www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 04, April 2023, pp : 519-524

e-ISSN : 2583-1062

Impact Factor : 5.725

MOTION CAPTURING OF A PERSON USING DEEP LEARNING AND CREATING A 3D MODEL

V. Ravi Kishore[1], S. S. Mahendra Reddy[2], M. Bhanu Prasad[3], K. S. S. Sita Reddy[4]

[1]Associate Professor Department of CSE AEC, Surampalem

[2,3,4]Student, Department of CSE AEC, Surampalem

ABSTRACT

The cinema industry invests significant amounts of money into creating CGI, including special effects, creatures, and scenery. This process is expensive due to the need for specialized equipment, such as suits and masks, to capture movements, as well as the requirement for substantial processing power and time investment. An alternative approach to 3D animation, which involves animating limb by limb, is also very time-consuming.For independent creators or those starting game development companies, the cost of specialized equipment is often prohibitive. One solution is to use Computer Vision to capture video footage of the desired action and then convert it into a 3D model using human pose estimation. This process involves determining the location of human body joints and how they are connected.Human pose estimation is an important step in action recognition, scene understanding, and human re-identification, as well as being a crucial component in potential smart camera systems and gaming and filming modeling. By working to design more robust 3D human pose estimators, we can advance this technology and enable further progress in the entertainment industry.

Keywords: Motion Capturing, OpenCV, Human Pose Estimation, Unity Hub, Coordinates, Pose Estimation

1. INTRODUCTION

Motion capture is a technology used to capture the movement of people, animals or objects and apply it to a 2D or 3D model to create animations. Initially developed for gait analysis in the life science market, motion capture is now widely used in various fields, such as VFX studios, sports therapy, neuroscience, computer vision, and robotics. The process involves generating motion capture data from a live actor and importing it into software like Motion Builder to fine-tune character movement.One way to capture 3D motion is to use Human Pose Estimation (HPE), a technique that automatically extracts key points from 2D input and creates 3D rendering. HPE identifies and classifies joints in the human body, capturing a set of coordinates for each joint, which are known as key points. The connection between these key points is known as a pair, and the coordinates of these points can be used to create 3D character animations.To make the process more accessible to small game developers or individuals who can't afford expensive motion capture equipment, this project proposes using deep learning to capture the motion of a person and save their coordinates in each frame. These coordinates can be used to create 3D animations and games without the need for expensive equipment. The project uses OpenCV, CV zone, and media pipeline to estimate 33 landmark points on the human body and describe their pose.

2. LITRATURE SURVEY

Rajendran, Gopika, and Ojus Thomas Lee conducted research on a system that utilizes deep learning algorithms to animate a character in 3D space based on mocap data generated from normal RGB video. They employed Human Mesh Recovery (HMR) scheme to extract mocap data from the input video and transferred the data to Blender for animation. The accuracy of their framework was evaluated using subjective assessment and observation factor.Jingtian, Sun, et al. provided an overview of human pose estimation approaches, including their evolution and differences, using the widely accepted classification of top-down and bottom-up techniques. The paper covered pipelines and seminal studies, and compared and discussed various concepts and techniques, including modern techniques beyond top-down or bottom-up thinking.Z. Chen, Y. Huang and L. Wang conducted research on a system similar to the one in the first paper, which uses deep learning algorithms to animate a character in 3D space based on mocap data generated from normal RGB video. They employed Human Mesh Recovery (HMR) scheme to extract mocap data from the input video and transferred the data to Blender for animation. Their framework was also subjectively assessed using observation factor, resulting in an accuracy value.Suwich Tirakoat examined an optimized motion capture system that can record full-body motion, such as walking, running, and jumping, by changing the quantity and location of cameras and reflect markers. The findings indicated that 4-6 cameras equipped with an Eagle Digital camera, and reflection markers with a minimum of 29 points, including important moving marks and referring markers, were sufficient to record actors' actions in a 10 square meter area. The study's findings could be useful even with limited data mobility.Matej Supej studied a system that uses a GNSS RTK to return a reference trajectory and an inertial sensor-equipped suit to detect subject segment mobility in alpine skiing. The accuracy of the system was tested through various experiments, including a forced pendulum, a walking experiment, a gate position experiment, and a skiing experiment. Segment movement validity was found to depend on the frequency of motion, and the orientation inaccuracy of the motion capture outfit was mostly due to geomagnetic secular fluctuation. The device can measure a full ski course with reduced cost and labor.

