \( \mathbf{x}_k^r \) and \( \mathbf{x}_{k+1}^r \) are unknown for the camera movement. So a 1 unit is always set as the distance between \( \mathbf{x}_k^r \) and \( \mathbf{x}_{k+1}^r \) thus it results in the scaled 3D points \( \mathbf{x}_k^r \).

### 3. Automatic Pavement Defects Detection

#### 3.1. Road surface representation of scaled 3D points

The scaled 3D points \( \mathbf{x}_k^r = [x_k^r, y_k^r, z_k^r]^T \) represent all the reconstructed 3D points by using the \( k \)th image and \((k+1)\)th image. Since the points are scattered, fitting a surface among the points can help better describe the pavement surface condition and detect defects on the road surface. The quadratic polynomial surface fitting is given as:

$$f_z = c_0 + c_1 x + c_2 y + c_3 x^2 + c_4 y^2 + c_5 x y + c_6 x^3 + c_7 y^3 + c_8 x^2 y + c_9 y^2 x + c_{10} x y^2 + c_{11} y x^2 + c_{12} x^2 y^2 + c_{13} y^2 x^2 + c_{14} x y^3 + c_{15} y x^3 + c_{16} x^3 y + c_{17} y x^3 + c_{18} x y^3 + c_{19} + c_{20}$$

where \( c \) are coefficients. The other surface fitting models such as using linear polynomial surface and the higher order polynomial surface will be compared and discussed in Section 4.

The coefficients \( c_0, c_1, ..., c_{20} \) are calculated based on the least squares method which is given as:

$$e = \arg \min \| [c_0 + c_1 x_i + c_2 y_i + c_3 x_i^2 + c_4 y_i^2 + c_5 x_i y_i + c_6 x_i^3 + c_7 y_i^3 + c_8 x_i^2 y_i + c_9 y_i^2 x_i + c_{10} x_i y_i^2 + c_{11} y_i x_i^2 + c_{12} x_i^2 y_i^2 + c_{13} y_i^2 x_i^2 + c_{14} x_i y_i^3 + c_{15} y_i x_i^3 + c_{16} x_i^3 y_i + c_{17} y_i x_i^3 + c_{18} x_i y_i^3 + c_{19} + c_{20}] \|^2$$

and the Equation 15 solves the optimal solutions for the coefficients:

$$[1 \Sigma x_1 \Sigma y_1 \Sigma x_1 y_1 \Sigma x_1^2 \Sigma y_1^2 \Sigma x_1 y_1^2 \Sigma y_1^2 \Sigma x_1^2 y_1^2 \Sigma x_1^3 \Sigma y_1^3 \Sigma x_1^2 y_1^3 \Sigma x_1 y_1^3 \Sigma y_1^3]^T \begin{bmatrix} c_0 \cdots c_{20} \end{bmatrix}$$

and the Equation 15 solves the optimal solutions for the coefficients:

$$[1 \Sigma x_1 \Sigma y_1 \Sigma x_1 y_1 \Sigma x_1^2 \Sigma y_1^2 \Sigma x_1 y_1^2 \Sigma y_1^2 \Sigma x_1^2 y_1^2 \Sigma x_1^3 \Sigma y_1^3 \Sigma x_1^2 y_1^3 \Sigma x_1 y_1^3 \Sigma y_1^3]^T \begin{bmatrix} c_0 \cdots c_{20} \end{bmatrix}$$

by substituting \( \mathbf{x}_k^r = [x_k^r, y_k^r, z_k^r]^T \) into Equation 15, the corresponding coefficients of \( e \) can be calculated for the scaled 3D points of the road surface.

