6DoF VR VIDEO CONTENT PRODUCTION, DISTRIBUTION AND MULTI-PLATFORM REAL-TIME INTERACTION

ACADEMY OF BROADCASTING SCIENCE, NRTA

ABSTRACT:
This project provides a full-chain solution for 6DoF VR multi-modal video content production, distribution and experience, which has achieved the effects of significantly reducing production complexity and production time, reducing the amount of distributed data, cross-platform terminal access, real-time interaction in multi-person virtual space, and strong immersion interaction.

KEYWORDS:
6DoF (Degree of freedom), Virtual Reality, Content Production, Multi-platform, Real-time Interaction

1. INTRODUCTION

6DoF VR video content generally has problems such as high production difficulty, low scene richness, large amount of data, immature interactive patten, and limited large-scale application. This project has developed 6DoF VR video content production and distribution technology based on multi-modal data intelligent conversion, immersive interaction technology based on fixation and distance guidance, real-time rendering technology of digital image animation based on reverse dynamics and multi-platform VR video interactive technology. The 6DoF VR multi-modal video content production, distribution and multi-platform real-time interaction system was built to solve the above problems.

On the content production end, in the past, it was necessary to use 3D engines such as Unity for individual scene design and production, which was highly professional and difficult. This system can reduce the dimension and template processing of complex 3D scene content production, reduce the production difficulty and time, and reduce the amount of data required for transmission, also can carry out intelligent conversion of multi-modal material content. It improves the richness of scene content. At the content experience end, through multi-modal content fusion and multi-sensory guided interaction, users' cognition of the virtual environment is enhanced. Local motion capture, global physical inference and other technologies are used to generate and render accurate digital images and actions in real time, combined with multi-user, low-latency interaction technologies enriched the interactive patten and enhance the natural and immersive sense of interaction. The system is compatible with OpenXR standard, suitable for multiple platforms, also can support large-scale dissemination and application of 6DoF VR video content.

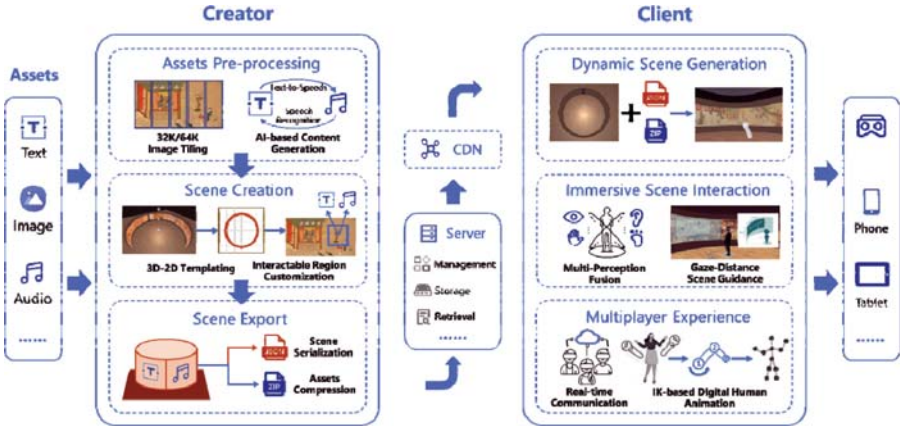

Figure 1 : *End-to-end system solution diagram*

2. ARCHITECTURE SOLUTION AND KEY TECHNOLOGIES

2.1 END-TO-END SOLUTION

This project has developed a 6DoF VR multi-modal video content production, distribution, and multi-platform real-time interaction system, which is divided into three parts: content production, content data distribution and content experience/interaction. The end-to-end system solution is shown in ***Figure 1***.

2.2 CONTENT PRODUCTION END

The content production end is responsible for organising, converting, positioning and editing the imported scene materials, generating 6DoF VR video scene description files, packaging the resources, and uniformly uploading them to the content data distribution end. The contend production end screenshot is shown in ***Figure 2***.

