Rethinking Road Surface 3-D Reconstruction and Pothole Detection: From Perspective Transformation to Disparity Map Segmentation

Rui Fan, *Member, IEEE*, Umar Ozgunalp, *Member, IEEE*, Yuan Wang,
Ming Liu, *Senior Member, IEEE*, and Ioannis Pitas, *Fellow, IEEE*

***Abstract*—Potholes are one of the most common forms of road damage, which can severely affect driving comfort, road safety, and vehicle condition. Pothole detection is typically performed by either structural engineers or certified inspectors. However, this task is not only hazardous for the personnel but also extremely time consuming. This article presents an efficient pothole detection algorithm based on road disparity map estimation and segmentation. We first incorporate the stereo rig roll angle into shifting distance calculation to generalize perspective transformation. The road disparities are then efficiently estimated using semiglobal matching. A disparity map transformation algorithm is then performed to better distinguish the damaged road areas. Subsequently, we utilize simple linear iterative clustering to group the transformed disparities into a collection of superpixels. The potholes are finally detected by finding the superpixels, whose intensities are lower than an adaptively determined threshold. The proposed algorithm is implemented on an NVIDIA RTX 2080 Ti GPU in CUDA. The experimental results demonstrate that our proposed road pothole detection algorithm achieves state-of-the-art accuracy and efficiency.**

***Index Terms*—Disparity map transformation, perspective transformation, pothole detection, simple linear iterative clustering.**

I. Introduction

A POTHOLE is a considerably large structural road failure, caused by the contraction and expansion of rainwater that permeates the ground under the road surface [1]. Frequently inspecting roads and repairing potholes is crucial for road maintenance [2]. Potholes are regularly detected and reported by certified inspectors and structural engineers [3]. However, this process is not only time consuming and costly but also dangerous for the personnel [4]. In addition, such detection is always qualitative and subjective because decisions depend entirely on an individual's experience [5]. Therefore, there is an ever-increasing need to develop a robust and precise automated road condition assessment system that can detect potholes both quantitatively and objectively [6].

Over the past decade, various technologies, such as vibration sensing, active sensing, and passive sensing, have been utilized to acquire road data and detect road damage [7]. Fox *et al.* [8], for example, developed a crowdsourcing system to detect and localize potholes by analyzing the accelerometer data obtained from multiple vehicles. While vibration sensors are cost effective and only require small storage space, the pothole shape and volume cannot be explicitly inferred from the vibration sensor data [4]. In addition, road hinges and joints are often mistaken for potholes [3]. Therefore, researchers have been focusing on developing pothole detection systems based on active and passive sensing. Tsai and Chatterjee [9], for instance, mounted two laser scanners on the Georgia Institute of Technology Sensing Vehicle (GTSV) to collect 3-D road data for pothole detection. However, such vehicles are not widely used, because of high equipment purchase costs and long-term maintenance costs [4].

The most commonly used passive sensors for pothole detection include Microsoft Kinect and other types of digital cameras [10]. A Kinect was used to acquire road depth information, and image segmentation algorithms were applied for pothole detection in [11]. However, the Kinect is not designed for outdoor use and often fails to perform well when exposed to direct sunlight, resulting in wrong (zero) depth values [11]. Therefore, it is more effective to detect potholes using digital cameras, as they are cost effective and capable of working in outdoor environments [4]. Given the dimensions of the acquired road data, passive sensing (computer vision) approaches [10] are generally grouped into

Manuscript received April 11, 2019; revised August 22, 2019 and August 6, 2020; accepted January 1, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant U1713211; in part by the Collaborative Research Fund by Research Grants Council Hong Kong under Project C4063-18G; and in part by the HKUST-SJTU Joint Research Collaboration Fund under Project SJTU20EG03, awarded to Prof. M. Liu. This article was recommended by Associate Editor N. R. Pal. *(Corresponding author: Rui Fan.)*

Rui Fan is with the Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093 USA, and also with the Department of Ophthalmology, University of California San Diego, La Jolla, CA 92093 USA (e-mail: rui.fan@ieee.org).

Umar Ozgunalp is with the Department of Electrical and Electronics Engineering, Cyprus International University, 518040 Mersin, Turkey (e-mail: uozgunalp@ciu.edu.tr).

Yuan Wang is with the Industrial R&D Center, SmartMore, Shenzhen, China (e-mail: yuan.wang@smartmore.com).

Ming Liu is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: eelium@ust.hk).

Ioannis Pitas is with the Department of Informatics, University of Thessaloniki, 541 24 Thessaloniki, Greece (e-mail: pitas@csd.auth.gr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2021.3060461.

Digital Object Identifier 10.1109/TCYB.2021.3060461

two categories: 1) 2-D vision-based and 2) 3-D reconstruction-based approaches [12].

The 2-D vision-based road pothole detection algorithms generally comprise three steps: 1) image segmentation; 2) contour extraction; and 3) object recognition [3]. These methods are usually developed based on the following hypotheses [12].

1) Potholes are concave holes.
2) The pothole texture is grainier and coarser than that of the surrounding road surface.
3) The intensities of the pothole region-of-interest (RoI) pixels are typically lower than those of the surrounding road surface, due to shadows.

Basic image segmentation algorithms are first applied on RGB or grayscale road surface images to separate the damaged and undamaged road areas. The most commonly used segmentation algorithms are triangle thresholding [13] and Otsu's thresholding [14]. Compared with the former, Otsu's thresholding algorithm minimizes the intraclass variance and exhibits better damaged road region detection accuracy [15]. Next, image filtering [16], edge detection [17], region growing [18], and morphological operations [19] are utilized to reduce redundant information (typically noise) and clarify the potential pothole RoI contour [5]. The resulting pothole RoI is then modeled by an ellipse [1], [9], [12], [16]. Finally, the image texture within this elliptical region is compared with that of the surrounding road region. If the elliptical RoI has a coarser and grainier texture than that of the surrounding region, a pothole is considered to have been detected [12].

Although such 2-D computer vision methods can recognize road potholes with low computational complexity, the achieved detection and localization accuracy is still far from satisfactory [11], [12]. In addition, since the actual pothole contour is always irregular, the geometric assumptions made in the contour extraction step can be ineffective. Furthermore, visual environment variability, such as road image texture, also significantly affects segmentation results [20]. Therefore, machine-learning methods [21]–[23] have been employed for better road pothole detection accuracy. For example, AdaBoost [21] was utilized to determine whether or not a road image contains a damaged road RoI. Bray *et al.* [22] also trained a neural network (NN) to detect and classify road damage. However, supervised classifiers require a large amount of labeled training data, and such data labeling procedures can be very labor intensive [5].

3-D pothole models cannot be obtained by using only a single image. Therefore, depth information has proven to be more effective than RGB information for detecting gross road damages, for example, potholes [6]. Therefore, the main purpose of this article is to present a novel road pothole detection algorithm based on its 3-D geometry reconstruction. Multiple (at least two) camera views are required to this end [24]. Images from different viewpoints can be captured using either a single moving camera or a set of synchronized multiview cameras [4]. In [25], a single camera was mounted at the rear of the car to capture the visual 2-D road footage. Then, scale-invariant feature transform (SIFT) [26] feature points are extracted in each video frame. The matched SIFT feature correspondences on two consecutive video frames are used to find the fundamental matrix. Then, cost energy related to all of the fundamental matrices was minimized using bundle adjustment. Each camera pose was, therefore, refined, and the 3-D geometry was reconstructed in a structure from motion (SfM) manner [24], [25]. However, SfM can only acquire sparse point clouds, which renders pothole detection infeasible. Therefore, pothole detection using stereo vision technology has been researched in recent years, as it can provide dense disparity maps [4].

