# Long-Awaited Next-Generation Road Damage Detection and Localization System is Finally Here

Rui Fan, Jintao Cheng, Yang Yu, Juncan Deng, George Giakos, *Fellow, IEEE*, Ioannis Pitas, *Fellow, IEEE*

***Abstract*— With the ever-increasing emphasis on road maintenance to a high standard, the need for automated and robust road damage inspection (detection and localization) systems is becoming greater than ever. In this paper, we introduce a real-time road damage inspection system, which has been embedded in a drone to reconstruct 3D road geometry using stereo vision, as well as to detect and localize road damage using disparity map segmentation and visual simultaneous localization and mapping. In addition, the 3D road map is built and updated, which enables road damage inspectors to assess road conditions in a more convenient way.**

## I. Introduction

A new poll of cyclists has found road damage is not just an inconvenience to road users, it is also a safety hazard [1]. It was reported in 2015 that an Olympic gold medalist incurred eight fractured ribs resulting in a punctured lung after hitting a pothole during a race [2]. Therefore, detecting, localizing and repairing road damage, *e.g.*, potholes and cracks, is crucial for road maintenance, traffic safety, and driving comfort [3].

Manual visual inspection is still the main form of road condition assessment [4]. This process is, nevertheless, exhausting, costly, time-consuming, and dangerous for the inspectors [5]. For example, the UK Asphalt Industry Alliance suggested a one-off investment of £12 billion in 2014 to improve the road conditions across England and Wales [6]. But after only three years, the UK Department for Transport had to announce another £17.5 billion funding scheme for road damage detection and repair across England [7]. Unfortunately, this investment is still far from enough [2]. Additionally, road damage inspection results are always qualitative and subjective, as decisions depend entirely on the inspectors' experience and judgment [8]. Therefore, there is an ever-increasing need to develop automated and embedded road inspection (detection and localization) systems, which can provide quantitative, objective and accurate road damage data in an efficient way [9].


Fig. 1. An illustration of embedded road damage inspection (detection and localization) system.

Over the past decade, various technologies, *e.g.*, the global positioning system (GPS), vibration sensing, active sensing and passive sensing, have been prevalently utilized to acquire road data and assist technicians in inspecting road damages [10]. For example, [11] developed a crowd-sourcing system to detect and localize potholes by analyzing the accelerometer data obtained from multiple vehicles, while [12] mounted two laser scanners on a Georgia Institute of Technology Sensing Vehicle (GTSV) to collect 3D road data for road damage detection. Furthermore, our previous works [10] and [13] employed an efficient dense stereo matching algorithm to reconstruct highly precise 3D road geometry models (accuracy: $\pm 3$ mm), and a disparity map transformation and segmentation algorithm to detect road damage.

With the recent advance of airborne technology, drones equipped with digital cameras provide new opportunities for intelligent road damage inspection [14]. For example, [15] designed a photogrammetric mapping system for drones, which can recognize different types of road damage, *e.g.*, potholes and cracks, from RGB images. Although such a system can detect road damage with very low computational complexity, the depth information is required to measure the volumes of the detected road damage [16]. To obtain such depth information, our recent work [8] embedded a real-time GPU-friendly dense stereo matching system on a drone to acquire dense 3D road damage geometry models. Additionally, compared with monocular road damage inspection systems, binocular systems can recover accurate absolute scales of the reconstructed 3D road geometry models, without considering complicated environmental hypotheses [17], [18]. Therefore, employing binocular road damage detection and localization systems is more feasible for drones [8].

