Closed-world tracking of multiple interacting targets for indoor-sports applications

Matej Kristan *, Janez Perš, Matej Perše, Stanislav Kovačič

Faculty of Electrical Engineering, University of Ljubljana, Tržaška 25, 1001 Ljubljana, Slovenia

Abstract

In this paper we present an efficient algorithm for tracking multiple players during indoor sports matches. A sports match can be considered as a semi-controlled environment for which a set of closed-world assumptions regarding the visual as well as the dynamical properties of the players and the court can be derived. These assumptions are then used in the context of particle filtering to arrive at a computationally fast, closed-world, multi-player tracker. The proposed tracker is based on multiple, single-player trackers, which are combined using a closed-world assumption about the interactions among players. With regard to the visual properties, the robustness of the tracker is achieved by deriving a novel sports-domain-specific likelihood function and employing a novel background-elimination scheme. The restrictions on the player's dynamics are enforced by employing a novel form of local smoothing. This smoothing renders the tracking more robust and reduces the computational complexity of the tracker. We evaluated the proposed closed-world, multi-player tracker on a challenging data set. In comparison with several similar trackers that did not utilize all of the closed-world assumptions, the proposed tracker produced better estimates of position and prediction as well as reducing the number of failures.

Key words: Tracking, Sport, Closed world, Particle filter, Multiple targets, Color histograms, Dynamic models, Local smoothing, Voronoi partitioning

* Corresponding author.
Tel: ++ 386 1 4768 876
Fax: ++ 386 1 4768 279
Email address: matej.kristan@fe.uni-lj.si
(Matej Kristan).

1 Introduction

Researchers have been studying the different aspects of human motion during athletic event for many years. Since the mid-1920s researchers like Hill [24] and later Keller [32] developed dynamic models of athletes that were used to predict the world records for linear races, such as the 100-meter sprint. As technology progressed, high-accuracy measurement devices were developed that allowed researchers to study the bio-mechanical properties of athlete's body [46]. While these devices were able to capture local features of the human body, i.e., the positions of the extremities, they were not appropriate for studying the athlete's performance on a larger scale. In particular, sports experts were interested in obtaining the positions of the players during a sports match. This would enable them to perform a tactical analysis of particular players or whole teams. Thus the early research in this field involved recording a sports match with a video camera and required many hours of tedious manual input to obtain only a small number of players' positions. However, even these crude approaches allowed researchers like Erdmann [18], and Ali and Farrally [3] to obtain valuable information about the load on soccer players during a match. The further development of specialized systems that could semi-automatically track the players throughout an entire match allowed sports experts to make a more in-depth tactical analysis. For example, by using a system that was specifically designed to track players in a squash match [42], Vučkovič et al. [55] were able to determine the behavioral features that distinguished the loser from the winner solely on the basis of positional data. However, the ability to obtain the trajectories of the players during a sports match is not only interesting to sports theorists. Coaches can also use this information to individualize the physical training plans of their athletes and plan the offencive/defencive strategies to achieve the best performance. Thus, the cornerstone of the various analyses that are interesting to sports experts is the ability to track the players on the court. In this paper we describe a tracking system, general enough to be applied to a variety of indoor team sports, that would allow the tracking of all or only a selection of the players during a match.

1.1 Related work

Many researchers from the field of computer vision have studied the problem of detecting and tracking people using visual data. This has given rise to a body of literature, of which surveys can be found in the work of Aggarval and Cai [1], Gavrila [21] and Gabriel et al. [20]. Here we give only a limited review of the work relevant to the problem of tracking players in sports match.

The approaches to tracking players in sports usually involve observing the

court with a single or multiple cameras. In some applications the cameras are static [41, 42, 36, 31, 58, 29, 19, 51], while other applications [26, 50, 38, 12, 39, 14, 10, 35] use a single pan-tilt-zoom camera. Most of the approaches apply a preprocessing step to eliminate the background and thus reduce the background clutter [1] in the images. These approaches are usually followed by morphological operations to filter out the noise and extract the image regions that possibly represent the players. An early approach to background elimination, introduced by Intille and Bobick [26], was based on differencing consecutive images in order to extract the moving players from the court. However, since this approach was designed to detect the changes in a sequence of consecutive images, it was not appropriate for the cases when players were standing still. As an alternative, Seo et al. [50] proposed a histogram-based approach, which was later adopted by many authors [50, 38, 31, 12] for tracking during a soccer match. The weakness of this approach was that it assumed a color-homogeneity of the court. Thus, it was not appropriate for indoor sports where the court often has advertisements in various colors. For this reason, several authors have applied more elaborate statistical models of the court [42, 36, 58, 19, 51] in order to be able to extract the players.

The efficiency of tracking depends a great deal on the cues that are used to encode the visual properties of the players. The early approaches [26, 50] utilized color templates, which were extracted at the estimated position of the player in one frame and used to localize the same player in the next frame. To make the visual representations more robust, Perš and Kovačič [41] decoupled the shape information from the color. They encoded the player's shape by utilizing 14 binary Walsh-function-like kernels, while the player's color was encoded using his average color. To capitalize on the shape information, Needham [36] encoded the shapes of the soccer players using a set of five pre-learned multi-resolution kernels. A similar approach was adopted by Lu and Little [35], who used a pre-learned set of dense grids of histograms of oriented gradients [15] to track hockey players and recognize their actions. An interesting attempt to develop a generic detector for hockey players was made by Okuma et al. [39]. They used a cascade of weak classifiers that were trained on a large number of manually extracted raw color patches containing the players. Ok et al. [38] noted that the player can usually be described by two colors: the color of the shirt and the color of the shorts. Therefore, they divided each player into two separate regions and then encoded each region by the mean value of the color within that region. The class of visual representations that can be viewed as a generalization of this approach is the color histograms [52]. They have been successfully applied in many tracking applications in sports [39, 12, 43] as well

[1] The term *background clutter* is used throughout this paper to refer to the background pixels that do not belong to a particular player. For example, when the player's texture is very similar to the texture of the court, we say the background clutter is severe.

as in the more general applications of visual tracking [40, 37, 54, 13]. Recently, Birchfield and Rangarajan [7] proposed a class of color histograms that also integrates the spatial information of the target's color.

Measurements based on visual data are known to be inherently ambiguous. Therefore, some researchers [29, 31, 58, 50] enforce a spatial continuity of the players' positions on the court by using the Kalman filter [30]. However, the assumptions that the Kalman filter imposes on the measurement process and the target's dynamics are usually too unrealistic for visual tracking and so result in a degraded performance. For this reason, many authors [10, 39, 38, 12, 36, 10] employ particle filters [4] instead. While these methods result in a more robust tracking than when using the Kalman filter, they usually increase the computational complexity [31].

A non-trivial task when tracking multiple players is maintaining the correct identities of the players throughout the match. In the estimation theory, a classical approach to tracking multiple targets involves a detection step followed by the target-to-measurement association. In addition to the Nearest Neighbor (NN) filter, techniques such as the Joint Probabilistic Data Association Filter (JPDAF) are common solutions to the association problem [23]. The applications of sports tracking based on the NN and JPDAF approaches can be found in [58, 29, 10] and [31], respectively. Some earlier applications of the JPDAF in the context of computer vision can be found in [45, 49]. The weakness of these approaches is that they involve an explicit detection and exhaustive enumeration of the associations, which leads to an NP-hard problem. Some attempts to reduce the complexity of the association problem include gating [58, 29, 10] and treating the associations as random variables which can then be assigned via sampling [25].

Another way to tackle the problem of tracking multiple targets is to concatenate the states of all the targets into a single joint-state. This makes it possible to apply particle-filtering techniques developed for single-target tracking [40, 36]. By introducing an additional variable that indicated the number of targets to the joint-state, the authors of BraMBLe [28] were able to track a varying number of visually similar targets. This approach was adopted by Czyz et al. [14] to track the soccer players of the same team. The weakness of the joint-state particle filters is that a poor estimate of a single target may degrade the entire estimation. For this reason, the number of particles needs to be increased, which may render the tracker computationally inefficient for more than three or four targets [33]. Recently, some efficient schemes based on Markov Chain Monte Carlo approaches have been proposed [59, 33] to solve this problem. Vermaak et al. [54] formulated the problem of tracking visually identical targets as the problem of maintaining the multi-modality of the estimated posterior distribution of the target states. This approach was later applied by Okuma et al. [39] and Cai et al. [10] to track players in a hockey

match.

A simple solution when the number of targets is known is to track each target with a separate tracker. This approach reduces the size of the state-space and allows the tracking of a specific target without the need to track all of the other targets as well, thus reducing the computational complexity of the tracker. However, this approach is rather naive, since the target with the highest score will often *hijack* the trackers of the nearby targets [33]. Solutions based on the *histogram back-projection* technique [50], *occlusion alarm probability* principle [38] and *template-based* methods [12] were proposed in the literature to cope with the problem of hijacking.

Some alternative approaches to tracking multiple interacting players in sports involve detecting all the players on the court and forming a graph structure among the detections. The problem of tracking is then posed as the task of finding the optimal path through the graph [19, 43, 51]. These approaches, however, rely on the explicit detection of all the players on the court using background-subtraction techniques. Since background-subtraction techniques are usually based on strong assumptions about the color of the players and the court, these approaches are constrained to a narrow subset of applications.

1.2 Our approach

In this paper we address the problem of tracking multiple interacting targets in indoor sports such as basketball, European handball, and squash. We present a novel system for tracking players using a static bird's eye view. The sporting event is regarded as a semi-controlled environment for which certain properties are known. In turn, these properties are used to construct a computationally efficient and robust sports-domain-specific tracker capable of tracking multiple players. The a-priori knowledge of the semi-controlled environment is formulated in the context of Intille and Bobick's [26] *closed worlds* as a set of closed-world assumptions. In order to cope with the uncertainties in the visual data, the proposed tracker is based on a statistical framework of particle filters [4].

Our original contribution is four fold. The first contribution is the derivation of the visual likelihood function for the particle filter. This likelihood function is estimated using a large amount of real-world data and reflects the dynamic visual properties of the players during a sports match. The second contribution is the approach to the dynamic estimation of the threshold for the background elimination. This allows simple models of the background to be used. The third contribution is the local-smoothing approach, which helps when modelling the inertia of the players. This allows robust tracking even

with a moderate number of particles, thus reducing the computational complexity of the tracker. The final contribution of this paper is the concept of managing multiple targets by jointly inferring the closed worlds of all the players using Voronoi partitioning. This makes it possible to track each player with a separate tracker and even further reduces the computational complexity of the multi-player tracker. We demonstrate the robustness of the proposed, closed-world, multi-player tracker using experiments that track a single player and multiple players.

The remainder of the paper is organized as follows. In Section 2.1 we present a set of closed-world assumptions that apply to tracking a single player in indoor sports. Section 2.2 describes the engine of the tracker. Section 2.3 deals with the visual properties of the players and shows how these are used for the tracking. We discuss the player's dynamical properties in Section 2.4, where the local smoothing scheme is presented. In Section 2.5 we augment the set of closed-world assumptions with an additional assumption, and present the closed-world, multi-player tracking scheme. Experiments with the proposed tracker are discussed in Section 3 and, finally, in Section 4 some conclusions are drawn.

2 Theoretical background

2.1 Closed world

Treating a sporting event as a semi-controlled environment is a concept that was originally introduced by Intille and Bobick [26] under the name *closed world* (CW). The main premise of this concept is that for a given region in space and time a specific context is adequate to explain that region. In our case the context is the following set of CW assumptions:

- **(CW1)** *The camera overlooking the court is static and positioned such that its optical axis is approximately perpendicular to the floor (Fig. 1).*
- **(CW2)** *The court is bounded, and its model can be calculated.*
- **(CW3)** *The players' textures can vary during the game; however, they are known at the beginning.*
- **(CW4)** *A player cannot change his/her position completely arbitrarily due to the effects of inertia.*

To make the situation clear, the relative position of the camera and an image of the court are shown in Fig. 1.

