SYSTEMS PAPER   **Open Access**

CrossMark

Constrained multi-target tracking for team sports activities

Rikke Gade* ⬤ and Thomas B. Moeslund

Abstract

In sports analysis, player tracking is essential to the extraction of statistics such as speed, distance and direction of motion. Simultaneous tracking of multiple people is still a very challenging computer vision problem to which there is no satisfactory solution. This is especially true for sports activities, for which people often wear similar uniforms, move quickly and erratically, and have close interactions with each other. In this paper, we introduce a multi-target tracking algorithm suitable for team sports activities. We extend an existing algorithm by including an automatic estimation of the occupancy of the observed field and the duration of stable periods without people entering or leaving the field. This information is included as a constraint to the existing offline tracking algorithm in order to construct more reliable trajectories. On data from two challenging sports scenarios—an indoor soccer game captured with thermal cameras and an outdoor soccer training session captured with RGB camera—we show that the tracking performance is improved on all sequences. Compared to the original offline tracking algorithm, we obtain improvements of 3–7% in accuracy. Furthermore, the method outperforms two state-of-the-art trackers.

Keywords: Tracking people, Counting people, Sports analysis, Soccer

1 Introduction

Sports analysis is an important research field, supporting a growing interest in data for statistical analysis of performance [1]. From recreational athletes wishing to track their own activities to professional teams, risking millions of dollars by losing a game, the interest in reliable performance measures is huge. Creating spatio-temporal trajectories of players is one of the essential steps in extracting statistics such as speed, distance and direction of motion. Manual annotation of video data used to be the only option, but it was very time-consuming and expensive. Thanks to research in computer vision, video analysis is increasingly automated, but even after several years of research on tracking algorithms, consistent tracking of multiple people is still very challenging [2]. Human motion can be erratic, and interactions between people substantially complicates the task.

In this paper, we focus on the application of tracking team sports activities. The challenges here are even more severe due to fast and erratic motion, close interactions between players, and the similar appearance of people.

Figure 1 shows an example image from a soccer training session.

For tracking purposes, the optimal camera view is a perpendicular top view. This is often not possible to obtain, e.g. at outdoor sports fields or in temporary indoor installations, so occlusions between people are a major challenge. Moreover, video captured from a long distance, with people wearing similar team uniforms, result in nearly identical appearances. This lack of distinct appearance information makes re-identification after full occlusions impossible. Thus, we must rely on motion information, even though some activities, especially in sports, often include fast and erratic motion. To overcome some of these challenges, we suggest utilising the fact that most team sports activities take place within a certain area, often with a constant number of people present over longer time periods.

The main contribution of this paper is a method for improving tracking precision of sports activities and similar activities with multiple people within a given area by integrating an automatic and robust counting algorithm. The estimated numbers act as constraints—guiding the tracking algorithm in these very challenging situations. We test the method on two challenging sports datasets

*Correspondence: rg@create.aau.dk
Visual Analysis of People Lab, Aalborg University, Rendsburggade 14, Aalborg, Denmark

Fig. 1 Example of an outdoor soccer training session captured by a fisheye camera (cropped to region of interest)

with people of similar appearance, as well as a more general tracking scenario with pedestrians in a courtyard environment.

The remaining part of this paper consists of the following sections: in Section 2, we discuss related work and then provide an overview of our proposed method in Section 3. Section 4 describes the counting algorithm, and Section 5 describes the tracking algorithm. Section 6 then combines those two methods in a constrained tracking approach. In the second-to-last section of the paper (7), the system is evaluated through tests and comparisons, and Section 8 concludes the paper.

2 Related work

Multi-target tracking is a popular area of research with fast progression and a large number of papers published each year [2]. Recent algorithms in this area can generally be divided into two main groups: online and offline approaches. Online methods are recursive, relying only on past observations, while offline approaches process a batch of frames in each iteration. Online methods include the classic Bayesian filters, such as Kalman filters [3] and Particle filters [4]. These are often applied in real-time applications, where processing time is crucial, and only past observations are available. However, in other applications, such as analysis of motion and behaviour, a time delay can be accepted in order to reach a higher precision. Batch processing approaches exploit more information and the possibility of running several iterations back and forth in time might help avoiding tracker drift.

