

# Pothole 3D Reconstruction With a Novel Imaging System and Structure From Motion Techniques

Adeel Ahmed, Moez Ashfaque, Muhammad Uzair Ulhaq<sup>id</sup>, Senthan Mathavan<sup>id</sup>, *Member, IEEE*,  
Khurram Kamal, and Mujib Rahman

**Abstract**—Machine vision based evaluation systems are receiving increased attention, day by day, for automated quality inspection of roads. Industrial pavement scanners consist of laser scanners and are very expensive, hence inaccessible for everyone. The proposed work presents a simple and novel approach for 3D reconstruction of potholes for an automated inspection and road surface evaluation. The technique utilizes a Structure from Motion based 3D reconstruction algorithm, along with laser triangulation, to generate 3D point clouds of potholes. Alongside, a novel low-cost system, consisting of a single camera and a laser pointer, is also proposed. Keypoint matching techniques are employed, with the 5-point algorithm, on successive image frames to generate a point cloud. However, this point cloud is not metric yet, without scale information. The scale ambiguity is solved by making use of the laser pointer, and using the principle of triangulation. The laser spot is also detected in the same image sequence that is used for point-cloud building, cutting down the image capturing and processing overhead. The system has been benchmarked on artificial indentations with known dimensions, proving the robustness of the measurement scheme and hardware. Static and dynamic tests have been performed. The mean depth errors for measurement made by the imager statically and at dynamic speeds of 10 km/hr, 15 km/hr, and 20 km/hr are 5.3%, 7.9%, 14.4%, and 26.6%, whereas for perimeter the errors are 5.2%, 6.83 %, 11.8%, and 27.8%. The proposed, low-cost technique shows promising results in generating 3D point clouds for potholes.

**Index Terms**—Pavement, imaging, potholes, structure from motion, keypoint, metrology.

## I. INTRODUCTION

**R**OADS and highways play a vital role in the economic development of a country, providing access to health, education, farming, food, and markets. An efficient road infrastructure provides a competitive edge to a country. Road surfaces are continuously subjected to distresses like potholes, cracks, and rutting due to severe weather conditions,

varying loads and aging. According to a survey, repair cost of distressed pavement can increase by seven times over a period of five years [1]. Therefore, constant monitoring and maintenance of road pavement surface is crucial to maintain good service and provide safe travelling conditions in an economical way. This can be achieved only through a system of timely inspection and action. Such systems recognize surface condition assessment of pavement as a factor that relies upon reliable, good-quality measurements for different types of distresses. These defects include potholes, cracks, rutting, and raveling, among others. Potholes are mostly bowl-shaped depressions, formed on the road surface, and can be up to 200 mm in depth. Potholes are the single most reason for causing damage to vehicle suspension systems, they are also the biggest contributor to bad ride quality and accidents [2].

Manual pavement inspection and distress evaluation is laborious and time consuming, leading to the conclusion that an automated distress evaluation system is most desirable. With varying degrees of accuracy and acceptance, a number of techniques for the assessment of road surface have been proposed over the years that include vibration based evaluation, image-based recognition, and 3D reconstruction. Vibration based methods use accelerometer data and GPS installed on vehicles to record impact of road distresses on vehicle dynamics. Although having low-cost and very fast processing capabilities as its advantages, vibration based methods are susceptible to noise. In addition, they can only assess the areas of wheel tracks of a road, which are right next to lane boundaries. Image recognition systems, which essentially gather 2D information, provide measurements of spatial properties and quantity of distresses for severity level assessment, but lack 3D information about distresses. To improve the quality of distress evaluation, depth of distresses is retrieved in 3D reconstruction methods. Laser based methods also produce 3D profiles of road surface using scanners. Whilst producing highly accurate results, laser equipment has the disadvantage of very high cost.

With the arrival of inexpensive, high-resolution cameras, a number of methods, using a single camera for 3D reconstruction have become realistic. To reconstruct potholes in 3D using a single camera, a new Structure from Motion (SfM) based vision technique is proposed in this article. Scale of reconstruction is determined using laser triangulation, by employing a novel hardware setup. A laser distance meter and LIDAR [3] have been used earlier for physical distance measurement and 3D surface scanning to aid reconstruction process in their

Manuscript received September 10, 2019; revised May 9, 2020, September 22, 2020, and November 29, 2020; accepted January 15, 2021. The Associate Editor for this article was K. Wang. (*Corresponding author: Muhammad Uzair Ulhaq.*)

Adeel Ahmed, Moez Ashfaque, Muhammad Uzair Ulhaq, and Khurram Kamal are with the Department of Mechatronics Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan (e-mail: adeel.ahmed86@mts.ceme.edu.pk; moizashfaque@yahoo.com; ulhaquzair@yahoo.com; khurram\_kamal@hotmail.com).

Senthan Mathavan is with Nobleo Technology, 5611 EB Eindhoven, The Netherlands (e-mail: s.mathavan@ieee.org).

Mujib Rahman is with Department of Civil Engineering, Brunel University London, Uxbridge UB8 3PH, U.K. (e-mail: mujib.rahman@brunel.ac.uk).

Digital Object Identifier 10.1109/TITS.2021.3054026

techniques. This research, for the first time, proposes the use of laser pointer to estimate scale of reconstruction through laser triangulation, hence providing a simple and low cost solution to address inherent scale ambiguity of single camera reconstruction.

## II. LITERATURE REVIEW

The status quo on pavement distress evaluation can be categorized into vibration-based sensing, 2D image-based distress detection and 3D pavement surface reconstruction [4]. Yu and Yu [5] proposed the vibration based sensing system which provides an evaluation of pavement distress by measuring the impact of road defects on the reactive dynamics of a moving vehicle [4]–[8]. Although they are low cost and can attain real time performance, vibration-based methods provide qualitative results only. The accuracy is hindered by the fact that joints and bumps of roads can be detected as potholes.

Image (i.e. 2D) vision-based methods employ image processing and machine learning algorithms on pavement images and video frames to automatically calculate 2D characteristics of potholes and other road defects. In earliest approach, Koch and Brilakis [9] proposed automatic detection of potholes from pavement images involving segmentation of an image into defect and non-defect regions. Comparison of defect regions with a predetermined model of pothole identifies potential potholes in image frames. In further improvement of their work [10], an enhanced pothole-recognition method, capable of video based pavement assessment, is presented. Similar video based approaches [10]–[12] target continuous tracking of potholes in subsequent frames to estimate total number of potholes. However, 2D vision-based methods do not provide depth information for severity assessment.

To include depth information in distress severity measurement, researchers proposed 3D pavement imaging based on Kinect sensor [13], [14], laser scanning [3], [20] and stereovision [22]–[24]. Hou *et al.* [15] explained the development of Digital Highway Data Vehicle (DHDV) at University of Arkansas. Equipped with several line scan and area-scan cameras installed at a fixed distance, DHDV captures images of pavement and its software reconstructs the pavement surface using the stereovision principle. Salari *et al.* [16] and Jog *et al.* [11] performed the integration of information from 2D images and 3D reconstruction to further improve the accuracy of pavement surface analysis. By taking an unconventional approach, Zhang and Elaksher [17] proposed a UAV-based imaging system for pavement condition monitoring which uses SfM based 3D reconstruction algorithm to determine image orientation parameters and 3D point cloud. The authors reported an accuracy of 0.5 cm for on surface and depth measurements of road distresses.

