Semi-automatic tracking of beach volleyball players

Gabriel Gomez[1], Andre Linarth[1], Daniel Link[2], Bjoern Eskofier[1]

Friedrich-Alexander Universität Erlangen-Nürnberg[1], Technische Universität München[2]

Introduction

For the optimal technical and tactical preparation of the national beach volleyball team, scouts are sent by the German Volleyball Association (DVV) to international tournaments to take videos from the top teams and make them available to the trainers. One challenging aspect to allow good video analysis is the software-based pre-structuring of these videos to specific game situations, such as serve direction, defense constellation or blocking technique. Structured videos are important for a good qualitative analysis of the video material, and are needed to generate statistical information about individual players. As the pre-structuring of the video material is very time consuming, it is only possible to analyze a restricted number of recorded matches, which leads to a possible loss of important information. To improve this, a game monitoring method was developed by the TU Munich together with the DVV in 2010, where ball exchanges can be classified based on their structure in time and space and positional data of the players (Link et al. 2010). By capturing only a restricted number of characteristic player positions at certain time points it is possible to classify and identify several different game situations. This method is the basis for the game monitoring used by the German national team and was successfully used in the Olympic Games 2012. While this method delivers very good results, player positions still need to be manually input, e.g. via a tablet pc.

With the increase of computational power in the last decades, tracking of objects or human beings has become a growing research area in the image processing and engineering fields. Its applications range from military and security to medicine and sports. In particular, tracking applications in sports have the main purpose of extracting useful information for the analysis of a player’s or team’s performance through video analysis based on statistical measures (Kristan et al. 2009). In the hard competitive environment of today’s professional sports, tactical discussions before and/or after a game based on the analysis of pre-structured videos of the own team as well as of the opponent teams can help to better prepare tactically for a game or for improving the techniques and strategies. Most tracking methods in professional sports have one or more advantages that improve tracking results. A) The use of multiple cameras such as commonly used in soccer games. B) The position of the camera being above the field facing right down or bird’s eye view. C) The background being partly or entirely homogeneous. D) Constant lighting conditions by artificial illumination. The mentioned advantages do not apply for beach volleyball, where tracking of the players is a difficult task due to severe player and net occlusion, perspective size changes, changing lighting conditions and blinking advertisement in the back of the field (see fig. 1). This paper presents a novel way of tracking beach volleyball players using Particle Filters and a single camera from a spectator’s perspective. The goal of this work is the implementation of a tracking method that allows positional determination of the four players at ball contacts. This enables the extraction of statistical measures for video analysis, and the classification of game situations such as serves, attacks or defense. This work will be integrated into the mentioned framework of the Technical University of Munich and the German Institute of Sport Science (Bundesinstitut für Sportwissenschaft, BISp), that allows the analysis of pre-structured beach volleyball videos based on real time tablet PC inputs by scouts during a game (Link et al. 2012).

Fig.1: Example frame from a typical beach volleyball video file used for the analysis and tracking. The extreme shadowy and sunny regions and the camera perspective make the tracking of the players challenging.

Previous work

Several methods for detection and tracking of players or pedestrians have been proposed, such as kernel-based object tracking (Comaniciu et al. 2003) or Bayesian and Markov chain tracking (Tzhao and Nevatia 2004). In visual tracking one often deals with non-linear and multi-modal problems that make Kalman-filtering unsuitable for the task (Perez et al. 2002). Particle Filters have become popular in recent years due to their robustness and good results compared to other methods, allowing arbitrary distributions, and due to their simplicity of implementation. There have been numerous works published on player tracking, mainly in sports like soccer, American football or ice hockey. Color based Particle Filter tracking of beach volleyball players using an integral histogram approach has been proposed in the past with good results (Mauthner et al. 2007).

Methods

In our framework, we target the tracking problem by means of Particle Filters, as proposed in (Perez et al. 2002). For each of the four players, a particle set is created to track one player over the entire game. The state variables of the particles represent the 2D position of the players on the field. Each particle is a weighted hypothesis of the current location of the player, and the set of particles represents the Probability Density Function (PDF) of the presence of a player on the field. For calculating the likelihood of a single hypothesis, the expected bounding volume of the player at each particle's position is projected into the image space. This is achieved through a homographic transformation between image and field planes (Hardley and Zisserman 2000). The bounding box is subdivided into four equally sized regions, for which color histograms are calculated (see fig. 2). These histograms are correlated with reference histograms extracted from the player, by means of the Bhattacharyya distance:

$$D\left(\boldsymbol{h}^{ref}, \boldsymbol{h}^{hyp}\right) = \sqrt{1 - \sum_{b \epsilon B} \frac{\sqrt{\boldsymbol{h}_b^{ref} \boldsymbol{h}_b^{hyp}}}{\sqrt{\sum_{b \epsilon B} \boldsymbol{h}_b^{ref} \cdot \sum_{b \epsilon B} \boldsymbol{h}_b^{hyp}}}} \tag{1}$$

Where B denotes the number of histogram bins, and $\boldsymbol{h}_b$ represents the value of the b-th bin of the reference or hypothesized histogram. The smaller this distance, the more likely it is that the respective particle's state corresponds to the player's position.

