Scientific Research Publishing

Journal of Computer and Communications, 2024, 12, 147-161
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

An Application of RGBD-Based Skeleton Reconstruction for Pedestrian Detection and Occlusion Handling

Ziyuan Liu

James Watt School of Engineering, University of Glasgow, Glasgow, UK
Email: fmadblaze@gmail.com

How to cite this paper: Liu, Z.Y. (2024) An Application of RGBD-Based Skeleton Reconstruction for Pedestrian Detection and Occlusion Handling. *Journal of Computer and Communications*, **12**, 147-161.
https://doi.org/10.4236/jcc.2024.121011

Received: November 29, 2023
Accepted: January 28, 2024
Published: January 31, 2024

Abstract

This study explores the challenges posed by pedestrian detection and occlusion in AR applications, employing a novel approach that utilizes RGB-D-based skeleton reconstruction to reduce the overhead of classical pedestrian detection algorithms during training. Furthermore, it is dedicated to addressing occlusion issues in pedestrian detection by using Azure Kinect for body tracking and integrating a robust occlusion management algorithm, significantly enhancing detection efficiency. In experiments, an average latency of 204 milliseconds was measured, and the detection accuracy reached an outstanding level of 97%. Additionally, this approach has been successfully applied in creating a simple yet captivating augmented reality game, demonstrating the practical application of the algorithm.

Keywords

AR, Pedestrian Detection, Occlusion Management, RGB-D, Azure Kinect, Unity

1. Introduction

Pedestrian detection stands as a pivotal challenge within the field of computer vision with several applications poised to positively impact our society ranging from surveillance, and robotics to automotive safety [1] [2]. Detecting pedestrians becomes even more complex with occlusions and varying scales coming into play. With an increasing incorporation of automated functions in vehicles and surveillance systems, an efficient system to detect pedestrians is imperative. Furthermore, pedestrian detection is particularly challenging in monocular images, which often suffer from low resolution and present pedestrians partially

occluded, both prevalent in the context of automotive systems [1]. The need for a novel approach leveraging RGBD-based skeleton reconstruction for pedestrian detection becomes critical due to these complexities. It has the potential to mitigate issues associated with the state of the art pedestrian detection systems, such as dealing with fluctuating scales and occlusions [2] [3]. Rather than attempting to solve pedestrian detection as a broad issue, this work's focus specifically lies in tackling this problem within the framework of Augmented Reality (AR) applications—it does this by reconstructing a three-dimensional human skeletal model and addressing occlusion handling. This unique approach could accelerate incorporating autonomous functionalities within AR technology.

During a multitude of pedestrian detection methodologies, the application of RGBD-Based Skeleton Reconstruction presents a unique and significant approach. Pedestrian detection is frequently hindered by issues related to occlusion and poor resolution especially in monocular images in urban settings [1] [2]. Conventional detection methods, despite their progress, struggle to efficiently handle complications arising from occlusions and multi-scale resolutions [3]. A viable solution to these predicaments lies in the concept of skeleton reconstruction, which uses RGBD-based technologies to recreate a human skeletal model. This not only enhances detection capabilities but also addresses challenges associated with scale and varying degrees of occlusion by providing a 3-dimensional perspective. Unlike conventional methods that operate on 2-dimensional data, skeleton reconstruction remains robust to varying lighting conditions and poses, making it a preferred choice for integration into augmented reality (AR) applications [4].

Exploring the domain of AR, the use of RGBD-Based Skeleton Reconstruction and Occlusion Handling turns out to be crucial. AR technology intricately merges virtual objects with real-world settings, making occlusion handling vital for visual realism. An efficient occlusion handling system can significantly augment the user experience by ensuring that virtual objects are appropriately inserted into the physical environment while considering perspective, depth and occlusion [1] [5]. Traditional AR systems, however, often fail to reach this level of refinement. The incorporation of a mechanism using RGBD-based Skeleton Reconstruction and occlusion handling can potentially address this problem, enhancing the realism integral to AR technology. By effectively embedding 3-dimensional human skeletal models into a virtual environment, managing occlusions becomes more effective, significantly boosting visual authenticity in AR applications [5]. Thus, the integration of pedestrian detection methodologies and AR through RGBD-based skeleton reconstruction and occlusion handling promises significant advances in future technology.

This research introduces an inventive dual-method approach that incorporates "RGBD-based 3D Skeleton Reconstruction" and "Occlusion Handling", designed primarily for pedestrian detection, yet holding vast potential for broader applications. Our creation of an interactive Augmented Reality (AR) game firm-

ly illustrates this adaptability. The fusion of these methods within an AR context allows for efficient translation of complex real-world visuals into immersive digital experiences. AR games prove to benefit in two significant ways from this merger. First, the 3D Skeleton Reconstruction enhances user engagement by rendering more accurate, tangible digital replicas of real-world objects. Secondly, the Occlusion Handling technique drastically improves AR gameplay by intelligently distinguishing and managing obstructed visuals, thereby presenting a clear, uninterrupted digital field to users. Though pedestrian detection and occlusion handling remain our primary research focus, this successful foray into AR game creation illuminates the broader, practical resonance of these methods across various domains.

