Online Human Proximity Detection Based on the Properties of Equirectangular Transformation of Spherical Image Data

Y. Shimbo
dpt. of Electronic Engineering
School of Engineering
Shibaura Institute of Technology
Tokyo 135-8548, Japan
ag21080@shibaura-it.ac.jp

H.A.H.Y. Sarathchandra*, A.M.N.N. Senevirathne, Chinthaka Premachandra
Graduate School of Engineering and Science
Shibaura Institute of Technology
Tokyo 135-8548, Japan
nb24504@shibaura-it.ac.jp, ma23501@shibaura-it.ac.jp, chintaka@shibaura-it.ac.jp

***Abstract*—Electric wheelchairs are categorised as pedestrians, which reclassifies accidents involving them as pedestrian-on-pedestrian incidents rather than traffic accidents, making it difficult for the problem to come to the forefront. Conventional proximity detection systems primarily measure distances in real time or merely identify nearby objects, leaving a gap in practical systems that focus on detecting the proximity of people. Additionally, it is more challenging to differentiate between dangerous situations, where a person or object approaches a wheelchair, and non dangerous situations, where a person or object merely crosses in front of it. In this study, we propose a novel method that can distinguish between approaching and crossing scenarios by utilizing images captured by an omnidirectional camera that encompasses all directions, along with accessing the distance from the camera to the object using the characteristics of the equidistant cylindrical view. By integrating images from the omnidirectional camera and object tracking, the proposed method can detect approaching objects from every direction in real time and assess their danger level. The performance evaluation of the proposed method has shown that it can achieve highly accurate detection of approaching objects, with an accuracy rate exceeding 95% at an average processing frame rate of more than 15 FPS, which is a satisfactory outcome for proximity detection.**

***Index Terms*—Equirectanguler images, object tracking, proximity detection, spherical data.**

I. Introduction

Assistive devices for locomotion and manipulation are increasingly becoming essential to maintain higher living standards. The growth of the global elderly population and differently abled communities intensifies this necessity. However, the current state-of-the-art developments are insufficient to adequately meet these needs, which drives us to further explore advancements in assistive robots. Electric wheelchair systems are well-known assistive devices that primarily aid mobility [1]. This work particularly focuses on enhancing electric wheelchair navigation using 360-degree vision sensors.

Literature indicates that wheelchair navigation is primarily controlled by several different modalities. However, most of the time, users must make high-level decisions and operate the electric wheelchair, which leads to quick fatigue and a significant cognitive load. For instance, these systems are operated using joysticks [2], gestures [3], voice commands [4], brain signals [5], etc., all of which require active user intervention to function. Autonomous wheelchair systems [6] can greatly reduce these challenges and provide a more comfortable life for users. However current knowledge lacks a detailed understanding of the surrounding environment, which results in poor decisions during autonomous driving.

The perception is crucial for making accurate decisions. In most electric wheelchair systems, a distinguished camera module serves as the primary vision sensor [7]. Perception can be further enhanced by integrating multiple cameras to observe the surrounding environment, surpassing human limitations [8], [9]. However, this leads to introducing additional computational overhead, and the system tends to become more complex due to the need for multiple processing modules and precise calibrations. To mitigate these specific issues, 360-degree cameras can be employed in these wheelchair systems. Nevertheless, this may present further challenges due to a lack of knowledge regarding 360-degree vision processing algorithms and techniques [10].

The data collected from these sensors assist the central system in making appropriate decisions, such as obstacle avoidance in autonomous navigation [11]. A major consideration here is that although there are other sensors like LIDAR for collision avoidance purposes, detecting human proximity within the entire 360 view of the wheelchair presents challenges [12], [13]. Even in most contributions, it is observed that extracting distance takes precedence over identifying whether a subject is on a collision course [14]. In addition, this data is integrated with various state-of-the-art mechanisms to enhance the systems' intelligence and illustrate the possibility of detecting relevant objects [15]. The YOLO model can be considered a well-established framework that can be utilised in such scenarios and is also suitable for real-time feature detections [16]–[18]. However, the simultaneous identification of obstacles and proximity detection has not been thoroughly discussed in previous efforts for omnidirectional viewing.

