International Journal of Internet, Broadcasting and Communication Vol.16 No.4 124-132 (2024)
http://dx.doi.org/10.7236/IJIBC.2024.16.4.124

IJIBC 24-4-14

Comparative Evaluation of AI Driven Markerless Motion Capture Tools for Efficiency

Balgum Song

Professor, Department of International College, Dongseo University, Korea
sbu1977@dongseo.ac.kr

Abstract

We explore the effectiveness of AI-driven markerless motion capture (MoCap) tools compared to the traditional marker-based OptiTrack system, known for its high accuracy in capturing precise movements. Through detailed comparative analysis, we assessed various free markerless MoCap tools, including Move One, Radical, Deep Motion, Plask, Rokoko, and Movmi, focusing on critical aspects such as pose accuracy, movement smoothness, and ground detection. Our findings indicate that Move One is the most versatile tool, offering excellent pose accuracy, smooth MoCap, and reliable ground detection, making it a strong contender for various animation tasks. We found that Radical excels in minimizing jitter, making it suitable for projects requiring smooth motion, while Deep Motion performs best in ground detection, which is crucial for accurate foot placement. Although markerless systems still do not fully match the precision of marker-based systems, we suggest that they present viable alternatives depending on the specific needs of a project. As AI technology continues to advance, we expect the gap between markerless and marker-based to narrow, expanding the potential applications of markerless MoCap in the industry.

Keywords: *Motion Capture, AI, Marker Base, Markerless*

1. Introduction

The evolution of motion capture (MoCap) technology has revolutionized the industry by enhancing digital capabilities, stimulating creativity, and establishing a solid foundation in motion knowledge. Initially, MoCap was primarily used for capturing detailed and precise movements, significantly contributing to fields such as professional animation and biomechanical research. Over time, the application of MoCap has expanded, providing an intuitive tool for exploring a wide range of possibilities and creating a large volume of valid actions [1]. This capability has made MoCap an indispensable asset in the industry, improving the efficiency and effectiveness of character expressions. By streamlining the process, MoCap allows for the creation of expressions more rapidly and precisely, thereby simplifying 3D animation for educators and facilitating the learning process [2]. In the 21st century, MoCap also became a core technology in special effects films, improving animation quality and production efficiency in comparison to traditional stunts, and bringing

Manuscript Received: September. 15, 2024 / Revised: September. 20, 2024 / Accepted: September. 26, 2024
Corresponding Author: sbu1977@dongseo.ac.kr
Tel: *** - **** - ****
Professor, Department of International College, Dongseo University, Korea

qualitative advancements to visual effects [3].

The advent of artificial intelligence (AI) has further advanced MoCap technology. The integration of AI into MoCap systems has opened up new possibilities, particularly in enhancing the precision and reliability of markerless MoCap. Through deep learning and computer vision, AI can automate the recognition and tracking of complex movements, significantly enhancing the speed and accuracy of data processing. This streamlines the animation production process, making motion capture more precise and efficient, while lowering costs, enabling more creators to access and utilize this technology [4]. Besides, traditional marker-based systems, such as OptiTrack, have long been the benchmark for high accuracy in capturing human movement, providing higher quality motion data, more accurate body joint angles, and better measurements of body segment lengths compared to markerless systems [5-8]. Nevertheless, the emergence of AI-driven markerless MoCap tools offers a promising alternative that could potentially match the quality of marker-based systems.

This study aims to analyze and compare the traditional marker-based OptiTrack system with several free markerless MoCap AI tools, including Move One, Radical, Deep Motion, Plask, Rokoko, and Movmi. The objective is to determine their suitability as alternatives to high-quality marker-based systems for efficiency.

Despite the rapid changes in the AI field, this research provides a foundational reference point for future innovations. By documenting the current state of markerless MoCap AI, this study offers valuable insights that future researchers can build upon, track progress, and identify emerging trends. The continuous improvement in AI algorithms promises to close the gap between markerless and marker-based systems, making markerless MoCap an increasingly viable option for a wide range of applications and beyond.

2. Background

In the realm of MoCap technology, two primary methodologies have shaped the industry: marker-based and markerless systems. These technologies, with their unique sets of advantages and limitations, significantly influence their adoption and application across various fields.