The first reported effort in employing stereo vision for road damage detection utilized a camera pair and a structured light projector to acquire 3-D crack and pothole models [27]. In recent years, surface modeling (SM) has become a popular and effective technique for pothole detection [28]–[30]. In [28], the road surface point cloud was represented by a quadratic model. Then, pothole detection was straightforwardly realized by finding the points whose height is lower than those of the modeled road surface. In [29], this approach was improved by adding a smoothness term to the residual function that is related to the planar patch orientation. This greatly minimizes the outlier effects caused by obstacles and can, therefore, provide more precise road surface modeling results. However, finding the best value for the smoothness term is a challenging task, as it may vary from case to case [30]. Similarly, random sample consensus (RANSAC) was utilized to reduce outlier effects, while fitting a quadratic surface model to a disparity map rather than a point cloud [30]. This helps the RANSAC-SM algorithm perform more accurately and efficiently than the methods in both [28] and [29].

Road surface modeling and pothole detection are still open research. One problem is that the actual road surface is sometimes uneven, which renders quadratic surface modeling somewhat problematic. Moreover, although comprehensive studies of 2-D and 3-D computer vision techniques for pothole detection have been made, these two categories are usually implemented independently [5]. Their combination, however, can possibly advance the current state-of-the-art to achieve highly accurate pothole detection results. For instance, our recent work [31] combined both iterative disparity transformation and 3-D road surface modeling for pothole detection. Although [31] is computationally intensive, its achieved successful detection rate and the overall pixel-level accuracy are much higher than those of both [28] and [30].

Therefore, in this article, we present an efficient and robust road surface 3-D reconstruction and pothole detection algorithm based on road disparity map estimation and segmentation. The block diagram of our proposed road surface 3-D reconstruction and pothole detection algorithm is shown in Fig. 1. We first generalize the perspective transformation (PT) algorithm proposed in [4], by incorporating the stereo rig roll angle into the PT parameter estimation process, which not only increases the disparity estimation accuracy but also reduces its computational complexity [4]. Due to its inherent parallel efficiency, semiglobal matching (SGM) [32] is utilized for dense subpixel disparity map estimation. A fast disparity transformation (DT) algorithm is then performed on the estimated subpixel disparity maps to better distinguish between damaged and undamaged road regions, where an energy function

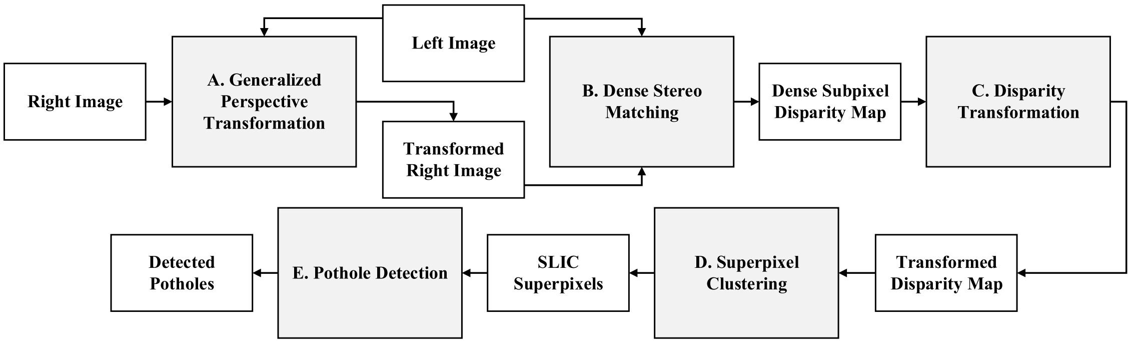


Fig. 1. Block diagram of our proposed road surface 3-D reconstruction and pothole detection system.

with respect to (w.r.t.) the stereo rig roll angle and the road disparity projection model is minimized. Finally, we use a simple linear iterative clustering (SLIC) algorithm [33] to group the transformed disparities into a collection of superpixels. The potholes are subsequently detected by finding the superpixels, whose values are lower than an adaptively determined threshold. Different potholes are also labeled using connected component labeling (CCL) [15].[1]

The remainder of this article continues in the following manner. Section II introduces the proposed road surface 3-D reconstruction and pothole detection system. Section III presents the experimental results and discusses the performance of the proposed system. In Section IV, we discuss the practical application of our system. Finally, Section V concludes this article and provides recommendations for future work.

II. Algorithm Description

A. Generalized Perspective Transformation

Road pothole detection focuses entirely on the road surface, which can be treated as a ground plane. Referring to the u-v-disparity analysis provided in [34], when the stereo rig is perfectly parallel to the road surface (stereo rig roll angle $\phi = 0$), the road disparity projections in the v-disparity domain can be represented by a straight line: $f(\mathbf{a}, \mathbf{p}) = a_0 + a_1 v$, where $\mathbf{a} = [a_0, a_1]^\top$ and $\mathbf{p} = [u, v]^\top$ is a pixel in the disparity map. $\mathbf{a}$ can be obtained by minimizing [35]

$$E_0 = \|\mathbf{d} - [\mathbf{1}_k \quad \mathbf{v}]\mathbf{a}\|_2^2 \tag{1}$$

where $\mathbf{d} = [d_1, \ldots, d_k]^\top$ is a k-entry vector of road disparity values, $\mathbf{1}_k$ is a k-entry vector of ones, and $\mathbf{v} = [v_1, \ldots, v_k]^\top$ is a k-entry vector of the vertical coordinates of the road disparities. However, in practice, ϕ is always nonzero, resulting in the road disparity map to be rotated by ϕ around the image center. This leads to a gradual disparity change in the horizontal direction, as shown in Fig. 2(d). Applying an inverse rotation by ϕ on the original road disparity map yields [36]

$$\mathbf{p}' = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \mathbf{p} \tag{2}$$

where $\mathbf{p}'$ represents the corresponding pixel of $\mathbf{p}$ in the rotated road disparity map. Therefore, the road disparity projections in the v-disparity domain can be represented by [37]

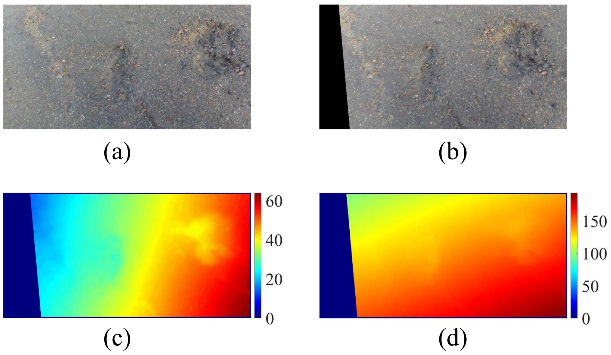


Fig. 2. (a) Original left road image. (b) Transformed right road image. (c) Estimated dense disparity map w.r.t. (a) and (b). (d) Estimated dense disparity map w.r.t. the original left and right road images.

$$f(\mathbf{a}, \mathbf{p}, \phi) = a_0 + a_1(v\cos\phi - u\sin\phi). \tag{3}$$

Compared to [4], (3) depicts PT in a more general way, as ϕ is also considered in the $\mathbf{a}$ estimation. The estimation of $\mathbf{a}$ and ϕ will be discussed in Section II-C. Equation (1) can, therefore, be rewritten as follows [36]:

$$E_0 = \|\mathbf{d} - \mathbf{T}(\phi)\mathbf{a}\|_2^2 \tag{4}$$

where

$$\mathbf{T}(\phi) = [\mathbf{1}_k \quad \cos\phi\mathbf{v} - \sin\phi\mathbf{u}] \tag{5}$$

and $\mathbf{u} = [u_1, \ldots, u_k]^\top$ is a k-entry vector of the horizontal coordinates of the road disparities. PT is then realized by shifting each pixel on row v in the right image $\kappa(\mathbf{a}, \mathbf{p}, \phi)$ pixels to the left, where κ can be computed using

$$\kappa(\mathbf{a}, \mathbf{p}, \phi) = \min_{x=0}^{W}[a_0 + a_1(v\cos\phi - x\sin\phi) - \delta_{\mathrm{PT}}] \tag{6}$$

where W denotes the image width and δ_{PT} is a constant set to ensure that the values in the disparity map [see Fig. 2(c)] w.r.t. the original left road image [see Fig. 2(a)] and the transformed road right image [see Fig. 2(b)] are non-negative.

