Human movements in virtual reality based on motion capture technology

Dániel Nagy
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
nagydani23@gmail.com

Károly Széll
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
szell.karoly@amk.uni-obuda.hu

***Abstract*— Motion tracking is an important aspect of virtual reality. The motion capture systems are the most famous interfaces to navigate and interact in 3D virtual worlds. Virtual reality is starting to blend into our everyday lives. Thus, the need is emerging to integrate the topic into the education. It is needed to teach the next generation of engineers how to use this tool. The paper introduces a motion capture experimental setup for research, development and educational purposes.**

***Keywords*— motion capture, virtual reality, augmented reality, practical education**

I. INTRODUCTION

In today's world, technologies of which digital imaging is an important part are gaining importance. It spans every aspect of life, from health, manufacturing, security, space exploration to the entertainment industry. Digital image processing is the process of transforming digitally stored images into mathematical algorithms to obtain the image information we need.

One of the most important paper on the subject of optical motion analysis is an Italian work [1]–[4] which introduces the basic human motion analysis tools with stereo-photogrammetry (multi-sided photographic optical recording) recording, with special emphasis on errors, their causes and ways to eliminate them.

II. PROBLEM DESCRIPTION

The subject of this paper is the development of a state-of-the-art motion capture (MoCap) experimental setup that can be used for educational, clinical, academic, and sports research purposes. The primary goal is to develop a framework that will allow the education of this engineering field in informatics and to meet emerging industrial development and research needs. The results of the system created are easy to use and scalable for future improvements.

The purpose of the measurements is to record human motion and to animate it in virtual reality. Time, angle and distance parameters calculated from measurement data can be widely used: athletes' techniques and performance can be analyzed, developed and scientific research performed. Biomechanical studies can be divided into two groups: kinematic measurements, which deal with the description and analysis of motions, and kinetic measurements, which study the forces that create and act on motion. This paper focuses on the first area. As technology advances, the requirements for motion testing are also increasing. A competitive system should enable fast, easy and accurate measurements [5]–[16].

Systems operating on an optical (also known as stereophotogrammetry) basis consist of two or more cameras and computer and software for processing their images that coordinate their operation. The basic principle is that the marker image, knowing the position and orientation of the camera that captures the image, defines a straight line in space that connects the camera center to the marker. If a given marker is on more than two camera images, two or more such lines will be defined that intersect at the spatial position of the marker. From the equations of the two lines the required coordinates can be calculated [2].

Unlike ultrasonic and electromagnetic systems, there are non-rigidly mounted sensors (cameras). Their position relative to one another and to the coordinate system of the measurement shall be determined during a calibration procedure. Calibration is usually easy with MoCap software and calibration tools. At the end of the process, the software gives an instrumental error characteristic of the calibration quality.

The cost of building optical-based systems is high compared to other methods, but in return they are highly accurate. Using the right amount of camera, one can create a measuring volume of any size, even outdoors, which can be important when studying for example animal movements. Coverage problems can be minimized by proper layout. The number of cameras that can be used is determined by the MoCap software and the PC hardware running it. There are several types of cameras, from which the user can choose the one that suits his purpose. Their sampling time, accuracy, field of view and angle of view vary from type to type.

Markers for optical-based motion detection systems do not need to be wired as they do not include sensors. The disadvantage of this is that the individual markers are indistinguishable. This could be done with markers of different shapes or sizes, which in turn would significantly increase the resource requirement of image processing, slow down the process and increase the size of the markers (see Fig. 1).

In order to process the information, it is necessary to know to which anatomical point the given data set belongs. This mapping can be done in the software, but the program can only follow this until the marker is displayed on every frame. If the marker is obscured for some reason and is not visible through one or more frames, it will reappear as a new marker after it has reappeared. As a result, lengthy post-processing is required after testing: a competent person must re-identify the marker at each such break. If three or more markers are rigidly connected to each other as rigid bodies, the system can automatically identify them based on the individual distances between the markers (symmetrical layouts should be avoided). These marker assemblies are commonly referred to as rigid bodies. From the position of the markers on the rigid body, the MoCap system can characterize it with one position and one orientation (a single marker only with a single position). The rigidbody is larger in size than the single markers and is therefore not capable of directly tracking anatomical points.

