Tracking People in Sport: Making Use of Partially Controlled Environment

Janez Perš and Stanislav Kovačič

Faculty of Electrical Engineering,
University of Ljubljana
Tržaška 25, SI-1000 Ljubljana, Slovenija
{janez.pers},{stanislav.kovacic}@fe.uni-lj.si,
Tel: +386 1 4768 876, Fax: +386 1 4768 279
WWW home page: http://vision.fe.uni-lj.si

Abstract Many different methods for tracking humans were proposed in the past several years, but only a few authors examined the accuracy of the proposed systems. As the accuracy analysis is impossible without the well-defined ground truth, some kind of at least partially controlled environment is needed. Analysis of an athlete motion in sport match is well suited for that purpose, and it coincides with the need of the sport research community for accurate and reliable results of motion acquisition. This paper presents a development of a multiple-camera people tracker, incorporating two complementary tracking algorithms. The developed system is suited for simultaneously tracking several people on a large area of a handball court, using a sequence of 384-by-288 pixel images from fixed cameras. This paper also examines the level of accuracy that this kind of computer vision system setup is capable of.
Keywords: computer vision, people tracking, controlled environment.

1 Introduction

People tracking is a rapidly developing field of computer vision. However, the most interesting locations to deploy computer vision based people trackers usually represent highly uncontrollable environments (railway stations, crowded halls, etc.), which poses significant difficulty in evaluating the performance of developed systems. On the other hand, many sport researchers struggle to obtain reliable data about movement of athletes, especially when sport activity covers a large area, for example in team sports. Sport matches, especially indoor ones, represent *partially controlled environment*, and are as such highly suitable as a test ground for development and testing of new people tracking methods.

The research in the fields of people tracking and analysis of sports-related video has flourished in the past several years [1–8]. However, the emphasis is still on development of tracking methods and improvement of reliability of the tracking itself. Only a few authors (for example [5]) examined the accuracy of their tracking systems or suggested both the method for evaluating the accuracy and obtaining the ground truth (for example [6]). On the other hand, use of

computers in gathering and analyzing the sport data is an established practice in sport science [9, 10]. One of important aspects of football, handball or basketball match analysis is the information about player movement [11], but due to limitations in available technology, the results obtained were often coarse and only approximate.

In this article, we present the method for tracking known number of people in a partially controlled environment - the handball court inside the sports hall. First, problems associated with image acquisition are discussed. Next, two algorithms for tracking athletes during the match are presented, and their combination which yields best results in terms of reliability and accuracy is presented. Next, the required post-processing of trajectories is briefly discussed. The collaboration with sports scientists enabled thorough evaluation of the accuracy of the developed system, which is described in a separate section. Finally, some conclusions about tracker performance are drawn.

2 Image Acquisition

Proper image acquisition significantly influences the performance of the tracking algorithms. In case of video annotation and highlighting [12] high accuracy is not required. In case of player motion acquisition and analysis, where certain measurements are performed and degree of uncertainty has to be specified, careful planning of image acquisition proves to be crucial for the success of the whole system [4]. Camera movement can add significant degree of difficulty to the tracking problem, as the objects and the sensor are independently moving with respect to the reference coordinate frame. Therefore, continuous calibration is required. To alleviate these problems, two stationary cameras with wide-angle lenses were chosen in our case. Their placement on the ceiling of the sports hall and the resulting combined image is shown in Figure 1.

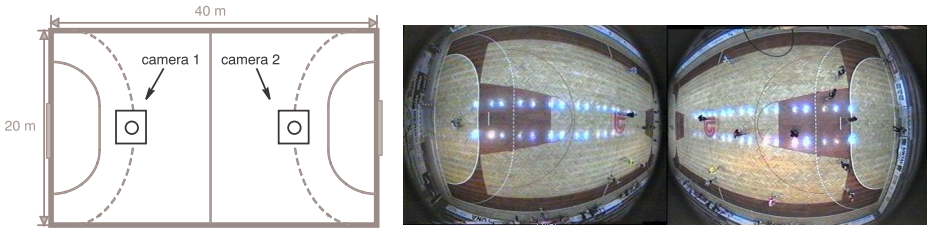


Figure 1. Handball playing court and camera placement (left). Example of combined image from two cameras, taken at the same instant of time (right).

