A Computer Vision Based Tracking System for Indoor Team Sports

Emad Monier, Per Wilhelm, Ulrich Rückert
System and Circuit Technology, Heinz Nixdorf Institute, University of Paderborn
Fürstenallee 11, 33102 Paderborn, Germany
{monier, wilhelm, rueckert@hni.upb.de}@hni.upb.de

ABSTRACT
This paper presents a video tracking system for tracking players in indoor sports using two high quality digital cameras. The tracking algorithm is based on template matching technique taking into consideration closed world assumptions. The output of the system can be visualized interactively for convenient analysis of player movements. The implementation has been efficiently done as a software system that can be used by coaches and sport scientists.

Keywords
Computer Vision, Tracking, Template Matching, Closed World

1. INTRODUCTION
There are a increasing demand in the sport community to have tracking information of the players which will help coaches and sport scientists to evaluate the performance of the team as well as the individual players during training and/or official games. The tracking of the player positions during the training session means finding the position of each player on the play ground in a sufficient accuracy and frequency so that the path information such as distance, speed profile and acceleration can be computed. This information is the basic input for further higher level analysis such as strategy, performance and fitness of the player and helps the coaches to design better training patterns.

The research in the aspect of human motion capturing during sport events (training or games) has begun relatively as early as mid 1920s. At this time researchers like Hill [1] and later Keller [2] developed dynamic model of athletes that were used to predict the world records for linear races, such as 100-meter sprint. As technology progressed, high-accuracy measurement devices were developed that allowed researchers to study the bio-mechanical properties of athlete's body for example Richards [3] in 1999 presented a comparison between some commercially available systems. The devices in these systems were able to measure only the local features of the human body such as positions of extremities so the analysis of athlete's performance in larger scale was not possible using these systems. The sport experts are always interested in obtaining the player positions during a sport match. Such information would allow them to analyze the tactical behavior or performance of certain players or the whole team. The early research in this field required recording a sport event (match or training) using a video camera and then many hours of tedious manual input to obtain some (not all) players' positions. Although a lot of effort is needed to obtain position information this allowed researchers like Erdmann [4], and Ali and Farrally [5] to obtain important information about the load on soccer players during a match. More in-depth analysis of player performance has been done using more advanced semi-automatic tracking systems. For examples in [6] a system specifically designed to track players in squash matches enabled Vučkovič et al. [7] to determine the behavioral features that distinguished the loser from the winner based on the position information. The system in [6] has been further developed to track handball players and named SAGIT [8].

The community which is interested in obtaining player trajectories during the game is not limited to the sport scientists. Coaches can also use this information to individualize the physical training plans of their players and to devise suitable offensive/defensive strategies to achieve the best performance. From above we can conclude that the players' tracking is a cornerstone of further analyses that are interesting to sport experts.

The proposed tracking system is based on two digital cameras mounted on the sport hall ceiling in order to cover the whole field. The cameras and the objectives are carefully selected to provide high quality images. The tracking method used here is template matching with some enhancements based on some assumptions taken from the rules of the game and the tracking environment. After the tracking process our system provides the trainer with visualization of analysis results of the position data.

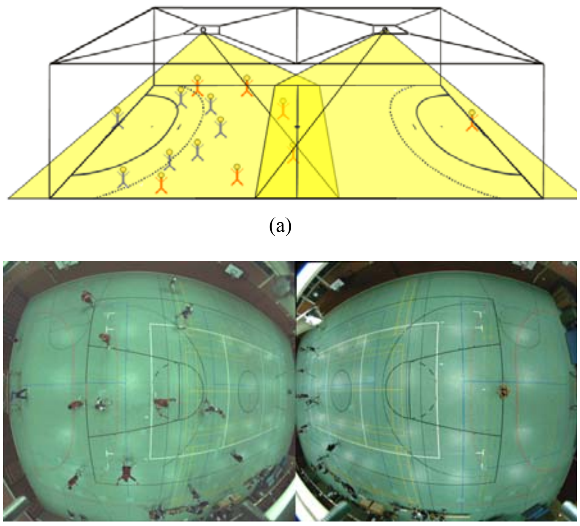
The rest of this paper is organized as follows: section 2 presents the idea of closed world assumptions. Section 3 is about image acquisition and experimental setup. Section 4 is talking about tracking including background elimination, the template matching, single and multiple players tracking, respectively. Section 5 presets the correction of camera distortion. In section 6 the procedure of calculating player positions is explained. Finally section 7 and 8 presents the discussion, conclusion and future work, respectively.