#### 3.2. Defects detection using 3D information

After a surface is fitted to all the points \( \mathbf{x}_k^r \), the relative positions of \( \mathbf{x}_k^r \) and the surface can be recognized. The 3D points of road surface can be classified as two groups: \( \mathbf{x}_k^r \) are the points representing the non-defect road surface, while \( \mathbf{x}_k^r \) are the points representing the defect region of the road surface. Usually by knowing the 3D information of the points, the distances between \( \mathbf{x}_k^r \) and the fitted surface are larger than the ones between \( \mathbf{x}_k^r \) and the fitted surface, and by setting several thresholds \( \mathbf{x}_k^r \) can be differentiated from \( \mathbf{x}_k^r \). However, since \( \mathbf{x}_k^r \) are up-to-scale 3D points instead of absolute ones, simply setting the thresholds to \( \mathbf{x}_k^r \) cannot work because the “scales” are not the same for \( \mathbf{x}_k^r \) at different values of \( k \).

In this case, the distance between the camera and the fitted surface is calculated to help classifying non-defect and defect points:

$$d_k^r = |c_0 + c_1 x_k^r + c_2 y_k^r + c_3 x_k^r y_k^r + c_4 x_k^r y_k^r + c_5 y_k^r + c_6 y_k^r^2 + z_k^r|$$

where \( d_k^r \) is the camera to fitted surface distance with the \( k \)th reconstructed 3D points. Similarly, \( \mathbf{x}_k^r \) represents the distances between the 3D points \( \mathbf{x}_k^r \) and the fitted surface which is given by:

$$d_k^r = |c_0 + c_1 x_k^r + c_2 y_k^r + c_3 x_k^r y_k^r + c_4 x_k^r y_k^r + c_5 y_k^r + c_6 y_k^r^2 + z_k^r|$$

since the height of the camera towards the road surface can be regarded as a constant, so by using \( d_k^r \) as a reference distance the ratio of \( d_k^r / d_k \) can be used to classify the 3D points to be defect or non-defect points. The points classification is given by:

$$\mathbf{x}_k^r \in \begin{cases} \{(D)\mathbf{x}_k^r, \text{ if } d_k^r < 0 \text{ and } d_k^r/d_k^r > T_{HP} \} \\ \{(N)\mathbf{x}_k^r, \text{ else } \} \end{cases}$$

where \( T_{HP} \) is a threshold to classify a point based on the ratio of its \( d_k^r/d_k \) and \( d_k^r \). If the ratio is greater than the threshold and the point \( \mathbf{x}_k^r \) locates below the road surface region, the point can be categorized as a defect point.

#### 3.3. Hybrid road defects detection technique

Fig. 2 shows the hybrid road defects detection technique which combines the camera’s motion model, the defect features in the 2D image, and the 3D information to refine the detection. The first step is the reconstruction of 3D points from 2D images which contain an overlapping region. The details are already shown in Section 2. The next step is the defect points detection by using the 3D information. Before the classification of points based on 3D information, the rejection of the mismatched points will help improving the 3D points reconstruction. The accuracy of the 3D points classification can then be improved by combining features of the defect points from 2D images, and finally the defects clustering process determines the defect regions and quantitatively estimates the severity of the current pavement defects.

#### 3.3.1 Mismatched points rejection

The corresponding feature points matching in two consecutive images’ overlapping region uses the SIFT feature\(^{(19)}\). Each feature point has a descriptor which is a 128-element vector describing itself. The Euclidean distance is used to match the corresponding features in the \( k \)th image and the \((k+1)\)th image, and a score is given to decide the matched points. The smaller the score, the better the matching. However, there may exists some mismatched points among the matched points. If by only setting a constant threshold, it is difficult to remove the mismatched points since if the threshold is set to be too small, there will be less corresponding points accepted as a correct matching, which also
means less reconstructed 3D points. In order to get denser 3D points to represent the road surface, a larger threshold could be used. But the problem of using a large threshold is that the mismatched points will also appear in the matching points, which will result in a wrong reconstruction of 3D points for road surface.

