Panoramic Video Generation by Multi View Data Synthesis

T. D'Orazio, M. Leo, N. Mosca

Istituto di Studi sui Sistemi Intelligenti per l'Automazione, C.N.R.
Via G. Amendola 122/D 70126 Bari, Italy
dorazio@ba.issia.cnr.it

Abstract

This paper presents a mosaic based approach for enlarged view soccer video production that can be provided to the audience as a complementary view for greater enjoyment of relevant events, such as offside, counter attack or goal, that spread out all over the playing field. Firstly, an enlarged view of the whole field is produced by fusing the images of six cameras placed on the two sides of the field. Then a color transformation is applied to have uniform colors on the parts of the playing field acquired from different cameras. Finally, the players are segmented by each camera and projected onto the enlarged view to produce videos of the most interesting events.

1 Introduction

Soccer is one of the most popular team sports in the world. For this reason sport broadcasters are very interested in producing videos that could attract and entertain the audience with more information and additional data on the observed events. In [8] a method for virtual view synthesis that allows viewers to virtually fly through real soccer scenes is presented. A mosaic based approach to event mining in sports video is presented in [7]. The specific problem of image-to-model homography estimation has been faced in [5] to avoid recourse to any kind of sensor on the cameras and on the players and to allow automatic and fast rectification of sport video sequences. Also in [9] a method for computing the position, the orientation and focal length of a camera is presented as a preliminary and strict requirement to be able to overlay graphics on the image. In order to identify objects, track them between frames and add enhanced content in multimedia applications, in [1], a novel approach to player segmentation and tracking, that operates on labeled regions and uses an amount of prior knowledge, is presented. Many papers have been presented in literature on panoramic image generation. An overview of the current state of research that points out the essential problems to consider and various approaches for their solutions is presented in [4]. The problem of stitching two or more non overlapping scenes together has been faced by the remote sensing community to generate wide image mosaic based on satellite scenes. In [2] the authors try to produce visually pleasing mosaics by placing the seam lines along salient image features present in all the overlapping images whose grey level difference is greatest. Corresponding feature points are selected in [6] in overlapped UAV (unmanned aircraft vehicle) snapshot images, then the transformation parameters are estimated to generate the panoramic view by adjusting detailed positioning without apparent errors.

In this paper we propose an automatic system that generates, by image mosaicing, soccer videos with an enlarged view that cover the whole playing field. Broadcast cameras cover just a portion of the field, while it could be interesting to have an overall view to see the arrangement of players during offside, counter attack or goal actions. These videos provide the audience with the ability to perceive the scene as a whole as if they were on the stands, since the view is enlarged and it is possible to see all the players moving on the field simultaneously.

2 System Overview

The main difficulties associated with the production of novel views starting from the images acquired by different cameras are related to different aspects: first of all, it is necessary to know the objects' depths in the scene in order to locate them in the new view. Second, different cameras provide different appearances of the same objects both for their intrinsic differences in the color calibration and both for their different positions in the field. Finally, the image projection and rectification that are generally applied in these contexts cannot

be used to project players moving on the field, since their appearances would be crushed onto the field as it happens to all the play-field features. In this paper we propose innovative solutions to the above and related problems.

In figure 1 the system overview is reported. The camera calibration phase is carried out off line, at the beginning when the camera setup is installed, and the parameters saved for all the successive processing. During the match the video streams acquired by the six cameras are processed by two parallel processes: the static panoramic image generation that is carried out periodically by integrating the background images which are continuously updated, and the players projection onto the panoramic images that is carried out for each acquired frame.

The paper is organized as follows: section 3 describes the panoramic view generation; in section 4 the projection of moving players onto the static panoramic image is explained. Some experimental results and discussion are reported in section 5.
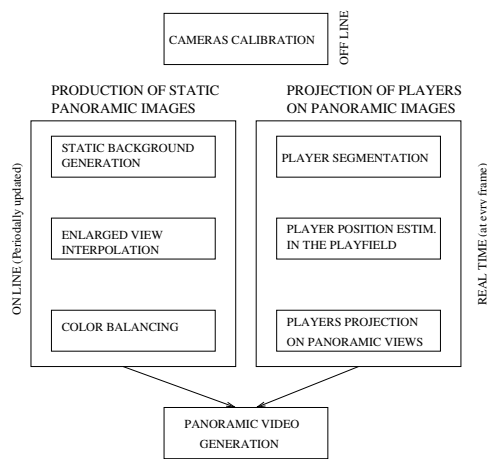