2. CLOSED WORLD ASSUMPTIONS
The closed world concepts introduced by Intille and Bobick [9] is based on the following assumption: given a region in space and time, a specific context is adequate to explain that region, i.e. determine all objects within the region. This region is called the closed world and the context is a boundary in space of knowledge, outside of which knowledge is not helpful in solving the tracking problem. If we consider the indoor sport hall as a closed world, we can define it by a set of assumption: (i) the two cameras are overlooking the playground and are fixed (ii) The playground is bounded, and its model can be calculated (iii) At a given time-step, two players can not occupy the same position.

We will show in section 4 how these assumptions are used in the tracking process.

3. IMAGE ACQUISITION

In order to keep the tracked players in the field of view all the time during the game two stationary digital cameras have been installed in the sport hall, one over each half of the field. A fisheye objective is used in order to catch the required view on the image sensor. The selection of the two cameras and the objective has been done after careful analysis and testing many kinds of camera with different sensors.



(a)

(b)

Figure 1. (a) Camera positioning (b) the images obtained image from the two cameras installed in University of Paderborn Sport hall.

The two cameras have 2/3” CCD sensors with two fisheye objectives each with 2.7mm focal length. Each camera has a Gigabit Ethernet interface and is capable of delivering 30 fps with a maximum resolution of 1932 x 1040. Figure 1 shows the positioning of the two cameras and the obtained images.

4. TRACKING

In this section we will describe the tracking method used in our system. The basic algorithm as well as its required preprocessing is presented in sections 4.1 and 4.2, respectively. The tracking of single player and managing multiple players is discussed in sections 4.3 and 4.4.

4.1 Background Detection

Background subtraction is a basic operation in computer vision in which each pixel in the current scene is classified as background or foreground so that the focus of processing will be on the foreground objects. In order to do this segmentation a model of the background should be computed or obtained. This can be done by saving a frame of an empty field. However different lightings conditions may require updating the background image which could be difficult during the game. A solution to this problem is to compute a background model during or before starting the analysis process and allow updating the model during the analysis if needed.

Our proposed tracking system allows the use of previously saved empty background image. It could also build a background model from several successive frames before the start of the tracking process. During the tracking process, the background model can be updated via convenient user interface. Figure 2 shows the algorithm for background detection and subtraction.

For each input frame

a. smooth the image by Gaussian filter
b. if the background update mode is on
    i. compute difference between current and previous image
    ii. perform threshold and dilation on the difference
    iii. copy the non-motion areas to the background model
c. compute difference between current frame and background model
d. threshold the difference

Figure 2. The background detection and subtraction algorithm

4.2 Template Matching

As an engine for tracking in our system we have used a template matching [10] technique. A template is a sub-image which contains the shape we are trying to find. In the template matching process the template is centered on an image point and the matched pixels are counted up. This process is repeated for the entire image and the point which gives the best match is supposed to be the position of the template in the image.

As a measure of similarity for the template matching operation we have used the normalized correlation coefficient [11] which is computed using the following equation:

$$R(x,y)=\frac{\sum_{x',y'} T'(x',y').I'(x+x',y+y')}{\sqrt{\sum_{x',y'} T'(x',y')^2.\sum_{x',y'} I'(x+x',y+y')^2}} \qquad (1)$$

where

$$I'(x+x',y+y')=I(x+x',y+y') -\frac{1}{wh}\sum_{x'',y''} I(x+x'',y+y'') \qquad (2)$$

and

$$T'(x',y')=T(x',y')-\frac{1}{wh}\sum_{x'',y''} T(x,y) \qquad (3)$$

where $R(x,y)$ is the result image, $T(x,y)$ is the template image and $I(x,y)$ is the searched image.

4.3 Single Player Tracking

Because of the high quality images provided by the cameras and after performing some experiments we have decided to track the player’s head plus a small part of the shoulder, because it’s relatively clear and easier to track than the whole body of the player. At the beginning of the tracking process the background model is computed using some frames which contains movements or an empty background as described in section 4.1. After that the user has to select a player to track by selecting the player head, which means initialization of the template for this player. The size of the template is based on the camera resolution; in our implementation we have used a template size of 20x20. Figure 3 shows the selection of the initial template for one player in a basketball game. In the next frame the template is searched as described in section 5 in a region of interest of 100x100 centered on the initial position of the player.