2.1 Marker Based MoCap

Marker-based MoCap systems are the industry standard for capturing highly precise and detailed movement data. However, this sophistication comes with several drawbacks. The equipment required for these systems is both complex and expensive. Multiple cameras, specialized software, and the markers themselves represent a significant financial investment. Additionally, the setup process is cumbersome, involving the meticulous placement of markers on the actor, requiring subjects to wear unnatural, skin-tight clothing, and the calibration of cameras [9]. This setup is not only time-consuming but also limits the flexibility of the capture environment. The capture area is restricted to spaces where the cameras have a clear view of the markers, which can be a significant constraint in dynamic or large-scale scenes.

2.2 Markerless MoCap

In contrast, markerless MoCap systems offer a more accessible and flexible alternative. These systems leverage computer vision techniques to track movements without the need for physical markers. A single camera can capture the necessary data, simplifying the setup process and reducing costs. This ease of use makes markerless systems particularly appealing for smaller projects, educational settings, and situations where budget constraints are a concern.

Early iterations of this technology were often criticized for their lack of precision compared to marker-based systems, particularly in the area of markerless facial motion capture animation, where no satisfactory industrial process has yet been established [10]. The absence of physical markers meant that the systems had to rely solely on visual data, which often produced noticeable visual artifacts such as foot skate and inter-frame jitter [11]. However, recent advancements in AI have significantly enhanced the accuracy and reliability of markerless MoCap. Most of these algorithms train neural networks using manually labeled image data and then estimate human posture, including joint centers and skeletons, when the user inputs images or videos to the trained network [12, 13]. Modern AI-driven systems are capable of analyzing and interpreting movement data with a level of detail that rivals, and in some cases approaches, that of marker-based systems.

While AI has improved their precision, these systems still generally fall short of the high fidelity offered by marker-based systems, such as higher-quality foot captures, the precise and robust calculation of joint rotations using marker positions, and the adaptation to different contexts or increased accuracy [14,15]. Nonetheless, the continuous improvement in AI algorithms promises to further close this gap, making markerless systems an increasingly viable option for a wide range of applications.

2.3 Comparative Analysis

Markerless methods have not been widely adopted due to the technical challenges of accurately capturing human movement without markers. However, recent advancements in computer vision technology offer promising new possibilities for effective markerless MoCap. This study aims to explore how modern markerless MoCap systems have improved in terms of precision and reliability, traditionally dominated by marker-based systems.

By comparing various AI-driven markerless tools with the established benchmarks of marker-based systems, this research seeks to assess their suitability for animation purposes. The evaluation focuses on critical aspects such as pose accuracy, movement smoothness, and ground detection, which often cause issues in markerless data [16]. Understanding these parameters is crucial for determining how well markerless systems can replicate the nuanced movements required for high-quality motions.

The findings from this study are expected to provide valuable insights into the potential of markerless MoCap technology, particularly in contexts where budget constraints and the need for flexible, easy-to-use systems are significant considerations.

3. Research Procedures

To evaluate the accuracy of markerless MoCap AI tools in detecting poses and movements compared to marker-based MoCap, this study uses the marker-based OptiTrack system, known for its high accuracy, as a benchmark. To maintain consistency, we recorded the same performer simultaneously using both an OptiTrack camera (marker-based) and a personal phone camera (markerless) as shown in Figure 1. The OptiTrack data served as our standard for accuracy measurement, while the personal phone video was used to apply the 2D video image to each markerless MoCap AI tool for comparison. We then exported the MoCap FBX files from both systems and compared them in Maya.

Figure 1 illustrates the comprehensive steps involved in comparing the marker-based OptiTrack system with various markerless MoCap AI tools. Initially, the performer's motion is captured simultaneously using the OptiTrack system and a smartphone. The OptiTrack data undergoes processing in Motive software, including calibration, creation, and capture phases, to ensure accurate motion data. Meanwhile, the smartphone video is imported into several markerless MoCap AI tools, including Move One, Deep Motion, Plask, RADiCAL, Rokoko, and Movmi. The motion data from both the OptiTrack system and the MoCap AI tools is then exported as FBX files. These files are imported into 3D software like Maya, where a detailed comparison is conducted. The comparison focuses on evaluating pose accuracy, movement smoothness, and ground detection, providing a robust assessment of the markerless MoCap AI tools' performance against the benchmark OptiTrack system.

