Detection and Distance Estimation against Motorcycles as Navigation Aids for Visually-impaired People

1st Indrabayu
Informatics Engineering Department
Hasanuddin University
Makassar, Indonesia
indrabayu@unhas.ac.id

2nd Nur Latifah Jamaluddin
Postgraduate Student of Electrical Engineering Department
Hasanuddin University
Makassar, Indonesia
nurlatifahjamaluddin@gmail.com

3rd Intan Sari Areni
Electrical Engineering Department
Hasanuddin University
Makassar, Indonesia
intan@unhas.ac.id

***Abstract*—Visually impaired people are limited to detecting obstacles when walking in an outdoor environment. This research aims to detect, estimate distance and relative position of the visually impaired people to the obstacle, particularly parked motorcycle. The proposed system uses the Single Shot Multibox Detector (SSD) Mobilenet for parked motorcycle detection by testing several different learning rates. The Pinhole Camera Model method is used to estimate the distances from 2 to 5 meters by comparing the actual motorcycle with the motorcycle in the image based on the similar triangel principle. The last stage is estimating the relative normal horizontal position of the visually impaired people to the obstacle to determine whether the position of the visually impaired people is in line with the right, the middle, or the left side of the parked motorcycle. This research uses 500 images, 80% as training data, and 20% as testing data. There are four scenarios based on the user's height with six parked motorcycles positions, i.e., the front, rear, right, left, right oblique, and left oblique. In addition, the tilting position of the smartphone camera uses two different angles, depending on the user's height. The highest parked motorcycle detection results obtained at the learning rate 10^{-3} with an accuracy of 95.04%, the average RMSE of the estimated distance is 0.11, and the highest accuracy of the estimated relative position is 83.26% at 2 meters distance.**

Keywords*—*visually impaired, motorcycle recognition, pinhole camera model

I. INTRODUCTION

For people with visual impairment, physical limitations provide plenty of daily activities. Visually impaired people need to get great attention to be able to move freely and safely. Based on the data from the World Health Organization (WHO) in October 2018, there are 217 million people globally experiencing moderate to severe vision problems, 36 million of which are blind [1]. Recorded by the Ministry of Manpower and Transmigration of the Republic of Indonesia, the number of persons with disabilities in 2010 was 7,126,409. From this number, 2,137,923 were visually impaired people, which was the highest number among the disabilities in Indonesia [2].

Limitation of vision causes visually impaired people to be limited in moving, especially when walking and recognizing objects throughout [3]. In the outdoor environment, visually impaired people unable to move freely and safely without the help of others or a stick [4]. Recognizing objects in the house is certainly quite different from objects in an outdoor environment. Inside the house, visually impaired people are familiar with the location of objects. Meanwhile, road conditions and obstacles are not easily predicted, such as the presence of trash cans, poles, or even vehicles that park on the roadside. Based on the results of the interview with people at the Blind foundation, eight from ten chose parked motorcycles as the most frequently hit vehicle. The low height of parked motorcycles cause visually impaired people difficult to reach using hands and stick.

In general, the visually impaired people use a stick when walking. However, the stick is very limited and inaccurate to provide information on the environmental situation, especially detecting obstacles or people a few meters ahead of them [3], [5]. This problem has become a topic of research that is being carried out in recent years. The researches focus on how to detect the object, detect object types, and estimate object distances with various types of methods utilizing modern technology.

Munir et al. developed the stick function to detect obstacles when walking. The voice-enabled stick designed with a sensor capable of detecting obstacles with sound instructions as output to move left, right, straight, or stop [6]. Sharma et al. also designed a smart stick with ultrasonic sensors and infrared to avoid obstacles and detect a puddle [7]. The smart stick is also able to estimate the distance of obstacles to visually impaired people. With the development of image processing, Heya et al. used the Python programming language and OpenCV library to detect objects in the indoor environment and estimate distances with 1 cm maximum error [8].

The use of stick still has many limitations, such as being unable to detect the type of object and estimating the distance and relative position of the visually impaired people to the object. Therefore, this research focuses on the use of image processing to help visually impaired people to detect and to estimate the distance and relative position to avoid the parked motorcycle when walking. In addition, this system is more comfortable compared to using a stick since they can free their hands.