The whole handball match that lasted for about an hour was recorded using two PAL cameras and two S-VHS videorecorders. A transfer to digital domain was carried out using the Motion-JPEG video acquisition hardware, at 25 frames per second and image resolution of 384x288 pixel. The combined image from both cameras is shown in Fig. 1.

3 Camera Calibration

To perform position *measurements* based on the acquired images, the relations between pixel coordinates in each of the images and world (court) coordinates have to be known. These relations are obtained by the camera calibration. The procedure is simplified due to rigid sport regulations, which precisely specify the locations and dimensions of various landmarks on the playing court. Unfortunately, due to the large radial distortion otherwise widely used calibration technique [13] fails to produce satisfactory results. We decided to build the model of radial image distortion, and couple it with simple linear camera model.

Fig. 2 illustrates the problem of radial distortion. For illustrative purposes only, let us imagine an ideal pinhole camera, mounted on a pan-tilt device. Point 0 is the point of intersection of optical axis of the camera with the court plane, when the pan-tilt device is in its vertical position. Point C denotes the location of the camera, and 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 α is the angle of the 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 α are given within the triangle on the enlarged part of Fig. 2 (left).

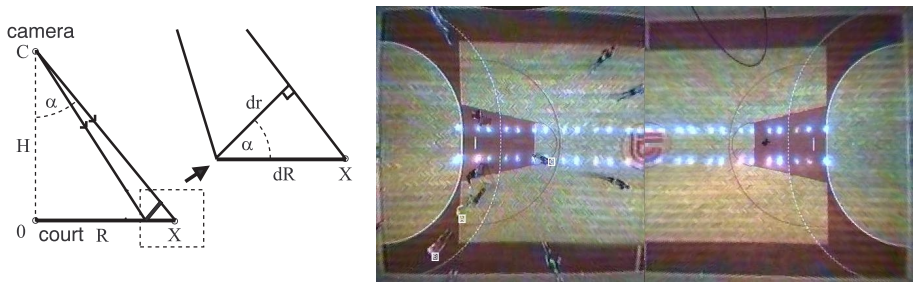


Figure 2. A model of radial distortion (left). A combined image from both cameras after the radial distortion correction (right).

Thus, we can write the following relations:

$$dr = cos(\alpha) \cdot dR, \quad \alpha = arctg(\frac{R}{H}), \tag{1}$$

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

Let us substitute the pan-tilt camera with a fixed camera, equipped with wide-angle lens. The whole area, which is covered by changing the angle α of the pan-tilt camera, is captured simultaneously to the single image of the stationary camera. Additionally, let us assume that the scaling factor between the dr and the image of dr on the image plane equals 1. Therefore, we can obtain the length

of the image of radius R on the image plane by integrating the left side of Eq. (2) over the interval $(0, r_1)$, and the right side over the interval $(0, R_1)$,

$$\int_0^{r1} dr = \int_0^{R1} cos(arctg(\frac{R}{H}))dR, \tag{3}$$

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. The solution of the inverse problem is then:

$$r_1 = H \cdot ln\left(\frac{R_1}{H} + \sqrt{1 + \frac{R_1^2}{H^2}}\right). \tag{4}$$

By solving Eq. (4) for R_1 we obtain the formula, which can be used for correcting the radial distortion:

$$R_1 = \frac{H}{2}\frac{(e^{-\frac{2r_1}{H}}) - 1}{e^{-\frac{r_1}{H}}}. \tag{5}$$

Parameters were obtained with the help of various marks, which are part of the standard court marking for handball matches (boundary lines, 6 and 9 m lines, etc.) and non-linear optimization. For illustrative purposes, a result of radial distortion correction is shown in Fig. 2 (right). Nevertheless, we decided to perform tracking on uncorrected images, and to correct the obtained player positions thereafter.