For multi-target tracking in RGB images, offline approaches have become increasingly popular, due to their superior accuracy. Compared to online (recursive) approaches, offline methods have great advances in that they optimise trajectories over batches of frames. These methods all operate on a set of detections as input and aim at reconstructing the trajectories by optimising an objective function. The main difference between these algorithms lies in the formulation of the objective function and the strategy for optimisation. Among others, the optimisation task has been formulated as integer linear programme [5, 6], network flow programme [7–10], quadratic Boolean programme [11], energy minimisation [12, 13], generalised clique graphs [14, 15] and maximum weight-independent set problem [16]. Other approaches include searching a hypergraph using a local-to-global strategy [17, 18] and using a hierarchical association of detection responses [19]. Some work also focuses on improving the appearance model for solving ambiguities, e.g. by implementing an online learning approach for discriminative appearance models [20, 21].

Despite the large amount of work conducted in this field, big challenges still remain in many applications due to noise and ambiguities. From a likely noisy set of detections, the algorithm must construct an unknown number of trajectories. This task causes ambiguities and thereby errors or inaccuracies.

Most work mentioned above designs algorithms for general pedestrian tracking. Benchmark datasets within this area often feature a continuous flow of people entering and leaving the scene. The focus of this work is the tracking of players in team sports, which has different properties in people's behaviour. One of these properties is that people mostly stay within the tracked area, compared to the continuous flow through the scene seen in typical pedestrian scenarios. The specific activities observed have been taken into account in the significant amount of research in multi-target tracking specifically for team sports videos [22, 23]. Recent methods developed for team sports suggest including context information like Game Context Features [24] and contextual trajectory information [25], improving tracking by modelling latent behaviour from team-level context dynamics [26] or by improving the detection step [27, 28].

In this work, we combine the generally well-performing offline tracking strategy with the knowledge of a constant number of players on the field over longer time periods. Specifically, we take advantage of automatic counting, which can help constrain the tracking problem by estimating the number of people present in the scene.

3 Overview

We propose a tracking algorithm for team sports applications that combines an existing counting algorithm and a modified offline tracking algorithm. It runs in two main iterations. The first iteration recognises time periods that can be characterised as stable periods (no people leaving or entering the scene) as well as estimating the probability of a given number of people present during that period. In the second iteration, the result is fed to a tracking algorithm in order to constrain the number of trajectories produced during each of the stable periods. The algorithm is illustrated in Fig. 2. The estimated numbers found during stable periods are fed to the tracking algorithm. During non-stable periods, no constraint is added, leaving the tracker to try to connect the paths using the original algorithm.

4 Counting people

In most applications, the recorded scene consists of an area where people move around freely and some possible entrance/exit areas. These entrance/exit areas might be only at the edge of the image or there might be doors in the scene. Assuming that people are not continuously moving in and out of the scene, the number of people observed in the scene will stay constant during several time periods. This is especially true in sports videos when capturing a well-defined court area with a constant number of players.

An estimation of this occupancy pattern can be calculated using the approach presented in [29], which will be described briefly in the remaining part of this section.

First, we must try to detect all people in each frame. As the cameras are static, background subtraction is applied for segmentations purposes, followed by automatic thresholding. The resulting binary objects are then examined and optionally split vertically or horizontal if they are likely to represent more than one person. This procedure is described in detail in [30].

An uncertainty about whether a true person is detected or not is related to each binary object. The probability of being a true detection is related to the ratio of white pixels within the bounding box and the ratio of white pixels observed on the edge of the bounding box. In experiments, the highest probability of white pixels is found at a ratio between 30 and 60%. Furthermore, less than 50% of the edge is allowed to be white. The weightings related to the ratio of white pixels in the rectangle (r_r) and the ratio of white pixels on the perimeter (r_p) are described in Eq. 1:

$$w_p(i) = \begin{cases} 0, & \text{if } r_p > 50\% \parallel r_r < 20\% \\ 0.8, & \text{if } r_r > 70\% \\ 0.9, & \text{if } r_r < 30\% \parallel r_r > 60\% \\ 1, & \text{otherwise} \end{cases} \tag{1}$$

The weighting of each detection is combined with a weight describing the uncertainty for each frame, caused by occlusions and clutter. Each frame counting is weighted like this:

$$w_f = a \cdot \prod_{i=1}^{n} w_p(i) + (1 - a) \cdot w_s \tag{2}$$

where n is the number of people, $w_p(i)$ is the probability of person i being a true detection (see Eq. 1), and w_s is a weight that decreases with the number of splits performed, indicating how cluttered the scene is. a controls the weighting of each part. The observed number in a frame are added to a histogram with the weight w_f, and after a stable period has ended, the histogram is scaled to an accumulated sum of 1. The circles in Fig. 3 illustrate the weighted histogram for each period.

