JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

REAL TIME 3D AVATAR ANIMATION

Akram Khan
Electronics and Computer Science
Rizvi College of Engineering
Mumbai , India

Swaraj Patil
Electronics and Computer Science
Rizvi College of Engineering
Mumbai , India

Danish Koradia
Electronics and Computer Science
Rizvi College of Engineering
Mumbai , India

Fahad Shaikh
Electronics and Computer Science
Rizvi College of Engineering
Mumbai , India

Ramkumar Maurya
Assistant Professor
Electronics and Computer Science
Rizvi College of Engineering
Mumbai , India

ABSTRACT - This paper proposes an efficient, lightweight framework for real-time dynamic avatar animation. Our framework can generate all facial expressions, gestures, and torso movements accurately. The key to our technique lies solely in utilizing the camera sensor of your laptop or an webcam. The magic happens through the combination of various technologies, including MediaPipe, Three.js, and Kalidokit. Specifically, MediaPipe within our framework calculates the 3D posture and facial landmarks of the captured subject. A mapping function then transforms this data to control the avatar's rigid skeleton. This approach eliminates the need for expensive motion capture sensors and the technical expertise required for traditional animation. As a result, our framework offers a simple and user-friendly solution for animating avatars in various applications. By providing open-access to the framework, our aim is to facilitate collaboration and innovation in the field of motion capture technology.

1.INTRODUCTION -

Motion capture (mocap) technology has revolutionized the way we interact and understand human movements. In recent years, the demand for real-time, high-fidelity motion capture solutions has surged, driven by advancements in computing power, sensor technology, and algorithmic innovation. This paper presents a comprehensive exploration of a novel motion capture framework, designed to meet the evolving needs of motion capture enthusiasts and professionals. The framework integrates cutting-edge techniques from computer vision, machine learning, to deliver robust, real-time motion tracking capabilities.

2. REAL TIME AVATAR ANIMATION FRAMEWORK

Our objective is to enable real-time animation of a high-fidelity avatar using only a single camera sensor . The framework comprises three main components:

2a. Motion capture : This involves tracking the models movement utilizing various key point such as for the face, hands, and body. To optimize computational efficiency, we employ cost-effective tracking methods like Blaze Pose.

2b. Skeleton Design : In this part , We use blazepose to create a skeleton that resembles the skeleton of a human body .The BlazePose provides a skeleton style that does not contain a spine, To get the spine information, we must use the four spots on the torso. The raw motion from BlazePose is difficult to process because the wrist has just four points. This could result in an inability to identify fine details, such as whether the back is straight when leaning over.On the other hand, virtual character 3D models usually have different formats (e.g. VRM, FBX, GLTF, etc.), coordinate systems, and other parameters.The Blazepose skeleton mainly contains 12 key bones, namely Hips, Neck, Chest, Spine, RightUpperArm, RightLowerArm, LeftUpperArm, LeftLowerArm, LeftUpperLeg, LeftLowerLeg, RightUpperLeg, RightLowerLeg.

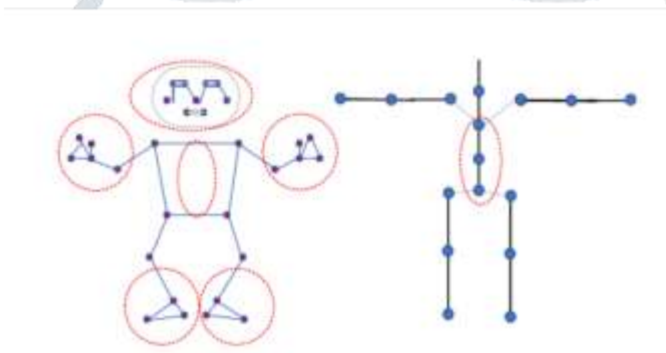


Figure 1.Comparisons between the BlazePose (left) and our proxy skeleton (right). It has simplifed skeleton structure and improved the torso to the standard of rigid animation.

2c . Adaptor Mapping for Avatar Animations :

Through the above measurements, models with different skeletal naming styles can be unified into a consistent naming rule. Our proxy skeleton form is based on the VRM coordinate system. There is no need to convert this type of model. Here, we calculate the rotation matrix to Euler angles. The following formula is used to calculate the rotation matrix *M* to Euler angles.

$$M(x,y,z) = \begin{matrix} f1(x,y,z) \\ f2(x,y,z) \\ f3(x,y,z) \end{matrix}$$

f1(x, y, z), f2(x, y, z), and f3(x, y, z) denote mapping the proxy skele- tons with three different types of avatars, such as VRM, Mixamo etc .