4 Player Tracking

Many general-purpose tracking algorithms could be used for the player tracking. However, in our setup, we are faced with the following difficulties:

- Players are small objects, typically only 10-15 pixels in diameter, which makes histogram-based identification techniques difficult.
- Players cast shadows, which overlap frequently, causing trouble for simple background subtraction techniques.
- Due to strict handball rules, any placement of markers is forbidden during the European Handball Federation (EHF) matches. However, players of different teams wear differently colored dresses.

4.1 Color-Based Tracking

Color as an identifying feature [14] can be also used for the task of tracking players. Color is generally largely independent of the view and resolution, and remains constant over long intervals of time. Therefore, colors of the players dresses could be input to the tracking system manually at the beginning of the tracking process without any adaptations later.

Color identification and localization, based on color histograms, was reported by Swain and Ballard [14]. However, given a small number of pixels that comprise each of the players, this technique is not appropriate. In most cases, there are only a few (3-6) pixels that closely resemble the reference color of the player's dress. The situation is illustrated by Fig. 4b. Therefore, different approach was needed.

The algorithm searches for the pixel most similar to the recorded color of the player. The search is performed in a limited area (9-by-9 pixels) around the previous player position. The three-dimensional RGB color representation was chosen instead of HSI, as some players wear dark dresses, which would result in undefined values of hue. The similarity measure is defined as euclidean distance:

$$S_{color}(x,y) = \sqrt{((I_R(x,y) - C_R)^2 + (I_G(x,y) - C_G)^2 + (I_B(x,y) - C_B)^2}, \quad (6)$$

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

The advantage of described algorithm is high reliability. The algorithm tracks players successfully even when the apparent player color is changed due to signal distortion during tape recording or lossy compression. The main problem is caused by diverse background with colored areas, which closely correspond to the color of the player's dress. The disadvantage of this method is also a high amount of jitter in the resulting player trajectories, which makes it inappropriate for a stand-alone use.

4.2 Template Tracking

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

Feature set, which can be used to successfully separate players from the background, needs to be found. Due to low resolution and rapidly changing appearance of the players it is extremely difficult to build an accurate model of a handball player. Instead, we have used a subset of modified Walsh functions and their complements, i.e. "templates", shown in Fig. 3, which extract the very basic appearances of the players.

First, the region of interest (ROI) which surrounds the position of the player is defined. Considering the size of the players in captured image, the region size was set to 16x16 pixels, with player position in the center of the region. Each channel of the RGB color image is processed separately and the vector $\mathbf{F}$, consisting of 14 features for each channel, is obtained using the following formula:

$$F_{i+14j} = \sum_{x=1}^{16} \sum_{y=1}^{16} K_i(x,y) \cdot I_j(x,y), \quad (7)$$

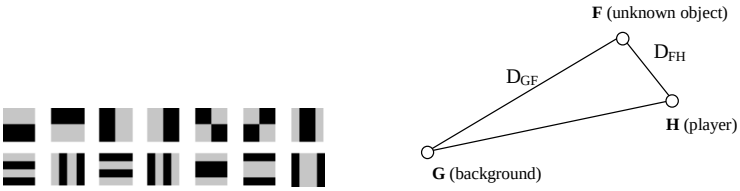


Figure 3. Basic templates of the player - modified Walsh functions. Black areas represent zeros, while gray areas denote value of 1 (left). Classification of an unknown object (right).

where K_i is one of the 14 template functions $(i = 0 \dots 13)$, and I_j is one of the three RGB channels $(j = 0, 1, 2)$, obtained with respect to ROI from the current image. Each channel yields 14 features, which results in 42-dimensional feature vector **F**.

Let vector **H** represent the estimated appearance of the player, and vector **G** represent the appearance of the background (empty playing court) at the same coordinates. Our goal is to classify the unknown object in the region of interest I, represented as vector **F**, either as a "player" or a "background". The simplified, two-dimensional case is shown in Fig. 3, right.