Fig. 1: MoCap in movie Ready Player One [17]

III. Simulation Environment

MoCap is the result of the coordinated work of cameras and target points. As a first step, a room with a solid color background should be provided during the recording, which is not essential for digitizing motion, only reducing the error factor. The error factor during recording is when the cameras cannot see a gesture that causes the skeleton and the moving body to fall apart, resulting in an incomplete or distorted gesture. This can also be caused by poorly configured cameras and the transition of moving body and poor background to nearly the same color. This requires manual adjustment of the recorded gesture, resulting in multiple playback adjustments, tuning, which is time consuming.

Fig. 2: PS Eye camera [16]

We used the iPi Recorder software to record the animation of our gestures, with a free version of 6 devices (camera, microphone, sensor). As a first step we surveyed the available cameras on the market, and we found the PS Eye cameras as the most cost-effective solution [18]–[20]. Not all software is capable of capturing images with these cameras, which is the reason of application of this software, which also supports most cameras on the market.

After installation, we connect the cameras, which will first take a picture of the empty room. The program thus creates the virtual space in which the cameras are located. With the application of a moving target point we can define the animator's space of movement. We used a flashlight for this purpose. The illuminating point was the target point we used to draw a cube around us. After processing this, the movement area was defined. Here we have the option of adjusting the virtual plane to the real plane if the program would have incorrectly delimited the space.

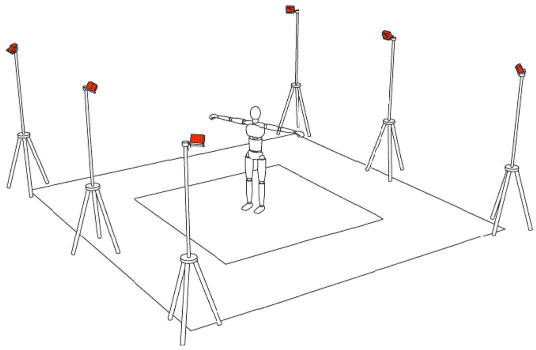

Fig. 3: MoCap room layout [21]

IV. Experimental Setup

After the recording space is calibrated, the animator is calibrated with the T-pose. After a delayed start of the program, the animator picks up the starting position and starts recording the desired movements.

Fig. 4: T-pose

Recording and virtual processing of video image data requires computing capacity on the computer's (VGA) video card module. Even a stable, powerful computer had difficulty processing data with more than three cameras. Thus connecting 6 cameras at the same time to the computer's USB ports failed. Subsequently, two computers were used to record data, where a master and slave were able to create a digitalized motion format using shared synchronized data processing.

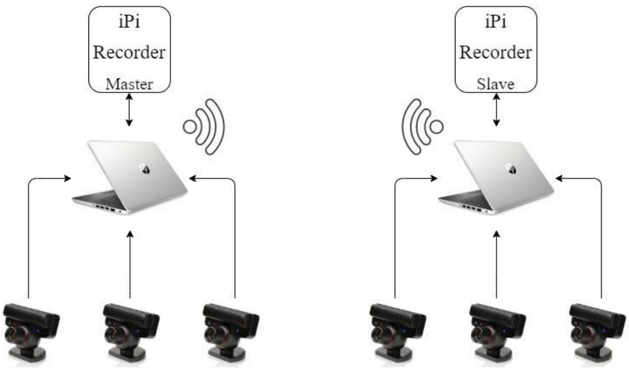

Fig. 5: Hardware architecture

Camera images are processed separately, and the slave sends the finished data to the master which creates a merged version of it. We calibrate the skeleton to the animator with T-pose and then, if possible, look at the results of the first movements at a step by step pace while still being able to adjust the skeleton. We can do this with automatic clean movements as well. Once you've made sure that the skeleton picks up the correct points on the animator and its motion is right for us, then we can start rendering.

