A Vision-Based Mobile Augmented Reality System for Baseball Games

Seong-Oh Lee, Sang Chul Ahn, Jae-In Hwang, and Hyoung-Gon Kim

Imaging Media Research Center,
Korea Institute of Science and Technology, Seoul, Korea
{solee,asc,hji,hgk}@imrc.kist.re.kr

Abstract. In this paper we propose a new mobile augmented-reality system that will address the need of users in viewing baseball games with enhanced contents. The overall goal of the system is to augment meaningful information on each player position on a mobile device display. To this end, the system takes two main steps which are homography estimation and automatic player detection. This system is based on still images taken by mobile phone. The system can handle various images that are taken from different angles with a large variation in size and pose of players and the playground, and different lighting conditions. We have implemented the system on a mobile platform. The whole steps are processed within two seconds.

Keywords: Mobile augmented-reality, baseball game, still image, homography, human detection, computer vision.

1 Introduction

A spectator sport is a sport that is characterized by the presence of spectators, or watchers, at its matches. If additional information can be provided, it will be more fun when viewing spectator sports. How about applying a mobile augmented-reality system (MARS) to spectator sports? Augmented Reality is widely used for sports games, like football, soccer, and swimming except baseball. Therefore, we want to focus on baseball games. Hurwitz and co-workers proposed a conceptual MARS that targets baseball games [1]. However, any implementation methods have not been presented.

Previous research literatures include several papers on Augmented Reality (AR) technology for sports entertainment. Demiris et al. used computer vision techniques to create a mixed reality view of the athletes attempt [2]. Inamoto et al. focused on generating virtual scenes using multiple synchronous video sequences of a given sports game [3]. Some researchers tried to synthesize virtual sports scenes from TV broadcasted video [4], [5]. These systems, however, were not designed for real-time broadcasting. Han et al. tried to build a real-time AR system for court-net sports like tennis [6]. Most of the previous works were applied to TV broadcasting of sports. There have been no AR systems for on-site sports entertainment.

In this paper we propose a MARS for baseball games in stadium environments. The overall goal of the system is to augment meaningful information with each player

R. Shumaker (Ed.): Virtual and Mixed Reality, Part I, HCII 2011, LNCS 6773, pp. 61–68, 2011.

on a captured playfield image during a game. This information includes name, team, position, and statistics of players and games, which are available via the Web and local information server installed in stadium. Our system is currently based on still images of playfields, which are taken by a mobile phone. The images are can be from different angles, having a large variation in size and pose of players and playground, and different lighting conditions.

The rest of this paper is structured as follows. Section 2 gives an overview and detailed description of the proposed system. The experimental results on the baseball field images are provided in Section 3, and Section 4 concludes the paper.

2 The Proposed System

Figure 1 shows the architecture of the proposed system. This system starts the processing with capturing a still image from a mobile device. We use a still image because of two reasons. The first reason is that users may have some difficulties in holding mobile devices without shaking for a long time while interacting with augmented contents on a live video frames. The second reason is that a still image has higher resolution than an image frame of a video sequence. In general, users take a picture in a long distance during a baseball game. For detecting players, we need a sufficient image resolution of the players. The captured still image is then analyzed to estimate a homography between a playfield template and the imaged playfield, and to detect the location of each player. If the analysis is performed successfully, the game contents are received by accessing the information server. A user can touch an interested player on the mobile phone screen. A best candidate of the corresponding player, then, is found by a simple method combining the detected player location and the game information with some boundary constraints. Finally, team and name of the player is augmented above the touched player. Detail information is displayed on a new screen when the user touches the player's name. We use a new screen, because the screen size is too small to display the whole information on the field image. Figure 5 shows an example of the results.

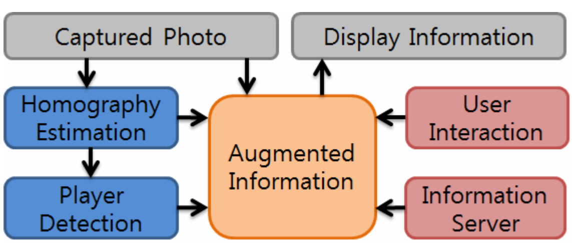

Fig. 1. The proposed system architecture

2.1 Planar Homography Estimation

One of the main techniques for AR technology is the homography estimation. In general, there are two different approaches for the homography estimation from a single image. First one is a marker-based approach that uses image patterns that are specially designed for homography estimation. Second one is a markerless approach that does not use those patterns, but is restricted to natural images that contain distinctive local patterns. However, baseball playfield images include formalized geometric primitives that are hard to distinguish between an input frame from the reference frame based on local patterns. Therefore, we propose a baseball playfield registration method by matching the playfield shape, which consists of geometric primitives. This method is divided into three steps. First, contours are extracted based on edges between the dominant color (e.g. grass) and others. Secondly, geometric primitives, like lines and ellipses, are estimated by using parameter estimation methods. Third, homography is estimated by matching those geometric primitives.