This paper presents an intelligent binocular road damage detection and localization system embedded on a drone, as

Rui Fan is with the Department of Control Science and Engineering, Tongji University, Shanghai 201804, China. email: rui.fan@ieee.org

Jintao Cheng and Juncan Deng are with the School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou 510006, China. email: chengjintao@unity-drive.come, dengjuncan@m.scnu.edu.cn

Yang Yu is with the Department of Electronic and Computer Engineering, the Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR, China. email: yyubj@connect.ust.hk

George Giakos is with the Department of Electrical and Computer Engineering, Manhattan College, New York, NY 10471, USA. email: giakos@manhattan.edu

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

This paper has received partial funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 871479 (AERIALCORE).

illustrated in Fig. 1. This system can 1) acquire dense road disparity/depth information, 2) reconstruct dense 3D road surface geometry models, 3) detect road damage, 4) build and update the environmental map, and 5) localize detected road damage on the environmental map.

The remainder of this paper is organized as follows: Section II reviews the state-of-the-art 3D road surface geometry reconstruction algorithms, road damage detection algorithms, and visual simultaneous localization and mapping (VSLAM) algorithms. Section III introduces our road damage inspection system. The experimental results are given in Section IV and the performance of the proposed system is also evaluated and discussed. Finally, Section V concludes the paper.

## II. Literature Review

### A. Visual 3D Road Surface Geometry Reconstruction

To reconstruct the 3D road surface geometry models, road images captured from different views are required [19]. This can be realized by using either a single movable camera or a group of synchronized cameras [20]. Structure from motion [21], [22] and optical flow [18], [23] are two typical types of algorithms that can be employed for 3D reconstruction when using a single moving camera. However, it is difficult for them to recover the absolute scales of the reconstructed 3D models if no environmental hypotheses are considered [21]. On the other hand, stereo vision is capable of recovering 3D model's absolute scale, as the two cameras are fixed.

Many stereo matching algorithms [10], [16], [24], [25], [26] have been proposed for 3D road surface geometry reconstruction. Among them, [10] proposed PT-SRP, an efficient and highly accurate iterative local stereo matching algorithm, designed specifically for road surface 3D reconstruction. It first transforms the perspective view of the target image to the reference view, which not only boosts the stereo matching speed, but also improves the disparity accuracy. An efficient stereo matching algorithm is then employed to estimate road disparities in an iterative manner. However, PT-SRP [10] is hard to implement in parallel on GPUs. Therefore, [8] proposed PT-FBS, a GPU-friendly road disparity estimation algorithm, embedded in a drone for real-time road surface 3D reconstruction. Nevertheless, PT-FBS [8] is still computationally intensive. Hence in [16], a generalized perspective transformation (GPT) algorithm was proposed and semi-global matching (SGM) was used for road disparity estimation. The experimental results suggest that GPT-SGM [16] outperforms both PT-SRP [10] and PT-FBS [8] in terms of both efficiency and accuracy. Therefore, GPT-SGM [16] is used in this paper to reconstruct 3D road geometry models for road damage inspection.

### B. Visual Road Damage Detection

The state-of-the-art visual road damage detection methods are developed based on either 2D image understanding/analysis [27] or 3D road surface modeling [5]. The former approaches typically utilize traditional image processing algorithms [4], [9], [12], [13], [28], [29] or modern deep convolutional neural networks (DCNNs) [2], [30], [31], [32], [33] to detect road damage, by performing either pixel-level image segmentation or instance-level object recognition on RGB or depth/disparity images. The image processing-based approaches typically consist of four main steps: 1) image pre-processing, 2) image segmentation, 3) shape extraction, and 4) object recognition, while the DCNN-based approaches address road damage detection in an end-to-end way, using semantic segmentation or object detection networks. On the other hand, 3D road surface modeling-based algorithms [5], [34], [35] typically formulate the 3D road point cloud as a quadratic surface, whose coefficients can be obtained by performing robust surface modeling.

In our recent work [5], a novel disparity image processing algorithm, referred to as disparity transformation, was proposed to better distinguish between the damaged and undamaged road regions. A robust 3D road surface modeling algorithm was then used to detect road damage from the transformed road disparity images. Later on, in [13], we introduced a more efficient visual road damage detection algorithm based on unsupervised disparity map segmentation, and it was proven that the energy minimization problem introduced in [29] has a numerical solution. Furthermore, in [16], we detected road damage by segmenting the transformed disparity map using simple linear iterative clustering [36], while in [2], we proposed a attention aggregation (AA) framework, which can take advantage of different types of attention models to enhance the performance of both single-modal and data-fusion DCNNs for semantic transformed disparity map segmentation.

