VR Viewing Test of 3D Reconstructed Content Generated by Markerless Motion Capture in Wide Area

Masaaki Matsumura, Kazuki Okami*, Hajime Noto, Hideaki Kimata

NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation
1-1 Hikarinooka Yokosuka-shi, Kanagawa, 239-0847, Japan
Keywords: VR viewing test, 3D reconstruction, human joint estimation, markerless motion capture

ABSTRACT

Recent years, the visualization techniques for wide area with AR and VR have been attracting attention. We propose the method to create a real-scaled VR viewing experience using images of actual handball game. And then, we test the experience can be entertained without feeling of discomfort using user questionnaires.

1 INTRODUCTION

Recent years, the visualization techniques for wide area with Augmented Reality and Virtual Reality (VR), such as highlight scenes in sports, has been attracting attention [1].

Intel® True View system [2] visualizes free-viewpoint videos using colored point cloud generated by triangulation for each pixel of calibrated high-resolution video cameras. However, in the use case of point cloud, when the virtual camera approaches to a player, not only the point resolution degrades, but the colored point reproducibility of a player that is not seen by any video camera may be degrade. Free Viewpoint Video System virtual camera system [3] proposed by Canon© visualizes free-viewpoint videos using polygon meshes generated by visual hull of silhouette derived players from each captured frame by video cameras. However, in the use case of the polygon meshes, when the virtual camera approaches to a player, not only the mesh and texture resolution degrade, but the mesh reproducibility of a player that is not seen by any video camera may be degrade. Therefore, these approaches require huge number of calibrated high-resolution cameras for maintain visual quality.

On the other hands, by estimating only the motion of human joints from the player captured in video and applying the detected motion to the pre-generated 3DCG model for the player, it is possible to maintain the quality of mesh and texture resolution reproducibility even with a few numbers of video cameras. However, since these estimated player's motions might have not enough reproducibility of motions, there remains a question as to whether the video generated by this method can be accepted by the general audience.

* present affiliation is NTT TechnoCross Corporation
Shinagawa Grand Central Tower.17F 2-16-4, Konan, Minato-ku, Tokyo, 108-8202, Japan

When reproducing a video often seen on a television using such a free-viewpoint video generation system, such as a video of a zoomed player or downward view of a playing field, it is expected that the discomfort will increase due to the inferior quality. However, if we provide real-scaled VR viewing experience from a viewpoint that is difficult to record by actual cameras such as the viewpoint on a playing field, we might obtain a good evaluation that exceeds the discomfort.

In this paper, we set our purpose is to generate a 3D reconstructed VR content that can provide a real-scaled VR viewing experience from a viewpoint that is difficult to record by actual video cameras, such as a viewpoint on a playing field, and to test whether the VR viewing experience can be entertained without feeling of discomfort using user questionnaires.

2 3D RECONSTRUCTION

We generate the VR content from real-captured video cameras using 3D reconstruction. Our VR content include players and environmental structures such as arena are generated by 3DCG creators. Our processing flow for human 3D reconstruction consists of three processes of “Player capture”, “Motion capture” and “Motion application into 3DCG” as shown in Fig. 1.

2.1 PLAYER CAPTURE

For generating the player on VR content, we capture the polygon mesh and texture of each player obtained by full-body scanning using Structure Sensor [4]. The skeletons and weight maps for each polygon mesh are set by 3DCG creators.

2.2 MOTION CAPTURE

Since it is difficult for players to demonstrate motion while wearing a motion capture suit, we capture player’s motions using markerless motion capture using wide-angle fixed-point cameras. 3D coordinate of each joint for each frame is derived by steps below:

(1) Extracts the bounding box for each player from time synchronized frame recorded by calibrated high-resolution cameras,

(2) Performs top-down 2D human joints estimation [5] on the player in the bounding box and obtains a heatmap of the joint,