Figure 1. The system overview

3   Panoramic view generation

We used six cameras placed on the opposite sides of the field (three for each side) to take the images shown in figure 2. Although these images have a good temporal and spatial resolution, they cannot be used in their original form for television purposes since they contain some overlapping zones, and miss other zones. Starting from these images we want to generate a panoramic view that contains the whole field as if it was observed by a unique camera placed on one side in the center of the field. In order to generate a realistic panoramic image we have to synthesize three neighbor images in

a unique image in which overlapping zones have to be discarded while missing areas have to be recovered by the opposite views. This step consists of three main tasks: the generation of static background images without moving players that must be the starting images for the generation of the panoramic view; the extraction of the areas of interest from all the images to interpolate a new image that shares the projective geometry constraints of the central cameras; the color calibration to generate uniform colors in all the six areas coming from different cameras.

In order to have a panoramic view that is realistic and as close as possible to the reality, it is necessary to start from static images that are consistent with the actual frames. These images have to be continuously updated to keep up with the continuous changes of lighting conditions, and above all have to contain only static regions. For this reason we have used a background modeling algorithm that is able to generate a reference image discarding moving points and considering the mean values of those points that, observed in a temporal window, have a low contents of energy. Details of this algorithm can be found in [3]. For the purpose of this paper it is necessary to say that this algorithm continuously updates the background models of the six views and these models are used to create at regular intervals the new updated panoramic view.

The projective geometry between the cameras and the field can be estimated during a calibration phase by using the knowledge of the positions of some points in the field and the estimation of the corresponding coordinates in the image planes. Let $P(kx_i, ky_i, k)$ be a generic point belonging to the ground plane (the playfield) and $p(u_i, v_i, 1)$ the corresponding point in the image plane. The relation between P and p can be expressed as follows $P = M \cdot p$, where the 9 elements of the matrix M are unknown and can be calculated knowing the intrinsic and extrinsic camera parameters. Anyway they can be computed if at least 4 pairs of points whose coordinates are known a priori both onto the ground planes and in the image plane solving an homogenous linear system (one of the elements of M can be set to 1). For this reason, during an off-line process some reference points have been placed onto all the fields and their coordinates have been measured both into a 3D world coordinate system and into the image plane coordinate systems of the six cameras. For each camera the elements of the matrices M are estimated solving the corresponding equation systems. In this way it is possible, knowing the position of one point in each image plane, to estimate its corresponding position onto the playing field, and inverting the matrices M it is possible also the opposite transformation. In the following

Figure 2. The global view of the field provided by the system at time t joining the images acquired from the six cameras placed in the stadium. We will call the three cameras on one side C6, C4 and C2, an the three cameras on the other side C5, C3, and C1 (from the left to the right, from the top to the bottom the corresponding images).

section we will indicate with M_i the direct matrix of the $i - th$ camera while with M_i^{-1} the corresponding inverted matrix.

In order to generate a panoramic view containing all the field as if it was observed by a unique central camera we have to produce a synthetic view starting from the six images reported in figure 2 and by using the estimated homographic matrices in the calibration phase. We decided to have a principal side of the field, that is the one with cameras 5, 3 and 1, and also to assign to camera 3 the role of principal view whose intrinsic color properties have to be obtained from the other cameras. Starting from the initial image dimension of $1920x1080$ we want to generate the new image I^n with a dimension of $(1920x3)x1080$. The (u, v) coordinates of the new image I^n vary in the range $[0, 5759]$ and $[0, 1079]$ respectively. The central part of this image is directly extracted by camera 3 (i.e. for $u \in [1920, 3839]$ $I^n(u, v) = I^3(u, v)$) while the other two parts have to be generated by the other views. We consider all the points of the left and right sides of the new image (i.e. with $u \in [0, 1919] \vee u \in [3840, 5759]$) and we project them onto the playing field by using the matrix M_3. According to the area in which these points fall we use the proper inverted matrices to re-project the points on the initial camera views to extract the corresponding color.

The knowledge of the areas of the play-field covered by adjacent cameras' field of views allows the color balancing between different views. Comparing the colors, a brightness transfer function is generated to balance the colors of the left and right parts of the image 3 in order to have a uniform appearance of the field. The most difficult task is the adjustment of the two cones

in the lower part of the image: since these two areas are generated by re-projecting points on the camera 4 image plane, the general appearance of the play-field is quite different. This is due to the fact that the grass of the play-field is mowed in two opposite directions that produce an inverse appearance of the same stripe from opposite cameras. In this case it is necessary a color transformation that re-produce the inversion of the light green stripe with the dark green stripe.