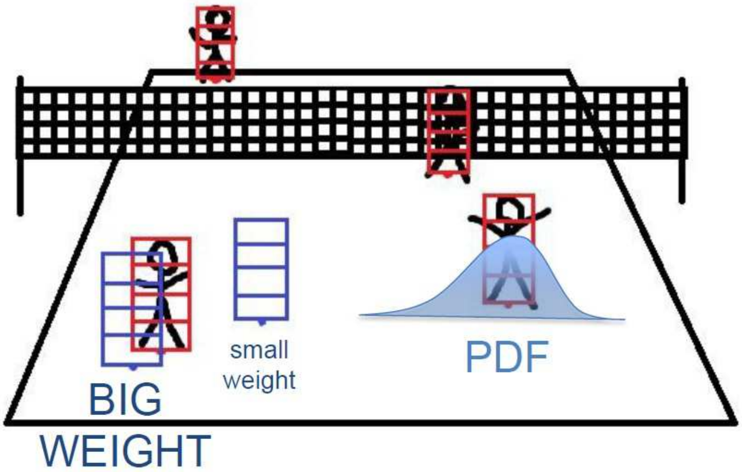

Fig.2: Schematic diagram of the field and the four players with their corresponding bounding boxes, two bounding boxes in blue corresponding to exemplary particles and the probability density function of a particle set.

A second cue applied to model the likelihood lies on the distinction between player and background. A background subtraction algorithm based on Gaussian Mixtures (Zivkovic and van der Heijden 2006) is applied to retrieve foreground pixels. The likelihood is modeled to be proportional to the sum of foreground pixels contained in the hypothesized bounding box. Therefore, by combining color histogram and foreground cues (see fig. 3), the likelihood of a single particle is modeled by:

$$\omega_j^i = \frac{c}{\sum_k D_{i,j}} + \sum_k F \qquad (2)$$

Where ω is the weight of the j-th particle corresponding to the i-th player; k is the number of subregions of the bounding box; $D_{i,j}$ is the Bhattacharyya distance given in (eq. 1), calculated with the reference histograms of the i-th player and the hypothesis given by the j-th particle; F is the normalized sum of foreground pixels at the sub-region with respect to its total area; and c a constant that balances the contribution of the color based cue with respect to the motion based foreground cue. The final model for each frame in the video, i.e. the predicted players' positions, is given by the average state of each particle set. For subsequent frames, a particle set is created for each player, according to the PDF retrieved in the previous step. In other words, particles are resampled proportionally to the likelihood distribution shown in the previous frame. After this resampling step, Gaussian noise is added to the new particles' positions to account for player movements on the field.

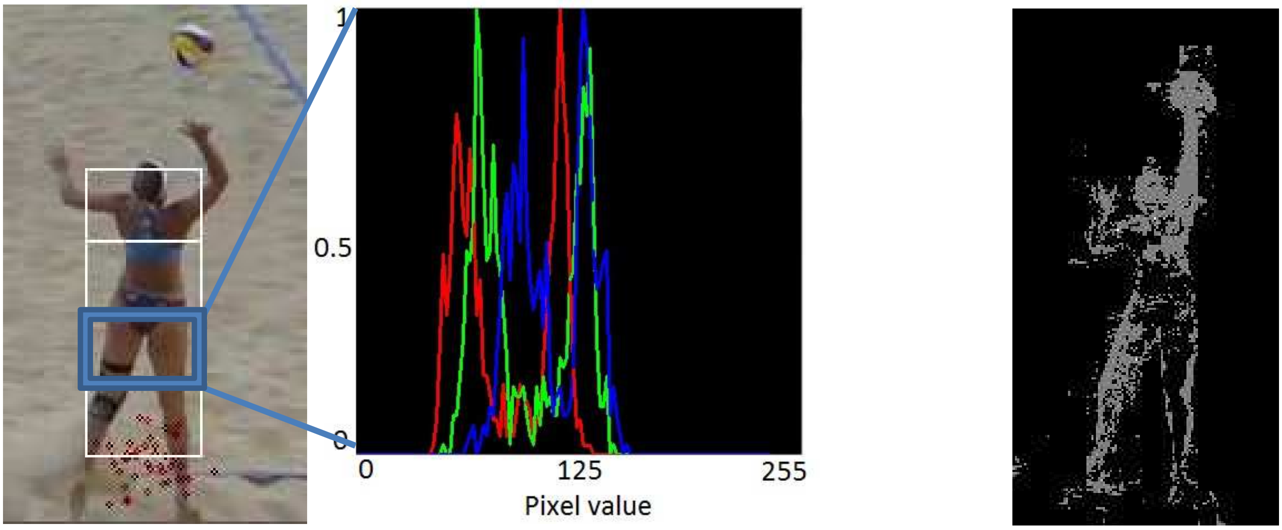

Fig.3: Serving player with its corresponding bounding box and subwindows (left); Sample normalized histogram of a subwindow for the R, G and B channels (middle); Binary image from the background subtraction algorithm where the white pixels show motion from subsequent frames (right).