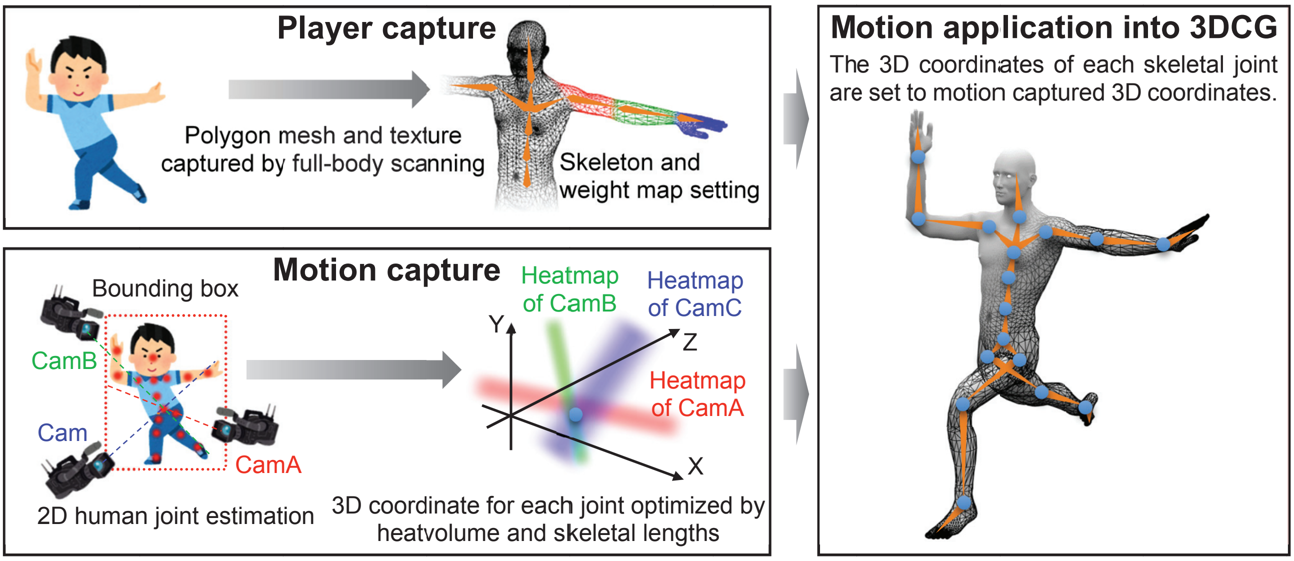

Fig. 1 Processing flow for human 3D reconstruction

The 3DCG character for VR content is generated by captured polygon mesh, texture and motion.

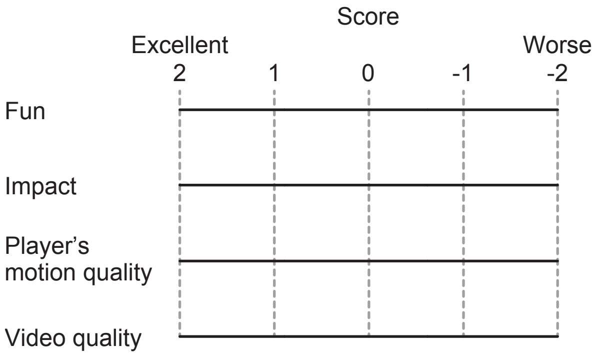

Fig. 2 Questionnaire for VR viewing test

User selects score for each question.

(3) Generates heatvolume by projecting (lens-distortion corrected) the heatmap from all cameras, optimizes 3D coordinates performed with the skeletal lengths of the player as a constraint,

(4) Denoises 3D coordinate by time-series filter.

2.3 MOTION APPLICATION INTO 3DCG

The 3D coordinates of each skeletal joint for the 3DCG model generated in "**2.1 PLAYER CAPTURE**" are set to denoised 3D coordinate.

3 VR VIEWING TEST

We captured all players with up to 14 cameras (4K 60 fps) on actual game of handball, reconstructed each player, court and ball into 3DCG, and developed a handball game viewing Application Software for HMDs that can switch the preset viewpoints in the handball court.

We generated each player on VR content using about 8,400 polygons, set the texture for each player assigning a mipmap up to 4096x4096 resolutions. Here, all the texture for each player is baked in single texture. Since the drawing performance of the HMD device was insufficient, the lighting effects for each player are baked into the texture in advance. Therefore, even if the player moves, the shade is drawn while being fixed.

The VR content consists of scenes that are shot three times within about 45 seconds, and virtual cameras were set in a position where the player's motions such as ball passes, runs, shots and keeper blocks are easy to see. Especially, in order to be able to feel visual impact, virtual cameras were also set at the viewpoint of the keeper who blocks shots.