B. Dense Stereo Matching

In our previous publication [4], PT-SRP, an efficient subpixel dense stereo matching algorithm was proposed to reconstruct the 3-D road geometry. Although the achieved 3-D

[1] Our project webpage is at ruirangerfan.com/projects/tcyb2021-rethinking.

geometry reconstruction accuracy is higher than 3 mm, the propagation strategy employed in this algorithm is not suitable for parallel processing on GPUs [38]. In [38], we proposed PT-FBS, a GPU-friendly road disparity estimation algorithm, which has been proven to be a good solution to the energy minimization problem in the fully connected Markov random field (MRF) models [39]. However, its cost aggregation process is still very computationally intensive. Therefore, in this article, we use SGM [32] together with our generalized PT algorithm for road disparity estimation.

In SGM [32], the process of disparity estimation is formulated as an energy minimization problem as follows:

$$E_1 = \sum_{\mathbf{p}} \Bigg(c(\mathbf{p}, d_{\mathbf{p}}) + \sum_{\mathbf{q}\in\mathcal{N}_{\mathbf{p}}} \lambda_1 \delta\big(|d_{\mathbf{p}} - d_{\mathbf{q}}| = 1\big) + \sum_{\mathbf{q}\in\mathcal{N}_{\mathbf{p}}} \lambda_2 \delta\big(|d_{\mathbf{p}} - d_{\mathbf{q}}| > 1\big) \Bigg) \tag{7}$$

where c denotes the stereo matching cost and $\mathbf{q}$ represents a pixel in $\mathcal{N}_{\mathbf{p}}$ (the neighborhood system of $\mathbf{p}$). $d_{\mathbf{p}}$ and $d_{\mathbf{q}}$ are the disparities of $\mathbf{p}$ and $\mathbf{q}$, respectively. λ_1 penalizes the neighboring pixels with small disparity differences, that is, one pixel; and λ_2 penalizes the neighboring pixels with large disparity differences, that is, larger than one pixel. $\delta(\cdot)$ returns 1 if its argument is true and 0 otherwise. However, (7) is a complex NP-hard problem [32]. Therefore, in practical implementation, (7) is solved by aggregating the stereo matching costs along all directions in the image using dynamic programming [32]

$$\begin{aligned} c^{\mathbf{r}}_{\text{agg}}(\mathbf{p}, d_{\mathbf{p}}) = c(\mathbf{p}, d_{\mathbf{p}}) &+ \min\Bigg(c^{\mathbf{r}}_{\text{agg}}(\mathbf{p}-\mathbf{r}, d_{\mathbf{p}}), \\ &\bigcup_{k\in\{-1,1\}} c^{\mathbf{r}}_{\text{agg}}(\mathbf{p}-\mathbf{r}, d_{\mathbf{p}}+k) \\ &+ \lambda_1, \min_i c^{\mathbf{r}}_{\text{agg}}(\mathbf{p}-\mathbf{r}, i) + \lambda_2 \Bigg) \end{aligned} \tag{8}$$

where $c^{\mathbf{r}}_{\text{agg}}(\mathbf{p}, d_{\mathbf{p}})$ represents the aggregated stereo matching cost at $\mathbf{p}$ in the direction of $\mathbf{r}$. $\tilde{d}_{\mathbf{p}}$, the optimum disparity at $\mathbf{p}$, can, therefore, be estimated by solving

$$\tilde{d}_{\mathbf{p}} = \min \sum_{\mathbf{r}} c^{\mathbf{r}}_{\text{agg}}(\mathbf{p}, d_{\mathbf{p}}). \tag{9}$$

The estimated road disparity map D_0 w.r.t. Fig. 2(a) and (b) is shown in Fig. 2(c). Since the right road image has been transformed into the left view using PT in Section II-A, the road disparity map D_1 w.r.t. the original left and right road images, as shown in Fig. 2(d), can be obtained by using

$$D_1(\mathbf{p}) = D_0(\mathbf{p}) + \kappa(\mathbf{a}, \mathbf{p}, \phi). \tag{10}$$

C. Disparity Transformation

As discussed in Section II-A, the road disparity projection can be represented using (3), which has a closed-form solution as follows [31]:

$$\mathbf{a}(\phi) = \Big(\mathbf{T}(\phi)^{\top}\mathbf{T}(\phi)\Big)^{-1}\mathbf{T}(\phi)^{\top}\mathbf{d}. \tag{11}$$

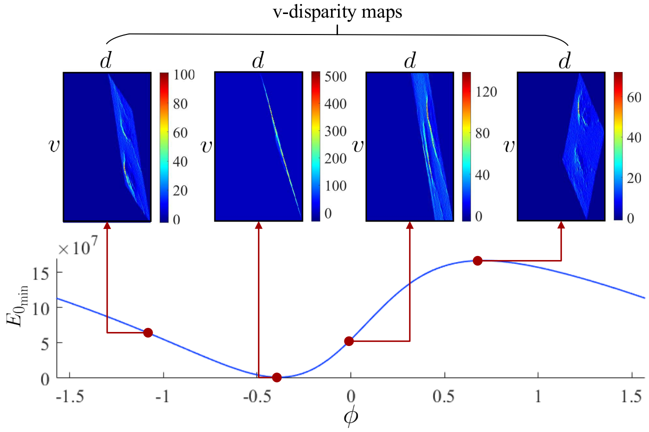
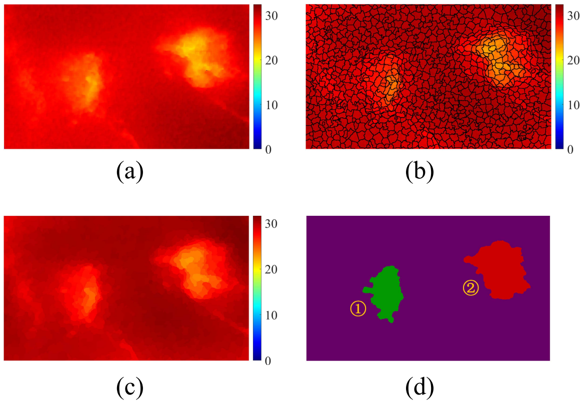


Fig. 3. $E_{0_{\min}}$ w.r.t. different estimated ϕ. Ground-truth $\phi \simeq -0.41$.



(a) (b)

(c) (d)

Fig. 4. (a) Transformed road disparity map. (b) Superpixel clustering result. (c) Superpixel-clustered transformed road disparity map. (d) Pothole detection result, where different potholes are shown in different colors.

$E_{0_{\min}}$ (the minimum E_0) has the following expression [35]:

$$E_{0_{\min}}(\phi) = \mathbf{d}^{\top}\mathbf{d} - \mathbf{d}^{\top}\mathbf{T}(\phi)\Big(\mathbf{T}(\phi)^{\top}\mathbf{T}(\phi)\Big)^{-1}\mathbf{T}(\phi)^{\top}\mathbf{d}. \tag{12}$$

Compared to the case when the stereo rig is perfectly parallel to the road surface, a nonzero roll angle results in a much higher $E_{0_{\min}}$ [38], as shown in Fig. 3. ϕ can be estimated by minimizing (12), which is equivalent to solving $\partial E_{0_{\min}}/\partial\phi = 0$ and finding its minima [37]. Its solution is given in [37]. The disparity transformation can then be realized using [40]

$$D_2(\mathbf{p}) = D_1(\mathbf{p}) - f(\mathbf{a}, \mathbf{p}, \phi) + \delta_{\text{DT}} \tag{13}$$

where D_2 denotes the transformed disparity map, as shown in Fig. 4(a), and δ_{DT} is a constant set to ensure that the transformed disparity values are non-negative. It can be clearly seen that the damaged road areas become highly distinguishable.