Edge Contours Extraction. Unlike other sports playfield that have well-defined white line structure, grass and soil colors are two dominant colors in baseball field [7]. The infield edge pixels define most of the shape structure. Actually, foul lines are designated with a white line. It is not enough to estimate the projective transformation. To detect the edge pixels, grass-soil playfield segmentation approach is considered [8]. However, according to the empirical analysis of the colors, grass pixels have dominant component of green in RGB color space. By setting a pixel with larger green component than red component as grass, we get a reliable pixel classification result as shown in Figure 2(b). Noise removal is followed by applying a Median filtering. Note that we do not filter out the background areas, such as sky and spectators, because the homography estimation step removes these areas automatically.

After pixel classification, an edge detection algorithm is applied to detect edge pixels. There are many methods in the literature to detect edges from an image. In this case, a simple edge detection method that detects pixels of grass area adjacent to other area is developed. We set as edge the pixels that have both of grass and other components in a 3x3 window. The detected edges are shown in Figure 2(c). Finally, edge pixels are linked together into lists of sequential edge points, one list for each edge-contour for discriminating the connectivity. Note that small segments and holes are removed by discarding contours that have smaller length than 50.

Geometric Primitives Estimation. The infield structure of a baseball field consists of two different types of shape, line and ellipse. Starting with the detected edge contours, line and ellipse parameters are extracted. Brief descriptions of the estimation methods are as follows.

Line segmentation method is used to form straight-line segments from an edge-contour by slightly modifying Peter Kovesis implementation [9]. The start and end positions of a line segment are determined, and the line-parameters are further refined with a least-square line fitting algorithm. Finally, nearby line segments with similar parameters are joined. The final line segmentation results are shown in Figure 2(d).

There are two possible ellipses, the pitcher's mound and the home plate, in a baseball field. It is hard to detect the elliptical shape of home plate in general, because it is not separated into a single edge-contour. Therefore, in our system, the pitcher's mound is considered as the best detectable ellipse in a playfield. A direct least squares ellipse fitting algorithm is utilized in each edge-contour for ellipse parameter estimation [10]. Then, we can find the pitcher's mound as the ellipse with minimum error smaller than a pre-defined threshold by using ellipse fitness function. Finally, the estimated ellipse is verified by fine matching based on sum of squared difference (SSD). Note that we assume that the observed image contains the pitcher's mound. The final detected ellipse is shown in red in Figure 2(d) (The figure is best viewed in color).

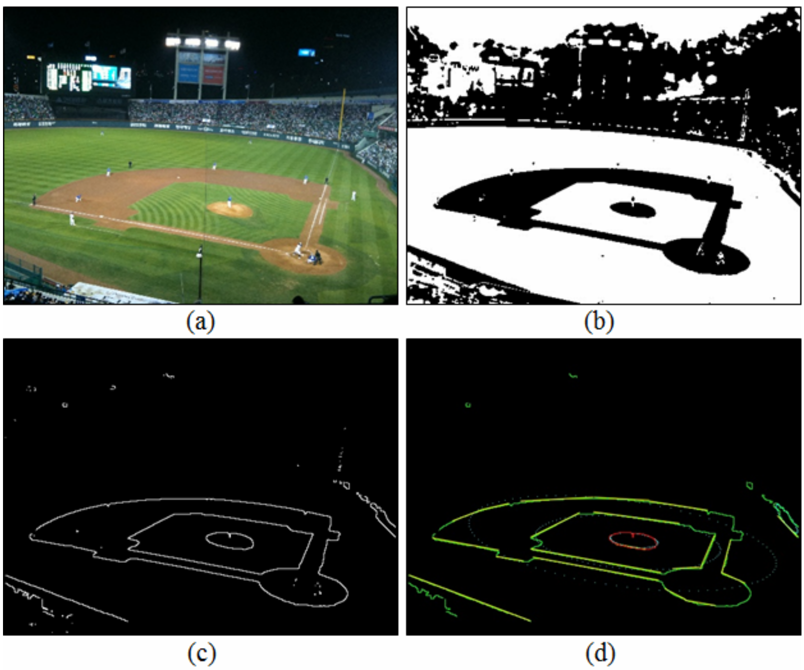

Fig. 2. Contours extraction and geometric primitives estimation: input image (a), classified grass pixels (white) (b), detected edges (c), detected lines (yellow) and an ellipse (red) (d)

