A Robust Multiple Object Tracking for Sport Applications [1)]

Thomas Mauthner, Horst Bischof

Institute for Computer Graphics and Vision

Graz University of Technology, Austria

{mauthner,bischof}@icg.tu-graz.ac.at

Abstract:
This paper deals with the task of tracking several similar looking athletes during a competition, using only a single camera. Due to the wide range of possible motions, non-rigid shape changes and mutual occlusions the tracking situations can become quite complex. We propose a novel method to use a high dimensional motion model for a particle filter without drastic increase in runtime. Integral histograms are extended to handle also rotated objects. In order to deal effectively with overlapping players we propose a re-weighting and re-sampling method for a particle filter. The paper is primarily designed for beach volleyball but is also evaluated on public available football sequences.

1 Introduction

Video based information about team sports can be used in different ways to increase the performance of teams and athletes [11]. Determining positions during interesting game situations, ratio of strain and relax time or physically exhausting actions like sprints or jumps would require an annotation of nearly every frame of the video sequence. For that reason computer vision and in particular tracking is of increasing importance for digital game analysis. Many different games e.g. soccer, hockey and other sports have used tracking in the past [7, 5]. Since its introduction into the computer vision the particle filter has been used in 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 to fuse different information cues in one tracker are major advantages of this tracking method [8, 9].

There are several weaknesses of particle filter based trackers. First, a state model which describes the tracked objects as good as possible has to be defined. A motion model for the state transition is needed as a prior probability. Second, the number of particles needed

[1)] This work was supported by the Austrian Science Fund (FWF P18600).

to adequately sample the probability space increases exponentially with the number of state space dimensions, leading to a high computational effort. Third, multiple object tracking requires special treatment if the objects are visually and spatially close together, due to the characteristic of the particles to drift to the strongest mode in the probability space.

Tracking athletes during beach volleyball games is the main objective of this work. To deliver accurate results the tracker has to handle rotations of players and follow players during jumps and other frequent motions, see example frames in Fig. 1, which leads to a high-dimensional state space. We propose an efficient method to keep the computational time short, based on integral histograms [10]. Separate color integral images are computed for every object but only over an area defined by the transition distribution of its particle set. This leads to an optimal use of the integral structure. Since the rotation of the object model is also needed, the rotation in the integral structure is approximated by sub-regions, similar to [3]. An additional re-sampling step provides the robustness for multi-object tracking and minimizes the number of lost trackers. The particles of several single object trackers are re-weighted using an overlap measure.

The remainder of the paper is organized as follows. In Section 2 the particle filter approach is summarized and the transition model used is explained. Section 3 shows the computation of the color properties and the likelihood function of the particles for the single object tracker. The additional re-sampling step for multi-object tracking is defined in Section 4. Results of the tracker in sports videos compared to the standard particle filter are shown in Section 5. Finally a conclusion and summary of the paper is given in Section 6.

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

2 Particle Filter Approach

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 Equ. 1 and 2 using the new measurement $\mathbf{z}_t$.

$$\textbf{predict}: p(\mathbf{x}_t \mid \mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} \mid \mathbf{z}_{1:t-1}) d\mathbf{x}_{t-1} \tag{1}$$

$$\mathbf{update}: p(\mathbf{x}_t \mid \mathbf{z}_{1:t}) \propto \frac{p(\mathbf{z}_t \mid \mathbf{x}_t)p(\mathbf{x}_t \mid \mathbf{z}_{1:t-1})}{p(\mathbf{z}_t \mid \mathbf{z}_{1:t-1})} \tag{2}$$

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

$$w_t^i \propto w_{t-1}^i \frac{p(\mathbf{z}_t \mid \mathbf{x}_t^i)p(\mathbf{x}_t^i \mid \mathbf{x}_{t-1}^i)}{q(\mathbf{x}_t^i \mid \mathbf{x}_{t-1}^i, \mathbf{z}_t)} \quad where \quad \sum_{i=1}^{N_p} w_t^i = 1 \tag{3}$$

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

2.1 State Space and Transition Model

During the tracking process a single fixed camera is assumed. Using a manual calibration step the homography between image coordinates and real world pitch coordinates is estimated. If a player is annotated once as a reference, for example during initialization of the tracker, the scale changes for field positions can be estimated using the homography. In image coordinates

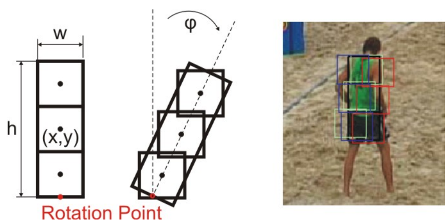


Figure 2: Motion model for players and fragmentation into sub-parts.

the state model is given by $\mathbf{x}_t = [x_t, y_t, v_x, v_y, \varphi_t]'$, where (x, y) are the center coordinates of a rectangle window, (v_x, v_y) are the velocities and φ is the rotation angle, see Fig. 2. Applying an autoregressive model the transition probability $p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$ can be represented by:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + \mathbf{v}_t \tag{4}$$

With this model the transition of particles is defined by a drift component defined in A, computed from the last position and velocities, and a random component in $\mathbf{v}_t = \mathcal{N}(0 \mid \Sigma)$ where σ_x, σ_y and σ_φ are the assumed variances. The possible scale changes of the objects are approximated by the homography.

