

Depth and Image Fusion for Road Obstacle Detection Using Stereo Camera

Oleg Perezyabov¹, Mikhail Gavrilenkov² and Ilya Afanasyev³

^{1,2}*Huawei St. Petersburg Research Center*

³*Independent Researcher*

¹perezyabov.oleg1@huawei.com, ²gavrilenkov.mikhail1@huawei.com, ³ilya.afanasyev@gmail.com

Abstract—This paper is devoted to the detection of objects on a road, performed with a combination of two methods based on both the use of depth information and video analysis of data from a stereo camera. Since neither the time of the appearance of an object on the road, nor its size and shape is known in advance, ML/DL-based approaches are not applicable. The task becomes more complicated due to variations in artificial illumination, inhomogeneous road surface texture, and unknown character and features of the object. To solve this problem we developed the depth and image fusion method that complements a search of small contrast objects by RGB-based method, and obstacle detection by stereo image-based approach with SLIC superpixel segmentation. We conducted experiments with static and low speed obstacles in an underground parking lot and demonstrated the successful work of the developed technique for detecting and even tracking small objects, which can be parking infrastructure objects, things left on the road, wheels, dropped boxes, etc.

I. INTRODUCTION

The intelligent transport monitoring system for roads, parking lots and highways must detect the sudden appearance of obstacles on the road that can lead to an emergency. Frequently objects, boxes, loads, etc. that have fallen from passing vehicles can become a source of such obstacles (that are especially dangerous on busy highways), as well as objects moving at low speed (such as animals).

According to European Agency for Safety and Health at Work (EU-OSHA), around a third of the deaths of people in workplace accidents in the EU are related to transport [1]. What is more, within the accidents that involve the people the reasons can be objects falling from vehicles. The EU-OSHA review of accidents and injuries to road transport drivers notes that "If a cargo is not adequately secured, it can be a danger to the driver and to others: the cargo can fall off the vehicle and form an obstacle that in turn may hurt or kill the driver or other road users... During strong braking or a crash the risk of cargo falling off the vehicle is increased" [1].

For this purpose, monitoring systems used at intersections or parking lots, consisting of surveillance cameras and, sometimes, depth sensors, it is logical to supply with the specific functions for analyzing the traffic situation, timely detection of both static and dynamic obstacles in real time, and warning drivers. It is assumed that sensor data processing can take place both on sensor systems if they have sufficient computing resources (like edge computing), and in the cloud if there is a high-speed data transmission channel (for cloud computing),

and, possibly, on the monitoring system server (if a computing cluster is provided).

It should be noted that different monitoring systems may have different restrictions on obstacle detection. So, Doppler millimeter radars cannot register static objects, lidar systems create dense point clouds that are difficult to process in real time, and also have a limited working distance (about 200 meters). Artificial intelligence systems with the detection and classification of objects can be inefficient, since neither the time of the appearance of the object, nor its nature, size, color, shape, etc. are known in advance. Moreover, it is difficult to even prepare the necessary data set for training a neural network, plus the performance of a ML/DL-based system for real-time high-resolution image processing may be questionable.

In this work, a stereo camera system was chosen for the study, since it meets such requirements as, on the one hand, it provides depth information by computing the disparity map (and, consequently, the distance to the object), on the other hand, the use of a video stream from one of stereo cameras gives us the opportunity to apply computer vision algorithms for object detection and tracking. Merging sensory information for both depth and vision yields more accurate object detection, positioning and characterization. Therefore, this article focuses on a depth and image fusion algorithm that takes advantage of the dual detection of road obstacles through the different processing channels of stereo and RGB information.

It should be noted that for monocular obstacle detection on the road with frame-by-frame image analysis, we used the OpenCV graph segmentation methodology with the following pre-processing: converting an RGB image to HSV with subsequent saturation increase, median filtering, erosion and dilation to remove noise. For depth-based obstacle detection, we use the recovery of disparity map from stereo (performed automatically by selected stereo camera software Stereolabs Zed 2 and Intel RealSense SDK), getting a dense point cloud and then applying the superpixel algorithm, to speed up point cloud processing and noise reduction to reduce the influence of background noise, which fluctuates greatly with time and illumination.

As restrictions on the experiments, we would note that the

area-of-interest for the underground parking lot was investigated with several types of objects such as boxes of different sizes, the orange-white traffic cones, and a person in the camera's field of view). Since we used Stereolabs Zed 2 and Intel RealSense stereo cameras available on the market and could not change the baseline (the distance between the cameras), the obstacle detection distance is limited to 10-15 meters. Due to significant stereo camera noise in the raw data streams from the SDK, we could not distinguish objects smaller than 10x10x10 cm without special RGB camera processing.

