Advanced Pedestrian Distance Estimation for ADAS with Canny Edge Detection and Stereo Vision

Oumayma Rachidi[1*], *Chafik* Ed-dahmani[1], and *Badr* Bououlid Idrissi[1]

[1]Electromechanical Engineering Department, ENSAM of Meknes, Moulay Ismail University, Morocco

Abstract. Pedestrian detection is a vital aspect of Advanced Driver Assistance Systems (ADAS), crucial for ensuring driving safety and minimizing collision risks. While detecting pedestrians is important, it must be paired with precise distance estimation to create a robust safety solution. Stereovision cameras are well-regarded for their effectiveness and affordability in measuring depth through disparity between two images. Despite this, research on pedestrian distance estimation using only stereovision remains sparse, with many studies relying on computationally heavy dense depth maps. This paper proposes an innovative method for computing object-level disparity specifically for pedestrian detection using stereo cameras. The approach integrates Canny edge detection with ORB (Oriented FAST and Rotated BRIEF) feature matching to efficiently identify and track keypoints within pedestrian bounding boxes. This method not only improves the accuracy of distance estimation but also reduces computational demands, making it suitable for real-time applications. The approach was thoroughly tested on a Raspberry Pi 4, a resource-constrained device, and achieved promising results, demonstrating its potential for practical use in ADAS.

Keywords: Stereovision, distance estimation, pedestrian detection, Canny edge detection, ORB

1 Introduction

Pedestrian detection is a fundamental element of Advanced Driver Assistance Systems (ADAS), significantly contributing to the improvement of road safety. With pedestrian-related accidents on the rise, the demand for reliable and effective detection systems is more pressing than ever. ADAS technologies are designed to aid drivers by providing timely warnings and, in some cases, by autonomously intervening to prevent accidents. To maximize the effectiveness of these systems, it is crucial not only to accurately detect pedestrians but also to precisely estimate their distance from the vehicle. This dual

* Corresponding author: oum.rachidi@edu.umi.ac.ma

functionality is key to enabling swift decision-making and appropriate responses to potential dangers. Modern ADAS platforms incorporate a variety of technologies, such as RADAR (Radio Detection and Ranging) [1], LIDAR (Light Detection and Ranging) [2], monocular cameras [3], and stereovision cameras [4], to achieve these goals. Among these, stereovision cameras stand out as an especially effective solution. They offer both accurate depth perception and detailed visual information, making them a compelling choice for pedestrian detection. Unlike RADAR, which may have difficulty with object classification, or LIDAR, which can be costly and sensitive to environmental factors, stereovision strikes an optimal balance. It provides precise distance measurements while being cost-effective and less susceptible to adverse weather conditions, making it particularly well-suited for real-time pedestrian detection in a wide range of driving environments.

The use of stereovision in pedestrian detection has not been as widely adopted as other sensor technologies. Most recent research has concentrated on utilizing RGB-D data to enhance pedestrian detection by generating 2D or 3D bounding boxes, which are particularly useful for managing occlusions [5,6]. Typically, these studies integrate depth maps with neural networks to improve detection accuracy. For example, T. Ophoff et al. [7] investigated the incorporation of depth maps to enhance the performance of the YOLOv2 model, particularly in detecting occluded pedestrians. Their findings revealed that mid-level fusion of data produced superior results compared to early or late fusion approaches, and their method was tested using an RGB-D pedestrian dataset obtained with a Kinect camera. Similarly, J. Zhang et al. [8] employed PSMNet to generate disparity maps, which were subsequently used to merge left and right images, thereby improving the pedestrian detection accuracy of the Faster-RCNN model.

Despite the advancements in pedestrian detection, research specifically aimed at estimating the distance to pedestrians using stereovision remains relatively scarce. Most existing approaches tend to focus on image-level disparity rather than object-level disparity, which can lead to increased computational demands [9]. Additionally, these studies have largely concentrated on enhancing object detection networks, often neglecting the unique challenges that arise in depth estimation within stereovision [10]. While optimizing detection networks can mitigate issues such as missed detections and false positives, it does not resolve fundamental challenges like feature point misalignment or inaccuracies in depth measurement between binocular images. Therefore, focusing solely on network optimization falls short of addressing the full spectrum of complexities associated with employing stereovision for pedestrian detection.

To overcome the aforementioned challenges, we introduce an object-level algorithm for pedestrian depth estimation that utilizes canny edge detection for extracting keypoints, paired with ORB (Oriented FAST and Rotated BRIEF) [11] for matching to compute disparity using stereo cameras. This algorithm is designed to be both efficient and lightweight, making it suitable for real-time applications, as confirmed by our testing.

