3D FREE-VIEWPOINT VIDEO CAPTURING INTERFACE BY USING BIMANUAL OPERATION

Tetsuya Watanabe, Itaru Kitahara, Yoshinari Kameda, Yuichi Ohta

Graduate School of Systems and Information Engineering, University of Tsukuba

ABSTRACT

This paper proposes an interface for capturing the 3D free-viewpoint video of a soccer game by using bimanual operation. Users virtually capture 3D free-viewpoint video by observing a virtualized miniature soccer stadium on a tabletop. To exploit the human hand's dexterous manipulation ability, we use a 3D tracker to input the 3D position and the orientation of a virtual camera that captures the 3D free-viewpoint video. The proposed system displays the overhead-view on a tabletop where a user manipulates the virtual camera. By browsing this overhead-view, the context can be easily understood (e.g., positions of players and the ball) in a virtualized soccer scene. We develop a proposed system and conduct on evaluations to confirm the effectiveness of our proposed method.

Index Terms — 3D free-viewpoint video, capturing interface, 3D position sensor, overhead-view, usability test

1. INTRODUCTION

As computer vision and image media technologies continue to develop, 3D free-viewpoint video generation is garnering much research attention [1]-[3]. 3D free-viewpoint video techniques can realize a novel way to enjoy such video media as television by providing audiences the opportunity to control the viewpoint. Moreover, in a large-scale space such as a soccer stadium, 3D free-viewpoint video has an attractive feature that can set the viewpoint anywhere, even where a real camera is rarely placed, e.g., a player's-eye view stepping onto the field or a bird's-eye view looking from the sky. Koyama developed a 3D free-viewpoint video broadcasting system in a soccer stadium using an effective 3D modeling technique called a player billboard that describes a soccer player with a single polygon [4].

The progress of 3D free-viewpoint video generation methods allows us to watch the virtualized world from an arbitrary viewpoint. However enjoying the 3D world remains difficulty since few people has much experience observing a video sequence and controlling the viewpoint. Developing a user friendly interface to capture 3D free-viewpoint videos can solve this problem.

In ordinary 2D video media, since the capturing person (camera person or film director) and the observer (user) are different, the user tend to accept the provided video, even if he or she is not satisfied with it (e.g., the field of view is too narrow to grasp the game context/strategy or a favorite player/team is not well captured). On the other hand, in 3D free-viewpoint video media, a user assumes both capturing and observing roles. Thus, the user can be satisfied with the provided video, when it is possible to control the capturing camera as he/she like. However, we have to take care about the difficulty to control a capturing camera for users. If the negative feeling of controlling the camera outweighs that given by watching boring videos that reflect the tastes of someone else, users may stop controlling the camera and abandon such 3D free-viewpoint video features.

With considering the problem described above, we introduce a user-friendly 3D user interface for capturing 3D free-viewpoint video as shown in Fig.1. One of the most important functions of the interface is that it allows users to easily capture a desired video sequence with intuitive interactions. Capturing 3D free-viewpoint video is realized by three sub-functions: "understanding scene context", "placing a capturing camera", and "setting a focusing point". The proposed interface can intuitively execute all three sub-functions.

Understanding scene context is useful to set a capturing camera at better viewing positions by estimating the object motion and the occurrence of interesting events. So that users can intuitively understand the capturing scene context, we installed an overhead-view monitor that shows the whole target space at a glance. If the 3D position of the capturing camera can be input using the body motion of users, it may be more intuitive than using an ordinary interface with a PC-I/O device such as a mouse. A 3D position sensor offers direct input. Since some focusing objects usually exist in a watching video, the focusing position is useful for setting the orientation of a capturing camera. A method to intuitively set a focusing point is developed by combining the overhead-view monitor and the 3D position sensor. Finally, the pose of the capturing camera is defined by setting the 3D points both of the camera and the focusing point.

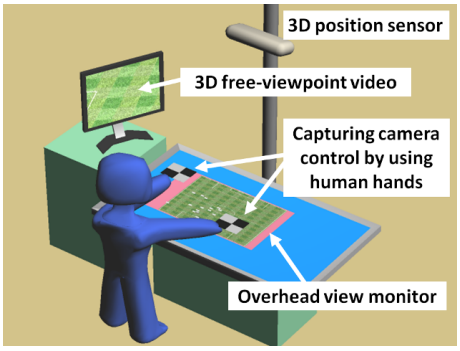

Fig. 1 3D free-viewpoint video browsing interface.