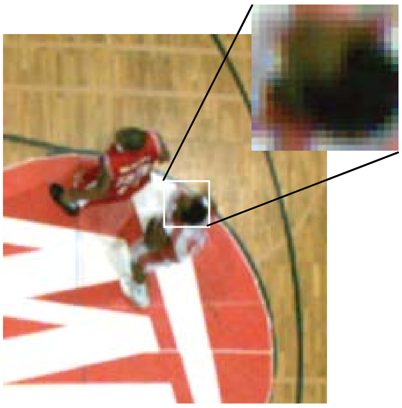


Figure 3. The selection of initial template for tracking a player in a basketball game.

The region of interest is based on the image resolution and the maximum speed of the player during a sprint. After locating the best matching position based on equation (1) the player template is updated by replacing the old one with the sub-image of the best match.

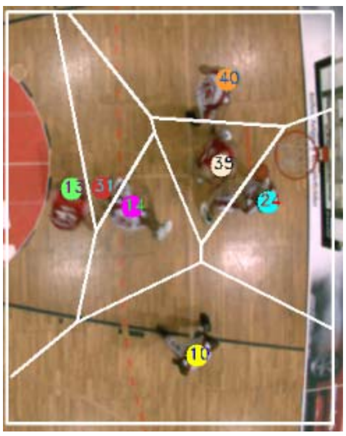


Figure 4. The Voronoi Partitioning used in multiple tracking

4.4 Multiple Players Tracking

In order to successfully track multiple players single trackers are used in a cooperative way. This means we use the information from other trackers to reduce the possibility of error. This is done by using the closed world assumption number (iii). To make things more clear let us assume that we track N players and at a given point of time the position of each player is known. It is possible to partition the image into N disjoint regions such that each region will contain only one player. Such kind of portioning is called Voronoi partitioning [11]. In our system we have used the implementation of Map Manager Library [13]. Figure 4 shows an example of Voronoi portioning in our system during multiple player tracking in a basketball game. This portioning is used to create a mask for each of the single trackers to be used in avoiding confusion when two or more players interact very closely. In order to reduce computational cost of creating these masks the distance matrix of all player positions is calculated and based on a certain threshold the Voronoi mask is computed. This means that players who are not close enough to confuse others are excluded from Voronoi mask calculations.

5. CORRECTING FISHEYE LENS DISTORTION

The video used for the visualization should be corrected because of the distortion caused by the fisheye lenses. The transformation from distorted to undistorted image is based on the following equations:

$$r_{Defish} = f\tan\left(\frac{r_{Fisheye}}{f}\right) \qquad [px] \qquad (4)$$

$$r_{Fisheye} = f\arctan\left(\frac{r_{Defish}}{f}\right) \qquad [px] \qquad (5)$$

Where r_{Defish} is the distance to the center of the corrected (*Defish*) image, $r_{Fisheye}$ is the distance to the center of the distorted (*Fisheye*) image and f is the focal length (in pixels). This transformation is based on the assumption that the lens is spherical and the distortion is only radial and not tangential.

Because the corrected image is bigger than the distorted one interpolation is required. To avoid interpolation the inverse mapping can be used to repeat some pixels form the original distorted image in the corrected one. In our implementation we have used lookup tables where the mapping for each camera is computed only once. In order to transform a pixel p' at location (x', y') from distorted image to a pixel p at location (x, y) in undistorted image we use the following equations:

$$x_{Fisheye} = (x - x_{Center})\frac{r_{Fisheye}}{r_{Defish}} + x'_{Center} \qquad [px] \qquad (6)$$

$$y_{Fisheye} = (y - y_{Center})\frac{r_{Fisheye}}{r_{Defish}} + y'_{Center} \qquad [px] \qquad (7)$$

$$r_{Defish} = \sqrt{(x - x_{Center})^2 + (y - y_{Center})^2} \qquad [px] \qquad (8)$$

Where the center of the distorted image is at pixel $p'_{Center} = (x'_{Center}, y'_{Center})$ and the center of the undistorted image is $p_{Center} = (x_{Center}, y_{Center})$.

Although the camera is fixed but it still requires calibration to correct small errors in its orientation. We do calibration when required based on points from the playing field which have known coordinates. Lookup tables are also computed and used to correct the camera orientation errors.