Thus, this effort illustrates the applicability of an spherical

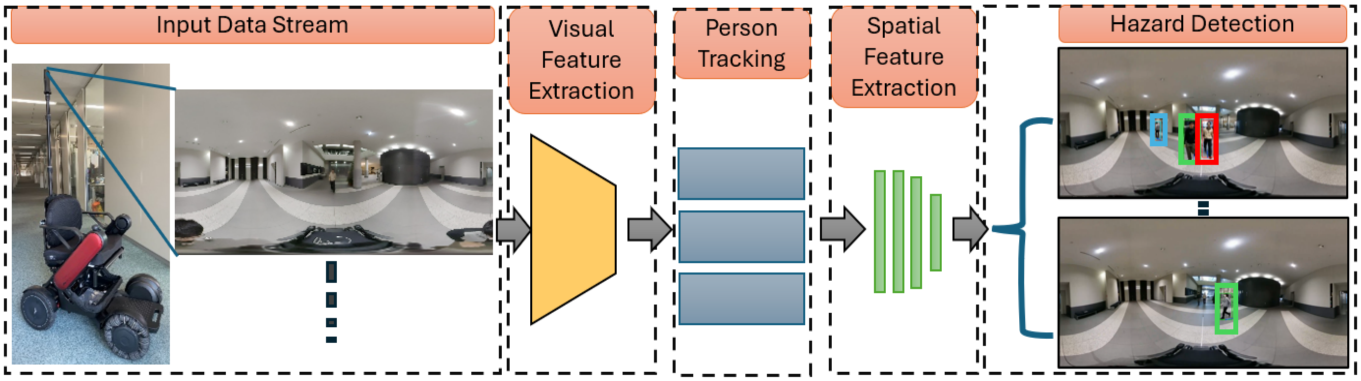

Fig. 1: To identify pedestrians approaching, an omnidirectional camera is mounted on the wheelchair, and the image stream is then transformed into an equirectangular format. At the first stage, YOLO11 is utilized to extract visual features, while ByteTrack serves as a tracking mechanism in the second stage. Finally, spherical features are extracted to determine whether the subject is in danger or not. The system's outputs are displayed using boxes: a red box represents the subject in danger, while the others(blue/green) indicate those not in danger at a given time.

camera data for achieving seamless autonomous navigation of electric wheelchairs while employing a YOLO11 model [19] for detecting subjects on a collision course with the wheelchair and avoiding potential hazards. Furthermore, this combination enhances the perception of these autonomous assistive devices by integrating conventional object-tracking mechanisms and advancing the systems to a higher level of intelligence. The remainder of the article is organised as follows, Section II provides a detailed overview of the proposed system architecture, Section III and IV presents the experiments and results. Section V make some discussions. Finally, Section VI concludes the article with future extensions.

II. System Architecture

Following chapter discuss the method proposed for the understanding surrounding peoples' proximity relative to a electric wheelchair system. The overall architecture of the proposed system is shown in Figure 1. As shown, overall pipeline composed of three major components.

A. Pedestrian Detection with Spherical Images

An omnidirectional camera was mounted directly above the wheelchair (Figure 1) and captures the surrounding environment. This facilitates capturing the entire environment around the wheelchair and makes the decision made by the system more accurate and reliable.

For "human" proximity detection, we created a dataset that can only detect people. Using a Ricoh THETA Z1[1] omnidirectional camera, we recorded videos of a sidewalk in 2K mode and produced approximately 600 images, with an average of about 4 people per frame, by cropping frames at regular intervals from the video. Moreover, all the videos were captured at 30 FPS, and the image size was set to 1920 x 960. The images were then annotated with labelImg2[2] to create a dataset for training and a dataset for validation at a ratio of 3:1. YOLO11 is equipped with several inference models by default, and "YOLO11n", which is the lightest model among them, was used as a weight for transition training. The results of iterative training under various conditions showed that the best accuracy was obtained when the number of training sessions was 100 and the batch size was 16.