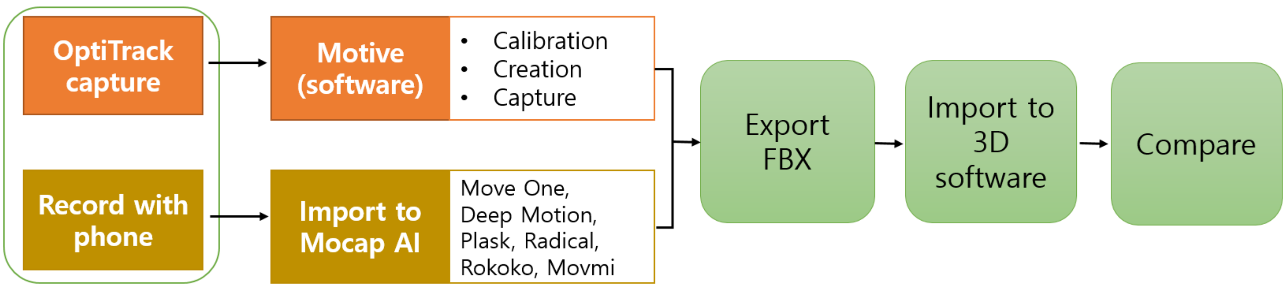


Figure 1. Procedure for marker-based and markerless MoCap analysis

The study examines movements and ground detection to assess the overall performance of these AI tools.

- **Pose Shape Accuracy**

Comparison of 13 extreme and passing pose shapes, including stand, walk (contact start, pass position, contact end), turn, jump (start, pass position, end, recovery), guard stance (start, end), punch, and kick.

- **Movement Analysis (Jitter Check and Action Smoothness)**

Evaluation of jitter, shake, and the natural flow of movements (arc).

- **Ground Detection**

Assessment of whether the foot stays on the ground, penetrates below it, or hovers above.

4. Comparative Analysis of Markerless MoCap AI Tools-Results

Each analysis was conducted using the OptiTrack poses as a benchmark and video footage showing commonly used poses such as walking, turning, jumping, punching, and kicking.

4.1 Pose Shape Accuracy

Accurately capturing poses for each body part is a fundamental and crucial skill in MoCap. This research selected 13 different extreme and passing poses commonly used in animation to evaluate how well each MoCap system AI can replicate precise poses from video.

Table 1 shows the comparison of pose shape accuracy among different MoCap systems is highlighted, with the poses outlined in red indicating significant deviations from the reference video. These deviations typically involve issues such as poorly executed leg and arm bends or tilted body positions.

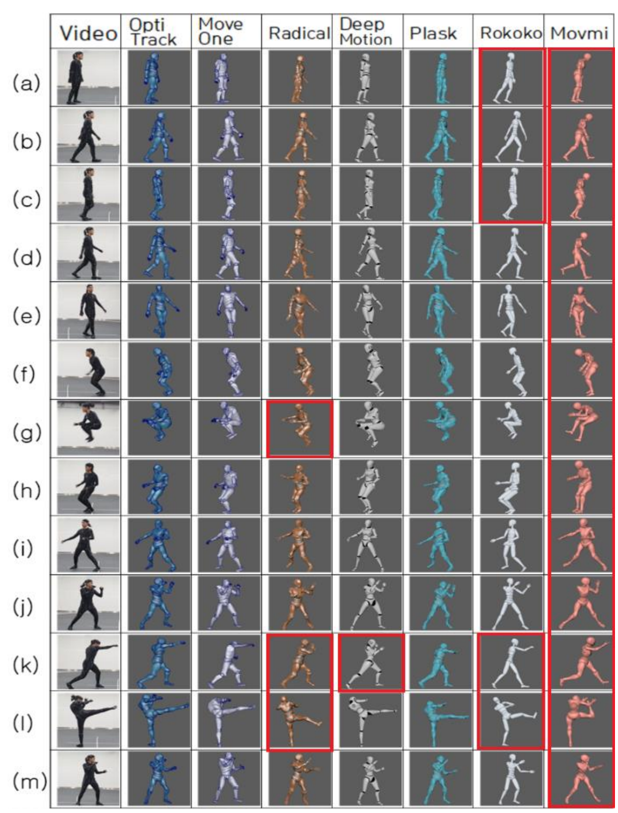
- **Move One**: Although generally accurate, Move One struggles with recognizing torso twists, particularly during punching motions, where the shoulder rotation is insufficient, resulting in a stiff appearance as shown in Table 1, row (k), column Move One.
- **Radical**: This system has issues with wrist offset, leading to unnatural hand and foot joint angles. Problems

