Video Synthesis at Tennis Player Viewpoint from Multiple View Videos

Kenji Kimura*
Department of Information and Computer Science
Keio University, Yokohama, Japan

Hideo Saito†
Department of Information and Computer Science
Keio University, Yokohama, Japan

ABSTRACT

In this paper, we propose a new method for synthesizing player-view from multiple view videos in tennis. Our method is divided into two key techniques: virtual-view synthesis and player's viewpoint estimation. In the former technique, the existing method "view interpolation", which is a technique of synthesizing images in an intermediate viewpoint from the real images in two viewpoints, restricts a viewpoint position between two viewpoints. To avoid this restriction, we propose a new method of the ability to set up a viewpoint position freely. In the latter, viewpoint position is computed from the center of gravity of a player region in captured videos using epipolar geometry. By applying the computed player's viewpoint to the former method, we can synthesize player-view image. Experimental results demonstrate that the proposed method can successfully provide tennis player-view video in tennis.

CR Categories: I.4.3 [Image Processing and Computer Vision]: Enhancement—Registration; I.3.3 [Computer Graphics]: Picture/Image Generation—Viewing algorithms;

Keywords: player view, virtual view, epipolar geometry, IBR

1 INTRODUCTION

New visual effects have recently been given to videos in broadcasts such as a sport. For example, there is "Eye Vision" system which gives the special image effect by changing continuously 30 or more cameras in American football.

The technique of generating a virtual view from two or more camera images is called IBR. IBR can be categorized into "Model Based Approach" and "Transfer Based Approach". In the former[6], a 3D shape model is reconstructed and a virtual view is generated by texture mapping to its surface. In the latter, a virtual view is synthesized by transferring of correspondences between real views. In this approach, there is a technique which reconstructs a 3D model of an object in order to estimate correspondences between multiple views, with strong calibration[1] or weak calibration[5]. These are not suitable for our purpose, since it is not expectable to reconstruct a model correctly in large-scale space.

Inamoto[2] has succeeded in generating a virtual view in soccer. In this method, correspondences between the reference views are estimated by classifying a soccer scene into sub regions, and virtual view is generated using "View Interpolation"[3], which dosen't require to reconstruct 3D object model. Therefore, this method is applicable to the sport scene in large-scale space.

However, View Interpolation has the drawback that viewpoint position is restricted between two viewpoints. For eliminating such restriction, we proposed the new technique, which is able to give a viewpoint position freely and set up it directly.

*e-mail: k-kimura@ozawa.ics.keio.ac.jp
†e-mail: saito@ozawa.ics.keio.ac.jp

2 PROPOSED METHOD

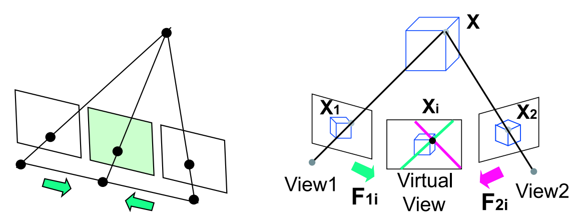
In our method, after dividing a tennis scene into dynamic regions (player/ball) and static regions (court/background), we synthesize a virtual view for every domain. By merging them, we can generate a virtual view of the whole tennis scene. In our synthesis process, two reference views near a virtual viewpoint position are chosen from input views. In addition, we generate player-view video, by given a estimated player's position as input of our virtual view method.

2.1 View Synthesis for Dynamic Region

In dynamic region, we generate a virtual view by the synthesis method of Transfer Based Approach. In "View Interpolation"[3], virtual viewpoint position is given by the relative weights to two reference viewpoints, and the correspondence points between reference views are interpolated by these weighs (Fig.1(a)). These positions are estimated to all over the correspondence pixels between the reference views, and then virtual view is generated by given a color of the reference views to these positions.

View Interpolation restricts a viewpoint between reference views in order to set up it by the relative weights to them. On the other hand, in our method, the position and posture of a virtual view should be able to be set up according to a player position. Therefore we come up with a method for synthesizing a virtual view image by F-Matrix between virtual view and reference view.

As shown in Fig.1(b), giving two F-Matrices between the virtual view and two reference views, two epipolar lines are drawn on the virtual view from a pair of correspondence points on the reference views. Then the two epipolar lines have an interection point, which is also corresponding to the point pair on the reference views. In this way, the two F-Matrices give mapping of corresponding points between two reference views onto the virtual view by giving the color of all correspondence points on the reference views to these intersecting pixels on the virtual view.



(a) View Interpolation (b) Proposed method

Figure 1: Transfer of corresponding points

[F-Matrix between Reference View and Virtual View]

We assume that 3D positions of some points on the standard parts of tennis court are known. Then, the coordinates on a virtual view of these points are estimated using the projection matrix of virtual view set up arbitrarily. Meanwhile, we estimate the correspondences of those points on the standard parts between a virtual view and the reference view, by giving the positions of them on the reference views manually. By these correspondences, the F-Matrix between the virtual view and the reference view is estimated.

IEEE Virtual Reality 2005
March 12-16, Bonn, Germany

2.2 View Synthesis for Static Region

In static region, a virtual view is synthesized using the correspondences of the standard parts of tennis court as **2.1**. Court and net regions are plane, and it can consider that a background region is also a plane by assuming that background exists far away. Therefore, we can synthesize a virtual view by transferring correspondences to virtual view using Homography, which is estimated from the correspondences between the virtual view and the reference view.