In conclusion, in order to reduce the overhead of classical pedestrian detection algorithms during training and for wilder application prospects, this article introduces an RGBD-based 3D skeleton reconstruction. Also, this article proposes an effective occlusion handling method.

The key contributions of this research's work are as follows:

- This article introduces a novel RGBD-based 3D human skeleton reconstruction method that addresses the complexities of pedestrian detection and sets a clear trajectory for future research.
- This article proposes an effective occlusion handling method using color frame and depth frame that further leverages the applicability of our pedestrian detection technique.
- The two technical methods yield impressive experimental results, reflected in the GIoU Value of 0.473 and high user satisfaction. Their potential usefulness is further highlighted in the development and implementation of a straightforward Augmented Reality (AR) shooting game, exemplifying the practical deployment of these methods.

The remainder of this article is organized as follows. Section II presents the related works in pedestrian detection, application of computer vision in AR gaming, and occlusion handling. Section III presents the methods of RGBD-based 3D skeleton reconstruction and occlusion handling. Section IV presents the evaluation of these two methods. Section V presents the application of these methods, an AR shooting game. Finally, this article concludes this work in Section VI.

2. Related Works

2.1. Pedestrian Detection

In the realm of computer vision, human body recognition is recognized as a complex and multi-dimensional topic. Years of exhaustive research have substantially honed our comprehension and proficiency; however, obstacles and difficulties remain. Concentrating on real-time human pose recognition, the research conducted by Jamie Shotton and his colleagues [6], delivered a methodology for swiftly and accurately discerning human postures within live settings. The primary advantage of their work lies in its real-time, precise human pose

recognition capabilities. Yet, its requirement for a depth camera diminishes its practicability, especially within vehicle cabins. Addressing the challenge of human key point detection, the study by Jing Zhang and her colleagues [7], approached the issue from the aspects of computational efficiency and precision. Although this study boasts a highly precise detection of human key points, the high computational cost limits its applicability, rendering it unfavorable in settings with resource constraints.

Within the field of 3D human pose estimations, Jinbao Wang and his associates [8] provided an in-depth review of various methodologies and their applications. Despite the comprehensive array of 3D pose estimation techniques presented in this review, it should be noted that a majority of them demand significant computational resources. A method for action recognition based on image sequences was presented by Qingdi Wei and his team [9]. Although their work demonstrates dependable identification of human movements within complex environments, the need for comprehensive image data and computational resources is challenging.

In conclusion, while significant progress has been made in the field of human body recognition technology, especially in terms of accuracy and real-time performance, deficiencies remain prominent, particularly in the areas of computational overheads and hardware requirements. These insights provide a critical benchmark for our own research, especially when considering practical applications of human body recognition within vehicle interiors.

2.2. Application of Computer Vision in AR Gaming

In the sphere of Augmented Reality (AR) gaming, the incorporation of computer vision technologies has notably escalated. These technologies not only heighten the immersive aspect of the gaming experience but also bridge the interactional gaps between the physical and the virtual realms. However, the field is constrained by specific limitations and challenges. AR gaming primarily utilizes computer vision for object recognition, tracking, and scene understanding. A study by Anastasiya Zharovskikh [10], under the auspices of InData Labs, sheds light on the broad application of computer vision in AR and VR, including object tracking and scene reconstruction. The benefits of these technologies lie in creating a more lifelike and immersive experience while the drawback pertains to the need for high-performance hardware and algorithms.

Conversely, Debiprasad Bandopadhyay underscores that AR gaming platforms such as Pokémon GO allow users to capture virtual creatures in the real world [11]. The strength of this application lies in the seamless blending of the virtual and real world. However, the precision of computer vision algorithms can present obstacles, especially when applied in complex environments like the outdoors or in low-light scenarios. Moreover, a paper by Nilesh Barla outlines the extensive possible applications of human pose estimation in AR gaming [12]. These applications often require highly accurate, real-time human pose data,

subsequently increasing computational load and energy usage. For instance, Trond Nilsen *et al.* in their work, "Tank War-Tabletop war gaming in augmented reality," used a simple webcam and leveraged computer vision technologies to convert a standard game board into an AR game [13]. This approach offers a realistic and immersive experience but needs high-performance hardware and complex algorithms. Similarly, a study by Feng Zhou *et al.* presents a comprehensive overview of dynamic areas of computer vision tracking research in AR gaming [14]. While the paper praises the virtues of computer vision applications in AR gaming, it also warns about the potential compromise in the accuracy of these algorithms in complex environments.

In summary, while the cited studies uniformly affirm the diverse applications and inherent potential of computer vision in AR gaming, they also acknowledge its drawbacks and challenges, particularly concerning hardware requirements and algorithmic accuracy. This provides a crucial reference for our research, especially for practical implementation of an AR gaming experience within the automotive interior environment.

2.3. Occlusion Handling

Occlusion presents a significant challenge within the field of computer vision, particularly in the context of augmented reality (AR). Often, AR in video-based displays merely superimposes virtual objects onto the real environment. This does not necessarily mirror the actual dynamics of an AR scene in many cases. When the reality is that a real object should occlude a virtual one, discrepancies in the augmented image may instead lead to user perception confusion. This discrepancy results in misunderstandings and possible errors during task execution. As described in "Occlusion in Augmented Reality" [15], an ideal situation is where real objects occlude virtual ones when the virtual content is nearer to the camera than the real content. Despite the crucial role it plays in maintaining the integrity of the AR experience, algorithms designed to handle these occlusion scenarios often demand significant computational resources.