As the main research contribution of this article, we would like to highlight the following:

- 1) A combined algorithm for detecting obstacles on the road using dual channels for processing stereo and mono information, followed by depth and image fusion with specially selected metrics.
- 2) An improved technique for detecting obstacles on the road applying a monocular camera with frame-by-frame image analysis, associated with graph segmentation and special pre-processing, using saturation increase, median filtering, erosion and dilation, rather than using neural networks and a priori known scene (i.e. a pre-recorded scene with an obstacle-free road).

The rest of the paper is organized as follows: The Section II considers research papers on road obstacles detection. The detailed methodology and implementation of our proposed algorithm is explained with examples in the Section III. The Section IV describes the experimental setup, tests and results. Finally, we discuss and conclude in the Section V.

II. RELATED WORK

The problem of obstacle detection on the road is devoted to many different studies and reviews [2]–[4]. Part of the research is devoted to the real-time detection of obstacles from on-board sensors of intelligent transport systems, driver assistance systems and automated vehicles [2], [5], [6]. And as noted in [2], the most common autonomous support functions for commercially available premium piloted vehicles are autonomous highway driving, semi-autonomous parking, and braking.

The methods that apply image processing are various. For example, the method described in article [5] uses morphological filtering of camera frames to detect obstacles in front of a moving vehicle in order to warn the driver about avoiding collisions (in this case, other road users can often be considered as an obstacle). In our case, we use traffic monitoring with a stationary sensor (i.e. stereo camera).

Frequently, the detection of known classes of objects on the road for transport tasks is carried out using ML/DL-based approaches [7], [8]. But, as we have already noted, if the type, shape and moment of an obstacle's appearance on the road is not known in advance, the efficiency of applying neural networks can be questionable. In addition, assembling a dataset for neural network training can be a difficult task.

Often in studies, to detect obstacles cameras and computer vision algorithms are used [3], [7]. However, sometimes depth sensors such as stereo cameras [6], [8], LiDARs [2], [9] and millimeter wave radars [2], [10] can be applied as well.

In our study, we use a stereo camera, as it combines the advantages of using depth information (and therefore the distance to the object) and the ability to process the video stream using computer vision algorithms.

III. METHODOLOGY AND ALGORITHM IMPLEMENTATION

Obstacles detection experiments have been performed on dataset collected by our own in underground parking scenario (see the Section IV). For ZED2 camera (Fig. 13a) the dataset includes video sequence from left and right RGB cameras and disparity map, retrieved via stereo camera SDK. For Intel RealSense Camera (Fig. 13b) the dataset consisted of left and right grayscale images, RGB image for color data and disparity map, obtained by camera SDK. The dataset scenario is as follows: parking spaces and parking passages with the boxes of 10x10x10 cm and 20x20x20 cm being thrown or brought by human in the camera field of view.

The proposed method of depth and RGB image fusion for road obstacle detection, and its detailed workflow is demonstrated in Fig. 1.

A. RGB image-based obstacle detection

We are faced with the task of searching for all the obstacles that can lead to traffic accidents. Therefore, our goal is to find both objects that were already on the ground when the camera was turned on, and objects that appeared in the field of view later. That is why we cannot apply a method that uses the so-called key frame, that is, a reference image in which we are sure that there are no obstacles. Also, we cannot implement a method that requires a priori information about the foreground and background, because we do not have enough information about the possible obstacle and its properties. Among the widespread image segmentation methods, the graph segmentation described in [11] seems to be very promising for solving our task, since it segments the image both in terms of spatial and color Euclidean distance (equation 1).

$$dist(p_i, p_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + \dots} \\ \sqrt{\dots + (r_i - r_j)^2 + (g_i - g_j)^2 + (b_i - b_j)^2} \quad (1)$$

where x and y are 2D coordinates of p_i and p_j image points, and r, g, b are the intensities of the red, green, and blue pixels, respectively.

The Fig. 2 shows the original test image (left) and its graph segmentation with the implemented color map (right) without preprocessing. Note that hereinafter we use the OpenCV implementation of graph segmentation with the following parameters: $\sigma = 0.6$, $k = 1074$ and $min_size = 185$.