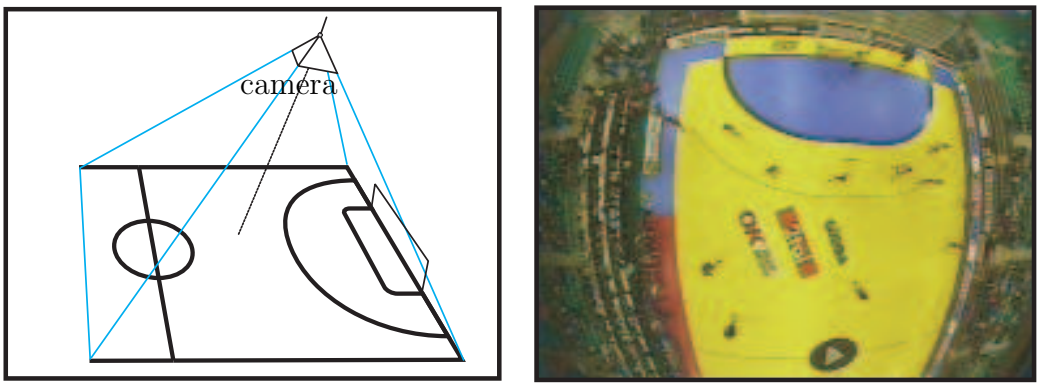


Fig. 1. The camera is placed above the court with its optical axis approximately perpendicular to the court (left). An image of a handball match obtained by the camera (right).

The four closed-world assumptions described above together with an additional assumption, which will be presented later, form the contextual basis on which a sports-domain-specific tracker will be presented in the following sections. First, however, we present the algorithmic basis of our tracker.

2.2 Particle filtering

In recent years, particle filters have been shown to provide an efficient means of visual tracking in various situations. Since their first appearance in the vision community [27] they have quickly increased their popularity by proving to provide a robust way of handling the uncertainties usually present in visual data. For these reasons we used a particle filter as the engine of our tracker. We provide here only the basic concept and notations, and refer the interested reader to [22, 27, 16, 4, 11] for more details.

Let $\mathbf{x}_{t-1}$ denote the state (e.g., the position and size) of a tracked object at time-step $t-1$, let $\mathbf{y}_{t-1}$ be an observation at $t-1$, and let $\mathbf{y}_{1:t-1}$ denote the set of all the observations up to $t-1$. From a Bayesian point of view, all of the interesting information about the target's state $\mathbf{x}_{t-1}$ is embodied by its posterior $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$. The aim of tracking is then to recursively estimate this posterior as new observations $\mathbf{y}_t$ arrive. This process is characterized by two steps: prediction (1) and update (2).

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \tag{1}$$

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \tag{2}$$

The recursion (1,2) for the posterior in its simplest form thus requires a specification of a dynamical model describing the state evolution $p(\mathbf{x}_t|\mathbf{x}_{t-1})$, and a

model that evaluates the likelihood of any state given the observation $p(\mathbf{y}_t|\mathbf{x}_t)$.

In our implementation we use the well-known CONDENSATION algorithm [27], which is, in essence, a boot-strap particle filter [22]. The posterior at the time-step $t-1$ is estimated by a finite Monte Carlo set of states $\mathbf{x}_{t-1}^{(i)}$ and their respective weights $w_{t-1}^{(i)}$

$$p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) \approx \{\mathbf{x}_{t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^{N}, \tag{3}$$

such that all the weights in the particle set sum to one. At time-step t, the particles are resampled according to their weights in order to obtain an unweighted representation of the posterior $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) \approx \{\tilde{\mathbf{x}}_{t-1}^{(i)}, \frac{1}{N}\}_{i=1}^{N}$. Next, they are propagated according to the dynamical model $p(\mathbf{x}_t|\tilde{\mathbf{x}}_{t-1}^{(i)})$, to obtain a representation of the prediction $p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \approx \{\mathbf{x}_{t-1}^{(i)}, \frac{1}{N}\}_{i=1}^{N}$. Finally, a weight is assigned to each particle according to the likelihood function, $w_t^{(i)} \propto p(\mathbf{y}_t|\mathbf{x}_t^{(i)})$. All the weights are normalized to sum to one, and the posterior of the time-step t is approximated by a new weighted particle set $p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx \{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^{N}$. The current state $\hat{\mathbf{x}}_t$ of the target can then be estimated as the minimum mean-square error (MMSE) estimate over the posterior $p(\mathbf{x}_t|\mathbf{y}_{1:t})$

$$\hat{\mathbf{x}}_t = \sum_{i=1}^{N} \mathbf{x}_t^{(i)} w_t^{(i)}. \tag{4}$$

2.3 Modelling the visual properties

Our goal is to obtain the positions of the players during a sports match solely by observing their visual properties. In this section we discuss how these visual properties are modelled and used for the tracking in our application.

2.3.1 Color histograms

The part of the player that is almost always seen by the camera is the player's torso, which is fairly elliptical; for example, see Fig. 2. For this reason we model each player with an elliptical region and encode the player's color properties with the color histogram [37] sampled inside that region. The state of the tracked player $\mathbf{x}_t$ is parameterized by an ellipse $\mathbf{x}_t = (x_t, y_t, a_t, b_t)$ with its center at (x_t, y_t), and with parameters a_t and b_t denoting the width and the height, respectively.

When constructing the color histogram it is beneficial to assign higher weights to the pixels that are closer to the center of the ellipse and lower weights to those farther away. This can help achieve some robustness, since the pixels that

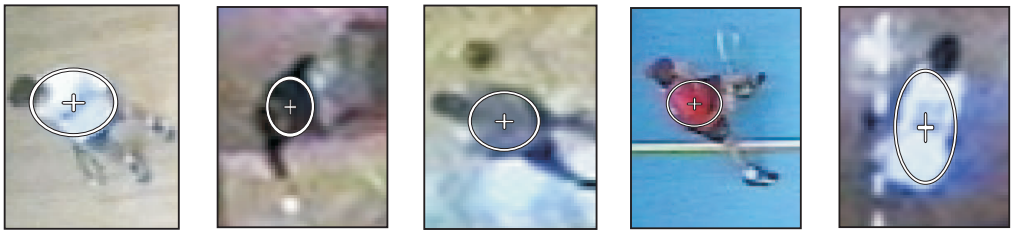


Fig. 2. The images show the players from different sports on the court. The dominant part of the player that is seen by the camera is the player's torso, which can be approximated with an ellipse (white).

are closer to the center are less likely to be affected by the clutter stemming from the background pixels or nearby players. Furthermore, if some a-priori knowledge of which pixels are not likely to belong to the player is available, it should be used in the construction of the histogram, i.e., those pixels should be ignored. An elegant way of discarding the pixels that do not belong to the player is to use a mask function, which assigns a prior weight of zero to those pixels that are not likely to have been generated by the player and a weight of one to all the other pixels.

Let $E = (x, y, a, b)$ be an elliptical region at some state $\mathbf{x}_t = (x, y, a, b)$. The RGB color histogram with $B = 8 \times 8 \times 8$ bins $\mathbf{h_x} = \{h_i\}_{i=1}^{B}$, sampled within the ellipse E, is then defined as

$$h_i = f_h \sum_{\mathbf{u} \in E} K(\mathbf{u}) M(\mathbf{u}) \delta_i(b(\mathbf{u})), \tag{5}$$

where $\mathbf{u} = (x, y)$ denotes a pixel within the elliptical region E. $\delta_i(\cdot)$ is the Kronecker delta function positioned at histogram bin i, and $b(\mathbf{u}) \in \{1...B\}$ denotes the histogram bin index associated with the color of a pixel at location $\mathbf{u}$. $K(\cdot)$ is an Epanechnikov weighting kernel, as in [13, 37], positioned at the center of the ellipse, $M(\mathbf{u})$ is the a-priori binary mask function, and f_h is a normalizing constant such that $\sum_{i=1}^{B} h_i = 1$. The mask function $M(\mathbf{u})$ can in general be composed of several mask functions. Indeed, in our application we define $M(\mathbf{u})$ as an intersection of two mask functions $M_D(\mathbf{u})$ and $M_V(\mathbf{u})$, which will be described in the following sections.

According to CW2, the background image can be modelled, thus we define the measure of the presence that evaluates whether a player with a predefined reference histogram $\mathbf{h}_t$ is present at some state $\mathbf{x}_t$ as

$$D(\mathbf{h}_A, \mathbf{h}_t; \mathbf{h}_B) = \beta^{-1} \rho(\mathbf{h}_A, \mathbf{h}_t; \mathbf{h}_B), \tag{6}$$

where $\mathbf{h}_A$ and $\mathbf{h}_B$ are histograms sampled at the state $\mathbf{x}_t$ on the current and the precalculated background image, respectively. β is the portion of the pixels within the elliptical region of $\mathbf{x}_t$ that are assigned to the foreground by the

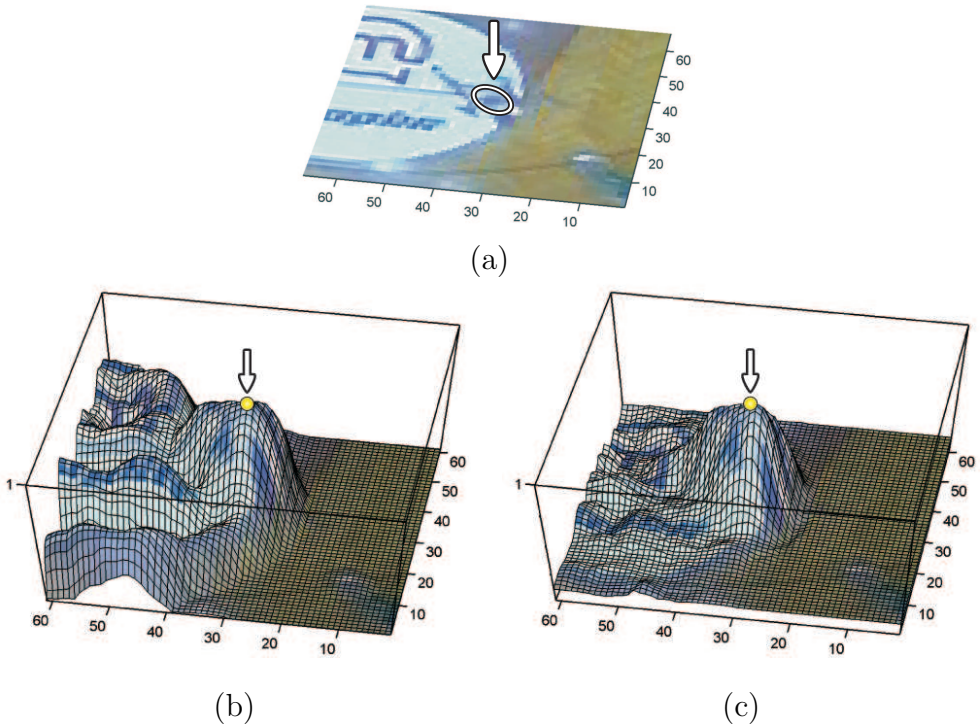


Fig. 3. The histogram of a basketball player was sampled within the ellipse (a). A non-normalized distance $\varrho(\mathbf{h}_A, \mathbf{h}_t)$ is shown in (b) with respect to different positions. The result for the proposed normalized-distance measure $\rho(\mathbf{h}_A, \mathbf{h}_t; \mathbf{h}_B)$ is shown in (c). For better visualization, one minus the distance measures are shown. The correct position of the player is depicted by a white arrow and a circle in each image. Notice how the mode corresponding to the selected player is more pronounced with respect to the background clutter when the normalized distance is used.

mask function $M(\mathbf{u})$. $\rho(\mathbf{h}_A, \mathbf{h}_t; \mathbf{h}_B)$ is the normalized distance between $\mathbf{h}_A$ and $\mathbf{h}_t$, given the background histogram $\mathbf{h}_B$, defined as

$$\rho(\mathbf{h}_A, \mathbf{h}_t; \mathbf{h}_B) = \frac{\varrho(\mathbf{h}_A, \mathbf{h}_t)}{\sqrt{\varrho(\mathbf{h}_B, \mathbf{h}_t)^2 + \varrho(\mathbf{h}_A, \mathbf{h}_t)^2}}, \tag{7}$$

where $\varrho(\mathbf{h}_1, \mathbf{h}_2) = 1 - \sum_i \sqrt{h_{1i} h_{2i}}$ is the Hellinger distance [48].