Laser scanners are also popular among researchers for 3D pavement imaging for distress measurement. Chang *et al.* [18] developed a 3D laser scanning system for pavement surface and extraction of an elevation model by means of a grid-based processing approach. Similar approaches [19], [20] perform the conversion of laser stripe deformation due to road distresses into 3D depth profiles by signal processing. Results achieved through laser scanning systems suggest that laser

generated 3D models can provide highly accurate measurements of distress features as well as estimation of area and volume of road distresses.

### A. Multi View Stereo

Vision based techniques that use two or more images and use point correspondences as their main cue are termed as Multi-view stereo (MVS) techniques. Stereovision, Structure-from-Motion (SfM), and Simultaneous Localization and Mapping (SLAM) are examples of MVS algorithms that share a basic processing pipeline, which is explained below.

1) *Stereovision*: In 3D vision-based approaches, stereovision is a most widely used technique. Basic operation of stereovision includes the computation of the 3D structure of the scene using images taken from at least two different viewpoints. One of the challenging research problems in stereo vision is detecting corresponding pairs of key points in stereo images [21]. Candidate points for matching must be different enough from its neighbors to rule out any ambiguity while matching. Researchers have proposed several algorithms and techniques to address this research problem [22]–[24]. Every algorithm makes use of a matching cost function or energy function to develop correspondence between two pixels.

2) *Visual Simultaneous Localization and Mapping (VSLAM)*: VSLAM is another widely used computer vision technique. Earlier work on SLAM was predominantly based on single Extended Kalman Filter (EKF) approach [25] for reconstruction of medium sized maps in robotic environment and demonstrated the necessity of maintaining estimate correlations. Owing to increased image resolution and image quality of digital cameras and increased computing power, Davison [26] introduced the first real-time monocular SLAM (MonoSLAM) algorithm.

3) *Structure From Motion (SfM)*: SfM is another vision-based technique that employs stereo correspondences to compute 3D geometry of the scene using more than two images. While taking a set of images and intrinsic camera parameters as inputs, SfM algorithms output two important things: (1) camera matrix for each image view (2) 3D points representing objects visible in the images. Apart from some additional steps, a number of state-of-the-art SfM algorithms have a set of common processing steps. Initial research on SfM was focused on the geometry of multiple views while assuming a rigid scene [21]. A technical overview of early work on visual reconstruction algorithms is presented by Tomasi [27]. Later on, the development and use of RANSAC [28] proved to be a key step forward, which enabled researchers to robustly estimate the epipolar geometry between two views.

4) *Parallel Tracking and Mapping (PTAM)*: Klein and Murray [29] proposed the PTAM algorithm to separate the two basic processes of reconstruction. It was proposed that camera pose tracking and environment map building be performed in two simultaneous pipelines. A separate mapping process enables the system to extract and match thousands of features in real time. The PTAM algorithm utilizes measurements of hundreds of features per frame to perform pose estimation. The methodology of PTAM can be summarized as follows:

(1) tracking and mapping are performed separately in parallel threads, (2) key-frame based mapping is performed using bundle adjustment, (3) map is initialized using the 5-point algorithm on a stereo pair, (4) new points are initialized using epipolar geometry, and (5) incremental map building to produce a dense point cloud. PTAM has been adopted by many researchers for highly robust tracking to provide sparsely mapped 3D points in real time [30], [31]. Dense 3D mapping requires intensive computations and therefore higher processing power. State-of-the-art SfM algorithms are currently achieving high quality dense stereo reconstruction in real time while making use of GPU implementation. In this regard, Newcombe *et al.* [32] presented DTAM (Dense Tracking and Mapping), which overcame the limitations in previous real-time monocular SLAM systems. DTAM generates a dense 3D surface model and immediately uses it for dense camera tracking via whole image registration. Similar approaches achieved dense mapping using GPU-accelerated techniques [31], [32].

5) *Scale Ambiguity*: Structure from Motion using a single camera tends to simultaneously obtain camera motion and 3D representation of the scene. Without any external knowledge about the environment, the resulting reconstruction can only be defined for a scale [32]. Scale information is essential to determine the real-world dimensions of reconstructed structure. Existing reconstruction algorithms employ inertial sensors and laser distance meters for scale estimation. In inertial measurement unit (IMU) supported reconstruction methods [33]–[37], feature based tracking and 3D map building is assisted by inertial sensing in relative position and orientation estimation. Consequently, the 3D model obtained after reconstruction represents the real-world dimensions of the objects being reconstructed. The algorithms are optimized to deal with noisy measurements, e.g. due to motion blur, rolling shutter artifacts, or low-quality IMU.

Wu *et al.* [38] proposed integration of a laser distance meter (LDM) with a camera for enhanced monocular visual odometry to be used for astronaut navigation. Using direct distance measurements from LDM, system recovers a global trajectory from monocular images. In another example, Fanfani and Colombo [3] evaluated the performance of a hybrid camera-LiDAR framework to estimate absolute camera displacements using a SLAM based method. LDM and LiDAR are expensive hardware, and their use pushes up the overall cost of the system. Moreover, bringing together different technologies into a single sensor package has also its drawbacks.

In this research, laser triangulation is achieved by using a low-cost laser pointer to recover scale information, thereby, metric 3D models. In this regard, it is proposed that the image from the camera is used for SfM based reconstruction as well as to recover the scale of the scene, by robustly detecting the laser spot in the scene. Even though stereo reconstruction is not possible in the road where the laser spot falls, the spot is significantly small to have any real effect on the 3D reconstruction. There can be instances where the laser falls into deep recesses, like thin cracks, thereby not being able to calculate the scale of the current image frame. To eliminate this absence of scale, the proposed algorithm makes use of the

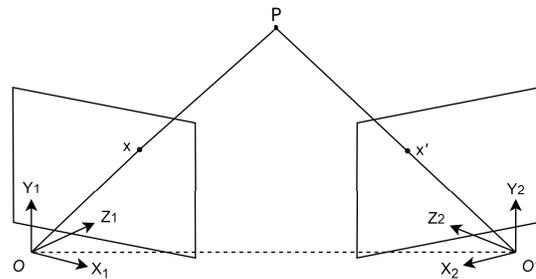


Fig. 1. Two view geometry.

scales of the scene measured from the previous and successive frames to reconstruct the current scene without issues; the retrospective correction from future frames is not an issue as the road profile need not be constructed in real-time. As the thin cracks, and similar features, are also not densely packed even in worst road surfaces, the proposed method works. The idea is novel and never applied before for SfM based monocular reconstruction, let alone its application to pavement imaging.