The structure of this paper is as follows: Section 2 elaborates on the proposed methodology, Section 3 presents the experimental results, and Section 4 concludes the discussion with insights and future directions.

2 Proposed method

The proposed algorithm is structured around two core elements: depth computation based on the triangulation method inherent to stereovision, and object-level disparity estimation, which is facilitated through keypoint extraction using edge detection and feature matching with ORB. Disparity estimation is specifically conducted within bounding boxes that are generated by a custom YOLOv5s model. The general framework of our approach is outlined as follows:

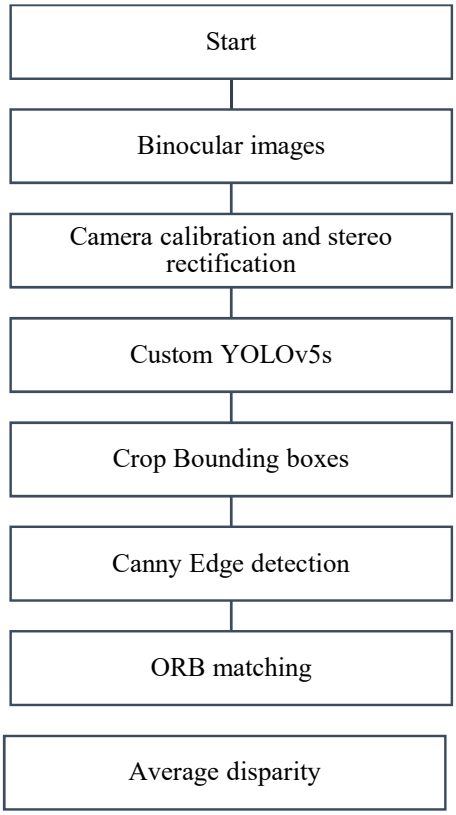

Fig. 1. Overall framework of our approach.

2.1 Stereovision Principle for depth estimation

Depth estimation using stereovision is based on the concept of triangulation, where two or more cameras, placed at different angles, capture images of the same scene. This setup produces two images from slightly different perspectives, leading to small variations known as disparities. By matching corresponding points in these images, the system calculates the disparity, which represents the positional difference of these points on the image planes. Disparity is inversely related to the object's distance from the cameras—objects closer to the cameras have higher disparity, while those farther away show lower disparity. With the known separation between the cameras (baseline) and the focal length of the lenses, the depth of each point can be determined through triangulation. Figure 2 [12] depicts a stereo vision system configuration where two cameras—left and right—are placed at a fixed baseline distance, B, similar to the spacing of human eyes in binocular vision. The distance, D, to the target object is calculated using Equation (1), which incorporates the focal length of the cameras, f, and the disparity, d. The disparity, d, is calculated by determining the difference in the object's position between the corresponding points in the left and right images.

$$D = {}^{B.f}/_{d} \tag{1}$$

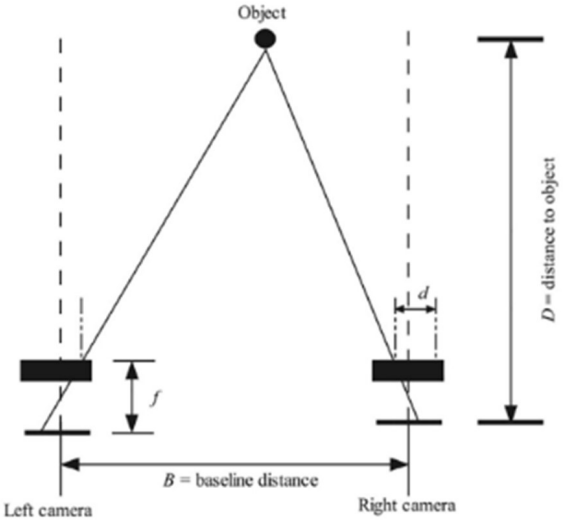

Fig. 2. Stereovision principle for depth estimation.