are especially evident during jumps, where the foot joints fail to bend correctly as shown in Table 1, row (g), column Radical, and during punches and kicks as shown in Table 1, rows (k) and (l), column Radical, where the limbs do not fully extend.

- **Deep Motion**: Similar to Radical, Deep Motion also struggles with wrist offset, leading to misalignment. It fails to properly recognize punching motions as shown in Table 1, row (k), column Deep Motion, leaving the arm bent instead of extended. Additionally, it shows excessive neck rotation.
- **Plask**: While Plask generally performs well, there is a slight issue with joints being slightly bent during punches as shown in Table 1, row(k), column Plask, although this does not significantly detract from overall accuracy.
- **Rokoko**: Rokoko often misinterprets 2D images when rendered in 3D space, resulting in the entire body being tilted. It also fails to recognize neck rotation and suffers from stiffness in the torso. The accuracy of punches and kicks is notably poor as shown in Table 1, rows (k) and (l), column Rokoko.
- **Movmi**: Movmi shows significant issues, starting with a hunched posture from the shoulders to the head as shown in Table 1, row (a), column Movmi. The joints do not coordinate well, leading to a disjointed appearance. This system struggles to distinguish between distinct poses such as punches and kicks, making it difficult to identify specific actions.

The analysis reveals that OptiTrack, Move One, and Plask are the most accurate in capturing precise poses that closely match the reference video. Radical and Deep Motion show reasonable performance but with noticeable deviations in certain poses. Rokoko exhibits various inaccuracies, and Movmi is the least accurate in maintaining precise poses. This detailed analysis underscores the varying performance levels of markerless MoCap systems, with some showing promise while others still have significant room for improvement.

Table 1. Comparing pose shape accuracy



4.2 Movement Analysis (Jitter Check and Action Smoothness)

Figure 2 illustrates the amount of jitter exhibited by each MoCap AI, focusing on key joint areas where jitter is most noticeable. Jitter is measured by the frequency of unnecessary, sudden movements in each frame. The graph shows that Radical, OptiTrack, Move One, and Deep Motion have the least jitter, indicating that they provide the most accurate and smooth MoCap. Interestingly, Radical exhibits even less jitter than OptiTrack. On the other hand, Movmi shows the highest levels of jitter, particularly in the legs and upper body, highlighting significant areas in need of improvement.

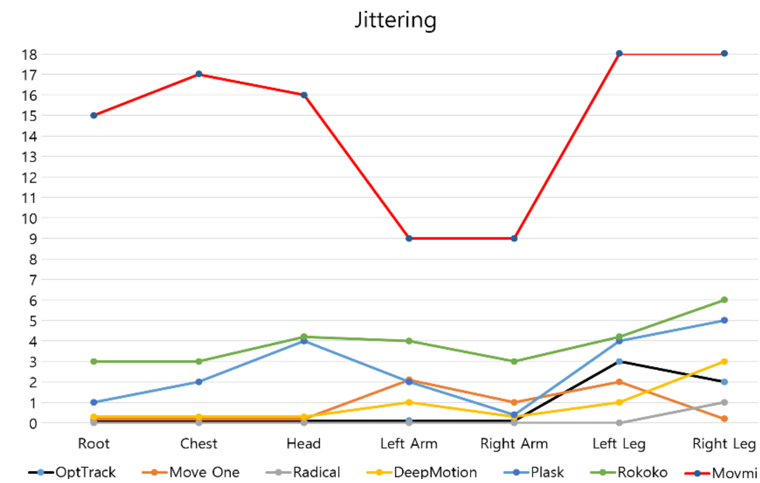


Figure 2. OptiTrack, Move One, Radical, Deep Motion, Plask, Rokoko, Movmi jitter comparison

Figure 3 shows the motion trail of the left arm from a top view, exported from Maya, to visualize the animation movements. Compared to OptiTrack, which delivers accurate results using 24 cameras, this image allows us to assess how closely other MoCap AIs replicate the OptiTrack flow, while also showing the rigidity or smoothness of the character's movements.

