

# AR Display for Observing Sports Events Based on Camera Tracking Using Pattern of Ground

Akihito Enomoto and Hideo Saito

Department of Information and Computer Science, Keio University  
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Yokohama 223-8522 Japan  
{akihito,saito}@hvrl.ics.keio.ac.jp

**Abstract.** We present an AR display system for observing sports events on a desktop stadium based on camera tracking using pattern of the desktop ground by overlaying players in real sports events captured with multiple cameras. In this paper, we take soccer as the sports event. In the proposed system, the pose and the position of an observing camera are estimated in real-time by using a soccer field pattern on the desk top and an AR marker. The soccer field pattern in the desktop stadium on which the object soccer game is observed via AR display is previously registered with the real soccer stadium in which the real soccer game is captured with multiple cameras. In the previous procedure, we also estimate camera parameters (projection matrices) of the multiple cameras capturing the real soccer game using planar structures in the soccer field. Positions of soccer player and ball are also previously estimated based on the camera parameters. In the on-line procedure for AR display, the textures of the players captured in the multiple soccer video are simply overlaid onto AR camera videos with CG models which are generated for giving additional visual information.

**Keywords:** Augmented Reality, Free viewpoint videos, Multiple cameras.

## 1 Introduction

Computer vision and image processing technologies have extensively been applied to provide novel ways of viewing sports videos for various purposes in recent years. Such application can provide an immersive observation and additional information to viewers in broadcasting of sports games. We can see some examples, such as a virtual off-side line superimposing in soccer game broadcasting [1], world record line superimposing in swimming races, etc. Those additional effects definitely give exciting observing experiences that cannot be obtained by watching captured videos without any additional information. The information superimposing in input videos such as off-side line and world record line are considered as Augmented Reality (AR) video synthesis, in which the key technology is the way of estimating the position and pose of cameras capturing the sports events.

Besides, capturing using multiple cameras of object sports events also provides novel experiences of observation. Eye Vision [2] provides a sort of fly through viewing effect by switching cameras that capture the same sports event from multiple viewpoints.

While Eye Vision is simply switching videos captured with multiple cameras, we can even synthesize a novel viewpoint video using videos captured with multiple cameras based on computer vision technology. Such method is called as free viewpoint synthesis, and has extensively been studied in recent 10 years [3,4,5,6]. Inamoto et al. [3] proposed a method for synthesizing intermediate viewpoint videos from videos captured with only weakly calibrated multiple cameras for real soccer match application. Jarusirisawad et al. [6] extended this method for applying to practical broadcasting of sports events.

Combining AR display and free viewpoint synthesis can also provide novel ways in order to provide immersive observation of such sports events. Uematsu et al. [7] has applied wide-area camera tracking method using multiple markers [8] to baseball observation onto a desktop baseball stadium model. Inamoto et al also proposed a system for AR observation of real soccer match onto a desktop soccer field model [9]. In this system, an user can observe objective soccer match on the soccer field model through HMD (Head Mounted Display) from arbitrary viewpoint. In this method, images of real players are registered with the field model image captured with the camera attached with HMD based on only projective geometry between the multiple viewpoint cameras capturing real soccer match for avoiding camera calibration procedure of the cameras. This implies that only image-based rendering can be augmented into the field model image, but virtual objects cannot be rendered in the AR display because of lack of 3D coordinate information. Such virtual object rendering is often useful for rendering objects that cannot be captured in input multiple viewpoint videos, as soccer goal, moving ball, etc.

In this paper, we propose a system for providing AR observation of soccer match onto a desktop field model using a video see through device, in which even CG rendering can also be augmented into the field model image. In this system, we utilize the pattern on the field model for real-time estimation of camera parameters (projection matrix) of the observing camera with the help of an AR toolkit marker. The estimation of projection matrix of the observing camera enables to augment both rendered images by captured videos with real stadium cameras and CG images into the observing image. Additionally, CG generated players can also be rendered into the observing image, which is especially useful when the viewpoint of the observing camera is much different from the viewpoints of the real stadium cameras.

