International Journal of Internet, Broadcasting and Communication Vol.16 No.1 106-123 (2024)
http://dx.doi.org/10.7236/IJIBC.2024.16.1.106

IJIBC 24-1-14

Application of Virtual Studio Technology and Digital Human Monocular Motion Capture Technology -Based on <Beast Town> as an Example-

YuanZi Sang, KiHong Kim, JuneSok Lee, JiChuTang, GaoHe Zhang, ZhengRan Liu, QianRu Liu, ShiJie Sun, YuTing Wang, KaiXing Wang

Doctor, Department of Visual Contents, Dongseo University, China.
Professor, Department of Visual Contents, Dongseo University, Korea.
Professor, Department of Software Contents, Dongseo University, Korea.
Master, Department of Visual Contents, Dongseo University, China.
Doctor, Department of Visual Contents, Dongseo University, China
Master, Department of Visual Contents, Dongseo University, China.
Master, Department of Visual Contents, Dongseo University, China.
Master, Department of Visual Contents, Dongseo University, China.
Doctor, Department of Visual Contents, Dongseo University, China.
Master, Department of Visual Contents, Dongseo University, China.

syzcc0306@gmail.com

Abstract

This article takes the talk show "Beast Town" as an example to introduce the overall technical solution, technical difficulties and countermeasures for the combination of cartoon virtual characters and virtual studio technology, providing reference and experience for the multi-scenario application of digital humans. Compared with the live broadcast that combines reality and reality, we have further upgraded our virtual production technology and digital human-driven technology, adopted industry-leading real-time virtual production technology and monocular camera driving technology, and launched a virtual cartoon character talk show - "Beast Town" to achieve real Perfectly combined with virtuality, it further enhances program immersion and audio-visual experience, and expands infinite boundaries for virtual manufacturing.

In the talk show, motion capture shooting technology is used for final picture synthesis. The virtual scene needs to present dynamic effects, and at the same time realize the driving of the digital human and the movement with the push, pull and pan of the overall picture. This puts forward very high requirements for

Manuscript Received: December. 21, 2023 / Revised: January. 12, 2024 / Accepted: January. 22, 2024
Corresponding Author: syzcc0306@gmail.com
Tel: +82-10-8379-3060
Doctor, Department of Visual Contents, Dongseo University, China.

multi-party data synchronization, real-time driving of digital people, and synthetic picture rendering. We focus on issues such as virtual and real data docking and monocular camera motion capture effects. We combine camera outward tracking, multi-scene picture perspective, multi-machine rendering and other solutions to effectively solve picture linkage and rendering quality problems in a deeply immersive space environment. , presenting users with visual effects of linkage between digital people and live guests.

Keywords: *Monocular Motion Capture, Virtual Studio, Virtual Shooting, Virtual character production*

1. Introduction

In the digital age, the deepening of intelligence and informatization has provided impetus for the iteration of film technology. How to further optimize the virtual production process of live animation, reduce the work tasks of compositors, animators and editors, and release the directors, photographers, actors and other creative directors The creativity and imagination of personnel are core issues that need to be solved urgently to produce high-quality special effects movies. This article combines the actual shooting process of the talk show "Beast Town" to study the specific technologies, processes and specifications of virtual broadcasting. By analyzing the role and impact of motion capture technology on the production of virtual animation images, this article sorts out the methods for shooting this type of image content. Feasible technical solutions, with a view to promoting the establishment of standardized processes and demonstration applications for the same type of image production.

The research process of this project is shown in Figure 1.

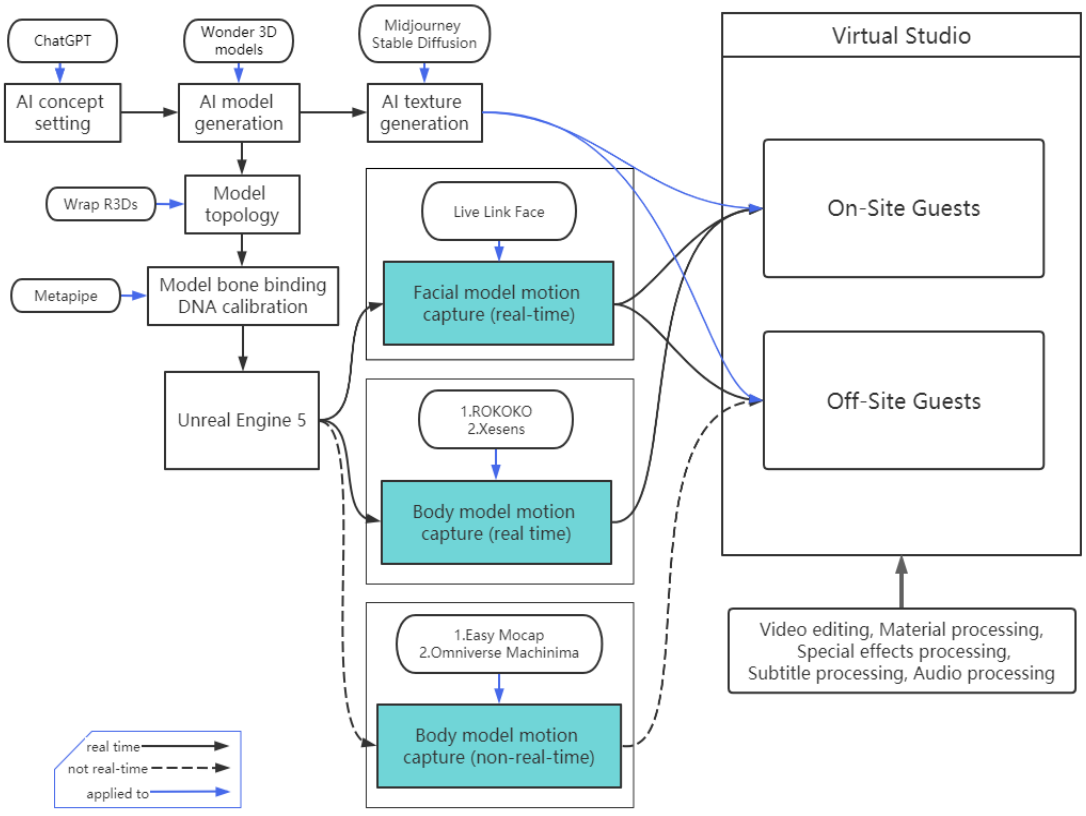

Figure 1. <Beast Town> Flow Chart

2. Motion Capture Technology

Common motion capture technologies are optical capture and inertial motion capture. The former relies on a calibrated volume camera, while the latter is an independent system [1]. Motion capture technology is widely used in entertainment, sports, military, robotics, and other fields [2]. Based on inertial motion capture technology, the natural demonstration movements of the human body are used to synchronously control virtual characters, and the virtual action characters are used as real-person stand-ins to achieve real-time action creation and the establishment of a live broadcast platform.

