EMPLOYING SSLG WITH V- DISPARITY MAPS TO SEGREGATE ROAD ANOMALIES AND DESIGN ROBOTIC WHEELCHAIRS FOR DRIVABLE AREAS

[1]Bhaskara Palakurthi, [2]Srinivasa Reddy Dumpa, [3]Sethu Kumar Thummalapenta, [4]Bolleddu Haritha
[1,2,3]Assistant Professor, [4]UG Student, [1,2,3,4]Department of Electronics and Communication Engineering, Brilliant Grammar School Educational Society Group of Institutions Integrated Campus, Hyderabad, India

ABSTRACT
ROBOTIC wheelchairs are made to increase the mobility of the elderly or crippled, hence enhancing their quality of life. In order to do this, autonomous navigation has undergone extensive research and evolved into a crucial feature for robotic wheelchairs. Identifying regions and anomalies in photos pixel-by-pixel is referred to as segmenting drivable areas and road abnormalities. It plays a key role in autonomous navigation. Robotic wheelchairs might bump or possibly roll over while travelling over road anomalies without properly separating drivable sections and road oddities, which could harm human riders. In this research, we propose a self-supervised learning strategy to describe the solution to this problem for the purpose of segmenting drivable areas and road anomalies. The proposed method can automatically generate segmentation labels for drivable areas and road anomalies. Then, we train RGB-D data based semantic segmentation neural networks and get predicted labels. We firstly develop a pipeline named Self-Supervised Label Generator (SSLG) to automatically label drivable areas and road anomalies. Then, we use the segmentation labels generated by the SSLG to train several RGB-D data-based semantic segmentation neural networks.
Keywords: SSLG, RGB-D

INTRODUCTION

Robotic wheelchairs are made to increase the mobility of the elderly or crippled, therefore improving their quality of life. In order to do this, autonomous navigation has undergone extensive research and evolved into a crucial feature for robotic wheelchairs. Identifying regions and anomalies in photos pixel-by-pixel is referred to as segmenting drivable areas and road abnormalities. It plays a key role in autonomous navigation. Robotic wheelchairs might bump or possibly roll over while travelling over road anomalies without properly separating drivable sections and road oddities, which could harm human riders. In this study, we define the drivable area as the space in which robotic wheelchairs of any size may go, whereas the road anomaly is the space with a height more than 5cm from the surface of the drivable area. The segmentation of drivable areas and road anomalies could be addressed using semantic segmentation techniques. RGB-D cameras, such as Kinect [1], are visual sensors that can stream RGB and depth images at the same time [2]–[4]. We use an RGB-D camera for the segmentation of drivable areas and road anomalies in this paper. The reason why we use RGB-D camera is that the depth difference between road anomalies and drivable areas could be useful to distinguish them. Recent development of deep learning techniques has brought significant improvements on the semantic segmentation using RGB-D cameras [5]–[7]. In order to train a deep neural network, we usually needa large-scale dataset with hand-labeled ground truth. However, generating such adataset is time-consuming and labor- intensive. In order to provide a solution for the excessive consumption of time and labor for manual labeling, we present a self-supervised approach to segment drivable areas and road anomalies for robotic wheelchairs with an Intel

Realsense D415 RGB-D camera. Fig. 1 shows our robotic wheelchair, RGB-D camera and mini PC. The Intel Realsense D415 is an active stereo camera that estimates distances with emittedinfrared lights. It is equipped with an Intel Realsense D415 RGB-D camera to collect data and a Mini PC to process data.

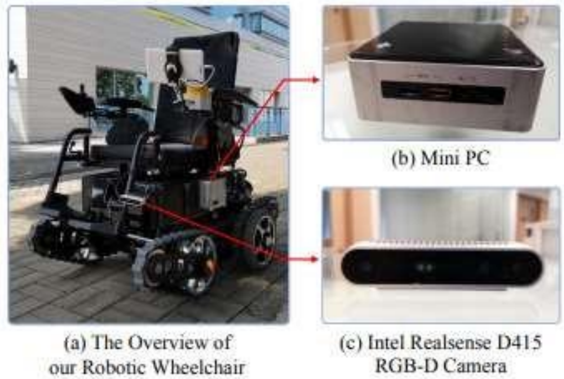

Fig. 1: The robotic wheelchair used inthis work.