### C. Visual Simultaneous Localization and Mapping

The state-of-the-art VSLAM algorithms are generally classified as direct [37], indirect [38], and hybrid [39]. These algorithms provide fundamental building blocks and can be easily employed to localize the detected road damages.

Monocular VSLAM algorithms often cannot recover reliable absolute scale information [40], and thus, many researchers have resorted to additional sensors, such as inertial measurement units (IMUs) or digital cameras, to improve the localization accuracy [41], [42]. The former are generally known as visual-inertial odometry (VIO) systems, which fuse the estimated visual states with the data acquired using IMUs to enhance the robustness of the VSLAM systems [43], while the latter typically utilize the depth information acquired using stereo rigs to improve the absolute scale recovery [41].

Recent VIO research uses either filtering or optimization techniques to improve the localization accuracy. For instance, [42] considered the IMU measurements and proposed a loosely-coupled filter to recover accurate absolute scales. Similarly, in [44], the 6-DoF poses estimated by a parallel tracking and mapping (PTAM) framework were fused with the IMU measurements to improve odometry system robustness. Additionally, [45] introduced a loosely-coupled fusion method to initialize the scale and bias parameters. [43] proposed a tightly-coupled filter to extract multi-level patch features along with the 3D landmarks during camera tracking. The camera poses were then estimated using a stan-

dard extended Kalman Filter [46]. Moreover, [47] presented a sliding window-based optimization framework, which is also capable of selecting key-frames. In their system, a cost function with respect to both visual re-projection error and inertial error was formulated for camera pose optimization.

In addition to VIO systems, a binocular SLAM framework with high computational efficiency and robustness was proposed in [41] for absolute scale recovery. In recent years, ORB-SLAM2 [48] and ORB-SLAM3 [49], two well-known VSLAM frameworks, have achieved compelling visual odometry results [18]. Due to their superior performance for mapping, loop-closure detection and relocalization, ORB-SLAM2 [48] is used in this paper for road damage localization. In addition, we also compare its performance with ORB-SLAM3 [49].

## III. Intelligent Road Damage Inspection System

As discussed in [16], PT-SRP [10] cannot fully exploit the parallel computing architecture, and PT-FBS [8] is very computationally intensive. Therefore, GPT-SGM [16] is utilized in this paper for dense road disparity information acquisition. It first transforms the target road image (stereo right image) $I_t$ into the reference view using

$$I'_t(u,v) = I_t(u + \kappa(u,v,\phi,\boldsymbol{a}), v), \tag{1}$$

where

$$\kappa(u,v,\phi,\boldsymbol{a}) = \min_{x=0}^{W} \Big[a_0 + a_1(v\cos\phi - x\sin\phi) - \delta\Big], \tag{2}$$

$I'_t$ is the transformed target image, $\boldsymbol{p} = (u;v)$ is an image pixel, $u$ and $v$ are the horizontal and vertical coordinates of $\boldsymbol{p}$, respectively, $\boldsymbol{a} = (a_0;a_1)$ stores the road disparity projection parameters, $\phi$ is the stereo rig roll angle [50], $W$ is the image width, and $\delta$ is a constant set to ensure that the disparities between the reference and transformed target images, *i.e.*, $I_r$ and $I'_t$, are non-negative [16]. $\boldsymbol{a}$ and $\phi$ can be estimated by minimizing [51]:

$$E = \left|\left|\boldsymbol{d} - \Big[\mathbf{1}_k \quad \boldsymbol{v}\cos\phi - \boldsymbol{u}\sin\phi\Big]\boldsymbol{a}\right|\right|_2^2, \tag{3}$$