In the late 1990s, computer animation (Computer Animation) technology gradually matured and was widely used in film production. As a result, virtual characters and spectacle scenes in special effects movies began to make their debut, and impacted traditional special effects makeup. industry. Special effects makeup artist Nick Dudman once said: “In the past, we only used traditional makeup methods and prosthetic props to handle special effects. However, when complex monster special effects became popular, we had to learn how to use invisible effects.

Wire control, equipment and other related technologies. When computers started creating magical monsters for movies—think Jurassic Park—we started to worry about how long makeup artists' actual makeup skills would last. In fact, while it may seem like computers will take over all traditional areas of practical makeup, we are now finding that there is room for collaboration between the two - both computer stunts and practical makeup technology can complement each other. "[5] This is indeed the case. Motion capture was only widely used in the field of film and television in the 21st century and became one of the core technologies of special effects movies. The emergence of motion capture technology improved the animation effects and production efficiency of subsequent special effects movies. Compared with traditional stunts, they have been qualitatively improved.

2.1 Traditional Optical Motion Capture

Sensor : bule trident, Set up hardware devices: Install the hardware components of the Vicon system, motion recognizers, including cameras (Valyrie) and human wear sensors (bule trident). Cameras are typically positioned within a room or area to capture the desired motion.

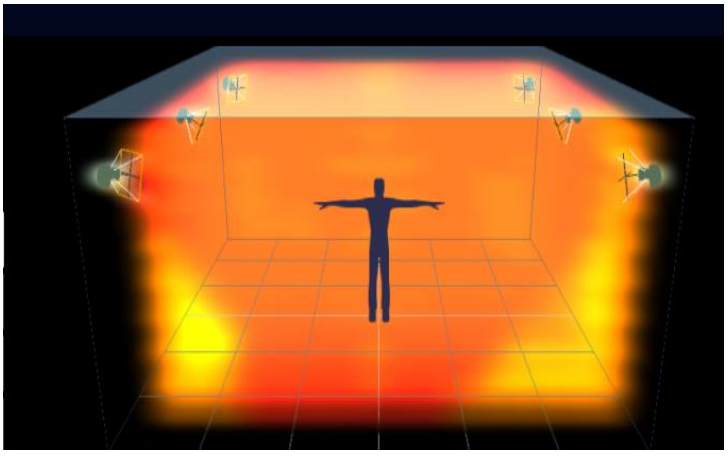

Figure 2. vison nexus

Calibrating the system: Before performing motion capture, the Vicon system needs to be calibrated using (vison nexus). At the same time, make sure to ensure the data collected by the camera, use mask to block unnecessary environmental detection sources to ensure that they can accurately capture the position of the marker point, establish a coordinate system, and modify the launch file.

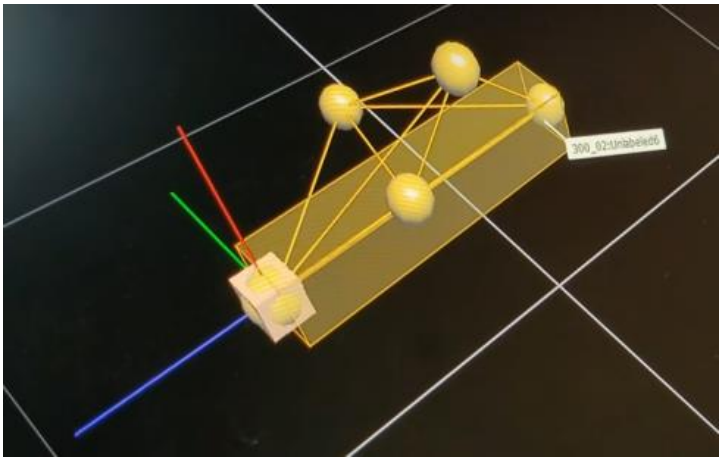

Figure 3. Create coordinate system

```
<launch>
        <node name="viconros" pkg="viconros" type="viconros" output="screen">
        <param name="host" value="192.168.2.100:801"/>
        <param name="model" value="quadmodel"/>
        <param name="segment" value="whole"/>
        </node>
        <node name="xbee_send" pkg="xbee_com" type="xbee_send" output="screen"/>
        <!--node name="record" pkg="rosbag" type="record" args="record -O /home/chengque/rec.bag -a "/-->
</launch>
```

Figure 4. Modify launch file

Capturing motion: Once the system is set up and calibrated, you can start capturing the desired motion. The marked object or human body moves within the camera's field of view, and the camera will record the movement of the marked point.

Data Analysis: Captured data will be recorded and stored via the Vicon system (amtinetforce). You can use the software provided by Vicon to analyze this data to obtain detailed information about the movement.

Data Applications: Motion captured data can be used in a variety of applications such as movie special effects, video game development, sports science research, biomedical research, etc. Depending on your specific project needs, you can use your data in different areas.

Advantages: Accuracy, film-level action, multiplayer

Disadvantages: large space, cumbersome operation, too much real-time cost, expensive

2.2 ROKOKO

The motion capture suit is put on according to the bone sensor to align the joints of the human body.

