Visual Tracking of Athletes in Beach Volleyball Using a Single Camera

Thomas Mauthner[1], Christina Koch[2], Markus Tilp[2], Horst Bischof[1]

[1]Institute for Computer Graphics and Vision, Graz University of Technology, Austria

[2]Institute for Sport Science, Karl-Franzens-University, Graz, Austria

Abstract

This paper aims at successful tracking of beach volleyball athletes during competition using only a single camera. Due to the wide range of possible motions and non-rigid shape changes, the tracking task becomes quite complex. We propose a novel method based on integral histograms, to use a high dimensional model for a particle filter without drastic increase in runtime. We extend integral histograms to handle rotated objects. Additionally to the tracking process, a segmentation of the lower body parts enables generating real world player positions from a single camera view. Comparisons to hand annotated position data revealed sufficient accuracy for classical sport scientific purposes. The paper focuses on beach volleyball but the proposed methods can be utilized in other sports and non sports applications.

KEY WORDS: VISUAL TRACKING, BEACH VOLLEYBALL, TIME-MOTION ANALYSIS

Introduction

When analyzing sports games the main aspects of interest are the used techniques, the played tactics and the physiological demands of athletes. All three characteristics are important to quantify skills and shortcomings of athletes or teams and define requirements for training and competition.

For the analysis of technique and tactics, video technology has become very common and is utilized in several ways to analyze these aspects. The simplest way is to use video recordings to provide feedback for athletes (Liebermann & Franks, 2004) or to study opponent teams during crucial game situations in replay. Beyond these attempts interactive video systems (Dartfish®, Fribourg, Switzerland or Statshot®, Graz, Austria) are used to gather further information e.g. by counting frequencies of special techniques and by evaluating the effectiveness of actions. Such an attempt was successfully used by Tilp, Koch, Stifter & Ruppert (2006) to generate video based statistics for the analysis and comparison of world class junior beach volleyball teams.

The positions of actions in sports are often crucial for success or defeat and therefore an essential information to rate the quality of an action. In order to determine the playing position it became quite common to define relevant zones of the court and to estimate in which zone an action occurred (Hughes & Franks, 2004). Getting accurate information about positions is complicated due to several factors. A distortion caused by the perspective view, missing marks and the transitions from one zone to another often causes wrong rating decisions and errors.

Exact position information is furthermore required to calculate covered distances, velocities and accelerations of athletes with which the physiological demands can be estimated. The main advantage compared to classical methods like heart rate monitoring or lactate testing is that interaction with the athletes can be avoided. Different methods for such time-motion analyses have been used since the early 1970's. Before adequate technology was available such analyses were made manually via observation (Reilly & Thomas, 1976) or via audio recording (Yamanaka, Haga, Shindo, Narita, Koseki, Matsuura & Eda, 1988).

Due to accuracy reasons, methods based on video data followed by a manual computer supported analysis have become the preferred method in the last years (for review see Spencer, Bishop, Dawson & Goodman, 2005 or Bangsbo, Mohr & Krustrup, 2006). Determining positions during interesting game situations, ratio of action and recovery time or the amount of physically exhausting actions like sprints or jumps would require an annotation of nearly each frame of a video sequence. In order to obtain this information with a feasible amount of user interaction, an automatic system for position computation is needed.

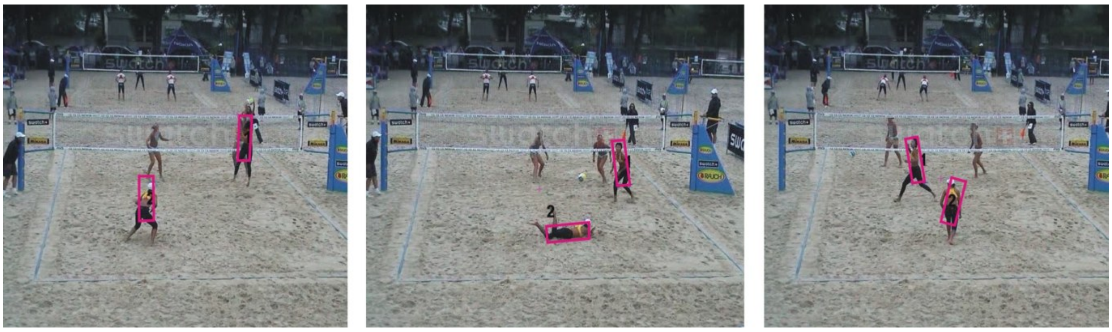

Figure 1. The proposed method can handle characteristic player motions like jumps and digs, which occur frequently during tracking.

Commercial applications

Existing commercial applications for more or less automated tracking and position estimation demonstrate the interest of teams and coaches in gathering such information. They can be roughly divided into two groups: systems using markers (active or passive) and markerless systems. Two representatives for the first group are Cairos® (Munich, Germany) and LPM® (Abatec AG, Regau, Austria). Both systems use a radio based method, where every tracked object is equipped with a transponder which position is measured by several base stations. This technique has the advantages that the 3D position is computed up to 1000 times per second with a high spatial accuracy and that the number of tracked objects can be high. A disadvantage of such systems is the possible influence of markers on the athlete's behavior though this problem has improved remarkably by minimizing marker size. However, as most of the sport rules prohibit the wearing of markers during competition the use of these techniques is very limited in sport practice.

Markerless systems are mainly based on video input. Their main advantages are that the influence on players during competition is zero and that also opponents can be observed. Furthermore, the obtained video data can also be used for feedback or tactical observations. One of the leading systems is Amisco Pro® (Nice, France). It is a commercial multi camera match analysis system (8 stable, synchronized and fixed camera orientations) approved by several European soccer clubs. Recently, the system has also been used scientifically to estimate covered distances and running velocities in international soccer as reported by Salvo, Baron, Tschan, Calderon Montero, Bachl & Pigozzi (2007). Although this system may provide interesting data the required technical and financial effort (especially for hardware) is an excluding factor for most type of sports. Therefore, it would be necessary to improve vision based tracking software to get accurate results without an enormous amount of technical effort.