The camera’s motion model can be used to remove the wrong matching. Since in \(x_k^c = [x_k^c, y_k^c, z_k^c]^T\), the camera is mounted at a certain height from the road surface (\(z_k^c\) can be treated as a constant), the camera’s motion from \(x_k^c\) to \(x_{k+1}^c\) can be represented as moving in a 2D \(x_k^c,y_k^c\) plane. Therefore, every corresponding point found in \(\mu_x\) should have a similar motion with other corresponding points:

\[
d_{kj}^c = x_{k+1}^c - x_k^c
\]

\[
d_{kj}^r = u_{k+1,j} - u_{k,j}
\]

where \(d_{kj}^c\) is the distance between two camera locations, \(u_{k,j}\) and \(u_{k+1,j}\) represent the \(j\)th feature point in image \(k\) matches the \(j\)th feature point in image \(k+1\), and \(d_{kj}^r\) is the distance between two matching points. Based on the distribution of all the matching points, \(d_{kj}^f\), \(t_1\) and \(t_2\) represent the lower quartile and the upper quartile of \(d_{kj}^r\), respectively. Since it is not practical to get the camera’s motion between state \(k\) and \(k+1\), the correct matching is defined as:

\[
(c_d_k^f = \{d_{kj}^f | t_1 - m \cdot 1QR < d_{kj}^f < t_2 + m \cdot 1QR\}
\]

where \(c_d_k^f\) represents the correct matching points, 1QR is the interquartile range which is calculated by \(t_2 - t_1\), \(m\) is a constant, and the mismatched points are removed by two threshold \(t_1 - m \cdot 1QR\) and \(t_2 + m \cdot 1QR\). By doing this, correct feature matchings are reserved using two thresholds which are unique between two consecutive images. The assumption in this section is that most of the \(d_{kj}^f\) are correct matching so that the lower quartile and upper quartile of the \(d_{kj}^r\) can be used as the threshold values.

The function \(\text{CORRECT DEFECTS}()\) takes two inputs \(l_k\) and \(\{x_k^r\}\), where \(l_k\) is the \(k\)th image captured by the camera. At line 1 and line 2 (L1 and L2) two parameters are initialized in which \(\{x_k^r\}\) represents the correct defect points and \(length(\{x_k^r\})\) is the number of points contained by \(\{x_k^r\}\). Then L3 convert \(l_k\) to the grayscale image \(l_k\text{gray}^r\). L4 conducts the image binarization using a threshold \(Th_b\), which means for every pixel if its intensity is less than \(Th_b\), its value becomes 1, and otherwise 0. From L5 to L6 the \(\text{erosion}\) and \(\text{dilation}\) are common operations to erode the image first, and then dilate it to remove the noise. Since inside the 3D defect region of the road surface it always has the shadow under natural lighting condition, L5 to L6 are the processes to find the shadow in the image which potentially appear together with defects. Since the shadow region (where \(l_k\text{binary} = \text{1}\) ) and its vicinity may contain defect, the \(\text{dilation}\) will enlarge the area of \(l_k\text{binary} = \text{1}\) in L7 to L10 keeps finding the correct defect points based on the updated 2D defect region. Although \(\{x_k^r\}\) are 3D points, they can be mapped from 3D to their 2D corresponding points in the image. So L9 checks the corresponding 2D positions of \(\{x_k^r\}\) and returns the correct defect points which are located in the region of \(l_k\text{binary} = \text{1}\), also called potential defects region (PDR). In L10, the \(l_k\text{binary-update}(\{x_k^r\})\) function update the PDR using the new defect points \(\{x_k^r\}\) by the same dilation procedure used in L6. One of the dilation structuring element is the square with width of \(W\), which chases the \(l_k\text{binary}\) Values to 1 at all the pixels within this square if the center pixel value is originally 1. Finally, if no more new \(\{x_k^r\}\) are found through this process, \(\{x_k^r\}\) is returned as the correct defect points.

### 3.3.2 Defects detection using both 2D and 3D information

Based on the 3D information of all the reconstructed points, the potential defect points \(\{x_k^r\}\) can be found. Although in Section 3.3.1 the mismatched points have already been removed from \(x_k^r\), there are still some points \(\{x_k^r\}\) in \(\{x_k^r\}\) which do not represent the true defect region on road surface because of the 3D reconstruction error.