3. EXISTING SYSTEM AND PROPOSED SYSTEM

A. Existing system:

There are two primary methods of motion capture: marker-based and markerless. Marker-based systems utilize wearable body suits with IMU sensors and trackers to capture the motion of an object, which is then recorded by cameras and used to create animations for entertainment or scientific purposes. Markerless systems, on the other hand, do not require any special equipment and can be recorded from a video data sequence using motion-based algorithms to track and detect objects.Marker-based systems come in several types, including acoustical, mechanical, and optical systems. Acoustical systems use sound transmitters and receptors to estimate the frequencies of emitters in 3D space, while mechanical systems utilize sliders and potentiometers to display joint locations. Optical systems, the most commonly used in cinema and video game settings, require specially designed suits with reflectors on key articulations to capture motion.Marker-less motion capture has the advantage of being cost-effective and eliminating computational limitations, but it does not capture motion as precisely as marker-based systems. Additionally, some marker-based systems only identify 2D motion and do not create a 3D model of the object being tracked.Proposed system:Utilizing Deep Learning techniques and video as an input format, our system is capable of identifying a person's motion. It accomplishes this by capturing the co-ordinates of each joint during the movement, generating a file with the corresponding data. This process leverages the 33 points available in Human Pose Estimation, each of which represents a joint in the human body. By utilizing these points, we can create a 3D model in various development platforms, facilitating the creation of 3D animations and games by referencing the motion of the individual. Some advantages of this system include a reduction in the time required to capture motion, no cost for motion identification, and accessibility for newly-formed game development startups. Popular development platforms like Blender and Unity can be utilized for the creation of 3D animations.To create a 3D model of a person using motion capture and deep learning, the following steps can be followed. Firstly, the input video needs to be loaded using the OpenCV library. Next, the Human Pose Estimation (HPE) technique can be used to identify the joints and limbs of the person in the video. Once this has been done, motion data can be extracted by comparing the coordinates of each joint between consecutive frames. With this motion data, a 3D model of the person can be generated using the 33 landmark points on the human body identified by OpenCV and CV zone to describe the pose of the person and the positions of the joints and limbs required to be created. Finally, the 3D model can be outputted in a file format that can be used with various 3D modeling and animation software for further processing, such as adding it to a game or creating an animation.

4. BLOCK DIAGRAM

The proposed system includes 7 steps to follow for implementing the motion capture system:Video Capturing: Record the motion of a person using a high-resolution camera and eliminate any background noise.Pose Estimation: An algorithm is used to identify the motion in each video frame and detect the joints of the human body in the video for each second.Temporal Filtering: A temporal filter is used to decrease the noise produced by the 3D pose estimation process. The filter adjusts to different movement speeds using adaptive control, resulting in better frame sequence reconstruction. 3D Co-Ordinates Capturing: After detecting the joints frame by frame, the coordinates for each second are recorded in a text file, which is utilized to construct the model. These coordinates are then processed into Unity Hub to create the 3D model.Skeleton Calibration: The skeleton calibration module is used to determine the lengths of the limbs of the person whose pose is being estimated. This step is critical in order to apply kinematic fitting to a calibrated skeleton, which will adjust the pose prediction to the user's body.Kinematic Fitting: The kinematic fitting module receives the 3D positions from the pose estimator and adjusts them to the user's limb lengths while considering the anatomic constraints of the human body. This is done using kinematic chains that model the skeleton pose with three-dimensional rotations and translations.Model Alignment: The model alignment module arranges the limbs and joints of the model in a position such that it mimics a skeleton model inheriting the motion of the person in the video. This model can then be utilized for animations