Computer vision related work

On the one hand one can see that computer vision and in particular tracking are increasingly important for digital game analysis. On the other hand many different games like soccer, hockey, tennis and other type of sports have been used as test data for new computer vision approaches.

To handle the unpredictable behaviour of objects of interest during tracking sport games, e.g. athletes and ball, particle filter based methods have become common in that area. Since its introduction into the computer vision by Isard & Blake (1998), the particle filter has been used for various tasks and is a common method for player tracking in sports. The simplicity of the method, the ability to recover from uncertainties during tracking, and the possibility of fusing different information cues in one tracker are major advantages of this tracking method (see Perez, Vermaak & Ganget, 2002; Perez, Vermaak & Blake, 2004).

For the analysis of handball and basketball games Kristan, Perš, Perše & Kovačič (2006) have developed an indoor tracking system. Due to ceiling mounted cameras, the mutual occlusions between players are minimized. The tracking method uses a color based particle filter, considering the player as an elliptical region. The position of an object is then estimated by the center of tracking ellipse. Okuma, Taleghani, De Freitas, Little & Lowe (2003) combined a particle filter tracker with the detection results of an offline trained classifier to track hockey players. The tracking region of a player was defined to be an upright rectangle. Although the tracking results were quite impressive, results on estimated ground positions were not published. A comprehensive framework for automatic annotation of tennis matches was made by the group of Joseph Kittler (Yan, Christmas & Kittler, 2005). The tracking of players was done by subtracting the current frame from a pre-computed background image and using a blob tracker on the results. In addition, a support vector machine is trained to detect tennis ball candidates and a particle filter is used to track the ball.

Our approach

This work presents our approach in beach volleyball for a vision based tracking system which can be used in practice by trainers and athletes without extensive technical effort.

To achieve these goals, the tracking algorithm should be able to track athletes during competitions only by the use of a single camera and without complex calibrations. This has the positive side effect that already existing beach volleyball videos can be analyzed as well. Tracking information should then be used to compute real world coordinates which provide exact position and enable time-motion analysis. We assume an offline annotation and tracking scenario, where the whole game video is available. The tracking and position estimation process must not be fully autonomous. Therefore a small amount of user interaction for correction and re-initialization is acceptable.

Due to the playing characteristics of beach volleyball (and most other game sports) and the constraints due to a single camera the tracking algorithms used have to handle rotations and scale changes of bodies, e.g. squatting during a receive action. Specifically for beach volleyball the tracking methods should provide position information to improve the rating of techniques as well as motion analysis to estimate physical load (e.g. by detecting jumping movements as seen in Fig. 1).

Our approach consists of in three main parts: configuration, tracking and position estimation. In the configuration step, the transformation between video image and court coordinates is calibrated. Furthermore color and scale references are predefined for each player, by marking them in a single image. A background model is created automatically from input video. A color based tracker which is computational efficient by using integral structures forms the second part. It allows rotations to follow players during all possible motions and estimates the size of a player using the data from the configuration step. The tracker is only applied on the upper part of a player, which stays more compact during motions. If player positions are needed an additional segmentation step, using a skin color classification which was trained beforehand, is applied. This segmentation is only performed within an area defined by the tracker, where the lower part of the body is assumed, and therefore the additional runtime is negligible. Real world coordinates are finally estimated using the calibration from the configuration step. Figure 2 visualizes the main parts and the work flow.

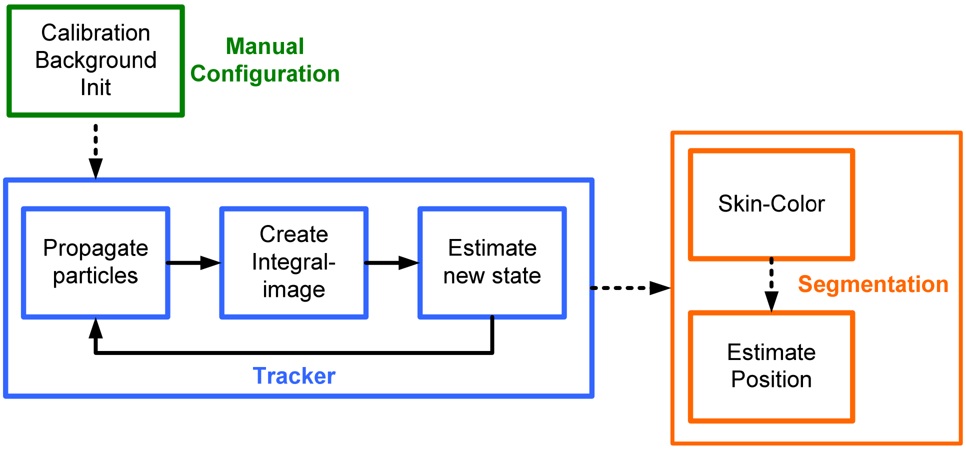

Figure 2.Vizualization of the main processing steps.

The remainder of the paper is organized as follows. First, the particle filter approach is summarized and the transition model used for tracking is explained. Furthermore, the computation of the color properties and the likelihood function of the particles for the single object tracker are described. Based on the tracking results, players are segmented from the background to allow the computation of real world coordinates. Experiments show an evaluation of tracking and position estimation results on manual annotated ground truth data. Finally, conclusion and summary are given at the end.

Methods

This section describes the tracking method used in this work. The particle filter concept is briefly explained in addition with the motion and appearance model used for player tracking. Object scale estimation and problems with position estimation from single view cameras are illustrated. In order to obtain real world coordinates, an additional segmentation step is performed.