2.2 Depth estimation method

Most depth estimation methods using stereovision focus on creating a depth map for the entire image. Traditional techniques often rely on computationally intensive local matching scores, such as Sum of Squared Differences (SSD) [13], Sum of Absolute Differences (SAD) [14], and Normalized Cross-Correlation (NCC) [15]. Recent advancements in this field involve end-to-end deep learning models specifically designed for depth estimation. While these newer approaches offer improved accuracy, they are computationally demanding and require a detailed understanding of the entire scene, which can hinder their efficiency in real-time applications. To address these limitations, our method introduces an object-level disparity estimation approach that calculates disparity by analysing the differences between ORB matches, derived from keypoints identified through edge detection. The proposed method offers several key advantages over existing pedestrian distance estimation techniques. Firstly, it employs object-level disparity estimation, allowing for more accurate distance measurements by focusing on individual pedestrians rather than relying on a global depth map. Secondly, our approach utilizes ORB feature matching, which is computationally efficient compared to traditional methods like SSD and SAD, making it well-suited for real-time applications.

2.2.1 Canny edge detection

When using ORB keypoints within bounding box regions, relying exclusively on ORB can be insufficient, especially as the distance from the camera grows, which significantly decreases the number of detected keypoints. To overcome this challenge, an edge detection technique is utilized to focus on the most relevant keypoints within the bounding boxes. By emphasizing the most pertinent edges, this technique helps identify critical structural features that are less affected by distance-related scaling issues. The edges provide valuable information about the shape and contours of objects, allowing for the prioritization of keypoints that contribute meaningfully to depth estimation. This combined approach ensures that even at varying distances, a sufficient number of reliable keypoints are maintained, ultimately improving the accuracy and robustness of the method in diverse scenarios. A

notable and widely used edge detection algorithm is the Canny edge detector [16]. The Canny algorithm is renowned for its effectiveness in identifying edges by employing a multi-stage process that includes noise reduction, gradient calculation, non-maximum suppression, and edge tracking through hysteresis. This method enhances the accuracy of keypoint detection within bounding boxes, as it effectively highlights significant edges that ORB might miss, particularly at greater distances. By integrating edge detection with ORB, our approach aims to maintain a robust set of keypoints, improving the reliability of disparity estimation even in challenging conditions. The overall flow chart of this Canny Edge detection algorithm is presented in Figure 3 [17].

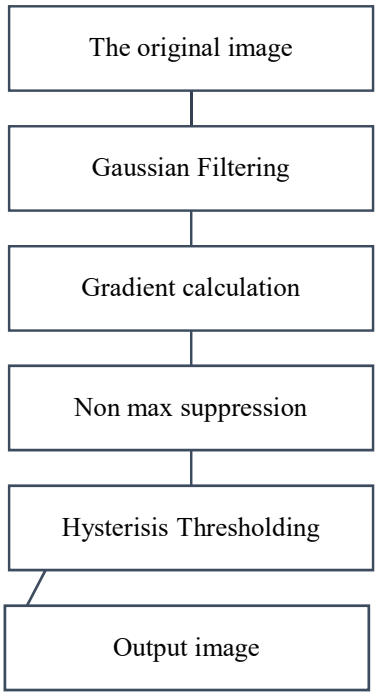

Fig. 3. Canny edge detection Flowchart.

Initially, Gaussian smoothing is applied to the image to reduce noise, using a Gaussian filter described by the formula:

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{2}$$

where σrepresents the Gaussian filter parameter. A smaller σ value increases edge detection accuracy but reduces the signal-to-noise ratio, while a larger σ improves the signal-to-noise ratio but decreases accuracy. Thus, selecting an appropriate σ is crucial for effective edge detection.

In the next step, the gradient magnitude and direction of the smoothed image are computed. This involves calculating the partial derivatives of the image in the x and y directions using:

$$A_x(i,j) = \frac{I(i,j+1) - I(i,j) + I(i+1,j+1) - I(i+1,j)}{2} \tag{3}$$

$$A_y(i,j) = \frac{I(i,j) - I(i+1,j) + I(i,j+1) - I(i+1,j+1)}{2} \quad (4)$$

From these, the gradient magnitude A(i,j) and direction $\theta(i,j)$ are computed as:

$$A(i,j) = \sqrt{A_x^2(i,j) + A_y^2(i,j)} \quad (5)$$

$$\theta(i,j) = \arctan\left(\frac{A_y(i,j)}{A_x(i,j)}\right) \quad (6)$$

The subsequent step involves non-maximum suppression, which refines edge detection by retaining only the local maximum gradient points. The algorithm compares the gradient value of the target pixel A(i,j) with those of its neighbours along the gradient direction. If the gradient magnitude of the target pixel exceeds those of the neighbouring pixels on either side of the gradient direction, the pixel is preserved; otherwise, it is set to zero.