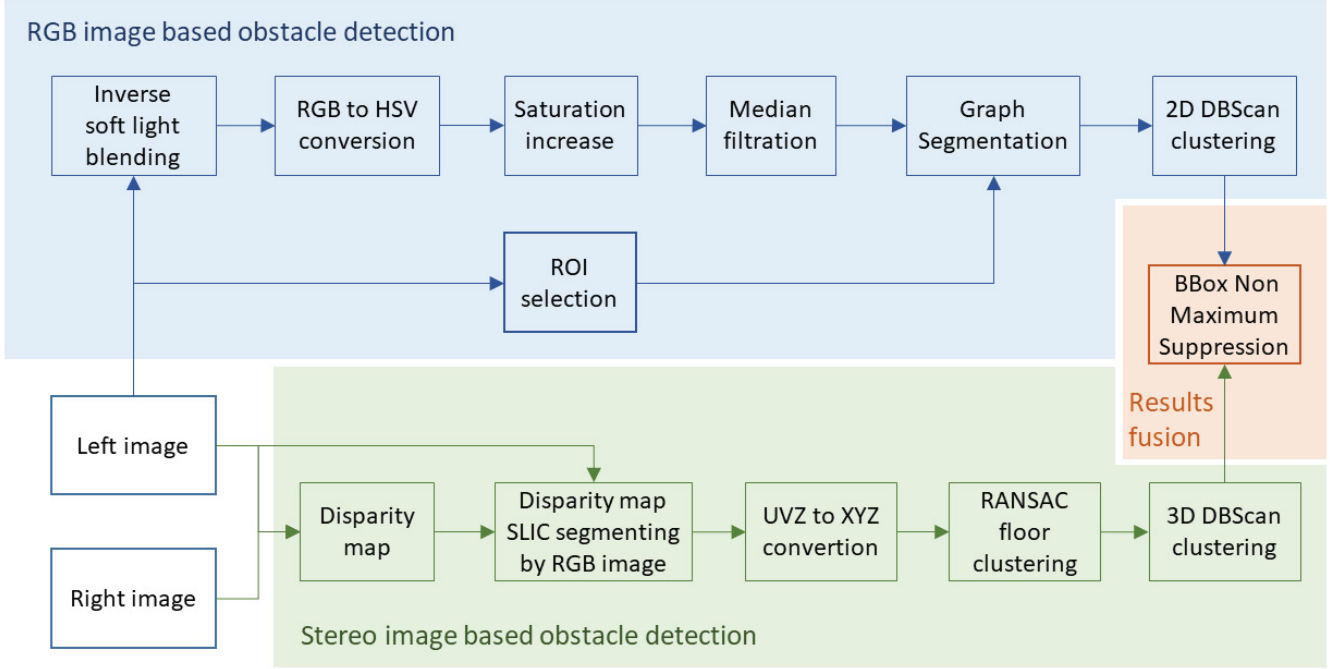


Fig. 1. The methodology of depth and image fusion for road obstacle detection using stereo camera

Fig. 2 demonstrates that the floor is segmented quite well. However, only the larger box from two target objects (light brown boxes) that are used here as obstacles, is segmented. But even this box is faded in the image with some cracks. The small box is not segmented at all.



Fig. 2. Original test image (left) and its graph segmentation with implemented color map (right) without pre-processing

Among the shortcomings of the method for our task, we can note its over-sensitivity to some local changes in light and color, when some random cracks and shadows can be identified by the algorithm as separate segments. To avoid this shortcoming, we implemented image pre-processing, which consists of several steps.

At the first stage, we tried to make local changes in color and light smoother. For this purpose we blended the image with its inverse copy using the so-called soft light blending (the equation 2) [12], [13]. To avoid the discontinuity at the 50% gray level that is inherent in the main method, the Pegtop formula was used [13].

$$f_{pegtop}(a, b) = (1 - 2b)a^2 + 2ba, \quad (2)$$

where a is the pixel intensity value for the image being processed, and b is the pixel intensity value for its inverse copy, $b = 255 - a$. Both images are of type *uint8* and all image pixels are in the range of intensities from 0 to 255.

Now that we have applied the Pegtop soft light blending, we can see in Fig. 3 the changes in the segmentation. The small box is still not segmented, but the big number of cracks in the large box have disappeared, as well as a part of the shadow near the column.



Fig. 3. Original test image (left) and its graph segmentation visualized as the colormap (right) using Pegtop soft light blending with an inverse copy of the image

The Fig. 4 shows another example of soft light blending with an inverse copy of the image with the same segmentation settings. The finally segmented floor area after such pre-processing is detected at a greater distance and does not fade with the wall. In addition, road poles that are located together in the foreground of the image are correctly segmented from each other.

At the second step of pre-processing, the conversion from the RGB color space to the HSV color space (the equation 3) was



Fig. 4. Original images (left) and images with the results of their graph segmentation (right). The bottom images are before pre-processing, and the top image - after Pegtop soft light blending (with its inverse copy) applied to the bottom original image

performed (see, the Fig. 5, left), since in most cases this provides better image contrast.

$$\begin{cases} V \leftarrow \max(R, G, B) \\ S \leftarrow \begin{cases} V - \min(R, G, B) & \text{if } V \neq 0 \\ 0 & \text{otherwise} \end{cases} \\ H \leftarrow \begin{cases} 60(G - B)/(V - \min(R, G, B)) & \text{if } V = R \\ 120 + 60(B - R)/(V - \min(R, G, B)) & \text{if } V = G \\ 240 + 60(R - G)/(V - \min(R, G, B)) & \text{if } V = B \\ 0 & \text{if } R = G = B \end{cases} \\ \text{if } H < 0 \text{ then } H \leftarrow H + 360 \end{cases} \quad (3)$$

where R , G and B are red, green and blue color space components respectively; H , S and V are Hue, Saturation and Value color space components, which are in the range of $0 \leq H \leq 360$, $0 \leq S \leq 1$, and $0 \leq V \leq 1$.