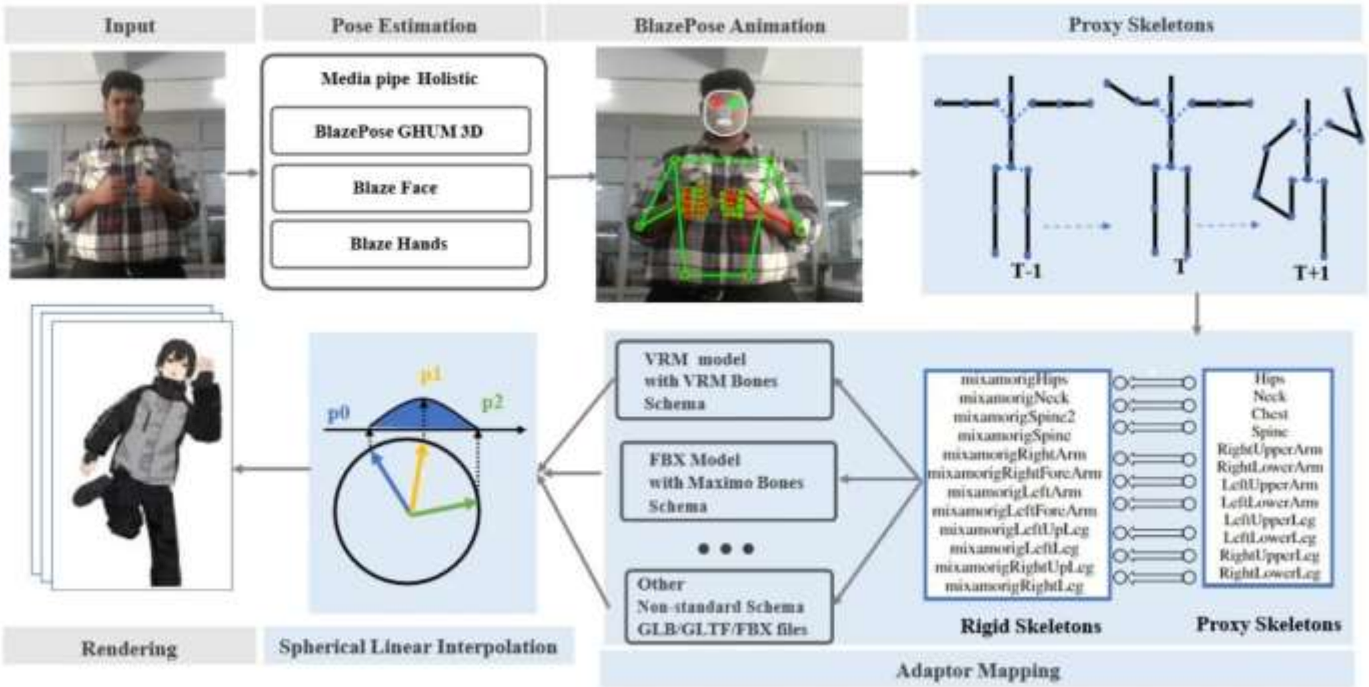


Figure 2: Our unified avatar animation framework consists of three core modules. Firstly, we utilize a single model for estimating key points of the face, hands, and body from monocular video footage through motion capture. Secondly, an adapter is employed to translate the raw motion skeletons into a simplified proxy topology that matches the structure of rigid characters. This process involves creating a simplified kinematic skeleton by excluding hand and head key points from the standard VRM, followed by converting motion data rotation matrices into Euler angles. Finally, virtual characters are animated using a kinematic key points map. This map provides a simplified skeleton representation of key points, compatible with various bone naming conventions and coordinate system standards such as VRM and Maximo.