II. LITERATURE STUDY

A. *Single Shot Multibox Detector (SSD) Mobilenet*

Image processing can detect objects using deep learning. Three deep learning methods have high performance, which is RCNN Faster, You Only Look Once (YOLO), and Single Shot Multibox Detector (SSD) [9]. The method with speed, accuracy, and adequate capacity is the most compatible method to be applied to smartphones with limited memory capacity and data processing. Based on these three deep learning methods, SSD is the best, even with a small number of the dataset [10]. In addition, the SSD algorithm also does not use the Region Proposal Network (RPN) network as in the

Faster RCNN method so that the processing becomes faster [11]. Deep learning combines SSD and Mobilenet to produce an efficient detection and tracking of objects. SSD Mobilenet is chosen in this study because it has a simple architecture, is easy to implement and very compatible for smartphone devices [12]. Unlike previous research, this research uses custom training data of parked motorcycle with relatively few data for training data using deep learning.

B. *The Distance Estimation*

The distance estimation of an object is currently one of the crucial things in various fields of science, such as the application of safety in driving, traffic surveillance, personal assistant for visually impaired people [13]. According to Megalingam et al., it is very important for drivers to know how close the distance of other vehicles is when they want to park, especially when the parking lot is crowded. After detecting the vehicle behind it, using the Pinhole Camera Model method can be done estimating the distance to make it easier for drivers to park the vehicle so that fatal accidents can be avoided [14].

Bianco et al. also estimated the distance of objects using perspective geometry from the image form. Using the Pinhole Camera Model, the research used a mono camera and classified the estimated distance of each object using data collected from different sensors [13]. Heya et al. also estimated the distance of obstacles in the indoor environment using the Pinhole Camera Model method. Indrabayu et al. used camera projections to calculate vehicle displacement distances in video frames using the Pinhole Camera Model method to estimate vehicle speeds [15]. These previous researches estimated the distance using a stationary camera and produced stable video data. However, this research uses a moving smartphone camera and produces unstable video data, which makes the processing more challenging.

C. *The Estimation of Relative Position*

While walking, there are several conditions of the visually impaired people position to the object. For parked motorcycle, the possible conditions are the position of the visually impaired people in line with the right, the middle, or the left side of the parked motorcycle. The distance to avoid the parked motorcycle is certainly different in each position. By knowing the relative position of the parked motorcycle, the visually impaired can estimate the distance to avoid it. Previous researches estimated the relative position to objects in various fields. One of them is in an intelligent transportation system, where a mono camera on a car being driven is used to detect cars and road boundaries to estimate the distance of the car driven by the car in front of it [16].

Siddart at al. also estimated a relative position for the visually impaired people using CMOS camera on glasses to detect and notify the object location to the visually impaired people using Fixed DO Solfege, a system that integrates each object syllable with one tone. If the visually impaired people reached the right hand to a point in front of them, the system tells the name of the object. However, this system is not compatible used in outdoor environments [17]. In this research, relative position estimation is carried out to determine whether the visually impaired is in line with the right, the middle, or the left side of the parked motorcycle.

III. Research Methodology

The process of retrieving video data shown in Fig. 1. Visually impaired people as user hang the smartphone horizontally using povie on the user's neck. Povie is a wearable accessory with a purpose to capture Point-of-View (POV) videos using the smartphone. Height and distance of the user to the parked motorcycle represent as h and d. Four users, which are *A, B, C,* and *D* have different height, 148, 154, 167, and 172 cm, respectively, with a distance of 2 to 5 meters. Two types of cameras used in this research have 13 and 16-megapixel image resolutions.

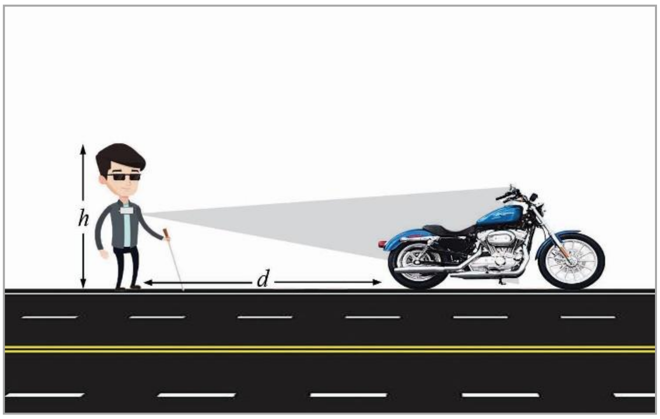


Fig 1. Illustration of data retrieval

Camera tilting of each user is set differently based on height for capturing all parts of the parked motorcycle. The camera tilting is measured against sea level or horizontal plane. The users with a height of 148 and 154 cm use 80º tilting camera and the users with a height of 167 and 172 cm use 70º tilting camera. The camera tilting settings illustrated in Fig 2.