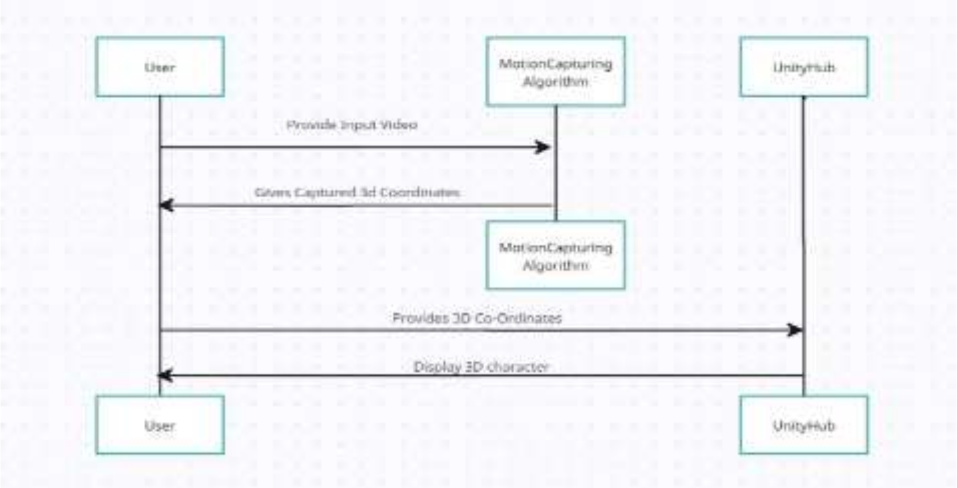

Fig. 1 Block Diagram

5. METHODOLOGY

The process of motion capture involves several important steps. Firstly, data acquisition is performed using a camera or webcam to capture the video. Then, video processing may be done to remove noise and adjust brightness and contrast. Landmark estimation is then used to identify key points on the object, and point extraction is carried out to obtain the coordinates of these landmarks. The resulting coordinate file is imported into Unity, where a wireframe or skeletal structure is created for the model to be animated. Finally, the points are projected onto the model per frame to create the animation.The process flow of motion capture starts by capturing video using a camera or webcam, followed by pre-processing the video to enhance its quality. Next, the video is fed into an algorithm to identify the motions of the person in the video. Landmark points are then tracked using mediapipe by human pose estimation, and the resulting 3D landmark points are saved in a text file. Finally, these points are mapped onto a 3D model in Unity to create an animation that mimics the motion of the person in the video.

6. RESULTS

A. TESTING-

To conduct the testing for this project, the following steps must be taken: Verify video input: The first step is to ensure that the motion capturing algorithm is capable of accepting various types of input videos with different resolutions, frame rates, and file formats. The tester should also ensure that the video quality is adequate to accurately capture the motion of the individual being recorded Test motion capturing: The next step is to test the motion capturing algorithm to ensure that it can accurately detect the motion of the individual in the video and generate 33 3D coordinates per frame. The output of the algorithm should be compared to the actual motion in the video to identify any discrepancies.Verify coordinate transfer: Once the 3D coordinates have been generated, the tester should confirm that they can be transferred to the Unity editor without any loss or distortion. The coordinates generated by the algorithm should be compared to those in the editor to ensure they match.   Test 3D model creation: Finally, the tester must test the 3D model creation process in the Unity editor to ensure that it accurately reflects the motion and actions of the individual in the video. The movements and actions of the model should be compared to those in the video.   Throughout the testing process, the tester must also check for errors, bugs, or inconsistencies that may occur and report them to the development team for resolution. Additionally, documenting the testing process and results is crucial for future reference and to confirm that the system is functioning as intended.

B.TESTCASES

Table 1: Tabular representation of Test Cases

Testcase Id	Testcase Name	Expected Output	Actual Output	Status
1.	First person video given as input	Captured the Motion of Person and created a 3D Model	Captured the Motion of Person and	PASS

			created a 3D Model	
2.	Second person video given as input	Captured the Motion of Person and created a 3D Model	Captured the Motion of Person and created a 3D Model	PASS
3.	Third person video given as input	Captured the Motion of Person and created a 3D Model	Captured the Motion of Person and created a 3D Model	PASS