Tracking with particle filter

The idea of the particle filter is to estimate the state $\mathbf{x}_t$ of a tracked object by using a set of weighted particles (Isard & Blake 1998). Each particle simulates the behavior of the object using Monte-Carlo simulations, a motion model and a measurement. Given a state space model $\mathbf{x}_{t-1}$ at time t-1 and all measurements up to t-1 known as $\mathbf{z}_{1:t-1}$ the posterior $p(\mathbf{x_t} \mid \mathbf{z}_{1:t})$ can be estimated by the recursion of Equations 1 and 2 using the new measurement $\mathbf{z}_t$.

Predict: $$p(\mathbf{x}_t \mid \mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1})\, p(\mathbf{x}_{t-1} \mid \mathbf{z}_{1:t-1})\, dx_{t-1} \qquad (1)$$

Update $$p(\mathbf{x}_t \mid \mathbf{z}_{1:t}) = \frac{p(\mathbf{z}_t \mid \mathbf{x}_t)\, p(\mathbf{x}_t \mid \mathbf{z}_{1:t-1})}{p(\mathbf{z}_t \mid \mathbf{z}_{1:t-1})} \qquad (2)$$

The required posterior density function p($\mathbf{x}_t$ | $\mathbf{z}_{1:t}$) of the new state can be approximated using sequential Monte Carlo simulations of a finite set of particles $\{x_t^i\}_{i=1\ldots Np}$. From an initial state, the weights $\{w_t^i\}_{i=1\ldots Np}$ associated with the particles are computed by sampling from a proposal distribution q($\mathbf{x}_t$ | $\mathbf{x}_{t-1}$,$\mathbf{z}_t$) (see Equation 3).

$$w_t^i \propto \frac{p(z_t^i \mid x_t^i)\, p(x_t^i \mid x_{t-1}^i)}{q(x_t^i \mid x_{t-1}^i, z_{1:t}^i)} \qquad \text{where} \qquad \sum_{i=1}^{N_P} w_t^i = 1 \qquad (3)$$

Using the state transition model p($\mathbf{x}_t$ | $\mathbf{x}_{t-1}$) as proposal distribution leads to the bootstrap filter, where the weights are directly proportional to the observation model p($\mathbf{z}_t$ | $\mathbf{x}_t$). Finally, the posterior density can be approximated by $p(\mathbf{x}_t \mid \mathbf{z}_{1:t}) \approx \sum_{i=1}^{N_P} w_t^i x_t^i$. To avoid the degeneracy of the particle set, resampling of the weights is done if necessary (see Arulampalam, Maskell, Gordon & Clapp, 2002, for more details).

State model used for players

During the tracking process players are described with rectangles given by center coordinates, size and rotation angle. In image coordinates the state model of an player at time t is defined by $\mathbf{x}_t$ = [x_t, y_t, υx_t, υy_t, φ_t] where (x_t,y_t) are the center coordinates of the rectangular window, (υx_t, υy_t) are the velocities and φ_t is the rotation angle of the player, see Figure 3. The size of a player (h,w) during tracking is computed directly, using the assumption of a fixed camera. The Homography **H** between image coordinates and real world court coordinates is determined with an initial manual calibration. If a player is annotated once as a reference, for example during initialization of the tracker, the scale parameters (h,w in Figure 3) for different court positions can be estimated using the Homography **H**.

The real state of a player $\mathbf{x}_t$ is estimated by a set of particles simulating possible states x^i_t. Applying an autoregressive model with an constant velocity assumption, the transition probability p($\mathbf{x}_t$ | $\mathbf{x}_{t-1}$) can be represented by:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{v}_t \qquad (4)$$

With this model the motion of particles is defined by a drift component defined in matrix **A**, equal for all particles of a player, and a random component in $\mathbf{v}_t$, which is assumed to be normally distributed for x, y and φ.

Using the homography

Assuming that image coordinates are given for each player, one would be interested in the real world coordinates. Reconstruction of full 3D coordinates of players or ball requires at least two cameras, which we do not have in our setup. However, under the assumption that the players are moving on a specified plane one can use a perspective mapping between two planes, defined by the Homography, to compute court positions of players (for details see Hartley & Zisserman, 2002).

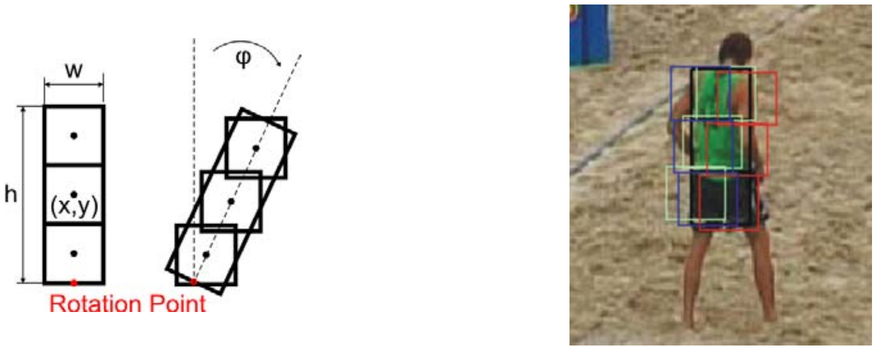

Figure 3. Left: Object description for players by a rectangular patch. For approximation of rotation the tracker is divided into three sub-parts. Right: Image shows 3 possible states of particles with different positions and orientations in blue, green, and red.

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \qquad (5)$$

Estimating the unknown Homography matrix H in Equation 5, requires at least 4 points in the image and in the target view, respectively. The linear transformation uses the homogenous coordinates of image points $[x\ y\ 1]^T$ to compute the transformation to the given target coordinates $[x'\ y'\ z']^T$. In the resulting rectified view perpendicular angles are reconstructed and, with the metric world coordinates given, one can reconstruct real world court coordinates (e.g. 8x16m for beach volleyball, visualization in Figure 4).

