Video platform for player tracking and enhanced visualization of sport events

E. Carrasco[a], I. Barandiarán[a]

[a] *VICOMTech, Paseo Mikeletegi 57, Donostia – San Sebastián, E-20009, SPAIN*

Abstract

In this paper, the recent developments of a video platform for sport event enhancement are described. Firstly, a new algorithm for on-line tracking of players moving in a playing ground from still camera video images is presented. Secondly, a method for realistic and seamlessly augmentation of the sport scene in on-line video with rendered 3D information and graphics is shown.

The on-line player tracking algorithm has been tested in real conditions and good performance has been proved. Comparison with other tracking technologies has shown satisfactory results. A demonstrator of on-line video augmentation has been built and tested in laboratory conditions. On-line 3D information overlaying capabilities have been successfully tested. The demonstrator has been evaluated with two different tracking systems: an optical tracking system based on the reflection on markers of infrared light and the on-line player tracking algorithm developed in the framework of this project.

Final results show that the proposed on-line video platform for player tracking and enhanced visualization of sport events will contribute to better and sooner evaluate the performance of sport teams and to make broadcasted sport events more appealing.

Keywords: Player On-line Video Tracking, On-line Enhanced Visualization, Sport Teams Performance Monitoring.

1. Introduction

In the last few years, many companies dedicated to the analysis of football (soccer) matches have emerged. These firms shoot the matches and then extract all relevant information, mainly consisting in the indexation of mayor events and the localization of each player in the playing ground. In many cases, most work is carried out manually, requiring an important number of operators to work full time to provide the trainer with the results in two or three days time.

The information resulting from this service is very valuable. For example, the trainer can use it to analyze the football match in order to evaluate the performance of the team. Nevertheless, the drawbacks of this service are its high cost and the long time they have to wait for the results.

In this scenario, computer vision techniques can be used to automate the tracking of the players in the playing ground from the video sequences of the match. With these techniques, manual input can be drastically reduced, and produce cheaper, quicker and more accurate results. Quite a lot research has been carried out in this field in last years. However, current work remains experimental and far from practical application.

Moreover, while the player tracking information is currently only delivered to the coachers, it is expected that in a near future some of it would be broadcasted on-line to the audiences of the football matches. We believe that this information would be very well received by the audience and could help to increase the appealing of the sport events. In order to deliver such information, software tools for the enhancement of on-line videos have to be developed and integrated with on-line tracking systems.

In this paper we present our developments of an on-line video platform for player tracking and enhanced visualization of sport events. In section 2, the state of the art will be summarized. Then, in section 3, the new algorithm for on-line tracking of players will be described. In section 4, the on-line 3D enhanced video visualization system will be shown. Next sections 5 and 6 will deal with the testing of the demonstrators in laboratory and real conditions. Finally, the results will be summarized in section 7 and conclusions will be drawn up in section 8.

2. State of the art

Actually, the two most important companies dedicated to the analysis of football matches using video information are Sport Universal Process [1] and ProZone [2]. Although major football teams are their customers, their requirements are not yet fully met. Main disadvantages consist in price, match analysis information retrieval time and usefulness of the given information [3].

It is largely assumed that computer vision is going to be the key technology in the upgrading of the actual sport analysis systems. Compared to other tracking technologies, major advantages of computer vision technology is that is not invasive and that the opponent team can be analyzed also.

In the lasts years, much research has been carried out in the field of tracking of players from video images. The key issue here consists of developing a fast, simple and robust segmentation technique of the players to track. To deal with this challenge, several computer vision techniques based on different concepts have been proposed: template segmentation, colour identification, border detection, image subtraction… Up to the moment, all these techniques have proved to have a limited success. So, the actual trend aims at combining these techniques to enhance the segmenting and tracking performance. Finally, more efforts have to be done to develop real time algorithms [4].

On the other hand, real-time tracking data of the players in the match is a very valuable information, both for the coachers and for the general public attending a sport event. Showing this information on-line, seamlessly integrated with the sport scenario and with rich 3D advanced graphics may increase the appealing of the sport events.

Currently, there are several companies working in this field, such as Orad [5], and enhanced display information is becoming quite usual in several TV broadcasted sports events. However, they are commonly simple, static 2D graphics. Advanced 3D graphics, realistically and seamlessly integrated in the scene and linked with on-line tracking systems are going to come next.

The state of the art in computer graphics knowledge is very advanced today. Actual hardware offers really high performance at very low cost and there are several programming languages and libraries that make building new systems simple, reliable and fast.

On the research side, the main topic here consists of the on-line integration of 3D computer graphics seamlessly in the sport video sequences shoot from cameras prone to change its pose and internal parameters (i.e. zooming for example) [6].