6. CALCULATING PLAYER POSITION

The output of the tracking algorithm is positions in image coordinate of the distorted (*Fisheye*) image which need to be converted to real world coordinate on the ground of the field in meters in order to calculate further information such as covered distance, speed and acceleration. To do this we have to exchange the focal length f with the height of the camera h_{Camera} in equation (4) so we will get this equation

$$r_{Defish} = h_{Camera} \tan\left(\frac{r_{Fisheye}}{f}\right) \qquad [m] \qquad (9)$$

$$r_{Fisheye} = \sqrt{(x' - x'_{Center})^2 + (y' - y'_{Center})^2} \qquad [m] \qquad (10)$$

$$x_{Head} = (x' - x'_{Center})\frac{r_{Defish}}{r_{Fisheye}} \qquad [m] \qquad (11)$$

$$y_{Head} = (y' - y'_{Center})\frac{r_{Defish}}{r_{Fisheye}} \qquad [m] \qquad (12)$$

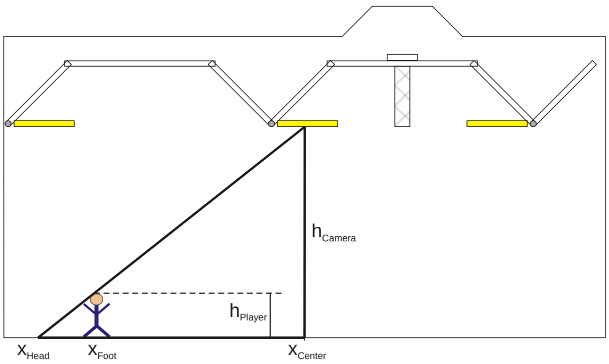


Figure. 6 Calculating real world coordinates of players

Figure 6 shows the position of the camera and the player. Looking at the triangle in the figure we can write this equation from elementary geometry:

$$\frac{h_{Camera}}{h_{Player}} = \frac{(x_{Head} - x_{Center})}{(x_{Head} - x_{Center}) - (x_{Foot} - x_{Center})} \qquad (13)$$

to calculate the feet position we solve equation (13) for x_{Foot} then we get:

$$x_{Foot} = x_{Head}\left(1 - \frac{h_{Player}}{h_{Camera}}\right) + x_{Center}\frac{h_{Player}}{h_{Camera}} \qquad [m] \qquad (14)$$

Substituting for x_{Head} from equation (11) we get:

$$x_{Foot} = (x' - x'_{Center})\frac{r_{Defish}}{r_{Fisheye}}\left(1 - \frac{h_{Player}}{h_{Camera}}\right) \qquad [m] \qquad (15)$$

the same holds for y_{Foot}.

Instead of calculating the feet position in two steps through the position of the projection of the head x_{Head} and then to the feet position x_{Foot}, we can calculate it in one step by subtracting the height of each player from the subject distance h_{Camera}. This means that equation (9) will be written as follows:

$$r^*_{Defish} = (h_{Camera} - h_{Player})\tan\left(\frac{r_{Fisheye}}{f}\right) \qquad [m] \qquad (16)$$

If we substitute r^*_{Defish} for r_{Defish} in (11) we get in (17) x_{Foot} (in meter) which is exactly the same result as in (15) but with less calculations.

$$x_{Foot} = (x' - x'_{Centre})\frac{r^*_{Defish}}{r_{Fisheye}} \qquad [m] \qquad (17)$$

y_{Foot} can be calculated in the same way.

7. DISCUSSION

We have used our video system for tracking players in two sport halls. One of them is the Maspernhalle in Paderborn where basketball games in the first German basketball league are played, the other one is the University of Paderborn sport hall where local games in handball and basketball are played.

One frame is processed in about 100 ms. Additional time will be spent in correcting the errors in tracking which occur because of small sliding of the template away from the head of the player or because of complex interaction between the players which can not be avoided even when using the scheme in section described in 4.4. Another source of errors is the inaccurate background which may occur because of the change in lighting conditions during the game which requires updating the background.

Figure. 5 Interactive Visualization of tracking data

The users of our system are sport trainers, sport medicine and sport science researchers. In order to make it easy to understand the tracking information we have developed an interactive video visualization tool [14], to present the tracking information. This tool allows the user to interactively choose the information that he likes to see on the video.

Figure 6 and 7 show the speed profile and the field coverage of one basketball player during one quarter of an official basketball game. Speed profile shows the percentage of standing, walking, jogging, running and sprinting of the player in the whole playing time, which is very interesting for the coach in evaluating the performance of the player with respect to the fitness. Field coverage shows the whole path (or part of it) of the player during the quarter which can very useful in evaluating the strategic performance of the player.