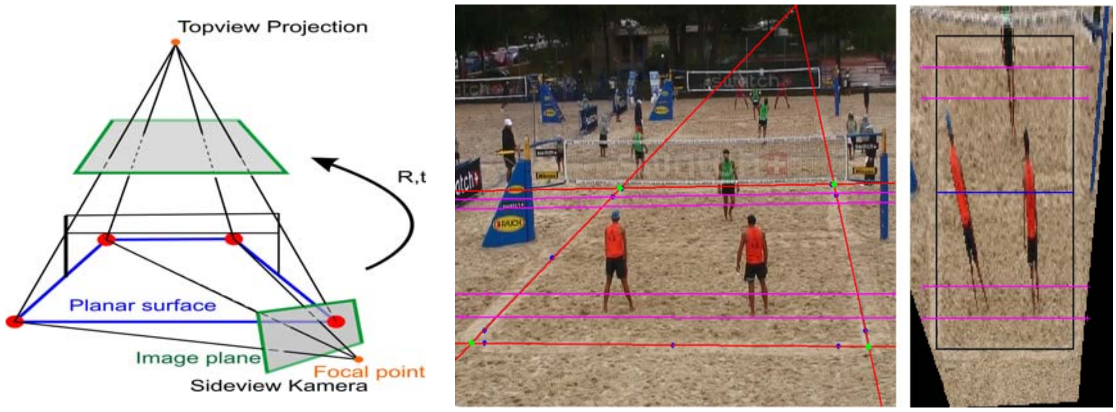

Figure 4. Left: Projection of real world points to the image plane of a camera and the transformation to a virtual top view projection. Right: The known coordinates of the playfield corners are used for calibration of the setup. Screenshot shows the perspective camera view and the undistorted top view image. The differences of the resolution in depth are shown by the parallel lines in both views.

With the world coordinates in meters and given the frame rate of the video stream, approximations of speed and acceleration can be calculated. The achievable accuracy of the field coordinates depends on the resolution of the camera as well as on the distance and orientation between player and camera. Players further away from the camera center have a lack of resolution, especially in the y-coordinates, and therefore less accurate positions can be calculated.

It has to be mentioned that the Homography transformation contains only the transformation between planes, as shown in Figure 4. Therefore, it only holds for points on the calibrated playfield. By tracking players who are not on the ground plane, the projected position would be wrong (see Figure 5).

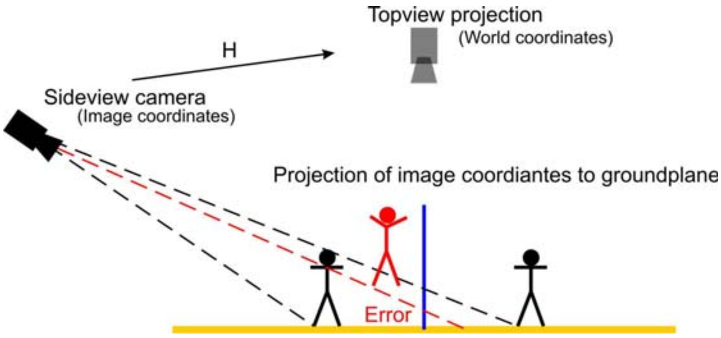

Figure 5. During jumps players are off the calibrated ground plane. The estimated positions contain an error especially in their y-coordinates, due to the assumed projection onto the ground plane.

Color tracking

To evaluate the set of particles, a measurement function has to be defined to see how good a particle fits to the real state of a player. Color information is a simple but powerful method to describe an object of interest. In contrast to shape description methods, which have also been used with particle filters, color information is less vulnerable to clutter. In particular, the intensive and distinct team colors in sports support the use of color histograms for our model description.

Using the HSV color space, an object is described with 3 independent N_B-bins histograms for the hue, saturation and value channel. An object, in our case a player, is initialized with three reference histograms [$\mathbf{h}^H_{ref}$, $\mathbf{h}^S_{ref}$, $\mathbf{h}^V_{ref}$] for the color channels. To compare candidate histograms [$\mathbf{h}^H_P$, $\mathbf{h}^S_P$, $\mathbf{h}^V_P$] sampled from a particle estimation with the reference histograms, the Bhattacharrya similarity coefficient D($\mathbf{h}_P$, $\mathbf{h}_{ref}$) is used. Combining the color channels the likelihood model $p(z^C \mid x)$ is finally assumed as exponentially distributed with a weighting constant λ as shown in Perez et al. (2002).

Histogram creation for each particle is a very time consuming task. Moreover, the particles overlap most of the time, so that many image areas are described several times. Porikli (2005) computed the histogram information of an image using the integral image approach, which leads to a drastic speed up. Additionally, the integral structure is only needed for the image area covered by particles, which is usually much smaller than the whole image.

Once the integral histogram is computed for an image, the histogram information of particles can be obtained using only three operations independent from position and scale of the particle. The disadvantage of using the integral structure is that it cannot be rotated. Lienhart & Maydt (2002), proposed a method to compute 45° rotations in the integral image which is not sufficient for our aims. Barczak, Johnson & Messom (2006) extended the set of possible rotations to any angle by approximating from pre-computed rotated images. Applying such an approach to a huge set of particles with different rotations would diminish the speed up achieved by the integral approach.

We decided to use an approximation approach similar to Grabner et al. (2006). The original tracking rectangle is divided into N_S subparts to approximate the rotation in the integral image (see Figure 3). Assuming that the subparts are independent, the color likelihood for a particle with state x and consisting of N_S subparts is finally computed by:

$$p(z^C \mid x) = \exp(-\lambda \cdot \sum_{j=1}^{N_S} \sum_{C \in \{H,S,V\}} D^2(h_{j,P}^C, h_{j,ref}^C)) \qquad (6)$$

The improper approximation of the object due to the rotated subparts is compensated by the high number of available particles. In addition, a spatial relation is integrated into the likelihood computation of the particles, which was also shown by Perez et al. (2002). This leads to more stable tracking results. Furthermore, the number of subparts and their spatial relation can be changed.