Fig. 5. Original test image after RGB to HSV color space conversion (left) and its graph segmentation visualized as the colormap (right)

Both target boxes are now segmented, however there are more unwanted segments in the image, especially in the floor area (Fig. 5, right). After increasing the saturation by 50%, the image contrast improved even more (Fig. 6, left). The quality of the segments for both boxes increased, and some unnecessary segments on the right side disappeared, but new large segments appeared on the left side (Fig. 6, right).

Although the graph segmentation filter has some parameters to adjust the size and number of segments, median filtering was performed to avoid false segments due to small cracks and spots on the ground. The Fig. 7 shows the results of this step. The kernel size in this case is 5×5 . The boxes now have fairly clean individual segments, and the number of incorrect segments is quite small.

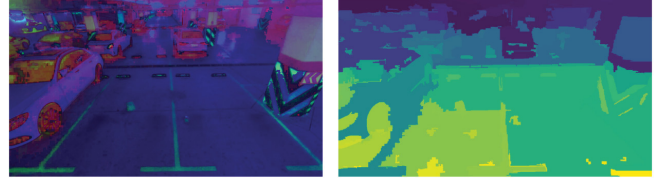


Fig. 6. Original test image after increasing the saturation in the HSV color space conversion (left) and its graph segmentation visualized as the colormap (right)

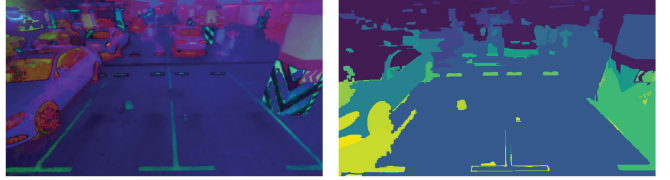


Fig. 7. Original test image after increasing the saturation in the HSV color space conversion and image blurring (left) and its graph segmentation visualized as the colormap (right)

However, to reduce further the number of unwanted segments, morphological erosion with a fairly large kernel (7×7 in our case) was implemented in the segmented image, as shown in Figure 8. It is a possible optional step (that is why we did not show it on the workflow in Fig. 1) that can be applied to the segmented image (7, right) if there is a lot of small spots.

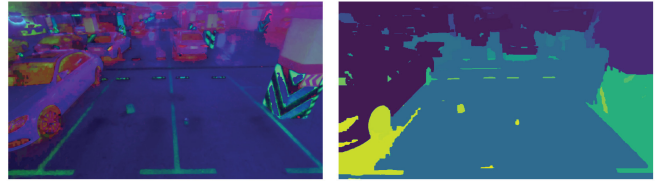


Fig. 8. Original test image after increasing the saturation in the HSV color space conversion and image blurring (left) and its graph segmentation after morphological erosion of the segmented image shown as a colormap (right)

Then the segments can be clustered using DBScan, which is a popular density-based clustering algorithm, or even found by looking for unique values. To improve the recognition rate and reduce the number of false alarms caused by cracks and marks on the ground, it is recommended to manually select the region of interest (ROI) that does not include stationary objects, road passages and parking lanes. The result of the algorithm implementation with selected ROI is shown in Fig. 9. The red color in the upper right figure indicates that obstacles are being detected.

B. Stereo-based obstacle detection

While RGB image segmentation techniques are quite robust when they detect even very small objects, they are still limited by contrast of object to background. For better detection of obstacles with low contrast to the background, a disparity map generated by a stereo camera was used. This method also has

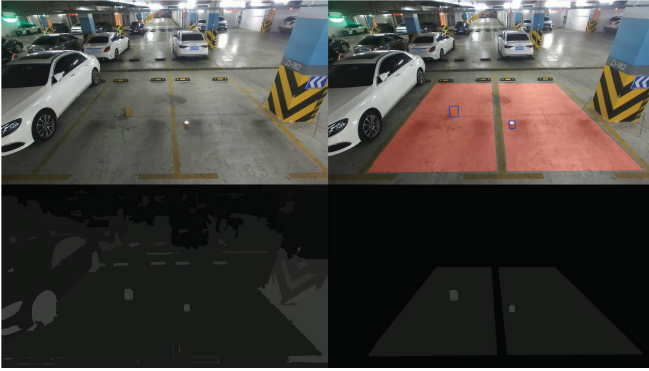


Fig. 9. Original test image (top left), and its segmented image (bottom left), the segmented image with selected ROI (bottom right), and the result of obstacle detection (top right)

