Stereo-Aided Blockage Prediction for mmWave V2X Communications

1st Shinsuke Bannai
Grad. School of Inform. and Eng.
The University of Electro-Communications
Tokyo, Japan
s.bannai@uec.ac.jp

2nd Katsuya Suto
Grad. School of Inform. and Eng.
The University of Electro-Communications
Tokyo, Japan
k.suto@uec.ac.jp

***Abstract*—Vision-aided blockage prediction has been recognized as a promising approach for stable Millimeter wave (mmWave) vehicle-to-everything (V2X) communications. For the supplement of channel-based prediction, the vision-aided approach can predict the communication condition in the near future, i.e., several hundred milliseconds; however, both the accuracy and prediction time of exiting work using monocular vision are not enough for mmWave V2X communications. To address the challenge, the paper proposes a stereo-aided blockage prediction that extracts explicit features for blockage prediction using a simple algorithm, i.e., stereo depth estimation. Through the emulation using a dataset generated by CARLA, we demonstrate that the proposal achieves four times faster computation than the existing method using monocular vision while improving the prediction accuracy by 15.688 %.**

***Index Terms*—Deep learning, Millimeter Wave, Computer vision, Stereo camera**

I. Introduction

Millimeter wave (mmWave) is attracting much attention for vehicle-to-everything (V2X) communications because it can support cooperative perception applications such as image and point cloud sharing. However, the loss of connection in mmWave communications occurs due to sudden blockages by surrounding mobility. Blockage prediction and beam management in mmWave communications are recognized as key challenges toward stable cooperative perception [1].

In recent years, several works have addressed the challenge of blockage predictions, which can be classified into light detection and ranging (LiDAR)-based [2], [3] and RGB camera-based approaches [4], [5]. In the work [2], the authors proposed a blockage prediction model that processes the point cloud generated by LiDAR data to capture the dynamics of communication environments. The work [3] proposed a static cluster removal algorithm for LiDAR data processing and showed that the LiDAR-based blockage prediction approach outperforms the wireless signature approach when the prediction interval is larger than 0.2 s.

Because LiDAR sensor needs a high cost and time-consume process, RGB camera-based approach has been studied to address the issue [5]. In the work [4], the authors proposed a blockage prediction using RGB camera. The proposed model has a deep structure consisting of object detection, bounding box embedding, and recurrent prediction; hence, it consumes high processing time. The work [5] extends the object detection of the aforementioned model, i.e., 3D object detection is used to extract the size and location of vehicles accurately. The work showed that 3D object detection, i.e., depth information, can improve prediction accuracy; however, it still suffers from computation costs.

This paper proposes a novel blockage prediction method to reduce the computation time while achieving high prediction accuracy. Specifically, we employ a stereo camera instead of an RGB monocular camera. The stereo camera can estimate the depth information with few computation costs, while the existing RGB camera-based approach [5] needs a deep neural network for depth information extraction. Also, the stereo camera-based depth estimator has a lower estimation error compared to 3D object detection with an RGB camera.

The major contributions of this paper are listed as follows.

- We introduce a deep neural network structure that is specialized for stereo-aided blockage prediction for mmWave V2X.
- The proposed model is evaluated using a dataset obtained by CARLA simulator [6]. We show that the stereo-aided blockage prediction can outperform the monocular camera-based approach in terms of computation time and prediction accuracy.
- The impact of maximum prediction frames is discussed. We show that there is an optimal setting to maximize the prediction accuracy.

II. Blockage prediction with stereo camera

We consider V2X networks where a roadside unit beside the road transmits a high volume of data to a connected vehicle for driving assistance. The roadside unit has mmWave antenna for data communications and a microwave antenna for dedicated short-range communications (DSRC) [7]. Assuming a heavy traffic road, mmWave communications often lose a connection due to the sudden blockage surrounding vehicles. The loss of connection in the mmWave communication is mainly due to a sudden blockage of the dominant links caused by surrounding vehicle mobility. Accurate estimation of blockage needs prior channels state information (CSI); however, the approach cannot achieve enough accuracy of the current and future states due to the high-speed mobility in V2X scenarios. This paper,