Vector of features **G** is calculated from the image of the empty playing court at the same coordinates as **F**. The reference vector **H** is obtained by averaging the last n vectors of features for a successfully located player, which allows certain adaptivity, as the player appearance changes over time. The value of n depends on the velocity at which the players move, but best results were obtained with the value of $n = 50$ which corresponds to two seconds of video sequence.

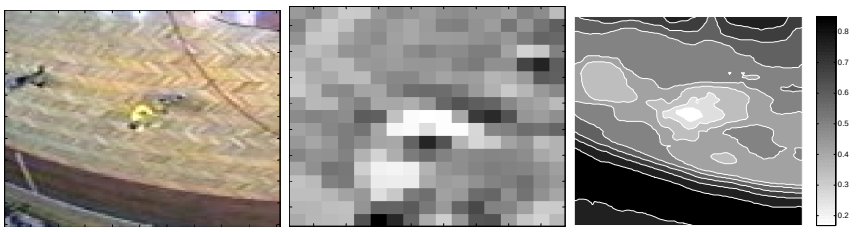
Similarity measure S is obtained using the following formula:

$$S_{template} = \frac{D_{FH}}{D_{GF} + D_{FH}}, \qquad S \in [0, 1], \tag{8}$$

where D_{GF} and D_{FH} are Euclidean distances. The domain of measure S is interval $[0 \dots 1]$. Low value of $S_{template}$ means high similarity between the observed object **F** from the region of interest I and stored appearance of the player in the vector **H**. Fig. 4 (c) shows the test result on a single player and demonstrates the ability of this technique to locate an object.

4.3 Tracking

At the very beginning of the tracking, human operator initializes player positions. Initial estimates for player positions for each subsequent frame are derived from the preceding image from image sequence. On each image from the sequence, the color tracking method, described in section 4.1 is used to roughly estimate player position. 3-by-3 pixel neighborhood of estimated position is examined for the minimum of similarity measure (8). The position of the minimum is used as the next estimate and the process is iteratively repeated up to 10 times. The maximum number of iterations limits the area that is being searched and is



(a) (b) (c)

Figure 4. Locating the player wearing a yellow dress. (a) The player is shown in the center of the image. (b) Distance S_{color} (Eq. 6) to the yellow color for a close-up view of a player from the image (a). White pixels mark the areas that match the yellow color, while darker ones mark areas of different colors. (c) Similarity measure $S_{template}$ of the whole image (a), as defined in (Eq. 8). Feature vector **H** (player reference) was obtained from the subimage of the same player, captured at different instants of time. The white area corresponds to the region of high similarity ($S_{template}$ is 0.2 or lower).

defined by the maximum expected player movement. However, initial estimate, provided by the color tracking algorithm is saved and used as the starting position for the next frame. This ensures both high reliability provided by the color tracking mechanism, and low amount of trajectory jitter, due to use of template tracking to correct initial estimates.

4.4 Trajectory Post-processing

The trajectories obtained using previously described method contain certain amount of noise, which makes player velocity calculation extremely difficult.

The spectrum of noise overlaps with the spectrum of rapid player movements and some data loss is expected when filtering player trajectories. An obvious way of trajectory filtering is by using a Gaussian filter, as shown in (9). We process x and y components of the trajectory separately, treating them as one-dimensional, time-dependent signals,

$$u(t) = \frac{1}{2N_F + 1} \sum_{i=-N_F}^{N_F} x(t+i) \cdot G(i), \tag{9}$$

where $2N_F + 1$ denotes the width of the filter, u is the filtered component of the trajectory, x is the component of the raw trajectory as provided by the tracking method and G is the array of Gaussian coefficients. The precalculated set of $N_F + 1$ coefficients in the range of Gaussian function $(-3\sigma, 3\sigma)$ was used.