C.SCREENSHOTS

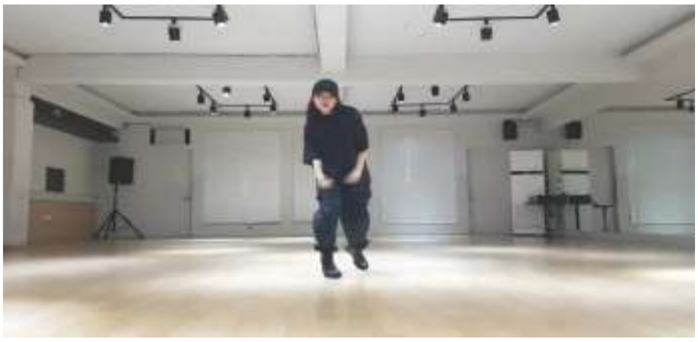

Fig 2: Input Recorded Video

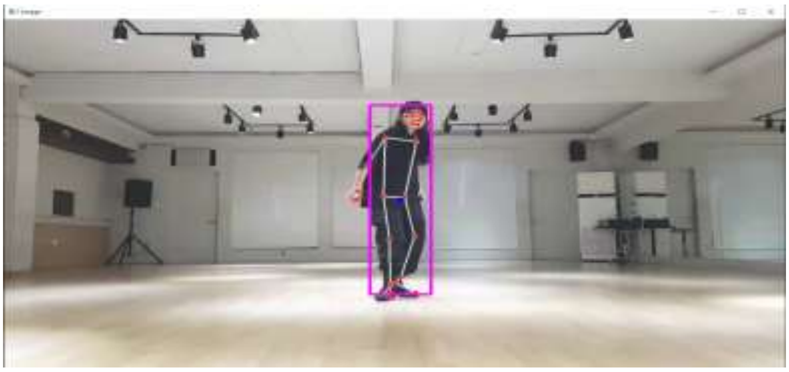

Fig 3: Pose and Motion Detection of Person in the video.