979-8-3503-7099-7/24/$31.00 ©2024 IEEE

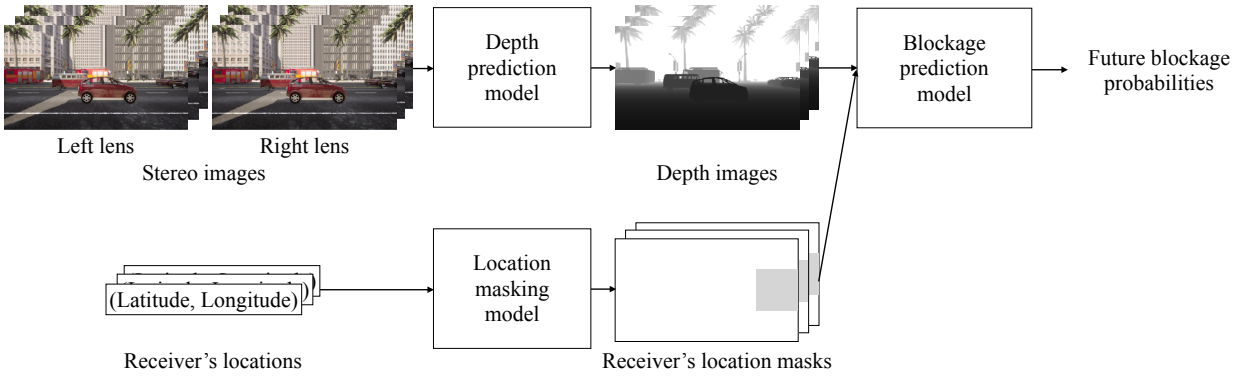


Fig. 1. Proposed blockage prediction system using a stereo camera.

therefore, aims to investigate future blockage prediction using computer vision. Compared to the existing work [5] using a monocular camera, we improve accuracy and prediction speed by introducing a DL-based blockage prediction model using a stereo camera.

Fig. 1 depicts the proposed blockage prediction system. The system needs m time-series images obtained at the roadside unit and GNSS information of the receiver to output blockage probabilities of future n frames. The proposed system consists of three models, i.e., the depth prediction model, location masking model, and blockage prediction model.

A. Depth prediction model

We first construct the m time-series depth images corresponding to the input m stereo images. The depth image is useful for DL-based prediction models because 1) the clear geometry of vehicles can be determined with this data, and 2) vehicles can be easily detected even if the sizes of vehicles are different. Hence, the depth image can improve the robustness of the DL-based prediction model compared to the RGB images.

In the paper, we use a traditional depth prediction model, i.e., stereo block matching (StereoBM) [8], instead of DL-based models [9]. This is because StereoBM is a fast computation speed with enough prediction accuracy. Stereo BM calculates features of blocks based on kernel operation for left and right images and calculates the distance z as follows:

$$z = \frac{f \times b}{d}, \tag{1}$$

where f is the focal length of cameras, d is the distance between points where the same feature exists in two images, and b is the distance between the right and left lens.

B. Location masking model

In addition to the depth image, we employ a location mask that indicates the receiver's location in the image domain. The mask can strengthen the feature of the receiver and its surroundings, which is preferred information for blockage prediction.

The roadside unit constructs the mask using the receiver's GNSS information transmitted from the receiver using the DSRC system. Note that this paper assumes the sensing cycles of the stereo camera and GNSS are the same and ignores the time gap. The size of the mask is the same as the depth image. The masking block (chunks of pixels) is calculated from the relative distance between the roadside unit and receiver and the view angle of the cameras. We set the block size to be 20×20, which is enough size to remain the feature in convolutional processes.

C. Blockage prediction model

We introduce our proposed blockage prediction model as shown in Fig. 2. The model uses m depth images and location masks to predict the blockage probability of future n frames. The model consists of 3D convolutional layers, ConvLSTM layers [10], 2D convolution layers, and linear (feedforward neural network) layers. At first, the 3D convolution layers aim to generate feature maps that indicate the location of objects in each frame. The ConvLSTM layers extract time-series features between the feature maps generated by 3D convolution layers. The 2D convolution layers generate the latent variables that characterize the blockage state. Finally, the linear layers predict the future blockage probability of mmWave communications between the roadside and the receiver. Batch normalization and rectified linear unit (ReLU) are employed for each layer. The sigmoid function is employed to output the prediction values.

The number of prediction frames n can be determined based on the computation time. Specifically, we should set higher prediction frames than the computation time. Therefore, the next section investigates the computation time and impact of n on prediction accuracy.

