Computer Vision System for Tracking Players in Sports Games

Janez Perš, Stanislav Kovacic
Faculty of Electrical Engineering, University of Ljubljana
Tržaška 25, 1000 Ljubljana
janez.pers@kiss.uni-lj.si, stanislav.kovacic@fe.uni-lj.si

Abstract

The development of computer vision system for tracking players in indoor team games is presented. Several image processing and tracking methods are described, along with camera calibration and lens distortion correction. The output of the system consists of spatio-temporal trajectories of the players, which can be further processed and analyzed by sport experts. In some critical situations the automatic tracking process must be manually interrupted. To correct miss-trackings, human supervision is required. Some experimental results are presented as well.

Keywords: object tracking, human motion analysis, camera calibration, handball

1. Introduction

Human motion analysis is receiving increasing attention from computer vision researchers. This interest is motivated by applications over a wide spectrum of topics (Aggarval and Cai, 1997). In this paper we concentrate on tracking people in team sports based on computer vision. If the trajectories are sufficiently accurately determined, a wealth of additional information, e.g. players speed, acceleration and players interactions can be obtained.

For many years the analysis of a sport event has been based on "observation sheets" filled-in during the match. Later, with a help of video recordings, motion acquisition and analysis were performed manually, a time consuming and tedious task. In the past, progress in introducing the computer vision technology to the team sports domain was slow, due to inadequate video and computational facilities and complexity of the tracking problem itself. The players strive to move rapidly, change direction unpredictably and collide with one another. They violate the smooth motion assumption, on which many tracking algorithms are based. Players appear in the images as highly non-rigid forms, especially due to the movements of their extremities. Many of the proposed approaches solved the motion acquisition and analysis problem only partially and were therefore unable to provide an adequate solution to the sports experts, i.e. tracking every player in the whole field in every instance of time (Erdmann, 1994).

2. Image acquisition

Careful planning of image acquisition could be crucial for the success of the whole project, and inappropriately placed cameras can add a significant degree of difficulty to the tracking problem (Intille and Bobick, 1995). To determine the trajectories of the players from the beginning to the end of the game, the objects in question have to be in the field of view for the duration of the whole match. Two stationary cameras, mounted directly above the court, as shown in Fig. 1 (a) were chosen as the most straightforward solution.
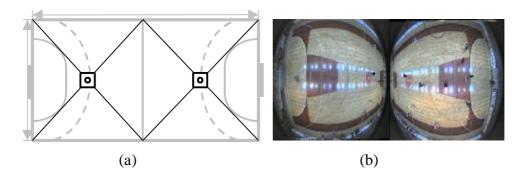
Fig. 1: (a) Camera placement. (b) Image obtained.

Both the handball match and the test video sequence of players moving on predefined paths were recorded using two PAL S-VHS videorecorders. The transfer to digital domain was carried out using the Motion-JPEG frame grabber at 25 frames per second with image resolution of 384x288 pixels. The result is shown in Fig. 1 (b).

3. Camera calibration

Due to significant radial distortion otherwise widely used calibration technique (Tsai, 1987) failed to produce satisfactory results. We decided to take different approach by modeling the radial image distortion more accurately.

Let us imagine an ideal pinhole camera *C*, mounted on a pan-tilt device above the point *0*, as illustrated in Fig 2. *X* is the observed point on the court plane, at distance *R* from the point *0*. *h* is the distance from the camera to the court plane. Angle *a* is the angle of pan-tilt device when observing the point *X*. The differential *dR* of radius *R* is projected to the differential *dr*, which is parallel to the camera image plane. The image of *dr* appears on the image plane. Relations between *dR*, *dr* and *a* are given within the triangle on the enlarged part of Fig. 2 (a).
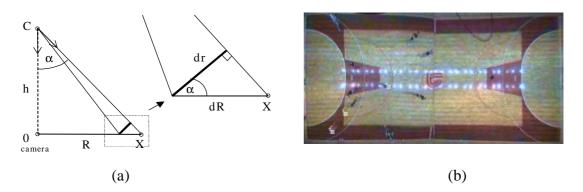

Fig. 2: (a) Model of radial distortion. (b) The resulting images after correction.