Including information about background

Usually, kernel or mask functions are applied to take into account that some background pixels are always included in the tracking window. To measure the influence of background pixels in our integral approach, a background probability $p(z^B \mid x)$ is included in the formulation of the measurement likelihood of the particles. Because of the static camera, the background image can be computed in a preprocessing step. Using Equation 6 also for the background similarity, the final observation model for a particle is given by:

$$p(z \mid x) = \frac{p(z^C \mid x\ )}{p(z^C \mid x\ ) + p(z^B \mid x)} \qquad (7)$$

For every particle i with state x_t^i at time-step t, the background similarity $D(\mathbf{h}_{j,P}^i, \mathbf{h}_{j,B}^i)$ is measured for each subpart j. The histogram $\mathbf{h}_{j,P}^i$ is sampled from the actual frame and $\mathbf{h}_{j,B}^i$ is computed for the same area in the background image. The integral structure for the background has to be computed only once beforehand (see Figure 6). Integrating the background probability prevents the tracker from drifting into background regions during mutual occlusions of the players.

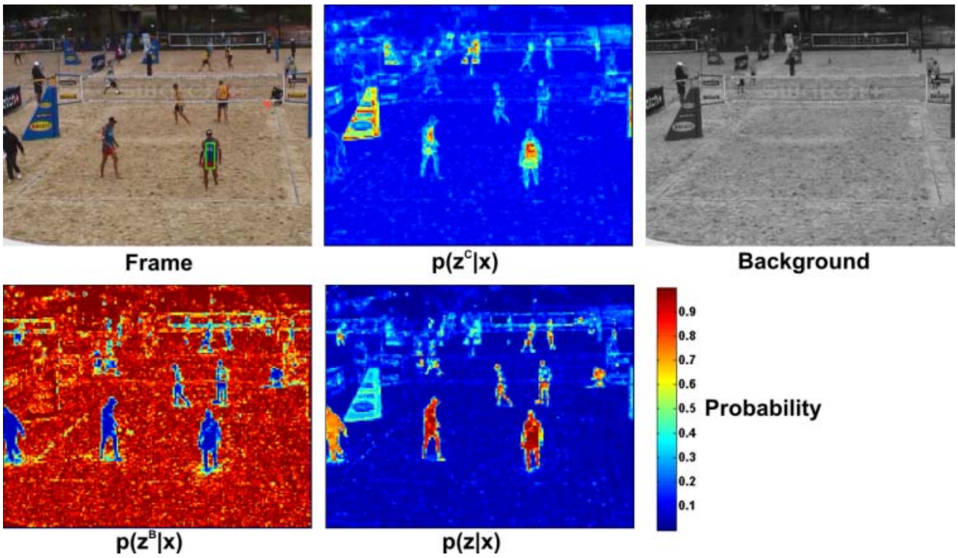

Figure 6. Top-row: Left: Actual input frame, where the green rectangle marks the patch used for the reference color histogram. Middle: Probabilities of pixels being the tracked object, using only color information. Similar colored objects in the background cause errors. Right: Pre-calculated background image. Bottom-row: Left: Probability of areas to be background. Right: Combined probabilities where background areas with colors similar to the tracked players, like advertisement spaces, are less likely now.

Segmentation and position estimation

Based on the obtained tracking results which are valid for the upper part of the athletes, the computation of the real field position is bounded on a smaller area. Knowing the position of the player, the rotation-angle of its body and an estimated scale, an area for segmentation is defined in the video frame.

The similarity between skin colored and sand colored pixels makes it hard to segment the patch in player and background. We chose to transform the RGB pixel into the YCbCr color space, which has shown good performance for face or skin segmentation tasks (Phung, Bouzerdoum & Chai, 2005). Using about 2 millions of skin, sand and background training pixels, a mixture of Gaussian models have been computed in an offline process to describe the different classes in the YCbCr color space (Figure 7 shows some segmentation results).

Figure 7. Left: Original input video frame. Middle: Segmented sand-colored pixel. Right: Skin segmentation can be used to create more accurate information about player positions.

Using the segmentation results of skin colored pixels and pixels containing to the player, known from tracking, the image can now be divided into player and background regions. Morphological operations are used to filter out small segmentation errors. The final segmentation result, which can be seen as an example in Figure 8, is a combination of the biggest segmented regions.

The estimated ground position in image coordinates (x,y) of a player is computed by the mean of all x-coordinates of the segmentation and the maximum y-coordinate. This is motivated by the fact that the mean represents the center of gravity of the region, respectively the player. The maximum y-coordinate is taken because of the used projection from image coordinates to real world coordinates.

One can see that the combination of tracking and segmentation results leads to more accurate results. Assuming a fixed size for the lower part of the player, or using only a rectangular window, would only be valid for players standing upright.

Evaluation experiments and first results

The following section contains an evaluation of the proposed tracker. The method is compared to manual annotations in terms of overlaps on players and estimated field positions. Therefore, a set of 12 test-sequences, each consisting of several hundred frames was used. All sequences, including male and female rallies, have been annotated manually at every third frame. Additional results for multi-object tracking in sport applications can be seen in Mauthner & Bischof (2007).