where $\boldsymbol{d} = (d_1;\cdots;d_k)$ stores a group of $k$ disparities (note here that in this stage the disparities are estimated using keypoint detection and matching algorithms), $\boldsymbol{u} = (u_1;\cdots;u_k)$ and $\boldsymbol{v} = (v_1;\cdots;v_k)$ store the horizontal and vertical coordinates of the disparities, respectively, and $\mathbf{1}_k$ is a $k$-entry vector of ones. A dense road disparity image $\ell$ can be estimated by solving

$$c^{\boldsymbol{r}}_{ag}(\boldsymbol{p},d_{\boldsymbol{p}}) = c(\boldsymbol{p},d_{\boldsymbol{p}}) + \min\Bigg(c^{\boldsymbol{r}}_{ag}(\boldsymbol{p}-\boldsymbol{r},d_{\boldsymbol{p}}), \bigcup_{k\in\{-1,1\}} c^{\boldsymbol{r}}_{ag}(\boldsymbol{p}-\boldsymbol{r},d_{\boldsymbol{p}}+k)+\lambda_1, \min_i c^{\boldsymbol{r}}_{ag}(\boldsymbol{p}-\boldsymbol{r},i)+\lambda_2\Bigg), \tag{4}$$

where $c$ denotes the stereo matching cost, $d_{\boldsymbol{p}}$ denotes the disparity of $\boldsymbol{p}$, $\lambda_1$ and $\lambda_2$ are two penalty parameters, and $c^{\boldsymbol{r}}_{ag}(\boldsymbol{p},d_{\boldsymbol{p}})$ represents the aggregated stereo matching cost at $\boldsymbol{p}$ in the direction of $\boldsymbol{r}$. A transformed road disparity map $\ell'$ can thus be obtained [52]:

$$\ell'(\boldsymbol{p}) = \ell(\boldsymbol{p}) + \kappa(u,v,\phi,\boldsymbol{a}) - a_0 - a_1(v\cos\phi + u\sin\phi) + \delta. \tag{5}$$

The road damage can subsequently be detected by segmenting the transformed disparity images using image segmentation techniques, such as semantic segmentation DCNNs.

Then, we employ ORB-SLAM2 [49] to localize the detected road damage, as well as build and update the environmental map. ORB-SLAM2 [49] consists of two components: a) tracking (front end) and b) mapping (back end). The input of this system is a reference road image and its corresponding depth information. The front-end tracker first extracts features from accelerated segment test (FAST) [53] feature points and computes their corresponding oriented fast and rotated BRIEF (ORB) descriptors [54]. The relative pose $\boldsymbol{T}_{k,k-1}$ between frame $k$ and frame $k-1$, *i.e.*, $\mathcal{F}_k$ and $\mathcal{F}_{k-1}$, can be estimated from a set of matched correspondence feature point pairs $\boldsymbol{P}_k = (\boldsymbol{p}_{0,k};\boldsymbol{p}_{1,k};\ldots;\boldsymbol{p}_{n,k})$ and $\boldsymbol{P}_{k-1} = (\boldsymbol{p}_{0,k-1};\boldsymbol{p}_{1,k-1};\ldots;\boldsymbol{p}_{n,k-1})$, where $\boldsymbol{p}_{i,k} = (u_{i,k};v_{i,k})$ represents the $i$-th matched feature point in the $k$-th frame. The expression of $\boldsymbol{T}_{k,k-1}$ is as follows:

$$\boldsymbol{T}_{k,k-1} = \big[\boldsymbol{R}_{k,k-1}|\boldsymbol{t}_{k,k-1}\big], \tag{6}$$

where $\boldsymbol{R}_{k,k-1}$ is a rotation matrix and $\boldsymbol{t}_{k,k-1}$ is a translation vector. A point $\boldsymbol{p}^C_{k-1} = [x_{k-1},y_{k-1},z_{k-1}]^\top$ in the $(k-1)$-th camera coordinate system (CCS) can be transformed to $\boldsymbol{p}^C_k = [x_k,y_k,z_k]^\top$ in the $k$-th CCS using

$$\boldsymbol{p}^C_k = \boldsymbol{R}_{k,k-1}\boldsymbol{p}^C_{k-1} + \boldsymbol{t}_{k,k-1}. \tag{7}$$