2. RELATED WORK

Inamoto proposed a 3D free-viewpoint video observing system using a Mixed-Reality (MR) technique [5] in which users observe a virtualized soccer scene overlaid on a miniature soccer stadium in the real world using a Head-Mounted Display (HMD). Since MR techniques control the capturing camera by watching the target scene, good viewing positions can be easily found by understanding the scene context. Intuitively controlling the

978-1-4244-6378-7/10/$26.00 ©2010 IEEE

capturing camera is possible by merging a real action (observed by the naked eye) and a virtual action (captured by a virtual camera) in MR space. However, the freedom for setting a capturing cam-era is limited, because its moveable area is constrained by the range of the human's head's movement. Since our hands have a much wider moveable area than the head, we realize a 3D free-viewpoint video browsing system that can observe the target space from anywhere by exploiting this wide area.

3. 3D FREE-VIEWPOINT VIDEO CAPTURING INTERFACE

To realize the sub-functions to place a capturing camera and a focusing point, our browsing interface employs an MR technique with which users can intuitively understand the spatial relation-ship between the real and virtualized worlds. In MR space, a user controls a capturing camera and a focusing point by using bimanual operation to achieve the embodiment effect. A 3D position sensor acquires accurate 3D positions in real time, and then the capturing camera's extrinsic parameters, which is used for the 3D free-viewpoint video rendering, is calculated.

Since we have much experience with 3D relationships, we can intuitively use a 3D position sensor to help us indicate points in 3D space. However, without understanding the cap-tured scene's context, finding a favorite viewpoint and a focus-ing point is not easy. Out browsing interface installs an over-head-view monitor that displays a bird's-eye view so that users can see the entire target space at a glance.

4. 3D FREE-VIEWPOINT VIDEO GENERATION

Our browsing system is based on the 3D free-viewpoint soccer video system proposed by Koyama [4] illustrated in Fig.2. That system executes all processes by capturing multiple images to render 3D free-viewpoint video in real time with an effective 3D modeling technique called a player billboard that describes a soccer player with a single polygon.

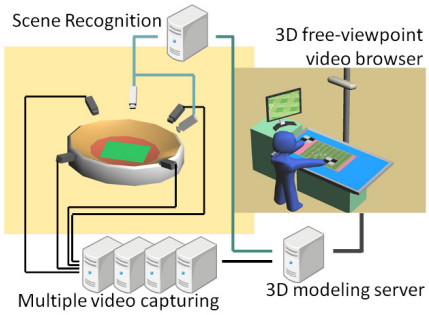

Fig. 2 3D free-viewpoint soccer video system [4].

In the scene recognition block, the system captures a target scene with two cameras set on such high places as rooftops and calculates the 3D positions of a soccer ball [6] and players [7]. In the multiple video capturing block, the system simultaneously captures multiple videos to extract the texture information of the players. To completely obtain all the texture information without a large appearance gap between the multiple cameras, the system should have layout more than eight cameras. The 3D modeling server generates a 3D model of the target objects using the estimated position and the texture information and transmits the 3D model, which is necessary for rendering an observer's view, to the 3D free-viewpoint browser. In the 3D free-viewpoint browser block, users input a desired viewpoint to observe the soccer action. Then the system calculates the camera parameters of the virtual camera and sends them to the 3D modeling server. When the block receives a 3D model responding to the request, a 3D free-viewpoint image is rendered by using the model.

5. CAPTURING CAMERA CONTROL

5.1 Position Input with a 3D Position Sensor

To realize an interface that can utilize the embodiment effect, our interface employs a 3D position sensor to input the capturing camera's viewpoint. As illustrated in Fig. 3, we define world coordinate system ***Cf*** as the basis for merging the real and virtual worlds. The origin is set at the right-bottom corner; the ***X***-axis is set along the sideline, and the ***Y***-axis is set along the goal line of the soccer field. A 3D position sensor has its own 3D coordinate system ***Cc***. To calculate the 3D position in the world coordinate system, ***Cf***, the projection matrix must be calibrated from ***Cc*** to ***Cf*** in advance.