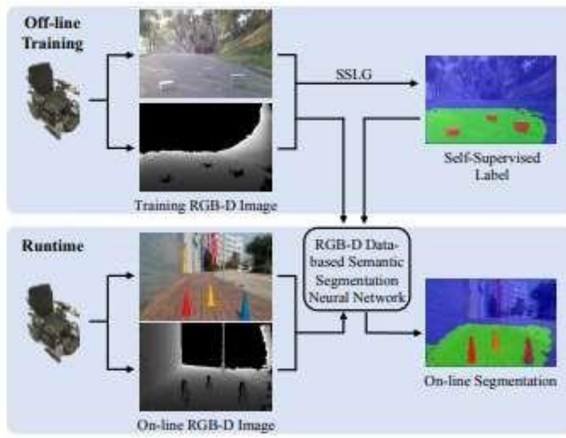

Fig. 2 illustrates the overview of ourproposed self supervised approach.

We firstly develop a pipeline named Self- Supervised Label Generator (SSLG) to automatically label drivable areas and road anomalies. Then, we use the segmentation labels generated by the SSLG to train several RGB-D data-based semantic segmentation neural networks. Experimentalresults show that although the segmentation

Fig. 2 illustrates the overview of our proposed self supervised approach [10]. We firstly develop a pipeline named Self-Supervised Label Generator (SSLG) to automatically label drivable areas and road anomalies. Then, we use the segmentation labels generated by the SSLG to train several RGB-D data- based semantic segmentation neural networks. Experimental results showthat although the segmentation labels generated by the SSLG present mis- labelings, the results of our proposed self-supervised approach still outperforms the state-of-the-art traditional algorithms and the state-of- the-art self-supervised algorithms. The contributions of this paper are summarized as follows:

We propose a self-supervised approach to segment drivable areas and road anomalies for robotic wheelchairs.

We develop a pipeline to automatically label drivable areas and road anomalies using RGB-D images. We construct an RGB-D dataset1 ,which covers 30 common scenes where robotic wheelchairs usually work

RELATED WORK

Drivable Area Segmentation Labayradeet al.[8] proposed an algorithm that converted the drivable are segmentation problem into astraight line detection problem. Following [8], some improvements [9]–[10]were developed to enhance the robustness and accuracy. In recent years,some researchers have tried to solve this segmentation problem from other perspectives. For example, Ozgunalp et al. [12] proposed to use the estimatedplanar patches and patch orientations to reduce the impact of outliers. The proposed an approach to fuse the information of light detection andranging (LIDAR) and vision data, respectively.

Road Anomaly Segmentation Early approaches of road anomaly segmentation mainly adopted traditional computer vision algorithms with handcrafted features. Cong et al. designed a feature based on the depth confidence analysis and multiple cues fusion. Lou et al. proposed an approach combining both regional stability and saliency for small road anomaly segmentation. With the development of deep learning, many work used deep neural networks to segment road anomalies. Peng et al. used a deep neural network to extract features from RGB-D images and fused these features in three levels to generate the predicted result. Chen et al. proposed a multiscale multi-path fusion network with cross-modal interactions, which could explore deep connections between RGB images and depth images. There also exist some RGB-D data- based semantic segmentation neural networks that fuse RGB and depth data together such as FuseNet [5] and Depth- aware CNN [6], which can achieve impressive performance. However, these methods rely on hand-labeled ground truth, which would become difficult to implement when there is no sufficienttraining dataset. Our proposed SSLG can automatically label drivable areas and road anomalies with RGB-D images, which greatly reduces the time and labor for manual labeling.