### III. THEORY

#### A. Stereo Reconstruction

A two-camera stereo geometry is shown in Figure 1. The real-world coordinates of a point that is seen by both the cameras is given by  $P$ . Let us assume that the image coordinates of point  $P$  in the first camera is given by  $x$ , in homogeneous coordinates. The same point is imaged as  $x'$  in the second camera. In stereo theory, the two points are related by the so-called fundamental matrix.

When the cameras are calibrated, the relationship between the two image coordinates is defined by the essential matrix. For a detailed treatment of the concepts and equations, the reader is directed to any standard textbook on computer vision or 3D imaging [39].

The five point algorithm is proposed by Nister [40] to compute the essential matrix from any 5 distinct point correspondences between the cameras and to solve the relative pose problem. The 5-point algorithm allows perfect stereo reconstruction, but does not estimate the scale of the 3D reconstruction. This article proposes a solution to the problem of scale estimation, by using laser triangulation, as explained below. Using a similar approach to un-calibrated 7 and 8-point methods, the 5-point algorithm can be modified for more than 5-points [40].

#### B. Laser Triangulation

Laser triangulation is a well-known technique used for 3-dimensional measurements in machine vision [39]. It is a simple and robust technique, where a laser beam is projected on object surface. Laser spot on the object surface is imaged by a lens with a numerical aperture. The lens focuses the image of laser spot on a CCD array sensor. The angle of triangulation is formed by the axes of illumination and observation. Planar geometry for the triangulation scheme is shown in Figure 2. The derivation of the depth of the point on the road where the laser points is as follows.

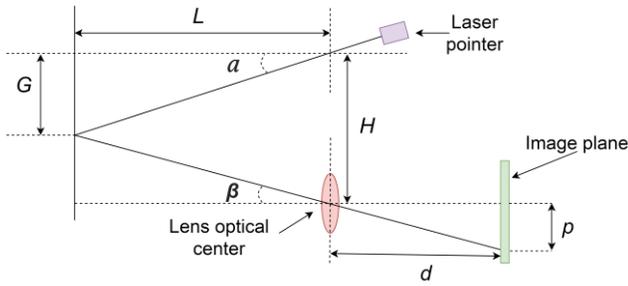


Fig. 2. Laser triangulation schematic.

The collection angle  $\beta$  is computed as:

$$\beta = \tan^{-1} \left( \frac{p}{d} \right)$$

From trigonometry it can be verified that:

$$L = \frac{H}{\tan \alpha + \tan \beta}$$

and,

$$G = L \tan \alpha \quad (3)$$

From Equations (1) and (3), it can be shown that

$$L = \frac{Hd}{p + d \tan \alpha} \quad (4)$$

where, also referring to Figure 2,  $L$  is the shortest distance of laser point from imaging plane,  $H$  is the baseline distance between camera lens and laser pointer,  $G$  is the shortest distance from the laser pointer location to the normal of the laser pointer plane,  $d$  is distance between lens and CCD array sensor, and  $p$  is the pixel distance of the located point from the image center.  $\alpha$  and  $\beta$  are projection and observation angles, respectively.

The position of the laser dot varies on CMOS array as distance between ground and setup increases or decreases. The image of laser dot is captured at different known distances from ground. The camera and the laser source are fixed rigidly with respect to each other at the fixed distance between them. Corresponding pixel position of laser dot is plotted against known distances from ground to obtain a relation between pixel position and ground distance.

### C. Imaging From a Moving Vehicle

Images are also obtained by mounting the imaging system on a vehicle. Pothole images acquired from video frames contain motion blur which reduces the quality of image, hence degrading keypoint matching. A low number of keypoint matches does not produce a dense 3D point-cloud upon stereo reconstruction, which in turn reduces the accuracy of system. However, to solve the blur issue in dynamic images, a technique devised by [22] is used in this work, where the dynamic images are passed through a series of operations involving high-pass filtering, histogram equalization, and Gaussian filtering respectively.

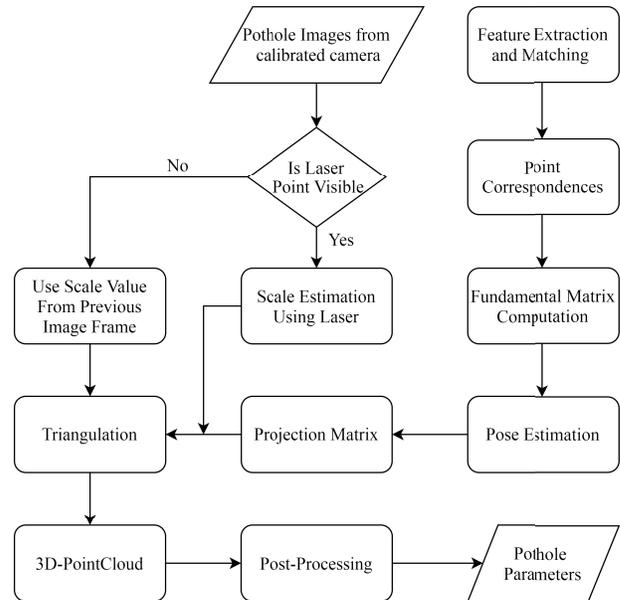


Fig. 3. Flow chart of proposed technique.

## IV. METHODOLOGY

To obtain the metric 3D models of potholes, the proposed system uses a SfM based reconstruction algorithm integrated with laser triangulation. The flow chart of proposed method is shown in Figure 3.

### A. SfM Based Reconstruction

The algorithm takes as input a set of pothole images and outputs the translation and rotation matrices between the cameras for each image, and 3D points visible in each image. The algorithm relies on the SIFT feature transform for 2D feature extraction and matching, which is performed using VL\_Feat library [41] in MATLAB. SIFT provides dimensional feature descriptors. Values of different parameters in SIFT algorithm are as stated in Lowe [42].

Matched 2D image points  $(x, x')$  are provided as input to the 5-point algorithm. As the intrinsic camera matrix is known the essential matrix can be calculated. Now, the projection matrices for both cameras can be recovered from the essential matrix. Projection matrices associated with each image frame pair map matched 2D points  $(x, x')$  to get 3D point cloud of scene using triangulation.

In order to determine the absolute scale of reconstruction,  $S$ , the laser triangulation provides distance  $L$  between camera center and ground. Pixel location of laser point in image frame is automatically detected in the following manner. A small window of size relative to the size of the laser point was passed over the image to detect the red point based on the image intensity values. The window in the image with highest red intensity was regarded as the pointer location. This location was then fed into the scale estimation process in (4). A square region  $40 \times 40$  around this pixel is searched for 2D SIFT feature point that has minimum Euclidean distance from the pixel location of laser point. This feature point serves as an estimate for the laser point and provides a measure of distance

$L'$  defined up to a scale between camera center and laser point projection on ground. The distance  $L'$  is then divided by absolute distance  $L$  provided by laser triangulation to get scale factor  $S$  in (5).

$$s = \frac{L'}{L} \quad (5)$$

The initial reconstruction is then multiplied by the scale factor to obtain the real-world reconstruction.