We display an overhead-view and set three landmark points (A, B, and C) on the overhead-view monitor (Fig. 3). Point C is set on the origin, point A is set on the X-axis, and point B is set on the Y-axis of the world coordinates. The 3D positions of the three landmark points are measured using the 3D position sensor, where all of the measured 3D information is described in the sensor's 3D coordinate system, ***Cc***. The 3D translation parameter from ***Cc*** to ***Cf*** is calculated by comparing the 3D positional information of point C in both coordinate systems. The 3D rotation parameters from ***Cc*** to ***Cf*** are calculated by comparing the three basis vectors (X, Y, and Z axes) in both coordinate systems. As the result, a matrix D that translates a 3D coordinate in *Cc* to one in *Cf* is estimated.

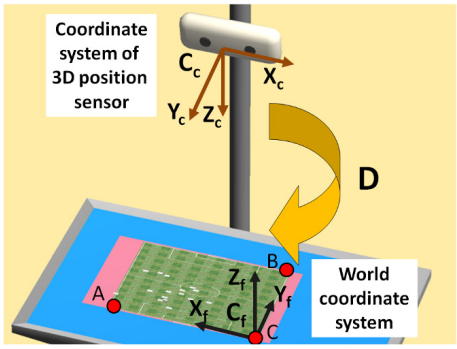

Fig. 3 Geometrical relationship between two 3D coordinate systems (Cf and Cc)

5.2 Focusing Point Input with Overhead-view Display

Our browsing interface displays an overhead-view so that users can understand the scene context (Fig. 4). The positions of a ball and soccer players are displayed by circles and squares and are colored differently for each team. A user observes with two 3D markers on both hands (e.g., on the right hand for a marker of a focusing point, and on the left hand for a capturing viewpoint). Then the 3D coordinates of a focusing point and a capturing viewpoint are defined as ***Pf*** and ***Pv***, respectively. Exciting soccer plays usually occur on the ground, so one of the easiest ways to set the focusing point is by putting a marker on the target icon. When a target object (e.g., ball or player) is moving on the soccer field, it is possible to continue watching it by simply sliding the focusing point marker on the overhead-view monitor. Finally,

the capturing camera's orientation is estimated by calculating a 3D vector from the camera position ***Pv*** to the focusing point ***Pf***.

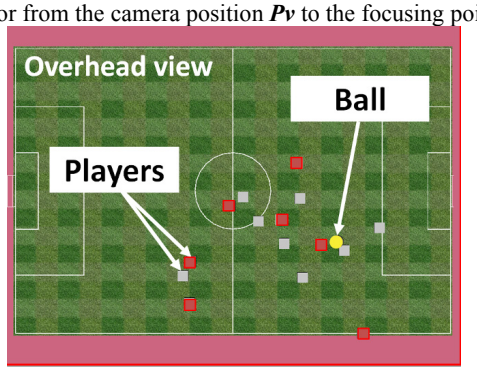

Fig. 4 Overhead-view: positions of ball and players displayed with circles and squares colored by team.

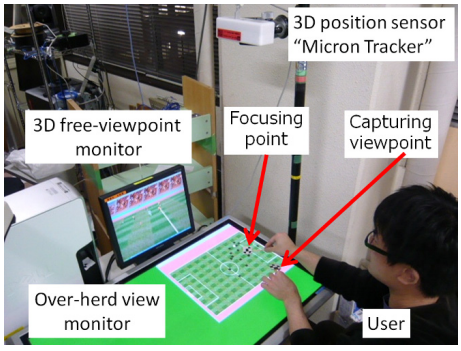

Fig. 5 A pilot system of our proposed method. Almost all users enjoyed 3D free-viewpoint video by utilizing features of the capturing interface.

Fig. 6 Examples of generated 3D videos

6. EXPERIMENTS

As shown in Fig.5, we developed a pilot system to confirm the effectiveness of our proposed method. We installed a 3D position sensor called Micron Tracker Sx60 (Claron Technology Corp.) that acquires accurate (less or equal 0.25 mm) 3D points with 30 frames per second. The coverage is 115 cm wide, 70 cm high, and 55 cm deep [8]. The overhead-view is displayed on a large screen monitor setting sideways on a desk. We used a 32-inch Liquid Crystal Display (LCD) monitor to display the 29.7 by 42.0 cm squared virtual soccer field in the overhead-view. In addition, existing system with a PC-I/O device such as mouse is remained on pilot system for evaluation experiments.