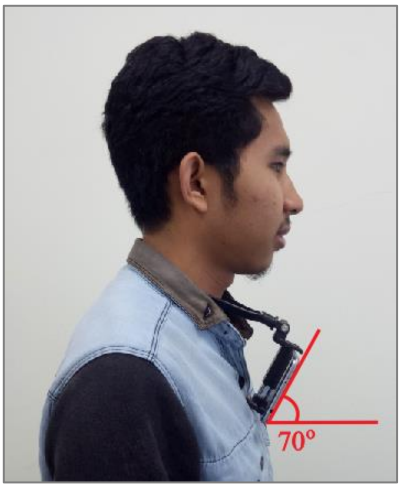


Fig 2. Camera tilting setting

The steps of the research illustrated in the block diagram Fig 3. The initial process is extracting frames from each parked motorcycle video. The training process is then carried out through transfer learning using the SSD Mobilenet pre-trained model. The training result is used to detect the parked motorcycle in various distances from 2 to 5 meters and six different positions. After that, distance estimation is done using the Pinhole Camera Model method and the final step estimating the relative position of the visually impaired people to the parked motorcycle so that they can estimate the steps to avoid it.

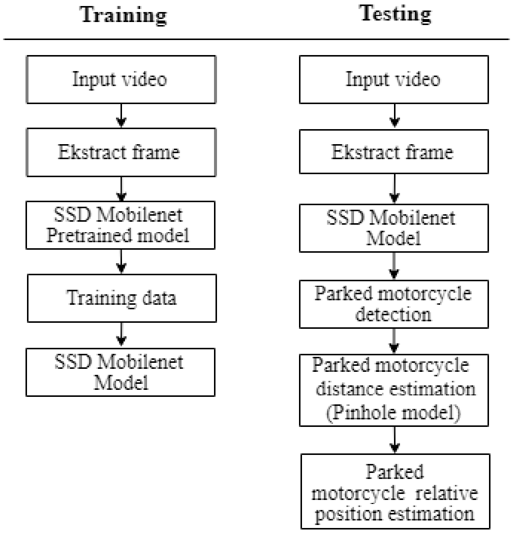


Fig 3. Research flow

A. *The Dataset*

The dataset uses parked motorcycle videos with six positions, which are the front, rear, right side, left side, right oblique, and left oblique with the user walking from 5 to 2 meters toward the parked motorcycle. Fig. 4 shows six parked motorcycle positions. Video extraction is carried out into the image frames. From 500 data image selected, 80% uses as training data and 20% as test data, each with 400 and 100 data images.

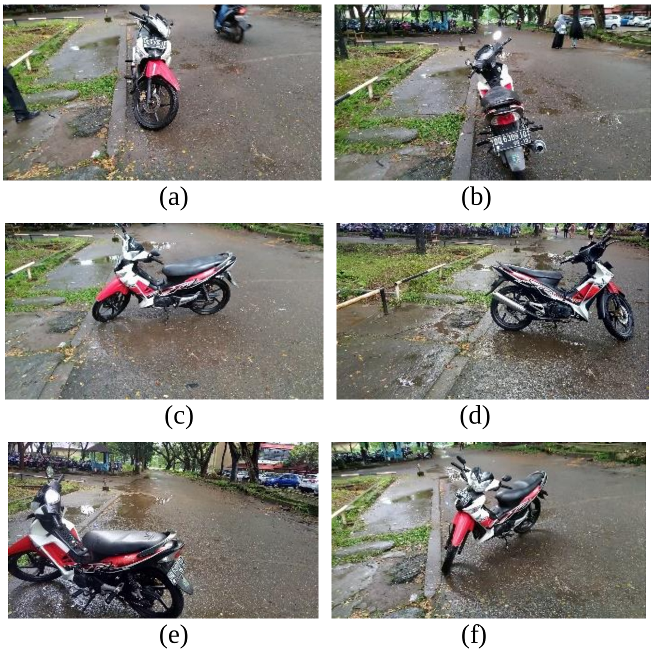


Fig 4. Parked motorcycle dataset in six positions (a) front side, (b) rear side, (c) right side, (d) left side, (e) right oblique side, (f) left oblique side