3. Implementation Details -

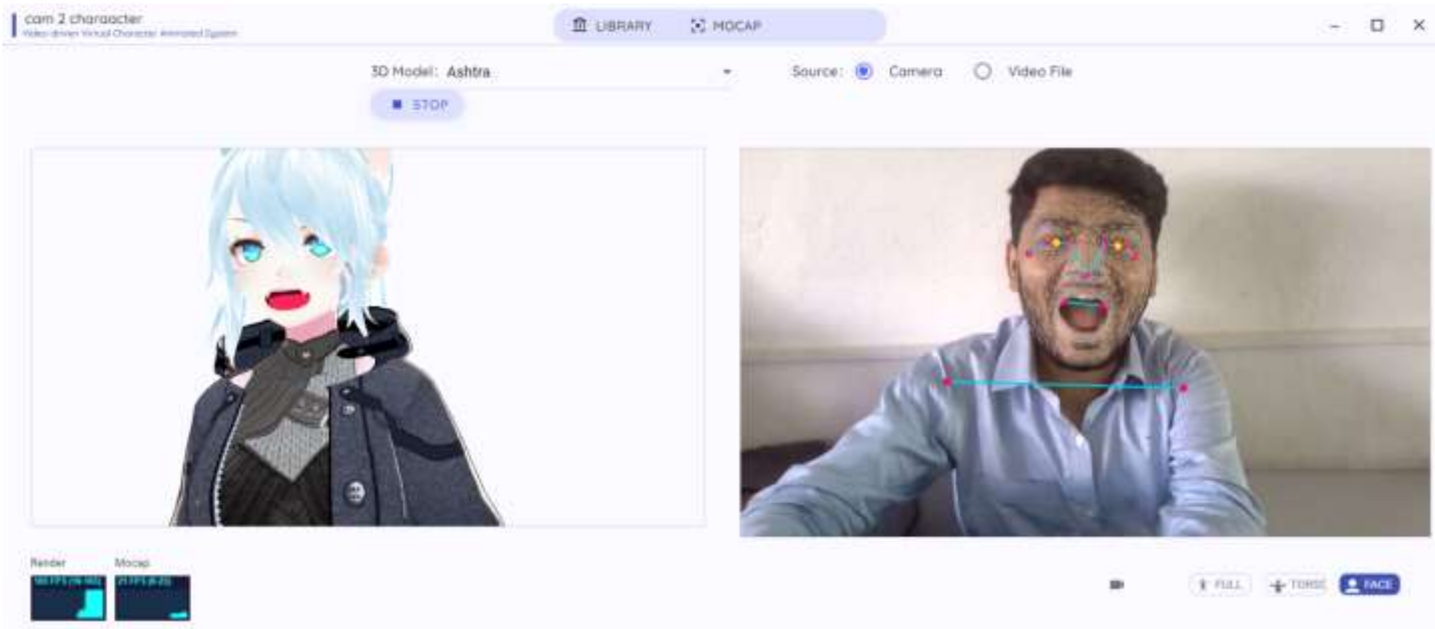
Our desktop application is a comprehensive solution for virtual character management, motion capture, and real-time avatar control. It leverages cutting-edge technology to provide a seamless user experience and powerful functionality.

Motion Capture with Media Pipe and BlazePose GHUM 3D: We utilize Media Pipe, a state-of-the-art machine learning framework developed by Google, to execute the BlazePose GHUM 3D algorithm. This allows us to capture human motion accurately and in real-time. The camera and video file decoder within the application generate events for each frame acquired. These events are then sent to the neural network, where the BlazePose GHUM 3D algorithm calculates and extracts skeletal point positions from the captured data.

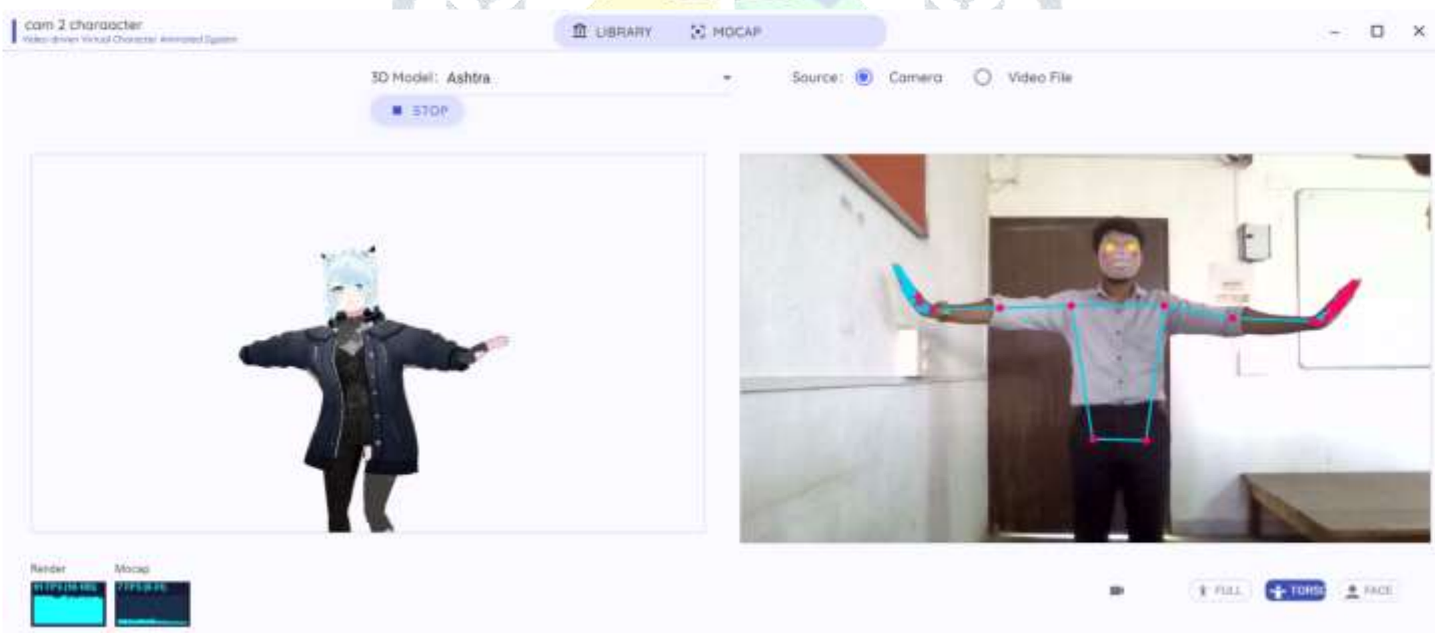
Rendering with Three.js and WebGL:Our rendering functionality is powered by Three.js, a popular JavaScript library for creating and displaying 3D graphics in web browsers. Three.js utilizes WebGL technology, enabling high-performance rendering directly in the browser without the need for plugins. This ensures smooth and responsive visualization of virtual characters and environments.