Note that the normalization term in (7) incorporates the distance between the reference color model and the background color. Such a normalization aids tracking when the player's color is similar to that of the background. In these situations the measure (7) favors those regions for which the reference-color model is closer to the color in the current image than to the background color. This effectively attenuates the background clutter and forces particles closer to the target. An example of the normalized and non-normalized distance measure is shown in Fig. 3.

2.3.2 The likelihood function

In order to carry out the update step (2) of the particle filter, the probability density function (pdf) of the presence measure (6) needs to be known. Due to the lack of a rigorous physical background with which the analytical form of this pdf could be derived, an experimental approach was chosen instead.

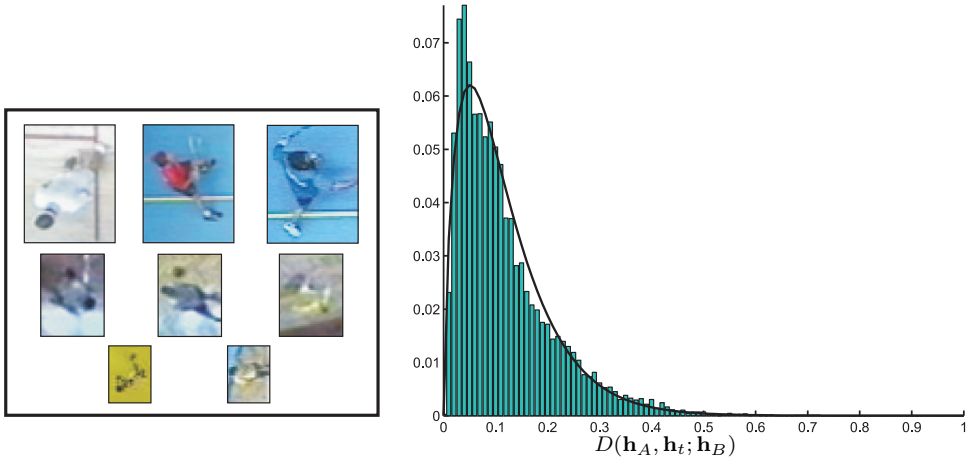
In order to estimate the pdf of (6), the values of the presence measures typical for players during a match were recorded. According to Bon et al. [8, 9] approximately forty percent of a player's motion in sports like handball, basketball and soccer can be classified as running and the other sixty, as walking or standing. For this reason we selected several sequences containing players from indoor sports; see Fig. 4a for examples. These players were tracked using a simple tracker from the literature [37] and by manual marking. This enabled us to obtain approximately 115,000 values of the measure (6). Forty percent of the recorded values corresponded to the fast-moving players, while sixty percent corresponded to the players that were moving slowly or standing still. These values are visualized using the histogram in Fig. 4b.

To identify the best model for the gathered data, a model selection was carried out using the Akaike information criterion (AIC) [2] among four models of the probability density functions: exponential, gamma, inverse gamma and zero-mean gaussian. The test with the AIC showed that the gamma function explained the data significantly better than the other functions. For this reason the probability density function of measure (6) was chosen in the form of

$$p(\mathbf{y}_t|\mathbf{x}_t) \propto D(\mathbf{h}_A, \mathbf{h}_t; \mathbf{h}_B)^{\gamma_1 - 1} e^{\frac{-D(\mathbf{h}_A, \mathbf{h}_t; \mathbf{h}_B)}{\gamma_2}}. \tag{8}$$

The parameters γ_1 and γ_2 were estimated from the data using the maximum-likelihood approach. The estimated values were $\gamma_1 = 1.769$ and $\gamma_2 = 0.066$.

Note that the gamma distribution assigns small probability values to those values of the measure (6) that are very close to zero. At first glance this may not seem reasonable for the purposes of object localization; however, if we observe a running player in two consecutive time-steps, it is more likely that the player's appearance will change within these two time-steps than stay the same. This is an inherent property of the player's *visual* dynamics that follows directly from assumption CW3 and is thus implicitly captured by the likelihood function (8).



(a) (b)

Fig. 4. The left-hand image shows some of the players that were used to estimate the empirical probability density function of measure (6). The right-hand image shows this function in the form of a histogram, and overlaid is the maximum-likelihood fitted gamma probability distribution model.

2.3.3 The background mask function

While color histograms are powerful color cues for tracking textured objects, they can fail when the object is moving on a similarly textured background. This is usually due to their inability to capture the spatial relations in the texture, and the fact that they are always sub-sampled in order to increase their robustness. There is, however, still some useful information left in the current and the background image – the difference between the two. By thresholding this difference image with some appropriate threshold κ_t, we can construct a mask image, that filters out those pixels that are likely to belong to the background. Since, in general, the illumination of the court is non-uniform in space and time, and since the visual properties among the players vary, the threshold has to be estimated dynamically for each player.

Let $\hat{\mathbf{x}}_t$ denote the estimated state of a player at time-step t. Let $\mathbf{h}_A$ and $\mathbf{h}_B$ be the histograms sampled at that state on the current image $A(\cdot)$ and the background image $B(\cdot)$, respectively. If the similarity between the player's visual model and the background is large enough ($\varrho(\mathbf{h}_A, \mathbf{h}_B) < \varrho_{thresh}$), then the mask in the next time-step is generated. To approximate the threshold κ_{t+1} for the next time-step, first a mask is calculated by thresholding the difference between the current and the background image

$$M_D(\mathbf{u}) = \begin{cases} 1 \ ; \ \|A(\mathbf{u}) - B(\mathbf{u})\| \geq \kappa_{t+1} \\ 0 \ ; \qquad otherwise \end{cases} . \tag{9}$$

A threshold value that assigns some predefined percentage η_0 of the pixels within the ellipse of the current state to the background is chosen as the approximation of κ_{t+1}. If the mask is not generated in the next time-step ($\varrho(\mathbf{h}_A, \mathbf{h}_B) \geq \varrho_{thresh}$), then $M_D(\mathbf{u}) = 1$ in (9) for all pixels $\mathbf{u}$.

The parameters η_0 and ϱ_{thresh} were estimated empirically by manually selecting players on a heavily cluttered background. Their values were set to $\eta_0 = 25\%$ and $\varrho_{thresh} = 0.8$, respectively.

2.3.4 Adaptation of the visual model

As the player moves across the court, his/her texture varies due to the non-uniform lighting conditions, influences of the background, and variations of the player's pose (CW3). Therefore, the color model, i.e., the player's current reference histogram $\mathbf{h}_t$, has to be able to adapt to these changes. In addition, if the current state of the player is likely to have been falsely estimated, then the reference histogram should be updated by a very small amount, or not at all. Otherwise, it should be adapted by some larger amount.

Let $\hat{\mathbf{x}}_t$ be the estimated state of a player at the current time-step. The histograms $\mathbf{h}_A$ and $\mathbf{h}_B$ are sampled at that state on the current and the background image, respectively. The adaptation equation then follows a simple auto-regressive form

$$\mathbf{h}_t = \alpha_t \mathbf{h}_A + (1 - \alpha_t)\mathbf{h}_{t-1}, \tag{10}$$

where $\mathbf{h}_{t-1}$ is the reference histogram from the previous time-step. The intensity of the adaptation is defined with respect to the normalized distance between (7) $\mathbf{h}_A$ and $\mathbf{h}_{t-1}$ as

$$\alpha_t = \Omega_{max} \cdot (1.0 - \rho(\mathbf{h}_A, \mathbf{h}_{t-1}; \mathbf{h}_B)), \tag{11}$$

where Ω_{max} denotes the maximal adaptation. Again, this parameter was estimated by means of a controlled experiment, and was set to $\Omega_{max} = 0.05$.

2.4 Modelling the target dynamics

Most of the time during a match the player's aim is to act in an unpredictable fashion in order to confuse the opponent. This implies that the dynamics could be modelled by a random-walk model [47] as $p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_{t-1}, \Lambda_t)$, where $\mathcal{N}(\cdot;\cdot,\cdot)$ denotes the normal distribution with mean $\mathbf{x}_{t-1}$ and a diagonal

covariance matrix

$$\Lambda_t = diag(\sigma_{xy}^2, \sigma_{xy}^2, \sigma_{ab}^2, \sigma_{ab}^2). \tag{12}$$

Note that the variances in Λ_t determine the amount by which the player's state is expected to change between consecutive time-steps and depend on the size of the player in the current image. We account for this relation by writing

$$\sigma_{xy} = H_{t-1} \cdot \alpha_{xy} \ , \quad \sigma_{ab} = H_{t-1} \cdot \alpha_{ab}, \tag{13}$$

where $H_{t-1} = \sqrt{a_{t-1}^2 + b_{t-1}^2}$ is a measure of the size of the elipse of the player's previous state $\mathbf{x}_{t-1}$. The equations in (13) require some reasonable estimates for α_{xy} and α_{ab}, and we derive these next.

Based on the findings of Bon et al. [9] who refer to Kotzamanidis [34], Erdmann [18] and Bangsbo [5] regarding the dynamics of handball/soccer players, we estimated the highest velocity of a player as $v_{max} = 8.0$m/s. At a frame rate of 25frames/s we can say $v_{max} = 0.32$m/frame. During tracking, the player is usually determined by an ellipse that is approximately the size of his/her shoulders, which is estimated to be $H_t \approx 0.4$m. Assuming a Gaussian form of the velocity distribution, the highest velocity can be estimated as three standard deviations. Thus, $v_{max} = 3\sigma_{xy}$/frame, and the parameter for σ_{xy} in (13) is then $\alpha_{xy} = \frac{0.32}{3 \cdot 0.4} \doteq \frac{1}{4}$. This means that the expected change in the position of the ellipse's center within two consecutive time steps is approximately one fourth of the ellipse's size.

The changes in the shape of the player's ellipse occur mainly due to the player leaning forward or to one side. Thus we can assume that within two consecutive time-steps the size along each axis can change, at most, by 15 percent. Following a similar line of thought, the parameter for σ_{ab} can be estimated as $\alpha_{ab} = 0.05$.

The intention of a player might indeed be to move in such a way as to appear unpredictable to the opponent. However, the motion itself is not entirely unpredictable, since it is constrained by the player's *task* and *physical limitations*, which enforce a sort of inertia on the motion (CW4). We therefore define the state evolution model to be

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_{t-1} + \mathbf{d}_t, \Lambda_t), \tag{14}$$

where $\mathbf{d}_t$ is a time-step constant drift modelling the influence of the inertia. Note that the model (14) uses the noise Λ_t (12), which was derived with a random-walk assumption. In reality, the effective noise of (14) should be

smaller, since a part of the random-walk noise is absorbed in the drift $\mathbf{d}_t$. Nevertheless, we use (12), since it presents an upper bound on the true noise of (14). Later, in the experiments section, we will show that even when an overestimated noise is used, the proposed dynamic model performs favorably in comparison to a widely used dynamic model. Next, we describe the method for estimating the drift $\mathbf{d}_t$ in (14).

2.4.1 Local smoothing

In order to satisfactorily estimate the drift $\mathbf{d}_t$ at time-step t, a reliable estimation of the past few states is needed. Since we are using a particle filter to recursively estimate the posterior of the target in time, the variance of the estimated state will usually depend on the number of particles used and the strategy by which the particles are propagated. For example, in order to cope with the sudden changes in motion, the common strategy is to increase the variance of the noise in the dynamic model. This, however, results in many particles having low values which contribute very little to the final estimation of the current state. The logical solution then is to increase the number of particles and/or use a clever strategy to concentrate the particles in the regions with high probability. Such strategies might be the application of an auxiliary-variable particle filter [44] or perhaps the methods of local likelihood sampling [53], to name just two. Even though any one of the above methods is likely to result in efficient tracking, they all introduce an additional computational complexity, which slows down the tracking. We propose here an alternative approach, where during each time-step the current state estimated from the particle filter is smoothed according to a locally-in-time learned conservative dynamic model. This model assumes that the player is not likely to change his/her velocity abruptly.