B. Pedestrian Tracking

As a prerequisite, the proposed system must be able to track objects in real time without any problems to work practically. "ByteTrack" [20] is one of the successor models of "SORT" [21], which is the MOT method that pioneered the tracking-by-detection type. SORT was also a pioneer in the MOT field, and the model had problems with real-time performance [22]–[24] because it enabled tracking by combining both location and appearance information for batch processing. SORT computes the prediction of the current frame in the Kalman filter [18], calculates the Intersection over Union (IoU) between the predicted detection box and the actual box and creates a cost matrix. Finally, it calculates the similarity by applying the Hungarian method [19] to the created matrix and identifies whether the previous frame and the current frame are identical or not. ByteTrack has become a practical object tracking system by performing this process in two steps without changing the fundamental principle of "computing similarity using the Kalman filter and the Hungarian method".

C. Spatial Feature Extraction

As mentioned, extracting spherical features from an omnidirectional image is challenging. However, we proposed the following simple method to extract features. Initially, spherical data were transformed into equirectangular images (Figure 1) and the spatial features from the transformed version.

The camera was calibrated to determine the subjects within the threshold distance(d), which is then recognized as the danger zone and identifies the corresponding coordinates concerning an equirectangular image. The following figure 2 shows the corresponding coordinates relative to the wheelchair position.

[1]https://thetaz1.com/ja/
[2]https://github.com/HumanSignal/labelImg

Fig. 2: Equirectangular image transformed from an omnidirectional capture and points(blue dots inside red circles) representing calibrated threshold line (3 ± 0.1m from the wheelchair) which utilized to detect the danger zone.

We monitor the interested subject's vertical motion relative to the base frame of the wheelchair's omnidirectional camera. To extract whether the interested subject is moving towards the wheelchair we monitor the sine value of the angle made by the line connecting two adjacent tracking frames with the threshold line which is identified with the calibration (Figure 3). Whenever the subject is identified within the danger zone the algorithm identifies whether the subject is moving towards or not.

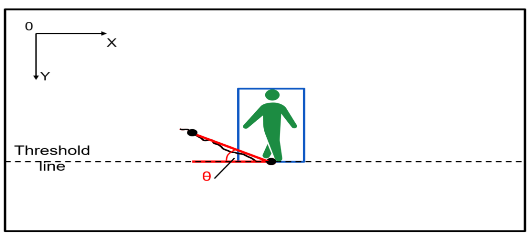

Fig. 3: Proximity detection with threshold line. When the person is below the threshold level both are in danger zone.

III. Experiments

As the system proposed, all the experiments were carried out for the spherical data captured with RICHO360 camera. Primarily, we utilize YOLO11 for pedestrian detection and extend the detection to track the pedestrians within the frame with ByteTracker [20]. Initially, the accuracy of the pedestrian detection was evaluated with precision against the recall plot (PR curve) for the custom dataset (Figure 4). It shows that the fit rate remains at 0.9 even when the reproduction rate is as high as 0.9, indicating that the system is capable enough to accurately detect as many people as possible.

The threshold of the proximity for determining danger zone was determined by capturing the images in all four major directions (left, right, front and back) relative to the wheelchair of a person standing at a distance(d) of 3 ± 0.1m by the camera which is fixed to the back of the wheelchair at a height of 1.5 m. Furthermore, set the angle $\theta = 45^\circ$ as the threshold value for determining whether a person is on a collision course or not (Figure 4). Then the overall system was transferred and tested with a Jetson Orin Nano Developer kit[3] which is attached to the electric wheelchair system.