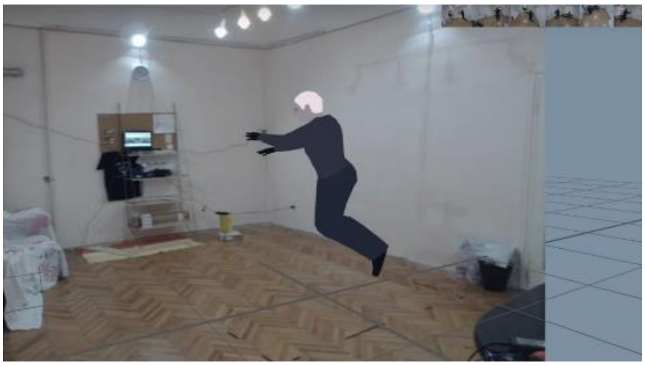

Fig. 6: Animated movement

The resulting animated motion becomes applicable for moving any similar skeleton shape. We integrated the image to move the body created in a 3D graphics program called Autodesk Maya. The resulting animated figure can be placed and controlled in a video game or on the virtual reality training platform we use.

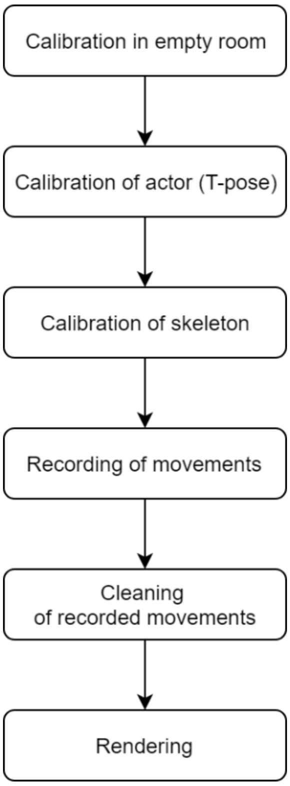

Fig. 7: Process workflow

V. Conclusion

The paper addressed the role of MoCap and VR technology in education. This paper presents a MoCap tool that allows testing and teaching the basic techniques. The measurements are based on six PS Eye cameras and the softwares iPi Recorder and Autodesk Maya. The purpose of our research was to review the area of motion capture technology, to research it and to gain a broad understanding of its depths, how this technology can be applied in healthcare and education. During our research, we built our own MoCap studio to digitize the movements and create an animated character as an educational process for university students.

The technology used in the project enabled us to digitize high-quality movements. Further research is needed to determine the accuracy of the system.

VI. Acknowledgement

The authors wish to thank the support to the Arconic Foundation.

VII. References

[1] A. Cappozzo, U. D. Croce, A. Leardini, and L. Chiari, "Human movement analysis using stereophotogrammetry: Part 1: theoretical background," *Gait Posture*, vol. 21, no. 2, pp. 186–196, 2005.

[2] L. Chiari, U. D. Croce, A. Leardini, and A. Cappozzo, "Human movement analysis using stereophotogrammetry: Part 2: Instrumental errors," *Gait Posture*, vol. 21, no. 2, pp. 197–211, 2005.

[3] A. Leardini, L. Chiari, U. D. Croce, and A. Cappozzo, "Human movement analysis using stereophotogrammetry: Part 3. Soft tissue artifact assessment and compensation," *Gait Posture*, vol. 21, no. 2, pp. 212–225, 2005.

[4] U. D. Croce, A. Leardini, L. Chiari, and A. Cappozzo, "Human movement analysis using stereophotogrammetry: Part 4: assessment of anatomical landmark misplacement and its effects on joint kinematics," *Gait Posture*, vol. 21, no. 2, pp. 226–237, 2005.