Let $\mathbf{o}_{t-T:t-1} = \{\mathbf{o}_k\}_{k=t-T}^{t-1}$ denote a sequence of the T past smoothed states of the tracked target. Let $\pi_{t-T:t-1} = \{\pi_k\}_{k=t-T}^{t-1}$ denote the set of their weights and let $\mathbf{v}_k = (\mathbf{o}_k - \mathbf{o}_{k-1})$ denote the shift between two consecutive smoothed states. We define a discrete local shift distribution based on the past smoothed states as

$$p(\mathbf{v}|\mathbf{o}_{t-T:t-1}) = \sum_{k=t-T}^{t-1} \delta(\mathbf{v}_k - \mathbf{v}) G_k(t), \tag{15}$$

where $\delta(\cdot)$ is the dirac-delta function. The weights $G_k(t)$ are defined as

$$G_k(t) = c_0 \pi^{(k)} \pi^{(k-1)} e^{-\frac{1}{2}\frac{(k-t+1)^2}{\sigma_o^2}}. \tag{16}$$

The first term c_0 in the above equation is the normalizing constant ensuring

that $\sum_{k=t-T}^{t-1} G_k(t) = 1$. The second and third terms reflect the likelihood of the states $\mathbf{o}^{(k)}$ and $\mathbf{o}^{(k-1)}$ used to calculate the shift $\mathbf{v}_k$, and the last term is a Gaussian that assigns higher a-priori weights to the more recent shifts. Note that, the Gaussian form was used for the last term exclusively to attenuate the importance of the older shifts in the distribution (15). In general, however, other forms that exhibit similar behavior (e.g., an exponential function) could have been used.

The current drift $\mathbf{d}_t$ is then estimated as the expected value over the local shift distribution

$$\mathbf{d}_t = \langle \mathbf{v} \rangle_{p(\mathbf{v}|\mathbf{o}_{t-T:t-1})}, \tag{17}$$

where $\langle \cdot \rangle_{p(\mathbf{v}|\mathbf{o}_{t-T:t-1})}$ denotes the expectation operator over $p(\mathbf{v}|\mathbf{o}_{t-T:t-1})$.

The number of the smoothed states used in (15) is set to $T = 3\sigma_o$ for practical applications, since the a-priori weights of all the older states are negligible. Assuming that a player cannot radically change his/her velocity within one half of a second, a value for the parameter σ_o is chosen to comply with this time frame. Since all our test sequences were recorded at a frame rate of 25 frames per second, we have chosen this parameter to be $\sigma_o = 4.3$. Thus in our application only $T = 13$ past smoothed states are considered.

The smoothed state is calculated as follows. At time-step t, the approximation to the distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ becomes available from the particle filter, and a MMSE estimate (4) $\hat{\mathbf{x}}_t$ of the state is calculated. This estimate is then fused with the prediction of the smoothed states $\tilde{\mathbf{o}}_t = \mathbf{o}_{t-1} + \mathbf{d}_t$ according to their likelihoods $w_{\hat{\mathbf{x}}_t} = p(\mathbf{y}_t|\hat{\mathbf{x}}_t)$ and $w_{\tilde{\mathbf{o}}_t} = p(\mathbf{y}_t|\tilde{\mathbf{o}}_t)$, respectively, as

$$\mathbf{o}_t = \frac{\tilde{\mathbf{o}}_t \cdot w_{\tilde{\mathbf{o}}_t} + \hat{\mathbf{x}}_t \cdot w_{\hat{\mathbf{x}}_t}}{w_{\tilde{\mathbf{o}}_t} + w_{\hat{\mathbf{x}}_t}}. \tag{18}$$

Finally, the corresponding weight of the new smoothed state $\mathbf{o}_t$ is evaluated using the likelihood function $\pi_t = p(\mathbf{y}_t|\mathbf{o}_t)$.

The evolution of the local shift distribution with respect to the player's motion is illustrated in Fig. 5. The first image (Fig. 5a) shows a squash player standing still. The samples of the local shift distribution spread around his center, indicating no preferable direction. As the player begins to move towards the center of the court (Fig. 5b), the samples gather around the direction of travel. In Fig. 5c, when the player suddenly stops, the samples spread around his center and as he begins to move towards the upper right-hand corner (Fig. 5d), the samples again gather in that direction.

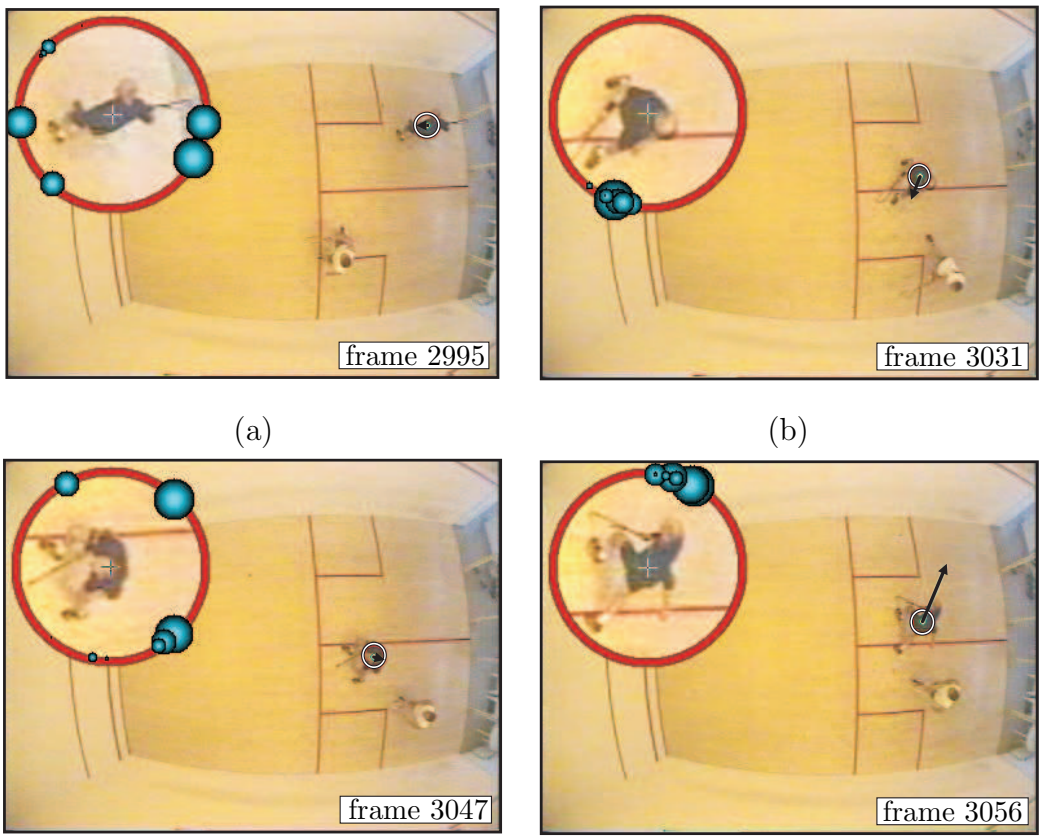


Fig. 5. Figures (a-d) show a tracked player during a squash match. The current smoothed state is depicted by an ellipse superimposed on the player and the arrow indicates the current drift ($\mathbf{d}_t$). The local shift distribution is represented by the dots on the circle of the player's enlarged image. The size of each dot represents the corresponding weight $G_k(t)$. For a better visualization, only the angular shift distribution is shown here, i.e., the $p(\mathbf{v}|\mathbf{o}_{t-T:t-1})$ integrated over the radius.

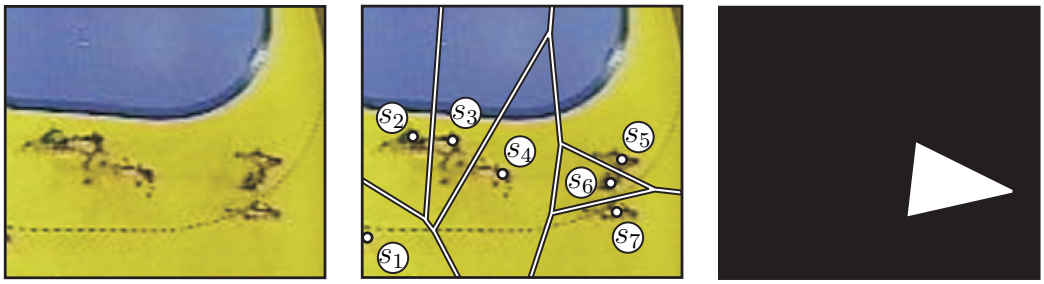
2.5 Closed-world multi-player tracking scheme

The closed-world assumption about the camera position (CW1) postulates that the players are viewed from above. Since one player cannot be located on top of another during a regular match, we can assume that it is unlikely to observe a complete occlusion at any time. This then forms a new closed-world assumption:

- **(CW5)** *At a given time-step, two players cannot occupy the same position.*

For now, let us assume that at a given time-step the true positions $\{{}^{(j)}\mathbf{s}\}_{j=1}^{N_p}$ of all N_p players are known. From the assumptions CW1 and CW5 it follows that the entire image can be partitioned into non-overlapping regions, such that each region contains only one player. One way to achieve such a partitioning

is to construct a Voronoi diagram [17] which is completely defined by the set of points/seeds $\mathbf{S} = \{^{(j)}\mathbf{s}\}_{j=1}^{N_p}$. The Voronoi diagram generates a set of N_p pairwise-disjoint convex partitions $\mathbf{V}_t = \{^{(j)}\mathbf{V}\}_{j=1}^{N_p}$, such that each partition contains exactly one seed. For every point in the particular partition the closest seed is then the one encapsulated by that partition. An example of the Voronoi diagram among $N_p = 7$ seeds corresponding to the positions of the seven players (Fig. 6a) is shown in Fig. 6b.



(a) (b) (c)

Fig. 6. Seven players of a handball match are shown on the court (a) and the centers of all the players with the corresponding Voronoi partitioning are shown in (b). The mask function $^{(6)}M_V(\mathbf{u})$ corresponding to the sixth partition $^{(6)}\mathbf{V}$ is shown in (c). The white depicts the area that is seen through the mask, while the black indicates the occluded area.

Let $\mathbf{X}_t$ denote the joint-state, i.e., the concatenation of the states of all the players $\mathbf{X}_t \stackrel{\Delta}{=} \{^{(j)}\mathbf{x}_t\}_{j=1}^{N_p}$. The aim of tracking multiple players is then to estimate the joint-state posterior $p(\mathbf{X}_t|\mathbf{y}_{1:t})$ through time. If the current partitioning $\mathbf{V}_t$ is known, then the players' states become conditionally independent, given the partitioning. Thus the posterior conditioned on $\mathbf{V}_t$ factors across all the players as

$$p(\mathbf{X}_t|\mathbf{y}_{1:t}, \mathbf{V}_t) = \prod_{j=1}^{N_p} p(^{(j)}\mathbf{x}_t|\mathbf{y}_{1:t}, \mathbf{V}_t), \tag{19}$$

where $p(^{(j)}\mathbf{x}_t|\mathbf{y}_{1:t}, \mathbf{V}_t)$ is the posterior of the j-th player conditioned on the partitioning $\mathbf{V}_t$. This implies that once the partitioning is known, each player can be tracked by a single-player tracker confined to the corresponding partition. The restriction of the j-th tracker to its partition $^{(j)}\mathbf{V}$ can be easily achieved by using an additional mask function $^{(j)}M_V(\mathbf{u})$, defined as

$$^{(j)}M_V(\mathbf{u}) = \begin{cases} 1 \ ; & \mathbf{u} \in {}^{(j)}\mathbf{V} \\ 0 \ ; & otherwise \end{cases}, \tag{20}$$

where $\mathbf{u}$ is a pixel contained by the region ${}^{(j)}\mathbf{V}$. The mask function ${}^{(j)}M(\mathbf{u})$ in (5) for the j-th player is then defined as an intersection of the mask functions ${}^{(j)}M_D(\cdot)$ (9) and ${}^{(j)}M_V(\cdot)$

$$ {}^{(j)}M(\mathbf{u}) = {}^{(j)}M_D(\mathbf{u}) \cap {}^{(j)}M_V(\mathbf{u}). \tag{21} $$

The superscript ${}^{(j)}(\cdot)$ in (21) emphasizes that all the masks are player-dependent. An example of the mask function ${}^{(6)}M_V$ for the sixth player from Fig. 6b is shown in Fig. 6c.