Occlusion incidents in AR scenarios can arise from various sources, such as object movement or inconvenient placement that obstructs the AR interface. Strategies such as depth order determination, hidden object visualization, and creation of occlusion-capable displays are discussed in "Occlusion Handling in Augmented Reality: Past, Present, and Future" as common ways to manage occlusion [16]. However, as clarified in "Occlusion Handling for Mobile AR Applications in Indoor and Outdoor Scenarios" [17], these strategies are often computationally intensive. To counter occlusion, deep learning methods are employed, which have the benefit of learning from complex scenarios, enhancing their resilience against occlusion but require high computational power.

Contrary to traditional interaction methods often found lacking within tangible AR environments, an article proposes an innovative occlusion-based interaction technique [18]. This approach uses the visual occlusion of physical markers

to enable intuitive two-dimensional interaction in tangible AR environments—providing an edge where the integration of multiple tools or platforms may be required.

In conclusion, despite the existence of powerful algorithms and innovative interaction methods to address occlusion, the unique restrictions resulting from their computational demands still present considerable challenges. These barriers highlight potential areas for future research, focusing specifically on optimizing these algorithms and interaction methods for environments with restricted computational resources.

3. Methods

3.1. RGBD-Based Skeleton Reconstruction for Pedestrian Detection

The capabilities of pedestrian detection and occlusion handling are significantly enhanced by RGBD-based 3D skeleton reconstruction. This method is based on a unique approach that involves creating, updating, and removing elaborate skeleton models representative of detected pedestrians.

The development of the skeleton structure involves a specific dictionary, encompassing all joints and structural connections. Skeleton model creation is accomplished through iterative mechanisms looping through a pre-set array of joint types. This approach facilitates the effective construction of independent joints and connections for different situation, seamlessly integrated into a broader skeletal structure.

This technique distinguishes itself through its dynamic management of joint positions and their corresponding associations, enabling precise real-time tracking of pedestrians. RGB images returned by the camera are processed using deep learning methods for human identification. The depth sensor's values are then used to extract the 3D positions of joints, which are transitioned into Unity world coordinates for joint position updates. Specific transformation matrices and rotation values are employed in this process, promoting alignment between the skeleton model and the actual pedestrian. Moreover, real-time updates are made to the connections between joints. These connections are regarded as evolving line segments. Position, direction, and size are their primary attributes, dictated by the respective joint positions to which they are coupled. **Figure 1** showed the training process of the model.

The comprehensive skeleton model also houses routines to routinely update encapsulated colliders, thus enhancing interaction with neighboring physical phenomena. An iterative function inbuilt in the algorithm maintains the centrality of colliders vis-à-vis the "waist" joint of the skeleton, irrespective of the pedestrian's motions or rotations.

Additionally, certain features work in tandem to ensure completeness of the skeleton model. Specifically, these features ensure seamless integration of joints and connections based on confidence levels, numerically quantified and scaled from 0 to 1. The impact of these levels reflects in the visibility factors of skeleton

joints. Only joints and connections surpassing a pre-set confidence level threshold are activated.

Figure 2 demonstrates the component of skeleton reconstruction for pedestrian detection. The algorithm is also able to track multiple human forms simultaneously. Each detected individual is assigned a unique identification number to coordinate this multiplexing capability. Consequently, independent tracking and skeleton generation can be executed for each identification number, enabling simultaneous and independent multi-body tracking and detection. Figure 3 shows the camera was set as the Skeleton Avatar's parent.

In conclusion, our research proposes an effective method for real-time pedestrian detection, RGBD-based 3D skeleton reconstruction. Implementing these strategies could prove instrumental in various real-world applications where precision and accuracy are paramount. Figure 4 demonstrates how the two skeletons were identified and tracked at the same time.

```
Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.449
Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.810
Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.491
Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = -1.000
Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.449
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.480
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.627
Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.640
Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = -1.000
Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.640
```

Figure 1. Training the model.