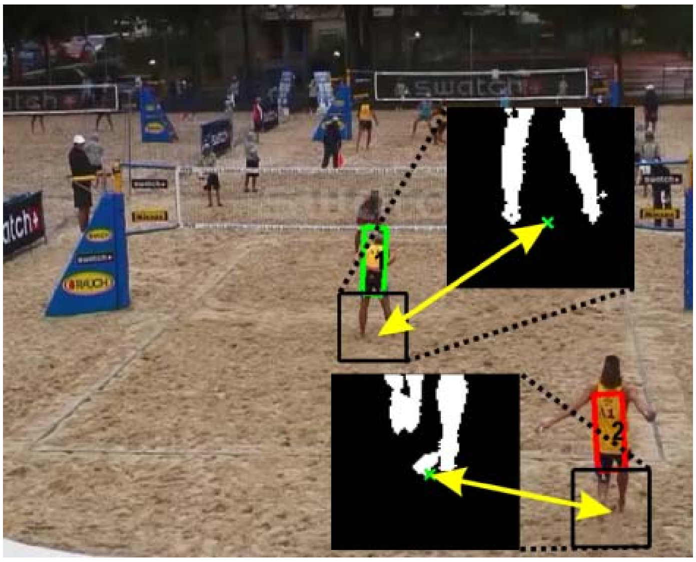

Figure 8. Segmentation of patches into player and background regions. Size and position of the segmented region are derived from the tracker result. Estimated positions are indicated by arrows.

Verification of tracker results

The reference ground-truth was manually annotated by experts familiar with beach volleyball. A rotated rectangle was placed over the upper part of the player by the annotators in every third frame of the test videos.

An overlap factor is computed from the shared area between the manual reference and the tracking result in relation to the total area of both rectangles (Figures 9 and 10). Total overlap of reference annotation and tracking result in an overlap factor of 1 and no overlap leads to a factor of 0.

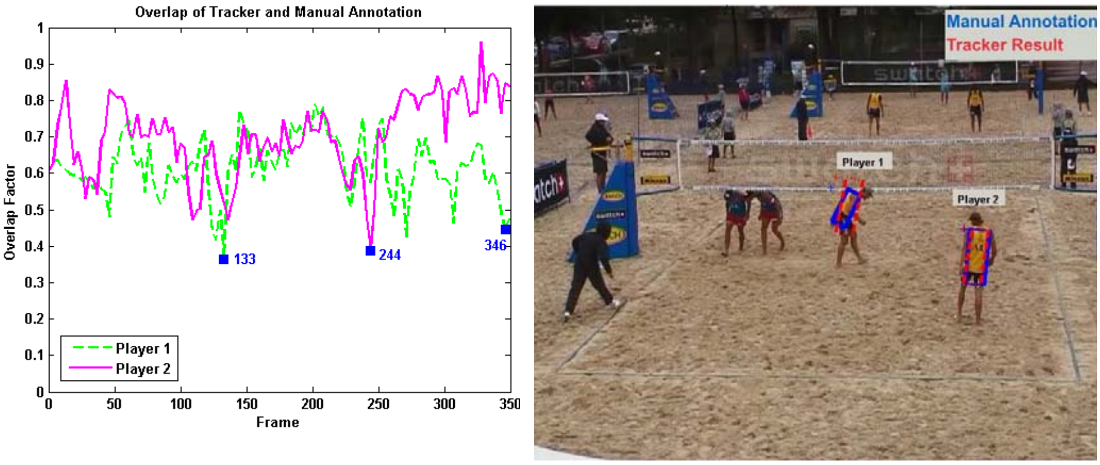

Figure 9. Left: Progression of the overlap factor during tracking of two players during sequence 1. Right: Manual annotation (blue rectangle) and tracking result (red dashed rectangle) for frame 346 in sequence 1. Both players are fully covered by their trackers, but according to the bending of player 1 the overlap factor is about 0.4, and lower than for player 2.

As described in the method section, the result of the tracker is computed over a weighted sum of the particles. The size of each particle is estimated from its position, and therefore, the size of the tracker result is always a combination of different scales. Additionally, the size of a torso is assumed to be fixed, which is not true, in image coordinates, if the players bend or crouch during the game. These mentioned difficulties do not exist if annotation is performed by a human, and therefore experts always scaled their reference rectangles to the visible part of the upper body. Low overlap values during such frames can be traced back on the differences between human perception and automatic computation (see Figure 9 and Figure 10).

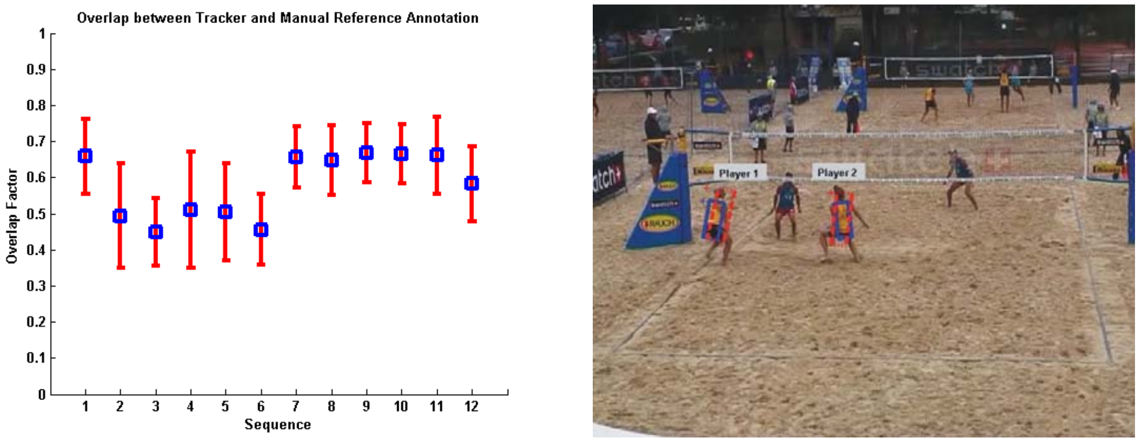

Figure 10. Left: Tracker overlaps on all test sequences, given by mean values and standard deviations. The sequences 2 - 6 belong to female games for which automatic tracking is less accurate due to the smaller amount of specific colors and the similarity between sand and skin color. Right: Manual annotation and tracking result for frame 133 of sequence 1.