In reality, prior to the tracking iteration, the true positions of the players are not known. The posterior of the joint-states thus involves an integration over all the possible Voronoi configurations

$$ p(\mathbf{X}_t|\mathbf{y}_{1:t}) = \int_{\mathbf{V}_t} p(\mathbf{V}_t|\mathbf{y}_{1:t}) \prod_{j=1}^{N_p} p({}^{(j)}\mathbf{x}_t|\mathbf{y}_{1:t}, \mathbf{V}_t). \tag{22} $$

This integral could in principle be approximated via a Monte Carlo integration; however, due to the complexity of the problem at hand, this may lead to a computational load that would be too large for practical applications. As an alternative, we propose a sub-optimal solution where prior to the tracking iteration the partitioning $\mathbf{V}_t$ is estimated and used to carry out the tracking recursions for each player independently and in a sequential manner:
Initially, the Voronoi partitioning is estimated via smoothed predictions

$$ {}^{(j)}\tilde{\mathbf{o}}_t = {}^{(j)}\mathbf{o}_{t-1} + {}^{(j)}\mathbf{d}_t, \tag{23} $$

where ${}^{(j)}\mathbf{o}_{t-1}$ and ${}^{(j)}\mathbf{d}_t$ are the smoothed estimate of the state (18) and the drift (17), respectively, of the j-th player from the previous time-step. We assume that the smoothed states with the larger weights π_{t-1} (Section 2.4.1) are more likely to have been properly estimated in the previous time-step than those with the smaller weights. Therefore, the player with the largest weight π_{t-1} is chosen and the single-player tracking iteration is carried out for that player using the initially estimated Voronoi partitioning. The current smoothed state of the player is then calculated and used to update the Voronoi partitioning. Next, the player with the second-largest weight π_{t-1} is selected and the procedure is repeated until all the single-player trackers are processed. A summary of the proposed, closed-world, multi-player tracker is given in Fig. 7.

In principle, the sequential recursing through single-player trackers described above could be repeated a few times in order to arrive at a better estimation of the current partitioning $\mathbf{V}_t$. This would then lead to better estimates of

- Calculate the background image, e.g., pixel-wise by means of a median filter along the temporal axis.
- Initialize the tracker by selecting the players. (e.g. manually)
- For t = 1, 2, 3 ... do:

(1) Sort the players into descending order in terms of the weights π_{t-1} of their corresponding smoothed states $\mathbf{o}_{t-1}$.
(2) Initialize all the seeds with the predicted states (23): $\mathbf{S} = \{s_j\}_{j=1}^{N_p}; s_j \leftarrow {}^{(j)}\tilde{\mathbf{o}}_t$
(3) For $j = 1 : N_p$
  (a) Construct a set of Voronoi partitions $\mathbf{V}_t = \{{}^{(j)}\mathbf{V}\}_{j=1}^{N_p}$ using the set of current seeds $\mathbf{S}$.
  (b) Construct the Voronoi mask ${}^{(j)}M_V(\mathbf{u})$ via (20) and calculate the single-player mask function ${}^{(j)}M(\mathbf{u})$ from (21).
  (c) Run the conventional CONDENSATION iteration (Section 2.2) using the dynamic model from (14) with the drift ${}^{(j)}\mathbf{d}_t$.
  (d) Calculate the current smoothed state ${}^{(j)}\mathbf{o}_t$ (18) and the corresponding weight ${}^{(j)}\pi_t$.
  (e) Sample the histogram at ${}^{(j)}\mathbf{o}_t$ and adapt the model to that histogram as in Sect. 2.3.4.
  (f) If needed, estimate the threshold for the mask function ${}^{(j)}M_D(\mathbf{u})$ in the next time-step (Section 2.3.3).
  (g) Update the j-th Voronoi seed with the smoothed average state:$s_j \leftarrow {}^{(j)}\mathbf{o}_t$.
(4) End For j

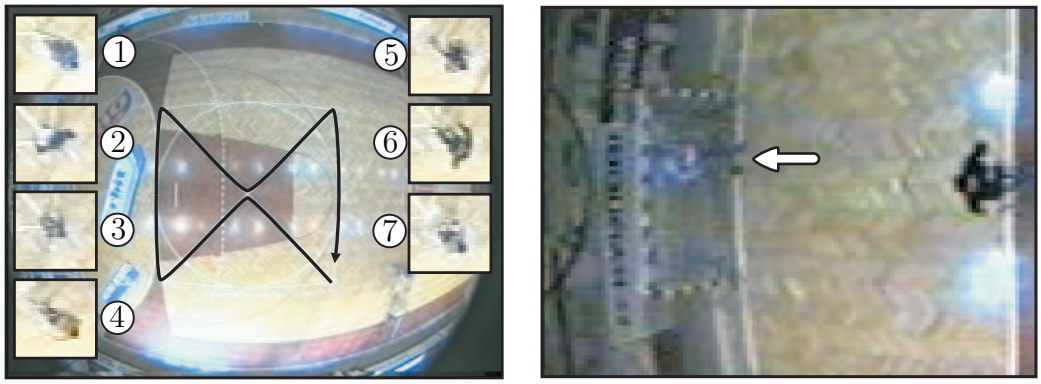
Fig. 7. Closed-world multi-player tracking algorithm.

the single-player posteriors. However, in our experience, a single iteration is sufficient to achieve satisfactory results.

3 Experimental study and results

Two sets of experiments were designed to evaluate two different aspects of the proposed multi-player tracker from Fig. 7. The first set of experiments, which is described in Section 3.1, involved tracking a single player. This set was used to evaluate the tracking capabilities of the proposed multi-player tracker in situations when the players do not interact.

The second set of experiments involved tracking multiple players, and was used to evaluate how the closed-world constraint CW5 influences the tracking capabilities of the proposed multi-player tracker when the players do interact. The latter set of experiments and their results are reported in Section 3.2. The videos demonstrating the results of the experiments are available online



(a) (b)

Fig. 8. The left-hand image shows seven players and the path used in the first experiment. The right-hand image shows a goal-keeper in front of the goal (indicated by an arrow) from the second experiment. The goalkeeper is barely distinguishable from the background due to the significant background clutter.

at *http://vision.fe.uni-lj.si/Research/trackp/articles/cviu06mk*.

3.1 Experiments with a single player

Two experiments were carried out involving the tracking of a single player. The first experiment was designed to quantify the effect of local smoothing (Section 2.4.1) in the proposed closed-world tracker from Fig. 7. We denote this tracker by $\mathbf{CW}_{\text{ls}}$. This experiment included seven players of different colors sprinting on a path drawn on the court (Fig. 8a) while performing sharp turns. The average visual size of each player was approximately 10×10 pixels. Each player was manually tracked thirty times and the average of the thirty trajectories obtained for each player was taken as the ground truth. In this way approximately 273 ground-truth positions $\mathbf{p}_t = (x_t, y_t)$ per player were obtained.

The $\mathbf{CW}_{\text{ls}}$ tracker was compared to a tracker that did not employ smoothing and where the adaptation and the background-subtraction steps (e) and (f) in Fig. 7 were carried out directly on the MMSE estimate (4). This tracker employed a discrete-time random-walk model for the ellipse size with a standard deviation of 5% of the current size, and a discrete-time nearly constant velocity (NCV) dynamic model [47] for the position. The variances of the NCV noise were learned from the ground-truth data and were set to $\sigma_{\dot{x}} = \sigma_{\dot{y}} = 1$pixel/frame. We denote this tracker by $\mathbf{CW}_{\text{ncv}}$. The ellipse width and height in both trackers were constrained to lie within the interval of [8,12] pixels.

All seven players from Fig. 8a were tracked $R = 30$ times with each tracker.

A standard one-sided hypothesis testing [6] was applied to determine whether $\mathbf{CW}_{\text{ls}}$ had superior performance to the $\mathbf{CW}_{\text{ncv}}$. The performance of the trackers in the r-th repetition was defined in terms of the root-mean-square (RMS) error as

$$C^{(r)} \triangleq \frac{1}{7}\sum_{k=1}^{7}\left(\frac{1}{T}\sum_{t=1}^{T}\|{}^{(k)}\mathbf{p}_t - {}^{(k)}\hat{\mathbf{p}}_t^{(r)}\|^2\right)^{\frac{1}{2}}. \tag{24}$$

In (24) ${}^{(k)}\mathbf{p}_t$ denotes the ground-truth position at time-step t for the k-th player, ${}^{(k)}\hat{\mathbf{p}}_t^{(r)}$ is the corresponding estimated position and $\|\cdot\|$ is the l_2 norm. At each repetition, a *sample performance difference*

$$\Delta^{(r)} = C_{ncv}^{(r)} - C_{ls}^{(r)} \tag{25}$$

was calculated. The terms $C_{ls}^{(r)}$ and $C_{ncv}^{(r)}$ were the cost values (24) of $\mathbf{CW}_{\text{ls}}$ and $\mathbf{CW}_{\text{ncv}}$, respectively.

In our case the null hypothesis H_0 was that $\mathbf{CW}_{\text{ls}}$ is *not* superior to $\mathbf{CW}_{\text{ncv}}$. For each tracker we calculated the sample performance difference mean $\bar{\Delta} = \frac{1}{R}\sum_{r=1}^{R}\Delta^{(r)}$ and its standard error $\sigma_{\bar{\Delta}} = \sqrt{\frac{1}{R^2}\sum_{r=1}^{R}(\Delta^{(r)} - \bar{\Delta})^2}$. The null hypothesis was then tested against an alternative hypothesis H_1, that $\mathbf{CW}_{\text{ls}}$ *is* superior to $\mathbf{CW}_{\text{ncv}}$, using the statistic $\frac{\bar{\Delta}}{\sigma_{\bar{\Delta}}}$. Usually, the alternative hypothesis is accepted at a significance level of α if $\frac{\bar{\Delta}}{\sigma_{\bar{\Delta}}} > \mu_\alpha$, where μ_α represents a point on the standard Gaussian distribution corresponding to the upper-tail probability of α. In our experiments we used $\alpha = 0.05$, which is common practice in hypothesis testing.

The results of the hypothesis testing on position and prediction with respect to a different number of particles in the particle filter are shown in Table 1. The second and third column in Table 1 show the test statistic $\frac{\bar{\Delta}}{\sigma_{\bar{\Delta}}}$. In all cases the test statistic is greater than $\mu_{0.05} = 1.645$. Thus we can accept the hypothesis that the tracker $\mathbf{CW}_{\text{ls}}$ is superior to $\mathbf{CW}_{\text{ncv}}$ when it comes to estimating the position and the prediction at the $\alpha = 0.05$ level. Note that when 100 particles were used the improvement of $\mathbf{CW}_{\text{ls}}$ over $\mathbf{CW}_{\text{ncv}}$ in estimating the prediction was large enough to be accepted at the $\alpha = 0.05$ level. However, it was not large enough to be accepted at the $\alpha = 0.025$ level ($\mu_{0.025} = 1.960$). In all the other cases the improvement of $\mathbf{CW}_{\text{ls}}$ over $\mathbf{CW}_{\text{ncv}}$ could have been accepted even at levels lower than $\alpha = 0.01$ ($\mu_{0.01} = 3.090$).