3 Color Model

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. Especially the intensive and distinct team colors support the use of color histograms for our model description.

Using the HSV color space, an object is described with independent N_B-bins histograms for the hue, saturation and value channel. An object is separated into subparts and each subpart is initialized with three reference histograms $[\mathbf{h}_{ref}^H, \mathbf{h}_{ref}^V, \mathbf{h}_{ref}^S]$ for the color channels. To compare a candidate histogram $[\mathbf{h}_P^H, \mathbf{h}_P^V, \mathbf{h}_P^S]$ sampled from a particle estimation with the reference histograms, the Bhattacharyya similarity coefficient $D(\mathbf{h}_P, \mathbf{h}_{ref})$ is used. Combining the color channels the likelihood model is finally assumed as exponentially distributed with a weigthing constant λ, see also [8].

$$p(z^C \mid x) = exp\left(-\lambda \cdot \sum_{C \in \{H,S,V\}} D^2\left(\mathbf{h}_P^C, \mathbf{h}_{ref}^C\right)\right) \tag{5}$$

3.1 Color Cues using Integral Structures

Histogram creation for each particle is a very time consuming task. Moreover, the particles overlap most of the time, therefore many image areas are described several times. In [10] Porikli showed the possibility of computing the histogram information of an image using the integral image approach. Once the integral histogram is computed for an image, the histogram information of particle areas can be obtained using only 3 operations, independent from position and scale of the particle. The disadvantage of using the integral structure is that it cannot be rotated. In [6] a method is proposed to compute 45° rotations in the integral image which is not sufficient for us. Barczak et.al. [2] 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 at different angles would diminish the speed up achieved by the integral approach.

Because the final state of an object is estimated using a weighted sum over a set of particle states, we decided to use an approximation approach also used by [3]. The original tracking rectangle is divided into N_S subparts to approximate the rotation in the integral image, see Fig. 2. Making the simplifying assumption that the subparts are independent, the color likelihood for a particle with state $\mathbf{x}$ and consisting of N_S subparts is finally computed by:

$$p(z^C \mid x) = exp\left(-\lambda \cdot \sum_{j=1}^{N_S} \sum_{C \in \{H,S,V\}} D^2\left(\mathbf{h}_{j,P}^C, \mathbf{h}_{j,ref}^C\right)\right) \tag{6}$$

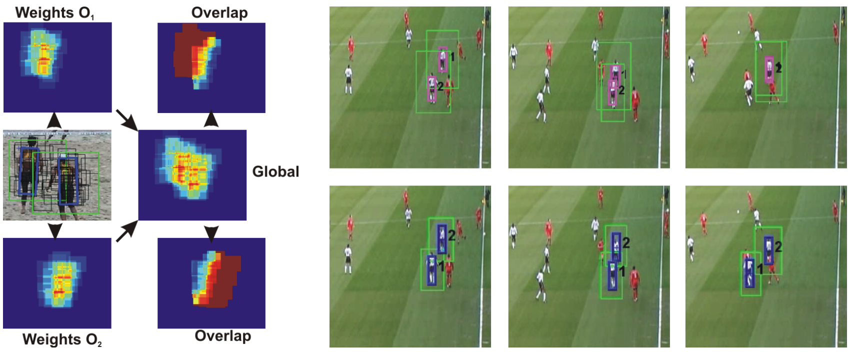
The improper approximation of the object due to the rotated subparts is compensated by the high number of available particles, and their weighted sum in the final tracker result. In addition a spatial relation is integrated into the likelihood computation of the particles, which as shown by [8], leads to more stable tracking results. Also the number of used subparts and their spatial relation can be changed. Usually kernel or mask functions are applied to take into account that some background pixel are always included in the tracking window. In our integral approach a background probability $p(\mathbf{z}^B \mid \mathbf{x}^B)$ 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 Equ. 5 also for the background similarity, the final observation model for a particle is given by:

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

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

4 Tracking with Multi-Object Resampling

The similar appearance of players that are tracked leads to a number of problems. During spatial proximity or partial occlusions the particle set of the individual object trackers overlap, which leads to multiple modes in the filtering distribution. Resampling the particle sets individually can cause losing track of objects, when several trackers will drift to the same player, see Fig. 3. Occlusion rules can be defined by using additional variables in the state model, which is practicable only for a small number of objects. In [4] MCMC iterations are used to describe interactions of objects at the cost of an additional computational complexity. Vermaak et.al [12] used clustering to maintain the multimodality of several similar objects in one particle filter. Because of the clustering approach split and merge operations are performed during overlap and therefore identities of objects are lost. In this work we decided to use an overlap criterion of the particle sets to resample the weights in a global context. An additional weighting step separates the particle sets and keeps particles from drifting to similar objects. The partition of particles after total occlusions is also encouraged. The particle sets are supervised individually for each team sharing a common appearance. First the single object trackers distribute their particles and evaluate their weights independently. The weights of all N_O objects are stored in individual weight maps, where each coordinate contains the sum of all particles which overlay this position, see Fig. 3. The division of single object weights by the global sum creates an overlap factor [0...1] for each coordinate and particle respectively. Particles separated from other objects get an overlap factor of 1, which means that their weights are not changed. Particles which share their location with particles form other objects get an overlap factor < 1, so they get less weight and are sorted out in the final resampling step. The computation of the weight maps and the overlap coefficient requires no additional runtime. During a complete occlusion between objects, no separation is possible but the trackers can share the same image area. Fig. 3 shows the effect of the multi-object resampling on an example sequence from the PETS [1)] dataset. Without the additional re-weighting and re-sampling step particles drift to the strongest mode and the second track is lost.



(a) Overlap factor. (b) Multiple Objects

Figure 3: **(a)Visualization of the overlap factor. (b)Soccer image sequence shows the typical results for several single particle filters. After the intersection particles of both trackers evolve to the same mode. Using the proposed additional resampling step with the overlay weights the particle sets stay separated. This can also be seen by the green rectangles which mark the areas of the associated integral histograms.**

5 Experiments

Fig. 4 shows the runtime of the proposed method in comparison to the use of standard color histograms. The blue line marks the computational time for the creation of only 10 standard color histograms of 30x60 pixel patches in Matlab without rotation as a reference. In contrast the histogram approach shows the runtime for a multi-object tracker handling two rotatable single object trackers. The approximation shown in Section 3.1, makes it possible to track rotating objects without a large overhead. The difference between the normal integral time and the runtime with rotation depends only on the number of subparts that have to be rotated. Three different beach volleyball video scenes have been used to evaluate the proposed method on multi-object handling situation. The sequences have been captured during the Grand Slam 2006 in Klagenfurt, and therefore show real competition behavior of male and female athletes, see Fig. 1 and 5. In a set of 9623 frames, 29 test scenes are selected, in which single object particle trackers fail during occlusions or spatially near actions. All multi-object tracking results have been achieved with 100 and 200 particles per object, each particle divided into three subparts. The results are summarized in Table 1. With the additional multi-object

[1)] http://ftp.pets.rdg.ac.uk/VS-PETS/

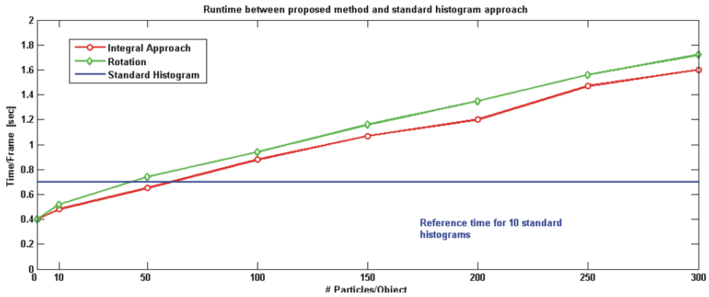


Figure 4: **Runtime comparison between proposed histogram approach and standard histogram method. The rotation in the integral histogram struct needs only the rotation of point coordinates, which makes it really fast compared to the standard histogram sampling.**

resampling step 79% of situations where single particle filters fail on tracking both players are solved correctly using 200 particles per object. In 68% of the cases the object is successfully identified during the intersection. In Fig. 5 an image sequence from a typical game situation summarizes the tracking results. Using the proposed method, the tracker does not drift away onto the background during a occlusion but follows the occluding player, which corresponds to a human annotation. After the intersection of the two tracked players the particles sets are separated but identities of objects are switched. The maintenance of the identities of two similar objects in identically states during a total occlusion can not be handled directly with the proposed method.