3. On-line video tracking of players

The objective of this section is to describe an algorithm which is capable of tracking on-line all the players of the same team that are in the field of view of a single still video camera shooting one defined area of a football playing ground. The tracking information extracted from the video image frames will consist of the 2D positioning of the players in the playing ground. This information will be estimated for each player of the same team, i.e. players dressed with the same colours. This information will be computed for each shoot video frame, so trajectories, distances and speeds of each player can also be computed in time.

The algorithm architecture starts with a manual initialisation. This initialisation consists of selecting the area to track, to calibrate the playing ground plane and to identify the colour of the players to track. Then, computations are done automatically on each new frame to segment all the players and to match the new positions in relation to the previous ones. The proposed architecture has been depicted in Figure 1.

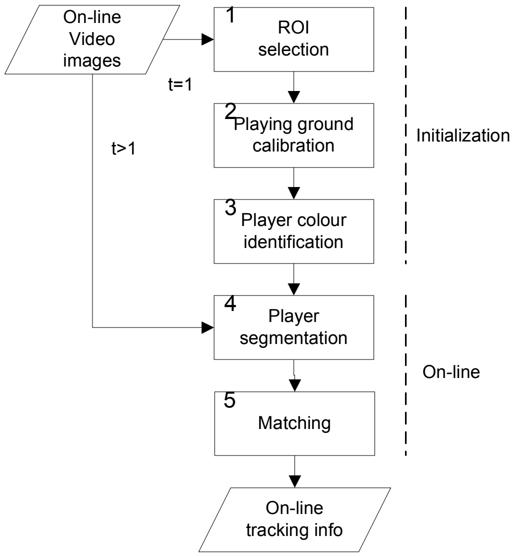

Fig. 1: Architecture of the proposed player tracking algorithm.

In the following sections each module of the proposed algorithm is described in detail.

3.1. Selection of the Region of Interest

First, the region of interest has to be selected. To accomplish this task, four non-collinear points have to be selected manually on the image plane. These four points define a quadrilateral. The region of the image outside this quadrilateral will not be taken into account in the following calculations.

3.2. Playing ground calibration

Then, the playing ground plane has to be calibrated in order to compute the real distances covered by the tracked players from the images taken by the video camera.

The simplest way to calibrate the playing ground consists of computing the 2D homography transformation between the playing ground plane and the image plane (Eq. 1).

$$m_c = \mathrm{H} m_r \quad (1)$$

where,
m_r : 2D point in playing ground plane
H : 2D homography transformation matrix
m_c : 2D projected point in the image plane

The 2D homography transformation matrix can be easily computed with only four known 2D point correspondences between the playing ground plane and the image plane using the Normalized Direct Linear Transformation algorithm [7]. The image plane points are selected manually directly on the video images, and the coordinates of the corresponding points from the playing ground plane are introduced manually.

3.3. Player colour identification

The next step consists of the identification of the colour of the players' shirts of the team to be tracked. For this task, a still frame is needed where some players to be tracked appear. Then, a representative shirt of a player has to be cropped, which is carried out using a manual region of interest segmenting tool.

Next, the main colour of the shirt has to be identified. This task is carried out applying several filters on the rectangular subimage of the player's shirt. This filter pipeline is addressed next.

First, a low-pass filter is applied to smooth the image. Next, three histograms are calculated corresponding to each RGB colour plane. Then, the maximum value of each histogram is computed. These values will be considered as the mean value of the colour of the players of the team to track. Finally, the variance values of the identified colour are computed. This is achieved by matching each histogram to a Gauss curve. More details on this techniques can be found at [8].

3.3. Real-time player segmentation

Provided the mean and the variance values of the colour of the players of the team to track, the players can be segmented from each frame of the real-time video sequence.

To accomplish this task, next steps have to be followed. First, the whole frame is smoothed by means of a low-pass filter. Then, the image is binarized using the mean and variance values computed in the previous stage as threshold values. In this binary image, regions to one are identified players, whereas zero is all the rest of the image. Finally, all 8-connected blobs are labelled and its areas and centroids are computed. The size of the area of each blob reflects the visibility of the corresponding player and the centroid will be considered as his centre of mass. [8]

3.4. Real-time player correspondence matching

Once the centroids of the players of the same team in the first frame are known, the correct correspondences in time with the centroids in next frames are computed. The technique described here uses the physical maximum speed of a player over the playing ground (10m/s) on two consecutive frames and the 2D homography transformation computed in Section 3.2 to estimate the maximum area that can be covered by a player in next frame. Finally, the matching centroid on the next frame will be the one that lies inside this area and has minimum distance with the centroid of the previous frame. With this technique, the number of possible candidates is highly reduced in order to accelerate the search for the right one.