III. Performance Evaluation

In the section, we evaluate the effectiveness of the stereo-aided blockage prediction, comparing it with the conventional work using a monocular camera [5]. Further, we investigate the impact of prediction frames on accuracy and discuss the optimal number of time frame settings.

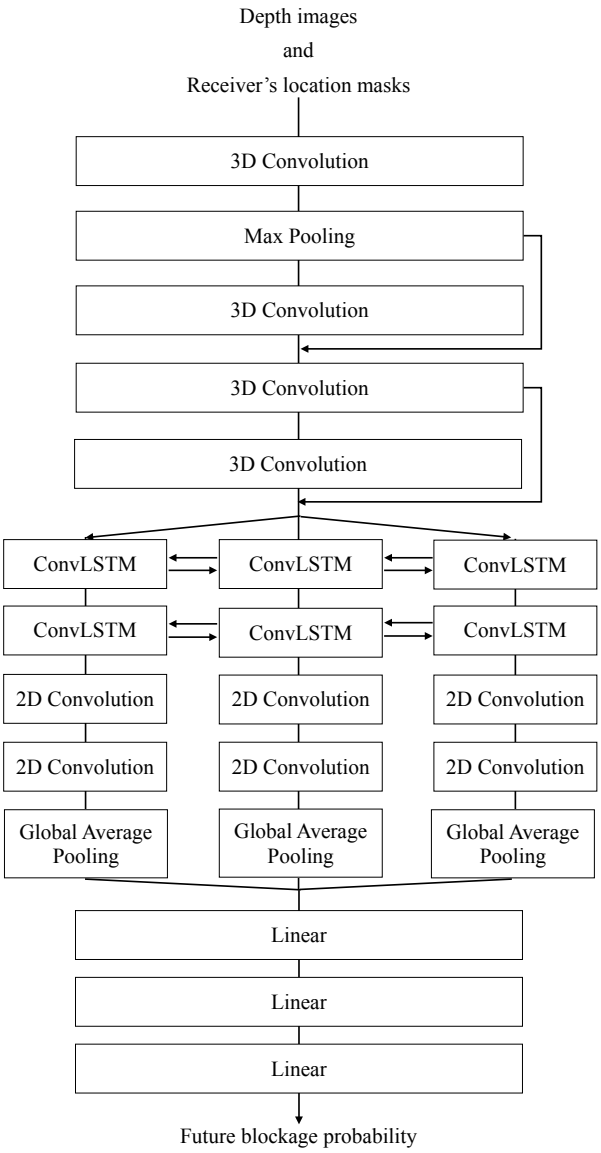


Fig. 2. Blockage prediction model

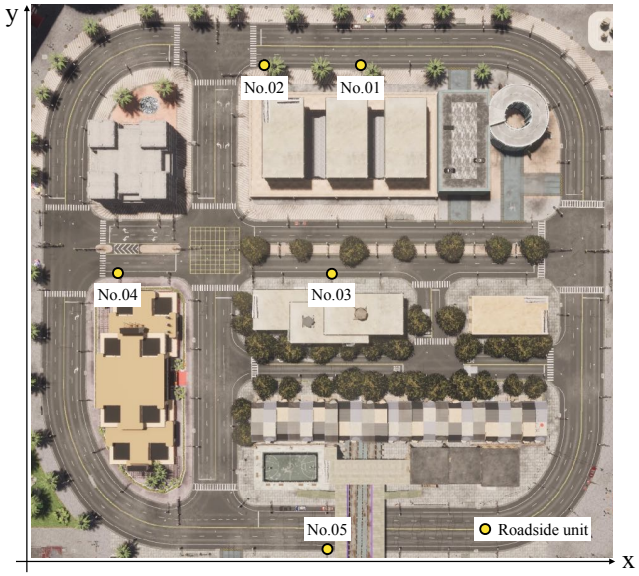


Fig. 3. Bird's-eye view of the envisioned city, where five roadside units deploy beside the roads.