Figure 5. RoKOKO

Enter rokoko and use body USB to connect to the host and calibrate the clothing and host (must be connected to the same wifi, enter the IP address of the host and confirm whether the clothing version is updated to the latest version)

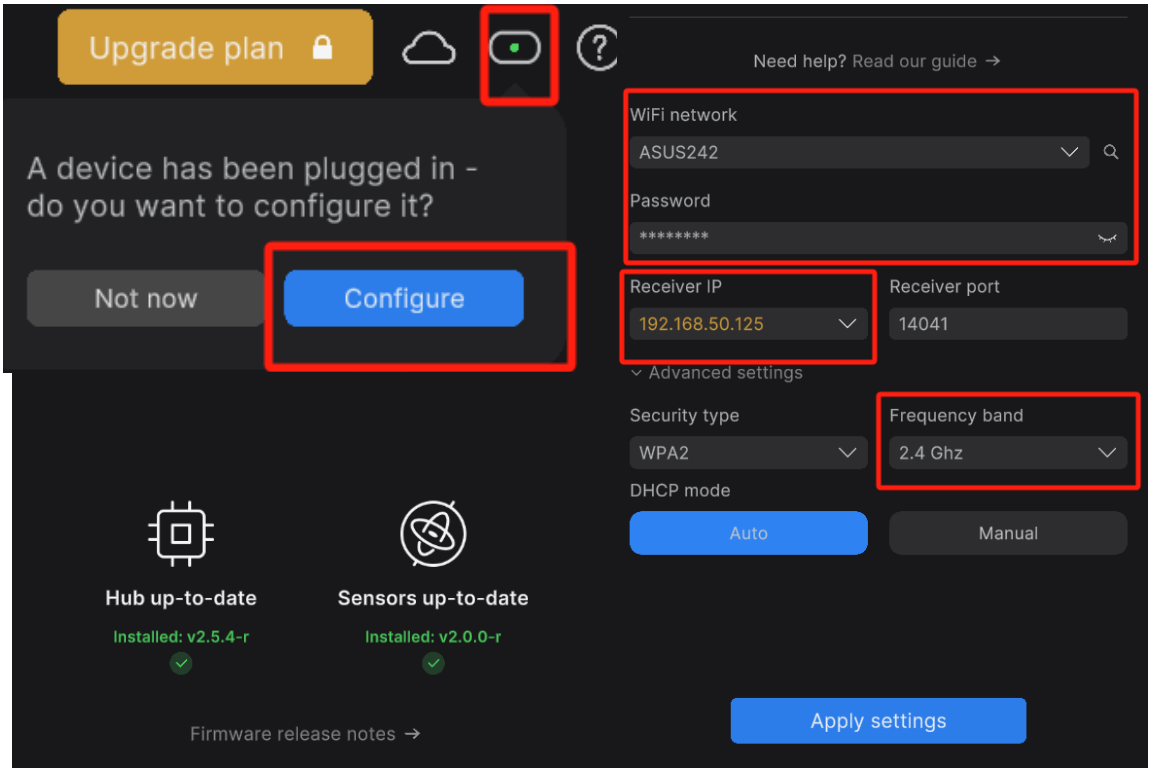

Figure 6. Enter rokoko

The Rokoko interface sets the height of the required motion capture actor and the measurement length of each part of the body to avoid mold wear when the character moves.

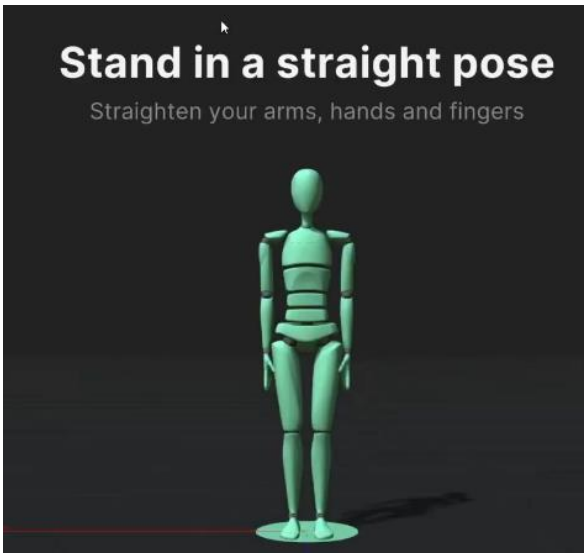

Figure 7. Stand in a straight pose

Figure 8. The actor stands upright and begins the calibration

After pressing the play button, the actor can freely make the required actions and the animation will be produced. After pausing, complete the animation production. Set the check box on the right side of the interface to another software you want to export (for example: UE5, MAYA, Unity, etc.)

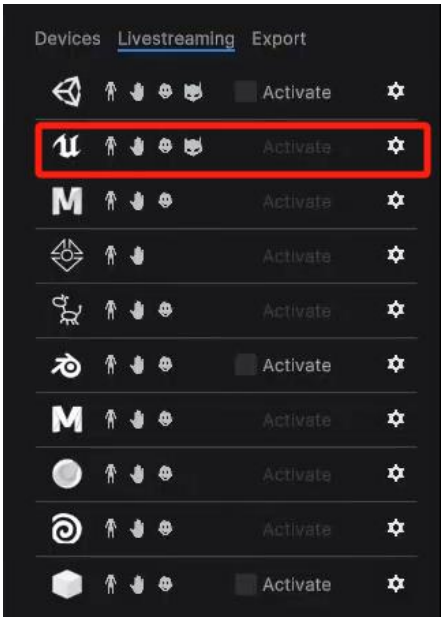

Figure 9. Export settings

2.3 EasyMocap

After the project is deployed, 4-8 cameras are deployed and the environment is deployed. After each camera is calibrated on the chessboard, the xy axis of each camera is established, and then the z-axis vertical calibration is performed.