During the reconstruction process, laser pointer may not be visible in some image frames. The algorithm successfully tackles the invisibility of laser point in one image frame by assuming the scale factor calculated in the previous image frame.

### B. Post Processing of Point Cloud

The processes listed below are applied to the point cloud obtained for better visualization and to obtain metrological parameters of pothole.

1) *Tilt Removal*: Due to varying camera angle and distance, the point cloud obtained after the 3D reconstruction contains a tilt, with respect to the global coordinate system as shown in Figure 7. This tilt needs to be removed, especially if the point cloud is to be processed further for obtaining the metrological features of the pothole. The tilt is removed in two steps. First, the surface equation of the road is calculated by using a RANSAC algorithm which uses the following plane fitting equation.

$$b_1x + b_2y + b_3z + b_4 = 0 \quad (6)$$

where  $x$ ,  $y$  and  $z$  are the coordinates of the point cloud of the road surface,  $b_1$ ,  $b_2$ ,  $b_3$  and  $b_4$  are the coefficients the plane equation.

Second, with reference to [43] and without any loss of generality, the global coordinate system is chosen, with respect to the point-cloud coordinate system, such that the translation vectors along  $x$ - and  $y$ -axis are zero. It is assumed that  $u = (u_x, u_y, u_z)$  is the unit vector around which the rotation of point-cloud takes place. Now, the transformation matrix  $T$  [43] between the global coordinate system and the point-cloud becomes as (7), shown at the bottom of the page, where,

$$\begin{aligned} u_x &= b_2 / \sqrt{b_1^2 + b_2^2 + b_3^2} \\ u_y &= -b_1 / \sqrt{b_1^2 + b_2^2 + b_3^2} \\ \cos\theta &= b_3 / \sqrt{b_1^2 + b_2^2 + b_3^2} \\ \sin\theta &= \sqrt{b_1^2 + b_2^2} / \sqrt{b_1^2 + b_2^2 + b_3^2} \end{aligned}$$

This transformation matrix is used to orient the point cloud into a tilt-less orientation using the following transformation equation:

$$P_0 = P_1 \times T^{-1} \quad (8)$$

where  $P_1$  represent the original orientation of the point cloud, and  $P_0$  represent the transformed, tilt-less orientation of the point cloud.  $P_0$  and  $P_1$  are both homogenous coordinates.

2) *Differentiation of Pothole From Road Surface*: Aligned 3D-pointcloud obtained after tilt removal process is used to separate the pothole points from the road surface. The 3D points which satisfy the plane equation described in the above section are regarded as road surface whereas all the other 3D points are taken as part of a pothole.

3) *Perimeter of Pothole*: To obtain the perimeter of pothole, the boundary function provided in MATLAB is used. The boundary function is based on the alpha-shape algorithm [44]. 3D points representing pothole points are projected on plane that represents the road surface. These 2D projections provide view of surface area covered by pothole. Joining the outermost points, boundary function draws a boundary of the pothole and the total length of the boundary is considered as the perimeter of pothole.

4) *Maximum and Average Depth of a Pothole*: Distances of each pothole inlier point along  $z$ -axis represent the depth of pothole at that location. The maximum  $z$ -value corresponds to the maximum depth of pothole whereas the mean value of these distances gives an estimate for the average depth of the pothole.

## V. EXPERIMENTATION

Images of potholes on different roads in Islamabad, Pakistan are captured using proposed imaging system shown in Figure 4. The proposed imaging system has the field of view that spans about 1 m, whereas lane width of a typical road is  $\sim 4$ m. Therefore, four such cameras, fitted in the lateral direction of the road, along with laser pointers, are required to cover a single lane. The calibrated camera has a resolution of  $1280 \times 720$  pixels. For algorithm development and processing, MATLAB is used. The computations are performed on Intel Ci5 6<sup>th</sup> Gen Processor with 8 GB RAM. The computation time of algorithm is found by iterating the algorithm on the test set of 30 potholes, and the average computation time is found to be 5.2 seconds. An alternate equivalent approach for 3D-reconstruction such as semi-global matching is used by [45]. They have reported the computation time of 19 ms where computations are performed on GTX 280 GPU. The computation time for proposed algorithm can be reduced by parallelizing the computations on GPU.

For an absolute scale of reconstruction, a novel hardware setup comprising a webcam and a laser pointer rigidly attached

$$T = \begin{bmatrix} \cos\theta + u_x^2(1 - \cos\theta) & u_x u_y(1 - \cos\theta) & u_y \sin\theta & 0 \\ u_x u_y(1 - \cos\theta) & \cos\theta + u_y^2(1 - \cos\theta) & -u_x \sin\theta & 0 \\ -u_y \sin\theta & u_x \sin\theta & \cos\theta & -b_4/b_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7)$$

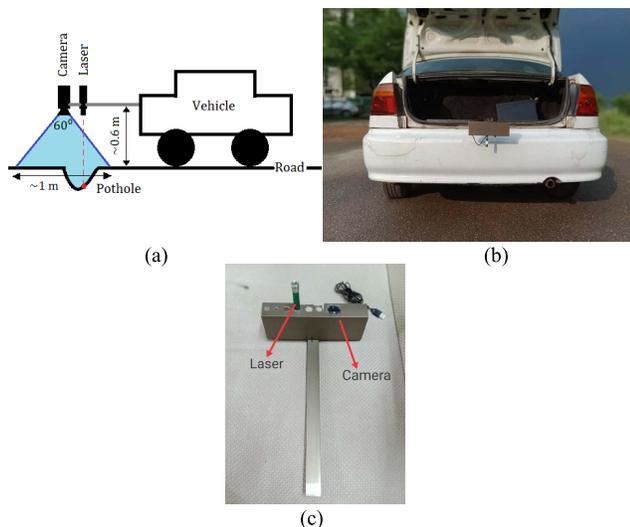


Fig. 4. (a) Dynamic imaging schematic, (b) Imaging system mounted to the rear of a car, (c) Proposed imaging system.

on a solid bar is used as shown in Figure 4 (c). Such arrangement allows camera to capture laser dot on pavement surface and determine distance of camera setup from ground using laser triangulation principle. Images of real potholes are captured using the proposed camera setup at a standoff distance of 40-60 cm above the road surface from different viewpoints.

For each natural pothole, average depth, and perimeter are manually measured; for these potholes, the ground truth of their metrology is established based on the procedure of Rajab *et al.* [46]. Depth values at different locations in a pothole are recorded and used to find the average depth of a pothole. To measure the perimeter of a pothole, a chord of wire is placed alongside the edges of the pothole whose length is then measured to get pothole perimeter.