D. Superpixel Clustering

In Section II-C, the disparity transformation algorithm allows better discrimination between damaged and undamaged road areas. The road potholes can therefore be detected by applying image segmentation algorithms on the transformed disparity maps. However, the thresholds chosen in these algorithms may not be the best for optimal pothole detection

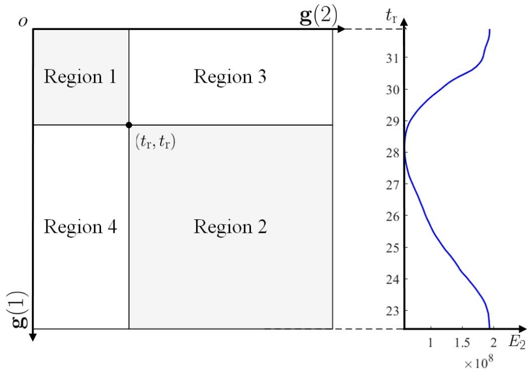


Fig. 5. Illustration of a 2-D histogram and E_2 w.r.t. t_r.

accuracy, especially when the transformed disparity histogram no longer exhibits an obvious bimodality. Furthermore, small blobs with low transformed disparity values are always mistaken for potholes, as discussed in [37]. Hence, in this article, we use superpixel clustering to group the transformed disparities into a set of perceptually meaningful regions, which are then used to replace the rigid pixel grid structure.

Superpixel generation algorithms are broadly classified as graph based [41], [42] and gradient ascent based [43]–[45]. The former treats each pixel as a node and produces superpixels by minimizing a cost function using global optimization approaches, while the latter starts from a collection of initially clustered pixels and refines the clusters iteratively until error convergence. SLIC, an efficient superpixel algorithm was introduced in [33]. It outperforms all other state-of-the-art superpixel clustering algorithms, in terms of both boundary adherence and clustering speed [33]. Therefore, it is used for transformed disparity map segmentation in our proposed road pothole detection system.

SLIC [33] begins with an initialization step, where p cluster centers are sampled on a regular grid. These cluster centers are then moved to the positions (over their eight-connected neighbors) corresponding to the lowest gradients. This not only reduces the chance of selecting noisy pixels as a superpixel but also avoids centering a superpixel on an edge [33]. In the next step, each pixel is assigned to the nearest cluster center, whose search range overlaps its location. Finally, we utilize k-means clustering to iteratively update each cluster center until the residual error between the previous and updated cluster centers converges. The corresponding SLIC [33] result is shown in Fig. 4(b), where it can be observed that each pothole consists of a group of superpixels.

E. Pothole Detection

After SLIC [33], the transformed disparity map is grouped into a set of superpixels, each of which consists of a collection of transformed disparities with similar values. Then, the value of each superpixel is replaced by the mean value of its containing transformed disparities, and a superpixel-clustered transformed disparity map D_3, as illustrated in Fig. 4(c), is obtained. Pothole detection can, therefore, be straightforwardly realized by finding a threshold t_s and selecting the superpixels, whose values are lower than t_s.

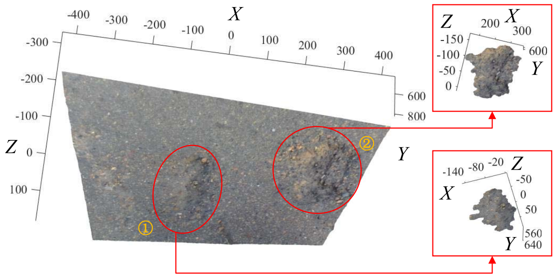


Fig. 6. Point clouds of the detected road potholes.

In this article, we first introduce a 2-D thresholding method based on k-means clustering. The proposed thresholding method hypothesizes that a transformed disparity map only contains two parts: 1) foreground (pothole) and 2) background (road surface), which can be separated using a threshold t_r. The threshold t_s for selecting the pothole superpixels is then determined as follows:

$$t_s = t_r - \delta_{\mathrm{PD}} \tag{14}$$

where δ_{PD} is a tolerance. To find the best t_r value, we formulate the thresholding problem as a 2-D vector quantization problem, where each transformed disparity $D_2(\mathbf{p})$ and its m-connected neighborhood system $\mathcal{N}_{\mathbf{p}}$ provide a vector

$$\mathbf{g} = \left[D_2(\mathbf{p}), \frac{1}{m} \sum_{\mathbf{q} \in \mathcal{N}_{\mathbf{p}}} D_2(\mathbf{q}) \right]^{\top}. \tag{15}$$

The threshold is determined by partitioning the vectors into two clusters $\mathbf{S} = \{\mathbf{S}_1, \mathbf{S}_2\}$. The vectors $\mathbf{g}$ are stored in a 2-D histogram, as shown in Fig. 5. According to the MRF theory [46], for an arbitrary point (except for the discontinuities), its transformed disparity value is similar to those of its neighbors in all directions. Therefore, we search for the threshold along the principal diagonal of the 2-D histogram, using k-means clustering. Given a threshold t_r, the 2-D histogram can be divided into four regions (see Fig. 5): regions 1 and 2 represent the foreground and the background, respectively; while regions 3 and 4 store the vectors of noisy points and discontinuities. In the proposed algorithm, the vectors in regions 3 and 4 are not considered in the clustering process. The best threshold is determined by minimizing the within-cluster disparity dispersion, as follows [47]:

$$\underset{\mathbf{S}}{\arg\min}\, E_2 = \underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{2} \sum_{\mathbf{g} \in \mathbf{S}_i} \left\| \mathbf{g} - \boldsymbol{\mu}_i \right\|^2 \tag{16}$$

where $\boldsymbol{\mu}_i$ denotes the mean of the points in $\mathbf{S}_i$. E_2 to t_r is shown in Fig. 5. The corresponding pothole detection result is shown in Fig. 4(d), where different potholes are labeled in different colors using CCL. In addition, the point clouds of the detected potholes are extracted from the 3-D road point cloud, as shown in Fig. 6.

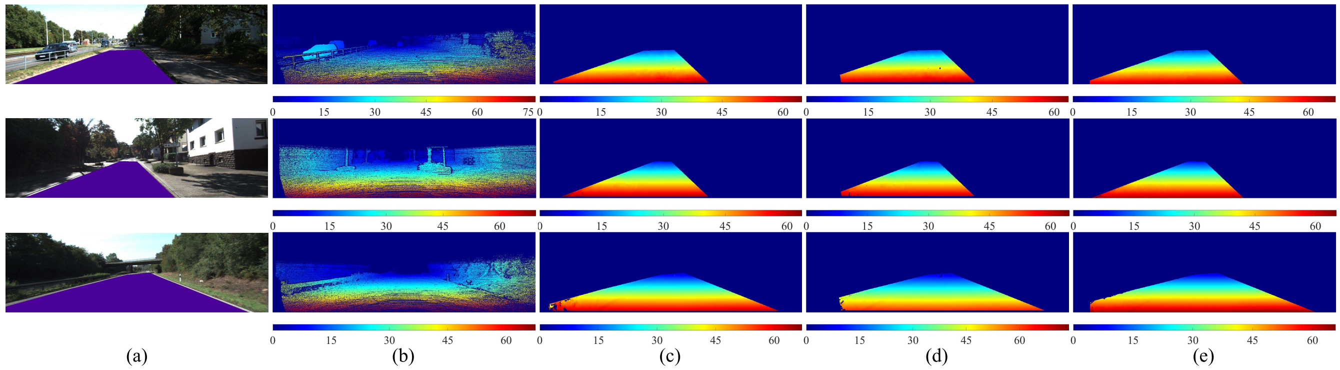


Fig. 7. Examples of dense stereo matching experimental results. (a) Left stereo images, where the areas in purple are our manually selected road RoIs. (b) Ground-truth disparity maps. (c)–(e) Disparity maps estimated using PT-SRP [4], PT-FBS [38], and GPT-SGM, respectively.