In order to split video sequences into stable and unstable periods, we must detect when people are close to the border of the scene and therefore likely to leave or enter the tracking area. The border and tracking areas must be predefined manually for each scene. Periods with people detected within the border area should be flagged as unstable and observed for people leaving or entering
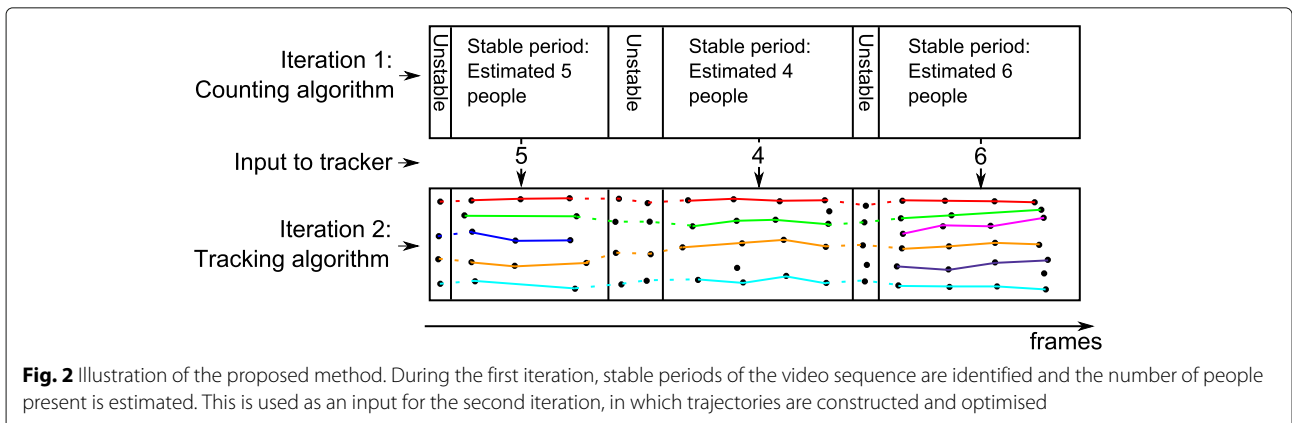

Fig. 2 Illustration of the proposed method. During the first iteration, stable periods of the video sequence are identified and the number of people present is estimated. This is used as an input for the second iteration, in which trajectories are constructed and optimised
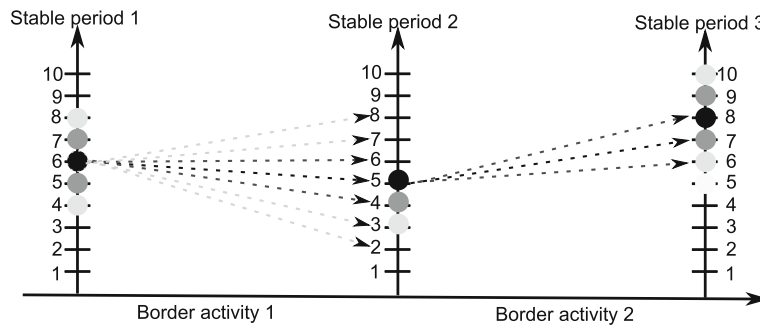
Fig. 3 Example of a simple graph. Dark nodes and edges have the highest weight. Edges exist between all nodes in two consecutive periods, but to simplify the illustration, the edges with the lowest weight are not drawn

the scene. All other periods are marked as stable periods and should contain the same number of people until the next period of border activity. Estimations of the number of people leaving and entering the scene during unstable periods are found by applying local tracking on people within the predefined area close to the border.

Estimating the number of people is done by frame-based detection succeeded by an graph optimisation algorithm, based on Dijkstra's algorithm [31]. The graph optimisation interprets the stable periods as nodes and transitions (people leaving or entering the scene) as edges. All nodes and edges have a weight factor based on the detection and tracking results.

Figure 3 illustrates the graph approach. For each stable period, the number of people is represented by circles where a darker colour indicates a higher weight. The lines between two stable periods represent the transitions, also coloured darker for a higher weight. The path through this graph is optimised to the highest total weight.

For each video sequence, this counting algorithm collects timestamps, numbers and probability weights, which are then transferred to the tracking algorithm.