5 Evaluation and results

5.1 Reliability and efficiency

A sequence of 750 images, corresponding to the 30 seconds of the handball match was used to test tracker reliability. Tracking was performed simultaneously for

the all 14 players, present at the playing court. Human operator, supervising the tracking process had to intervene 14 times during the tracking process. Intervention consisted of stopping the tracking process, manual marking of the player which was tracked incorrectly and restarting the tracking process. However, in comparison to full manual tracking, which would require 10500 (750*14) manual annotations for the same sequence, the tracker proved to be extremely useful. Processing speed averaged 4.5 frames per second.

5.2 Accuracy

There are several sources of errors that can influence the overall accuracy of the tracking: movement of player extremities, VCR tape noise, image compression artifacts, imperfect camera calibration and quantization error. However, thorough analysis of the error propagation would be difficult in our case. Therefore, a set of experiments was designed to determine tracker accuracy. All experiments included several handball players, differently dressed, which performed various activities, according to the purpose of each experiment.

Ground truth was obtained simply by drawing a pattern of lines near the middle of the one half of the handball court. The pattern, shown in Fig. 5 was measured using a measuring tape.

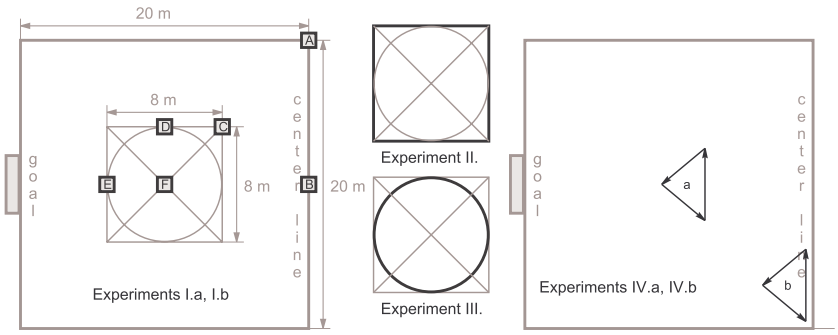


Figure 5. Setup for experiments I.-IV. on the one half of the handball court. Left: player positions during the experiment I. are marked with black boxes. Middle: Reference player trajectories for experiments II. and III, shown with thick lines. Right: Approximate player trajectories for the experiment IV.

The following experiments were performed:

Experiment I. Players were instructed to stay stil at the predefined places - 3 players near the court centre, 2 players near the court boundary. In the second part of experiment, they were instructed to perform various activities (passing ball, jumping on the spot, etc.) but they were not allowed to move across the court plane. Reference position was obtained from the drawn pattern. Reference velocity and path length were exactly zero, since the

players never left their positions. RMS (Root Mean Square) error in player position and player velocity was calculated, as well the error in path length. Effect of trajectory smoothing was also evaluated.

Experiment II. Players were instructed to run and follow the square trajectory. Influence of trajectory filtering was observed. RMS error in player position (distance from measured player position to the square reference trajectory) with respect to filter width was calculated. The results confirmed that heavy smoothing hides rapid changes in player trajectory and is therefore inappropriate from this viewpoint.

Experiment III. Players were instructed to run and follow the circular trajectory with constant velocity. RMS error in player velocity was observed, and the reference velocity was simply calculated from the length of the circular path and the time each player needed for one round. As players were unable to move with *exactly* the same velocity all the time, part of the velocity variation can be contributed to them. In this way we made sure that our tracker was performing even better than the actual measurements have shown.

Experiment IV. We compared our system to widely used manual, video-based kinematic analysis tool - APAS (Ariel Performance Analysis System, [10]), which was used as a ground truth this time. Results were consistent with previous experiments, except in the level of detail that APAS captured. Velocity graph obtained using APAS has clearly shown accelerations and deccelerations of the player, associated with each of his steps. This is the feature that our system was designed to avoid, since it is not useful in match analysis.

Accuracy of the designed system can be summarized as shown in Table 1.