B. *Transfer Learning using Single Shot Multibox Detector (SSD) Mobilenet*

Four hundred dataset motors with the same number of steps, which are 50,000 and different learning rates of 10-1 to 10-5 are trained using the SSD Mobilenet method. The system uses Tensorflow as an open source library, which is a framework for modeling training results for use on smartphone devices. Discretion the priors is the bounding box (*bbox*) of each feature map on several different scales and ratios, Mobilenet SSDs can detect parked motorcycles in various scales and ratios by finding the highest score and overlap between priors with the predicted motor shape.

C. *Distance and Relative Position Estimation to The Parked Motorcycle*

The distance estimation process is carried out using the Pinhole Camera Model method. This method estimates the distance using the camera projection of the object by comparing the actual object with the object in the image. The Pinhole Camera Model method uses the principle of a similar triangle. This method estimates the distance by comparing the position of the *y* coordinate as the distance of the parked motorcycle to the image. Table 1 shows a comparison of the distance (d_r) with the reference position (p_r) of the parked motorcycle in the image.

TABLE I. COMPARISON OF DISTANCE AND POSITION OF PIXEL REFERENCES IN IMAGES

d_r (meter)	p_r (pixel)
2	905
3	675
4	505
5	475

The closest reference pixel position with bbox (D_d) calculate the maximum *y* coordinate difference in the bbox (p) with each reference position (p_r) using equation (1). After obtaining the closest reference position, the closest reference distance (d_r) can be determined. Using the reference position (p_r) and the closest reference distance (d_r) obtained through the calculation of equation (1), the calculation of the actual distance estimation (d_a) can be calculated using equation (2).

$$D_d = \min \sum_{i=1}^{4} |p - p_r(i)| \quad (1)$$

$$d_a = \frac{p_r \times d_r}{p} \quad (2)$$

Furthermore, the estimated horizontal normal position of the visually impaired people to the parked motorcycle illustrate in Fig. 6. The bbox is divided into three parts, which are bbox 1, 2, and 3 as bbox on the right, the middle and the left side of the parked motorcycle. To determine which side is in line with the visually impaired people, the frame width (w) is divided into two to determine the midpoint of the frame as the position of the visually impaired people in the image (P_a).

$$P_a = \frac{w}{2} \quad (3)$$

$$D_p = \min \sum_{i=1}^{3} |P_a - x_p(i)| \quad (4)$$

The P_a calculation is done using equation (3). Then, P_a is compared with each x coordinate point (x_p) in bbox 1, 2, and 3 using equation (4) to get the closest distance between the visually impaired people and parked motorcycle (D_p).