Automatic Labeling

Automatic labeling aims to generatelabeling data automatically or in an unsupervised way without hand-labeled training data. Barnes et al. proposedan approach to generate training data for the segmentation problem of path proposals by exploiting the information of odometry and obstacle sensing. Mayr et al. designed a pipeline to automatically generate training data for the segmentation problem of drivable areas. However, neither of these two approaches can generate the label of drivable areas and road anomalies simultaneously. With the development of graphic rendering engines, some researchers utilized the data from real- world environments to construct simulation environments, where it is effortless to acquire segmentation labels.Gaidon et al. proposed a real-to- virtual world cloning approach to generate photorealistic simulation environments. Xia et al. proposed Gibson Environment for developing real-world perception for active agents. However, there is a certain gap between the data collected in simulation environments and in real-worldenvironments. For example, the depth images collected in simulation environments are very dense, but the actual depth images collected by RGB-D cameras have some invalid pixels. Besides, these simulation environments cannot cover the common indoor and outdoor scenes where robotic wheelchairs usually work and the road anomalies that robotic wheelchairs may encounter in real environments.Furthermore, in order to construct such a photo-realistic simulation environment for robotic wheelchairs, either large amounts of paired data for target-source domains or physical measurements of important objects in the scene are needed, which is time-consuming and labor-intensive.

Self-Supervised Semantic Segmentation Zeng et al. proposed a self-supervised approach to semantic segmentation by utilizing the knownbackground information. However, theirapproach is only suitable for single confined scenes with objects that can be moved. In common scenes where robotic wheelchairs usually work, there aremany anomalies (e.g., street lamps) thatcan not be moved. Besides, some researchers designed proxytasks (e.g., image colorization) to extractmeaningful representations for self supervised learning. Recently, Zhanet al. presented a mix-and-matchtuning approach for self supervised semantic segmentation based on theexisted proxy tasks.

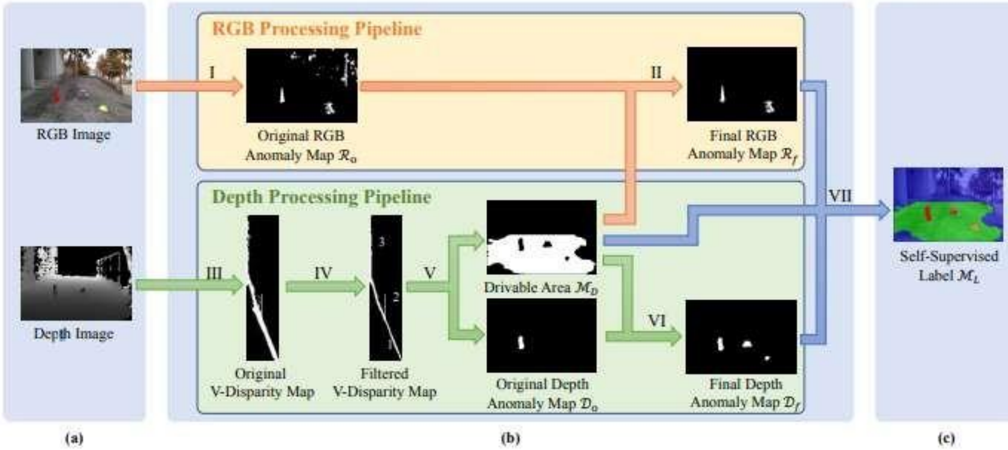

Fig. 3: The overview of our proposed SSLG

PROPOSED METHOD

Figure 3 consists of (a) Input of RGB-Dimages(b) Processing pipeline of RGB-D images and (c) Output of self-supervised labels. (b) is composed of RGB Processing Pipeline shown in the orange box, DepthProcessing Pipeline shown in the green box and (VII) Final Segmentation Label Generator shown in the blue lines. The RGB Processing Pipeline consists of (I) Original RGB Anomaly Map Generator and (II) Generation of final RGB anomaly maps. The Depth Processing Pipeline consists of

(III) Computation of original v-disparity maps, (IV) Filtering of original v-disparity maps, (V) Extraction of the drivable areaand original depth anomaly maps as well as

(VI) Generation of final depth anomaly maps.