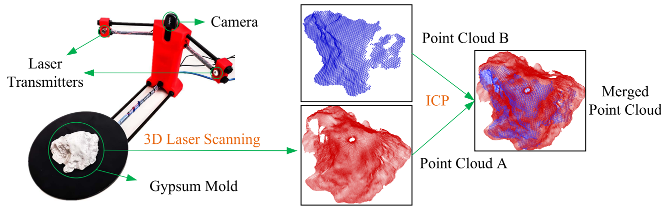


Fig. 8. Evaluation of road pothole 3-D geometry reconstruction. Point cloud *A* is acquired using a BQ Ciclop 3-D laser scanner; pothole could *B* is generated using our proposed road disparity map estimation algorithm.

TABLE I
COMPARISON OF e_p AND e_r AMONG PT-SRP [4], PT-FBS [38] AND OUR PROPOSED DENSE STEREO ALGORITHM

Algorithm	e_{p} (%)			e_{r} (pixels)
	$\varepsilon = 1$	$\varepsilon = 2$	$\varepsilon = 3$	
PT-SRP [4]	5.0143	0.3913	0.0588	0.4237
PT-FBS [38]	**4.5979**	0.2174	0.0227	0.4092
GPT-SGM (proposed)	4.6069	**0.1859**	**0.0083**	**0.4079**

III. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of our proposed pothole detection algorithm both qualitatively and quantitatively. The proposed algorithm was implemented in CUDA on an NVIDIA RTX 2080 Ti GPU. The following two sections, respectively, evaluate the performances of the road surface 3-D reconstruction and road pothole detection subsystems.

A. Road Surface 3-D Reconstruction Evaluation

In our experiments, we utilized a stereo camera to capture synchronized stereo road image pairs. Our road pothole detection datasets are publicly available at: https://github.com/ruirangerfan/stereo_pothole_datasets.

1) Dense Stereo Matching Evaluation: Since the road pothole detection datasets we created do not contain the disparity ground truth, the KITTI stereo 2012 [48] and stereo 2015 [49] datasets are used to evaluate the accuracy of our proposed stereo matching algorithm (GPT-SGM). As GPT-SGM only aims at estimating road disparities, we manually selected a road RoI (see the purple areas in the first column of Fig. 7) in each road image to evaluate the disparity estimation accuracy.

Two metrics are used to measure disparity estimation accuracy.

1) Percentage of error pixels (PEP) [50]

$$e_{\mathrm{p}} = \frac{1}{q}\sum_{\mathbf{p}} \delta(|D_1(\mathbf{p}) - D_4(\mathbf{p}))| > \varepsilon) \times 100\% \tag{17}$$

where q denotes the total number of disparities used for accuracy evaluation, D_4 represents the ground-truth disparity map, ε denotes the disparity error tolerance, and δ is introduced in Section II.

2) Root mean-squared error (RMSE) [51]

$$e_{\mathrm{r}} = \sqrt{\frac{1}{q}\sum_{\mathbf{p}}(D_1(\mathbf{p}) - D_4(\mathbf{p}))^2}. \tag{18}$$

Furthermore, we also compare our algorithm with PT-SRP [4] and PT-FBS [38]. The experimental results are given in Fig. 7. Their comparisons w.r.t. different e_{p} and e_{r} are shown in Table I, where we can observe that GPT-SGM outperforms PT-SRP [4] and PT-FBS [38] in terms of e_{p} when $\varepsilon = 2$ and $\varepsilon = 3$, while PT-FBS [38] performs slight better than GPT-SGM when $\varepsilon = 1$. Compared with PT-FBS [38], the value of e_{p} obtained using GPT-SGM reduces by 14.5% ($\varepsilon = 2$) and 63.4% ($\varepsilon = 3$). In addition, compared with PT-FBS [38], when $\varepsilon = 1$, e_{p} obtained using GPT-SGM increases by only 0.2%. Furthermore, GPT-SGM achieves the lowest e_{r} value (~0.4079 pixels). Therefore, the overall performance of GPT-SGM is better than both PT-SRP [4] and PT-FBS [38]. Our proposed GPT-SGM algorithm runs at a speed of 98 fps on an NVIDIA RTX 2080 Ti GPU.

2) 3-D Road Geometry Reconstruction Evaluation: To acquire the pothole point cloud ground truth, we first poured gypsum plaster into a pothole and dug the gypsum mold out, when it had become dry and had hardened. Then, the 3-D pothole model was acquired using a BQ Ciclop 3-D laser scanner. The laser scanner is equipped with a Logitech C270 HD camera and two one-line laser transmitters. The camera captured the reflected laser pulses from the gypsum mold and constructed its 3-D model using the laser calibration parameters. An example of the BQ Ciclop 3-D laser scanner and the created pothole ground truth is shown in Fig. 8. Next, we

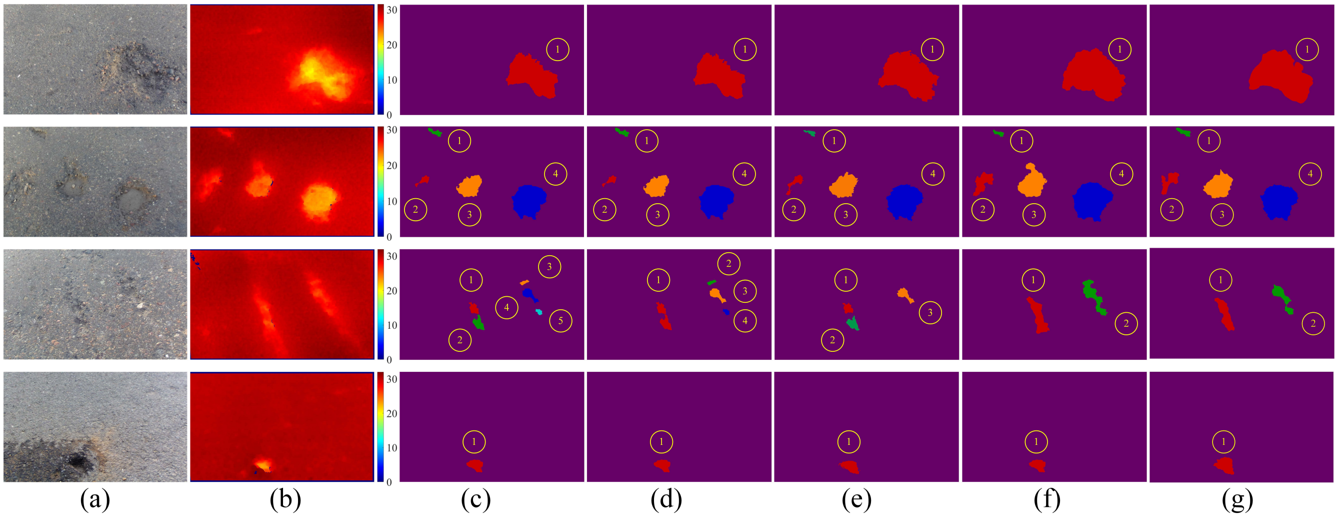


Fig. 9. Experimental results of road pothole detection. (a) Left road images. (b) Transformed road disparity maps. (c)–(f) Results obtained using [28], [30], [31], and the proposed algorithm, respectively. (g) Road pothole ground truth.