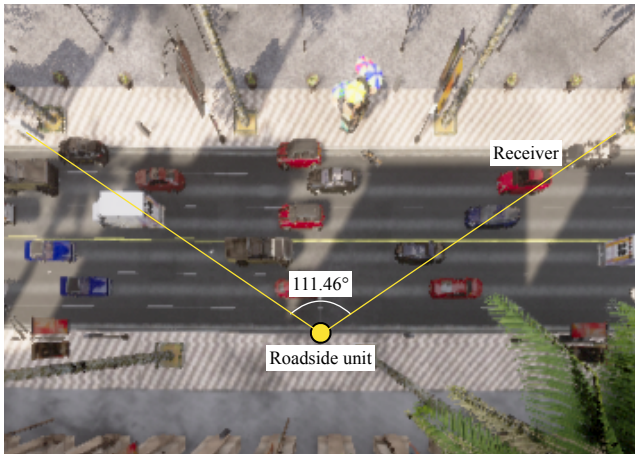


Fig. 4. An example of V2X communications scenario.

A. *Dataset construction*

We simulate mmWave V2X communications in urban areas using CARLA [6] to construct the dataset of blockages. CARLA is an autonomous driving simulator that allows us to freely place vehicles and many kinds of sensors on some maps. Fig. 3 shows the envisioned city and location of roadside units. We simulate based on the [5] which represents an urban area. Specifically, we use a town10 map in CARLA to simulate a realistic urban area. We deploy five roadside units with a mmWave antenna and a stereo camera. Table I summarizes the parameters of CARLA simulations.

As for the parameters of a stereo camera, we refer to the commercial product, i.e., ZMP RoboVision3 [11]. The frame rate is 30 FPS, the resolution is 1920×1080, and the angle of view is $111.46°$. The distance between the left and right lens is 20 cm.

As for wireless communications configurations, the wireless band is 60 GHz. The transmitter is installed next to the stereo camera. Note that we decide whether blockage occurs or not based on visible line of sight without consideration of the Fresnel zone. A single receiver vehicle (Toyota Prius) is deployed into the network while 25 other vehicles (random types other than Toyota Prius) may become a blockage. All vehicles move around with the speed of 30 ± 5 km/h. The received antenna is installed beside of wing mirror; thus, the height is around 0.7 m.

Fig. 4 shows an example of V2X communications scenarios. The roadside unit captures 300 frame images for each simulation. The blockage dataset for each roadside unit is summarised in Table II. The dataset contains 42.2 % blockage data and 57.8 % non-blockage cases.

TABLE I
PARAMETER SETTINGS IN SIMULATION

Parameter	value
Frame rate	30 FPS
Camera angle of view	111.46°
Resolution	1920×1080
lens distance	20 cm
Height of stereo camera	1.5 m
Frequency	60 GHz
Height of transmitter	1.5 m
Moving speed of receiver	30 ± 5 km/h
Height of receiver	0.7 m
Number of receiver vehicles	1
Number of surrounding vehicles	25

TABLE II
BLOCKAGE DATASET GENERATED BY CARLA

	No.1	No.2	No.3	No.4	No.5
Blockages	131	101	161	162	78
Non-blockages	169	199	139	138	222

TABLE III
MODEL PARAMETERS

Parameter	value
Input image size	240×135
Number of input frames m	3
Number of prediction frames n	10
Normalization	$[0, 1]$
Epoch	300
Batch size	128
Learning rate	5×10^{-4}
Loss function	Mean squared error (MSE)
Optimization	Adam

TABLE IV
ACCURACY OF BLOCKAGE PREDICTION

	Accuracy	Precision	Recall
Proposal	77.248 %	47.000 %	83.932 %
Conventional	61.560 %	32.804 %	83.036 %

B. Model training

Table III shows the parameter setting for model training. The number of input frames m is set to 3, and the number of prediction frames n is set to 10. Input frames are reshaped to 240×135 resolution, and each pixel value is normalized to the range of $[0, 1]$. Note that the proposal uses both the left and right images, while the conventional method uses only the left image.

The number of epochs, batch size, and learning rate are set to 300, 128, 5×10^{-4}, respectively. Mean squared error (MSE) is used to calculate the loss function for back proportion. Further, we use Adam with the setting of $\beta_1 = 0.5$, $\beta_2 = 0.999$. We train five models and select one with the best estimation accuracy in the training data as the prediction model in testing. For training, we use the dataset of roadside units of No. 1, 2, 3, and 4. The dataset of roadside unit No. 5 is used for testing. The model training and testing are executed by the computer consisting of AMD Ryzen 7 5800X3D and Nvidia RTX A6000.