## 2 Proposed System

The data processing in the propose system can be divided into off-line phase and on-line phase. In the off-line phase, an objective soccer match is captured with multiple uncalibrated cameras from multiple view points in the stadium (we call those cameras as stadium cameras in this paper). Next, player region segmentation should be performed. The segmentation might be possible in automatic process by using state-of-art segmentation algorithms, but we performed with semi-automatic process with manual intervention. In addition to this segmentation, we also need to detect features in the videos for calibrating the capturing cameras such as lines of field pattern and the positions of the ball with manual operation. Such manual operation can be accepted because those procedures do not have to be performed in on-line phase.

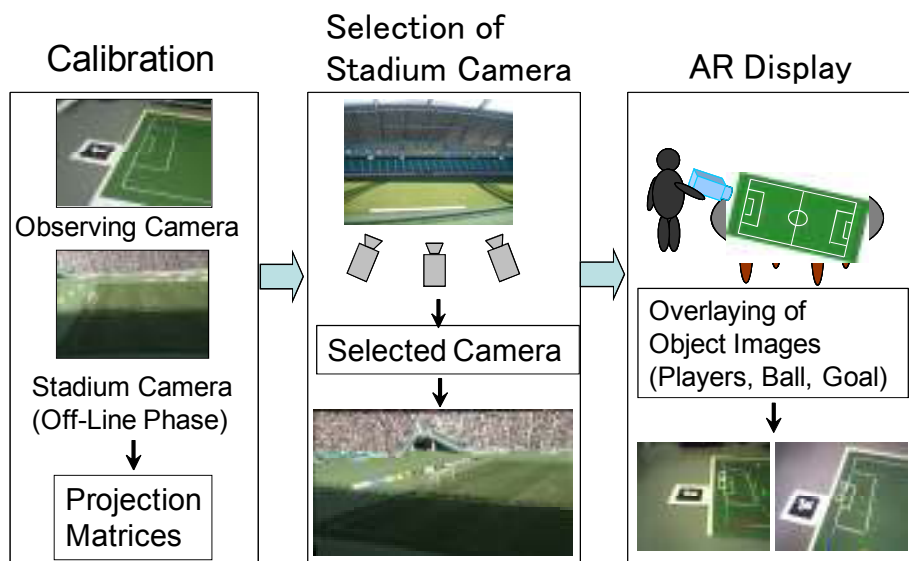


Fig. 1. Overview of the proposed method

In on-line phase, the segmented images of the players captured with the multiple cameras are overlaid onto the image of the field model on the desktop captured with the observing camera. Fig.1 shows the overview of the on-line phase. First, four corner positions of the penalty area on the field pattern are detected in the image of the observing camera. Then using corresponding positions of the detected corners in the real stadium, the projection matrix of the observing camera onto the field of the real stadium is computed.

Next, according to the computed projection matrix of the observing camera, the stadium camera closest to the observing camera is selected. Then, the positions of players captured in the selected stadium camera are computed and transferred into the observing camera coordinates, so that the segmented player image of every player can be overlaid at the correct position into the observing camera.

## 2.1 Player Region Segmentation (Off-Line Phase)

The background subtraction is a generic way for player region segmentation, but it is not suitable especially when the shadow of the player exists. Recently, the graph-cut is often used for achieving robust segmentation [6], but it needs some manual intervention for each frame of input videos. For avoiding such problems, we employ a method for player region segmentation using plane constraint of the soccer field ground. We can assume all players do not lay down on the ground, but stand on the ground plane. This means that only the pixels of the ground can be transferred between different stadium cameras each other using homography matrix for the ground plane. Based on this assumption, we can assume that a pixel color  $P$  in a player region in a camera is not same as pixel colors  $P_i$  at transferred positions in the other cameras using homography matrices of the plane. This implies

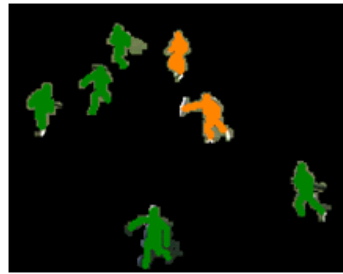
$$P \neq P_1 \approx P_2 \quad (1)$$

Therefore, by transferring a pixel onto the other cameras using homography matrices of the ground and comparing the colors of them, we can decide that the pixel is inside in the player region if the colors are different. If we do not take this method, but just apply the background subtraction method, we will get the segmentation result as shown in Fig.2 (a), where the cast shadow regions are also segmented as the player region. On the other hand, we can segment the player region without shadow region as shown in Fig.2 (b).