Evaluation Process Specification We defined an evaluation process that is depicted in Figure 4. This process consists of five steps: Route selection; Dataacquisition; Anomaly recognition; Anomaly identification; and

Results

With the objective of evaluating the anomaly detection model described in Section 3 with data collected in real-world conditions, we designed, developed and deployed a cloud-based road condition/anomaly information managementservice based on vehicle context data, collected during driving activities using smartphones, and a Web application for identifying road anomalies. The overallfunctional process is as follows.Smartphones collect inertial data, i.e., accelerometer data for X, Y, and Z axes,GPS coordinates, speed, and bearing (thisattribute is not used for anomaly detection), during driving activity. For this,smartphones are mounted on the vehicle's windshield in a vertical portrait orientation.

Self-Supervised Label Generator

Our proposed Self-Supervised Label Generator (SSLG) is designed to generateself-supervised labels of drivable areas and road anomalies automatically. The overviewof the SSLG is shown in Fig. 3. We firstly elaborate the depth processing pipeline inspired by [8]. As derived in [8], for anRGB-D camera consisting of two cameras, the projection of the real world point P with coordinates of (X, Y, Z) on the image coordinates (U, V) can be computed by (1)–(3):

$$U_l = u_l - u_0 = f\frac{X + b/2}{Y\sin\theta + Z\cos\theta} \qquad (1)$$

$$U_r = u_r - u_0 = f\frac{X - b/2}{Y\sin\theta + Z\cos\theta} \qquad (2)$$

$$V = v - v_0 = f\frac{Y\cos\theta - Z\sin\theta}{Y\sin\theta + Z\cos\theta} \qquad (3)$$

where b is the distance between the optical centers of two cameras; f is the focal length; (u0, v0) is the center of the image plane; ul , ur are the projection of the point P on two cameras, respectively; θ is the pitch angle with respect to the ground plane. Then, the disparity Δ can be calculated by (4):

$$\Delta = u_l - u_r = f\frac{b}{Y\sin\theta + Z\cos\theta} \qquad (4)$$

Horizontal planes in the real world coordinates can be represented by Y = m,which leads to:

$$\Delta\frac{m}{b} = V\cos\theta + f\sin\theta \qquad (5)$$

Similarly, vertical planes in the real world coordinates can be represented by Z = n,which leads to:

$$\Delta\frac{n}{b} = V\sin\theta + f\cos\theta \qquad (6)$$

Equation (5) and (6) show that horizontal planes and vertical planes in the real world coordinates can be projected as straight linesin the v-disparity map. Actually, [8] proposed that this conclusion applies to all planes. The intuition behind the depth processing pipeline is that drivable areas canbe regarded as planes in most cases and road anomalies can also be regarded as planes approximately. Then, the segmentation problem can be converted into a straight lineextraction problem. The original v-disparity map can be obtained by computing the depthhistogram of each row in the depth image (Fig. 3 III). Since the computed v-disparity map often contains much noise, the steerablefilter with the second derivatives ofGaussian as the basis function [9] is applied to filter the original v-disparity map (Fig. 3 IV). Then, the Hough Transform algorithm [27] is applied to extract straight lines in the filtered v-disparity map (Fig. 3 IV). Gao et al. [9] concluded that the drivable area is dominant in v-disparity maps; the straight line with the smallest disparity is the projection of the infinity plane; the remaining straight lines except the two straight lines mentioned above are marked as road anomalies. According to these conclusions, we firstly filter out the straight lines representing road anomalies with a height smaller than 5cm from the surface of the drivable area according to their lengths. Then, it is easy to find that in the filtered v- disparity map, straight line No.1, No.2 and No.3 represent the drivable area, the road anomaly and the infinity plane, respectively. After that, we extract the drivable area MD and the original depth anomaly map Do according to the straight line detectionresults (Fig. 3 V).