Users watch the generated 3D free-viewpoint videos displayed on a 19-inch LCD monitor set in front of them. Fig. 6 shows examples of videos generated by this system's user. Users observed the 3D space from various viewpoints, confirming that they could discover their favorite viewpoints.

In order to compare the effectiveness of our proposed capturing method with the ordinal one which controls the virtual viewpoint by using a mouse device, we conducted on two evaluation experiments with focusing on "retrieval capability" and "tracking capability" to capture 3D free-viewpoint video. The eleven subjects who familiar with using computers in the daily life. We give them 5-minutes advance practice to understand the capturing systems and the contents of the evaluation scene.

6.1 Evaluation for retrieval capability

For confirming the advantage on the retrieval capability of our method, we conduct on experiments to time a capturing procedure to find out the ball aligned in the virtual 3D space and frame it by using a virtual camera. The ball is aligned at the following three positions P1, P3 and P3.

- P1: Far from the virtual camera, it is necessary to move position of the camera to find out the ball.
- P2: Far from the virtual camera and it is necessary to move position and orientation of the camera to find out the ball.
- P3: Near from the virtual camera, it is necessary to much larger move position and orientation of the camera to find out the ball.

We carried out t-test and the results are shown Fig. 7. All results have significant difference (0.01). In all alignments P1-3, the consuming time for proposed method tend to shorten in comparison with the ordinary one. We consider that there are two reasons for the significant difference. One is that the overhead-view display makes the user to easily understand the ball position even while it does not appear in the 3D free-viewpoint video. The other is that it is possible to easily track the target with sliding the focusing marker on the target icon in proposed system.

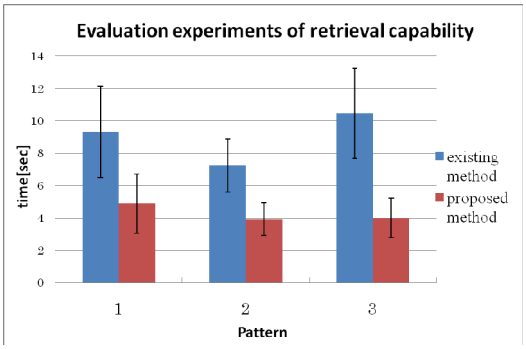

Fig. 7 Results of evaluation of retrieval capability. In all alignments P1-3, the consuming time for proposed method tend to shorten in comparison with the ordinary one.

6.2 Evaluation for tracking capability

For confirming the advantage on the tracking capability of our method, we conduct on experiments to track and capture a ball which is moving around the virtual 3D space. We prepare the following four types of ball trajectories as shown in Fig. 8. The all ball trajectory are displayed to the subjects before the evaluations.

- T1: Uniform linear motion. The ball goes away from the initial virtual camera position.
- T2: Uniform linear motion. The ball cuts across in front of the initial virtual camera position.
- T3: Uniform motion with following a sin-curve.
- T4: Non-uniform motion with following a sin-curve.

The evaluation score is given by the equation (1). Here, the P_{ball} is the position of a target ball; $P_{viewpoint}$ is the position of a capturing camera, F_{total} is the total frames of scene for the evaluation. The distance between P_{ball} and $P_{viewpoint}$ becomes larger, the score becomes better (higher). We inform the score calculation rule to subjects before the evaluations.

$$score = \frac{\left(\sum^{F_{total}} \frac{1}{\left| P_{ball} - P_{viewpoint} \right|} \right)}{F_{total}} * 100 \qquad (1)$$

We carried out t-test , the result of that are shown Fig. 9. The results of T2, T4 have significant difference (0.01) and also T3 has the difference (0.05). However, the result of T1 does not have the difference. About the T1, it is easier to track the ball than others because the ball motion is simple and the moving direction is same as the camera motion. Therefore it is possible to well track the ball by using simplified device such as a mouse.