So in order to remove the \(\{x_k^r\}\), the 2D image information of the road surface is applied together with the 3D information given in points \(\{x_k^r\}\). The defect points corrected by both the 2D information and the 3D information is given below:

\[
\text{CORRECT DEFECTS}(l_k, \{x_k^r\})
\]

1. \(length.pre \leftarrow -1\)
2. \(length(\{x_k^r\}) \leftarrow 0\)
3. \(l_k\text{gray} \leftarrow \text{grayscale}(l_k)\)
4. \(l_k\text{binary} \leftarrow l_k\text{gray} < Th_b\)
5. \(l_k\text{binary-erosion}\)
6. \(l_k\text{binary-dilation}\)
7. \(\text{while} length(\{x_k^r\}) > length.pre\)
8. \(length.pre \leftarrow length(\{x_k^r\})\)
9. \(\{x_k^r\} \leftarrow l_k\text{binary-inside}(\{x_k^r\})\)
10. \(l_k\text{binary-update}(\{x_k^r\})\)
11. \(\text{Return } \{x_k^r\}\)

### 4. Experimental Results

In the authors’ previous work, a system which could capture high-resolution road surface images at a high speed had been built. Fig. 3 shows the road surface capturing system. The camera is vertically facing towards the road surface, and the images captured by the camera are saved in the PC on the vehicle.
surface images are captured with only one of the cameras in the system when the vehicle is driving on the public road. The camera’s frame rate is set to be 100 Hz and the shutter speed is 0.04 ms. The height of the camera towards the road surface is 1.4 m. Since the camera’s frame rate and shutter speed are all set to be high, the camera captures images with no motion blur and more than 50% overlapping region between two consecutive images at up to 100 km/h driving speed. It was discovered that 50% to 80% overlapping region would be reasonable for the algorithm.

This section presents three experiments to evaluate the efficacy of the proposed technique of automatic 3D pavement defects detection at the high speed. In the first experiment, different surface fitting models based on the 3D points of the road surface are compared by their performance using the labeled images. The second experiment deals with the mismatched points rejection and the choice of parameters. The last subsection shows examples of using the proposed automatic pavement defects detection technique on five typical road surface images, and then analyze the performance of the technique based on 3500 road surface images.

### 4.1. Evaluation of surface fitting on scaled 3D points

The surface fitting to the scaled 3D points were examined by the linear, quadratic and cubic surface fitting models. The quadratic model was already shown in Equation 13, and the linear and cubic models were given as:

\[
f_x = c_0 + c_1 x + c_2 y + z = 0
\]

\[
f_y = c_0 + c_1 x + c_2 y + c_3 x^2 + c_4 y^2 + c_5 x^2 y + c_6 x y^2 + c_7 y^3 + c_8 x^3 + c_9 y^3 + z = 0
\]

and these three models were compared when they were applied to the surface fitting of the scaled 3D points in order to find the appropriate surface fitting model. 100 images labeled as flat road surface without defects were used to evaluate the performance of each surface fitting model.

Fig. 4 shows the mean value and the standard deviation of \( \frac{d_{k+1}^f}{d_k^f} \) among the 100 labeled image. From Equations 16, 17, and 18, \( \frac{d_{k+1}^f}{d_k^f} \) represents the ratio between the distance of points to fitted surface and the distance of the surface to the camera. Since the \( \frac{d_k^f}{d_k^s} \) is almost unchanged, the smaller the \( \frac{d_{k+1}^f}{d_k^f} \), the closer the points are to the fitted surface. \( \frac{d_{k+1}^f}{d_k^f} \) can be shown from Equation 16 and 17:

\[
\frac{d_{k+1}^f}{d_k^f} = \frac{|c_0 + c_1 x_k + c_2 y_k + c_3 x_k^2 + c_4 y_k^2 + c_5 x_k^2 y_k + c_6 x_k y_k^2 + c_7 y_k^3|}{c_0 + c_1 x_k + c_2 y_k + z_k}
\]

which shows that if \( \frac{d_k^f}{d_k^s} \) is close to zero, the points are close to the surface \( f_x \). It can be found from the plot that the mean values and the standard deviation of the \( \frac{d_{k+1}^f}{d_k^f} \) using the linear surface fitting model are larger than the ones using the quadratic or cubic models.