some limitations. First, the accuracy of the stereo image is directly proportional to the distance, the maximum value of which depends on the baseline between the cameras. In our case, the baseline is about 15 cm, so the maximum distance is limited to about 10 m, and the best accuracy is in the range of 1.5-4 m. Second, all disparity map acquisition methods are not very accurate in the presence of structureless and mirror surfaces, therefore, in this case, the disparity map appears to fluctuate on these surfaces, making smaller obstacle detection harder or even impossible (Fig. 10).

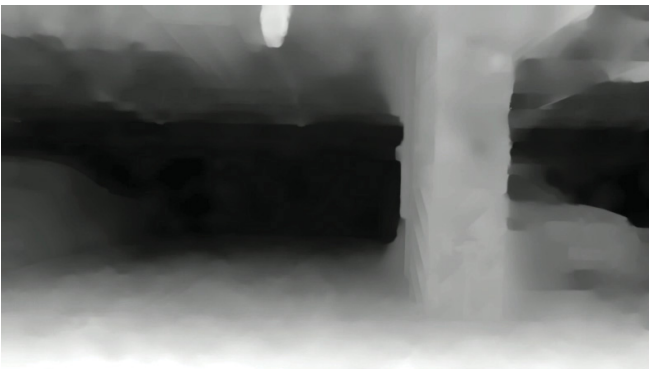


Fig. 10. The example of fluctuations in disparity map obtained from Stereolabs ZED 2 camera

To improve the accuracy of obstacle recognition, we used the disparity map averaging by relevant RGB image segments, obtained using SLIC segmentation [14]. To improve the accuracy of obstacle recognition, we used the disparity map averaging by relevant RGB image segments, obtained using SLIC segmentation [3]. In other words, in this particular case, the RGB image is used as a source for for disparity map segmentation. The positive feature of most superpixel segmentation methods is that they provide high-quality detection of object edges. It also speeds up the next steps because you can work with fewer superpixels instead of working with all the pixels in the image, which can be very slow at higher resolutions.

Although the disparity map contains depth information, we intend to find 3D clusters therefore we need to convert it to 3D point cloud by converting UVZ coordinates to XYZ coordinates. An obstacle can be defined as a volume above the ground level, so to search for such an object, we should find the ground points, and detect all points above this level that represent an obstacle in our case. To find the ground level, which is usually a plane, we use RANSAC algorithm [15] that is a robust model fitting method to detect the ground using a plane model.

Point cloud clustering is performed by DBScan [16] as this algorithm is very efficient for clustering points by density. In Fig. 11, the red 3D superpixels represent the result of RANSAC clustering, and other cluster colors show various obstacles in 3D space.

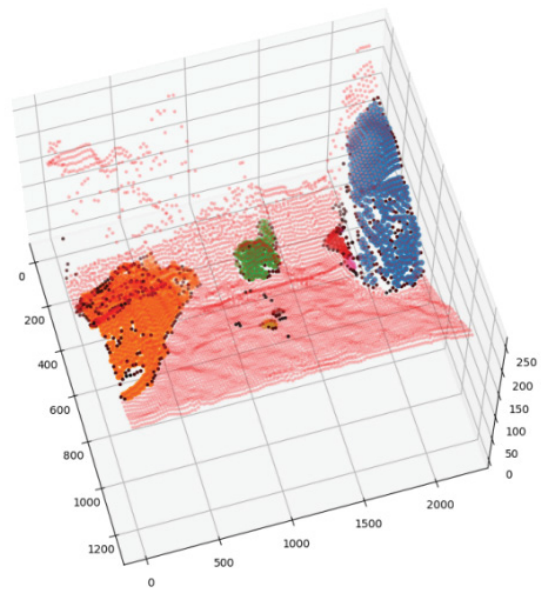


Fig. 11. Point cloud after RANSAC and DBScan clustering

In Fig. 12 the clusters of found obstacles projected onto the image. The large objects, such as cars and big boxes are detected robustly, however the smaller boxes are faded with a floor cluster due to the poor quality of the disparity map.

C. Fusion

In order to take advantage of both methods: searching small contrast obstacles in the RGB-based method and finding larger obstacles without the influence of contrast, we fused the bounding boxes from both algorithms using non-maximum suppression by central point distance. This will allow us to gain the obstacle detection rate in the scene, using both stereo and image analysis techniques.