Intrisic parameter contains K, dist.

\[K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix},\] \[\mathbf{dist}=(k_1, k_2, p_1, p_2, k_3).\]

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{dist} = (k_1, k_2, p_1, p_2, k_3).$$

Figure 10. Intrisic parameter

The definition of dist can be found in OpenCV,
Extrinsic parameter contains R,T.
\[X_{cam} = R X_{world} + T.\]

$$X_{cam} = RX_{world} + T.$$

Figure 11. Extrinsic parameter

In a motion capture system, there are several coordinate systems. One is the world coordinate system V(w), which is a coordinate system that the attitude capture module can directly capture and output. It represents the attitude of the object in the three-dimensional world; due to the different hardware of the sensor, the world coordinate system is represented by the output of the sensor. V(w) can be different; Vi(w)(i=1,2,...n) can be used to represent the world coordinate system determined by each sensor. The other is the coordinate system where the captured object is located. The coordinate system of each object may be different. The coordinate system of each captured object can be simply defined as a unified coordinate system, that is, the coordinate system. V(o). In addition, the posture capture module needs to be installed on the captured object to capture the movement of the object. Since when installing the posture capture module, it is not guaranteed that the tilt angle and orientation of each posture capture module are the same. Therefore, the captured object and There is an installation coordinate system V(a) between the installed attitude capture modules. The installation coordinate system of each captured object is different. Vi(a)(i=1,2,...n) can be used to represent the installation coordinate system of the i-th object.

Figure 12. EasyMocap

After setting the internal and external parameters of the camera, process the video and perform multi-angle action extraction in the cmd command box. At the same time, we noticed that in terms of identification of character IDs, only four cameras are guaranteed to be calibrated to a complete person ID. There will not be too

much movement, otherwise the character ID will be confused, resulting in confusion of the skeletal information. After exporting the motion capture data, you will get a json file, and then use blender to convert the json file to bvh file, and assign the information to the model.

3. Production process

3.1Model

3.1.1. Making models

In our project, we are making the virtual anchor, which includes character design and three-view drawing. We use a range of specialist tools to achieve this. First, we used tools such as Midjourney to design the characters, including appearance, clothing, facial features, etc. These design requirements engage our audience while being consistent with our content and themes. In addition, we also used tools such as Stable Diffusion's Contornet plug-in to create three views of the character, including front, side and back views, to ensure that the character could be accurately drawn and presented at different angles. During this process, we paid attention to detail definition and color design to maintain consistency and appeal.

Figure 13. character conception

3.1.2 Modify the model

To modify the model to match the vertex data of the character, the topology of the model needs to be adjusted to ensure that the number of vertices matches the number of faces and structure of the model, which provides the basis for the high fidelity and smoothness of digital characters. However, in some cases, as described in this article, we need to deal with unusual structures, such as incorporating dog ears into the model.

Integrating it directly into the topology can cause routing and binding issues. To solve this challenge, we chose to detach the ears from the head and reintegrate them after the topology work was completed. In this process, the ear is treated as a wearable device on the head. This avoids changing the rhythm of the original wiring and allows the ear to be perfectly integrated in the model.

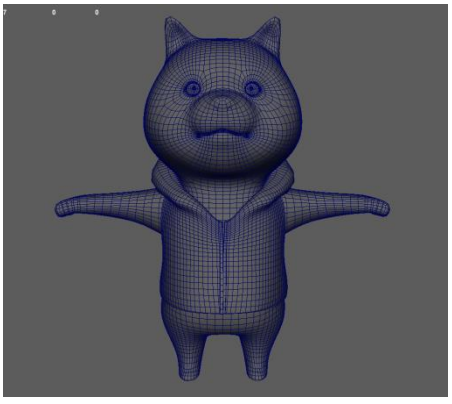

Figure 14. character model

Also confirm the UV information of the model.

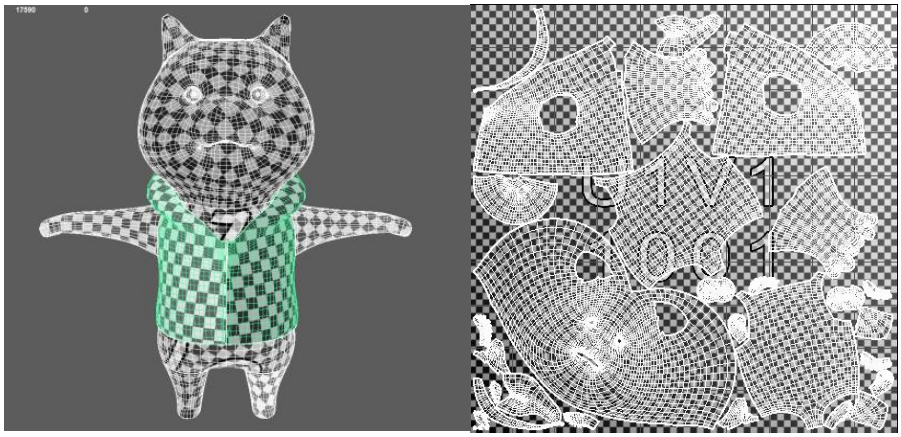

Figure 15. Model UV

3.2.Rigging

3.2.1 Build motion capture skeleton data

Use advanced skeleton to bind bones. Before binding, confirm whether the wiring of the model conforms to the trend of movement. Use weight drawing to modify the Blend shape of the bones to obtain a better motion model.