The homography matrices between the stadium cameras are computed in the off-line phase by specifying correspondence points on the ground plane between the cameras.



(a) Background Subtraction



(b) Segmentation using the ground plane constraint

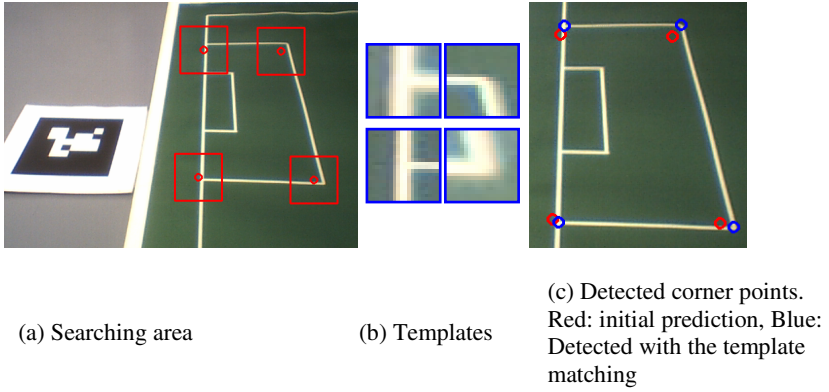
**Fig. 2.** Comparison of segmented region

## 2.2 Detecting Penalty Area in Observing Camera (On-Line Phase)

In this system, we estimate the projection matrix  $P$  via the homography matrix  $H$  of the field plane. For estimating the homography matrix, the four corners of the penalty area of the field pattern are detected as the following procedure.

First, the positions of the corners are initially predicted by using AR marker that can easily be detected so that we can get the initial prediction of corner positions of the penalty area robustly in the on-line real-time processing. We employ ARTag [10] in this paper's experiment. According to the initial prediction, we detect accurate position of the corners using template matching.

In the template matching, we first set the searching area in the neighboring area of the initial predicted four points as shown in Fig.3 (a). We then transfer the captured image to the top-view image using the predicted four points. In the top view image, we take small window images at the initial predicted four points as template images, which are shown in Fig.3 (b). Then, each template image is searched within the searching area, so that we can get accurate corner positions of the penalty area as shown by the blue points in Fig.3 (c).



**Fig. 3.** Detection of four corners of penalty area

If the initial predicted four points cannot be obtained because of miss-detection of AR marker, the detected four points in the previous frame are used as the predicted four points. This enables continuous estimation of the projection matrix of the observing camera even when the AR marker cannot be detected by the camera.

For estimating the projection matrix, we first compute homography  $H$  using the four corner points of the penalty area detected as described above. Then we do not use the positions of the corners of the penalty area in the stadium model, but in the real stadium, so that we can compute the projection matrix of the observing camera onto the real stadium position. This virtually connects the observing camera to the real stadium coordinate.

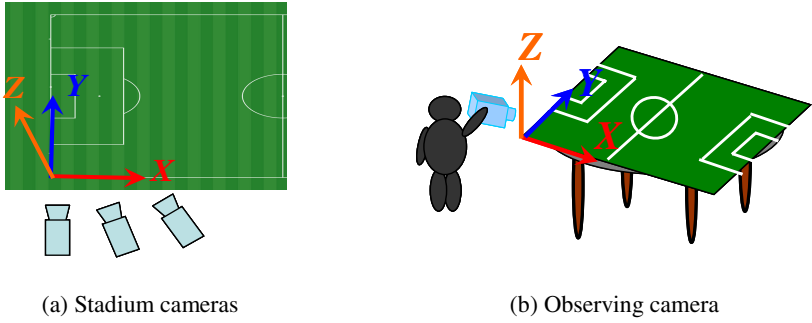
For upgrading homography to projection matrix, we employ the following equation [11][12].

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \sim P \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \quad (2)$$

As shown in Eq. (2), homography is equivalent with projection matrix when  $Z$  component is zero. This implies that we need to compute rotation vector for  $Z$  axis, which can be computed from rotation vectors for  $X$  and  $Y$  axis using the property of rotation matrix. Accordingly, we can get projection matrix of the observing cameras.