Finally, double-thresholding and edge connectivity are applied to further refine the edge map. High (T_h) and low (T_l), where $T_l = 0.6\,T_h$ thresholds are established. Pixels with gradient magnitudes below the low threshold are classified as non-edges, while those above the high threshold are identified as edges. For gradient magnitudes between the two thresholds, the pixel is evaluated based on its connectivity to other edge pixels in its 8-neighborhood. If connected to edge pixels, it is classified as an edge point; otherwise, it is discarded. This process ensures the accuracy and continuity of the detected edges, enhancing the overall quality of edge detection. The results of this algorithm on a real time frame are presented on Figure 4.

2.2.2 Keypoints matching

In this research, the ORB (Oriented FAST and Rotated BRIEF) algorithm is employed to match keypoints identified within bounding box regions using edge detection techniques. ORB integrates the BRIEF (Binary Robust Independent Elementary Features) descriptor, which encodes local image patches around each keypoint into concise binary strings, providing a computationally efficient and compact representation.

The BRIEF descriptor generates binary strings by performing a series of binary tests on pairs of pixel intensities within a local image patch. The binary test is defined by:

$$b_k = \begin{cases} 1 & if\ I(p_i) > I(p_j) \\ 0 & otherwise \end{cases} \quad (7)$$

Where $I(p_i)$ and $I(p_j)$ are the intensity values at positions p_i and p_j in the patch.

The results of these binary comparisons are concatenated to form a binary string, which serves as the keypoint descriptor. To match descriptors, the Hamming distance is utilized, which counts the number of differing bits between two binary strings, thereby enabling fast and efficient keypoint matching. Using ORB in this manner supports effective feature matching, which is crucial for accurate depth estimation in the specified regions.

Fig. 4. Canny edge detection results using a real time frame.

3 Experimental results

To assess the efficacy of our proposed method, we carried out real-time experiments utilizing two OV5647 5MP cameras paired with a Raspberry Pi4. These experiments demonstrated that our approach is feasible for deployment on a resource-constrained platform, highlighting its computational efficiency.

Figure 5 illustrates a comparative analysis of keypoints detected by the ORB algorithm versus those identified by our method, showcasing performance variations across different conditions. The figure clearly shows that ORB's keypoint detection significantly diminishes when the person is positioned at a lateral position. This reduction becomes more severe as the bounding boxes shrink in such scenarios, resulting in fewer detectable features. This limitation of ORB presents challenges for applications that rely on consistent feature detection, such as depth estimation, where a high density of keypoints is essential for reliable results.

In contrast, our method effectively addresses this issue by preserving a robust number of keypoints irrespective of the subject's orientation or bounding box size. This stable keypoint detection is crucial for ensuring the accuracy of subsequent feature matches and, consequently, the precision of depth estimation. Notably, our approach maintains a sufficient number of keypoints even as the distance between the person and the camera increases, where traditional methods like ORB may encounter difficulties.

Alongside the keypoint comparison shown in Figure 5, Figure 6 consolidates the analysis of match quantities generated by both ORB and our method at a fixed distance. This figure demonstrates that our approach yields a significantly higher number of matches due to its superior keypoint detection. This abundance of matches is essential for robust depth estimation.

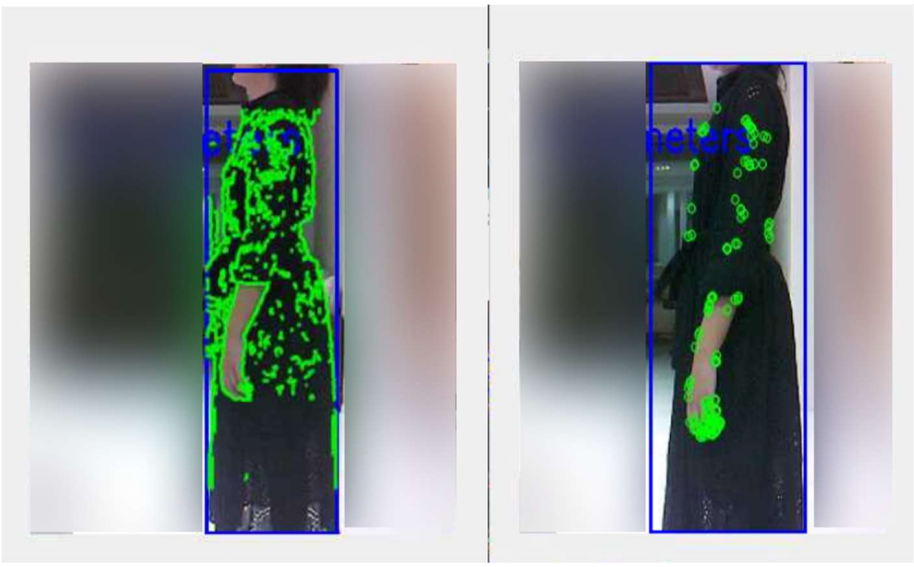

Fig. 5. Generated Keypoints: Left image represents the keypoints generated by canny edge detector while the right image represents ORB keypoints.