In the prediction, we execute binary classification based on the continuous value of the output. We employ a blockage threshold for the binary classification that optimizes precision P and recall R in validation. The precision P and recall R is defined as

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{2}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{3}$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively. We define blockage and non-blockage as positive and negative, respectively. We calculate a harmonic mean of precision and recall to determine the threshold.

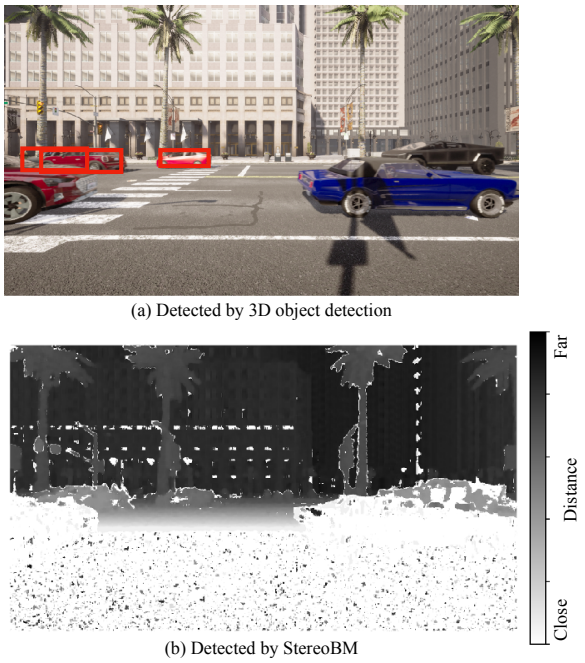
C. Accuracy

Table IV summarizes the accuracy of blockage prediction methods. Accuracy is expressed as $\mathrm{TP} + \mathrm{TN}/(\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN})$. The proposal improves the accuracy by 15.688 %, compared to the conventional method. The performance gain happens due to the improved precision, i.e., a decrease in FP (the probability of the false prediction of the blockage in non-blockage cases). As the conventional method treats all extracted 3D bounding boxes equally for blockage prediction, it tends to predict the blockage based on the density of vehicles, thereby causing the false prediction of the blockage in case of higher density. The proposal can address the issue for the following reasons: i) It employs a location mask to prioritize the feature of vehicles closer to the receiver, ii) The depth information provides the explicit feature on blockage prediction that can reduce the FP.

We can see a tiny improvement in recall, i.e., the probability of the false prediction of non-blockage in blockage cases. The main reason for the decrease in the recall is the extraction error of the blockage feature. Indeed, the conventional method can not detect some vehicles, as shown in Fig. 5(a). As 3D object detection is a complex algorithm, it is difficult to train a model robust to the size of objects. Because the model is trained to fit the size of objects far from the transmitter, the conventional method may cause false predictions due to the error of object detection. Meanwhile, the proposal employs depth images as input features of blockage prediction. The depth images can provide distance from the transmitter and the shape of the vehicles. Because we can recognize the 3D geometry of the V2X communications, the proposal improves the probability of the false prediction of non-blockage. In the paper, the proposal employs a prediction model similar to the conventional method to justify the effectiveness of stereo-vision. We can improve the accuracy by designing an adequate deep neural network model for depth images.

D. Computation time

Table V shows the computation times. The computation time consists of feature construction time and blockage prediction time. The proposal employs depth prediction with stereo vision as feature construction, while the existing method



(a) Detected by 3D object detection

(b) Detected by StereoBM

Fig. 5. Examples of blockage features that are used for blockage predictions.

TABLE V
COMPUTATION TIME

	Total	Feature construction	Blockage prediction
Proposal	19.663 ms	15.012 ms	4.651 ms
Conventional	94.023 ms	89.232 ms	4.791 ms

employs 3D object detection with monocular vision. The proposal significantly reduces the total computation time thanks to lightweight depth prediction, i.e., 15.012 ms. Meanwhile, 3D object detection of the existing method needs a vast computation time, i.e., 89.232 ms, because the neural network architecture should be deep and complex to achieve enough detection accuracy. The proposal and the existing method employ similar neural network architecture for blockage prediction; hence, the computation time is almost the same.