As the VSLAM framework used here employs a constant-velocity motion model, $\boldsymbol{T}_{k,k-1}$ is first initialized as $\boldsymbol{T}_{k-1,k-2}$ [48]. Then, $\boldsymbol{T}_{k,k-1}$ is optimized by minimizing the following re-projection error:

$$e_{i,k}(\boldsymbol{p}_{i,k}) = \left|\left|\pi\left((\boldsymbol{\xi}\oplus\boldsymbol{T}_{k,k-1})\,\boldsymbol{T}^{-1}_{k,k-1}\boldsymbol{p}^C_{i,k-1}\right) - \boldsymbol{p}_{i,k}\right|\right|_\gamma, \tag{8}$$

where $||\cdot||_\gamma$ denotes the robust Huber norm and $\boldsymbol{\xi}$ represents the increment twist [48]. $\pi(\cdot)$ is the projection function. The optimum increment twist $\boldsymbol{\xi}$ can be estimated by minimizing

$$\boldsymbol{\xi} = \arg\min_{\boldsymbol{\xi}} \sum_{\boldsymbol{p}_{i,k}\in\boldsymbol{P}_k} w_{i,k}e_{i,k}(\boldsymbol{p}_{i,k}), \tag{9}$$

where $w_{i,k}$ represents the optimization weighting for the $i$-th matched feature point in the $k$-th frame. The increment twist is then updated using

$$\boldsymbol{\xi} \leftarrow \log\left(\delta\boldsymbol{\xi}\oplus\mathbf{T}(\boldsymbol{\xi})\right). \tag{10}$$

Following the coarse-to-fine strategy, a motion-only bundle adjustment (BA) is performed to optimize the $k$-th local map $\{\mathcal{F}_k,\boldsymbol{Q}_k\}$, where $\boldsymbol{Q} = [\boldsymbol{p}^C_{0,k},\boldsymbol{p}^C_{1,k},\ldots,\boldsymbol{p}^C_{m,k}]^\top$ stores a set of 3D feature points in the $k$-th CCS. Furthermore, the local map $\{\mathcal{F}_k,\boldsymbol{Q}_k\}$ is also updated in the mapping

TABLE I

ATE COMPARISON BETWEEN ORB-SLAM2 [48] AND ORB-SLAM3 [49].

| Algorithm | Dataset ID | Mean [m] | Median [m] | RMSE [m] | STD [m] |
|---|---|---|---|---|---|
| ORB-SLAM2 [48] | 1 | 2.51 | 1.88 | 2.94 | 1.53 |
| ORB-SLAM3 [49] | 1 | 2.36 | 1.92 | 2.77 | 1.43 |
| ORB-SLAM2 [48] | 2 | 4.51 | 4.60 | 4.71 | 1.36 |
| ORB-SLAM3 [49] | 2 | 4.79 | 5.09 | 5.02 | 1.51 |

TABLE II

RDE COMPARISON BETWEEN ORB-SLAM2 [48] AND ORB-SLAM3 [49].

| Algorithm | Dataset ID | Mean [m] | Median [m] | RMSE [m] | STD [m] |
|---|---|---|---|---|---|
| ORB-SLAM2 [48] | 1 | 0.07 | 0.06 | 0.09 | 0.06 |
| ORB-SLAM3 [49] | 1 | 0.08 | 0.07 | 0.14 | 0.11 |
| ORB-SLAM2 [48] | 2 | 0.17 | 0.11 | 0.24 | 0.19 |
| ORB-SLAM3 [49] | 2 | 0.07 | 0.05 | 0.10 | 0.07 |

process using a full BA. Similarly, an optimum global joint set $\mathcal{X} = \{\boldsymbol{Q}, \boldsymbol{T}\}$ can be estimated using:

$$\mathcal{X} = \arg\min_{\mathcal{X}} \sum_{\boldsymbol{p}_{i,j-1} \in \boldsymbol{P}_j} \sum_{\mathcal{F}_j \in \boldsymbol{F}_i} w_{i,j} e_{i,j}, \tag{11}$$

where $\boldsymbol{F}_i$ stores a set of keyframes where $\boldsymbol{p}_i^C$ can be observed. Finally, we eliminate outliers using $\mathcal{X}^2$-test [48], where each feature point is constrained with one-pixel variance. The detected road damages can therefore be localized, and the 3D map is built and updated, as shown in Fig. 2.

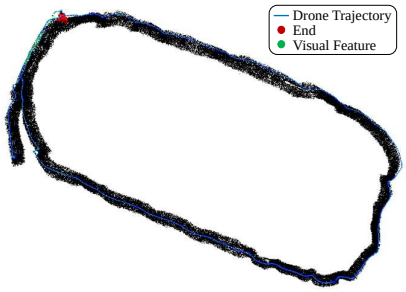


Fig. 2. Road damage localization and environmental map creation result.

## IV. EVALUATION AND DISCUSSION

In this section, we first use the datasets in [55] to compare ORB-SLAM2 [48] and ORB-SLAM3 [49] for road damage localization. Then, we discuss the existing problems in the proposed road damage detection and localization system.

We use two matrices: a) absolute trajectory error (ATE) and b) relative distance error (RDE) to quantify the accuracy of road damage localization, as shown in Tab. I and II, respectively. It can be observed that ORB-SLAM2 [48] and ORB-SLAM3 [49] perform similarly on these two datasets, where the scenarios contain only road images (no obstacles on road). The mean ATEs achieved are $2.36 - 2.51$ m and $4.51 - 4.79$ m, respectively, while the mean RDEs achieved are $0.07 - 0.08$ m and $0.07 - 0.17$ m, respectively. Therefore, ORB-SLAM2 [48] and ORB-SLAM3 [49] are shown to be two effective techniques to be used for road damage localization.

However, the introduced road damage detection and localization system has several issues. First of all, with GPT-SGM [16], it is difficult to obtain accurate road disparity information when the drone is flying too high, as it is difficult to extract reliable visual features for perspective transformation. Therefore, it is more feasible to use (unsupervised) DCNNs trained for end-to-end road disparity estimation. Moreover, as discussed in [2], [5], [13], a transformed disparity map is a more informative type of visual feature for road damage detection, but its realization requires empty roads. When the stereo rig captures obstacles, an obstacle removal algorithm needs to be conducted before performing roll angle and road disparity projection model estimation, as an obstacle on the road can lead to a significant error.

## V. CONCLUSION

This paper first reviewed the state-of-the-art visual road surface 3D geometry reconstruction algorithms, visual road damage detection algorithms, and VSLAM algorithms. Then, we introduced an intelligent binocular road damage detection and localization system based on dense road disparity estimation, road disparity transformation and segmentation, and VSLAM. The proposed system has been successfully implemented on a drone for real-time road damage inspection. We evaluated the performance of road damage localization using our previously published road visual odometry datasets. The experimental results demonstrated that road damage localization can achieve an ATE of $2.36 - 4.79$ m and an RTE of $0.07 - 0.17$ m. Furthermore, we discussed the existing challenges of the proposed road damage detection and localization system. With these challenges overcome, such drone-based road damage inspection systems will be feasible for production.

## REFERENCES


[1] A. Ballinger, "New research says potholes are not just an inconvenience, theyre a safety risk," *BristolLive*, Mar 2018.

[2] R. Fan, H. Wang, M. J. Bocus, and M. Liu, "We learn better road pothole detection: from attention aggregation to adversarial domain adaptation," in *European Conference on Computer Vision*. Springer, 2020, pp. 285–300.