Thus, we can write the following relations:

$$dr = \cos(\boldsymbol{a}) \cdot dR, \quad \boldsymbol{a} = \mathrm{arctg}(\frac{R}{h}), \tag{1}$$

$$dr = \cos(\mathrm{arctg}(\frac{R}{h}))dR. \tag{2}$$

Now, let us substitute the pan-tilt camera with fixed camera, equipped with wide-angle lens. The whole area, which would be covered by changing the angle *a* of a pan-tilt camera, is captured simultaneously to the single image of a stationary camera. Additionally, let us

assume the scaling factor between the dr and the image of dr on the image plane to be 1. Therefore, we can obtain the length of the image of radius R on the image plane by integrating left side of the eq. (2) over the interval $(0, r_1)$, and over the interval $(0, R_1)$ on the right side. With R_1 being the distance from the observed point X to the point 0, and r_1 being the distance from the image of point X to the image of point 0 on the image plane, we have the solution of the inverse problem:

$$r_1 = h \cdot \ln\left(\frac{R_1}{h} + \sqrt{1 + \frac{R_1^2}{h^2}} \right) \qquad (3)$$

By solving eq. (3) for R_1 we obtain the formula, which can be used to correct the radial distortion:

$$R_1 = \frac{h}{2} \frac{(e^{-\frac{2r_1}{h}}) - 1}{e^{-\frac{r_1}{h}}}. \qquad (4)$$

4. Player tracking

We developed three different algorithms for use in player tracker: motion detection, template tracking and color based tracking. We tested different combinations of these algorithms; each one has its own advantages and disadvantages.

4.1 Motion detection

The most straightforward approach to motion detection is image subtraction. Each frame C in the sequence was subtracted from the "reference frame" R, i.e. an image of the empty court, yielding the difference image D,

$$D = \left| R_R - R_C \right| + \left| G_R - G_C \right| + \left| B_R - B_C \right|, \qquad (5)$$

where R, G and B denote the red, green and blue components of the current frame and the reference frame, respectively. The difference image D is then thresholded using fixed threshold value and subsequently filtered to reduce image noise. Filtered image is shown in Fig. 3. Resulting blobs in the filtered image are then counted, labeled and their centers of gravity calculated. Some of blobs correspond to the players, while others are caused by noise, shadows and other distracting objects. This limits the efficiency of this technique. Collisions of more than two players and intense player shadows contacting other players require almost always an intervention by human operator. Typical problem is shown in Fig. 3 (b),(c).
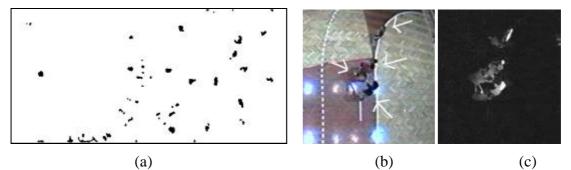

(a)   (b)   (c)

Fig. 3: (a) Blobs detected in the motion detection stage. (b) Typical scene, which confuses the motion detection algorithm. Four players are marked by white arrows. (c) Corresponding difference image.

4.2 Template tracking

Visual differences between the players and the background objects were exploited to further improve tracking process.

The feature set, which could be used to successfully separate objects from background was needed. This is a difficult problem, as shape of the players varies over time and the image resolution is low. Players are represented by relatively small area of only 10-15 pixels in diameter. We have defined a set of 2D functions, i.e. "templates" $\{K_j, j=1,..,14\}$, which extract the very basic appearances of the player. They resemble the Walsh functions, but they are not orthogonal. These functions are shown in Fig. 4 (a).
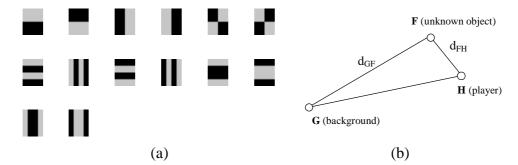

(a)   (b)

Fig. 4: (a) Basic appearances of the player. Black areas represent zeros, while grey areas represent values of 1. (b) Classification of an unknown object.

Each channel of RGB color image is processed separately and the vector of 14 features F for each channel is obtained using the following expression:

$$F_j = \sum_{x=1}^{16}\sum_{y=1}^{16} K_j(x,y) \cdot I(x,y), \quad (j=1,..,14),\qquad (6)$$

where K_j is one of templates, and I is one of the three RGB channels, obtained from the current image. The result is 42-dimensional feature vector F. Additionally, similar vector of features G is calculated from the reference image of the empty playing court at the same coordinates. The third vector H is obtained by averaging the last n vectors of features F, which allows certain adaptivity, as the player appearance changes over time. The simplified, two-dimensional case of classification is shown in Fig. 4 (b).
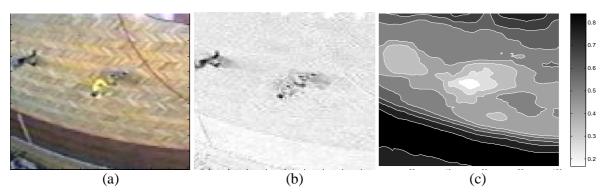

(a)   (b)   (c)

Fig. 5: Locating the player wearing the yellow dress using the set of templates. (a) Player is shown in the center of image. (b) The difference image, as defined in (5) for that particular case. c) Calculated similarity measure s as defined in (7). The white region (showing high similarity) marks the player location more accurately than the difference image, shown in (b).