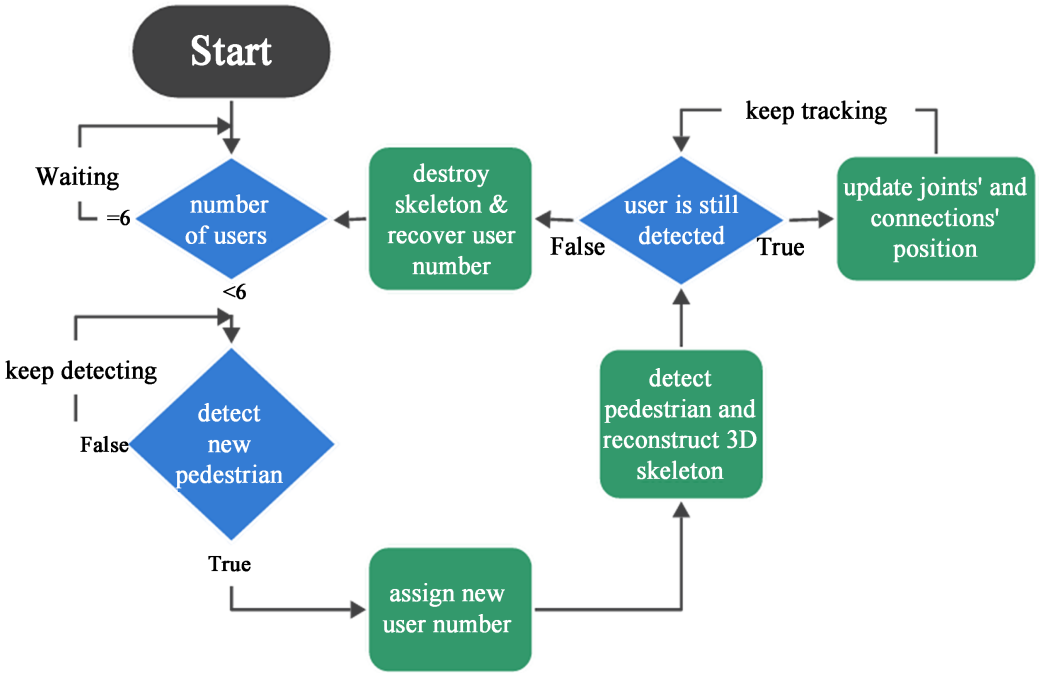

Figure 2. Component of skeleton reconstruction for pedestrian detection.

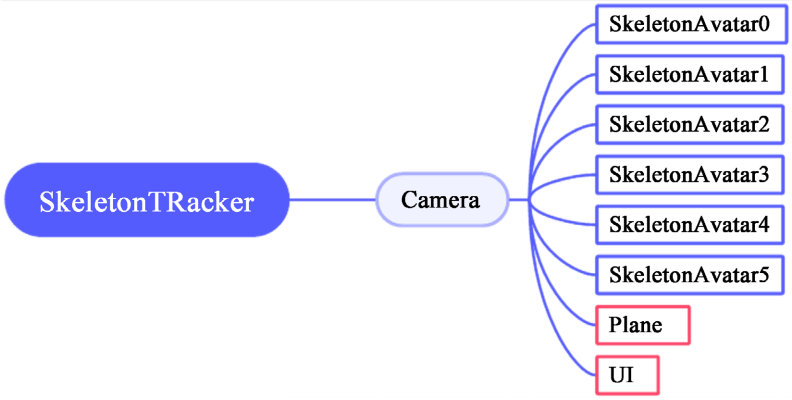

Figure 3. Set "Camera" as "SkeletonAvatar"s' parent.

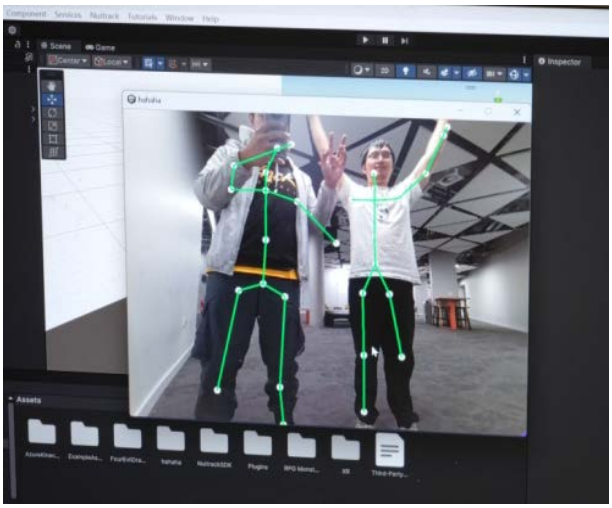

Figure 4. Identifying and tracking two skeletons at the same time.

3.2. Implementing Occlusion Handling by Integrating Depth and Color Frames

Through a novel amalgamation of color and depth frames, our research presents an innovative algorithm adept at overriding occlusion challenges. Color frames, colloquially termed as RGB frames, are proficient in capturing vivid visual inputs. In contrast, depth frames effectively register depth information respective to each pixel of the frame. The dual functionality offered by the paired use of these frames has necessitated the design of a script, meticulously detailed to handle both these facets competently.

Initially, the algorithm introduces the color frame data, which is minimally processed and has direct access to graphical hardware resources. The colorful visual data are transformed into 2D texture using a byte array. The significance of this mechanism, a pioneering one in its field, lies in its ability to hasten the data transformation process, essentially accelerating data transmission from the Central Processing Unit (CPU) to the Graphical Processing Unit (GPU). This operational enhancement not only escalates overall system performance but also conserves the integrity of the original image data effectively.

While the color frame is essential for acquiring the visually salient details, depth frames are integral for understanding the spatial configuration of the canvas. Our approach incorporates these frames to construct a height map. Given the restrictions of most graphical APIs, our technique methodically transforms the information contained in these frames into a byte array. This ensures an unambiguous data transmission to the GPU and concurrently avoids unnecessary computations and memory allocations.