A notable difference between sequences extracted from male and female games can be seen in Figure 10. Sequences from 2 to 6 are taken from female games, while sequence 1 and sequences 7 to 12 are from male competitions. This effect can be explained by the different appearances of the athletes. Male athletes wear shirts which, most of the time, are distinguishable from the background due to specific team colors. Because of the similarity between skin and sand color, the appearance of female athletes is not that precise during tracking which makes it harder to estimate the correct scale of a player. Nevertheless, the overall performance is balanced over the sequences. Considering the example frames given for sequence 1, an overlap factor of 0.45 still means an acceptable tracking result. Please note that no manual interaction was needed during the 12 test scenes.

Estimation of field positions

Similar to the tracker evaluation the reference position data consists of manual annotations. For every third frame of the test sequences, an image coordinate per player was defined as the reference position. The difference between manual reference and automatic position result is given as the Euclidian distance in image and real world coordinates.

Results of estimated positions are directly compared with the manual ground truth, without any additional filtering. The variance in the results could be reduced, if situations like jumps and occlusions between players would be excluded or corrected manually. Note that even varying human annotations can result in different position results. Figure 11 shows a visual comparison between estimated positions and reference annotations. Trajectories of one team during a rally are show individually in two separate top-view projections.

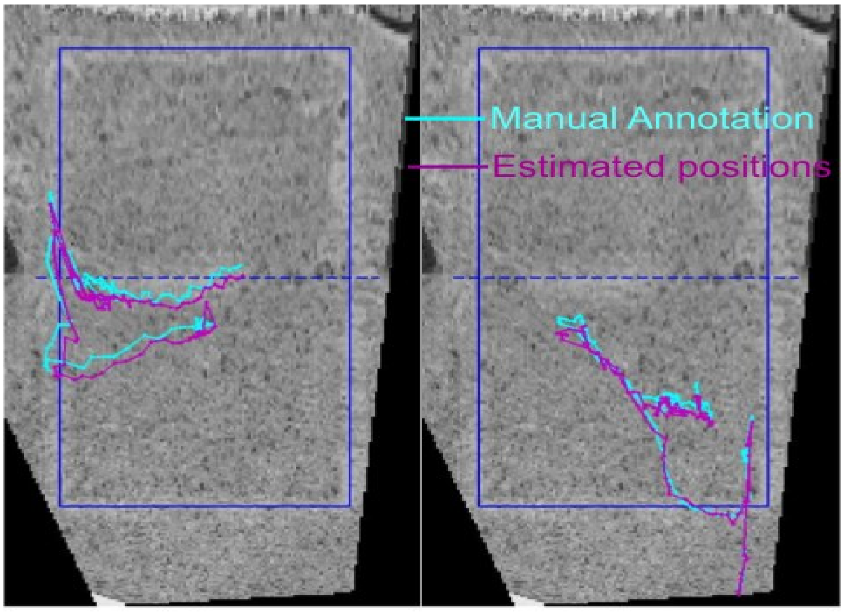

Figure 11. Comparision between manual annotation and results of our method. Positions are shown for every third frame only. Left and right images show the manually in comparison with automatically generated trajectories of two different players during a rally. Note left image: Projection causes a wrong position estimation into the opponent field during a jump movement (left side of the field). After the landing the player position was estimated correctly again.

As previously observed, several problems may occur when computing an exact athlete position from a single camera view. Depending on the point which is projected to real world field coordinates, the position can vary. Even the distance between both legs can be around half a meter. Additionally, the accuracy of the projection depends on the geometric resolution per pixel. In the used test sequences, the resolution varies between 3 cm/pixel and 10 cm/pixel, depending on the distance to the camera. Considering this fact, the results shown in Figure 12 are satisfying. Compared to the tracking results no difference between male and female games is observed. This effect can be led back to our special skin segmentation step.

Nevertheless, the results are accurate enough to answer several sport scientific questions and can be used for further analysis. Based on this position data, other parameters such as velocity and acceleration can be derived. Using the projected coordinates of the tracker, the resulting speed during jumps is not realistic due to projection errors (see Figure 11). Furthermore, a characteristic motion occurs, consisting of acceleration away from the camera followed by the inverse motion back towards the camera. The whole jump motion takes place in a maximum timeslot of about 30 frames. Such a shaped pattern can easily be found in the provided velocity data of each player and therefore could be exploited to detect jumps.

Conclusion and further work

A simple and yet effective method has been presented for tracking multiple objects within the scope of sport applications. The presented approach aims at obtaining position and motion information using the video input of a single camera, as this is the typical situation in sports practice. This aim could be achieved by combining several computer vision methods. By dividing the tracker window into subparts, the approximation of rotations in the integral histogram is possible. Therefore, sport specific motions can be followed with almost no additional runtime compared to using only an upright rectangle. Tracking results and segmentation of skin colored regions are combined to estimate real world court positions of athletes. Together with the possibilities of using the calibration for scale estimation during tracking and computation of real world coordinates, we are able to create useful tracking and position results for beach volleyball games.

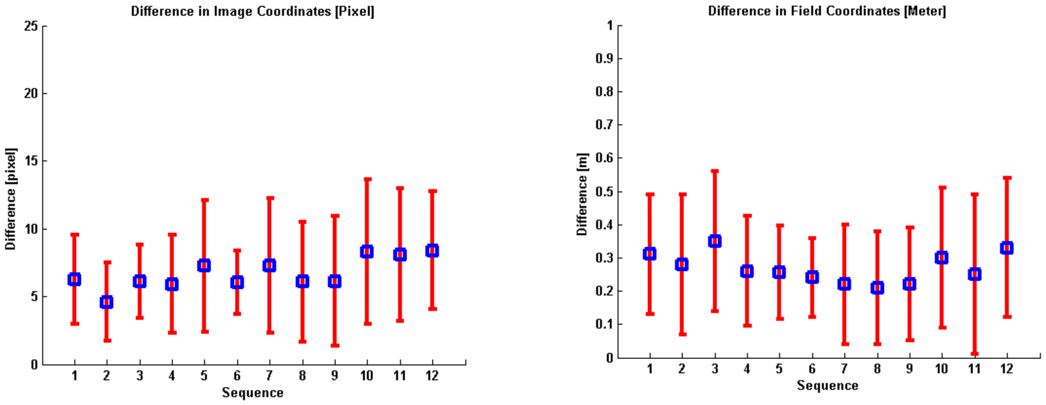

Figure 12. Comparison of manual annotated positions and the results of the automatic method in image and real-world coordinates.