## VI. RESULTS AND DISCUSSION

### A. Calibrations

1) *Camera Calibration*: The camera calibration is performed within MATLAB. The camera matrix determined in camera calibration [47] is given as,

$$K = \begin{bmatrix} 1414.33 & 0 & 630.89 \\ 0 & 1414.70 & 343.03 \\ 0 & 0 & 1 \end{bmatrix}$$

2) *Laser Calibration*: To calibrate the laser pointer, images of a straight plane were captured from the setup at known heights from the flat surface. Heights of 10 to 110 cm were marked with successive intervals of 10 cm on a ruler. At each mark the ground was imaged with camera axis of the acquisition setup maintained perpendicular to the ground. The pixel location of the laser points in each image was located. A relation between laser pointer location in the image and the corresponding height was obtained. Figure 5 shows the relation between the height of experimental setup and the pixel locations of the laser pointer. Using the pixel location of the laser dot, the depth of the point can hence be estimated through a look-up table.

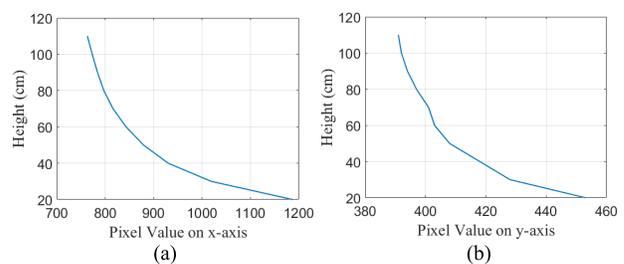


Fig. 5. (a) Position of laser pointer on x-axis of image plane, (b) Position of laser pointer on y-axis of image plane.

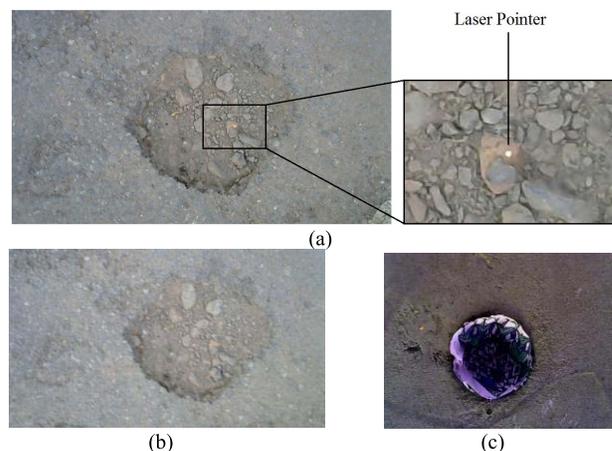


Fig. 6. (a) Real pothole N1 (static image) with detected laser pointer, (b) Real pothole N1 (dynamic image), (c) Artificial pothole-like indentations A1.

### B. Pothole Reconstruction

Figure 6 shows the results of raw pothole images as obtained from image sensors. For reconstruction, one image pair is found enough to capture the horizontal extents of all potholes. However, for very large potholes, the algorithm can also be implemented with multiple image frames as well. For a given input image, a SIFT based keypoint extraction is performed. In the subsequent image, the same procedure is carried out. Then the keypoints from both the images are matched using SIFT features. In this regard, a window size of  $40 \times 40 \times 40$  pixels is used to search for the best matching candidates from the second image. The 40 pixel window side itself has been chosen based on the typical depth range of pothole depth and the standoff between the imaging setup and the road. In a parallel manner, to the SIFT keypoint detection, the laser point is also detected in the images in order to do triangulation. These two processes start the chain of processing activities given in Figure 3, resulting in an intermediate point cloud as shown in Figure 7.

### C. Tilt Removal

A subsequent part of the algorithm removes the alignment issues between the imaging system and the road, called tilt removal in this article (section IV-B-1), and the horizontally aligned point cloud of the pothole is shown in Figure 8. For fitting a plane through the surface points, a distance threshold is used to remove any outlier data point from the plane. Any point lying at a distance beyond this threshold value is regarded as an outlier. As this threshold value is manually

TABLE I  
MEASURED PARAMETERS FOR ARTIFICIAL POTHOLE-LIKE INDENTATIONS

	Perimeter (cm)		Percent Error (%)	Depth (cm)				Percent Error (%)
	Measured	Ground Truth		Max. Depth (cm)		Avg. Depth (cm)		
				Measured	Ground Truth	Measured	Ground Truth	
Hemispherical Indentation A1	28.7	31	-7.3	5.43	5.00	3.60	3.56	2.8
Conical Indentation	49.6	54	-8.1	7.53	8.16	5.00	5.17	3.2

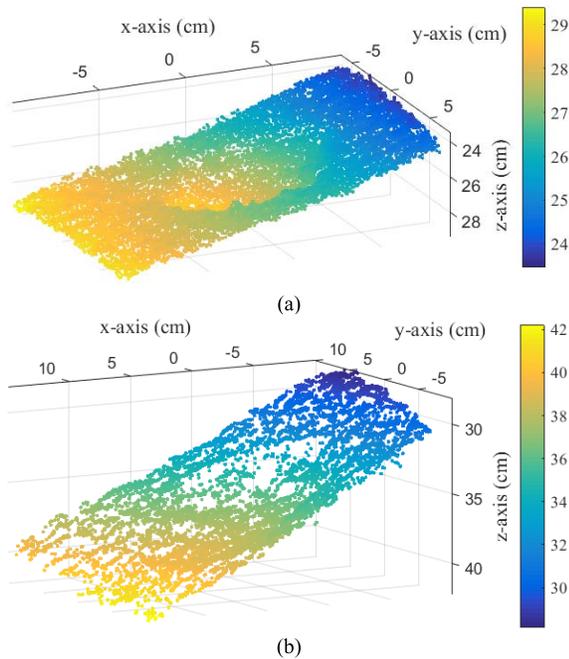


Fig. 7. 3D-reconstruction of Pothole N1, (a) Static image, (b) Dynamic image.

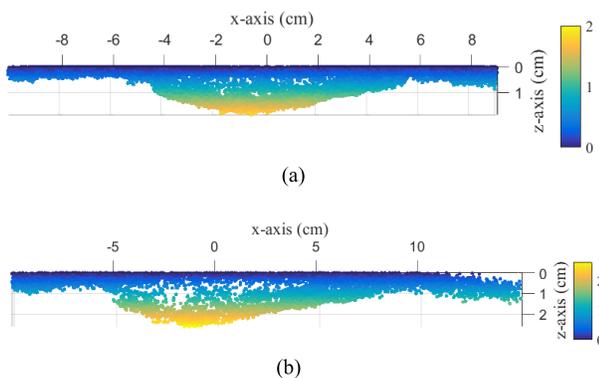


Fig. 8. Aligned Pothole N1 (a) Static, (b) Dynamic.

chosen, its sensitivity regarding the final results obtained by the algorithm is studied in the following manner.