Following the successful application and synchronization of color and depth frames, the dynamic adjustment of the 3D mesh size commences trialed against changes in the camera's field of view (FOV) and the dimensions of the input frame. Constructing a realistic rendering of a 3D mesh necessitates the derivation of a target aspect ratio and a scale factor, both calculated based on the camera's vertical FOV. This ensures the 3D mesh aligns as closely as possible to the real-world proportions, further enhancing the algorithm's effectiveness.

Additionally, our research methodology carves an optimized approach to 3D mesh generation. A schematic of vertices and a UV coordinate list are conceptualized and subsequently exploited for constructing the mesh's triangles. Post the completion of the triangle's construction, the mesh undergoes a series of computational optimizations. This leads to a re-calculation of normals and bounds, which bolsters operational performance and significantly improves the algorithm's speed and efficiency as Figure 5 shows.

In summary, the application of our algorithm enables dynamic mesh adjustment to create a realistic rendering of occlusions, thereby providing an impression of object distances from the viewer. By utilizing color and depth frames, this research has introduced a methodology that not only promises high performance and superb rendering quality but also provides a solid foundation for

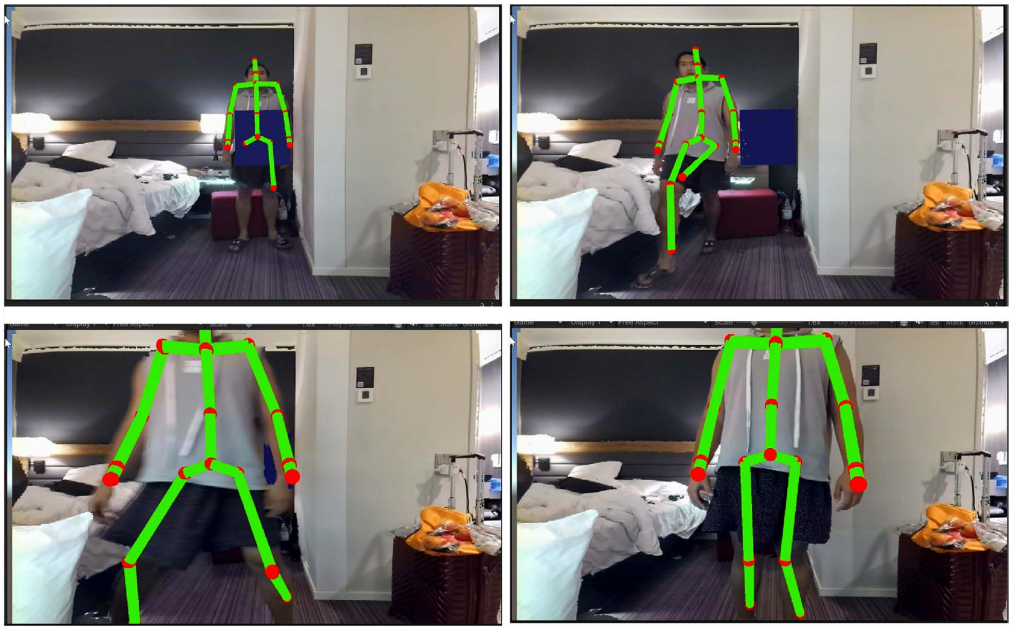

Figure 5. "Occlusion".

future advancements in this realm.

4. Results and Discussion

System Performance Evaluation

Dataset: This research used Human3.6 M as the dataset to train our deep learning skeleton tracking algorithm. The Human3.6 M dataset is a dataset for 3D human pose recognition, captured by four calibrated cameras. It includes annotations for the positions of 24 body parts and joint angles in 3D human poses. And the Human 3.6 M dataset comprises 3.6 million 3D human pose images, featuring 11 professional actors (6 males and 5 females) and 17 scenes (discussing, smoking, taking photos, making phone calls, etc.).

Evaluation metric: Precision is calculated as the ratio of true positive (TP) objects, which are correctly classified as positive, to all objects classified as positive by the classifier (TP + FP).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

Recall, also known as sensitivity, refers to the proportion of objects that are correctly identified as positive (TP) among all objects, including those that were not classified as positive but are indeed positive (TP + FN).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

Average Precision (AP) refers to the area enclosed by the curve with Recall on the horizontal axis and Precision on the vertical axis. The value of AP is the area covered by the curve.

$$\text{AP} = \sum_{i=1}^{n-1} \left(r_{i+1} - r_i \right) p\left(r_{i+1} \right) \tag{3}$$

IoU, which stands for Intersection over Union, is the most commonly used metric in object detection. It reflects the detection performance of predicted bounding boxes compared to the ground truth bounding boxes. First, calculate the area of the smallest closed bounding box, denoted as A_c (in simple terms, it is the area of the smallest box that contains both the predicted box and the ground truth box). Then, calculate the IoU, and calculate the proportion of the area in the closed bounding box that does not belong to either of the two boxes. Finally, subtract this proportion from IoU to obtain GIoU.

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \tag{4}$$

$$\text{GIoU} = \text{IoU} - \frac{|A_c - U|}{|A_c|} \tag{5}$$

Accuracy is the proportion of correctly predicted quantity to the total quantity.