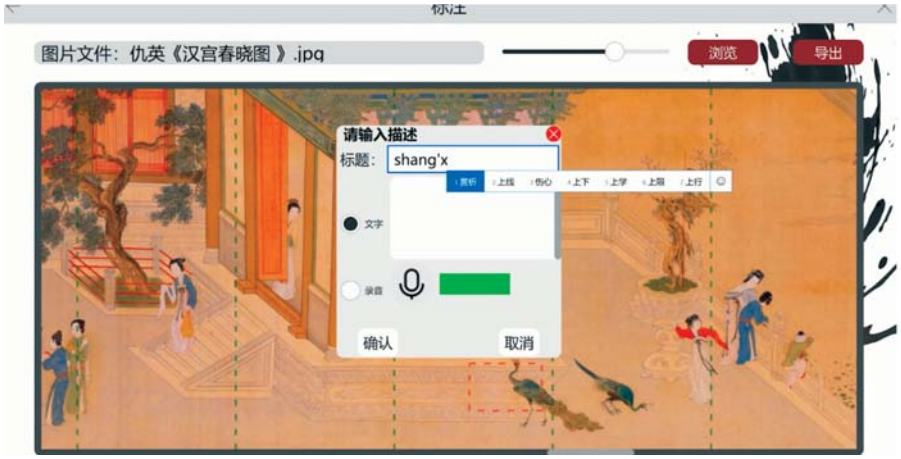

Figure 2 : *Screenshot of the content production end*

The key technologies of the content production end are as follows:

a) The typical three-dimensional scenes in the actual application requirements are templated and presented in two-dimensional form, which reduces the complexity of scene design, understanding and production;
b) Images, texts, audio and other multi-modal materials can import VR image support 64K and above ultra-high resolution. By dividing the original image into several block images with lower resolution, it can avoid decoding and rendering failure, maintain high fidelity and high definition;
c) Can set user interaction points and place material content at different locations in a scene, and the system can automatically generate multi-sensory guided interaction areas in three-dimensional space;
d) Though AI technology, it can convert text to audio, audio to text, to achieve multi-modal content conversion, so that the scene content can be selected by the user in the most appropriate mode of presentation;
e) When the content is completed, the content location and interactive information can be text serialised, the scene description file is generated, and only the original material and part of the generated content are compressed and packaged, avoiding the direct transmission of three-dimensional scene and object data, which greatly reduces the amount of data required for distribution.

2.3 CONTENT DATA DISTRIBUTION END

The content distributor receives the produced scene content file through the network server, and with the optional content distribution network (CDN), provides the access, retrieval, addition, deletion and other capabilities of the scene description file and resource file, to achieve the rapid distribution of the produced content in many places.

2.4 CONTENT EXPERIENCE AND INTERACTION END

After the content experience end obtains the scene description file and resource file from the content distribution end, it generates the 6DoF VR video scene conforming to the description in real time, dynamically and asynchronously. Users can enter the scene through different terminals, experience the scene in six degrees of freedom, and communicate and interact with other users in the same scene in real time with low delay. The screenshot of museum scene shown in ***Figure 3***.

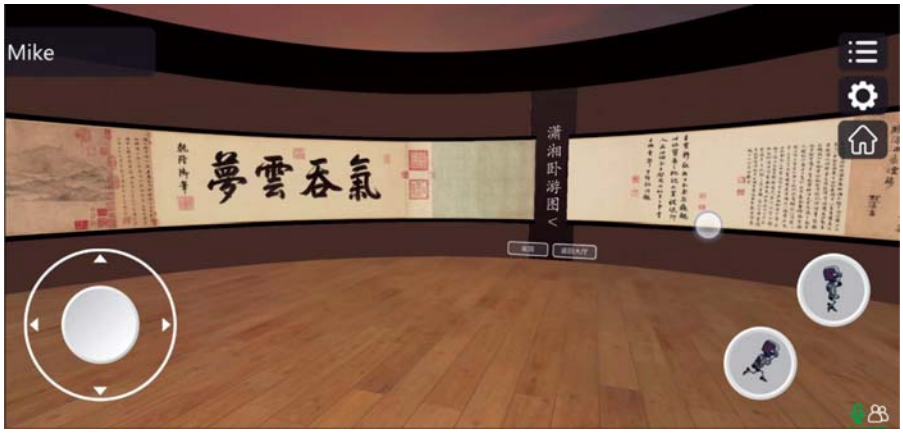

Figure 3 : *Screenshot of the museum scene*

The key technologies of the content experience end are as follows:

a) Users can browse, retrieve, and access all the scenes stored on the server. After selecting a scene, the system automatically downloads and decompresses the corresponding resource file to load the scene template (shown in ***Figure 4-1***);
b) The experiential software can asynchronously and dynamically fill the material, set the location and various properties, add interactive areas and behaviors, and then generate 6DoF VR video scenes in real time (shown in ***Figure 4-2***);
c) The sensor of VR headset display and other devices can capture the user's gaze angle and spatial position information in real time, and then dynamically match and display/hide the interaction area inside/outside the current field of view, guide the user's scene, and improve the concentration of its experience;
d) The experience end supports multi-sensory integration of multi-modal content, and enhances the user's immersion experience and the awareness of scene information through ultra-high-definition image rendering and spatial audio playback;
e) Through real-time network data exchange, support multiple users to experience the same scene at the same time, and through voice and other low-latency interaction;
f) Use the sensor of VR headset display and other devices to obtain the key point pose of the user's part bone, deduce the parent bone pose based on reverse dynamics, and then adjust the whole bone chain data to generate natural digital image action animation (shown in ***Figure 5***);
g) Compatible with the OpenXR standard, generate content experience terminals running on different devices and system platforms, adaptive adjustment of interaction interface, interaction mode, operation parameters, etc., to achieve a multi-user, cross-platform real-time interactive experience of 6DoF VR video content, so that the wide range of 6DoF VR video content promotion and application is no longer limited to VR hardware equipment.

Figure 4-1 : *Users choose the scene*

Figure 4-2 : *Users choose the viewing area*

3. HIGH-LIGHT FEATURES (DATA BASED ON ACTUAL TESTS)

a) The equivalent video resolution is more than 64000×3840, and the spatial audio effect is equivalent to 32 channels;
b) The 6DoF VR video frame rate reaches 120fps, the number of scene vertices exceeds 11000, and the number of triangular pieces exceeds 14500;
c) Compared with the ordinary 6DoF VR system, the user's memory ability, understanding ability and immersion score of the 6DoF scene are improved by 12%, 27% and 13% on average;
d) The server is configured with 16GB memory, which can support 300 users online in the same scene at the same time. The delay of multi-user voice interaction is less than 200ms;
e) The system supports real-time calculation of 144 skeletal pose animations with a frame rate of 75fps;
f) Terminal type supports computer, mobile phone, tablet, VR headset and so on.

4. APPLICATION SCENARIOS

In the 6DoF VR video content production, the complex and time-consuming 3D scene production is converted to 2D graphic design and editing, which greatly reduces the difficulty of 6DoF video content production and saves production time and cost. At the same time, the generated scene content data can be described by text serialisation, avoiding the direct transmission of 3D scene data, and also reduces the amount of data required for distribution.

In terms of 6DoF VR video content experience, it supports multi-modal data multi-sensory experience, multi-user real-time experience, and low-latency interaction, and supports cross-platform experience of VR headsets, mobile phones, tablets, computers, and other devices.

Figure 5 : *Generate natural digital image motion based on reverse dynamics*

Figure 6 : *Application case of Peking University Golden Autumn Education Group*

The system can be applied to a variety of pattens such as VR+ cultural travel, VR+ education, VR+ training, VR+ entertainment and so on.

At present, the system has been applied to some offline teaching points of Peking University Golden Autumn Education Group (shown in ***Figure 6***), Migu Virtual Museum project, etc., and is promoting the application with Henan Museum, Bohua Ultra-HD Innovation Center, Shanghai Jing Qing Hui Vision Technology Co., LTD.

AUTHOR

Wei Na holds a master's degree in electronic and communication engineering, and is engaged in high-tech video, virtual reality, TV production broadcast, and other fields of technology research and application work. As the project leader, she completed several SARFT scientific research projects and took the lead in compiling several broadcast and television industry virtual reality technology standards.