For varying values of the threshold, the error of the systems performance is calculated by plotting the error values obtained for average and maximum depths and, perimeter by using the corresponding ground truth values available. These error plots can be seen in Figure 9. Based on these plots, the value of the threshold corresponding to the minimum errors is chosen

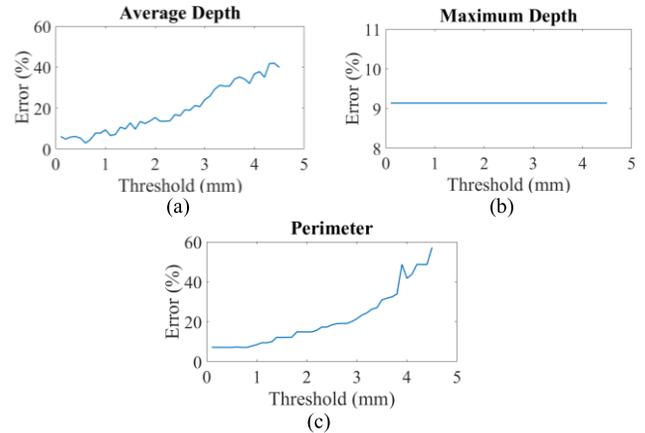


Fig. 9. Error plots for different threshold values used in RANSAC for (a) average depth, (b) maximum depth, (c) perimeter.

as the best value of the threshold. This best value is found to be 0.7 mm.

#### D. Metrological Features/Evaluation of Potholes

The last portion of the algorithm proposed in this article deals with extracting the metrological properties of potholes, namely the maximum depth, average depth and the perimeter.

The maximum depth is measured straight from the point cloud data in Figure 8. The results obtained on multiple potholes found in the roads of Islamabad. The measured values against the ground truth values are shown in Table II and Table III.

#### E. Benchmarking With Artificial Indentations in Concrete

To benchmark the algorithm, it is first used to recover 3D models of artificial potholes with known dimensions. For benchmarking purposes, artificial potholes are covered with a textured paper therefore enabling the algorithm to detect maximum number of distinct feature points for better matching and accurate reconstruction. Figure 12 shows a tilt removed point-cloud of the hemispherical artificial indentation (numerical results are in Table I). Comparison between measurements obtained for artificial and real potholes, and their respective ground truths, show that reconstruction of natural pothole is reasonably accurate. The mean percent error for the average depth of artificial potholes dataset is found to be 3.0%, whereas the mean percentage error for the perimeter of is found to be 7.7%. The proposed hardware and algorithm can become

TABLE II  
MEASURED AVERAGE DEPTH FOR REAL POTHOLES

	Ground Truth (cm)	Static Imaging		Dynamic Imaging						Mean Error (%)
				10 km/hr		15 km/hr		20 km/hr		
		Calculated (cm)	Error (%)							
Pothole N1	1.95	1.81	7.17	2.13	-9.23	2.28	-16.9	2.31	-23.5	16.5
Pothole N2	1.12	1.06	5.35	1.03	8.03	1.34	-19.6	1.41	-25.8	16.1
Pothole N3	3.13	3.32	-6.07	3.34	-6.70	3.53	-12.7	3.88	-27.6	15.7
Pothole N4	2.64	2.57	-5.60	2.68	-7.20	2.79	-11.6	3.37	-20.8	13.2
Pothole N5	2.87	2.80	2.43	3.12	-8.71	2.55	11.1	2.34	18.4	12.7

TABLE III  
MEASURED PERIMETER FOR REAL POTHOLES

	Ground Truth (cm)	Static Imaging		Dynamic Imaging						Mean Error (%)
				10 km/hr		15 km/hr		20 km/hr		
		Calculated (cm)	Error (%)							
Pothole N1	31.5	29.5	6.24	28.8	8.46	26.3	16.4	22.5	28.3	17.7
Pothole N2	42.3	40.2	5.00	39.2	7.38	35.2	14.4	28.4	32.9	19.0
Pothole N3	57.6	60.3	-4.59	54.1	6.03	52.1	9.60	40.8	29.2	14.9
Pothole N4	51.1	49.6	2.89	54.8	-6.04	47.3	8.50	37.8	26.0	13.5
Pothole N5	60.5	64.8	-7.17	56.7	6.24	54.3	10.2	46.6	22.8	13.1

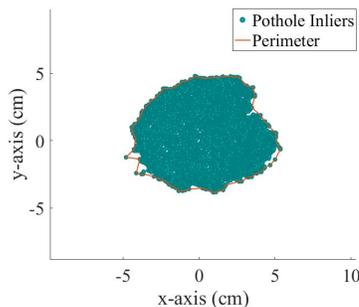


Fig. 10. Perimeter of Pothole N1.

a simple and inexpensive alternative to the industrial laser scanners used exclusively to scan roads presently. A laser scanner system uses laser line to scan a surface. Variations in surface depth cause deflections in the laser line which are captured by the camera and used for surface reconstruction. Therefore, its surface reconstruction is very dense and can be accurate up to sub-millimeter level. However, laser equipment is very expensive, costing in the excess of US\$1 million. Likewise, Kamal *et al.* [14] proposed the Kinect sensor technology for pavement depth evaluation and have reported the error percentage of 5.47% for volume calculation of artificial pothole like indentations. In case of stereovision technology, the accuracy of 3 mm is achieved with the error percentage of 6.72% for pothole evaluation [22]. In the present proposal, the laser pointer and the camera are aligned so that laser point falls

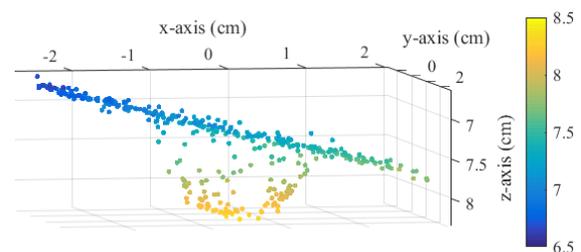


Fig. 11. 3D-reconstruction of Pothole A1.

on middle rows of pixels in camera frame. The algorithm is designed to search for laser point in these specific rows, which is very efficient processing.

#### F. Imaging of Real Potholes

To establish the accuracy of proposed hardware, 30 real-world potholes are imaged. These potholes on roads are imaged both statically and also from a moving vehicle at speeds of 10 km/hr, 15 km/hr, and 20 km/hr. For static imaging the mean percent error for perimeter and average depth of pothole is found to be 5.3% and 5.2%. The mean percent error for average depth of pothole under dynamic conditions is found to be 7.9%, 14.4% and 26.6% for vehicle speeds of 10 km/hr, 15 km/hr, and 20 km/hr respectively. For perimeter measurements of potholes, the mean error percentage is

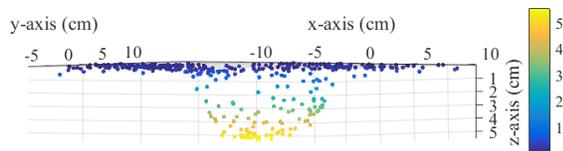


Fig. 12. Aligned Pothole A1.

around 6.83%, 11.8% and 27.8% at the vehicle speeds of 10 km/hr, 15 km/hr, and 20 km/hr respectively.