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6}$$

Experimental Design: The presented experiment's design was methodical, involving Latency Tests to measure the system's response time and compute average latency, Accuracy Tests to evaluate the Azure Kinect sensor's data precision under varying conditions like lighting, distance, and speed, and Resource Usage Tests monitoring the system's CPU, memory, battery, and network resource usage across diverse operational stages such as startup, active running, and shutdown.

Results and Analysis: The test results yielded desirable outcomes; the latency test revealed an average latency of 204 milliseconds adequately catering to real-time interaction needs. The accuracy test drove home that Azure Kinect's average accuracy, under a plethora of test conditions, touched 97%, rendering it evidently superior to similar offerings. The Resource Usage Test displayed stable outcomes; the CPU usage did not surpass 39% under peak-load operation, memory usage persistently fell below 2.3 GB, and battery longevity remained unswerving as Figure 6 showed.

5. Application

The overarching concept involves employing human recognition technology to track up to six persons and devise 3D skeletal models for each one of them, with colliders generated based on the skeletons' positions. Within a defined area in the virtual world, various monster types are randomized in their generation. A brief delay post-generation incites these monsters to initiate movements towards the camera. Each monster type boasts its own model—equipped with colliders and rigid bodies—animations, pace, point value, and health. Remarkably,

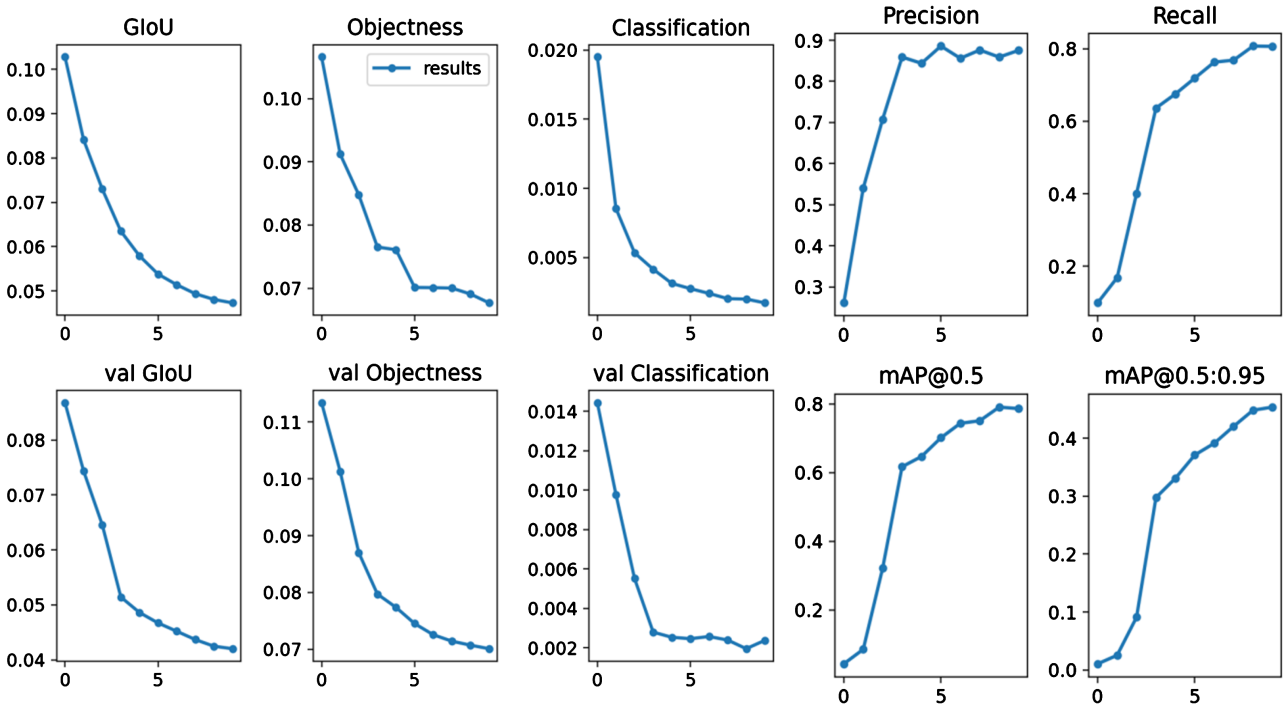

Figure 6. Model performance.

defeating certain monsters results in replenishment of the player's health. The course of the gameplay revolves around regulating the camera's rotation in the virtual world via mouse movements. Players engage in shooting by clicking the left mouse button, thereby earning points for each monster killed. Conversely, shooting a human form leads to a decrement in the player's life points. The game culminates either when the score threshold is achieved or the player's life points dwindle to zero. Figure 7 is the demonstration of the system components.