Support for Various 3D Model Formats: The application includes loaders for various 3D model formats, such as VRM, GLB/GLTF, and FBX. This flexibility allows users to import and work with a wide range of 3D assets, including characters, props, and environments. Whether it's industry-standard formats like FBX or the lightweight GLB/GLTF, our application seamlessly integrates them into the workflow, enhancing creativity and productivity.

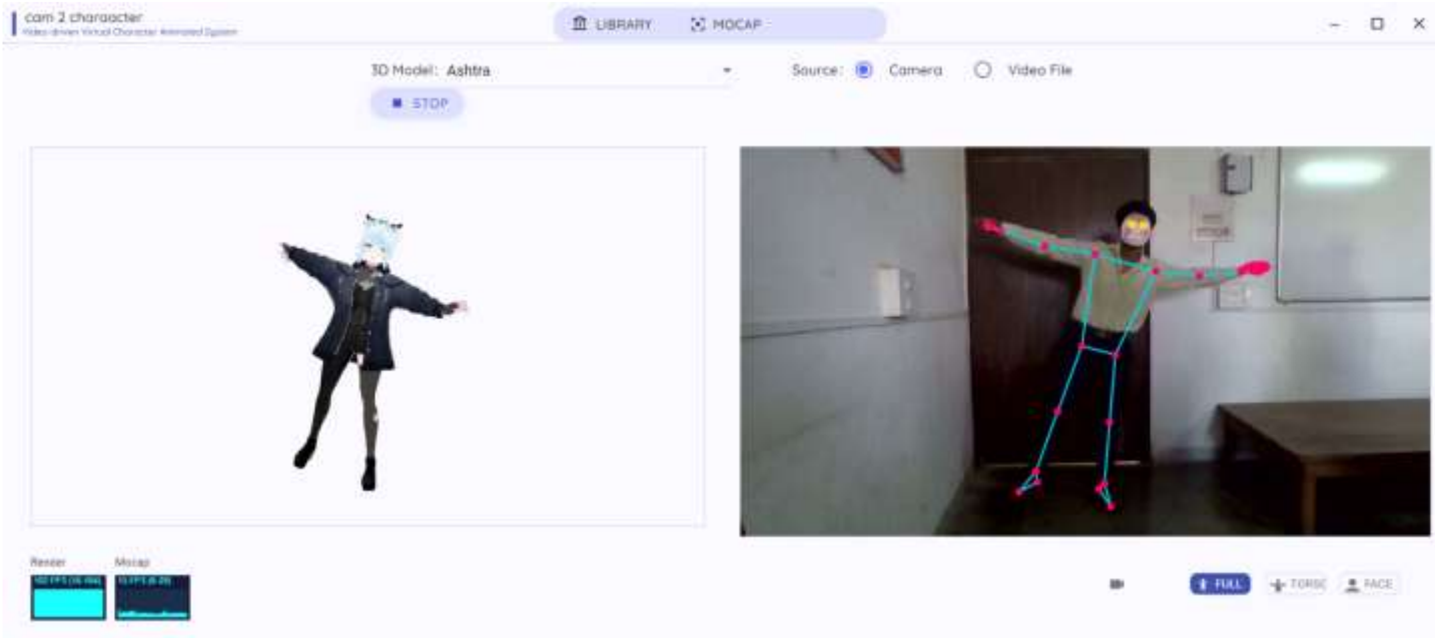
OUTPUT :