IV. EXPERIMENTAL SETUP, TESTS AND RESULTS

We conducted various experiments in the region of interest of the underground parking with several types of objects, such as boxes of different sizes (10x10x10cm and 20x20x20cm),



Fig. 12. The results of object segmentation by stereo image-based method with SLIC superpixel segmentation

orange-white traffic cones and, sometimes, a person in the camera's field of view. Thus, we collected our own dataset for such objects by moving the boxes in space and placing them on the floor at different angles to the camera, and sometimes throwing them into the camera's field of view so that they rolled along the floor at low speed.

For our experiments on road obstacle detection, we used the Intel RealSense and Stereolabs Zed 2 stereo cameras available on the market (the stereo cameras' photo in the Figure 13 and characteristics in the Table I). Intel RealSense Depth Camera D455 is the stereoscopic camera for both indoor and outdoor applications with possibility to observe the guaranteed measuring range from 0.6 to 6 m (in some cases, over 10m that varies with lighting conditions) [17], [18]. For Intel RealSense camera the dataset consisted of left and right grayscale images, RGB images for a color camera and disparity map, obtained with camera' SDK. Intel RealSense SDK 2.0 provides an on-chip self-calibration option that allows D455 stereo camera calibration in less than 15 seconds without the need for specialized targets [17]. Stereolabs Zed 2 is also the stereoscopic camera for indoor and outdoor applications with 120° Wide-Angle Field of View, and the measuring depth range of 0.2 to 20 m (since it has wider baseline) [19], [20]. Its SDK has the lightweight neural network for stereo matching that can bring some benefits for stereo depth sensing, spatial object detection and Positional Tracking [19]. For Stereolabs ZED 2 camera dataset includes video streams from left and right RGB cameras and a disparity map, retrieved from stereo camera SDK.

Since we could not adjust the baseline (the distance between the cameras), the obstacle detection range is limited to 10-15 meters. Due to the significant stereo camera noise in the raw data streams from the cameras' SDK, the use of the stereo algorithm alone without RGB camera processing (discussed in the Section III) is not sufficient to distinguish objects smaller than 10x10x10cm.

The experiments were carried out on a conventional computer with an integrated graphics card, the characteristics of which



(a) Stereolabs ZED 2

(b) Intel RealSense D455

Fig. 13. The stereocameras that we used to collect the road obstacle datasets in a parking lot scene

TABLE I. DEPTH CAMERA SPECIFICATION FOR INTEL REALSENSE D455 [18] AND STEREO LABS ZED 2 [20]

Intel RealSense Depth Camera D455	Configuration
Stereo depth resolution	2x (1280x720) @30fps
Depth streaming	up to 90 fps
Depth Field of View (H x V)	87° x 58°
Depth range	0.6 to 6 m
Depth Accuracy	<2% up to 4m
Baseline	95 mm
RGB streaming resolution	1920x1080 @30fps
RGB sensor FOV (H x V)	90° x 65°

Stereolabs Depth Camera ZED 2	Configuration
Stereo depth resolution	2x (2208x1242) @15fps 2x (1920x1080) @30fps 2x (1280x720) @60fps 2x (672x376) @100fps
Depth Field of View (H x V)	110° x 70°
Depth Range	0.2 - 20 m
Depth Accuracy	<1% up to 3m <5% up to 15m
Baseline	120 mm

are given in the Table II.

TABLE II. SPECIFICATIONS OF MACHINE

System	Configuration
Operating System	Windows 10
Processor	3.2 GHz Intel core i7-8700
Video adapter	Intel UHD Graphics 630 (350 - 1200 MHz)
RAM	16 GB
Hardware memory	1000 GB
DirectX version	12
OpenGL version	4.5

Let's analyze the results of field experiments in a parking lot scenario. The Fig. 14 shows the intermediate and final results of frame-to-frame processing with our dual channel obstacle detection method for three boxes of different sizes as obstacles. The Figure 15 demonstrates the results of video processing for a moving box and a moving person that were taken as target objects. The green bounding boxes in the 'FINAL' frames (right bottom pictures) in Fig. 14 and 15 are the result of the stereo camera-based obstacle detection, whereas the blue ones are the RGB image-based obstacle detection. Here, you can see the interesting example when each channel detects definite obstacles, whereas combining them into Depth and Image Fusion method helps to detect each obstacle in the region of interest (ROI) that increases the robustness of obstacle detection.