Accuracy using:	11 samples wide filter	25 samples wide filter
Position, still player:	0.2 (0.5) m RMS	0.2 (0.5) m RMS
Position, active player:	0.3 (0.6) m RMS	0.3 (0.6) m RMS
Velocity, uniform motion:	0.4 m/s RMS	0.2 m/s RMS
Velocity, uniform motion (%)	12%	7%
Path length, still player:	+0.9 m/min	+0.6 m/min
Path length, active player:	+10 m/min	+6 m/min

Table 1. Tracker accuracy. Numbers in parentheses indicate accuracy for player position near the court boundary.

6 Conclusion

Use of the controlled environment for our people tracker has enabled us to develop a tracking system which suits its purpose and, most importantly, it has enabled us to test its accuracy. We are confident that the obtained accuracy does not hit the limits of our tracking system, but rather the limits of possible definition of player position, velocity and path length itself. Our system observes the movement of the people across a plane at some level of detail. From computer

vision point of view, the handball players are large, non rigid objects and in many cases reporting player position with uncertainty as low as shown in Table 1 does not make any sense, simply due to the lack of *exact* definition of player position and velocity.

References

1. J. K. Aggarval and Q. Cai. Human motion analysis: A review. In *IEEE Nonrigid and Articulated Motion Workshop*, pages 90–102, Puerto Rico, June 17-19 1997.
2. Q. Cai and J. K. Aggarval. Tracking human motion using multiple cameras. In *Proceedings of the 13th International Conference on Pattern Recognition ICPR'96*, volume 3, pages 68–72, Vienna, 1996.
3. I. Haritaoglu, D. Harwood, and L. S. Davis. An appearance-based body model for multiple people tracking. In *Proceedings of the 15th International Conference on Pattern Recognition - ICPR 2000*, volume 4, pages 184–187, Barcelona, Spain, September 3-8 2000.
4. S. S. Intille and A. F. Bobick. Visual tracking using closed-worlds. In *Proceedings of the Fifth International Conference on Computer Vision ICCV '95*, pages 672–678, MIT, Cambridge, MA, June 20-23 1995.
5. G. Pingali, A. Opalach, and Jean. Y. Ball tracking and virtual replays for innovative tennis broadcasts. In *Proceedings of the 15th International Conference on Pattern Recognition ICPR 2000*, volume 4, pages 152–156, Barcelona, Spain, September 3-8 2000.
6. A. Pujol, F. Lumbreras, X. Varona, and J. Villanueva. Locating people in indoor scenes for real applications. In *Proceedings of the 15th International Conference on Pattern Recognition - ICPR 2000*, volume 4, pages 632–635, Barcelona, Spain, September 3-8 2000.
7. B. A. Boghossian and S. A. Velastin. Motion-based machine vision techniques for the management of large crowds. In *IEEE 6th International Conference on Electronics, Circuits and Systems ICECS 99*, Cyprus, September 5-8 1999.
8. S. Murakami and A. Wada. An automatic extraction and display method of walking person's trajectories. In *Proceedings of the 15th International Conference on Pattern Recognition - ICPR 2000*, volume 4, pages 611–614, Barcelona, Spain, September 3-8 2000.
9. A. Ali and M. Farrally. A computer-video aided time motion analysis technique for match analysis. *The Journal of Sports Medicine and Physical Fitness*, 31(1):82–88, March 1991.
10. Ariel dynamics worldwide. Internet URL. http://www.arielnet.com.
11. W. S. Erdmann. Gathering of kinematic data of sport event by televising the whole pitch and track. In *Proceedings of 10th ISBS symposium*, pages 159–162, Rome, 1992.
12. N. Nitta, N. Babaguchi, and T. Kitahashi. Extracting actors, actions and events from sports video - a fundamental approach to story tracking. In *Proceedings of the 15th International Conference on Pattern Recognition - ICPR 2000*, volume 4, pages 718–721, Barcelona, Spain, September 3-8 2000.
13. Y. R. Tsai. 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*, RA-3(4):323–344, 1987.
14. M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, November 1991.