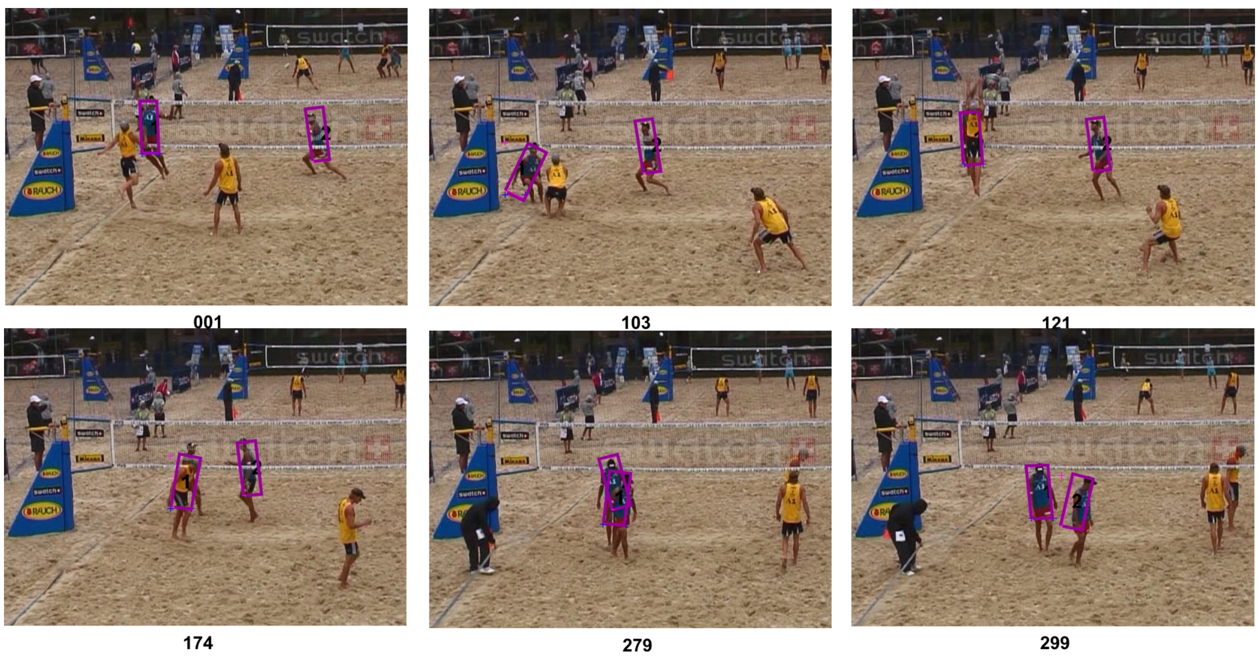


Figure 5: **Tracking results of the final method. During occlusions from opponent players, the tracker follows the occluding player, frames 121 and 174. After an intersection in frame 279 particle sets are separated again but identities are switched.**

#Particles	Multi-Object	Correct ID
100	68%	55%
200	79%	68%

Table 1: Results of successful multi-object tracking.

6 Conclusion

A simple, fast and yet effective method was presented for tracking multiple similar objects within the scope of sport applications. 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. The experiments showed the performance gain and the additional robustness during multi-object tracking.

References

[1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[2] A. L. C. Barczack, M. J. Johnson, and C. H. Messom. Real-time computation of Haar-like features at generic angles for detection algorithms. *Research Letters in the Information and Mathematical Scienece*, 9:98–111, 2006.

[3] M. Grabner, H. Grabner, and H. Bischof. Fast approximated SIFT. In *Proceedings Asian Conference on Computer Vision*, pages 918–927, 2006.

[4] Z. Khan, T. Balch, and F. Dellaert. An mcmc based particle filter for tracking multiple interacting targets. In *Proceedings European Conference on Computer Vision*, pages 279–290, 2004.

[5] M. Kristan, J. Perš, M. Perše, and S. Kovačič. Towards fast and efficient methods for tracking players in sports. In *CVBASE '06- Proceedings of ECCV Workshop on Computer Vision Based Analysis in Sport Environment*, pages 14–25, 2006.

[6] R. Lienhart and J. Maydt. An extended set of haar-like features for object detection. In *Proceedings International Conference on Image Processing*, pages 900–903, 2002.

[7] K. Okuma, A. Taleghani, N. De Freitas, J.J. Little, and D.G. Lowe. A boosted particle filter: Multitarget detection and tracking. In *Proceedings European Conference on Computer Vision*, pages 28–39, 2003.

[8] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *Proceedings European Conference on Computer Vision*, pages 661–675, 2002.

[9] P. Pérez, J. Vermaak, and A. Blake. Data fusion for visual tracking with particles. In *Proceedings of IEEE (issue on State Estimation)*, volume 92, pages 495–513, 2004.

[10] F. Porikli. Integral histogram: A fast way to extract histograms in cartesian spaces. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 829–836, 2005.

[11] M. Tilp, C. Koch, and G. Ruppert. Digital game analysis in beach volleyball. *International Journal of Performance Analysis in Sport*, 6(1):140–148, 2006.

[12] J. Vermaak, A. Doucet, and P. Pérez. Maintaining multi-modality through mixture tracking. In *Proceedings International Conference on Computer Vision*, volume 2, pages 1110–1116, 2003.