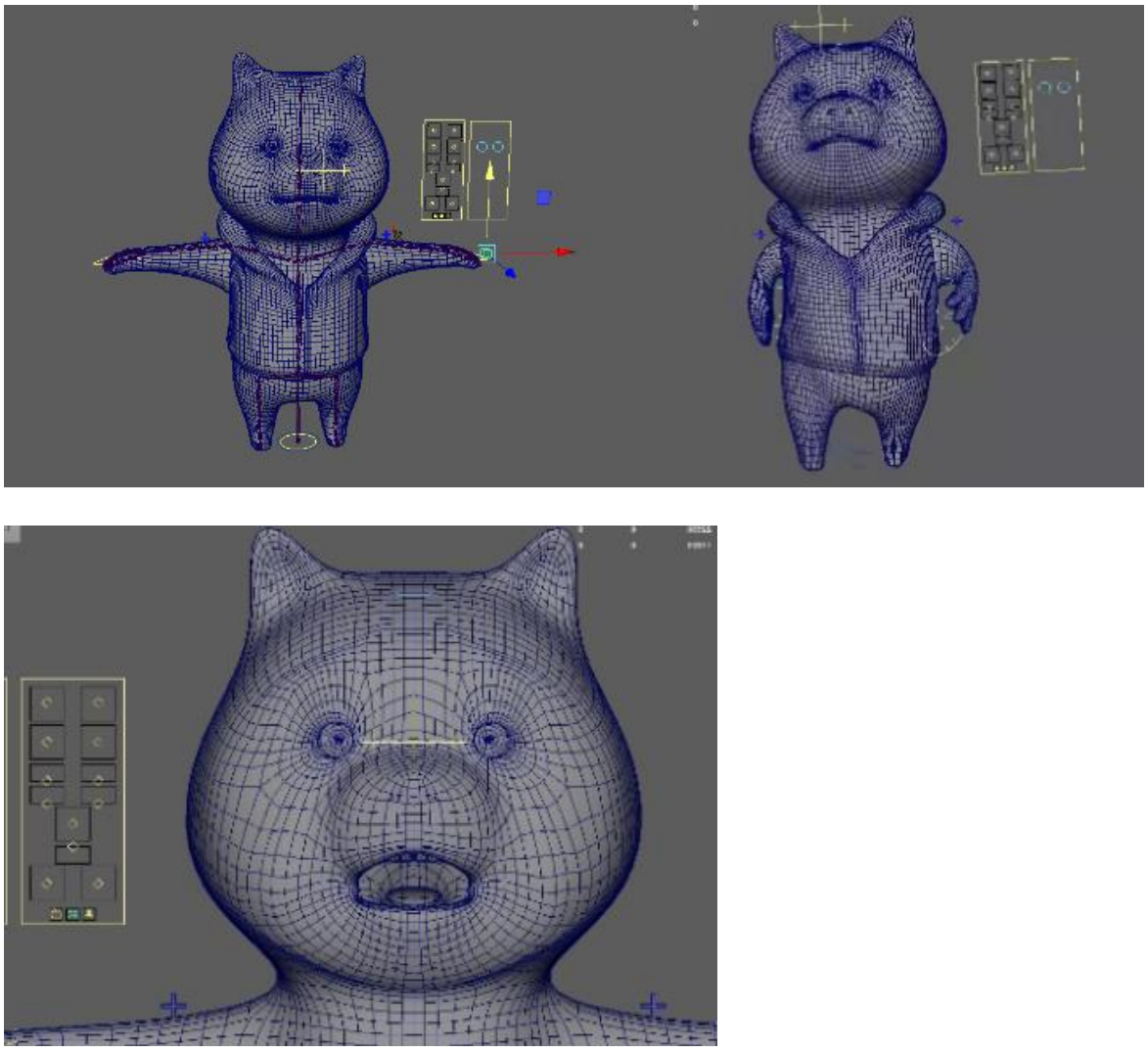

Figure 16. Model facial motion capture

The files produced are skeletal model fbx files and a phython control rig control file.

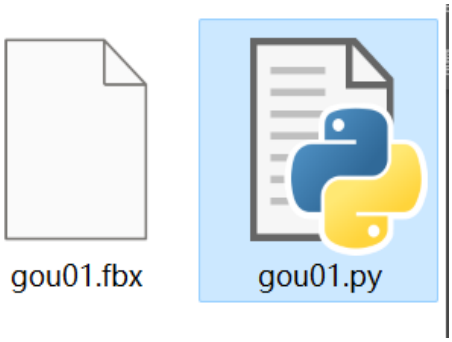

Figure 17. Generated Files

3.2.2. Modify DNA information

After modifying the DNA information, the expression of the DNA model obtained after modifying the DNA information will still be distorted at certain times, so we use Yanustudio's plug-in to modify the blendshape of each controller in Maya to improve the effect of each controller. Improve the naturalness during motion capture. After getting the model, you can perform dynamic compensation.

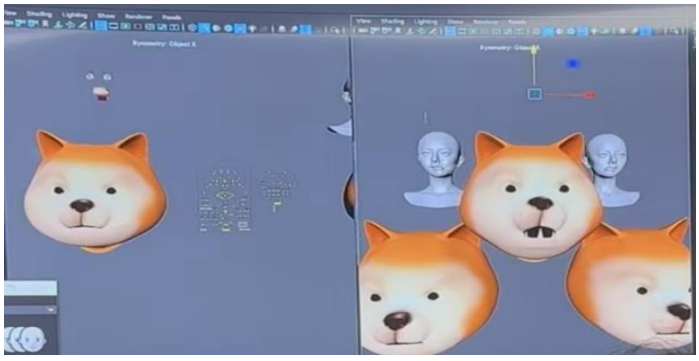

Figure 18. Modify DNA information

3.3. Role to UE

3.3.1. Character import UE

When importing the UE, I got that because the expression information on the head is produced using blend shape, the target deformation option needs to be checked.

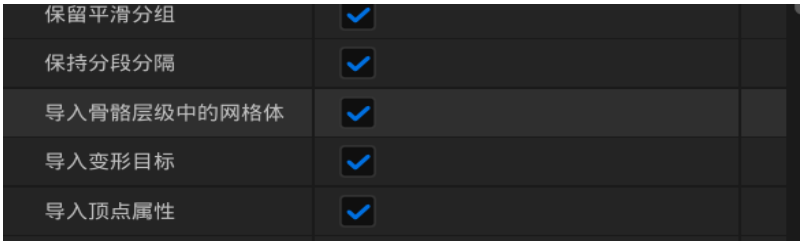

Figure 19. Character import UE

Since the included facial expression control is made using blendshape, you also need to import the phthon file into the engine. After selecting the bone target, choose to create a controlrig and import the python file into the target deformation.

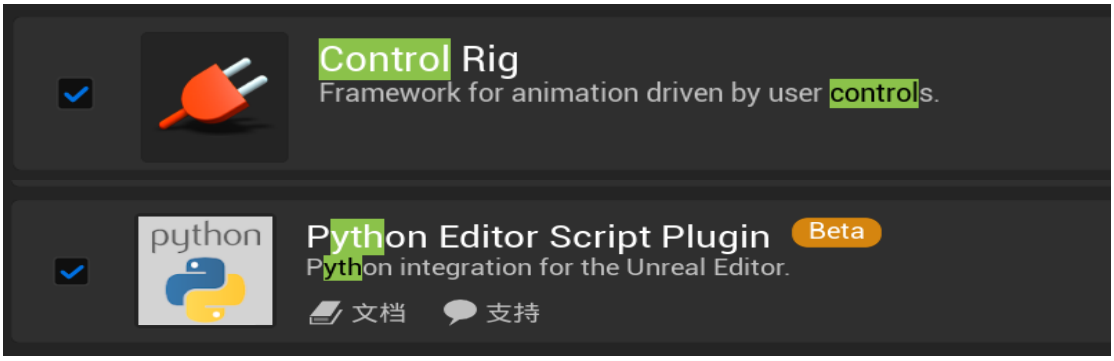

Figure 20. open the engine's controlrig and python plug-in.