Player's Behavior Management: This section's authentic replication of real-life player movements within the game environment is achieved with a script developed in Unity's API and C#, which shapes the core logic and sequences of the game. It encompasses player interactions and intricate game mechanics. The introduction of state management, event triggering, and UI updates enriches the

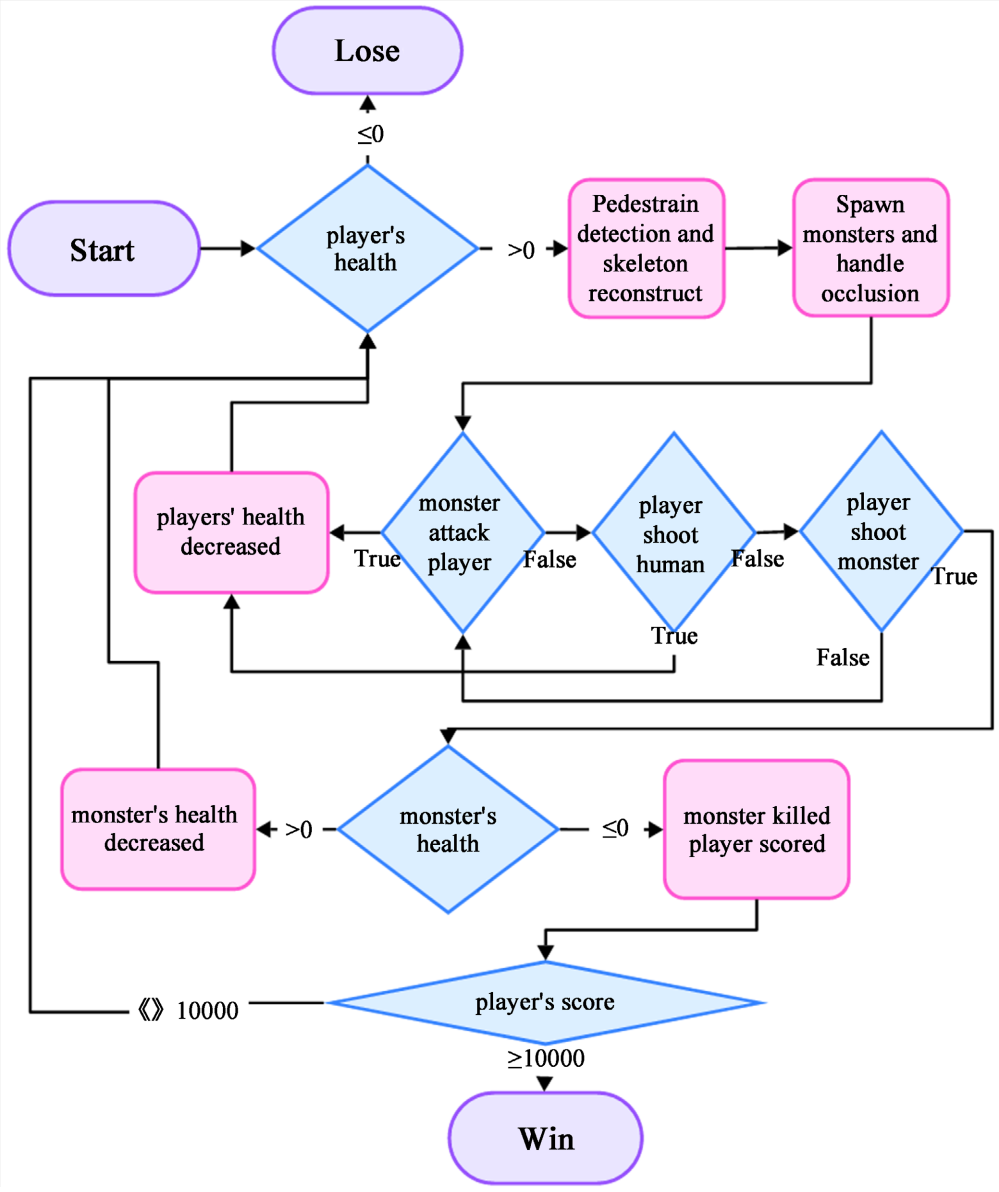

Figure 7. System components.

gameplay experience. The effective integration of skeleton reconstruction and occlusion handling enables a seamless blend of real-world actions with their in-game counterparts, resulting in an immersive and user-friendly gaming dynamic.

Spawn Monsters: This section describes functionalities pertaining to the generation, attribute setting, and behavioral patterns of monsters. This script adeptly organizes data and algorithms to execute a dynamic and adaptive system for monster generation and behaviour management. It demonstrates the application of Unity's API and the C# programming language in complex game-logic creation, holding substantial scientific value in fields such as Procedural Content Generation (PCG), Behavior Trees, State Machines, and AI decision models. Figure 8 shows the properties of monsters.

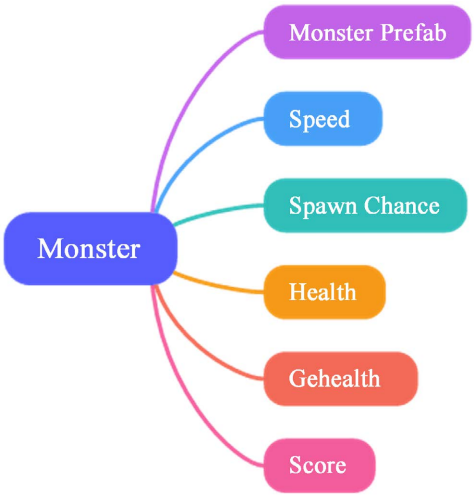

Figure 8. "Monsters".