[3]https://www.nvidia.com

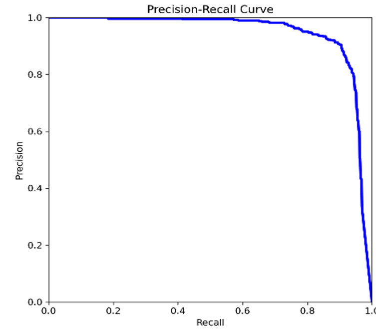

Fig. 4: PR curve for custom-made inference model.

Since it proposed an MOT-based object tracking, experiments were conducted for a few different scenarios. In the first instance, a single pedestrian is approaching others who are passing by without heading toward the wheelchair. Then, the number of people who approached directly towards the wheelchair was increased up to a maximum of three. Since the maximum number of pedestrians who can approach a wheelchair is about three at most, we decided to experiment with no more than three persons approaching the wheelchair. The figure 6 shows four specific cases where the experiments were conducted.

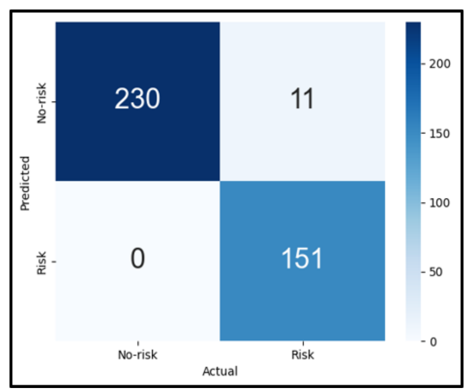

Fig. 5: The results for cases where up to three individuals either approaching towards or passing nearby.

During inference, a subject passing nearby without heading towards the wheelchair (safe) judgment was classified as positive, and an approaching towards wheelchair (risk) judgment was classified as negative. Based on this classification, a confusion matrix was constructed and evaluated. In addition, recall, precision and accuracy are calculated to evaluate the proposed mechanism for detecting pedestrians who are coming towards a wheelchair or passing nearby but heading away.

The processing speed of the proposed architecture was examined to assess the model's feasibility. Additionally, for the experiments, the trained model developed in Section II for pedestrian tracking was converted to FP16 format using "TensorRT", which numerous studies [25], [26] have shown to optimize models and significantly enhance inference speed. However, this conversion results in a slight drop in accuracy.

(a) Case I (b) Case I (c) Case II (d) Case II

Fig. 6: The judgment process for two scenarios involving multiple individuals approaching and passing nearby. (a) When two out of three individuals are approaching towards(blue) one passing nearby(green) (b) When two out of three individuals are approaching towards and in danger(red), (c) When two out of three individuals are passing(green) and one approaching(blue), (d) When two out of three individuals are passing(blue/green) one approaching toward and in danger(red).

IV. Results

The videos captured in Section III were inferred using the proposed system, and the resulting confusion matrix is shown in Figure 5. Recall, precision, and accuracy values derived from figure are presented in Table I. From these results, it is evident that all metrics achieved high accuracy, exceeding 95%. Furthermore, it reveals that misdetections occurred only when the ground truth was "approaching towards" in other words, misdetections happened exclusively in cases where the system incorrectly judged "approaching towards" as "passing nearby".

TABLE I: Metrics calculated for proposed architecture.

Recall	**Precision**	**Accuracy**
100%	95.4%	97.2%

The results of the processing time measurements are presented in Table II. In this analysis, processing times were measured for two scenarios involving three individuals moving: first case, when two of the three individuals approaching towards and second case, when two of them passing nearby. Examples of inference results for these two scenarios are shown in Figure 6.

TABLE II: Processing time for scenarios with single subject or three subjects moving in the scene in seconds.

	Single subject[s]	**Three subjects[s]**	
		Two passing	**Two approaching**
Average	0.064	0.066	0.065
Max	0.070	0.073	0.071
Min	0.060	0.063	0.061
FPS	15.74	15.20	15.49

From Table II, it can be observed that a processing frame rate of at least 15 FPS was achieved in all scenarios. When comparing the patterns, the FPS was higher when one individual was moving compared to when three individuals were moving. Furthermore, within the three-person scenarios, the FPS was lower in the case where two individuals were passing nearby without approaching.