Results

The tracking algorithm was tested on videos taken from professional beach volleyball games made available by the BISp. A good measure for the accuracy of the tracking algorithm is the ratio of the number of frames for which the algorithm tracks the players correctly, to the total number of frames. In total, 23 video sequences of an average of 247 frames each were tested. For all the sequences, ground-truth player positions were extracted by manual annotation. The results are summarized in table 1.

	Average number of frames tracked (rounded)	Percent of correctly tracked frames
Player 1 (front left)	237	95.9 %
Player 2 (front right)	236	95.4 %
Player 3 (back right)	228	92.3 %
Player 4 (back left)	218	88.3 %
Average players front	237	95.7 %
Average players back	223	90.3 %

Table 1: Average results of the tracking algorithm for 23 video sequences with an average of 247 frames.

Discussion

As can be seen in table 1, tracking players in the front shows to be more effective compared to tracking players in the back. With 95.4 percent correctly tracked frames, players in the front can be robustly followed even when the players cross each other, i.e. in the presence of partial occlusions. On the other side, tracking of the players in the back shows less reliable results for a number of reasons. First, the presence of the net between- player and camera considerably alters the color histograms. Second, players at the back appear smaller in the image due to the perspective projection. This smaller resolution directly influences on the likelihood computation. Finally, players closer to the camera severely occlude the ones in the back. While this can be dealt with for the foreground players, in conjunction with the other aforementioned problems it becomes more problematic for players in the back, leading to lower tracking accuracy. Nevertheless, the shown results demonstrate that the proposed tracking method delivers a high tracking accuracy when using videos with good lighting conditions. Although the players are not correctly tracked in all frames, this method can help increase the number of videos analyzed. Therefore more games can be analyzed in less time, saving time to structure the video material and get more accurate statistical information about the games. Also, some features that could not be analyzed due to time constraints before can now be included in the video analysis.

The use of an appropriate likelihood model is the most crucial point while designing a Particle Filter. It is sought to find a PDF that delivers a high value in the position of the tracked player, but very low values everywhere else, as proposed in eq. 2. Since the main cue used for the weighting of the particles is a color histogram, the probability density in the image will hardly be high for a single player and low everywhere else. On the contrary, the probability density map will give highest values for the selected player, but also high values for other players or spectators wearing similar colors. Therefore, the selection of the weighting scheme is done such that the regions of high probability density are spatially confined and do not overlap.

To improve the results of the tracking, especially for the back players, a better weighting algorithm that includes more hints will be implemented in future work. Also, tracking of the ball to find the time points of ball contact will be done in the future to be able to integrate the tracking into the TUM/BISp video analysis framework.

References

M. Kristan, J. Perš, M. Perše, S. Kovačič, Closed-world tracking of multiple interacting targets for indoor-sports applications, Computer Vision and Image Understanding, Volume 113, Issue 5, May 2009

D. Comaniciu, V. Ramesh and P. Meer, Kernel-Based Object Tracking, IEEE Trans. PAMI, vol.25, no.5, 2003.

P. Perez, C. Hue, J. Vermaak, and M. Gangnet. Color-Based Probabilistic Tracking. In ECCV, pages 661–675, 2002.

T. Zhao and R. Nevatia, “Tracking Multiple Humans in Crowded Environment,” proc. IEEE Conf. Computer Vision and PatternRecognition, 2004.

Link, D., Müller, J., & Lames, M. (2012). Der Beachscouter – Das Erfassungstool des Deutschen-Volleyballverbandes für Spieldaten im Beachvolleyball. In H. Wagner (Hrsg.), *NeuroMotion – Aufmerksamkeit, 9. gemeinsames Symposium der dvs-Sektionen Biomechanik, Sportmotorik und Trainingswissenschaft*, Tagungsband (S. 129), Münster.

Link, D., Ahmann, J. et al. (2010). Leistungsdiagnostik im Beachvolleyball auf Basis der Positionsdaten. Höchstleistungen in den Sportspielen: Vom Talent zur Expertise. Abstractband, p. 47, Münster, 2010.

Z. Zivkovic and F. van der Heijden. Efficient Adaptive Density Estimation per Image Pixel for the Task of Background Subtraction. In Pattern Recognition Letters, vol. 27, no. 7, pages 773-780, 2006.

R. Hardley and A. Zisserman, Multiple view geometry in computer vision, Cambridge University Press, 2000.

Mauthner, T., Koch, C., Tilp, M., & Bischof, H. Visual tracking of athletes in beach volleyball using a single camera. In L. Katz, R.M. Sorrentino (Edts.), Proceedings of 6th International Symposium on Computer Science in Sport, University of Calgary, 2007