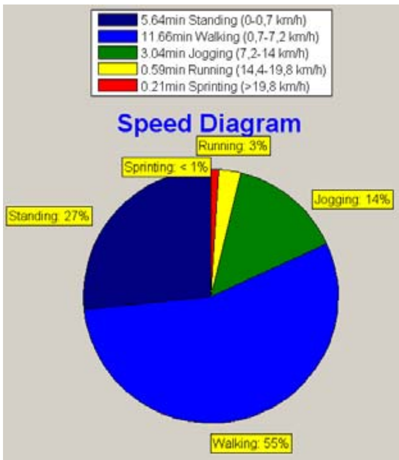


Figure 6: Speed profile of a basketball player in one quarter

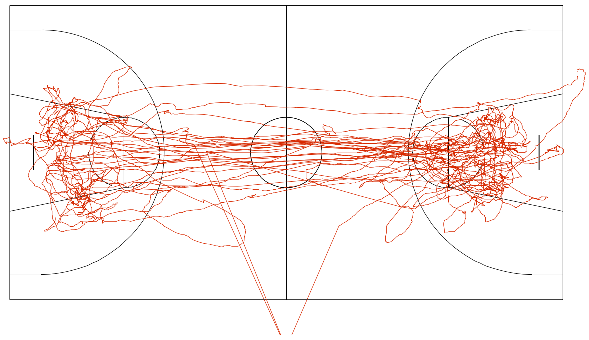


Figure 7: Field coverage of a basketball player in one quarter

8. CONCLUSION AND FUTURE WORK

A computer vision system for tracking players in indoor sport games has been presented. The output of the system is player positions. From player positions important information such as covered distance, speed and acceleration can be calculated and visualized. Intended users of the system are coaches and researchers in sport and coaching science. The engine of tracking in our system is template matching based head tracking. Using closed world assumptions helps reducing errors in tracking and to manage multiple tracking targets. Calculating feet position in real-world coordinate may require many calculations especially with fisheye lenses used for image acquisition. A simple procedure of calculating player feet position has been described. As a future work more tests will be done to assure the accuracy of the system in tracking player positions. Also other tracking methods, such as particle filters will be implemented and tested.

9. REFERENCES

[1] Hill, A. V., *The physiological basis of athletic records*. The Scientific Monthly Volume 21**,** Issue 4, 1925, pp. 409-428.

[2] Keller, J. B., *Optimal velocity in a race*. American Math. Monthly, 81, 1974, pp. 474-480.

[3] Richards, J. G. *The measurement of human motion: A comparison of commercially available systems*. Human Movement Science 18, 1999, pp. 589--602.

[4] W. S. Erdmann, *Gathering of kinematic data of sport event by televising the whole pitch and track*. In '10th ISBS symposium', 1992, pp. 159-162

[5] Ali, A. & Farrally, M., *A computer-video aided time motion analysis technique for match analysis*. Sports Medicine and Physical Fitness, 13, 1991, pp. 82-88.

[6] J. Pers, G. Vuckovic, S. Kovacic, and B. Dezman, *A low-cost real-time tracker of live sport events*. In International Symposium of Image and Signal Processing and Analysis, 2001, pp. 362-365

[7] G. Vučkovič, B. Dezman, F. Erculj, S. Kovacic, and Pers J., *Differences between the winning and the losing players in a squash game in terms of distance covered*. In 'the Eighth International Table Tennis Federation Sports Science Congress and The Third World Congress of Science and Racket Sports', 2004, pp. 202-207

[8] Bon M, Šibila M, Pori P. *Sagit – computer vision system for tracking handball players during the match*. EURO 2004 Coaches' Seminar during the 2004 Men's European Championship in Slovenia.

[9] S. S. Intille and A. Bobick. *Closed-world tracking*. In Int. Conf. of Computer Vision, pages 672–678, 1995.

[10] Von Mark S. Nixon, Alberto S. Aguado, *Feature Extraction and Image Processing*, p.p, 164- 168, Newnes publisher, 2002.

[11] Lewis, J. P. *Fast normalized cross-correlation*. In Proceedings of *Vision Interface 95*, Canadian Image Processing and Pattern Recognition Society, 1995, pp. 120-123.

[12] J. R. Sack and J. Urrutia. *Handbook of Computational Geometry*. Elsevier Science Pub Co, 2000.

[13] Manly, KF, Cudmore, Jr, RH, Meer, *Map Manager QTX, cross-platform software for genetic mapping*. JM (2001) Mammalian Genome 12: 930-932.

[14] Tanoto, A.; Du, J. L.; Kaulmann, T. & Witkowski, U., *MPEG-4-Based Interactive Visualization as an Analysis Tool for Experiments in Robotics*. In 'Int. Conf. on Modeling, Simulation & Visualization Methods', 2006, pp. 186-192