### 2.3 Selection of Stadium Camera for Overlay (On-Line Phase)

For each frame, we select the stadium camera closest to the observing camera using the estimated projection matrix as described above. In the experimental setting in this paper, we use the stadium cameras with the coordinate as shown in Fig. 4 (a), in which the goal area should be captured by all stadium cameras. In this case, the vertical axis on the field on the real stadium is set as  $Z$  axis, which implies that the most



**Fig. 4.** Arrangement of cameras

significant direction of the stadium camera is the rotation around Z axis. Therefore, we evaluate only the direction around Z axis for finding the stadium camera closest to the observing camera. The direction can be computed from the projection matrix of every camera.

#### 2.4 Computing Positions of Players and Ball (Off-Line Phase)

For overlaying the players and the ball onto the images captured with the observing camera, we need to compute the 3D positions of them in the coordinate of the field in the real stadium.

For the players' positions, we assume that the players are always standing on the ground by their foot, so that we can say that the position along Z axis for every player is always zero:  $Z=0$ . Then we can compute the positions on XY plane for every player from the foot position  $(x,y)$  of the player in the image captured in the stadium camera selected as described in the previous section by using the projection matrix of the stadium camera with the condition of  $Z=0$ .

For the ball position, we compute the 3D position from the ball positions detected in two images of arbitrarily selected two stadium cameras by using the projection matrices of both cameras. The 3D position can be computed according the following equation

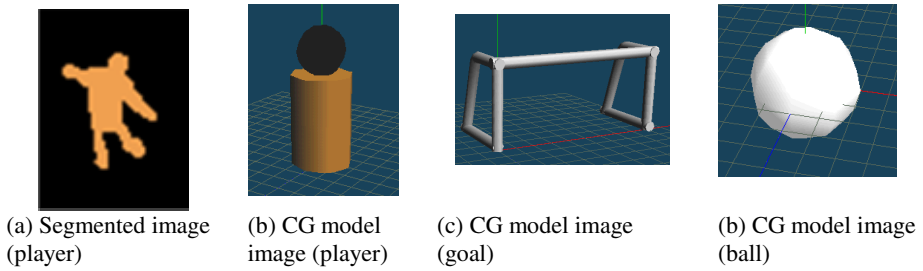
$$\begin{pmatrix} p_{31}x - p_{11} & p_{32}x - p_{12} & p_{33}x - p_{13} \\ p_{31}y - p_{21} & p_{32}y - p_{22} & p_{33}y - p_{23} \\ p'_{31}x' - p'_{11} & p'_{32}x' - p'_{12} & p'_{33}x' - p'_{13} \\ p'_{31}y' - p'_{21} & p'_{32}y' - p'_{22} & p'_{33}y' - p'_{23} \end{pmatrix} \begin{pmatrix} X_B \\ Y_B \\ Z_B \end{pmatrix} = \begin{pmatrix} p_{14} - p_{34}x \\ p_{24} - p_{34}y \\ p'_{14} - p'_{34}x' \\ p'_{24} - p'_{34}y' \end{pmatrix} \quad (3)$$

where the projection matrices of the stadium cameras are represented as  $\mathbf{P}$  and  $\mathbf{P}'$ , the ball positions in both cameras are indicated as  $(x,y)$ ,  $(x',y')$ , and the ball positions is  $(X_B, Y_B, Z_B)$ .

#### 2.5 AR Display (On-Line Phase)

For AR display onto the image captured with the observing camera, the players, the ball and the goal are overlaid at the projected positions from the computed 3D

positions of them using the projection matrix of the observing camera. The position of the goal is previously computed by manual operation. For the images of the ball and the goal, we use simple CG models as shown in Fig.5 (c) and (d) because of CG image are more clear in visual appearance than segmented image from the input image captured with the stadium camera. For the player, we use the segmented image of each player shown in Fig.5 (a) or the CG model image as shown in Fig.5 (b). The CG model image is used particularly when the viewpoint of the observing camera is close to the top-down direction, because there is no image captured with the stadium cameras for such direction of observing camera.



**Fig. 5.** Images of various objects that are overlaid onto observing camera image

The size of the overlaid image is also very important for realistic AR display of those objects. The magnifying size is determined based on the position of the observing camera. The magnifying ratio  $s$  of the object images is decided based on the ratio of the norms of translational vectors  $\mathbf{T}$  and  $\mathbf{T}'$  of the selected stadium camera and the observing camera, respectively, by the following equation.