4 Players and Ball Projections

Details of the ball and player segmentation and tracking algorithms can be found in [3]. When the ball and the players are observed in the areas covered by camera 1, 3 and 5 they are projected directly on the panoramic images. When the players are in areas 2, 4, and 6 covered only by the opposite cameras, they cannot be projected as they appear, but a shape resizing is required to adapt them to the new appearances in the opposite field of view. Indeed players that appear very far from the cameras on one side are really very near when observed by the opposite points of view of the cameras on the other side. This problem can be solved by using the homographic transformations among the cameras and the playing field again. From the blob position and dimension of the segmented player in an image plane it is possible to recover its position in the field and to estimate its height. Then, the re-projection of this shape in the opposite camera allows the proper sizing of players in the different field of view.

Figure 3. The image obtained by re-projecting the players segmented in the six views on the panoramic view of the field.

5 Experimental Results on Automatic Video Production

In figure 3 one of the images of the resulting video is shown. It has been extracted from the same frame as the 6 images shown in figure 2. The two players with dark uniforms that are not visible in cameras 1 and 3, but only in the opposite cameras 2 and 4, have been projected with a resizing of their shapes. It is clear how it is possible to appreciate with this new view the positions of all the players in the field. The observer has the impression of observing the whole field as if he is positioned in the central part of the stand. This feeling is more evident when a video is observed. It is possible to see all the players, moving simultaneously in the field and appreciate the team strategies in different playing actions.

This panoramic view can be very useful for the audience of broadcast programs, especially with the introduction of full HD tvs and broadcasting or monitors with higher resolutions. There is added value for soccer technicians as well, given the possibility to evaluate globally players movements, team strategies and so on. Certainly some problems remain to be solved. The uniformity of colors among different cameras is not always guaranteed especially during sunny days in which opposite cameras have different lighting conditions due to the presence of the sun. Another problem introduced by opposite cameras is that what is perceived as a left movement on one camera, will be viewed as a right movement on the other camera, and therefore the image must be flipped horizontally. In the same way, players that face frontally one of the cameras, show their back when viewed by the opposite camera. This problem does not have a simple solution if the player is seen only by the opposite camera. One solution, not currently employed, would be to find suitably similar views of the same or similar players, but showing the other side of him. Techniques such as chamfer matching, working on players silhouettes, along with the construction of a multi-view player database can help to counteract this problem, but it is being considered as future work.

6 Acknowledgements

The authors thank dr. Liborio Capozzo for technical support in the setup of the six cameras used in this work.

References

[1] E. L. Andrade, J. Woods, E. Khan, and M. Ghanbari. Region-based analysis and retrieval for tracking of semantic objects and provision of augmented information in intercative sport scenes. *IEEE Transaction on Multimedia*, 7(6), december 2005.

[2] C. Bileski, J. Grazzini, and P. Soille. Automated morphological image composition for mosaicing large image data sets. *IEEE International Geoscience and Remote Sensing Symposium*, 7(6), 23-28 July 2007.

[3] T. D'Orazio, M. Leo, P. Spagnolo, P. Mazzeo, N. Mosca, M. Nitti, and A. Distante. An investigation into the feasibility of real-time soccer offside detection from a multiple camera system. *IEEE Trans. on Cirtuit and System for Video Technology*, 19(12):1804–1818, December 2009.

[4] D. Farin, M. Haller, A. Krutz, and T. Sikora. Recent developments in panoramic image generation and sprite coding. *10th IEEE Workshop on Multimedia Signal Processing*, pages 64–69, 8-10 October 2008.

[5] J. Hayet, J. Piater, and J. Verly. Robust incremental rectification of sport video sequences. *Proc. of the British Machine Vision Conference*, pages 687–696, 2004.

[6] Y. Huang, J. Li, and N. Fan. Image mosaicing for uav application. *KAM*, pages 663–667, 21-22 December 2008.

[7] T. Mei and X. Hua. Structure and event mining in sports video with efficient mosaic. *Multimedia Tools Application*, 40(1):89–110, October 2008.

[8] N.Inamoto and H.Saito. Virtual viewpoint replay for soccer match by view interpolation from multiple cameras. *IEEE Trans. on Multimendia*, 9(6):1155–1166, 2007.

[9] G. Thomas. Real time camera tracking using sport pithc markings. *Journal of Real Time Image Processing*, 2:117–132, 2007.