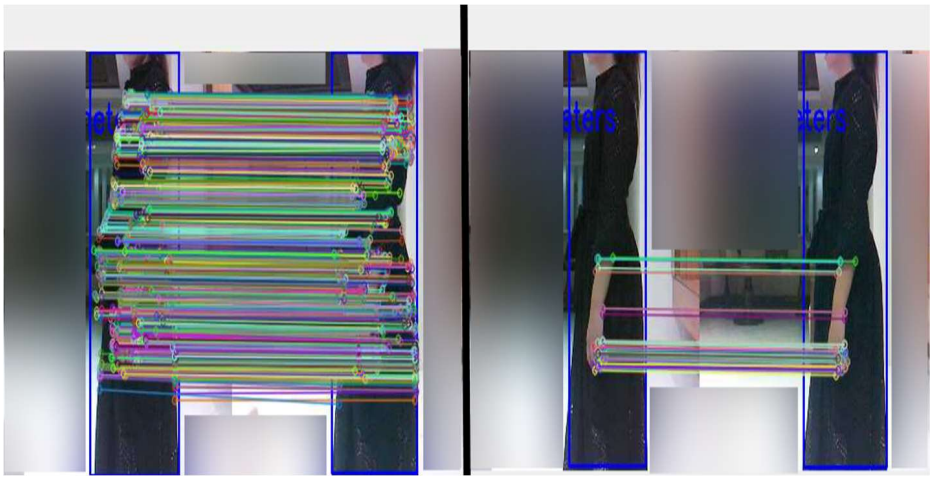

Fig. 6. Generated matches: Left image represents matches produced using canny edge detection keypoints while right image represents matches from ORB keypoints.

Table 1 provides a quantitative analysis of keypoints and matches produced by both ORB and our method at varying distances from the camera. This table supports the visual data by highlighting our method's superior ability to maintain keypoint and match quality, even under challenging conditions.

To evaluate our depth estimation approach thoroughly, we compared real-time results from our system with actual distances measured using a precision meter. The results indicate that our method achieves high accuracy at close range, demonstrating its robustness in such scenarios. However, accuracy diminishes as the distance increases, primarily due to the limited baseline of our stereo camera setup, which is only 5 cm. This short baseline restricts the effective range for accurate depth estimation. Table 2 details these findings.

Table 1. Comparison of keypoints and matches: Canny Edge detection VS ORB.

Distance(m)	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**
Number of ORB Keypoints(Left/Right)	95/61	32/18	13/15	7/12	None	None
Number of ORB matches	27	11	7	4	None	None
Our method: Number of Keypoints(Left/Right)	6027/6125	3709/3874	2314/2334	1554/1441	1393/1421	1070/1045
Our method : Number of matches	1298	974	655	419	489	456

Table 2. Comparison between original and predicted depth.

Original Depth(m)	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**
Predicted Depth(m)	0.53	1.1	1.7	2.3	2.9	3.7

4 Conclusion

Pedestrian detection plays a crucial role in enhancing safety within advanced driver assistance systems. While detecting pedestrians using bounding boxes is an important step, it is not sufficient alone for comprehensive safety; accurate distance estimation is also required. This paper introduces a new methodology that combines stereovision with ORB feature matching to estimate distances, utilizing keypoints obtained through edge detection. Experiments performed in real-time on a Raspberry Pi 4 highlight the practicality of implementing this method on resource-constrained hardware, achieving high performance with minimal error. Future work will focus on evaluating the approach with more capable hardware and assessing its effectiveness with a larger baseline to determine its performance over greater distances.