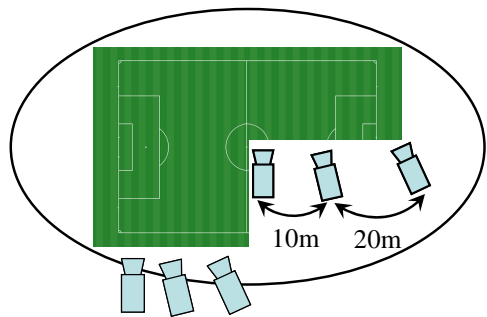
$$s = a \frac{|\mathbf{T}|}{|\mathbf{T}'|} \quad (a \text{ indicates a controlling coefficient}) \quad (4)$$

According to this equation, we can change the size of the object images by the distance between the stadium model and the observing camera.

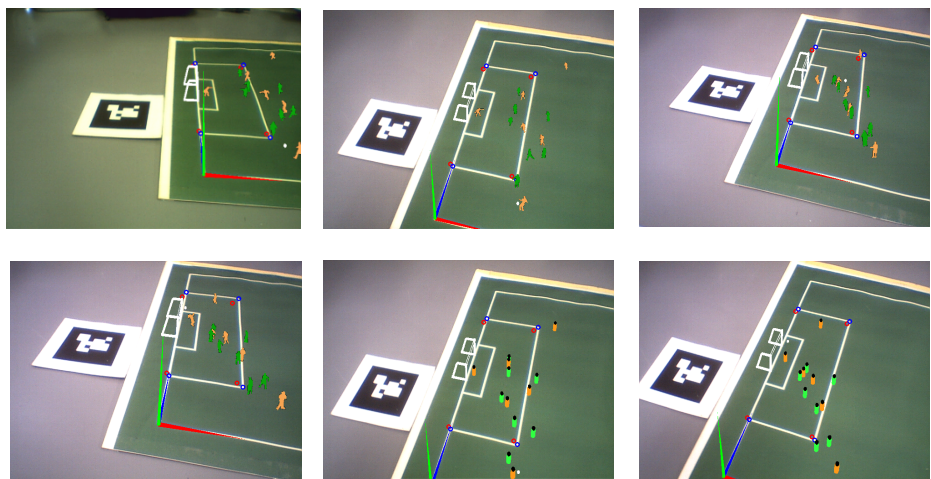
### 3 Experiments

As the videos captured by the stadium cameras, we use the multiple viewpoint videos of the professional soccer match that are captured in Ajinomoto Stadium in Tokyo, Japan. We used three cameras that were located as shown in Fig. 6. The distance between camera 1 and camera 2 was 10m, and the distance between camera 2 and 3 was 20m. We also synchronized those cameras in frame.

As for the observing camera, we used WebCam Live!Pro of Creative. The size of captured images is 640 x 480pixels.