E. Prediction frames analysis

We finally analyze the impact of maximum prediction frames n on the accuracy. We evaluate the accuracy of four models with different n settings, i.e., 5, 10, 15, and 20 frames. As shown in Fig. 6, all models have an optimal point around the middle of n. This is because i) the considered images have temporal correlation due to the mobility of vehicles, and ii) the model fits into the middle of time-series data with the data-driven learning approach. Therefore, we should prepare a model with an adequate n setting based on the cycle of beam alignment and tracking. On the other hand, the setting of lower n can improve the maximum accuracy. Considering the computation time constraints, the proposal can use $n \geq 1$; however, the conventional method should use $n \geq 3$. Thus, the proposal can further improve the accuracy.

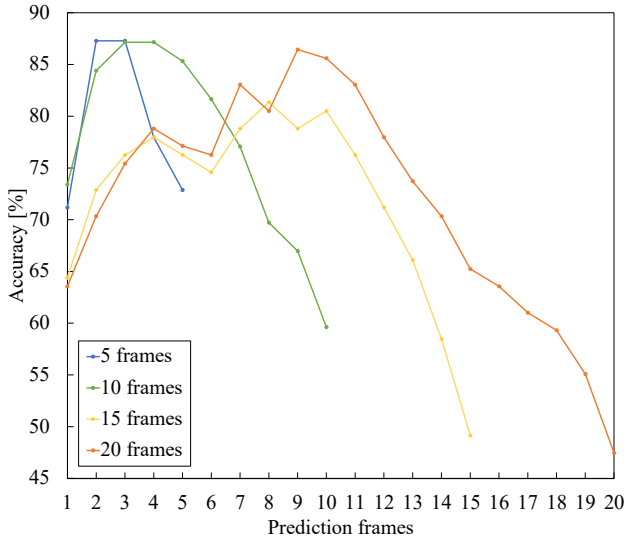


Fig. 6. The impact of maximum prediction frames n on the accuracy at each prediction frame.

IV. CONCLUSION

In this paper, we proposed a stereo-aided blockage prediction model for mmWave V2X communications. By using stereo vision, the proposed model can improve the accuracy by 15.688 %, compared to the existing model using a monocular camera. Further, we demonstrated that the proposal achieves around four times faster than the existing model. The proposal can execute predictions in around 20 ms cycle. To the best of our knowledge, the proposal is the fastest blockage prediction.

REFERENCES

[1] T. Shimizu *et al.*, "Millimeter wave v2x communications: Use cases and design considerations of beam management," in *Proc. APMC 2018*, Kyoto, Japan, Nov. 2018, pp. 183–185.
[2] D. Marasinghe, N. Rajatheva, and M. Latva-aho, "Lidar aided human blockage prediction for 6g," in *Proc. IEEE GC Wkshps 2021*, Madrid, Spain, Dec. 2021, pp. 1–6.
[3] S. Wu, C. Chakrabarti, and A. Alkhateeb, "Lidar-aided mobile blockage prediction in real-world millimeter wave systems," in *Proc. IEEE WCNC 2022*, Austin, TX, USA, Apr. 2022, pp. 2631–2636.
[4] G. Charan, M. Alrabeiah, and A. Alkhateeb, "Vision-aided 6g wireless communications: Blockage prediction and proactive handoff," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10 193–10 208, 2021.
[5] W. Xu *et al.*, "Computer vision aided mmwave beam alignment in v2x communications," *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, pp. 2699–2714, 2023.
[6] A. Dosovitskiy *et al.*, "CARLA: An open urban driving simulator," in *Proc. CoRL 2017*, Mountain View, CA, USA, Nov. 2017, pp. 1–16.
[7] J. B. Kenney, "Dedicated short-range communications (dsrc) standards in the united states," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1162–1182, 2011.
[8] OpenCV, *cv::StereoBM Class Reference*, 2023. [Online]. Available: https://docs.opencv.org/3.4/d9/dba/classcv_1_1StereoBM.html
[9] N. Mayer *et al.*, "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation," in *Proc. IEEE CVPR 2016*, Las Vegas, NV, USA, Jun. 2016, pp. 4040–4048.
[10] X. SHI *et al.*, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Proc. NIPS 2015*, vol. 28, Montreal, Quebec, Canada, Dec. 2015.
[11] Z. INC., *Stereo camera*, 2021. [Online]. Available: https://www.zmp.co.jp/en/products/sensor/robovision/robovision3