## VII. CONCLUSION AND FUTURE WORK

The proposed work presents a novel approach for metric reconstruction of potholes using single camera. The proposed method used two main techniques namely Structure from Motion (SfM) and Laser triangulation. The SfM algorithm involves feature extraction, feature matching, pose estimation and triangulation. SIFT feature transform is used for feature extraction and matching while 5-point algorithm is employed in pose estimation to determine fundamental matrix. Laser point triangulation is used to determine unknown scale of monocular reconstruction. Images of natural potholes from different roads in Islamabad are captured using a calibrated camera and laser apparatus. Reconstruction results for 30 natural pothole images are obtained. An artificial pothole rig with two potholes of different size and shapes is prepared to benchmark the algorithm. The proposed technique using laser triangulation shows promising results for metric reconstruction of natural potholes.

## REFERENCES

- [1] (2019). *ALARM Survey*. Accessed: Jun. 2018. [Online]. Available: <https://www.asphaltuk.org/alarm-survey-page/>
- [2] (2017). *KALTIRE Pothole Survey*. Accessed: Jun. 2018. [Online]. Available: <https://info.kaltire.com/news-releases/kal-tire-stores/pothole-survey-highlights-safety-risks-drivers-know/>
- [3] M. Fanfani and C. Colombo, "LaserGun: A tool for hybrid 3D reconstruction," in *Proc. Int. Conf. Vis. Syst.*, Jul. 2013, pp. 274–283.
- [4] T. Kim and S.-K. Ryu, "Review and analysis of pothole detection methods," *J. Emerg. Trends Comput. Inf. Sci.*, vol. 5, no. 8, pp. 603–608, 2014.
- [5] B. X. Yu and X. Yu, "Vibration-based system for pavement condition evaluation," in *Proc. 9th Int. Conf. Appl. Adv. Technol. Transp.* Chicago, IL, USA: ASCE, Aug. 2006, pp. 183–189.
- [6] K. De Zoysa, C. Keppitiyagama, G. P. Seneviratne, and W. W. A. T. Shihan, "A public transport system based sensor network for road surface condition monitoring," in *Proc. Workshop Netw. Syst. Developing Regions (NSDR)*, Aug. 2007, pp. 1–6.
- [7] J. Erikson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, "The pothole patrol: Using a mobile sensor network for road surface monitoring," in *Proc. 6th Int. Conf. Mobile Syst., Appl., Services*, Jun. 2008, pp. 29–39.
- [8] A. Mednis, G. Strazdins, R. Zviedris, G. Kanonirs, and L. Selavo, "Real time pothole detection using Android smartphones with accelerometers," in *Proc. Int. Conf. Distrib. Comput. Sensor Syst. Workshops (DCOSS)*, Jun. 2011, pp. 1–6.
- [9] C. Koch and I. Brilakis, "Pothole detection in asphalt pavement images," *Adv. Eng. Inform.*, vol. 25, no. 3, pp. 507–515, 2011.
- [10] C. Koch, G. M. Jog, and I. Brilakis, "Automated pothole distress assessment using asphalt pavement video data," *J. Comput. Civil Eng.*, vol. 27, no. 4, pp. 370–378, Jul. 2013.
- [11] G. M. Jog, C. Koch, M. Golparvar-Fard, and I. Brilakis, "Pothole properties measurement through visual 2D recognition and 3D reconstruction," in *Proc. Int. Conf. Comput. Civil Eng.*, Dec. 2012, pp. 553–560.
- [12] H. Lokeshwar, L. K. Das, and S. Goel, "Robust method for automated segmentation of frames with/without distress from road surface video clips," *J. Transp. Eng.*, vol. 140, no. 1, pp. 31–41, Jan. 2014.
- [13] Y. Zhang *et al.*, "A Kinect-based approach for 3D pavement surface reconstruction and cracking recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 12, pp. 3935–3946, Dec. 2018.
- [14] K. Kamal *et al.*, "Performance assessment of Kinect as a sensor for pothole imaging and metrology," *Int. J. Pavement Eng.*, vol. 19, no. 7, pp. 565–576, Jul. 2018.
- [15] Z. Hou, K. C. P. Wang, and W. Gong, "Experimentation of 3D pavement imaging through stereovision," in *Proc. Int. Conf. Transp. Eng.*, Jul. 2007, pp. 376–381.
- [16] E. Salari, E. Chou, and J. J. Lynch, "Pavement distress evaluation using 3D depth information from stereo vision," Michigan Ohio Univ., Ann Arbor, MI, USA, Final Rep. MIOH UTC TS43, 2012.
- [17] C. Zhang and A. Elaksher, "An unmanned aerial vehicle-based imaging system for 3D measurement of unpaved road surface distresses," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 27, pp. 118–129, Feb. 2012.
- [18] K. T. Chang, J. R. Chang, and J. K. Liu, "Detection of pavement distresses using 3D laser scanning technology," in *Proc. ASCE Int. Conf. Comput. Civil Eng.*, Jun. 2005, pp. 1–11.
- [19] Q. Li, M. Yao, X. Yao, and B. Xu, "A real-time 3D scanning system for pavement distortion inspection," *Meas. Sci. Technol.*, vol. 21, pp. 15702–15709, Nov. 2009.
- [20] W. Ouyang and B. Xu, "Pavement cracking measurements using 3D laser-scan images," *Meas. Sci. Technol.*, vol. 24, no. 10, Aug. 2013, Art. no. 105204.
- [21] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [22] M. U. U. Haq, M. Ashfaq, S. Mathavan, K. Kamal, and A. Ahmed, "Stereo-based 3D reconstruction of potholes by a hybrid, dense matching scheme," *IEEE Sensors J.*, vol. 19, no. 10, pp. 3807–3817, May 2019.
- [23] H. Hirschmuller, "Stereo vision in structured environments by consistent semi-global matching," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New York, NY, USA, Jun. 2006, pp. 2386–2393.
- [24] L. D. Stefano, M. Marchionni, and S. Mattoccia, "A fast area-based stereo matching algorithm," *Image Vis. Comput.*, vol. 22, no. 12, pp. 983–1005, 2004, doi: [10.1016/j.imavis.2004.03.009](https://doi.org/10.1016/j.imavis.2004.03.009).
- [25] A. J. Davison and N. Kita, "Sequential localisation and map-building for real-time computer vision and robotics," *Robot. Auto. Syst.*, vol. 36, no. 4, pp. 171–183, Sep. 2001.
- [26] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1403–1410.
- [27] C. Tomasi, "Visual reconstruction: Technical perspective," *Commun. ACM*, vol. 54, no. 10, pp. 104–105, 2011.
- [28] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. Assoc. Comput. Mach.*, vol. 24, no. 6, pp. 381–395, 1981.
- [29] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. 6th IEEE ACM Int. Symp. Mixed Augmented Reality*, Nov. 2007, pp. 225–234.
- [30] V. Pradeep, C. Rhemann, S. Izadi, C. Zach, M. Bleyer, and S. Bathiche, "Monofusion: Real-time 3D reconstruction of small scenes with a single Web camera," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Oct. 2013, pp. 83–88.
- [31] H. Lim, J. Lim, and H. J. Kim, "Real-time 6-DOF monocular visual SLAM in a large-scale environment," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 1532–1539.
- [32] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "DTAM: Dense tracking and mapping in real-time," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2320–2327.
- [33] M. Pizzoli, C. Forster, and D. Scaramuzza, "REMODE: Probabilistic, monocular dense reconstruction in real time," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 2609–2616.
- [34] R. A. Newcombe and A. J. Davison, "Live dense reconstruction with a single moving camera," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 1498–1505.
- [35] M. Lourakis and X. Zabulis, "Accurate scale factor estimation in 3D reconstruction," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, 2013, pp. 498–506.
- [36] J. Mustaniemi, J. Kannala, S. Särkkä, J. Matas, and J. Heikkilä, "Inertial-based scale estimation for structure from motion on mobile devices," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 4394–4401.