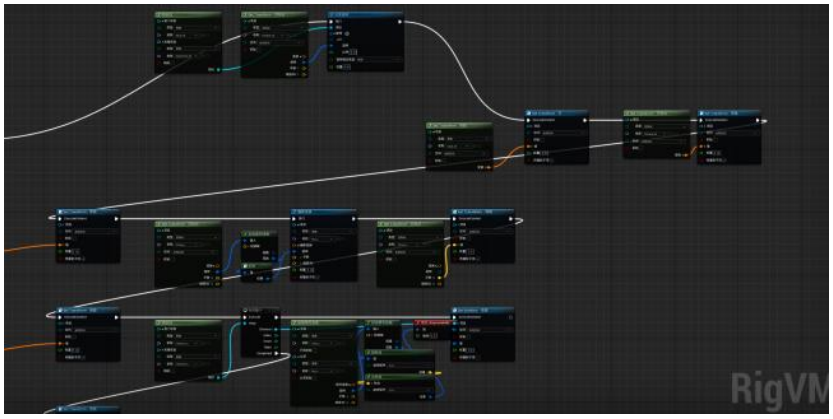

Figure 21. Drag the target python file into the target binding control.

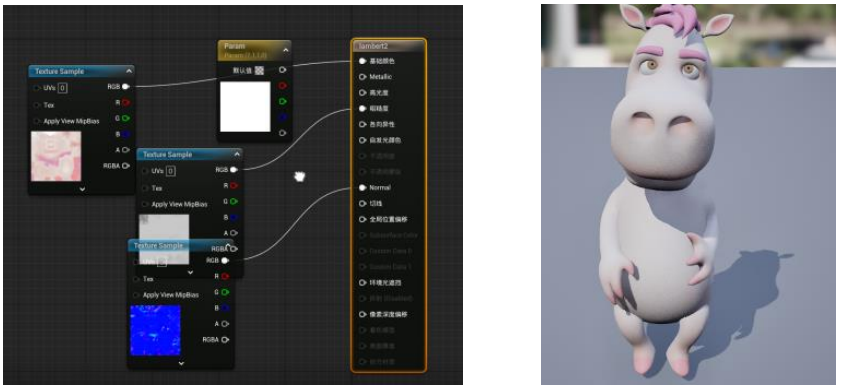

Figure 21. Connect the textures and adjust the UV value to get the result

3.4. Motion capture data processing

3.4.1 ik redirect (blender)

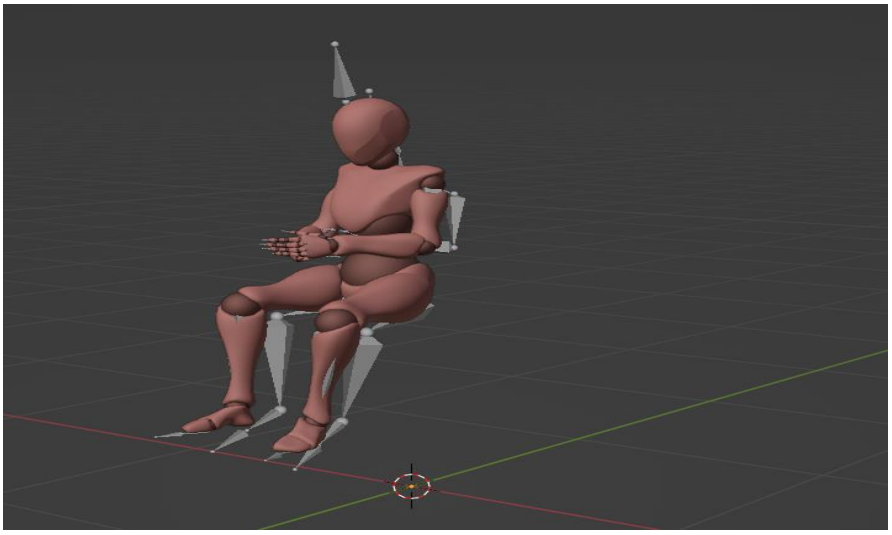

Figure 22. redirect the exported fbx file in ue to get the final result

3.5. Virtual character driver

3.5.1. Retargeting personas

Connect the obtained bvh motion capture file with the corresponding animal character file to create ik control. Since the animal character only has three fingers, we bind the thumb, index finger and ring finger of the character motion compensation data to each other, because in all In hand movements, the movement trends of these three fingers are more consistent with the movement trends of the five fingers.

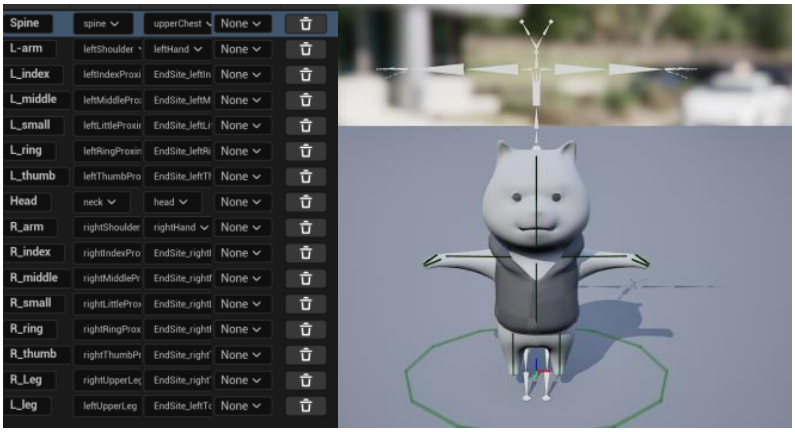

Figure 23. Retargeting personas

3.5.2 Modify the animation sequence

Since the animal characters are relatively obese, putting the character's motion capture data directly into the model will cause problems. After exporting the animation sequence, we rotated the character's hand bones forward and set key frames, and applied compressed animation.