In case of T2, T3 and T4, that the direction of the ball does not correspond to the direction of the camera. The score of our proposed method is higher than the ordinary method. Two features seem to contribute to this result. One is that the users can operate the several parameters such as the position and orientation of the virtual camera, and the other is that they can perform the intuitive operation with using the 3D marker. It indicates that our proposed method could flexibly apply to the change in velocity of the ball by using the two features.

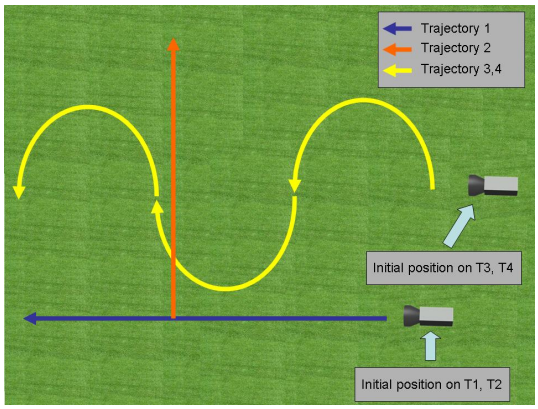

Fig. 8 An example of the ball trajectory.

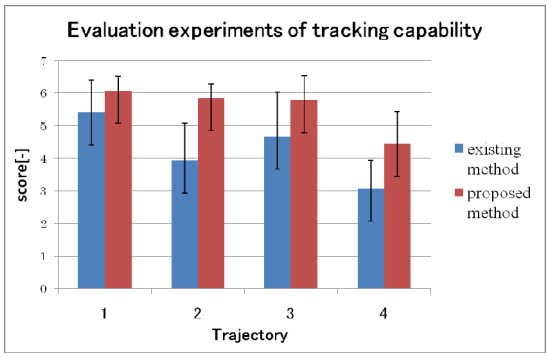

Fig. 9 Results of evaluation of tracking capability

7. CONCLUSION

We proposed an interface for browsing 3D free-viewpoint video of soccer action. To exploit the human hand's dexterous manipulation, we used a 3D position sensor to input the 3D position and orientation of a camera that captures 3D free-viewpoint video. The proposed system displayed an overhead-view on a tabletop where a user manipulates the capturing camera. By browsing this overhead-view, users easily understood the context (positions of players and a ball) in a virtualized soccer scene. We developed a pilot system and confirmed its effectiveness by usability evalua-tion.

8. REFERENCES

[1] T. Kanade, P. Rander, and P. J. Narayanan, Virtualized Reality: Constructing Virtual Worlds from Real Scenes, *IEEE Multimedia,* Vol. 4, No 1, pp. 34-47, 1997

[2] W. Matusik, C. Buehler, R. Rasker, S. J. Gortler, and L. McMillan, Image-Based Visual Hulls, *ACM SIGGRAPH 2000*, pp.369-374, 2000

[3] J. Carranza, C. Theobalt, M. A. Magnor, and H. P. Seidel, Free-Viewpoint Video of Human Actors, *ACM Transaction on Graphics*, Vol.22, No. 3, pp. 569-577, 2003

[4] T. Koyama, I. Kitahara, and Y. Ohta, Live Mixed-Reality 3D Video in Soccer Stadium, *Proc. of International Symposium on Mixed and Augmented Reality (ISMAR2003)*, pp. 178-187, 2003

[5] N. Inamoto and H. Saito, Immersive Observation of Virtualized Soccer Match at Real Stadium Model, *Proc. of International Symposium on Mixed and Augmented Reality (ISMAR2003)*, pp. 188-197, 2003

[6] N. Ishii, I. Kitahara, Y. Kameda, Y. Ohta, 3D Tracking of a Soccer Ball Using Two Synchronized Cameras, *Pacific-Rim Conference on Multimedia (PCM2007)*, pp. 196-205, 2007

[7] N. Kasuya, I. Kitahara, Y. Kameda, and Y. Ohta, Robust Trajectory Estimation of Soccer Players by Using Two Cameras, *The 19th International Conference on Pattern Recognition (ICPR2008)*, CDROM Proceedings, 4 pages, (2008)

[8] http://www.clarontech.com/measurement.php, Claron Technology Micron Tracker