- [37] P. Tanskanen, K. Kolev, L. Meier, F. Camposeco, O. Saurer, and M. Pollefeys, "Live metric 3D reconstruction on mobile phones," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 65–72.
- [38] K. Wu, K. Di, X. Sun, W. Wan, and Z. Liu, "Enhanced monocular visual odometry integrated with laser distance meter for astronaut navigation," *Sensors*, vol. 14, no. 3, pp. 4981–5003, Mar. 2014.
- [39] N. Pears, Y. Liu, and P. Bunting, *3D Imaging, Analysis and Applications*. London, U.K.: Springer, 2012, pp. 95–103.
- [40] D. Nister, "An efficient solution to the five-point relative pose problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 756–770, Jun. 2004.
- [41] (2007). *VLFeat*. Accessed: Feb. 10, 2019. [Online]. Available: <https://www.vlfeat.org/index.html>
- [42] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [43] I. R. Cole, "Modelling CPV," Ph.D. dissertation, Dept. Mech. Eng., Loughborough Univ., Loughborough, U.K., 2015. [Online]. Available: <https://repository.lboro.ac.uk/articles/ModellingCPV/9523520>
- [44] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, "On the shape of a set of points in the plane," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 4, pp. 551–559, Jul. 1983.
- [45] I. Haller and S. Nedeveschi, "GPU optimization of the SGM stereo algorithm," in *Proc. IEEE 6th Int. Conf. Intell. Comput. Commun. Process.*, Aug. 2010, pp. 197–202, doi: 10.1109/ICCP.2010.5606438.
- [46] M. I. Rajab, M. H. Alawi, and M. A. Saif, "Application of image processing to measure road distresses," *WSEAS Trans. Inf. Sci. Appl.*, vol. 5, no. 1, pp. 1–7, 2008.
- [47] (2015). *Camera Calibration Toolbox for MATLAB*. Accessed: Jan. 10, 2019. [Online]. Available: [https://www.vision.caltech.edu/bouguetj/calib\\_doc/index.html](https://www.vision.caltech.edu/bouguetj/calib_doc/index.html)



**Adeel Ahmed** was born in Haripur, Pakistan. He received the B.S. degree in mechatronics engineering from the University of Engineering and Technology at Peshawar, Peshawar, Pakistan, in 2014, and the M.S. degree in mechatronics engineering from the National University of Sciences and Technology, Islamabad, Pakistan, in 2018. His research interests include image processing, machine vision, and intelligent machines.



**Moez Ashfaq** was born in Gujrat, Pakistan. He received the B.E. degree in mechatronics engineering from the National University of Sciences and Technology, Islamabad, in 2017, where he is currently pursuing the master's degree in robotics and intelligent machine learning. His research interests include computer vision, machine learning, 3D-reconstruction, and robotics.



**Muhammad Uzair Ulhaq** was born in Islamabad, Pakistan. He received the B.E. degree in mechatronics engineering and the M.S. degree in biomedical engineering from the National University of Sciences and Technology, Islamabad, in 2017 and 2020, respectively.

From 2018 to 2020, he has worked as a Research Assistant with the Research Center for Modeling and Simulation, National University of Sciences and Technology. Since 2019, he has been working as a Computer Vision Engineer with Strada Imaging. His research interests include computer vision, deep learning, and robotics.



**Senthan Mathavan** (Member, IEEE) was born in Jaffna, Sri Lanka. He received the B.Sc. degree in mechanical engineering from the University of Peradeniya, Peradeniya, Sri Lanka, in 2005, and the Ph.D. degree in mechatronics from Loughborough University, Loughborough, U.K., in 2009.

From 2005 to 2006, he was an Assistant Lecturer with the Mechanical Engineering Department, University of Peradeniya. From 2010 to 2011, he held a Rolls-Royce funded post-doctoral research position with Loughborough University. He was a Mechatronics Engineer with ASML, The Netherlands, for five years. He is currently a Systems Architect with Nobleo Technology, The Netherlands. Since 2012, he has been also a Visiting Fellow with Nottingham Trent University, Nottingham, U.K. He has carried out research and development in the aerospace, semiconductor, and transportation industries. His research interests include mechatronics, machine vision, and applied artificial intelligence.



**Khurram Kamal** was born in Karachi, Pakistan. He received the B.E. degree in mechanical engineering from the NED University of Engineering and Technology, Karachi, in 2001, the master's degree in computer science from the University of Karachi, Karachi, in 2003, and the Ph.D. degree in mechatronics from Loughborough University, Loughborough, U.K., in 2008.

Prior to his Ph.D. research, he has been the Assistant Manager of production planning and control with the Press Shop Department, Pakistan Suzuki Motor Company Ltd., from 2003 to 2004. From 2009 to 2010, he was an Assistant Professor with the Department of Mechatronics Engineering, National University of Sciences and Technology, Islamabad, Pakistan. From 2010 to 2012, he was a Senior Lecturer with the Department of Mechatronics Engineering, Universiti Teknologi, Johor Bahru, Malaysia. He is currently an Assistant Professor with the Department of Mechatronics Engineering, National University of Sciences and Technology. His research interests include artificial intelligence, image processing, embedded systems, intelligent machines, and condition-based maintenance.



**Mujib Rahman** received the B.S. degree in civil engineering from Bangladesh Engineering University and Technology, Dhaka, Bangladesh, in 1995, the M.Sc. degree in structural engineering from City University London, London, U.K., in 1998, and the Ph.D. degree from The University of Nottingham, Nottingham, U.K., in 2004.

From September 2000 to August 2003, he was with the Nottingham Transportation Engineering Centre, The University of Nottingham, first as a Research Assistant and then as a Research Associate. He then spent five years in the industry, as the Project Manager and a Lead Engineer, and coordinated pavement design teams for several design and rehabilitation projects for national and international clients. He is currently a Senior Lecturer of civil engineering with the School of Engineering and Design, Brunel University London, Uxbridge, U.K. He has been actively involved in pavement engineering related research, consultancy, and teaching for more than 12 years.

Dr. Rahman is a member of the Institution of Civil Engineers and The Chartered Institution of Highways and Transportation and a fellow of The Higher Education Academy, U.K. He is also a Chartered Engineer.