[3] S. Mathavan, K. Kamal, and M. Rahman, "A review of three-dimensional imaging technologies for pavement distress detection and measurements," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2353–2362, 2015.

[4] C. Koch, G. M. Jog, and I. Brilakis, "Automated pothole distress assessment using asphalt pavement video data," *Journal of Computing in Civil Engineering*, vol. 27, no. 4, pp. 370–378, 2013.

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

[6] B. News, "Councils in england face huge road repair bills," *BBC*, Jan 2015. [Online]. Available: https://www.bbc.com/news/uk-england-30684854

[7] Department for Transport, "Annual report and accounts 2016/17," 2017.

[8] 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 *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2019, pp. 535–543.

[9] 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," *Advanced Engineering Informatics*, vol. 29, no. 2, pp. 196–210, 2015.

[10] R. Fan, X. Ai, and N. Dahnoun, "Road surface 3d reconstruction based on dense subpixel disparity map estimation," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3025–3035, 2018.

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

[12] Y.-C. Tsai and A. Chatterjee, "Pothole detection and classification using 3d technology and watershed method," *Journal of Computing in Civil Engineering*, vol. 32, no. 2, p. 04017078, 2018.

[13] R. Fan and M. Liu, "Road damage detection based on unsupervised disparity map segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, pp. 4906 – 4911, Nov. 2019.

[14] E. Schnebele, B. Tanyu, G. Cervone, and N. Waters, "Review of remote sensing methodologies for pavement management and assessment," *European Transport Research Review*, vol. 7, no. 2, p. 7, 2015.

[15] C. Zhang *et al.*, "An uav-based photogrammetric mapping system for road condition assessment," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*, vol. 37, pp. 627–632, 2008.

[16] R. Fan, U. Ozgunalp, Y. Wang, M. Liu, and I. Pitas, "Rethinking road surface 3d reconstruction and pothole detection: From perspective transformation to disparity map segmentation," *IEEE Transactions on Cybernetics*, 2021.

[17] Y. Dang, P. Chen, R. Liang, C. Huang, Y. Tang, T. Yu, X. Yang, and K.-T. Cheng, "Real-time semantic plane reconstruction on a monocular drone using sparse fusion," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7383–7391, 2019.

[18] H. Wang *et al.*, "Atg-pvd: Ticketing parking violations on a drone," in *European Conference on Computer Vision*. Springer, 2020, pp. 541–557.

[19] R. Fan, L. Wang, M. J. Bocus, and I. Pitas, "Computer stereo vision for autonomous driving," 2020.

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

[21] R. Roberts, L. Inzerillo, and G. Di Mino, "Developing a framework for using structure-from-motion techniques for road distress applications," *Eur. Transp. Transp. Eur*, vol. 77, pp. 1–11, 2020.

[22] L. Inzerillo, G. Di Mino, and R. Roberts, "Image-based 3d reconstruction using traditional and uav datasets for analysis of road pavement distress," *Automation in Construction*, vol. 96, pp. 457–469, 2018.

[23] H. Wang, R. Fan, and M. Liu, "Cot-amflow: Adaptive modulation network with co-teaching strategy for unsupervised optical flow estimation," *Conference on Robot Learning (CoRL)*, 2020.

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

[25] R. Fan, Y. Liu, X. Yang, M. J. Bocus, N. Dahnoun, and S. Tancock, "Real-time stereo vision for road surface 3-d reconstruction," in *2018 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2018, pp. 1–6.

[26] S. Sengupta, E. Greveson, A. Shahrokni, and P. H. Torr, "Urban 3d semantic modelling using stereo vision," in *2013 IEEE International Conference on robotics and Automation*. IEEE, 2013, pp. 580–585.

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

[28] S. Li, C. Yuan, D. Liu, and H. Cai, "Integrated processing of image and gpr data for automated pothole detection," *Journal of computing in civil engineering*, vol. 30, no. 6, p. 04016015, 2016.