4. On-line 3D enhanced video visualization system

The aim of this system is to enhance the visualization of on-line video of sport events through the use of advanced 3D computer graphics realistically and seamlessly inserted in the scene and which are updated with the on-line tracked information of the players (position, distance, trajectory, speed…).

The system is designed for one still camera shooting one defined part of the whole football playing ground. Dealing with moving cameras is a more challenging problem and it will be faced in our future work. Besides, the camera is shooting only one part of the whole playing ground because otherwise resolution might be lost and the tracking would therefore become less robust and accurate.

Then, in the next sections the architecture of the system is described. The still camera is first calibrated, then its pose is estimated, and finally, camera calibration and pose are transferred to a graphical engine which generates all the 3D computer graphics and mixes them with the on-line video of the sport event.

4.1. Camera calibration

First, a model for the camera is needed. In this sense, the pinhole camera model is chosen because is the most common one for this type of applications [7]. Pinhole camera model is shown in Fig.2.

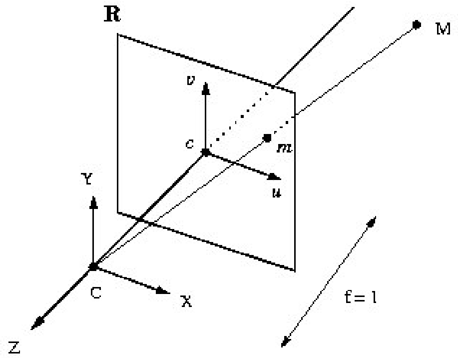

Fig. 2: Pinhole camera model.

Then, the camera model has to be calibrated, i.e., the internal parameters of the camera have to be estimated. Once the camera is calibrated, the projection of 3D points in the image plane can be computed. Mathematically this is expressed in Eq.2.

$$m_c = \mathrm{K} M_c \qquad (2)$$

where,
M_c : point in 3D space
K : internal parameter matrix of the camera
m_c : projected point in the image plane

To estimate de values of K matrix, the algorithm described in [9] is used. This calibration requires several snapshots of a check board drawing. This calibration is done only once provided that no camera parameters are not going to be changed.

4.2. Camera position and orientation computation

Provided that all the internal parameters of the camera are known (Sec. 4.1), then the position and orientation (pose) of the camera has to be estimated next. This information is needed because it is necessary to express all 3D points of the real scene in the coordinate frame of the camera before projecting them in the image plane. This concept is shown is Eq.3.

$$M_c = \mathrm{T}^{c}_{r} M_r \qquad (3)$$

where,
M_r : 3D point in playing ground reference
T^{c}_{r} : Transformation matrix from playing ground to camera reference
M_c : 3D point in camera reference

A minimal solution for a calibrated camera requires only three known image points with their corresponding points in the real coordinate frame. These three points are manually identified in the image frames shoot by the camera and their corresponding 3D real point coordinate values are introduced manually.

Mathematical solutions to the problem of estimation of the pose of a camera from three point correspondences reoccur often in the literature; our approach has consisted this time of using the Normalized DLT algorithm [7,10,11].

Now, once pose, T^{c}_{r}, has been estimated, from Eq.2 and Eq.3, final Eq. 4 can be obtained.

$$m_c = KT^{c}_{r}M_r \qquad (4)$$

4.3. Virtual scene generation

Once the internal parameters and the pose of the camera are estimated, a virtual scene is generated. This virtual scene is parallel to the real setting and is built to generate the virtual information and 3D computer graphics with the same perspective of the real scenario. All this computer generated information is overlaid on the real video images in the next step. Finally, the virtual scene has been built using *OpenGL* and many different virtual advanced graphical objects can be loaded into it (arrows, shadows, etc.). These objects may remain static in the scene, may move in a predefined way or may be updated with the coordinates of the players that are being tracked in the real setting.

4. Laboratory tests

All the techniques described in this paper have been tested first in *Matlab* and then implemented in *C++* and *OpenCV*. The final prototype has been tested first in laboratory conditions at the premises of VICOMTech.

The tests have been scheduled in two stages. First, the on-line 3D enhanced video visualization system has been tested. The aim of the test consisted of evaluating the graphic possibilities vs. hardware requirements of the system. A snapshot of the tests carried out on the system is shown in Figure 3.

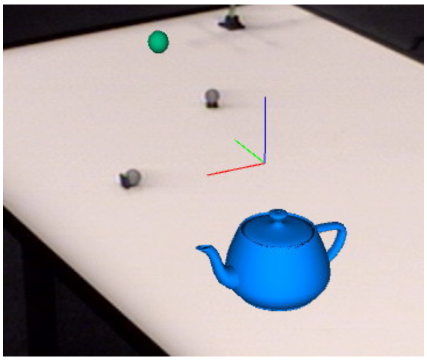

Fig. 3: Snapshot of the video output of the tests carried out on the 3D enhanced video visualization system.