V. Discussions

As shown with the evaluation metrics, most of the time, the proposed architecture brought better results with spherical image data, except for very few misjudgments. However, human natural sway brings a few misleading judgments with the proposed algorithm. Humans naturally sway their center of gravity from side to side while walking [27], especially at the moments when they switch their legs. As a result, particularly for people taking large steps, the bounding box is likely to move side to side with each step, which misleads to detection as a horizontal movement. To avoid this issue, a few frames can be considered, and then the trajectory can be plotted with mean values. This approach helps to minimize unnecessary sway by smoothing the data over time.

Furthermore, as shown in Table II, the proposed system demonstrate high performance in terms of online processing. When three individuals are moving simultaneously, processing time is higher than when more people are passing nearby. This can be attributed to the fact that lateral movement when passing creates more significant displacement in an equirectangular projection map compared to frontal approaches. When considering the movement of objects between adjacent frames, frontal approaches can be immediately recognized as the same object, whereas lateral movements result in larger displacements per frame, making identification more time-consuming. However, when looking at the overall system performance, the FPS does not drop significantly depending on the situation, suggesting that it is not a critical issue.

VI. Conclusion

This work proposed a online proximity detection system that can prevent collisions between electric wheelchair users and pedestrians moving nearby. Instead of relying on real-time distance measurement methods using traditional sensors, the proposed system combines equirectangular projection images obtained from an omnidirectional camera and this approach enables the differentiation between objects approaching the wheelchair and those simply passing nearby but not approaching, something that is challenging with conventional methods. In addition, a custom data set was created to train the pedestrian detection model and test with the overall architecture deployed on a electric wheelchair, where all the metrices exceed 95%. In future, further investigation of spherical data for navigation and scene understanding with assistive robots, including intelligent wheelchair systems, can be carried out to bring better user satisfaction.

REFERENCES

[1] O. Atoyebi, A. Wister, J. Mattie, J. Beadle, G. Gutman, H. Chaudhury, C. J. Sparrey, O. Y. Jones, W. B. Mortenson, E. O'Dea *et al.*, "Power assist add-ons for adult manual wheelchair users: A scoping review," *Assistive Technology*, pp. 1–12, 2025.

[2] E. Jang, S.-H. Eom, S. Park, S. Kim, and E. H. Lee, "A study on joystick-based collision avoidance methods for improving the operational convenience of intelligent wheelchairs," in *2025 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2025, pp. 1–6.

[3] R. B. M. Othman and E. M. Yuniarno, "Smart wheelchair control based on spatial features of hand gesture," in *2024 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*. IEEE, 2024, pp. 1–6.

[4] T. Williams and M. Scheutz, "The state-of-the-art in autonomous wheelchairs controlled through natural language: A survey," *Robotics and Autonomous Systems*, vol. 96, pp. 171–183, 2017.

[5] M. Y. Naser and S. Bhattacharya, "Towards practical bci-driven wheelchairs: a systematic review study," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1030–1044, 2023.

[6] S. A. Kumar, J. Vanualailai, and A. Prasad, "Assistive technology: autonomous wheelchair in obstacle-ridden environment," *PeerJ Computer Science*, vol. 7, p. e725, 2021.

[7] J. G. Torres-Vega, J. C. Cuevas-Tello, C. Puente, J. Nunez-Varela, and C. Soubervielle-Montalvo, "Towards an intelligent electric wheelchair: Computer vision module," in *Intelligent Sustainable Systems: Selected Papers of WorldS4 2022, Volume 1*. Springer, 2023, pp. 253–261.

[8] Y. Yang, D. Tang, D. Wang, W. Song, J. Wang, and M. Fu, "Multi-camera visual slam for off-road navigation," *Robotics and Autonomous Systems*, vol. 128, p. 103505, 2020.