Meanwhile, the performance of using the quadratic and cubic surface fitting models are equally matched. So using the quadratic surface fitting model as Equation 13 is reasonable.

### 4.2. Evaluation of mismatched points rejection

A mismatched points rejection method was evaluated based on two consecutive images. From Equation 21, the detected correct matched points were given by \( m^f_{k+1} \). Therefore, the detected mismatched points were given as:

\[
\cap_{a}^f \cap_{b}^f = \{ d_k^f | d_k^f < (t_1 - m \cdot 1QR) \cap d_k^f > (t_2 + m \cdot 1QR) \}
\]

where \( \cap_{a}^f \cap_{b}^f \) represented the detected mismatched points.

Fig. 5 shows the mismatched points rejection process and its performance by analyzing two consecutive images. Fig. 5(a) and 5(b) illustrates the 4th image and the \((k+1)\)th image captured by the camera. Fig. 5(c) demonstrates the feature points matching between the 4th image and the \((k+1)\)th image. It can be found that there are some obvious mismatched points among the matching. Fig. 5(d) shows the detected mismatched points based on Equation 24 where \( m = 3 \). The red rectangular areas contain the correct matched points although they are detected as mismatched points. Fig. 5(e) then evaluates the performance of mismatched points rejection by using different \( m \). The total number of all the matched feature points is 4362 and the number of true mismatched points in the image is 32. From \( m = 1 \) to 4.5, all the 32 true mismatched points are detected and the number of
matched points among the detected mismatched points \( |N_d|^k \) is decreasing. When \( m = 5 \), the number of true mismatched points becomes 30 which means 2 of them are not detected by Equation 24. So a smaller \( m \) can detect all the true mismatched points but will classify more matched points into \( |N_d|^k \) which makes the 3D reconstruction points sparser. Meanwhile, a larger \( m \) can reduce the number of matched points in the \( |N_d|^k \) while the detection for true mismatched points may be missed, which will lead to the 3D reconstruction failure. Since starting from \( m = 1.5 \) 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. So from Fig. 5(e) the value of \( m \) can be between 1.5 to 2.5 in order to keep more points meanwhile remove all the mismatched points.

4.3. Performance of road defects detection based on both 2D and 3D information

The proposed automatic road detects detection technique was illustrated through several examples at first. Then 3500 road surface images captured at the high speed were processed using the proposed automatic defects detection technique to analyze the performance. There were 308 images which contained defect regions and 3192 images represented flat road surface. In the experiment, the \( Th_p \) which has a value in the range of (0, 255) from the function \( CORRECT\_DEFECTS() \) was set to be 25, and the value of \( m \) from Equation 21 was 1.5. Meanwhile, from Equation 18 \( Th_p \) was 0.0071 in order to detect defects with more than 10 mm in depth. The performance of the proposed technique was also compared with other existing methods.

Fig. 6 shows the automatic 3D defects detection process using both 2D and 3D information on five typical scenarios. Fig. 6(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 area of the defect. The last one is the flat road surface and has no defect in it. Then Fig. 6(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. Then in Fig. 6(c), all the detected defects \( D^k \) are registered at the corresponding position in the image. It can be found that although most of the \( D^k \) are located in the defect region, there are still several \( D^k \) which are located not at the defect region. Therefore in Fig. 6(d), the binary 2D image is also used together with the information from 3D points. The function \( CORRECT\_DEFECTS() \) is applied to refine the defects detection by using both 2D and 3D information of the road surface. Finally, Fig. 6(e) demonstrates the refined defects detection results. It shows the removal of the incorrect \( D^k \) keeps and the true defect points in \( D^k \).