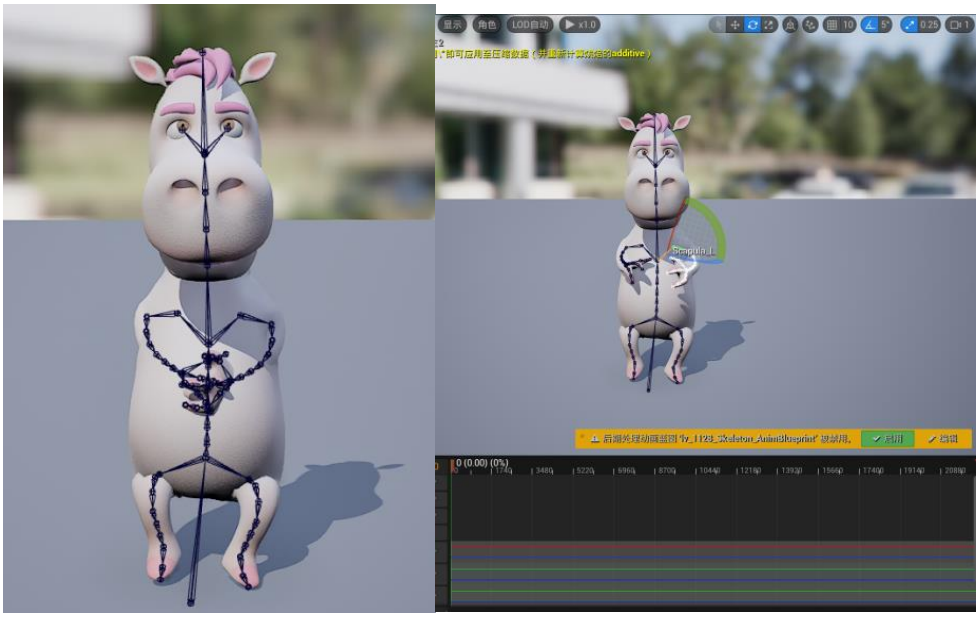

Figure 24. Modify the animation sequence

3.5.3 Configure facial data

Facial data uses livelink Face for face capture. After enabling the livelink Face plug-in, use the mobile phone to connect. After setting the same IP address, the connection is successful.

Figure 25. Configure facial data

Set up the facial data capture blueprint and add the blendshape of the livelinkface node-driven character to the blueprint.

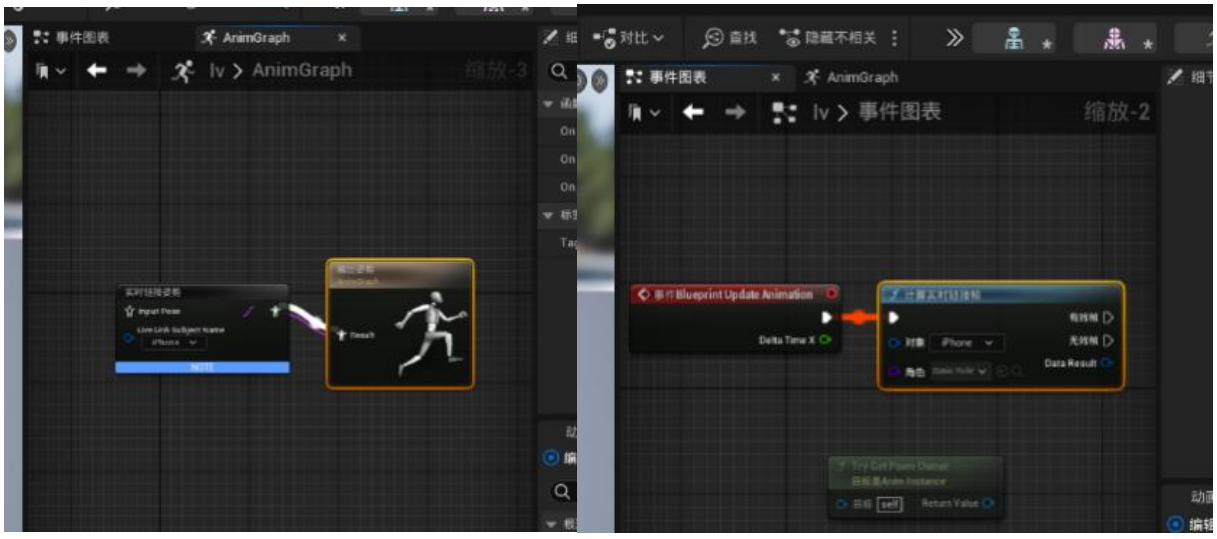

Figure 26. Set up the facial data capture blueprint

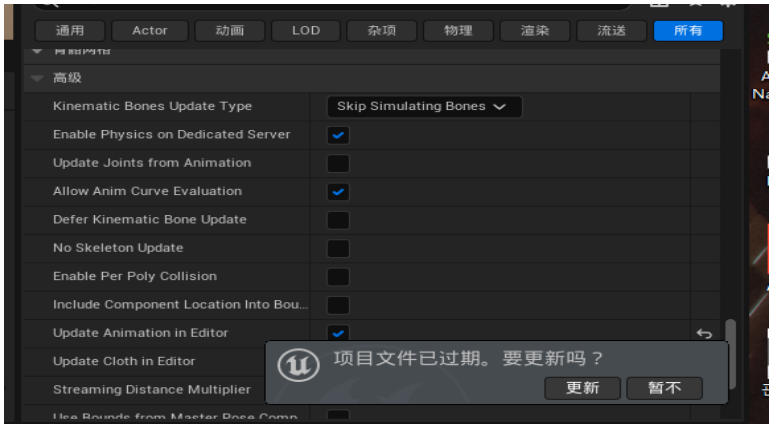

Figure 27. Open Update animation layer in editor in Settings

Mix facial animation and body animation because facial animation and body animation are not an animation sequence. Therefore, using the hybrid animation method, I tried to use the blend multi node for mixing at the beginning, but the resulting animation amplitude would be reduced. After trying several nodes, I used the apply mesh space additive node to perfectly solve the problem of loss of motion amplitude.