[9] Y. Wang, Y. Ng, I. Sa, A. Parra, C. Rodriguez-Opazo, T. Lin, and H. Li, "Mavis: Multi-camera augmented visual-inertial slam using se 2 (3) based exact imu pre-integration," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 1694–1700.

[10] M. Xu, C. Li, S. Zhang, and P. Le Callet, "State-of-the-art in 360 video/image processing: Perception, assessment and compression," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 5–26, 2020.

[11] F. Pacini, P. Dini, and L. Fanucci, "Design of an assisted driving system for obstacle avoidance based on reinforcement learning applied to electrified wheelchairs," *Electronics*, vol. 13, no. 8, p. 1507, 2024.

[12] J. Cui, S. Yu, Y. Shang, Y. Dai, and W. Zhang, "Research on outdoor navigation of intelligent wheelchair based on a novel layered cost map," in *Actuators*, vol. 14, no. 2. MDPI, 2025, p. 46.

[13] S. Jeyabal, W. Sachinthana, S. Bhagya, P. Samarakoon, M. R. Elara, and B. J. Sheu, "Hard-to-detect obstacle mapping by fusing lidar and depth camera," *IEEE Sensors Journal*, 2024.

[14] Y. Liu, L. Zhang, J. Liu, Y. She, B. Zhang, and F. Min, "Dynamic window approach using goa search algorithm for collision avoidance based on obstacle trajectory prediction," *IEEE Access*, 2025.

[15] T.-W. Wang, H.-P. Huang, and Y.-L. Zhao, "Vision-guided autonomous robot navigation in realistic 3d dynamic scenarios," *Applied Sciences*, vol. 15, no. 5, p. 2323, 2025.

[16] L. Qin, Y. Shi, Y. He, J. Zhang, X. Zhang, Y. Li, T. Deng, and H. Yan, "Id-yolo: Real-time salient object detection based on the driver's fixation region," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 15 898–15 908, 2022.

[17] R. Srinivasan, D. Rajeswari, A. Arivarasi, and A. Govindasamy, "Real-time vehicle classification and license plate recognition via deformable convolution based yolo v8 network," *IEEE Sensors Journal*, 2024.

[18] W. Xia, P. Li, H. Huang, Q. Li, T. Yang, and Z. Li, "Ttd-yolo: A real-time traffic target detection algorithm based on yolov5," *IEEE Access*, 2024.

[19] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," *arXiv preprint arXiv:2410.17725*, 2024.

[20] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "Bytetrack: Multi-object tracking by associating every detection box," in *European conference on computer vision*. Springer, 2022, pp. 1–21.

[21] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE international conference on image processing (ICIP)*. Ieee, 2016, pp. 3464–3468.

[22] Y. Du, Z. Zhao, Y. Song, Y. Zhao, F. Su, T. Gong, and H. Meng, "Strongsort: Make deepsort great again," *IEEE Transactions on Multimedia*, vol. 25, pp. 8725–8737, 2023.

[23] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE international conference on image processing (ICIP)*. IEEE, 2017, pp. 3645–3649.

[24] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "Bot-sort: Robust associations multi-pedestrian tracking," *arXiv preprint arXiv:2206.14651*, 2022.

[25] S. Manzoor, E.-J. Kim, S.-H. Joo, S.-H. Bae, G.-G. In, K.-J. Joo, J.-H. Choi, and T.-Y. Kuc, "Edge deployment framework of guardbot for optimized face mask recognition with real-time inference using deep learning," *Ieee Access*, vol. 10, pp. 77 898–77 921, 2022.

[26] M. Liu, Z. Li, Y. Li, and Y. Liu, "A fast and accurate method of power line intelligent inspection based on edge computing," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[27] R. C. Woledge, D. B. Birtles, and D. J. Newham, "The variable component of lateral body sway during walking in young and older humans," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 60, no. 11, pp. 1463–1468, 2005.