[29] R. Fan, M. J. Bocus, and N. Dahnoun, "A novel disparity transformation algorithm for road segmentation," *Information Processing Letters*, vol. 140, pp. 18–24, 2018.

[30] R. Fan *et al.*, "Road crack detection using deep convolutional neural network and adaptive thresholding," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 474–479.

[31] H. Chen, M. Yao, and Q. Gu, "Pothole detection using location-aware convolutional neural networks," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 4, pp. 899–911, 2020.

[32] J. Fan *et al.*, "Multi-scale feature fusion evolves semantic segmentation for road pothole detection," in *The 2021 IEEE International Conference on Autonomous Systems (ICAS)*. IEEE, 2021.

[33] W. Ye, W. Jiang, Z. Tong, D. Yuan, and J. Xiao, "Convolutional neural network for pothole detection in asphalt pavement," *Road materials and pavement design*, pp. 1–17, 2019.

[34] Z. Zhang, "Advanced stereo vision disparity calculation and obstacle analysis for intelligent vehicles," Ph.D. dissertation, University of Bristol, 2013.

[35] U. Ozgunalp, "Vision based lane detection for intelligent vehicles," Ph.D. dissertation, University of Bristol, 2016.

[36] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[37] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 611–625, 2017.

[38] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[39] C. Forster, M. Pizzoli, and D. Scaramuzza, "Svo: Fast semi-direct monocular visual odometry," in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 15–22.

[40] M. U. M. Bhutta *et al.*, "Loop-box: Multiagent direct slam triggered by single loop closure for large-scale mapping," *IEEE transactions on cybernetics*, 2020.

[41] I. Cvišic, J. Cesic, I. Markovic, and I. Petrovic, "Soft-slam: Computationally efficient stereo visual slam for autonomous uavs," *Journal of field robotics*, 2017.

[42] S. Weiss, M. W. Achtelik, S. Lynen, M. C. Achtelik, L. Kneip, M. Chli, and R. Siegwart, "Monocular vision for long-term micro aerial vehicle state estimation: A compendium," *Journal of Field Robotics*, vol. 30, no. 5, pp. 803–831, 2013.

[43] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct ekf-based approach," in *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2015, pp. 298–304.

[44] G. Klein and D. Murray, "Parallel tracking and mapping for small ar workspaces," in *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE Computer Society, 2007, pp. 1–10.

[45] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[46] G. Welch *et al.*, "An introduction to the kalman filter," 1995.

[47] S. Leutenegger, P. Furgale, V. Rabaud, M. Chli, K. Konolige, and R. Siegwart, "Keyframe-based visual-inertial slam using nonlinear optimization," *Proceedings of Robotis Science and Systems (RSS) 2013*, 2013.

[48] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[49] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "Orb-slam3: An accurate open-source library for visual, visual-inertial and multi-map slam," *arXiv preprint arXiv:2007.11898*, 2020.

[50] N. Vinogradov and et al., "Implementation of stereo rig roll angle estimation on a tms320c6678 dsp," in *2020 9th Mediterranean Conference on Embedded Computing (MECO)*. IEEE, 2020, pp. 1–4.

[51] R. Fan, H. Wang, P. Cai, J. Wu, J. Bocus, L. Qiao, and M. Liu, "Learning collision-free space detection from stereo images: Homography matrix brings better data augmentation," *IEEE/ASME Transactions on Mechatronics*, 2021.

[52] H. Wang, R. Fan, Y. Sun, and M. Liu, "Dynamic fusion module evolves drivable area and road anomaly detection: A benchmark and algorithms," *IEEE transactions on cybernetics*, 2021.

[53] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, "Adaptive and generic corner detection based on the accelerated segment test," in *European conference on Computer vision*. Springer, 2010, pp. 183–196.

[54] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Proc. Int. Conf. Computer Vision*, Nov. 2011, pp. 2564–2571.

[55] H. Huang *et al.*, "A robust pavement mapping system based on normal-constrained stereo visual odometry," in *2019 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2019, pp. 1–6.