Homography Estimation. A diamond shape which consists of four line segments is located inside of the infield and a circle, pitcher's mound, exists at the very center of the diamond. Outside the diamond, two foul lines are located. Hence, we define the playfield model composed of six line segments and a circle. The defined model is shown in Figure 3(a). Now, homography estimation is thought as a matter of finding correspondences between the model and a set of extracted shapes from the observed image. Our solution utilizes four line correspondences with two sets of parallel lines in the diamond shape of the playfield. A transformation matrix is determined immediately by using the normalized direct linear transformation [11].

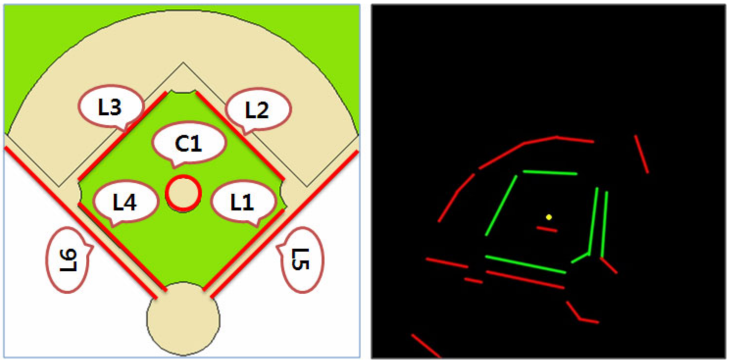

Fig. 3. The defined playfield model (6 lines and a circle) (left) and the geometrical constraints (green: selected lines, red: removed lines) (right)

Searching for the best correspondence requires a combinatorial search that can be computationally complex. Hence, we try geometrical constraints. Since we don't know any shape correspondences except the pitcher's mound, metric and scale properties are recovered roughly using relationship between a circle and an ellipse without considering perspective parameters. Then, all the extracted line segments are sorted in counter-clock wise order by an absolute angle of line joining the center of the ellipse and the center of each line segment. And we remove the line segments beyond the scope of pre-defined length from the center of the ellipse to the center of each line segment. We also applied some minor constraints by utilizing similar techniques proposed in the literature [7]. These constraints resolve the image reflection problem and reduce the number of search significantly as shown in Figure 3, where many lines are removed by geometrical constraints. For each configuration that satisfies these constraints, we compute the transformation matrix and the complete model matching error as described in [7]. The transformation matrix that gives the minimum error is selected as the best transformation. Finally, the estimated homography is verified by fine matching based on SSD. Figure 4 shows a transformed playfield model that is drawn over an input image using the estimated homography.

2.2 Player Detection

For automatic player detection in our framework, we use the AdaBoost learning based on histograms of oriented gradients that gives somewhat satisfied detection rate and fast search speed [13]. Dalal & Triggs show experimentally that grids of Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for human detection [14]. A feature selection algorithm, AdaBoost, is performed to automatically select a small set of discriminative HOG features with orientation information in order to achieve robust detection results. More details of this employed approach can be found in [13].

This approach is designed to use a larger set of blocks that vary in sizes, locations and aspect ratios. Therefore, it is possible to detect variable-size players in images. If we know the search block size, it improves the detection accuracy and reduces the searching time. In the proposed system, the search block size is calculated by using the average height of Korean baseball players (i.e. 182.9 *cm*) [12]. At first, a camera pose is estimated using a robust pose estimation technique from a planar target [15]. Input parameters for the pose estimation are the four corresponding lines that are used to estimate a homography. Next, a search block size at each pixel location is calculated approximately using a given camera pose and average height. In a baseball game, most of the interested players are inside a baseball field. The detected players outside the field are not considered. An example of player detection results is shown in Figure 4.

Fig. 4. Homography estimation and player detection: a transformed playfield model (left), detected players (green box) and a player outside the playfield (red box) (right)

3 Experimental Results

We have tested the proposed algorithm using photos taken with Apple iPhone 3GS and 4 on a PC with an Intel 2.67 GHz Core I7 CPU. The pictures were taken at Jamsil and Mokdong baseball stadiums in Seoul, Korea. Images were resized two different resolutions, 640 x 480 and 960 x 720, that are used to estimate homography and to detect players including outfielders respectively. Homography estimation time always remained between 50 and 100 *ms*. The time costs to detect all players are much longer than this. However, there is no need to search all the pixels inside the baseball field, because only an interested player is searched within the small region that is selected by a user.