Figure(3a) Facial expression and gesture animations



Figure(3b): Torso movement animation



Figure(3c): Full body tracking animation

4. APPLICATIONS – Our proposed framework can generate facial expressions, gestures, and torso movements in real-time, making it useful in the fields of VR/AR animation and entertainment. Here are some potential applications of this project:

- **Virtual Try-On:** The proposed framework can be used to animate virtual avatars in real-time, allowing users to try on virtual clothes and accessories in a more immersive way.
- **Virtual Reality (VR) and Augmented Reality (AR) Applications:** The proposed framework can be used to animate avatars in VR and AR applications, providing a more engaging and interactive experience for users.
- **Telecommunication:** The proposed framework can be used to animate virtual characters in telecommunication applications, allowing for more expressive and engaging communication.
- **3D Games:** The proposed framework can be used to animate characters in 3D games, providing a more realistic and immersive gaming experience.
- **Entertainment Production:** The proposed framework can be used to animate virtual characters in entertainment production, allowing for more expressive and engaging performances.
- **Virtual Anchors:** The proposed framework can be used to animate virtual anchors in live streaming and recording applications, providing a more engaging and interactive experience for viewers.
- **Virtual Fitting Rooms:** The proposed framework can be used to animate virtual avatars in virtual fitting rooms, allowing users to see how clothes fit and move on their virtual bodies.
- **Virtual Assistants:** The proposed framework can be used to animate virtual assistants in various applications, providing a more engaging and interactive experience for users.

Overall, the proposed framework has a wide range of applications in various fields, including VR/AR animation, entertainment, telecommunication, and gaming.

5. CONCLUSION -

This research introduces a novel framework for animating avatars in real-time. It prioritizes efficiency and fidelity while remaining cost-effective. Unlike traditional motion capture methods that require expensive equipment and specific body configurations, this system utilizes only a single camera. While the current implementation excels in efficiency, there's room for improvement. The accuracy of pose estimation, particularly in scenarios with significant obstructions in the video feed, can be enhanced. Additionally, complex gestures and finger movements may not be captured with the same level of precision as other motion capture solutions. However, the system effectively conveys the user's overall movements.

6. REFERENCES

1. Alterecho: Loose avatar-streamer coupling for expressive vtubing. In 2021 IEEE Inter- national Symposium on Mixed and Augmented Reality (ISMAR), pages 128–137. IEEE, 2021.
2. Breannan Smith, Chenglei Wu, He Wen, Patrick Peluse, Yaser Sheikh, Jessica K Hodgins, and Takaaki Shiratori. Constraining dense hand surface tracking with elasticity. ACM Transactions
3. 3.Abhishek Sharma and Eric Rombokas. Improving imu-based prediction of lower limb kinematics in natural environments using egocentric optical flow. *IEEE Transactions on Neural Systems and Rehabilitation Engineering (TNSRE)*, 30:699–708, 2022.
4. 4.Ramakrishna Perla and Ramya Hebbalaguppe. Google cardboard dates augmented reality : Issues, challenges, and future opportunities, 2017.
5. 5.Ruizhi Shao, Hongwen Zhang, He Zhang, Mingjia Chen, Yanpei Cao, Tao Yu, and Yebin Liu. Doublefield: Bridging the neural surface and radiance fields for high-fidelity human reconstruction and rendering. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2022.
6. 6. Zhe Li, Tao Yu, Zerong Zheng, Kaiwen Guo, and Yebin Liu. Posefu- sion: Pose-guided selective fusion for single-view human volumetric capture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14162–14172, 2021.
7. Ammar Qammaz and Antonis A Argyros. Towards holistic real-time human 3d pose estimation using mocapnets. In British Machine Vision Conference (BMVC). BMVA, November 2021.
8. Ammar Qammaz and Antonis A. Argyros. Occlusion-tolerant and personalized 3d human pose estimation in rgb images. In IEEE In- ternational Conference on Pattern Recognition (ICPR), (to appear), January 2021.
9. Ammar Qammaz and Antonis A Argyros. Mocapnet: Ensemble of snn encoders for 3d human pose estimation in rgb images. In British Machine Vision Conference (BMVC), Cardiff, UK, September 2019. BMVA.
10. Ramakrishna Perla and Ramya Hebbalaguppe. Google cardboard dates augmented reality : Issues, challenges and future opportunities, 2017.