To further illustrate the performance of the trackers the RMS errors (24) were averaged over all thirty repetitions for each tracker. Figure 9 shows the results when the number of particles used in the particle filter is varied. Using only 25 particles, the $\mathbf{CW}_{\text{ls}}$ achieved similar average RMS errors for position (Fig. 9a) and prediction (Fig. 9b) as the $\mathbf{CW}_{\text{ncv}}$ with 75 particles and outperformed

Table 1
Results for the comparison of $\mathbf{CW}_{\text{ls}}$ and $\mathbf{CW}_{\text{ncv}}$ from 30 runs using the test statistic $\frac{\bar{\Delta}}{\sigma_{\bar{\Delta}}}$

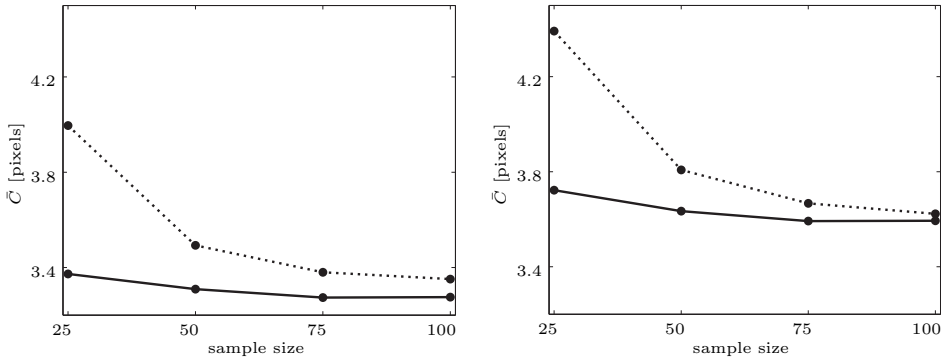
no. particles	Position ($\frac{\bar{\Delta}}{\sigma_{\bar{\Delta}}}$)	Prediction ($\frac{\bar{\Delta}}{\sigma_{\bar{\Delta}}}$)
25	17.134	16.907
50	9.416	8.337
75	6.991	4.836
100	5.081	1.955

the $\mathbf{CW}_{\text{ncv}}$ even when the number of particles in both trackers was increased. An important point to note here is that the $\mathbf{CW}_{\text{ls}}$ outperformed the $\mathbf{CW}_{\text{ncv}}$ even though the noise of the $\mathbf{CW}_{\text{ncv}}$ was estimated from the test data, while the noise used in the $\mathbf{CW}_{\text{ls}}$ was over estimated (see Sect. 2.4). This implies powerful generalization capabilities when using local smoothing in the closed-world tracking.

The aim of the second experiment was to evaluate how the proposed likelihood function and the dynamic background subtraction (Sect. 2.3) influence the tracking in the presence of background clutter. For this reason, tracker $\mathbf{CW}_{\text{ncv}}$ was compared to the simpler tracker $\mathbf{T}_{\text{ref}}$ from [37]. The trackers were compared in terms of the number of failures encountered while tracking a goalkeeper in a 733-images-long sequence of a handball match (Fig. 8b). The difference between $\mathbf{T}_{\text{ref}}$ and $\mathbf{CW}_{\text{ncv}}$ was that $\mathbf{T}_{\text{ref}}$ did not make use of the proposed likelihood function and the dynamic background subtraction. The number of particles used in this experiment was 25 and the parameters of the NCV dynamic models were set to the same values as in the previous experiment. The other parameters of $\mathbf{T}_{\text{ref}}$ were set as in [37]. The goalkeeper was tracked thirty times with each tracker. On average the tracker $\mathbf{CW}_{\text{ncv}}$ required only one user intervention per sequence, while the tracker $\mathbf{T}_{\text{ref}}$ required approximately 10 interventions to maintain a successful track throughout the sequence. We thus conclude that the proposed likelihood function and dynamic background subtraction increase the performance of the tracking in the presence of substantial background clutter.

3.2 Experiments with multiple players

To evaluate the effectiveness of the multi-player interaction scheme from Section 2.5, the proposed closed-world, multi-player tracker $\mathbf{CW}_{\text{ls}}$ from Fig. 7 was compared to the so-called *naive tracker*, which we denote by $\mathbf{CW}_{\text{naive}}$. The *naive tracker* was conceptually equal to $\mathbf{CW}_{\text{ls}}$, with the only difference being that the Voronoi mask functions ${}^{(j)}M_V$ (20) were always set to unity for all the players. Thus the tracker $\mathbf{CW}_{\text{naive}}$ was essentially a set of closed-world, single-



(a) (b)

Fig. 9. Average RMS errors (denoted by $\bar{C}$) of position (a) and prediction (b) for the $\mathbf{CW}_{\text{ncv}}$ (*dotted*) and $\mathbf{CW}_{\text{ls}}$ (*solid*) as a function of the number of particles.

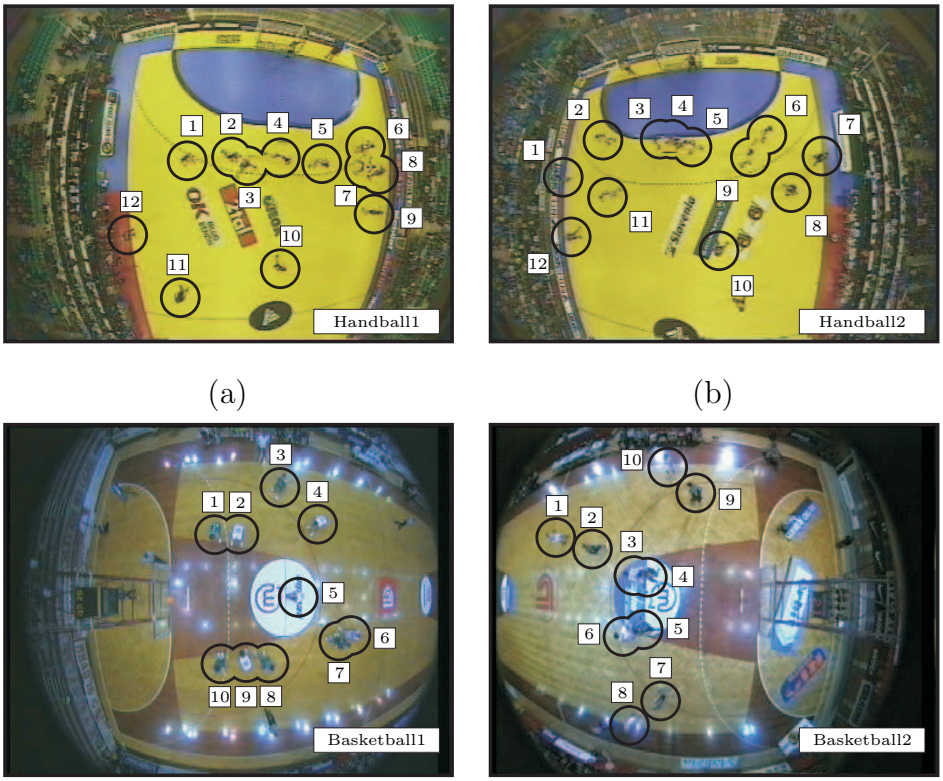
player trackers that did not interact according to the closed-world assumption CW5.

The trackers were compared on two recordings of a handball match and two recordings of a basketball match. Throughout the rest of this section we will refer to the handball and the basketball recordings as *Handball1*, *Handball2*, *Basketball1* and *Basketball2*. A typical image from each recording is shown in Fig. 10.

3.2.1 Description of the recordings

Two teams, each consisting of six players, were tracked in the recordings of the handball matches (Fig. 10a,b). The players of one team were wearing white shirts, and the players of the other team were wearing black shirts. The color of the court was mainly yellow and blue, with a few advertisement stickers on it. Because of the reflective properties of the material from which the court was made, and because of the side effects associated with using S-VHS tape for the video recording, the textures of the players varied significantly across different parts of the court. For example, white players appeared yellow on the yellow part of the court and blue on the blue part of the court. The textures of the black players were less affected by the color of the court.

In the experiments with the basketball matches, two teams, each consisting of five players, were tracked. In both recordings (Fig. 10c,d) the colors of the players were not influenced by the background as severely as they were in the recordings of the handball. Since all four recordings were originally recorded on an analog VHS recorder prior to digitization, they suffered from an effect called *color bleeding*. This resulted in bright colors spreading into the adjacent darker areas. For example, in Fig. 10a,b, the yellow patch of the court seems to be shifted to the right by a few pixels. Please see the online version of the



(a) (b)

(c) (d)

Fig. 10. Typical images from the four recordings used in the experiments for tracking multiple players. The first two images (a,b) show the twelve players of a handball match, while the second two images (c,d) show the ten players of a basketball match. All the players are depicted by a circle and a numeric label.

paper for the color images. Further information regarding the recordings is given in Table 2.

Table 2
Data for the recordings used in the experiments with multiple players

recording	frame rate [/s]	number of players	length [frames]	image size [pixels]
Handball1	25	12	950	348×288
Handball2	25	12	1257	348×288
Basketball1	25	10	740	368×288
Basketball2	25	10	305	368×288

3.2.2 Evaluation

The players were initialized manually and tracked throughout the entire recording. When a particular player was lost, the tracker was manually reinitialized

for that player and the tracking proceeded. The number of particles used for the tracking was 25 particles per player. In the recordings of the handball, the widths and heights of the players' ellipses were constrained to lie within the interval [6,8] pixels. In the case of the basketball recordings the interval [6,10] pixels was used. All the players were tracked five times with both trackers, and for each repetition the number of times the tracker failed was recorded. The results, averaged over the five repetitions, are shown in Table 3.

Table 3
The results for tracking multiple players with the proposed and the naive multi-player tracker

	average number of failures [/match]		failure rate per player [/min]	
recording	$\mathbf{CW}_{\text{naive}}$	$\mathbf{CW}_{\text{ls}}$	$\mathbf{CW}_{\text{naive}}$	$\mathbf{CW}_{\text{ls}}$
Handball1	34.0	4.0	4.47	0.53
Handball2	41.0	11.0	4.08	1.10
Basketball1	15.0	3.0	3.04	0.61
Basketball2	9.0	0.2	4.43	0.10

The naive tracker is denoted by $\mathbf{CW}_{\text{naive}}$, while the proposed multi-player tracker is denoted by $\mathbf{CW}_{\text{ls}}$. The second and the third columns show the average number of failures encountered by each tracker during the experiment. The last two columns show the same results recalculated to represent the number of times each tracker is expected to fail per player during one minute of tracking.

The second and the third column of Table 3 show the average number of failures encountered by the trackers $\mathbf{CW}_{\text{ls}}$ and $\mathbf{CW}_{\text{naive}}$ during the experiment. These columns show that in all cases the introduction of the last closed-world assumption (CW5) substantially reduced the number of failures and thus significantly improved the tracking. The results of the two columns could not be compared directly across different recordings because the recordings differed in their lengths as well as in the number of players. For this reason the results for each experiment were recalculated into failure rates per player and then normalized to a time-frame of one minute. These results are shown in the last two columns of Table 3. From the fourth column we see that the failure rates were approximately equal for all four experiments with the $\mathbf{CW}_{\text{naive}}$ tracker. While in comparison to $\mathbf{CW}_{\text{naive}}$ the proposed multi-player tracker $\mathbf{CW}_{\text{ls}}$ significantly reduced the failure rates, there were still some small residual failure rates present. These varied across the four recordings, as can be seen by comparing the results in the last column of Table 3. After a further inspection of the tracking results, we concluded that the residual failures could be assigned to one of the following four groups of errors:

(1) Some of the failures arose solely due to a heavily cluttered background,

and were not caused by interactions among the neighboring players. A substantial number of the failures in the recording *Handball2* could be attributed to the background clutter. In this match the player with the identification number 1 (Fig. 10b) could hardly be distinguished from the background. Some examples are shown in Fig. 11.