Table 1 shows the comparison of performance between using only 3D information and using both 2D and 3D information. Based on the values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), the accuracy is defined as \( (TP+TN)/(TP+TN+FP+FN) \), while precision is defined by \( TP/(TP+FP) \) and recall is represented by \( TP/(TP+FN) \). For the use of only 3D information, the recall is 97.7%, the precision is 64.1%, and the accuracy is 95.0%. However, when the proposed technique which uses both 2D and 3D information is applied, the recall is still 97.7% but the precision becomes 92.9% and the accuracy increases to 99.1%. The accuracy, precision and recall rate for the proposed method and the 3D only method are listed in Table 2.

Table 1. The performance of the proposed technique comparing with the technique using only the 3D information.

| Detected | Measured (3D only) | Measured (Proposed 2D and 3D combination) |
|----------|-------------------|------------------------------------------|
| Pothole (positive) | 301 (TP) | 301 (TP) |
| No Pothole (negative) | 7 (FN) | 7 (FN) |
| Pothole (positive) | 168 (FP) | 23 (FP) |
| No Pothole (negative) | 3024 (TN) | 3169 (TN) |
Table 2. Accuracy, precision, and recall rate for proposed method and 3D only method.

| Method                             | Accuracy | Precision | Recall |
|------------------------------------|----------|-----------|--------|
| Measured (3D only)                 | 95.0%    | 64.2%     | 97.7%  |
| Measured (Proposed 2D and 3D combination) | 99.1%    | 92.9%     | 97.7%  |

Table 3 shows the comparison between the proposed technique and the techniques of Eriksson [2], Koch [7], and Jo [8]. The ‘-’ in the table means “not available” in the corresponding paper. Although they do not provide their dataset and this comparison are based on different dataset, the proposed technique can still be found as achieving more than 90% accuracy, precision, and recall rate.

| Proposed | Eriksson | Koch | Jo  |
|----------|----------|------|-----|
| TP       | 301      | -    | 42  | 22 |
| FN       | 7        | -    | 8   | 9  |
| Accuracy | 99.1%    | -    | 99.8%| -  |
| Precision| 92.9%    | 90%  | 75% | 88%|
| Recall   | 97.7%    | -    | 84% | 71%|

5. Conclusion

A technique of automatic 3D road defects detection using both 2D and 3D information is presented in this paper. Images with overlapping regions and no motion blur are captured by a system at the high speed. Structure from motion algorithm reconstructs the road surface into 3D scaled points by using the overlapping regions in 2D images. Camera’s motion model is applied to help removing the mismatched 3D reconstruction points. Then the 3D information classifies the points into defect and non-defect points which are used to find the defect region on the road surface. Finally the defect region detection are based on both the characteristics of 2D images and the information of 3D points.

Three experiments were presented to evaluate the efficacy of the proposed automatic 3D pavement defects detection technique. The first experiment examined three surface fitting models and the quadratic surface fitting model was chosen for the proposed technique. Next, the experiment on the mismatched points rejection procedure illustrated that a value of m between 1.5 to 2.5 was efficient to remove the mismatched points and keep as many matched points as possible. The proposed technique was first applied to five typical road surface images to show the qualitative results of using both 2D images and 3D points information in defects detection. Then the proposed technique was applied to 3500 road surface images which contained 308 images with defects and 3192 images without defect regions. The result shows that the accuracy, precision, and recall are all above 90%. The results have demonstrated the efficacy and effectiveness of the proposed technique in automatic 3D pavement defects detection.
Fig. 6. (a) Original images of the road surface. (b) Reconstruction of 3D points from consecutive images with overlapping region. Red and blue boxes represent two camera positions. (c) Detected defects (red circles) $D^k_r$ based only on 3D information. (d) The binary 2D image marked with detected defects. (e) Refined defects $(\hat{D}^k_r)$ detection after combining 2D and 3D information together.