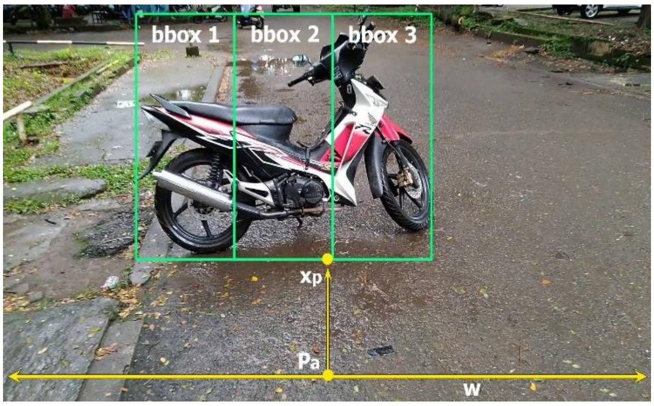


Fig 6. Relative position estimation of the parked motorcycle

IV. Result and Discussion

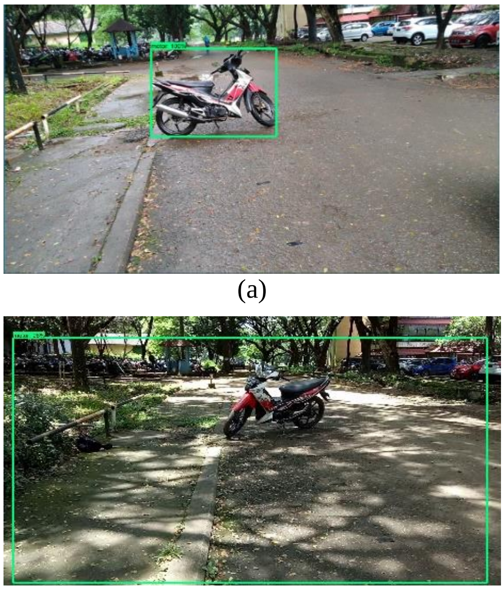
The learning transfer method uses the SSD Mobilenet pre-trained model. The training processes 400 parked motorcycle data images with six positions using 10^{-1} to 10-5 learning rate (LR) and distance of parked motorcycle (DT) 2 to 5 meters. The results of parked motorcycle detection accuracy based on learning rate and distance are shown in Table 2.

TABLE II. ACCURACY OF PARKED MOTORCYCLE DETECTION RESULTS

DT \ LR	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	Average Accuracy (%)
2 m	82.76	94.83	98.28	99.14	94.83	93.97
3 m	58.62	92.24	99.14	96.55	89.66	87.24
4 m	41.38	89.66	91.38	88.79	82.76	78.79
5 m	34.48	63.79	91.38	81.03	57.76	65.69

The 10^{-3} learning rate produces the highest average accuracy for parked motorcycle detection. With the right learning rate, the generated bbox is exactly fit the size of the detected parked motorcycle. Conversely, a poor learning rate produces a larger bbox than the actual size of the parked motorcycle, not covering the entire part, and even not detecting the parked motorcycle. The detection with different learning rate results is shown in Fig. 7.

The situation with less lighting and shadow on the road are the cause of error detection in learning rate with low percentage accuracy. With the right learning rate, SSD Mobilenet can detect parked motorcycle from a distance of 2 to 5 meters without being affected by lighting and shadows. The highest average accuracy of parked motorcycle detection based on distance is at a distance of 2 meters with a percentage of 93.97%, while the lowest is at a distance of 5 meters with a percentage of 65.69%.



(a)

(b)

Fig 7. Parked motorcycle detection result (a) good (b) poor

Furthermore, bbox as a result of parked motorcycle detection is used to estimate the distance using the Pinhole Camera Model method. From the four distance estimation results, the Pinhole Camera Model method can estimate the parked motorcycle distance close to the actual distance. Users with different heights and camera tilting produce an average RMSE at distances of 2 to 5 meters shown in Table 3.

TABLE III. AVERAGE RMSE OF PARKED MOTORCYCLE DISTANCE ESTIMATION RESULTS

Distance \ User	A	B	C	D	Average
2 m	0.07	0.08	0.22	0.08	0.11
3 m	0.13	0.1	0.15	0.12	0.13
4 m	0.15	0.15	0.17	0.14	0.15
5 m	0.37	0.14	0.23	0.33	0.27

The average RMSEs at distances of 2, 3, 4, and 5 meters are 0.11, 0.13, 0.15 and 0.27, respectively. Some estimation errors caused by the povie position used on the user's neck often shift or move when walking. In addition, the video capture results are unstable because the smartphone camera shifts when the user walks. The difference in camera mounting height also affects the parked motorcycle's initial reference point in the frame.

The final stage is estimating the relative position of the user towards the parked motorcycle with the average accuracy shown in Table 4. The farther the position of the parked motorcycle from the user, its position tends to be in the middle of the frame, even if the user is in line with the right or the left side of the parked motorcycle. The highest accuracy is 83.26 at 2 meters distance, while the lowest accuracy is 63.6 at 5 meters distance. Estimation errors are also caused by the instability of the user's smartphone camera when walking, which makes the parked motorcycle position video not accurately captured.

TABLE IV. ACCURACY OF PARKED MOTORCYCLE RELATIVE POSITION RESULTS

User / Distance	**A**	**B**	**C**	**D**	**Average (%)**
2 m	86.67	77.78	90	78.57	83.26
3 m	86.67	61.11	92.5	64.29	76.14
4 m	76.67	50	92.5	57.14	69.08
5 m	83.33	50	67.5	53.57	63.6

V. CONCLUSION

The biggest problem for visually impaired people is detecting objects in the outdoor environment. The use of a stick still cannot help optimally to avoid obstacles. Therefore, this research aims to detect and estimate distance and relative position of the visually impaired people to the obstacle, particularly parked motorcycle. The system can detect parked motorcycle with the highest accuracy of 96.04% at 10^{-3} learning rate. The distance estimation using the Pinhole Camera Model method gets the highest average RMSE of 0.11 and can be used by users with different heights and camera tilting. The highest accuracy of the estimated relative position is 83.26 at 2 meters distance.