Then, tests have been done using two different tracking systems: a commercial optical tracking system based on reflection of infrared light through use of body attached markers [12], and the player on-line video tracking described in these lines. The objective of this test consisted of comparing the performance of the proposed algorithm vs. a commercial solution.

In Figure 4, a snapshot of the monitor output is shown. In this test, the optical tracking system is used to re-create in the lab the conditions of the real football playing ground.

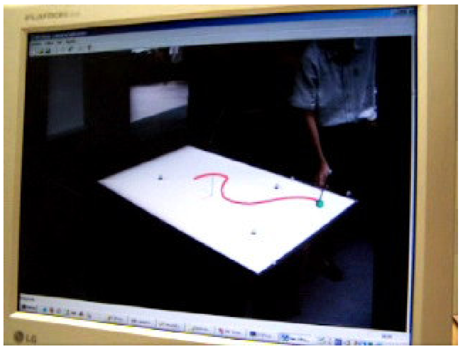

Fig. 4: Laboratory re-creation of tracking of players using a marker based optical infrared light tracking device.

5. Tests on real conditions

The testing of the prototype in real conditions took place in the premises of Zubieta, the training site of the Real Sociedad de Fútbol of San Sebastián, Spain. The test consisted mainly in tracking in real-time the players during different training events. A snapshot of the players being tracked in real-time is shown at Figure 5.

Fig. 5: Players being tracked during a training session.

6. Results and discussion

The results achieved during the tests show that the on-line player video tracking algorithm described in these lines work as good as the commercial optical tracking system, with the advantage of being more simple and no-invasive. An important point to highlight here is that the proposed strategy for tracking players can also be used to analyze the opponent team.

On the other hand, the on-line 3D enhanced video visualization system proposed in these lines can render advanced graphics over on-line video streams using low cost on-the-shelf common hardware. Moreover, these graphics can be realistically inserted, and quickly and smoothly moved in the real 3D scene along with the tracked information.

7. Conclusions

In this paper we presented a tracking algorithm allowing to greatly accelerate the tracking of the players over the playing ground in real matches traditionally done by manual operators. Moreover, the information extracted from the tracking system can be displayed on-line and seamlessly with the assistance of advanced 3D graphics in the video images. These contributions will help to better and sooner evaluate the performance of the sport teams and will make the broadcasted sport events more appealing.

Acknowledgements

First, we would like to express our gratitude to Vicente Serrulla, Consultant of the Real Sociedad de Fútbol, for his expert insight and valuable advices in technology development applied to professional football. Then, we would like to thank our colleagues in GAIA – Asociación Cluster de Teleco-municaciones, Bilbomática and Thaumat – Global Technology Systems for partnering with us within this project and for making such a good working atmosphere. Finally, we would like to thank the Spanish Ministry of Industry, Tourism and Trade for economically supporting this project in the framework of FIT-330100-2005-212.

References

[1] Sport Universal Process website: http://213.30.139.108/sport-universal/uk/index.asp

[2] ProZone website: http://www.pzfootball.co.uk/

[3] Setterwall D. Computerised Video Analysis of Football – Technical and Commercial Possibilities for Footbal Coaching, CID-247, ISSN 1403-0721, Department of Numerical Analysis and Computer Science, KTH – Royal Institute of Science, Stockholm, Sweden, 2003.

[4] Wang J. R. & Parameswaran N. Survey of sports video analysis: research issues and applications. In Proceedings of the Pan-Sydney Area Workshop on Visual information Processing - Volume 36 M, 2004, 87-90.

[5] Orad website: http://www.orad.tv/

[6] Li. M., Lavest J.M. Some Aspects of Zoom-Lens Calibration. IEEE-PAMI Trans. On Pattern Analysis and Machine Intelligence, Vol. 18. No11, pp 1105-1110, Nov 1996.

[7] Hartley R. & Zisserman A. Multiple View Geometry in Computer Vision (2nd edn). Cambridge University Press, 2004.

[8] Gonzalez R.C. & Woods R.E. Digital Image Processing (3rd edn), Prentice Hall; 2006.

[9] Zhang Z. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330-1334, 2000.

[10] Wolfe W.J., Mathis D.,Weber Sklair C. and Magee M. The perspective view of three points. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(1):66-73, January 1991.

[11] Haralick R. M., Lee C., Ottenberg K. and Nölle M. Analysis and solutions of the three point perspective pose estimation problem. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 592-598, 1991.

[12] STT Ingeniería y Sistemas website: http://www.simtechniques.com/