References

1. A. Lazaro, M. Lazaro, R. Villarino, and D. Girbau, 'Smart Spread Spectrum Modulated Tags for Detection of Vulnerable Road Users with Automotive Radar', *Sensors*, vol. 23, no. 5. 2023. doi: 10.3390/s23052730.
2. K.-I. Na and B. Park, 'Real-time 3D multi-pedestrian detection and tracking using 3D LiDAR point cloud for mobile robot', *ETRI Journal*, vol. 45, no. 5. pp. 836–846, 2023. doi: 10.4218/etrij.2023-0116.
3. A. N. Tabata, A. Zimmer, L. dos Santos Coelho, and V. C. Mariani, 'Analyzing CARLA 's performance for 2D object detection and monocular depth estimation based on deep learning approaches', *Expert Systems with Applications*, vol. 227. 2023. doi: 10.1016/j.eswa.2023.120200.

4. O. Rachidi, E.-D. Chafik, and B. Bououlid, 'Design of a Real-Time-Integrated System Based on Stereovision and YOLOv5 to Detect Objects', 2023, pp. 283–297. doi: 10.4018/979-8-3693-0497-6.ch016.
5. X. Song, G. Li, L. Yang, L. Zhu, C. Hou, and Z. Xiong, 'Real and Pseudo Pedestrian Detection Method with CA-YOLOv5s Based on Stereo Image Fusion', *Entropy*, vol. 24, no. 8. 2022. doi: 10.3390/e24081091.
6. Y. Xu and B. Zhang, 'BFOP3D: binocular feature based occluded pedestrian 3D detection', *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 13163. 2024. doi: 10.1117/12.3030753.
7. T. Ophoff, K. V. Beeck, and T. Goedeme, 'Improving Real-Time Pedestrian Detectors with RGB+Depth Fusion', *Proceedings of AVSS 2018 - 2018 15th IEEE International Conference on Advanced Video and Signal-Based Surveillance*. 2018. doi: 10.1109/AVSS.2018.8639110.
8. J. Zhang, Z. Ma, and N. Nuermaimaiti, 'A pedestrian detection model based on binocular information fusion', *2019 28th Wireless and Optical Communications Conference, WOCC 2019 - Proceedings*. 2019. doi: 10.1109/WOCC.2019.8770601.
9. A. Tupper and R. Green, 'Pedestrian Proximity Detection using RGB-D Data', *International Conference Image and Vision Computing New Zealand*, vol. 2019-December. 2019. doi: 10.1109/IVCNZ48456.2019.8961013.
10. V. Harisankar and R. Karthika, 'Real time pedestrian detection using modified Yolo V2', *Proceedings of the 5th International Conference on Communication and Electronics Systems, ICCES 2020*. pp. 855–859, 2020. doi: 10.1109/ICCES48766.2020.09138103.
11. E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, 'ORB: An efficient alternative to SIFT or SURF', *Proceedings of the IEEE International Conference on Computer Vision*. pp. 2564–2571, 2011. doi: 10.1109/ICCV.2011.6126544.
12. G.-S. Hong and B.-G. Kim, 'A local stereo matching algorithm based on weighted guided image filtering for improving the generation of depth range images', Displays, vol. 49, pp. 80–87, 2017, doi: https://doi.org/10.1016/j.displa.2017.07.006.
13. R. A. Hamzah et al., 'A study of edge preserving filters in image matching', Bulletin of Electrical Engineering and Informatics, vol. 10, no. 1. pp. 111–117, 2021. doi: 10.11591/eei.v10i1.1947.
14. R. A. Hamzah, M. N. Z. Azali, Z. M. Noh, M. Zahari, and A. I. Herman, 'Development of depth map from stereo images using sum of absolute differences and edge filters', Indonesian Journal of Electrical Engineering and Computer Science, vol. 25, no. 2. pp. 875–883, 2022. doi: 10.11591/ijeecs.v25.i2.pp875-883.
15. K. P. Kshirsagar and R. A. Shinde, 'Depth estimation using stereo medical imaging', Robotics and Automation in Healthcare: Advanced Applications. 2024. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85197124434&partnerID=40&md5=ca08d849e85754aa228936291d14ace1
16. H. Shahverdi, P. F. Moshiri, M. Nabati, R. Asvadi, and S. A. Ghorashi, 'A CSI-Based Human Activity Recognition Using Canny Edge Detector', Human Activity and Behavior Analysis: Advances in Computer Vision and Sensors: Volume 2. 2024. doi: 10.1201/9781032636054-5.
17. K. Liu, K. Xiao, and H. Xiong, 'An Image Edge Detection Algorithm Based on Improved Canny', in Proceedings of the 2017 5th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2017), Atlantis Press, Apr. 2017, pp. 533–537. doi: 10.2991/icmmct-17.2017.114.