**Fig. 6.** Arrangement of stadium cameras

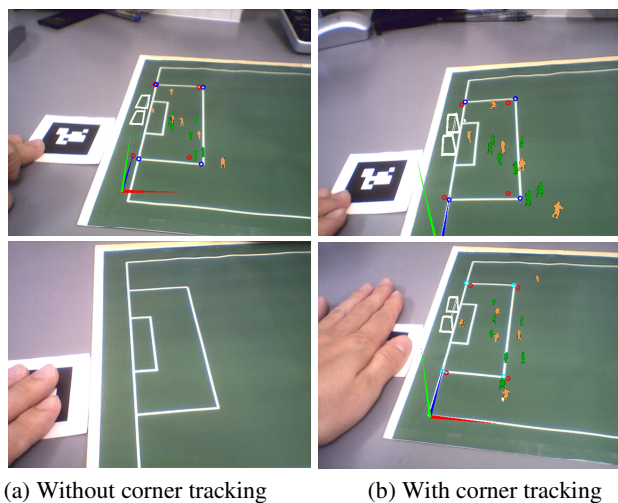


**Fig. 7.** The examples of AR display images

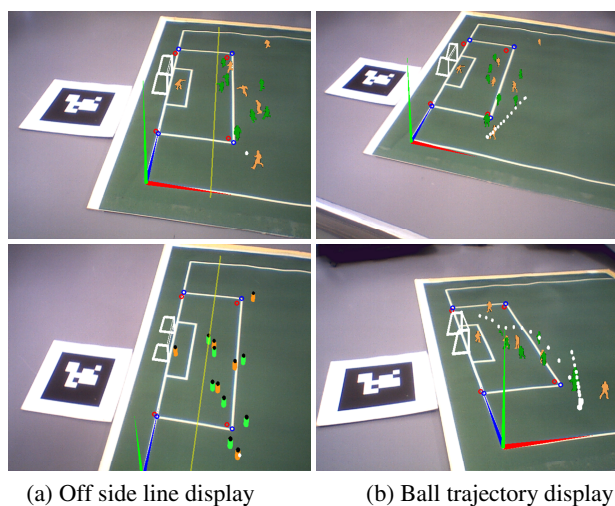
The examples of AR display images are shown in Fig.7. Please note that the player images in this example were intentionally painted by homogenous color by manually. This is because of the limitation of the video copyright of the professional soccer match. The example images demonstrate that we can correctly overlay the object onto the images captured with moving observing camera.

Fig. 8 demonstrates that the propose system can correctly track the observing camera even when the marker is occluded. Fig.8 (a) is the AR display on the observing camera without the template matching of the corners, in which the object images are overlaid at wrong positions in the observing images because of the occlusion of the maker. On the other hand, Fig. 8 (b) demonstrates that the projection matrix of the observing camera can be correctly estimated even when the marker is occluded, because the corners of the penalty area can correctly be detected.

Fig. 9 shows examples of the way to providing additional information of the soccer match, such as the off-side line and the trajectory of the ball.



**Fig. 8.** AR display in case of marker occlusion



**Fig. 9.** Adding information

## 4 Conclusion

In this paper, we present an AR display system for observing soccer match on a desktop stadium based on camera tracking using pattern of the desktop ground by overlaying players captured in real sports events with multiple stadium cameras. In the proposed system, the projection matrix of an observing camera is estimated in real-time by using a soccer field pattern on the desk top and an AR marker.

## References

1. CyberPlay, <http://www.orad.co.il>
2. <http://www.ri.cmu.edu/events/sb35/tksuperbowls.html>
3. Inamoto, N., Saito, H.: Virtual viewpoint replay for a soccer match by view interpolation from multiple cameras. *IEEE Trans. Multimedia* 9(6), 1155–1166 (2007)
4. Kitahara, I., Saito, H., Akimichi, S., Ono, T., Ohta, Y., Kanade, T.: Large-scale virtualized reality. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2001)*, Technical Sketches (2001)
5. Yaguchi, S., Saito, H.: Arbitrary viewpoint video synthesis from multiple uncalibrated cameras. *IEEE Trans. on Systems, Man and Cybernetics, Part B* 34(1), 430–439 (2004)
6. Jarusirisawad, S., Hayashi, K., Saito, H., Inamoto, N., Kawamoto, T., Kubokawa, N., Fujiwara, T.: The intermediate view synthesis system for soccer broadcasts. In: *Proceedings of ASIAGRAPH 2008*, Shanghai (2008)
7. Uematsu, Y., Saito, H.: AR baseball presentation system with integrating multiple planar markers. In: Pan, Z., Cheok, D.A.D., Haller, M., Lau, R., Saito, H., Liang, R. (eds.) *ICAT 2006*. LNCS, vol. 4282, pp. 163–174. Springer, Heidelberg (2006)
8. Uematsu, Y., Saito, H.: Vision-based registration for augmented reality with integration of arbitrary multiple planes. In: Roli, F., Vitulano, S. (eds.) *ICIAP 2005*. LNCS, vol. 3617, pp. 151–162. Springer, Heidelberg (2005)
9. Inamoto, N., Saito, H.: Immersive Observation of Virtualized Soccer Match at Real Stadium Model. In: *The Second International Symposium on Mixed and Augmented Reality (ISMAR 2003)*, pp. 188–197 (2003)
10. Fiala, M.: ARTag, a fiducial marker system using digital techniques. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2005)
11. Simon, G., Fitzgibbon, A.W., Zisserman, A.: Markerless tracking using planar structures in the scene. In: *Proc. of the International Symposium on Augmented Reality*, pp. 120–128 (2000)
12. Simon, G., Berger, M.: Reconstructing while registering: a novel approach for markerless augmented reality. In: *Proc. of the International Symposium Mixed and Augmented Reality*, pp. 285–294 (2002)