[5] G. Lu, Y. Li, S. Jin, Y. Zheng, W. Chen, and X. Zheng, "A real-time motion capture framework for synchronized neural decoding," in *2011 IEEE International Symposium on VR Innovation*, 2011, pp. 305–310.

[6] S. Yonemoto, A. Matsumoto, D. Arita, and R.-I. Taniguchi, "A real-time motion capture system with multiple camera fusion," in *Proceedings 10th International Conference on Image Analysis and Processing*, 1999, pp. 600–605.

[7] T. B. Moeslund and E. Granum, "A Survey of Computer Vision-Based Human Motion Capture," *Comput. Vis. Image Underst.*, vol. 81, no. 3, pp. 231–268, 2001.

[8] F. W. Da Silva, L. Velho, P. R. Cavalcanti, and J. Gomes, "An architecture for motion capture based animation," in *Proceedings X Brazilian Symposium on Computer Graphics and Image Processing*, 1997, pp. 49–56.

[9] Y. Ehara, H. Fujimoto, S. Miyazaki, S. Tanaka, and S. Yamamoto, "Comparison of the performance of 3D

camera systems," *Gait Posture*, vol. 3, no. 3, pp. 166–169, 1995.

[10] Y. Ehara, H. Fujimoto, S. Miyazaki, M. Mochimaru, S. Tanaka, and S. Yamamoto, "Comparison of the performance of 3D camera systems II," *Gait Posture*, vol. 5, no. 3, pp. 251–255, 1997.

[11] B. Michoud, E. Guillou, H. Briceno, and S. Bouakaz, "Real-time marker-free motion capture from multiple cameras," in *2007 IEEE 11th International Conference on Computer Vision*, 2007, pp. 1–7.

[12] A. Andreadis *et al.*, "Real-time motion capture technology on a live theatrical performance with computer generated scenery," in *2010 14th Panhellenic Conference on Informatics*, 2010, pp. 148–152.

[13] X. Zhu and K. F. Li, "Real-Time Motion Capture: An Overview," in *2016 10th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*, 2016, pp. 522–525.

[14] M. Windolf, N. Gotzen, and M. Morlock, "Systematic accuracy and precision analysis of video motion capturing systems--exemplified on the Vicon-460 system.," *J. Biomech.*, vol. 41, no. 12, pp. 2776–2780, 0 2008.

[15] H. Yoshimoto, N. Date, and S. Yonemoto, "Vision-based real-time motion capture system using multiple cameras," in *Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, MFI2003.*, 2003, pp. 247–251.

[16] A. Braun, T. Dutz, M. Alekseew, P. Schillinger, and A. Marinc, "Marker-Free Indoor Localization and Tracking of Multiple Users in Smart Environments Using a Camera-Based Approach," in *Distributed, Ambient, and Pervasive Interactions*, vol. 8028, N. Streitz and C. Stephanidis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 349–357.

[17] "Ready Player One Breakdown reel Archives - Movie VFX." [Online]. Available: https://www.movievfx.net/tag/ready-player-one-breakdown-reel/. [Accessed: 04-Nov-2019].

[18] S.-P. Chao, Y.-Y. Chen, and W.-C. Chen, "The Cost-Effective Method to Develop a Real-Time Motion Capture System," in *2009 Fourth International Conference on Computer Sciences and Convergence Information Technology*, 2009, pp. 494–498.

[19] 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," *J. Appl. Biomech.*, vol. 29, 2012.

[20] J. G. Richards, "The measurement of human motion: A comparison of commercially available systems," *Hum. Mov. Sci.*, vol. 18, no. 5, pp. 589–602, 1999.

[21] "Web-based avatar represented lecture viewer toward interactive e-Lecture performed by 3D avatar - IEEE Conference Publication." [Online]. Available: https://ieeexplore.ieee.org/document/7095984. [Accessed: 04-Nov-2019].