utilized the iterative closest point (ICP) algorithm [52] to register point cloud A and point cloud B, which were acquired using laser scanning and dense stereo matching, respectively. In order to improve the performance of the ICP algorithm, we first transformed the road surface point cloud to make it as close as possible to the XZ plane. This transformation can be straightforwardly realized using the camera height, roll, and pitch angles. The merged pothole point cloud is shown in Fig. 8. To quantify the accuracy of the pothole 3-D geometry reconstruction, we measured the root-mean-squared closest distance error e_c

$$e_c = \sqrt{\frac{1}{q}\sum_{i=1}^{q}\left\|\mathbf{P}_{A_i} - \mathbf{P}_{B_i}\right\|_2^2} \tag{19}$$

where $\mathbf{P}_B$ denotes a 3-D point in the generated pothole point cloud, $\mathbf{P}_A$ denotes the closest point to $\mathbf{P}_B$ in the ground truth, and q denotes the total number of points used for evaluation. The average e_c value we achieved is 2.23 mm, which is lower than what we achieved in [4].

B. Road Pothole Detection Evaluation

Examples of the detected potholes are shown in Fig. 9. In our experiments, the potholes that are either located at the image corners or composed of only one superpixel are considered to be the fake ones. To evaluate the performance of the proposed pothole detection algorithm, we first compared the detection accuracy of the proposed method with those of the algorithms in [28], [30], and [31]. The results obtained using [28], [30], and [31] are shown in (c), (d), and (e), respectively, of Fig. 9. The successful detection rates w.r.t. different algorithms and datasets are given in Table II, where we can see that the rates of [28] and [30] are 73.4% and 84.8%, respectively. The proposed algorithm can detect potholes with a better successful detection rate (98.7%). The incorrect detection occurs because the road surface at the corner of the image has a higher curvature. We believe this can be avoided by reducing the view angle.

Furthermore, we also compare these pothole detection algorithms in terms of the pixel-level precision, recall, accuracy, and F-score, defined as: precision $= [n_{\text{tp}}/(n_{\text{tp}} + n_{\text{fp}})]$, recall $= [n_{\text{tp}}/(n_{\text{tp}} + n_{\text{fn}})]$, accuracy $= [(n_{\text{tp}} + n_{\text{tn}})/(n_{\text{tp}} + n_{\text{tn}} + n_{\text{fp}} + n_{\text{fn}})]$, and F-score $= 2 \times [(\text{precision} \times \text{recall})/(\text{precision} + \text{recall})]$, where n_{tp}, n_{fp}, n_{fn} and n_{tn} represents the numbers of true-positive, false-positive, false-negative, and true-negative pixels, respectively. The comparisons w.r.t. these four performance evaluation metrics are also given in Table II, where it can be seen that the proposed algorithm outperforms [28], [30] and [31] in terms of both accuracy and F-score when processing datasets 1 and 3. Reference [31] achieves the best results on dataset 2. Our method achieves the highest overall F-score (89.42%), which is over 3% higher than that of our previous work [31].

In Table II, we also provide the runtime of [28], [30], [31], and our proposed method on the NVIDIA RTX 2080 Ti GPU. It can be seen that the proposed system performs much faster than [31]. Although our proposed method performs slower than [28] and [30], SLIC [33] takes the biggest proportion of the processing time. The total runtime of DT and pothole detection is only about 3.5 ms. Therefore, we believe by leveraging a more efficient superpixel clustering algorithm, the overall performance of our proposed pothole detection system can be significantly improved. Moreover, as discussed above, the proposed pothole detection performs much more accurately than both [28] and [30], where an increase of approximately 9% is witnessed on the F-score.

Many recent semantic image segmentation networks have been employed to detect freespace (drivable area) and road pothole/anomaly [40], [56], [57]. Therefore, we also compare the proposed algorithm with three state-of-the-art deep convolutional NNs (DCNNs): 1) fully convolutional network (FCN) [53]; 2) SegNet [54]; and 3) DeepLabv3+ [55]. Since only a very limited amount of road data are available, we employ k-fold cross-validation [59] to evaluate the performance of each DCNN, where k represents the total number of images. Each DCNN is evaluated k times. Each time,

TABLE II
ROAD POTHOLE DETECTION PERFORMANCE COMPARISON OF [28], [30], [31], AND THE PROPOSED METHOD

Dataset	Method	Correct Detection	Incorrect Detection	Misdetection	Recall	Precision	Accuracy	F-score	Runtime (ms)
Dataset 1	[28]	11	11	0	0.5199	0.5427	0.9892	0.5311	33.19
	[30]	22	0	0	0.4622	**0.9976**	0.9936	0.6317	22.90
	[31]	22	0	0	0.4990	0.9871	0.9940	0.6629	117.72
	Proposed	21	1	0	**0.7005**	0.9641	**0.9947**	**0.8114**	47.21
Dataset 2	[28]	42	10	0	0.9754	0.9712	0.9987	0.9733	30.77
	[30]	40	8	4	0.8736	**0.9907**	0.9968	0.9285	21.39
	[31]	51	1	0	**0.9804**	0.9797	**0.9991**	**0.9800**	124.53
	Proposed	52	0	0	0.9500	0.8826	0.9920	0.9150	45.32
Dataset 3	[28]	5	0	0	0.6119	0.7714	0.9948	0.6825	35.72
	[30]	5	0	0	0.5339	0.9920	0.9957	0.6942	26.24
	[31]	5	0	0	0.5819	0.9829	0.9961	0.7310	132.44
	Proposed	5	0	0	**0.7017**	**0.9961**	**0.9964**	**0.8234**	49.90
Total	[28]	58	21	0	0.7799	0.8220	0.9942	0.8004	33.23
	[30]	67	8	4	0.6948	**0.9921**	0.9954	0.8173	23.51
	[31]	78	1	0	0.7709	0.9815	**0.9964**	0.8635	124.90
	Proposed	78	1	0	**0.8903**	0.8982	0.9961	**0.8942**	47.48

TABLE III
COMPARISON OF THREE STATE-OF-THE-ART DCNNS TRAINED FOR ROAD POTHOLE DETECTION

DCNN	Disp		TDisp		RGB	
	accuracy	F-score	accuracy	F-score	accuracy	F-score
FCN [53]	**0.971**	0.606	0.983	0.797	0.949	0.637
SegNet [54]	0.966	0.516	0.979	0.753	0.894	0.463
DeepLabv3+ [55]	0.968	**0.673**	**0.987**	**0.856**	**0.977**	**0.742**

$k-1$ subsamples (disparity maps, transformed disparity maps, or RGB images) are used to train the DCNN, and the remaining subsample is retained for testing DCNN performance. Finally, the obtained k groups of evaluation results are averaged to illustrate the overall performance of the trained DCNN. The quantification results are given in Table III, where it can be observed that the DCNNs trained with the transformed disparity maps (abbreviated as TDisp) outperform themselves trained with either disparity maps (abbreviated as Disp) or RGB images (abbreviated as RGB). This demonstrates that DT makes the disparity maps become more informative. Furthermore, Deeplabv3+ [55] outperforms all other compared DCNNs for pothole detection. It can further be observed that our proposed pothole detection algorithm outperforms all the compared DCNNs in terms of both accuracy and F-score. We believe this is due to the fact that only a very limited amount of road data are available and, therefore, the advantages of DCNNs cannot be fully exploited.