The instability of the smartphone camera used by the visually impaired people when walking results in an inaccurate video capture of the parked motorcycle. The instability becomes the limitation of this system. To reduce the instability of video capture, a povie type that is more stable and does not easily shift at the user's neck is needed. For further research, another method can test, and the system can be made real-time and portable by implementing it as an embedded system on Android smartphones based on vision stereo so that it is easier to be used by the visually impaired people.

REFERENCES

[1] "Vision impairment and blindness." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. [Accessed: 13-Mar-2019].

[2] S. D. Poerwanti, "Manpower Management Disabled to Realize Workplace Inclusion," *INKLUSI*, vol. 4, no. 1, pp. 1–24, Jun. 2017.

[3] S. Rahman, S. Ullah, and S. Ullah, "Obstacle Detection in Indoor Environment for Visually Impaired Using Mobile Camera," *J. Phys. Conf. Ser.*, vol. 960, p. 012046, Jan. 2018.

[4] S. Tosun and E. Karaarslan, "Real-Time Object Detection Application for Visually Impaired People: Third Eye," in *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, 2018, pp. 1–6.

[5] W. C. S. S. Simões and V. F. de Lucena, "Blind user wearable audio assistance for indoor navigation based on visual markers and ultrasonic obstacle detection," in *2016 IEEE International Conference on Consumer Electronics (ICCE)*, 2016, pp. 60–63.

[6] H. M. U. Munir, F. Mahmood, A. Zeb, F. Mehmood, U. S. Khan, and J. Iqbal, "The voice enabled stick," in *2017 20th International Conference of Computer and Information Technology (ICCIT)*, 2017, pp. 1–5.

[7] T. Sharma, T. Nalwa, T. Choudhury, S. C. Satapathy, and P. Kumar, "Smart Cane: Better Walking Experience for Blind People," in *2017 3rd International Conference on Computational Intelligence and Networks (CINE)*, 2017, pp. 22–26.

[8] T. A. Heya, S. E. Arefin, A. Chakrabarty, and M. Alam, "Image Processing Based Indoor Localization System for Assisting Visually Impaired People," in *2018 Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS)*, Wuhan, 2018, pp. 1–7.

[9] S. Gianani, A. Mehta, T. Motwani, and R. Shende, "JUVO - An Aid for the Visually Impaired," in *2018 International Conference on Smart City and Emerging Technology (ICSCET)*, 2018, pp. 1–4.

[10] Y. Wang, C. Wang, and H. Zhang, "Combining single shot multibox detector with transfer learning for ship detection using Sentinel-1 images," in *2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA)*, 2017, pp. 1–4.

[11] J. Ryu and S. Kim, "Chinese character detection using modified Single Shot Multibox Detector," in *2018 18th International Conference on Control, Automation and Systems (ICCAS)*, 2018, pp. 1313–1315.

[12] G. Chandan, A. Jain, and H. J. and, "Real Time Object Detection and Tracking Using Deep Learning and OpenCV," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2018, pp. 1305–1308.

[13] S. Bianco, M. Buzzelli, and R. Schettini, "A unifying representation for pixel-precise distance estimation," *Multimed. Tools Appl.*, Aug. 2018.

[14] R. K. Megalingam, V. Shriram, B. Likhith, G. Rajesh, and S. Ghanta, "Monocular distance estimation using pinhole camera approximation to avoid vehicle crash and back-over accidents," in *2016 10th International Conference on Intelligent Systems and Control (ISCO)*, 2016, pp. 1–5.

[15] Indrabayu, R. Y. Bakti, and I. S. Areni, "A modified pinhole camera model for automatic speed detection of diagonally moving vehicle," *J. Eng. Sci. Technol.*, vol. 13, pp. 1722–1734, Jun. 2018.

[16] A. A. Ali and H. A. Hussein, "Distance estimation and vehicle position detection based on monocular camera," in *2016 Al-Sadeq International Conference on Multidisciplinary in IT and Communication Science and Applications (AIC-MITCSA)*, 2016, pp. 1–4.

[17] S. Kalra, S. Jain, and A. Agarwal, "Fixed DO Solfège based object detection and positional analysis for the visually impaired," in *2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2017, pp. 594–598.