Let us comment in details the processing steps in the frames of Fig. 14 and 15, focusing on the text annotations in the upper left corner of each picture:

- 'base' is the image from the left camera of the stereopair (Intel RealSense Depth Camera D455);
- 'depth' is the depth map;
- 'segmentation' is the results of the subsequent SLIC, RANSAC and DBSCAN clustering;
- 'bbox' is the bounding boxes around separate clusters of the subsequent SLIC, RANSAC and DBSCAN clustering;
- 'average bbox' - the results of averaging 'bboxes' over 5 frames (that was done to increase robustness of moving obstacle detection);
- 'graph detection' - bounding boxes around separate segments based on RGB image-based segmentation,
- 'ROI' shows the selected region of interest (ROI) in pink color that was used as a mask to compute there bounding boxes around separate clusters of subsequent SLIC, RANSAC and DBSCAN clustering and bounding boxes averaged over 5 frames;
- 'graph ROI' demonstrates the bounding boxes around separate segments of RGB image-based segmentation with selected ROI and bounding boxes averaged over 5 frames;
- 'FINAL' is the fusing result of 'ROI' and 'graph ROI' bounding boxes that shows the outcome of our Depth and Image Fusion method.

Since the performance of RGB image-based obstacle detection approach was not optimized well and takes more time than stereo camera-based obstacle detection, we tested the obstacle detection with Depth and Image Fusion method for moving objects at low speeds.

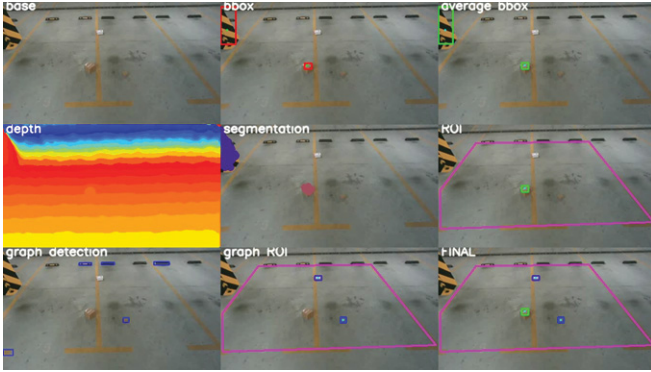


Fig. 14. The results of test frame processing using our dual channel obstacle detection method in a parking lot scenario with three boxes as obstacles. The bottom row shows the work of the 1st channel of the RGB image-based obstacle detection (the detected boxes are in blue color), and the right column represents the 2nd channel of stereo camera-based obstacle detection (the detected box is in green color). The bottom right picture demonstrates the result of Depth and Image Fusion where all boxes are detected with the combination of dual channel processing that increases the robustness of obstacle detection. The region of interest is shown in pink color

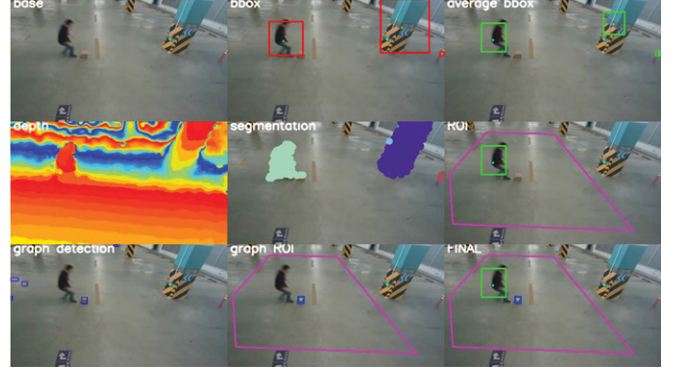


Fig. 15. The results of test frame processing using our dual channel obstacle detection method in a parking lot scenario with a moving box and a person as obstacles. The bottom row shows the work of the 1st channel of the RGB image-based obstacle detection (the detected box is in blue color), and the right column represents the 2nd channel of stereo camera-based obstacle detection (the detected person is in green color). The bottom right picture demonstrates the result of Depth and Image Fusion where all obstacles are detected. The region of interest is shown in pink color

V. CONCLUSIONS AND DISCUSSION

A. Conclusions

In this paper, we presented a methodology for dual detection of obstacles on the road recorded with Stereolabs Zed 2 or Intel RealSense D455 depth camera. We have shown with case study for detection objects with sizes of 10x10x10cm and 20x20x20cm that combining sensory information for both depth data and video, provides more accurate detection, positioning and characterization of objects. The depth and image fusion algorithm developed with the methodology demonstrates the advantages of dual detection of road obstacles using various stereo and RGB processing channels that complement each other.

To detect obstacles on the road using RGB streaming with frame-by-frame image analysis, the OpenCV graph segmentation technique was used, followed by pre-processing: converting an RGB image to HSV with subsequent saturation, median filtering, erosion, and dilation to remove noise. For stereo-based obstacle detection, we use disparity map reconstruction from stereo (performed automatically by selected stereo camera SDK), get a dense point cloud, and then apply a superpixel algorithm based on SLIC segmentation to speed up the point cloud processing, compute a floor plane using the RANSAC algorithm and detect objects using 3D DBSCAN clustering.