The main advantages compared to existing methods are the rotation sensitivity, which delivers tracking results more similar to human annotation, and the combination of tracking and segmentation. As shown in the results, our methods deliver more information to analyze player positions than common rectangular or elliptical shaped trackers. We believe that the presented methods, together with a reasonable amount of manual interaction, are sufficient for motion analysis and the evaluation of physical demands in beach volleyball. Preliminary results indicate that it will be possible to detect frequency of jumping movements automatically in future. Furthermore, position data can be used to improve accuracy of action annotations and tactical analyses. The presented results are valid for beach volleyball but can be principally transferred to similar outdoor sports where fixed multiple camera systems are not available. A successful application of the presented method would be a great relief for the annotation process in several types of sports.

The integration of the proposed methods into existing game analysis software is the next step. Tracking and position data should be combined with expert annotations about player behaviour and used techniques.

Acknowledgement

This work has been supported by the FWF (Austrian Science Fund, P18600).

References

Arulampalam, S., Maskell, S., Gordon, N., & Clapp, T. (2002). A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transaction on Signal Processing, 50(2),* 171-188.

Bangsbo, J., Mohr, M., & Krustrup, P. (2006). Physical and Metabolic Demands of Training and Match-Play in the Elite Football Player. *Journal of Sports Science*, *24(7)*, 665–674.

Barczack, A., Johnson, M., & Messom, C. (2006). Real-time Computation of haar-like Features at Generic Angles for Detection Algorithms. *Research Letters in the Information and Mathematical Science, 9*, 98-111.

Grabner, M., Grabner, H., & Bischof, H. (2006), Fast approximated SIFT. *Proceedings of Asian Conference on Computer Vision*, 918-927.

Hartley, R., & Zisserman, A. (Eds.) (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hughes, M. & Franks, I.M. (2004). How to develop a notation system. In *Notational Analysis of Sport, 2nd Ed*., (edited by M. Hughes, & I.M. Franks), 118–140. New York, Routledge.

Isard, M., & Blake, A. (1998). Condensation—Conditional Density Propagation for Visual Tracking. *International Journal of Computer Vision. 29(1)*, 5-28

Kristan, M., Perš, J., Perše, M., & Kovačič, S. (2006). Towards Fast and Efficient Methods for Tracking Players in Sports. *CVBASE'06 Proceedings of ECCV Workshop on Computer Vision Based Analysis in Sport Environment*, 14-25.

Liebermann, D. G., & Franks, I. M. (2004) The use of Feedback-based Technologies. In: Hughes, M., Franks, I. M. *Notational analysis of sport*. Routledge, London/New York. 40-58.

Lienhart, R., & Maydt, J. (2002). An Extended set of haar-like Features for Object Detection. *Proceedings International Conference on Image Processing*, 900-903.

Mauthner, T., & Bischof, H. (2007). A Robust Multiple Object Tracking for Sport Applications. *Proceedings of Austrian Association for Pattern Recognition*, 81-89.

Okuma, K., Taleghani, A., De Freitas, N., Little, J., & Lowe, D. (2004). A Boosted Particle Filter: Multitarget detection and tracking. *Proceedings European Conference on Computer Vision*, 28-39.

Perez, P., Vermaak, J., & Ganget, M. (2002). Color-based Probabilistic Tracking. *Proceedings European Conference on Computer Vision*, 661-675.

Perez, P., Vermaak, J., & Blake, A. (2004). Data Fusion for Visual Tracking with Particles. *Proceedings of IEEE (issue on State Estimation)*, 92, 495-513.

Phung, S., Bouzerdoum, A., & Chai, D. (2005) Skin Segmentation using Color Pixel Classification: analysis and comparison. *IEEE Transaction on Pattern Analysis and Machine Intelligence*. *27(1)*, 148-154.

Porikli, F. (2005). Integral Histograms: A Fast Way to Extract Histograms in Cartesian Spaces. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition, 1*, 829-836.

Reilly, T., & Thomas, V. (1976) A Motion Analysis of Work-rate in Different Positional Roles in Professional Football Match-play. *Journal of Human Movement Studies, 8*, 159-176.

Salvo, Di, V., Baron, R., Tschan, H., Calderon Montero, F.J., Bachl, N., & Pigozzi, F. (2007) Performance CharacteristicsAccording to Playing Position in Elite Soccer. *International Journal of Sports Medicine, 28*, 222-227.

Spencer, M., Bishop, D., Dawson, B., & Goodman, C. (2005). Physiological and Metabolic Response of Repeated-spring Activities – Specific for Field-based Team Sports. *Sports Medicine*, *35(12)*, 1025-1044.

Tilp, M., Koch, C., Stifter, S., & Ruppert, G. S. (2006). Digital Game Analysis in Beach Volleyball. *International Journal of Performance Analysis in Sport, 6(1)*, 140–148.

Yamanaka, K., Haga, S., Shindo, M., Narita, J., Koseki, S., Matsuura, Y. & Eda, M. (1988). Time and Motion Analysis in Top Class Soccer Games. In T. Reilly, A. Lees, K. Davids, W. J. Murphy (Eds.). *Science and football*, 334-340, London: Spon.

Yan, F., Christmas, W. & Kittler, J. (2005). A Tennis Ball Tracking Algorithm for Automatic Annotation of Tennis Match. *Proceedings of the British Machine Vision Conference,* 2, 619-628.