Figure 28. mesh space additive node

The overall nodes are as follows.

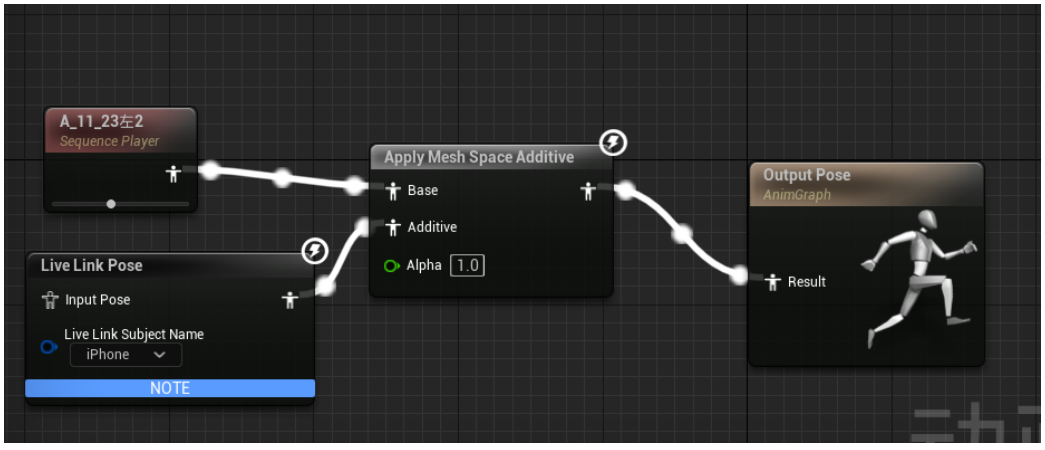

Figure 29. overall nodes

3.6. Post-processing

After adding the background, import the created characters and background into the engine, import the audio files into the final content, and set the parameters of the ambient light and camera to get the picture. After the obtained character is retextured and made into a shaper, the character presented on the screen is imported into the scene and the rendering pipeline is unified. Create mg animation (opening, transition, cutscene, end) to match subtitles.

Figure 30. Post preview effect video screenshot

Conclusion

This article takes the talk show "Beast Town" as an example to introduce the overall technical solution, technical difficulties and countermeasures for the combination of cartoon virtual characters and virtual studio technology, providing reference and experience for the multi-scenario application of digital humans. Compared with the live broadcast that combines reality and reality, we have further upgraded our virtual production technology and digital human-driven technology, adopted industry-leading real-time virtual production technology and monocular camera driving technology, and launched a virtual cartoon character talk show - "Beast Town" to achieve real Perfectly combined with virtuality, it further enhances program immersion and audio-visual experience, and expands infinite boundaries for virtual manufacturing.

Compared with multi-angle visual motion capture software, there is rokoko vision, which uses rokoko. At the beginning, rokoko was used for motion capture. After repeated debugging, the results were quite satisfactory. Thanks to the character model of rokoko studio, the parameters of each joint can be set, so the results were better. Compared with the original video, the movements were more accurate, but rokoko The results obtained after the word recording time exceeds 4 minutes will produce errors, and the Google information will be offset. The reason is that inertial motion capture systems generally use MEMS three-axis gyroscopes, three-axis accelerometers and three-axis magnetometers. An inertial measurement unit (IMU, Inertial Measurement Unit) is used to measure the motion parameters of the sensor. However, the sensor motion parameters measured by the IMU have serious noise interference, and the MEMS device has obvious bias and drift. The inertial motion capture system cannot accurately track the human posture for a long time. Only by solving this problem can the inertial motion capture system fully play its role in the VR industry.

After that, I used EasyMocap to export the motion capture data produced by Esaymocap to json format. What is more complicated is the confusion of IDs when capturing multiple characters at the same time, which

makes some IDs unrecognizable and leads to ID overlap. After the test Select the characters to sit separately and change the id python file in the json at the same time. This solves the id problem of the exported file and the results are ideal. However, due to the need for hand animation, the results obtained after adding hand animation are deviated, resulting in hand animation. The hand animation was disordered, and after adding hand animation, the output time increased three times. We called the kuda core, but it did not start. As a result, the calculation time increased to 3 seconds per stitch. At the same time, in order to ensure quality, we set up 8 When adding multiple platforms for motion capture, it takes up to 9 hours to output a 10-minute motion capture file (cpu accounts for 40%, gpu accounts for 4%). At the same time, when converting json files to bvh files, it is easy to cause confusion. . The software I finally chose to use was xr animator, and got better results. The live link face is used uniformly on the face. After redirection, the motion capture data is assigned to the character to obtain the final virtual character.

References

[1] WEI Y. Deep-learning-based motion capture technology in film and television animation production [J]. Security and communication networks, 2022(11):1-9.B. Sklar, Digital Communications, Prentice Hall, pp. 187, 1998.

[2] WANG YS, LANG X, DU YD, et al. Construction of live animation platform based on motion capture technology[C]// 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). Dalian, China: IEEE, 2021.

[3] Mahmood, Naureen, et al. "AMASS: Archive of motion capture as surface shapes." Proceedings of the IEEE/CVF international conference on computer vision. 2019.

[4] Van der Kruk, Eline, and Marco M. Reijne. "Accuracy of human motion capture systems for sport applications; state-of-the-art review." European journal of sport science 18.6 (2018): 806-819.

[5] Zordan, Victor Brian, et al. "Dynamic response for motion capture animation." ACM Transactions on Graphics (TOG) 24.3 (2005): 697-701.

[6] Moeslund, Thomas B., and Erik Granum. "A survey of computer vision-based human motion capture." Computer vision and image understanding 81.3 (2001): 231-268.

[7] YuanZi Sang, KiHong Kim, Yang Pan, From Technology to Content: Research on the Development of VR Flow Experience, The International Journal of Advanced Smart Convergence, (2003), Vol.11, No.3, pp.93-101. DOI: http://dx.doi.org/10.7236/IJASC.2022.11.3.93