2.3 Estimation of Player Viewpoint Position

It is difficult to track the correct direction of the player's gaze, and the 3D position of player's eyes. For estimating player viewpoint position, we assume that a player's eyes are in fixed height to a tennis court surface and always turn player's gaze in the direction of an opponent. Concretely speaking, the player's 2D positions on a ground plane is computed, and the player viewpoint position is estimated by adding these positions and constant value of player's height.

Next, we discuss how to estimate the player's 2D positions on ground plane correctly. There is a method of estimating these positions, by the basis of assumption that player is in contact with the ground. These positions in this method are estimated by projecting the bottom point of a player domain on a ground plane using Homography. However, an error may arise in the reason detectetd side of the leg is randomly switched in frames. Therefore, in our method, the player's 2D positions on the ground plane are estimated from the center of gravity of a player with two player-tracking cameras, which is considered to be the element estimated most stably in a player domain.

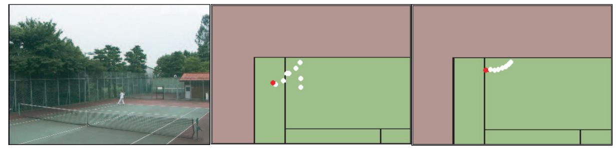
We assume that player-tracking camera capture a player so that the axis of the player is almost parallel to the vertical axis of the image plane. This assumption is equivalent to player's standing vertically to the ground plane and player-tracking camera's being parallel to the ground plane. By this assumption, we can consider that player positions on the ground plane are on the vertical line from the center of gravity of a player, and these positions in two player-tracking cameras are correspondence points between these cameras. Therefore, a vertical line is taken down from the center of gravity on the image, and this line is projected on a ground plane image using Homography between a player-tracking camera and the ground plane image, which is computed from correspondences in the standard parts. The intersection of these line estimated in two player-tracking cameras gives the player's 2D positions.

3 Experimental Results

We performed experiments to generate a player-view from multiple cameras in tennis. Three cameras for virtual view synthesis (Fig.3) and two cameras for player viewpoint estimation were set up.

First, we verify the result of the player viewpoint estimation (Fig.2). In this scene, the player is moving to the diagonal right from the player. However, the player's trajectory estimated by the method using the bottom point of a player domain seems random (Fig.2 (b)). This is because that detectetd side of the leg is randomly switched in frames. On the other hand, in our method, the trajectory according to the player's motion can be obtained (Fig.2 (c)).

Next, we verify the result of the virtual view generated by our method. For comparison, the virtual view generated using View Morphing[4], which is the technique of extending View Interpolation, are shown in Fig.4. View Morphing improves the defect of generating of a geometric distortion in View Interpolation. However, View Morphing is essentially equivalent to View Interpolation, and a virtual viewpoint in this method are set up by the relative weight between reference views. View Morphing is not to suitable for generating the player-view (Fig.4), since a virtual viewpoint can be set up only on the straight line which connected the input viewpoints. On the other hand, our method can generate the player-view on a court as shown in Fig.5.



(a) (b) (c)

Figure 2: Results of player viewpoint estimation [(a) input image, (b) using a bottom point of a player domain, (c) proposed method]

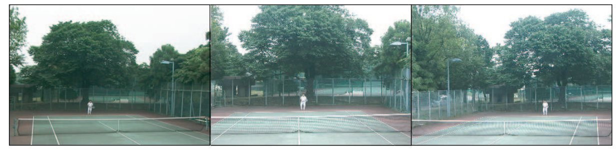


Figure 3: Input images (three cameras for virtual view synthesis)

Figure 4: Virtual view using View Morphing

Figure 5: Player-view using our technique

4 Conclusion

We generate a player-view video from multiple cameras, which capture tennis scene. In this paper, we propose a novel synthesis method, which can set up virtual viewpoint position freely, to avoid restriction of it in View Interpolation. In addition, we also proposed an efficient and robust method of estimating player viewpoint position using projective geometry.

References

[1] S.Baba H.Saito and T.Kanade. Appearance-based virtual view generation from multicamera videos captured in the 3-d room. In *IEEE Trans. on Multimedia*, volume 5, pages 303–316, 2003.

[2] N.Inamoto and H.Saito. Intermediate view generation of soccer scene from multiple videos. In *Proceedings of ICPR '02*, volume 2, pages 713–716, 2002.

[3] S.E.Chen and L.Williams. View interpolation for image synthesis. In *Proceedings of SIGGRAPH '93*, pages 279–288, 1993.

[4] S.M. Seitz and C.R. Dyer. View morphing. In *Proceedings of SIGGRAPH '96*, pages 21–30, 1996.

[5] S.Yaguchi and H.Saito. Arbitrary viewpoint video synthesis from multiple uncalibrated cameras. In *IEEE Trans. on Systems, Man and Cybernetics, PartB*, pages 430–439, February.

[6] P.J.Narayanan T.Kanade and P.W.Rander. Virtualised reality: Concepts and early results. In *IEEE Workshop on Representation od Visual Scenes*, pages 69–76, 1995.