IV. DISCUSSION

Potholes are typically detected by experienced inspectors in fine-weather daylight and are an extremely labor-intensive and time-consuming process. The proposed road pothole detection algorithm can perform in real time on a state-of-the-art graphics card. Compared with the state-of-the-art DCNN-based methods, our algorithm does not require labeled training data to learn a pothole detector. The accuracy we achieved is much higher than that of the existing computer vision-based pothole detection methods, especially those based on 2-D image analysis. Although computer vision-based road pothole detection has been extensively researched over the past decade, very few researchers have considered applying computer stereo vision in road pothole detection. Therefore, we created three pothole datasets using a stereo camera to contribute to the research and development of automated road pothole detection systems. In our experiments, the stereo camera was mounted at a relatively low height to the road surface, in order to increase the accuracy of disparity estimation. Our datasets can be used by other researchers to quantify the accuracy of their developed road surface 3-D reconstruction and road pothole detection algorithms.

V. CONCLUSION AND FUTURE WORK

In this article, we presented an efficient stereo vision-based road surface 3-D reconstruction and road pothole detection system. We first generalized the PT algorithm [4] by considering the stereo rig roll angle into the process of PT parameter estimation. DT made the potholes highly distinguishable from the undamaged road surface. SLIC grouped the transformed disparities in a collection of superpixels. Finally, the potholes were detected by finding the superpixels, which have lower values than an adaptive threshold determined using k-means clustering. The proposed pothole detection system was implemented in CUDA on an RTX 2080 Ti GPU. The experimental results illustrated that our system can achieve a successful detection rate of 98.7% and an F-score of 89.4%.

A challenge is that the road surface cannot always be considered as a ground plane, resulting in the wrong detection. Therefore, in future work, we will design an algorithm to segment the reconstructed road surface into different planar patches, each of which can then be processed separately using the proposed algorithm.

References

[1] J. S. Miller and W. Y. Bellinger, "Distress identification manual for the long-term pavement performance program," Infrastruct. Res. Develop., Federal Highway Admin., McLean, VA, USA, Rep. FHWA-HRT-13-092, 2014.

[2] R. Fan, "Real-time computer stereo vision for automotive applications," Ph.D. dissertation, Dept. Elect. Electron. Eng., Univ. Bristol, Bristol, U.K, 2018.

[3] 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.

[4] R. Fan, X. Ai, and N. Dahnoun, "Road surface 3D reconstruction based on dense subpixel disparity map estimation," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3025–3035, Jun. 2018.

[5] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth, "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure," *Adv. Eng. Informat.*, vol. 29, no. 2, pp. 196–210, 2015.

[6] S. Mathavan, K. Kamal, and M. Rahman, "A review of three-dimensional imaging technologies for pavement distress detection and measurements," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2353–2362, Oct. 2015.

[7] H. Wang, R. Fan, Y. Sun, and M. Liu, "Dynamic fusion module evolves drivable area and road anomaly detection: A benchmark and algorithms," *IEEE Trans. Cybern.*, 2021. doi: 10.1109/TCYB.2021.3064089.

[8] A. Fox, B. V. K. V. Kumar, J. Chen, and F. Bai, "Multi-lane pothole detection from crowdsourced undersampled vehicle sensor data," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3417–3430, Dec. 2017.

[9] Y.-C. Tsai and A. Chatterjee, "Pothole detection and classification using 3D technology and watershed method," *J. Comput. Civil Eng.*, vol. 32, no. 2, 2017, Art. no. 04017078.

[10] P. Wang, Y. Hu, Y. Dai, and M. Tian, "Asphalt pavement pothole detection and segmentation based on wavelet energy field," *Math. Problems Eng.*, vol. 2017, Feb. 2017, Art. no. 1604130.

[11] M. R. Jahanshahi, F. Jazizadeh, S. F. Masri, and B. Becerik-Gerber, "Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor," *J. Comput. Civil Eng.*, vol. 27, no. 6, pp. 743–754, 2012.

[12] 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, 2012.

[13] C. Koch and I. Brilakis, "Pothole detection in asphalt pavement images," *Adv. Eng. Informat.*, vol. 25, no. 3, pp. 507–515, 2011.

[14] E. Buza, S. Omanovic, and A. Huseinovic, "Pothole detection with image processing and spectral clustering," in *Proc. 2nd Int. Conf. Inf. Technol. Comput. Netw.*, 2013, pp. 48–53.

[15] I. Pitas, *Digital Image Processing Algorithms and Applications*. New York, NY, USA: Wiley, 2000.

[16] A. Tedeschi and F. Benedetto, "A real-time automatic pavement crack and pothole recognition system for mobile android-based devices," *Adv. Eng. Informat.*, vol. 32, pp. 11–25, Apr. 2017.

[17] S.-K. Ryu, T. Kim, and Y.-R. Kim, "Image-based pothole detection system for its service and road management system," *Math. Problems Eng.*, vol. 2015, pp. 1–10, Sep. 2015.

[18] Y. O. Ouma and M. Hahn, "Pothole detection on asphalt pavements from 2D-colour pothole images using fuzzy C-means clustering and morphological reconstruction," *Autom. Construct.* vol. 83, pp. 196–211, Nov. 2017.

[19] S. Li, C. Yuan, D. Liu, and H. Cai, "Integrated processing of image and GPR data for automated pothole detection," *J. Comput. Civil Eng.*, vol. 30, no. 6, 2016, Art. no. 04016015.

[20] E. Salari and G. Bao, "Automated pavement distress inspection based on 2D and 3D information," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, 2011, pp. 1–4.

[21] A. Cord and S. Chambon, "Automatic road defect detection by textural pattern recognition based on adaboost," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 27, no. 4, pp. 244–259, 2012.

[22] J. Bray, B. Verma, X. Li, and W. He, "A neural network based technique for automatic classification of road cracks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2006, pp. 907–912.

[23] Q. Li, Q. Zou, and X. Liu, "Pavement crack classification via spatial distribution features," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, 2011, Art. no. 649675.

[24] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[25] G. Jog, C. Koch, M. Golparvar-Fard, and I. Brilakis, "Pothole properties measurement through visual 2D recognition and 3D reconstruction," in *Proc. Comput. Civil Eng.*, 2012, pp. 553–560.

[26] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[27] A. Barsi *et al.*, "Mobile pavement measurement system: A concept study," in *Proc. ASPRS Annu. Conf. , Baltimore, MD, USA*, 2005, p. 8.

[28] Z. Zhang, "Advanced stereo vision disparity calculation and obstacle analysis for intelligent vehicles," Ph.D. dissertation, Electrical & Electronic Engineering, Univ. Bristol, Bristol, U.K., 2013.

[29] U. Ozgunalp, X. Ai, and N. Dahnoun, "Stereo vision-based road estimation assisted by efficient planar patch calculation," *Signal Image Video Process.*, vol. 10, no. 6, pp. 1127–1134, 2016.

[30] A. Mikhailiuk and N. Dahnoun, "Real-time pothole detection on TMS320c6678 DSP," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, 2016, pp. 123–128.

[31] R. Fan, U. Ozgunalp, B. Hosking, M. Liu, and I. Pitas, "Pothole detection based on disparity transformation and road surface modeling," *IEEE Trans. Image Process.*, vol. 29, no. 1, pp. 897–908, Aug. 2019.

[32] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[33] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[34] Z. Hu, F. Lamosa, and K. Uchimura, "A complete U-V-disparity study for stereovision based 3D driving environment analysis," in *Proc. 5th Int. Conf. 3-D Digit. Imag. Model. (3DIM)*, 2005, pp. 204–211.

[35] R. Fan *et al.*, "Learning collision-free space detection from stereo images: Homography matrix brings better data augmentation," 2020. [Online]. Available: arXiv:2012.07890.