However, the original depth anomaly map lacks robustness and accuracy because the straight lines representing small roadanomalies are too short and easy to be filtered out together with the noise. For instance, there are three road anomalies in the example (Fig. 3), but there is only one straight line representing road anomalies in the filtered v-disparity map and thus one road anomaly detected in the original depth anomaly map. The other two small road anomalies (i.e., the brick and the road sign) are filtered out together with the noise. In order to solve this problem, we utilize the drivable area that we have already generated. We can find that there are some holes inside the drivable area, which containthe missing road anomalies in the original depth anomaly map. Therefore, we extract holes in the drivable area and then combine the hole detection results with the original depth anomaly map to generate the final depth anomaly map Df , which is furthernormalized to the range [0, 1] (Fig. 3 VI). Although this method will bring some noise to the depth anomaly map, it greatly increases the detection rate of road anomalies to ensure the safety of the riders and we will correct it again with theinformation of RGB images.

Algorithm 1: Original RGB Anomaly Map Generator

Input: $\mathcal{L}$, h, w, σ_s.

Output: $\mathcal{R}_o$.

$\sigma = \min(\mathrm{h}, \mathrm{w})/\sigma_s$

initialize $\mathcal{L}_\omega$ with three channels $(l_\omega, a_\omega, b_\omega)$

construct a Gaussian kernel $\mathcal{G}$ with the size $3\sigma \times 3\sigma$ and the standard deviation σ

$\mathcal{L}_\omega = \mathcal{G}(\mathcal{L})$

$\mathcal{R}_o = \|\mathcal{L} - \mathcal{L}_\omega\|^2$

Now we elaborate the RGB image processing pipeline inspired by. The intuition behind the RGB processing pipeline is that the areas with different colors from surrounding areas are often marked as road anomalies. Based on thisprinciple, we design an original RGB anomaly map generator (Fig. 3 I), which is described in Algorithm 1. Let L denote the image in the Lab color space transformed from the RGB image, and h and w denote the width and the height of the RGB image, respectively.

We generate the original RGB anomaly mapRo by computing the difference between theLab color vector of each pixel and its Gaussian blurred result. In order to suppress the pattern of the drivable area, we choose a large filter scale for the Gaussian kernel to blur each channel of the Lab color space. The size of the kernel is 3σ ×3σ with the standard deviation σ and σs is chosen to be

12 to control the strength of weighting. However, the original RGB anomaly map lacks robustness and accuracy because of theinterference outside the drivable area. In order to solve this problem, we utilize the drivable area that we have already generatedto filter out the noise outside the drivable area (Fig. 3 II). After normalized to the range [0, 1], the final RGB anomaly map Rf is generated. The last step of our proposed SSLG is to combine two anomaly maps and the drivable area to generate the self supervised label. We design a final segmentation label generator (Fig. 3 VII), which is described in Algorithm 2. As for road anomalies, we firstly generate the final anomaly map MA.

Then, we set a threshold κ and the area inthe final anomaly map where the value is greater than κ is marked as road anomalies in the self-supervised label. In our case, we set α to 0.5 and κ to 0.3. The drivable area inthe self-supervised label is the same as the drivable area MD, and the rest area except drivable areas and road anomalies marked above is labeled as the unknown area. Finally the self-supervised label ML isgenerated, which is used for training RGB-Ddata-based semantic segmentation neuralnetworks as described in the followingsections.

Algorithm 2: Final Segmentation Label Generator

Input: $\mathcal{R}_f$, $\mathcal{D}_f$, $\mathcal{M}_D$, h, w, κ.
Output: $\mathcal{M}_L$.
compute $\mathcal{M}_A$ using (7)
for $i \leftarrow 1$ ***to*** h **do**
  for $j \leftarrow 1$ ***to*** w **do**
    if $\mathcal{M}_A(i,j) > \kappa$ **then**
      label $\mathcal{M}_L(i,j)$ as road anomalies
    else if $\mathcal{M}_D(i,j)$ is labeled positive **then**
      label $\mathcal{M}_L(i,j)$ as the drivable area
    else
      label $\mathcal{M}_L(i,j)$ as the unknown area
    end
  end
end