The both distances (Fig. 4 (b)) and corresponding similarity measure s are defined as follows:

$$d_{GF} = \sqrt{\sum_{j=1}^{42}(F_j - G_j)^2} \quad d_{FH} = \sqrt{\sum_{j=1}^{42}(F_j - H_j)^2} \; , \quad s = \frac{d_2}{d_1 + d_2} \tag{7}$$

The results of player localization using this approach are shown in Fig. 5.

4.3 Color tracking

Color as an identifying feature can be also used for tracking players. The three-dimensional RGB color representation was chosen, as some players wear dark dresses, which would result in undefined values of hue. Color identification and localization, based on color histograms (Swain and Ballard, 1991) is not appropriate, as there are only a few (3-6) pixels that closely resemble the reference color of the player's dress. To ensure reliable tracking, only the location of the pixel that is most similar to the player's color becomes new player position. The similarity measure used is defined as follows:

$$S(x, y) = ((I_R(x, y) - C_R)^2 + (I_G(x, y) - C_G)^2 + (I_B(x, y) - C_B)^2)^{\frac{1}{2}} \; , \tag{8}$$

where I is the image and C is the color of the player. Subscripts R, G and B denote the red, green and blue channel, respectively.

The algorithm searches for the pixel most similar to the recorded color of the player in limited area around the previous player position. Colors of the players are defined by the operator supervising the tracking process, before the actual tracking is started.

5. Results

The following tracking methods were tested: the motion detection algorithm (A), the color tracking algorithm (B) and the combination of color and template tracking algorithm (C). The use of color tracking in C avoids drift in player position, caused by the template tracking. Tests were performed on two video sequences. Their lengths were 30 and 50 seconds, respectively.

First sequence was taken from the recording of handball match and was used to evaluate reliability of the algorithms by counting the human interventions required to maintain error-free tracking process. The second one was the test sequence, in which players were standing still. The measured distances therefore directly correspond to noise added by particular tracking method. Additionally, the processing time per frame on PC equipped with 500 MHz Pentium III processor was measured. The results are shown in Table 1.

Table 1: Comparison results for three different combinations of tracking methods.

Method	Interventions	Noise	Time
A	45	80 meters	0.424 sec/frame
B	12	249 meters	0.175 sec/frame
C	14	55 meters	0.229 sec/frame

The advantages and disadvantages for each of the methods can be clearly seen. The motion detection (A) introduced little noise to the trajectories. On the other hand, more than one intervention per one second of playing time is required, which puts enormous pressure on

human operator supervising the tracking process. This approach is the slowest. Figure 6 shows the trajectories obtained using this method. The method B, the color tracking, required small number of interventions, as expected, it is fast, but the trajectories contain extremely high amount of noise.

The combination C is nearly as fast as color tracking alone, contains low amount of noise and requires little intervention from operator. We can conclude that the combination C is the most suitable for use in an automated player tracker.
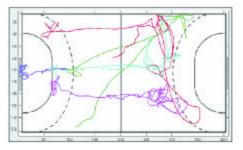

Fig. 6: Trajectories of several players for the first few minutes of the match, obtained by motion detection algorithm (A) and post-filtered using the Gaussian filter.

6. Conclusions

We presented the development of the automated tracking system for tracking players in the sports games. Although the system was only tested in the handball match, we can assume that it would be equally useful in other team sports, provided that the cameras could be placed directly above the playing court.

The system is capable of outputting the positions of every player in court coordinates in every instance of time. It was used to track players in the handball match to perform analysis of their movement. It forms a base of the complete sports analysis system, which will include advanced data processing, statistical analysis and presentation capabilities.

Acknowledgements

References

1. Aggarval, J. K. and Cai, Q. (1997), "Human Motion Analysis: A review", *IEEE Nonrigid and Articulated Motion Workshop*, pp. 90-102.
2. Erdmann, W. S. (1992), "Gathering of kinematic data of Sport Event by Televising the Whole Pitch and Track", *Proc. of 10th ISBS symposium*, pp. 159-162.
3. Intille, S. S. and Bobick, A. F. (1995),"Visual Tracking Using Closed- Worlds", *ICCV '95.*
4. Swain, M. J. and Ballard, D. H. (1991),"Color Indexing", *International Journal of Computer Vision*, Vol. 7, No. 1, November 1991.
5. Tsai, R. Y. (1987),"A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 4.