Acknowledgement

The authors would like to thank Murata Manufacturing Co., Ltd. for their support of this work. This paper is written based on a proceeding presented at JSAE FAST-zero'17 Meeting.

References

(1) Steer Davies Gleave - Roberta Frisoni, Francesco Dionori, Lorenzo Casullo, Christoph Vollath, Louis Devenish, Federico Spano, Tomasz Sawicki, Soutra Carl, Rooney Lidia, João Neri, Radu Silaghi, Andrea Stanghellini. “EU road surfaces: economic and safety impact of the lack of regular road maintenance.” European Union, 2014.

(2) Slone, Sean, and August Monday. Condition of US Roads & Bridges. Council of State Governments, 2010.

(3) Eriksson, Jakob, et al. "The pothole patrol: using a mobile sensor network for road surface monitoring." Proceedings of the 6th international conference on Mobile systems, applications, and services. ACM, 2008. pp. 29-39 (2008).

(4) Tai, Yu-chin, Cheng-wei Chan, and Jane Yung-jen Hsu. "Automatic road anomaly detection using smart mobile device." conference on technologies and applications of artificial intelligence, Hsinchu, Taiwan. (2010).

(5) Mednis, Artis, et al. "Real time pothole detection using android smartphones with accelerometers." Distributed Computing in Sensor Systems and Workshops (DCOSS), Proceedings of the 2011 International Conference on IEEE, pp. 1-6 (2011)

(6) Xue, Guangtao, et al. "Pothole in the Dark: Perceiving Pothole Profiles with Participatory Urban Vehicles." IEEE Transactions on Mobile Computing, Vol. 18, No. 5, pp. 1408-1419 (2017).

(7) Koch, Christian, and Ioannis Brilakis. "Pothole detection in asphalt pavement images." Advanced Engineering Informatics, Vol. 25, No. 3, pp. 507-515 (2011).

(8) Koch, Christian, Gauri M. Jog, and Ioannis Brilakis. "Automated pothole distress assessment using asphalt pavement video data." Journal of Computing in Civil Engineering, Vol. 27, No. 4, pp. 370-378 (2012).

(9) Jo, Youngtae, and Seungho Ryu. "Pothole detection system using a black-box camera." Sensors, Vol. 15, No. 11, pp. 29316-29331 (2015).

(10) Doycheva, Kristina, Christian Koch, and Markus König. "GPU-enabled pavement distress image classification in real time." Journal of Computing in Civil Engineering, Vol. 31, No. 3, pp. 04016061 (2016).
Azhar, Kanza, et al. "Computer vision based detection and localization of potholes in asphalt pavement images." Electrical and Computer Engineering (CCECE), 2016 IEEE Canadian Conference on, pp. 1-5 (2016).

Chang, K. T., J. R. Chang, and J. K. Liu. "Detection of pavement distresses using 3D laser scanning technology." Computing in Civil Engineering, pp. 1-11 (2005).

Hou, Zhiqiong, Kelvin CP Wang, and Weiguo Gong. "Experimentation of 3D pavement imaging through stereovision." International Conference on Transportation Engineering, pp. 376-381 (2007).

El Gendy, Amin, et al. "Stereo-vision applications to reconstruct the 3D texture of pavement surface." International Journal of Pavement Engineering, Vol. 12, No. 03, pp. 263-273 (2011).

Ahmed, Mahmoud, C. T. Haas, and Ralph 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).

Antol, Stanislaw, Kunjin Ryu, and Tomonari Furukawa. "A New Approach for Measuring Terrain Profiles." Proceedings of the ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. V02AT02A026 - V02AT02A026 (2013).

Moazzam, I., et al. "Metrology and visualization of potholes using the microsoft kinect sensor." Intelligent Transportation Systems-(ITSC), Proceedings of the 2013 16th International Conference on IEEE, pp. 1284-1291 (2013).

Hartley, Richard, and Andrew Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003.

Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision, Vol. 60, No. 2, pp. 91-110 (2004).

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, SAE Paper 2017-01-0082 (2017).