SIMULATION RESULTS

Fig.4. Input image

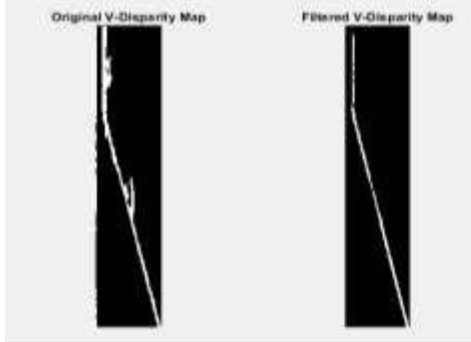

Fig.5. V- disparity map

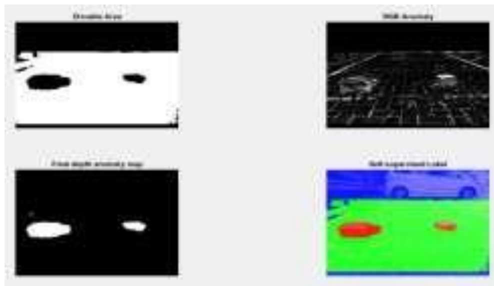

Fig.6. Output result

Conclusion

In this work, we presented a comprehensive study on the drivable area and road anomaly segmentation problem for roboticwheelchairs. A self-supervised approach was proposed, which contains an automatic labeling pipeline for drivable area and road anomaly segmentation. Experimental results showed that our proposed automatic labeling pipeline achieved an impressive speed-up compared to manual labeling. In addition, our proposed self-supervised approach exhibited more robust and accurate results than the state-of-the-art traditional algorithms as well as the state-of-the-art self-supervised algorithms. In our future work, we plan to investigate our work with planning algorithms for robotic wheelchairs to achieve autonomous navigation.

REFERENCES

1. Z. Zhang, "Microsoft kinect sensor and its effect," IEEE multimedia, vol. 19, no.2, pp. 4–10, 2012.
2. Y. Sun, M. Liu, and M. Q.-H. Meng, "Active Perception for Foreground Segmentation: An RGB-D Data-Based Background Modeling Method," IEEE Transactions on Automation Science and Engineering, pp. 1–14, 2019.
3. Motion removal for reliable rgb-dslam in dynamic environments," Robotics and Autonomous Systems, vol. 108, pp. 115 – 128, 2018.
4. Improving rgb-d slam in dynamic environments: A motion removal approach," Robotics and Autonomous Systems, vol. 89, pp. 110 – 122, 2017.
5. C. Hazirbas, L. Ma, C. Domokos, and Cremers, "Fusenet: Incorporating depth into semantic segmentation via fusion-based cnn architecture," in Asian Conference on Computer Vision.Springer, 2016, pp. 213–228.
6. W. Wang and U. Neumann, "Depth- aware cnn for rgb-d segmentation," in The European Conference on Computer Vision (ECCV), September 2018.
7. F. Lateef and Y. Ruichek, "Survey on semantic segmentation using deep learning techniques," Neurocomputing,2019.
8. R. Labayrade, D. Aubert, and J.-P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through" v-disparity" representation," in Intelligent Vehicle Symposium, 2002. IEEE, vol. 2. IEEE, 2002, pp. 646–651.
9. Y. Gao, X. Ai, Y. Wang, J. Rarity, and N. Dahnoun, "Uv-disparity based obstacle detection with 3d camera and steerable filter," in Intelligent Vehicles Symposium (IV), 2011 IEEE. IEEE, 2011, pp. 957–962

10. Bhavana Godavarthi, D Hemanth Kumar, Paparao Nalajala," Driver Behavior Model with Stability in the Presence of Driver Uncertainty using Embedded Design", International Journal of Pure and Applied Mathematics, Volume 117, Issue 20, Pages 837-844, 2017.