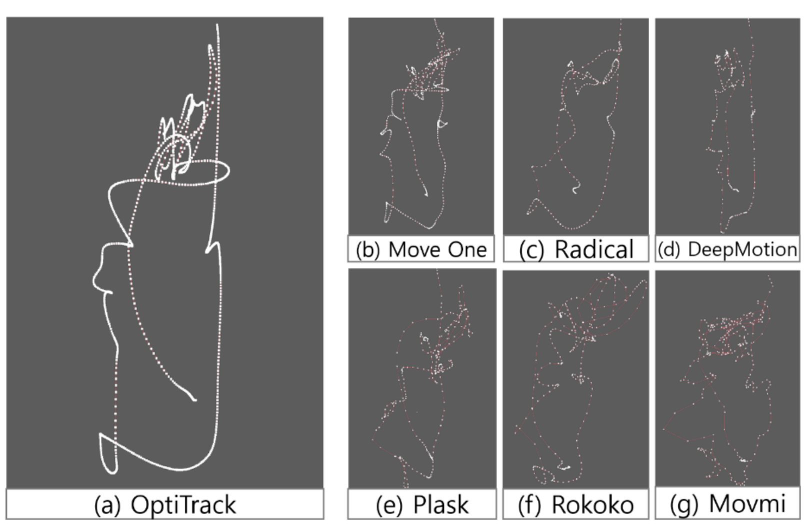


Figure 3. OptiTrack, Move One, Radical, Deep Motion, Plask, Rokoko, Movmi motion flow comparison

These observations suggest that Move One, Radical, and Deep Motion MoCap AIs closely approach the accuracy of OptiTrack. However, Plask, Rokoko, and Movmi still have considerable room for improvement in capturing accurate and smooth motion paths. In particular, Movmi produces some hard, jagged lines in the animation flow, indicating areas that require further enhancement.

4.3 Ground Detection

An essential factor in evaluating MoCap systems is their ability to accurately detect ground contact. This research added a ground plane to test whether each exported MoCap data correctly shows the foot in contact with the ground. First, we tested whether the foot penetrates the ground plane during slight and extreme movements. Then, we examined whether there were any recognizable motions where the feet were off the ground. Finally, we checked for foot sliding along the ground and recorded the results in the chart shown in Figure 4.

OptiTrack and Deep Motion exhibit excellent ground detection with minimal issues, while Move One shows moderate stability, experiencing some slight penetration and sliding. Radical, Plask, Rokoko, and Movmi face significant sliding problems that need to be addressed. Movmi, in particular, demonstrates the highest occurrences of foot sliding and penetration, indicating a need for substantial improvements in its ground detection capabilities.

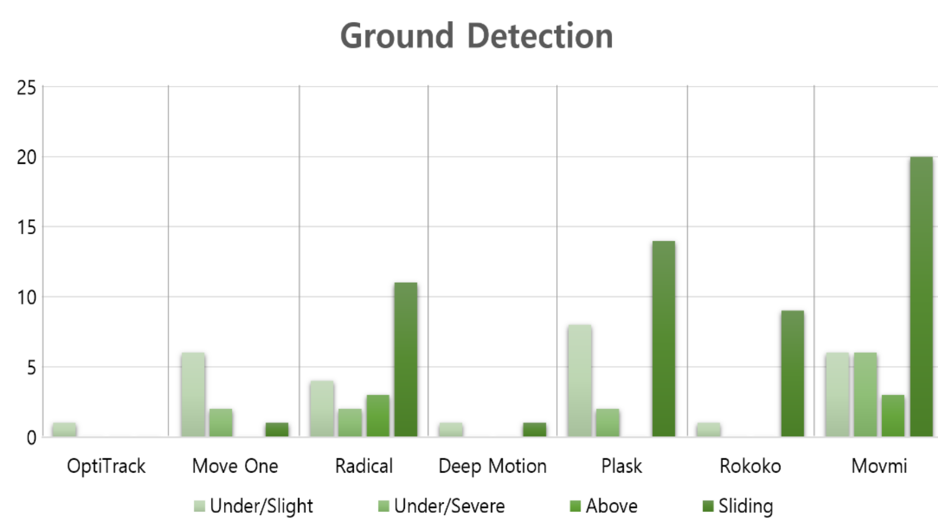


Figure 4. OptiTrack, Move One, Radical, Deep Motion, Plask, Rokoko, Movmi ground detection comparison