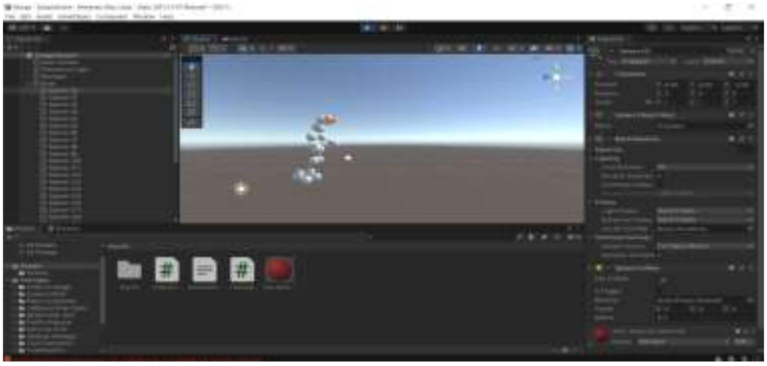

Fig 4: Importing the coordinates into unity Hub

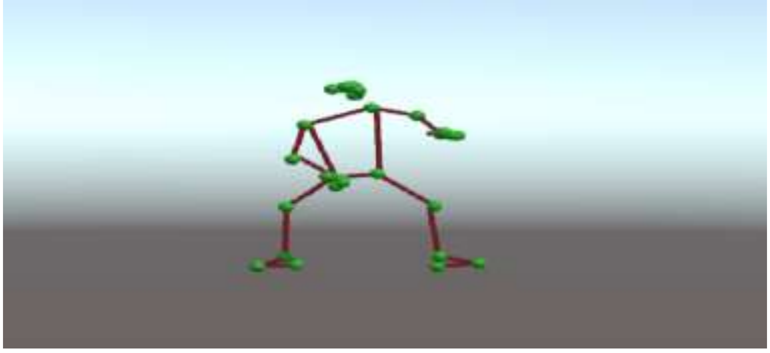

Fig 5: Final 3D Model having the pose and motion as Input Video

7. CONCLUSION AND FUTURE SCOPE

A. Conclusion: Our aim is to create a cost-effective technique to extract motion capturing data from a single videoandapply it to a 3D character animation that replicates human motion. The process includes extracting frames, estimating 3D pose, and mapping the 3D animation in Blender or Unity. The data is then imported for mapping and used as motion capture information to animate the character. The framework uses deep learning techniques for position estimation and open-source tools like Blender and Unity, making it user-friendly and economical. This low-cost motion capture system has the potential to democratize access to motion capture technology. It eliminates the high barriers to entry that are often associated with traditional motion capture methods by using available open-source tools and deep learning techniques.Moreover, the system accurately captures and maps human motion to a 3D character animation, which could have significant implications for various industries such as gaming, film, animation, and virtual reality.In conclusion, this project is a significant step forward in making motion capture technology more affordable and accessible while producing high-quality outcomes. The framework has the potential to transform the digital content creation process, revolutionizing the animation industry and beyond.B.Future Scope:Our objective is to create an affordable technique for extracting motion capturing information from a single video clip and using it to create 3D character animations that imitate human motion. The input data undergoes several procedures, including frame extraction, 3D pose estimation, and mapping of 3D character animation in Unity or Blender, before the animation file is produced. The motion capture data is imported as predicted coordinate values and used to animate the character based on the video input provided. The framework employs deep learning techniques for position estimation and uses free, open-source tools such as Blender and Unity, making it user-friendly and reasonably priced.This low-cost system for motion capture from a single video clip has enormous potential to democratize access to motion capture technology. It eliminates the high entry barriers that typically come with traditional motion capture techniques by utilizing readily available open-source tools and deep learning techniques.In addition, the system's capacity to accurately capture and map human motion to a 3D character animation has significant implications for a variety of industries, including gaming, film, animation, and virtual reality. In summary, this project represents a significant advance in making motion capture technology more accessible and affordable while delivering high-quality results. The framework has the potential to revolutionize the animation industry and other fields by changing how digital content is created.

8. REFERENCES

[1] Rajendran, Gopika, and Ojus Thomas Lee. "Virtual character animation based on data-driven motion capture using deep learning technique." 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS). IEEE, 2020.

[2] Jingtian, Sun, et al. "2D Human Pose Estimation from Monocular Images: A Survey." 2020 IEEE 3rd International Conference on Computer and Communication Engineering Technology (CCET). IEEE, 2020.

[3] Z. Chen, Y. Huang and L. Wang, "On the Robustness of 3D Human Pose Estimation," 2020 25th International Conference on Pattern Recognition (ICPR), 2021, pp. 5326-5332, doi: 10.1109/ICPR48806.2021.9413204.

[4] Tirakoat, Suwich. "Optimized motion capture system for full body human motion capturing case study of educational institution and small animation production." 2011 Workshop on Digital Media and Digital Content Management. IEEE, 2011.

[5] Supej, Matej. "3D measurements of alpine skiing with an inertial sensor motion capture suit and GNSS RTK system." Journal of sports sciences 28.7 (2010): 759-769.

[6] Sharma, Shubham, et al. "Use of motion capture in 3D animation: motion capture systems, challenges, and recent trends." 2019 international conference on machine learning, big data, cloud and parallel computing (comitcon). IEEE, 2019.

[7] S. Schwarcz and T. Pollard, "3D Human Pose Estimation from Deep Multi-View 2D Pose," 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 2326-2331, doi: 10.1109/ICPR.2018.8545631.

[8] A. Singh, S. Agarwal, P. Nagrath, A. Saxena and N. Thakur, "Human Pose Estimation Using Convolutional Neural Networks," 2019 Amity International Conference on Artificial Intelligence (AICAI), 2019, pp. 946-952, doi:10.1109/AICAI.2019.8701267.

[9] T. L. Munea, Y. Z. Jembre, H. T. Weldegebriel, L. Chen, C. Huang and C. Yang, "The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation," in IEEE Access, vol. 8, pp. 133330-133348, 2020, doi:

[10] 10.1109/ACCESS.2020.3010248.L. Zhao, J. Xu, C. Gong, J. Yang, W. Zuo and X. Gao, "Learning to Acquire the Quality of Human Pose Estimation," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 4, pp. 1555-1568, April 2021,doi: 10.1109/TCSVT.2020.3005522.