We conducted VR viewing test at a shopping mall, and a simple questionnaire with 5 digitized ratings (excellent to worse) such as Fig. 2 was obtained for many and unspecified users who visited our event hall.

4 RESULTS

Fig. 3 shows aggregate result of the questionnaires. For all questions, we were able to get mean score of more than 1.0 points. Because the content we generated can be viewed from the front of the handball player's shot, so we obtained an answer that 90% of users felt impact. Although it was not confirmed whether the real-scaled impactive VR viewing experience is a main reason, nearly 97% of user was enjoyed our VR content. About the discomfort by estimated player's motion quality, nearly 80% of users obtained positive responses. The score of "Player's motion quality" suggests a possibility that some estimation errors might not give a discomfort in motions such as sports athletes who can move quickly. Since our VR content was generated for a demonstration and proof of concept, we could not provide enough quality for "Video quality". However the score of "Video quality" has high correlation with the rendering performance of the HMD and the time spent on 3DCG production. The progress for rendering techniques such as video games might improve this score of "Video quality".

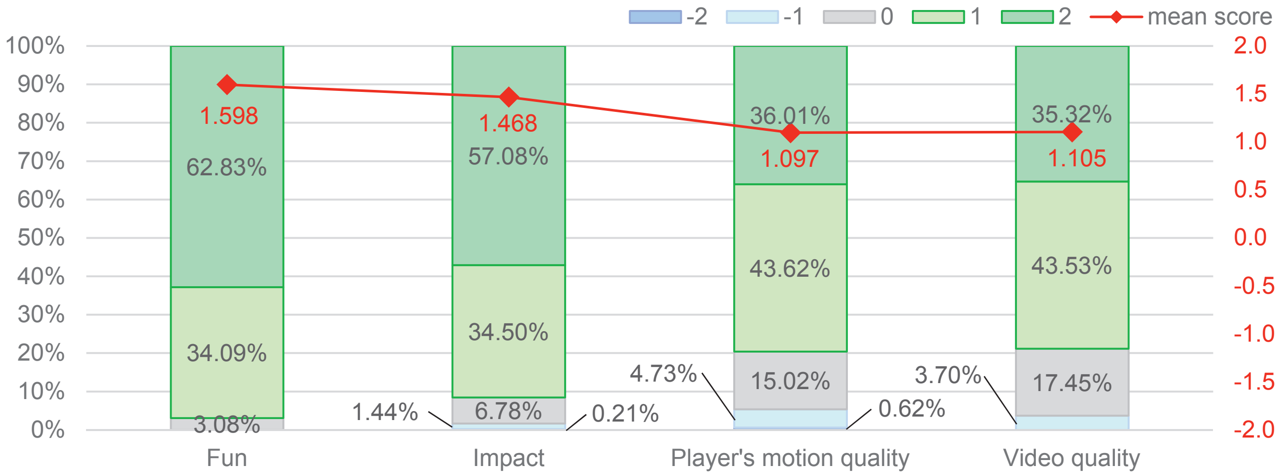

Fig. 3 Result for VR Viewing Test

Each question is corresponding to the question shown in Fig. 2.

Score 2 (green) and -2 (blue) correspond to "Excellent" and "worse" respectively.

5 CONCLUSION

We generated a VR content that can provide a real-scaled VR viewing experience from a viewpoint that is difficult to shoot by actual video cameras, such as a viewpoint on a playing field, and tested whether the VR viewing experience can be entertained without feeling of discomfort using user questionnaires.

As our future work, cross tabulation utilized age group or gender may derive high correlations between each question and them, and improve the viewing experience by 3D reconstruction technology that considered analysis results.

REFERENCES

[1] Konstantinos Rematas, "Watching Sports in Augmented Reality." in IEEE Potentials Volume: 38, Issue: 3, May-June 2019, pp-20-23 (2019).

[2] Intel®, "True View." https://www.intel.com/content/www/us/en/sports/technology/true-view.html.

[3] Canon©, "Free Viewpoint Video System virtual camera system." https://global.canon/en/news/2017/20170921.html.

[4] Occipital, "Structure Sensor." https://structure.io/structure-sensor.

[5] Kazuki Okami, Masaaki Matsumura, Hajime Noto, Hideaki Kimata, "Investigation of Robust Pose Estimation Technique with Occulusion for Space Reconstruction" CSVC2018-20, MR2018-9, (2018).