5. Discussion

The comparative analysis of various markerles MoCap AI tools against the marker-based OptiTrack system reveals significant insights into the strengths and limitations of these technologies. The findings demonstrate that different markerless MoCap systems excel in various aspects, making the choice of tool highly dependent on the specific needs of the project.

Move One emerges as the best overall choice for markerless MoCap due to its excellent pose shape accuracy, closely matching the reference video, along with above-average smooth MoCap, minimal jitter, and reliable ground detection. This makes it a versatile option for a wide range of animation tasks. For projects where smooth motion with minimal jitter is the primary concern, Radical presents itself as a strong candidate, even outperforming OptiTrack in jitter reduction. On the other hand, if ground detection is the most critical factor, Deep Motion offers the best performance in this area, although it does show some deviations in pose accuracy.

These findings suggest that the selection of a markerless MoCap tool should be guided by the specific priorities of the industry project. Each system has its strengths, and understanding these can help animators and researchers choose the most appropriate tool for their particular needs. While these markerless systems still have room for improvement, particularly in matching the precision of marker-based systems like OptiTrack, they offer viable alternatives depending on the focus of the capture process. This analysis underscores the importance of aligning the tool's capabilities with the project's specific requirements to achieve the best results.

6. Conclusion

We evaluated the effectiveness of AI-driven markerless MoCap tools compared to the traditional marker-based OptiTrack system. Our findings highlight that the choice of markerless MoCap tools should be guided by the specific priorities of a project, whether it's precise pose accuracy, smooth MoCap, or effective ground detection. We identified Move One as a strong overall contender, offering excellent pose accuracy, good motion smoothness, and reliable ground detection, making it a versatile option for various animation tasks. We found that Radical excels in minimizing jitter, making it ideal for scenarios where smooth motion is crucial, while Deep Motion demonstrates the best performance in ground detection, making it suitable for projects where accurate foot placement is a priority.

Although these markerless systems still face challenges in fully matching the precision and reliability of traditional marker-based systems like OptiTrack, we suggest that they present viable alternatives depending on the focus of the capture process. As AI-driven technologies continue to advance, we expect the gap between markerless and marker-based systems to narrow, expanding the potential applications of markerless MoCap in the industry.

Acknowledgement

This work was supported by Dongseo University, “Dongseo Frontier Project” Research Fund of 2023

References

[1] T. Y. Mou, "Keyframe or Motion Capture? Reflections on Education of Character Animation." *EURASIA Journal of Mathematics, Science and Technology Education*, Vol. 14, No. 12, pp. 1-16, Oct 2018. DOI: https://doi.org/10.29333/ejmste/99174

[2] A. Humphrey, J. Adams, and R. Hoetzlein, "Developing Techniques for Rigging and Motion Capture to Simplify 3D Animation." *Aquila–The FGCU Student Undergraduate Research Journal*, pp. 36-45, April 2022. DOI: https://doi.org/10.24049/aq.7.1.6

[3] Y. Z. Sang, K. H. Kim, J. S. Lee, G. H. Zhang, Z. R. Liu, Q. R. Liu, S. J. Sun, Y. T. Wang, and K. X. Wang, "Application of Virtual Studio Technology and Digital Human Monocular Motion Capture Technology-Based on< Beast Town> as an Example." *International Journal of Internet, Broadcasting and Communication 16*, no. 1, pp. 106-123, Feb 2024. DOI: http://dx.doi.org/10.7236/IJIBC.2024.16.1.106