(2) Sometimes two similar players came very close to one another and were switched by the tracker despite the use of Voronoi partitioning. Such failures occurred only rarely, usually when just before the (near) collision the position and prediction of at least one of the colliding players was poorly estimated. An example of the switching of two black players in the recording *Basketball2* is shown in Fig. 12.

(3) Objects that were not tracked caused problems when they were in close proximity to the visually similar tracked players. One such situation consistently caused failures in the recording *Basketball1* when a white player was moving close to a white referee (see Fig. 13a). This was a typical problem of *tracker-hijacking*, discussed in Section 1.1. To demonstrate how such failures could be prevented, we have tracked the referee from Fig. 13 and the tracker was able to maintain a correct track of the white player; results are shown in Fig. 13b.

(4) Sometimes the collision of several players on a cluttered part of the court resulted in failures of the tracker. This was the case in the recording *Handball2*, where three players collided and crossed the goal-area line (Fig. 14). The situation was especially difficult because this was the place where the color of the court changed from yellow to blue. Because of the previously mentioned effect of color bleeding and the court's reflectance properties (Section 3.2.1), the players appeared to change their colors very quickly as they crossed the line. This introduced additional ambiguities and ultimately caused a failure.

Fig. 11. Figures show the handball player from Fig. 10b with the identification number 1, who is hardly distinguishable from the court due to the background clutter. The player is depicted by a white circle.

As we have pointed out in point (3) above, tracking may fail in situations where players move close to other visually similar objects that are not tracked. We have seen in Fig. 13 that in some cases these situations can be resolved within the proposed tracking framework by tracking those objects as well. We have thus repeated the experiment with the recording *Basketball1* where we

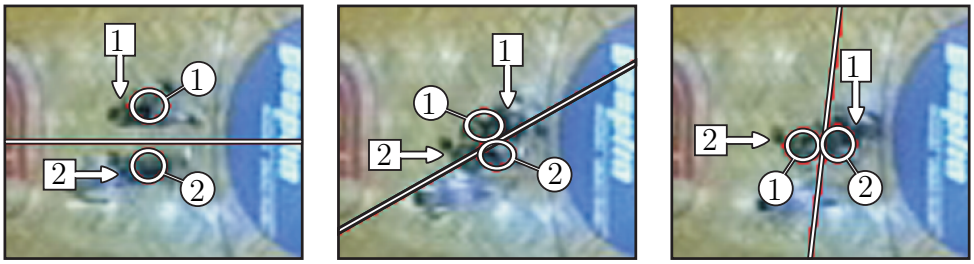


Fig. 12. Figures show two visually similar players in a basketball match during a near collision. The players' true identities are indicated by the numbers in the white squares. The tracker-estimated states and the corresponding identities are depicted by white ellipses and the Voronoi partitioning is indicated by a white line separating the players. Note that before the collision the markers with the same identification number denote the same players (left). Just before the players pass by one another, the state of the player with the identification number 2 is badly estimated (middle), and the tracker switches their identities (right).

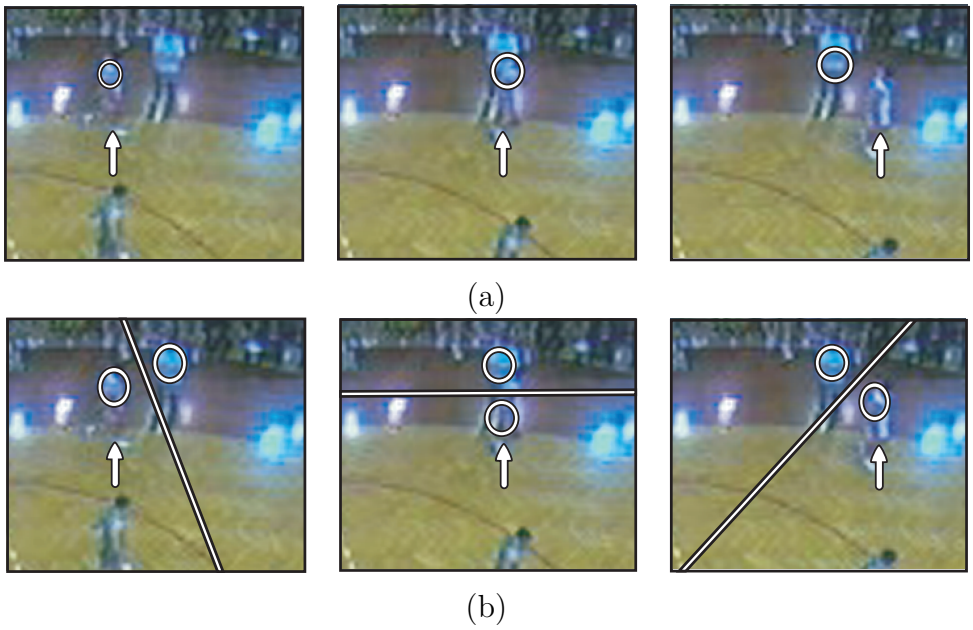


Fig. 13. Figures show a white player passing by a white referee. The upper row (a) shows results when only the player is tracked. The location of the player is depicted by an arrow, while the estimated state is depicted by the ellipse. As the player passes by the referee (middle) the tracker is *hijacked* by the referee and the tracking fails (right). The bottom row (b) shows results when both, the player and the referee, are tracked and tracking does not fail. The white line between the players depicts the Voronoi partitioning.

have also tracked the referee that was responsible for the failure described in Fig. 13. The results for $\mathbf{CW}_{\text{ls}}$ have improved by reducing the number of failures per experiment by one failure. In our experience, however, sport experts are usually interested in the teams, or a selection of players, rather than *everyone* on the court. We have observed that sports experts usually track a selection

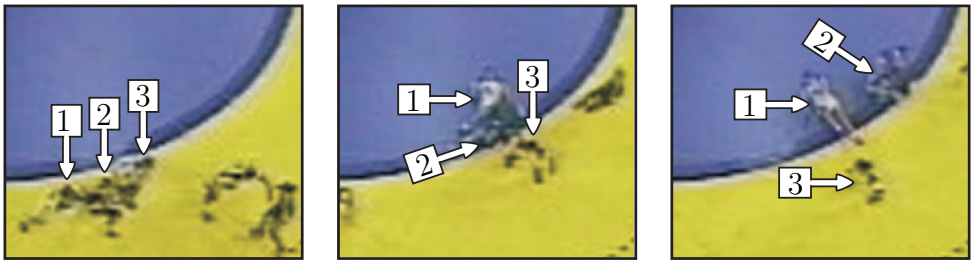


Fig. 14. Three players from the recording *Handball2* are shown as they collide along the goal-area line. Each player is depicted by a numeric label (1,2 and 3) and an arrow (left). The players change their color as they cross from the yellow part of the court to the blue part (middle and right). Please see the online version of the paper for the color images.

of players at a time. The reason is that switching two players, or improper tracking of a single player, could have a devastating effect on the subsequent analysis that sports experts perform. Therefore, situations where a player is tracked and the referee (or even another player) is not are common in practice. In those situations, failures like the one described in Fig. 13a can be expected.

In general, the proposed, closed-world, multi-player tracker $\mathbf{CW}_{\text{ls}}$ exhibited a robust performance and maintained a successful track even through the persistent collisions of several visually similar players. A sequence of three images from the recording *Handball1* (Fig. 15) shows an example, where several players collide and remain in collision. The tracker $\mathbf{CW}_{\text{ls}}$ successfully tracks all the players throughout the collision while maintaining their identities.

The trackers used in the experiments were implemented in C++ and tested on an Intel Pentium 4 personal computer with a 2.6-GHz CPU. A one time-step iteration for tracking a single player took approximately 7 ms of processing time. Since the bottleneck of the algorithm is the construction of the histograms, the processing time increases with the player's size. When tracking multiple players with $\mathbf{CW}_{\text{ls}}$, the processing time was proportional to the number of players plus the time required to construct the Voronoi regions. For example, a single iteration to track the twelve players of the handball match with $\mathbf{CW}_{\text{naive}}$ took approximately 86 ms, while $\mathbf{CW}_{\text{ls}}$ took approximately 108 ms. This means that approximately 22 ms was spent on the construction of the Voronoi regions and the corresponding mask functions.

4 Discussion and conclusion

A computationally efficient algorithm for tracking multiple players in indoor sports was presented in this paper. The effectiveness of the algorithm was achieved by considering a sporting event as a semi-controlled environment for

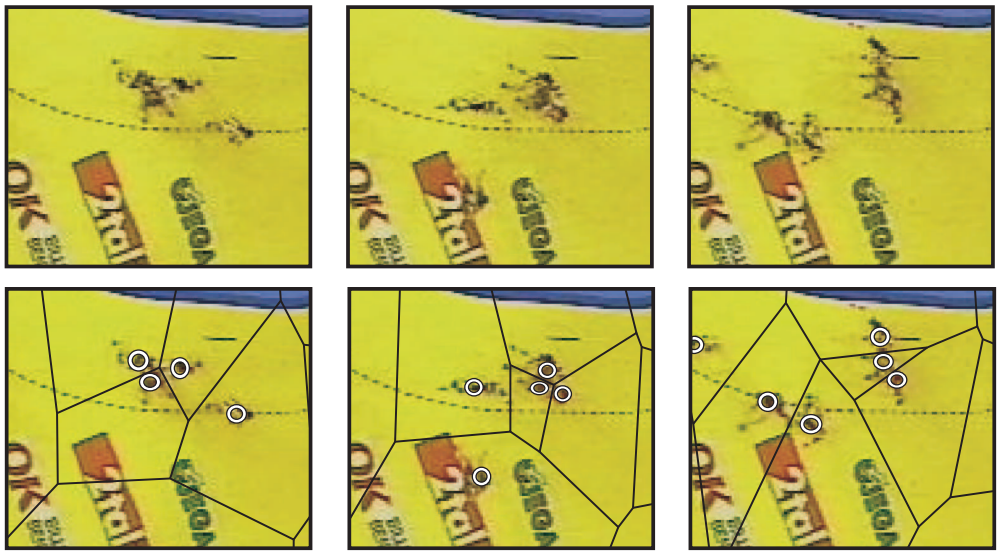


Fig. 15. The top row shows consecutive frames with numbers 681, 699 and 713 from the recording *Handball1*, where multiple players clashed or moved close to each other. The tracking result is shown in the bottom row where the Voronoi partitioning is drawn with black lines and the players are depicted by the ellipses.

which certain closed-world assumptions were derived. These assumptions considered the camera placement, the visual as well as the dynamic properties of the players, and the interactions among the players. The parameters concerning the visual properties of the players were estimated using recordings of real-life sports matches. On the other hand, the parameters of a player's dynamic model were estimated by consulting the sports literature. The restrictions imposed by the camera placement and the nature of the interactions among the players allowed the proposed closed-world multi-player tracker to be formulated as a set of single-player closed-world trackers. These were combined by jointly inferring a set of partitions, such that each partition contained only one player. This made it possible to track each player using a separate single-player tracker.

Extensive tests showed that the closed-world assumptions significantly increased the tracking performance. This was observed as a reduction of the failure rate and an increase in the accuracy of the position and the prediction. Even when a moderate number of particles was used in the particle filter (i.e. 25 particles), the proposed tracker maintained a satisfactory track. The results of the experiments with multiple players showed that the tracking performance increased in terms of a lower failure rate when the closed-world assumption about the interactions among the players was used.

Our original contribution is four fold. The first contribution is a derivation of the visual likelihood function for the particle filter, which reflects the dynamic visual properties of the players during a sports match. The second contribution is the approach of the dynamic estimation of the threshold for

the background elimination, which allows simple models to be used for the background. The third contribution is the approach of local smoothing, which helps to model the inertia of the players. This allows robust tracking even with a moderate number of particles, thus reducing the computational complexity of the tracker. The final contribution of this paper is the concept of managing multiple targets by jointly inferring the closed-worlds of all the players. This in effect allows the tracking of each player with a separate tracker and further reduces the computational complexity of the multi-player tracker.