6. Conclusion

In conclusion, this study introduces an innovative approach to pedestrian detection and occlusion handling, harnessing the power of RGBD-based skeleton reconstruction. Our dual-technique system not only tackles the existing complexities in pedestrian detection but also exhibits considerable promise for wider applications. One such application is evidenced in the creation of an AR game, a striking example of our technology's ability to convert intricate real-world scenarios into immersive digital interactions. The persuasive experimental results coupled with high user acceptability validate our system's efficacy and functionality. Nonetheless, regarding occlusion handling, instances persist where certain pixels at boundaries are inaccurately rendered in scenarios involving complex colors or shapes. Future research endeavors may propose solutions to these limitations.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Dollar, P., Wojek, C., Schiele, B. and Perona, P. (2011) Pedestrian Detection: An Evaluation of the State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**, 743-761. https://doi.org/10.1109/TPAMI.2011.155

[2] Dollár, P., Wojek, C., Schiele, B. and Perona, P. (2009) Pedestrian Detection: A Benchmark. 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 20-25 June 2009, 304-311. https://doi.org/10.1109/CVPR.2009.5206631

[3] Xiao, Y., Zhou, K., Cui, G., Jia, L., Fang, Z., Yang, X. and Xia, Q. (2021) Deep Learning for Occluded and Multi-Scale Pedestrian Detection: A Review. *IET Image Processing*, **15**, 286-301. https://doi.org/10.1049/ipr2.12042

[4] Zhang, S., Benenson, R., Omran, M., Hosang, J. and Schiele, B. (2016) How Far Are We from Solving Pedestrian Detection? 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 27-30 June 2016, 1259-1267. https://doi.org/10.1109/CVPR.2016.141

[5] Guo, X., Wang, C. and Qi, Y. (2017) Real-Time Augmented Reality with Occlusion Handling Based on RGBD Images. 2017 *International Conference on Virtual Reality and Visualization* (*ICVRV*), Zhengzhou, 21-22 October 2017, 298-302. https://doi.org/10.1109/ICVRV.2017.00069

[6] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., *et al.* (2011) Real-Time Human Pose Recognition in Parts from Single Depth Images. *CVPR* 2011, Colorado Springs, CO, 20-25 June 2011, 1297-1304. https://doi.org/10.1109/CVPR.2011.5995316

[7] Zhang, J., Chen, Z. and Tao, D. (2021) Towards High Performance Human Keypoint Detection. *International Journal of Computer Vision*, **129**, 2639-2662. https://doi.org/10.1007/s11263-021-01482-8

[8] Wang, J., Tan, S., Zhen, X., Xu, S., Zheng, F., He, Z. and Shao, L. (2021) Deep 3D Human Pose Estimation: A Review. *Computer Vision and Image Understanding*, **210**, 103225. https://doi.org/10.1016/j.cviu.2021.103225

[9] Wei, Q., Hu, W., Zhang, X. and Luo, G. (2007) Dominant Sets-Based Action Recognition Using Image Sequence Matching. 2007 *IEEE International Conference on Image Processing*, San Antonio, TX, 16 September-19 October 2007, VI-133-VI-136. https://doi.org/10.1109/ICIP.2007.4379539

[10] Zharovskikh, A. (2020) The Role of Computer Vision in AR and VR. https://indatalabs.com/blog/computer-vision-ar-vr

[11] Bandopadhyay, D. (2023) The Future of AR and Computer Vision: What to Expect. https://www.linkedin.com/pulse/future-ar-computer-vision-what-expect-debiprasad-bandopadhyay/

[12] Barla, N. (2021) A Comprehensive Guide to Human Pose Estimation. https://www.v7labs.com/blog/human-pose-estimation-guide

[13] Nilsen, T. and Looser, J. (2005) Tankwar-Tabletop War Gaming in Augmented Reality. 2nd International Workshop on Pervasive Gaming Applications, PerGames, 5.

[14] Zhou, F., Duh, H. B. L. and Billinghurst, M. (2008) Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR. 2008 *7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, Cambridge, 15-18 September 2008, 193-202. https://doi.org/10.1109/ISMAR.2008.4637362

[15] Shah, M. M., Arshad, H. and Sulaiman, R. (2012) Occlusion in Augmented Reality. 2012 *8th International Conference on Information Science and Digital Content Technology* (*ICIDT*2012), **2**, 372-378.

[16] Macedo, M.C.D.F. and Apolinario, A.L. (2021) Occlusion Handling in Augmented Reality: Past, Present and Future. *IEEE Transactions on Visualization and Computer Graphics*, **29**, 1590-1609. https://doi.org/10.1109/TVCG.2021.3117866

[17] Alfakhori, M., Sardi Barzallo, J.S. and Coors, V. (2023) Occlusion Handling for Mobile AR Applications in Indoor and Outdoor Scenarios. *Sensors*, **23**, 4245. https://doi.org/10.3390/s23094245

[18] Lee, G. A., Billinghurst, M. and Kim, G. J. (2004) Occlusion Based Interaction Methods for Tangible Augmented Reality Environments. *Proceedings of the* 2004 *ACM SIGGRAPH International Conference on Virtual Reality continuum and Its Applications in Industry*, 419-426. https://doi.org/10.1145/1044588.1044680