[4] C. Xi, and J. H. Chung, "A Case Study on AI-Driven< DEEPMOTION> Motion Capture Technology." *International Journal of Internet, Broadcasting and Communication 16*, no. 2, pp. 87-92, May 2024. DOI: https://doi.org/10.7236/IJIBC.2024.16.2.87

[5] G. Nagymáté, and R. M. Kiss, "Application of OptiTrack motion capture systems in human movement analysis: A

systematic literature review." *Recent Innovations in Mechatronics 5*, no. 1, pp. 1-9. April 2018 DOI: https://doi.org/10.17667/riim.2018.1/13

[6] D. Thewlis, C. Bishop, N. Daniell, and G. Paul, "Next-Generation Low-Cost Motion Capture Systems Can Provide Comparable Spatial Accuracy to High-End Systems." *Journal of Applied Biomechanics*, Vol. 29, No. 1, pp. 112-117, Feb 2013. DOI: https://doi.org/10.1123/jab.29.1.112

[7] J. Li, H. Gao, F. Zhang, and M. Zheng, "A Method of Human Motion Feature Extraction and Recognition Based on Motion Capture Device." *Transactions on Edutainment XVI*, pp. 184-195, April 2020. DOI: https://doi.org/10.1007/978-3-662-61510-2_18

[8] S. C. Puthenveetil, C. P. Daphalapurkar, W. Zhu, M. C. Leu, X. F. Liu, A. M. Chang, J. K. Gilpin-Mcminn, P. H. Wu, and S. D. Snodgrass, "Comparison of Marker-Based and Marker-Less Systems for Low-Cost Human Motion Capture." *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 55867, American Society of Mechanical Engineers, pp. 1-9, Aug 2013. DOI: https://doi.org/10.1115/detc2013-12653

[9] E. D. Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H. Seidel, and S. Thrun, "Performance capture from sparse multi-view video." *In ACM SIGGRAPH 2008 papers*, pp. 1-10, Aug 2008. DOI: https://doi.org/10.1145/1399504.1360697

[10] Z. D. Hou, K. Hong. Kim, D. J. Lee, and G. H. Zhang, "Real-time markerless facial motion capture of personalized 3D real human research." *International Journal of Internet, Broadcasting and Communication 14*, no. 1, pp. 129-135, Feb 2022. DOI: https://doi.org/10.7236/IJIBC.2022.14.1.129

[11] V. Marek, L. Sigal, J. Hodgins, and O. Jenkins, "Video-based 3D motion capture through biped control." *ACM Transactions On Graphics (TOG) 31*, no. 4, pp. 1-12, July 2012. DOI: https://doi.org/10.1145/2185520.2335378

[12] N. Nakano, T. Sakura, K. Ueda, L. Omura, A. Kimura, Y. Iino, S. Fukashiro, and S. Yoshioka, "Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras." *Frontiers in sports and active living 2: 50*, pp. 1-9, May 2020. DOI: https://doi.org/10.3389/fspor.2020.00050

[13] G. Papandreou, T. Zhu, L. Chen, S. Gidaris, J. Tompson, and K. Murphy, "Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model." *In Proceedings of the European conference on computer vision (ECCV)*, pp. 269-286, Mar 2018. DOI: https://doi.org/10.1007/978-3-030-01264-9_17

[14] S. Scataglini, E. Abts, C. V. Bocxlaer, M. V. Bussche, S. Meletani, and S. Truijen, "Accuracy, Validity, and Reliability of Markerless Camera-Based 3D Motion Capture Systems versus Marker-Based 3D Motion Capture Systems in Gait Analysis: A Systematic Review and Meta-Analysis." *Sensors 24*, no. 11, pp. 1-27, June 2024. DOI: https://doi.org/10.3390/s24113686

[15] L. Wade, L. Needham, P. McGuigan, and J. Bilzon, "Applications and limitations of current markerless motion capture methods for clinical gait biomechanics." *PeerJ 10*, pp. 1-27, Feb 2022. DOI: https://doi.org/10.7717/peerj.12995

[16] M. Menolotto, D.S. Komaris, S. Tedesco, B. O'Flynn, and M. Walsh, "Motion capture technology in industrial applications: A systematic review." *Sensors 20*, no. 19 , pp. 1-23, Oct 2020. DOI: https://doi.org/10.3390/s20195687