Since the proposed tracker is based on a simple particle filter, i.e. the CONDENSATION algorithm, it is expected to improve in performance if a more efficient particle filter is considered. The tracking could also be made more robust by using a more sophisticated method for maintaining the background model. Currently, a single dynamic model is used to describe the player's motion. Alternatively, a mixture of models [57] describing different modes of behavior could be used and perhaps more sports-specific dynamic models could be considered. For example, a stochastic variant of the model developed by Keller [32] could be applied. While these extensions are likely to increase the performance, some of them may also significantly increase the computational complexity of the tracking. This might render the tracker inappropriate for practical applications. With time, however, these drawbacks could be compensated for by advances in the computational power of modern computers.

The multi-player tracker presented in this paper relies solely on one camera mounted above the court. Such a camera position is geometrically nearly optimal for estimating the position of a player on the court. However, it sometimes does not provide enough visual data to reliably track the player throughout the entire match. For example, when the player moves over a texturally very similar part of the court, the measurements might become ambiguous, and the tracking can fail. One way to cope with this problem would be to incorporate a separate detection scheme to automatically re-initialize the tracker once the player has been lost. Another, conceptually different, solution would be to introduce additional side cameras to increase the amount of visual information.

Currently, color histograms are used to encode the visual properties of the players. Alternatively, other appearance models, such as the recently introduced SMOGs [56], could be applied. These might increase the tracker's capability to discriminate the player from the background. However, replacing color histograms with another appearance model would require re-estimating the visual likelihood function. We expect these topics will be the focus of further research.

References

[1] J. K. Aggarwal and Q. Cai. Human motion analysis: A review. *Comp. Vis. Image Understanding*, 73(3):428–440, 1999.

[2] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automatic Control*, 19:716–723, December 1974.

[3] A. Ali and M. Farrally. A computer-video aided time motion analysis technique for match analysis. *J. Sports Medicine and Physical Fitness*, 31:82–88, 1991.

[4] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking. *IEEE Trans. Signal Proc.*, 50(2):174–188, February 2002.

[5] J. Bangsbo. The physiology of soccer: With special reference to intense intermittent exercise. *Acta Physiologica Scandinavica*, 619:1–155, 1994.

[6] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*, chapter 11, pages 438–440. John Wiley & Sons, Inc., 2001.

[7] S. T. Birchfield and S. Rangarajan. Spatiograms versus histograms for region-based tracking. In *Proc. Conf. Comp. Vis. Pattern Recognition*, volume 2, pages 1158–1163, 2005.

[8] M. Bon, J. Perš, and S. Kovačič. Computer vision system for tracking players during the handball match. In *Perspectives and profiles. Ann. Con. European College of Sport Science*, page 549, 2001.

[9] M. Bon, J. Perš, M. Šibila, and S. Kovačič. *Analiza gibanja igralca med tekmo*. Faculty of Sport, University of Ljubljana, 2001.

[10] Y. Cai, N. de Feritas, and J. J. Little. Robust visual tracking for multiple targets. In *Proc. European Conf. Computer Vision*, volume IV, pages 107–118, 2006.

[11] Z. Chen. Bayesian filtering: From kalman filters to particle filters, and beyond. Technical report, McMaster University, Hamilton, Ontario, Canada, 2003.

[12] K. Choi, Y.D. Seo, and S.W. Lee. Probabilistic tracking of soccer players and ball. *Proc. Asian Conf. Computer Vision*, 1:27–30, January 2004.

[13] V.D. Comaniciu and P. Meer. Kernel-based object tracking. *IEEE Trans. Patter. Anal. Mach. Intell.*, 25(5):564–575, May 2003.

[14] J. Czyz, B. Ristic, and B. Macq. A color-based particle filter for joint detection and tracking of multiple objects. In *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, March 2005.

[15] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. Conf. Comp. Vis. Pattern Recognition*, volume 1, pages 886–893, June 2005.

[16] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, January 2001.

[17] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*, chapter Nonparametric Techniques, pages 177–178. Wiley-Interscience Publica-

tion, 2^{nd} edition, 2000.

[18] W. S. Erdmann. Gathering of kinematic data of sport event by televising the whole pitch and track. In *Proc. Int. Soc. Biomech. Sports Symposium*, pages 159–162, 1992.

[19] P.J. Figueroa, N.J. Leite, R.M.L. Barros, I. Cohen, and G. Medioni. Tracking soccer players using the graph representation. In *International Conference on Pattern Recognition*, volume 4, pages 787–790, 2004.

[20] P. Gabriel, J. Verly, J. Piater, and A Genon. The state of the art in multiple object tracking under occlusion in video sequences. In *Proc. Advanced Concepts for Intelligent Vision Systems*, page 166173, 2003.

[21] D. M. Gavrila. The visual analysis of human movement: A survey. *Comp. Vis. Image Understanding*, 73(1):82–98, 1999.

[22] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/non-gaussian Bayesian state estimation. In *IEE Proc. Radar and Signal Processing*, volume 40, pages 107–113, 1993.

[23] R. Hemlick. *Multitarget-Multisensor Tracking: Applications and Advances*, volume 3, chapter IMM Estimator with nearest-neighbour Joint Probabilistic Data Association, pages 175–178. Artech House, inc., 2000.

[24] A. V. Hill. The physiological basis of athletic records. *Sci. Monthly*, 21:409–428, 1925.

[25] C. Hue, J.-P. Le Cadre, and P. Pérez. A particle filter to track multiple objects. In *IEEE Workshop on Multi-Object Tracking*, pages 61–68, Vancouver, Canada, July 2001.

[26] S. S. Intille and A. F. Bobick. Visual tracking using closed-worlds. In *Proc. Int. Conf. Computer Vision*, pages 672–678, June 1995.

[27] M. Isard and A. Blake. CONDENSATION – conditional density propagation for visual tracking. *Int. J. Comput. Vision*, 29(1):5–28, 1998.

[28] M. Isard and J. MacCormick. Bramble: A Bayesian multiple-blob tracker. In *Proc. Int. Conf. Computer Vision*, pages 34–41, 2001.

[29] S. Iwase and H. Saito. Parallel tracking of all soccer players by integrating detected positions in multiple view images. In *International Conference on Pattern Recognition*, volume 4, pages 751–754, 2004.

[30] R. E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Engineering*, 82:34–45, 1960.

[31] J. Kang, I. Cohen, and G. Medioni. Soccer player tracking across uncalibrated camera streams. In *IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance In conjunction with ICCV03*, pages 172–179, 2003.

[32] J. B. Keller. Optimal velocity in a race. *American Math. Monthly*, 81:474–480, 1974.

[33] Z. Khan, T. Balch, and F. Dellaert. MCMC-based particle filter for tracking a variable number of interacting targets. *IEEE Trans. Patter. Anal. Mach. Intell.*, 27(11):1805–1819, November 2005.

[34] C. Kotzamanidis, K. Chatyikoloutas, and A. Gianakos. Optimization of the training plan of the handball game. *Handball: periodical for coaches,*

referees and lecturers, 2:65–71, 1999.

[35] W. L. Lu and J. J. Little. Tracking and recognizing actions at a distance. In *Proc. Workshop on Compter Vision Based Analisys in Sport Environments In conjunction with ECCV06*, pages 49–60, May 2006.

[36] C.J. Needham. *Tracking and Modelling of Team Game Interactions*. PhD thesis, School of Computing, The University of Leeds, October 2003.

[37] K. Nummiaro, E. Koller-Meier, and L. Van Gool. Color features for tracking non-rigid objects. *Chinese J. Automation*, 29(3):345–355, May 2003.

[38] H. Ok, Y. Seo, and K. Hong. Multiple soccer players tracking by CONDENSATION with occlusion alarm probability. In *Int. Workshop on Statistically Motivated Vision Processing*, 2002.

[39] K. Okuma, A. Taleghani, N. De Freitas, J. J. Little, and D. G. Lowe. A boosted particle filter: Multitarget detection and tracking. In *Proc. European Conf. Computer Vision*, volume 1, pages 28–39, 2004.

[40] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *Proc. European Conf. Computer Vision*, volume 1, pages 661–675, 2002.

[41] J. Perš and S. Kovačič. Tracking people in sport : Making use of partially controlled environment. In *Int. Conf. Computer Analysis of Images and Patterns*, pages 374–382, 2001.

[42] J. Perš, G. Vučković, S. Kovačič, and B. Dežman. A low-cost real-time tracker of live sport events. In *Proc. Int. Symp. Image and Signal Processing and Analysis*, pages 362–365, 2001.

[43] F. Pitié, S.A. Berrani, A. Kokaram, and R. Dahyot. Off-line multiple object tracking using candidate selection and the viterbi algorithm. In *Proc. IEEE Int. Conf. Image Processing*, volume 3, pages 109–112, 2005.

[44] M. K. Pitt and N. Sheppard. Filtering via simulation: Auxiliary particle filters. *J. American Statistical Association*, 94(446):590–599, 1999.

[45] C. Rasmussen and G. D. Hager. Probabilistic data association methods for tracking complex visual objects. *IEEE Trans. Patter. Anal. Mach. Intell.*, 23(6):560–576, 2001.

[46] J. G. Richards. The measurement of human motion: A comparison of commercially available systems. *Human Movement Science*, 18:589602, 1999.

[47] X. Rong Li and V. Jilkov P. Survey of maneuvering target tracking: Dynamic models. *Trans. Aerospace and Electronic Systems*, 39(4):1333–1363, October 2003.

[48] N. Saito and R.R. Coifman. Improved local discriminant bases using empirical probability density estimation. In *Comput. Section of American Statistical Association*, pages 312–321, 1997.

[49] D. Schulz, W. Burgard, D. Fox, and A.B. Cremers. Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. In *Proc. IEEE Int. Conf. Robotics and Automation*, volume 1, pages 665–670, 2001.

[50] Y. Seo, S. Choi, H. Kim, and K. S. Hong. Where are the ball and players? soccer game analysis with color based tracking and image mosaick. In *Proc. Int. Conf. Image Analysis and Processing*, volume 2, pages 196–203, 1997.

[51] J. Sullivan and S. Carlsson. Tracking and labelling of interacting multiple targets. In *Proc. European Conf. Computer Vision*, number 3, pages 619–632, 2006.

[52] M.J. Swain and D.H Ballard. Colour indexing. *Int. J. Comput. Vision*, 7(1):11–32, November 1991.

[53] P. Torma and Cs. Szepesvari. On using likelihood-adjusted proposals in particle filtering: Local importance sampling. In *Proc. Int. Symp. Image and Signal Processing and Analysis*, September 2005.

[54] J. Vermaak, A. Doucet, and P. Pérez. Maintaining multi-modality through mixture tracking. In *Proc. Int. Conf. Computer Vision*, volume 1, pages 110–116, 2003.

[55] G. Vučković, B. Dežman, F. Erčulj, S. Kovačič, and Perš J. Differences between the winning and the losing players in a squash game in terms of distance covered. In *Science and racket sports III: Int. Table Tennis Federation Sports*, pages 202–207, 2004.

[56] H. Wang, D. Suter, and K. Schindler. Effective appearance model and similarity measure for particle filtering and visual tracking. In *Proc. European Conf. Computer Vision*, volume 3851, pages 328–337, May 2006.

[57] J. Wang, D. Zhao, W. Gao, and S. Shan. Interacting multiple model particle filter to adaptive visual tracking. In *Proc. Int. Conf. Image and Graphics*, pages 568–571, 2004.

[58] Ming Xu, J. Orwell, and G. Jones. Tracking football players with multiple cameras. In *Proc. IEEE Int. Conf. Image Processing*, volume 5, pages 2909– 2912, 2004.

[59] T. Zhao and R. Nevatia. Tracking multiple humans in crowded environment. In *Proc. Conf. Comp. Vis. Pattern Recognition*, volume 2, pages 406–413, 2004.