We also implemented the system on a mobile platform (Apple iPhone 4). The whole steps were processed within two seconds. The information server manages contexts of baseball games held in Korea. The mobile device connects to the information server via wireless network after the image processing step. As we know, the information server does not provide the exact location of each player. Therefore, we roughly matched the detected player with the given information by inference based on the team, the position, and the detected location. Figure 5 shows results of the system after touching an interested player on the mobile phone screen.

Fig. 5. The implemented system on a mobile platform: the upper screen displays team, name (over the player), and position (upper-right) in Korean text after touching an interested player and the lower screen displays the detail information of the player

4 Conclusion and Future Work

We have described the vision-based Augmented Reality system that displays supplementary information of players on a mobile device during a baseball game. Since homography estimation plays an important role in this system, we propose a new estimation method to fit a baseball field. As a player detection method, we employ the fast and robust algorithm based on Adaboost learning that gives somewhat satisfied detection rate and search speed. However, sometimes we fail to detect players. Further improvement of the detection rate remains as a future work. We have successfully implemented the system on a mobile platform and tested the system two different stadiums.

Our current system does not cover every baseball stadiums, because the proposed pixel classification algorithm is based on the playfield consists of grass and soil. However, we found that there are various types of playfield in the world. For example, some stadiums have a playfield that is painted white lines on a green field. Therefore, our next goal is to develop a system that satisfies these various types of playfield.

References

1. Hurwitz, A., Jeffs, A.: EYEPLY: Baseball proof of concept - Mobile augmentation for entertainment and shopping venues. In: IEEE International Symposium on ISMAR-AMH 2009, pp. 55–56 (2009)
2. Demiris, A.M., Garcia, C., Malerczyk, C., Klein, K., Walczak, K., Kerbiriou, P., Bouville, C., Traka, M., Reusens, E., Boyle, E., Wingbermuhle, J., Ioannidis, N.: Sprinting Along with the Olympic Champions: Personalized, Interactive Broadcasting using Mixed Reality Techniques and MPEG-4. In: Proc. of BIS 2002, Business Information Systems (2002)
3. Inamoto, N., Saito, H.: Free viewpoint video synthesis and presentation of sporting events for mixed reality entertainment. In: Proc. of ACM ACE, vol. 74, pp. 42–50 (2004)
4. Matsui, K., Iwase, M., Agata, M., Tanaka, T., Ohnishi, N.: Soccer image sequence computed by a virtual camera. In: Proc. of CVPR, pp. 860–865 (1998)
5. Kammann, T.D.: Interactive Augmented Reality in Digital Broadcasting Environments. Diploma Thesis, University of Koblenz-Landau (2005)
6. Han, J., Farin, D., de With, P.H.N.: A Real-Time Augmented-Reality System for Sports Broadcast Video Enhancement. In: Proc. of ACM Multimedia, pp. 337–340 (2007)
7. Farin, D., Han, J., de With, P.: Fast Camera Calibration for the Analysis of Sport Sequences. In: IEEE Int. Conf. Multimedia Expo (ICME 2005), pp. 482–485 (2005)
8. Kuo, C.-M., Hung, M.-H., Hsieh, C.-H.: Baseball Playfield Segmentation Using Adaptive Gaussian Mixture Models. In: International Conference on Innovative Computing, Information and Control, pp. 360–363 (2008)
9. Nguyen, T.M., Ahuja, S., Wu, Q.M.: A real-time ellipse detection based on edge grouping. In: Proc. of the IEEE International Conference on Systems, Man and Cybernetics, pp. 3280–3286 (2009)
10. Halir, R., Flusser, J.: Numerically stable direct least squares fitting of ellipses. In: 6th International Conference on Computer Graphics and Visualization (1998)
11. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
12. Korea Baseball Organization: Guide Book (2010), `http://www.koreabaseball.com`
13. Zhu, Q., Avidan, S., Yeh, M.C., Cheng, K.T.: Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. In: IEEE Conf. on CVPR, pp. 1491–1498. IEEE Computer Society Press, Los Alamitos (2006)
14. Dalal, N., Triggs, B.: Histogram of Oriented Gradients for Human Detection. In: IEEE Conf. on Computer Vision and Pattern Recognition, CVPR, vol. 2, pp. 886–893 (2005)
15. Schweighofer, G., Pinz, A.: Robust Pose Estimation from a Planar Target. IEEE Trans. on Pattern Analysis and Machine Intelligence 28, 2024–2030 (2005)