[36] R. Fan, M. J. Bocus, and N. Dahnoun, "A novel disparity transformation algorithm for road segmentation," *Inf. Process. Lett.*, vol. 140, pp. 18–24, Dec. 2018.

[37] R. Fan and M. Liu, "Road damage detection based on unsupervised disparity map segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 11, pp. 4906–4911, Nov. 2019.

[38] R. Fan, J. Jiao, J. Pan, H. Huang, S. Shen, and M. Liu, "Real-time dense stereo embedded in a UAV for road inspection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, 2019, pp. 535–543.

[39] M. G. Mozerov and J. van de Weijer, "Accurate stereo matching by two-step energy minimization," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1153–1163, Mar. 2015.

[40] R. Fan, H. Wang, M. J. Bocus, and M. Liu, "We learn better road pothole detection: From attention aggregation to adversarial domain adaptation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 285–300.

[41] O. Veksler, Y. Boykov, and P. Mehrani, "Superpixels and supervoxels in an energy optimization framework," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 211–224.

[42] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[43] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.

[44] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "TurboPixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, Dec. 2009.

[45] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[46] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[47] R. Fan *et al.*, "Road crack detection using deep convolutional neural network and adaptive thresholding," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2019, pp. 474–479.

[48] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The kitti vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

[49] M. Menze, C. Heipke, and A. Geiger, "Joint 3D estimation of vehicles and scene flow," in *Proc. ISPRS Workshop Image Sequence Anal. (ISA)*, vol. II-3/W5, 2015, pp. 427–434.

[50] R. Fan, Y. Liu, M. J. Bocus, L. Wang, and M. Liu, "Real-time subpixel fast bilateral stereo," in *Proc. IEEE Int. Conf. Inf. Autom. (ICIA)*, 2018, pp. 1058–1065.

[51] S. Daniel and S. Richard, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, 2002.

[52] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Sensor Fusion IV: Control Paradigms and Data Structures*, vol. 1611. Bellingham, WA, USA: Int. Soc. Opt. Photon., 1992, pp. 586–607.

[53] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[54] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[55] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[56] R. Fan, H. Wang, P. Cai, and M. Liu, "SNE-RoadSeg: Incorporating surface normal information into semantic segmentation for accurate freespace detection," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 340–356.

[57] H. Wang, R. Fan, Y. Sun, and M. Liu, "Applying surface normal information in drivable area and road anomaly detection for ground mobile robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2020, pp. 2706–2711.

[58] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2015, pp. 234–241.

[59] S. Geisser, "The predictive sample reuse method with applications," *J. Amer. Stat. Assoc.*, vol. 70, no. 350, pp. 320–328, 1975.

Rui Fan (Member, IEEE) received the B.Eng. degree in automation (control science and engineering) from the Harbin Institute of Technology, Harbin, China, in 2015, and the Ph.D. degree in electrical and electronic engineering from the University of Bristol (supervised by Prof. J. G. Rarity and Dr. N. Dahnoun), Bristol, U.K., in 2018.

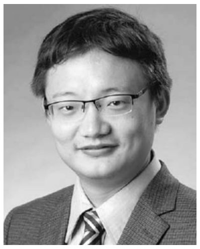
From 2018 to 2020, he was the Deputy Director of the Robotics and Multiperception Laboratory, as well as the Research Associate with the Robotics Institute and the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong. He co-found ATG Robotics in 2019 and has been working as their Chief Scientist since 2019. Since 2020, he has been a Postdoctoral Fellow, supervised by Prof. L. Zangwill and Prof. D. J. Kriegman, with the Department of Ophthalmology and the Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA. He will join the Department of Control Science and Engineering, School of Electronic and Information Engineering, and Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai, China, as a Research Full Professor in Fall 2021. He is the General Chair of the Autonomous Vehicle Vision (AVVision) community and the Director of Machine Intelligence and Autonomous Systems Research Group. His research interests include computer vision, machine/deep learning, image/signal processing, autonomous driving, and bioinformatics.

Umar Ozgunalp (Member, IEEE) received the B.Sc. degree (Hons.) in electrical and electronic engineering from Eastern Mediterranean University, Famagusta, Cyprus, in 2007, the M.Sc. degree in electronic communications and computer engineering from the University of Nottingham, Nottingham, U.K., in 2009, and the Ph.D. degree from the University of Bristol, Bristol, U.K., in 2016.

He is currently an Assistant Professor with the Department of Electrical and Electronics Engineering, Cyprus International University, Mersin, Turkey. His research interests include computer vision, pattern recognition, and intelligent vehicles.

Yuan Wang received the B.Sc. degree in telecommunication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2016, and the M.Sc. degree in telecommunication engineering from the Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2017.

From 2018 to 2019, he worked as a Research Assistant with the Robotics and Multiperception Laboratory, HKUST. Since 2020, he has been working with SmartMore, Shenzhen, China. His research interests include semantic segmentation, 2-D/3-D object detection, and graph neural networks.



Ming Liu (Senior Member, IEEE) received the B.A. degree in automation from Tongji University, Shanghai, China, in 2005, and the Ph.D. degree from the Department of Mechanical and Process Engineering, ETH Zurich, in 2013, supervised by Prof. R. Siegwart.

He is currently an Associate Professor with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong. He is involved in several NSF projects, and National 863-Hi-Tech-Plan projects in China. He is PI of over 20 projects, including projects funded by RGC, NSFC, ITC, and SZSTI. He has published over 90 papers in major international journals and conferences. His research interests include dynamic environmental modeling, 3-D mapping, machine learning, and visual control.

Mr. Liu won the second place of EMAV'09 (European Micro Aerial Vehicle Competition). He received two awards from IARC 14 (International Aerial Robotics Contest). He won the Best Student Paper Award as the first author for MFI 2012 (IEEE International Conference on Multisensor Fusion and Information Integration), the Best Paper Award in Information for ICIA 2013 (IEEE International Conference on Information and Automation) as first author and Best Paper Award Finalists as co-author, the Best RoboCup Baper Award for IROS 2013 (IEEE/RSJ International Conference on Intelligent Robots and Systems), the Best Student Paper Award of IEEE ICAR 2017, and the Best Paper Award in Automation for ICIA 2017. He twice won the innovation contest Chunhui Cup Winning Award in 2012 and 2013. He won the Wu Wenjun AI Innovation Award in 2016. He was the General Chair of International Conference on Computer Vision Systems in 2017, and the Program Chair of IEEE International Conference on Real-Time Computing and Robotics (IEEE-RCAR) in 2016 and the International Robotic Alliance Conference in 2017.

Ioannis Pitas (Fellow, IEEE) received the Diploma and Ph.D. degrees in electrical engineering from the Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece.

Since 1994, he has been a Professor with the Department of Informatics, AUTH, where he is the Director of the Artificial Intelligence and Information Analysis Lab. He served as a Visiting Professor at several universities. He has published over 1000 papers, contributed in 47 books in his areas of interest and edited or (co-)authored another 11 books. He has also been member of the program committee of many scientific conferences and workshops. He participated in 71 R&D projects, primarily funded by the European Union and is/was Principal Investigator in 42 such projects. He leads the big European H2020 R&D Project MULTIDRONE. He is an AUTH Principal Investigator in H2020 R&D Projects Aerial Core and AI4Media. He is the Chair of the Autonomous Systems Initiative. He is the Head of the EC funded AI doctoral school of Horizon2020 EU-funded R&D project AI4Media (1 of the 4 in Europe). He has over 32 000 citations to his work and an h-index of more than 85 (Google Scholar). His current interests are in the areas of computer vision, machine learning, autonomous systems, and image/video processing.

Prof. Pitas served as an Associate Editor or a Co-Editor of nine international journals and is the General or the Technical Chair of four international conferences. He is an IEEE Distinguished Lecturer and a Fellow of EURASIP.