Since we have the significant background noise, which fluctuates greatly with time and illumination in the raw data streams from the SDK (we observe it for both Stereolabs Zed 2 or Intel RealSense D455 depth camera output data, and suppose that it goes from disparity map estimation), we could not distinguish between objects smaller than 10x10x10 cm without the special processing of RGB channel. Thus, the applying dual detection of obstacles on the road with the developed

depth and image fusion algorithm gives encouraging results.

The main research contribution of this paper:

- A stereo camera-based depth and image fusion algorithm for road obstacle detection that take advantages in searching small contrast objects by RGB-based method and finding larger obstacles without the influence of contrast by stereo image-based approach with SLIC superpixel segmentation.
- An improved road obstacle detection technique using a monocular camera with frame-by-frame image analysis that applies the OpenCV graph segmentation after the preliminary processing: converting an RGB image to HSV color space with saturation increase, median filtering, morphological erosion and dilation, instead of using neural networks and a pre-recorded scene.

B. Discussion

In some cases, when the angular size of the detected object is close to the size of all sorts of different random cracks or spots, in the first part of the algorithm (by image) there may occur false positive alarms of system. Since they are not constant, but for units of frames, in general one can try to filter them by detection time, which, in its turn will for sure affect the quality and detection rate of objects, actually being in the camera frame.

Although the presented method has only been tested in a parking scenario at relatively short distances, it is reasonable to assume, based on the theory of stereo vision and epipolar geometry, that the method can be used at longer distances by increasing the baseline for stereo camera (which is only about 15 cm in our case) and focal length. In our further research we suppose to compare this detection method with existing ones in terms of accuracy, algorithm performance, and robustness to the various object detection.

REFERENCES

- [1] N. Copsey, L. Drupsteen, J. v. Kampen, L. Kuijt-Evers, E. Schmitz-Felten, and M. Verjans, "A review of accidents and injuries to road transport drivers," 2010.
- [2] X. Yu and M. Marinov, "A study on recent developments and issues with obstacle detection systems for automated vehicles," *Sustainability*, vol. 12, no. 8, p. 3281, 2020.
- [3] T. Rateke and A. von Wangenheim, "Passive vision road obstacle detection: a literature mapping," *International Journal of Computers and Applications*, pp. 1–20, 2020.
- [4] J.-w. Hu, B.-y. Zheng, C. Wang, C.-h. Zhao, X.-l. Hou, Q. Pan, and Z. Xu, "A survey on multi-sensor fusion based obstacle detection for intelligent ground vehicles in off-road environments," *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 5, pp. 675–692, 2020.
- [5] S. Godha, "On-road obstacle detection system for driver assistance," *Asia Pacific Journal of Engineering Science and Technology*, vol. 3, no. 1, pp. 16–21, 2017.
- [6] A. Dhiman and R. Klette, "Pothole detection using computer vision and learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3536–3550, 2019.
- [7] T. Rateke and A. v. Wangenheim, "Road obstacles positional and dynamic features extraction combining object detection, stereo disparity maps and optical flow data," *Machine Vision and Applications*, vol. 31, no. 7, pp. 1–11, 2020.
- [8] L. Chen, J. Yang, and H. Kong, "Lidar-histogram for fast road and obstacle detection," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 1343–1348.
- [9] X. Zhang, M. Zhou, P. Qiu, Y. Huang, and J. Li, "Radar and vision fusion for the real-time obstacle detection and identification," *Industrial Robot: the int. journal of robotics research and application*, 2019.
- [10] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International journal of computer vision*, vol. 59, no. 2, pp. 167–181, 2004.
- [11] "Wikipedia: Blend modes," https://en.wikipedia.org/wiki/Blend_modes, accessed: 2022-03-03.
- [12] "Pegtop soft light mode," <http://www.pegtop.net/delphi/articles/blendmodes/softlight.htm>, accessed: 2022-03-03.
- [13] 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 PAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [14] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, p. 381–395, jun 1981. [Online]. Available: <https://doi.org/10.1145/358669.358692>
- [15] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of Int. Conf. on Knowledge Discovery and Data Mining*, ser. KDD'96. AAAI Press, 1996, p. 226–231.
- [16] "Intel realsense depth camera," <https://www.intelrealsense.com/depth-camera-d455/>, accessed: 2022-03-03.
- [17] "Intel realsense. product family d400 series. datasheet. depth camera d455," <https://www.intelrealsense.com/wp-content/uploads/2020/06/Intel-RealSense-D400-Series-Datasheet-June-2020.pdf>, 2020.
- [18] "Stereolabs zed 2 depth camera," <https://www.stereolabs.com/zed-2/>.
- [19] "Zed 2 camera and sdk overview. datasheet," <https://cdn.stereolabs.com/assets/datasheets/zed2-camera-datasheet.pdf>, 2020.