5 Tracking by energy minimization
As a starting point for the offline tracking algorithm, we use the algorithm proposed by Milan et al. [13], which has shown very good results for pedestrian tracking on public datasets. It has publicly available source code[1], which we will use for further testing. The aim of this method is to find the optimal solution for multi-target tracking over an entire video sequence, given a set of coordinates for all detections. The core part of this algorithm is to minimise the following global energy function, given a set of detections X:

$$E(X) = E_{\text{det}} + \alpha E_{\text{app}} + \beta E_{\text{dyn}} + \gamma E_{\text{exc}} + \delta E_{\text{per}} + \epsilon E_{\text{reg}} \tag{3}$$

E_{det} aims to keep the solution close to the detections. E_{app} utilises the appearance of different objects to disambiguate data association. E_{dyn} is the dynamic model, using a constant velocity model. E_{exc} is a mutual exclusion term, introducing the physical constraint that two objects cannot be present in the same space at the same time. The target persistence term E_{per} penalises trajectories with start or end points far from the image border. The last term E_{reg} is a regularisation term that favours fewer targets and longer trajectories. For an exact definition of each term, we refer to [13]. E_{app} will be discarded in this work, as no appearance information is extracted.

E_{reg} is a term that considers the number of targets, and we investigate if the constraint can be integrated in this term. The original E_{reg} term proposed in [13] is defined as follows:

$$E_{\text{reg}}(X) = N + \sum_{i=1}^{N} \frac{1}{F(i)} \tag{4}$$

where F is the temporal length of trajectory i in frames and N is the total number of trajectories. Thus, the first part of the equation infers that the energy directly increases with the number of trajectories. The second part is the sum of the inverse length of all trajectories; hence, in the minimisation process, it favours long trajectories.

This tracking algorithm takes a detection file as input; thus, it can be applied on both RGB and thermal video, utilising our detection method described in Section 4.

6 Constraining the tracking algorithm
We aim to constrain the tracking algorithm to construct approximately n trajectories, where n is the number with the highest probability, estimated by the counting algorithm described in Section 4. Two relevant parameters can intuitively be formulated: the number of targets tracked per frame and the total number of trajectories in each stable period. Ideally, since we are only concerned about stable periods, the total number of

Fig. 4 A frame from the indoor thermal dataset

trajectories within a period should correspond to the number of targets tracked in each frame. However, if the trajectory of one person is fragmented into shorter tracks, the total number of trajectories will increase while the correct number of targets can still be tracked in every frame. Likewise, if the target is lost during the sequence, the total number of trajectories might be correct, while some frames have fewer targets. Therefore, both measures might be valid parameters to include in the optimisation.

In Eqs. 5 and 6, A and B represent similarity measures between the number of targets and the estimated number, per frame and per stable period, respectively:

$$A = \frac{1}{F} \sum_{i=1}^{F} P(s, n(i)) \tag{5}$$

$$B = \frac{1}{S} \sum_{s=1}^{S} P(s, N(s)) \tag{6}$$

where $P(s, n)$ is a discrete probability function constructed from the results of the counting algorithm, which returns the probability of n number of targets in stable period s. The number of targets is given either per frame i in $n(i)$ or per stable period s in $N(s)$. F is the total number of frames, and S is the total number of stable periods.

Including the original two terms, we now have four possible terms with the following purposes:

1. Minimise number of targets (orig.)
2. Maximise length of tracks (orig.)
3. Constrain number of targets per frame (A)
4. Constrain number of tracks per stable period (B)

Since we now know the estimated number of people during each period, the original term 1, which minimises the number of targets, conflicts with the purposes of terms 3 and 4, which add more specific constraints on the number. As a result, we discard term 1 and propose a new E_{reg} term, including terms 2, 3 and 4:

$$E_{\text{reg}}(X) = \sum_{i=1}^{N} \frac{1}{F(i)} - w_1 \frac{1}{F} \sum_{i=1}^{F} P(s(i), n(i))$$
$$- w_2 \frac{1}{S} \sum_{x=1}^{S} P(S(x), N(x)) \tag{7}$$

A negative sign is applied to the two new terms in order to make the optimal solution a minimum value. A weight (w_1, w_2) is added to each term, adjusting the influence from each term. These weights will be fitted during an optimisation process, described in Section 7.

7 Evaluation
7.1 Datasets
To prove the robustness of our proposed method, we test on two different sports datasets. One is captured with a thermal camera at an indoor sports arena, while the other is captured with an RGB fisheye camera at an outdoor soccer field. Thermal imaging is used for privacy reasons in the public indoor sports arena. Both datasets demonstrate the typical challenge with the similar appearance of sports players.

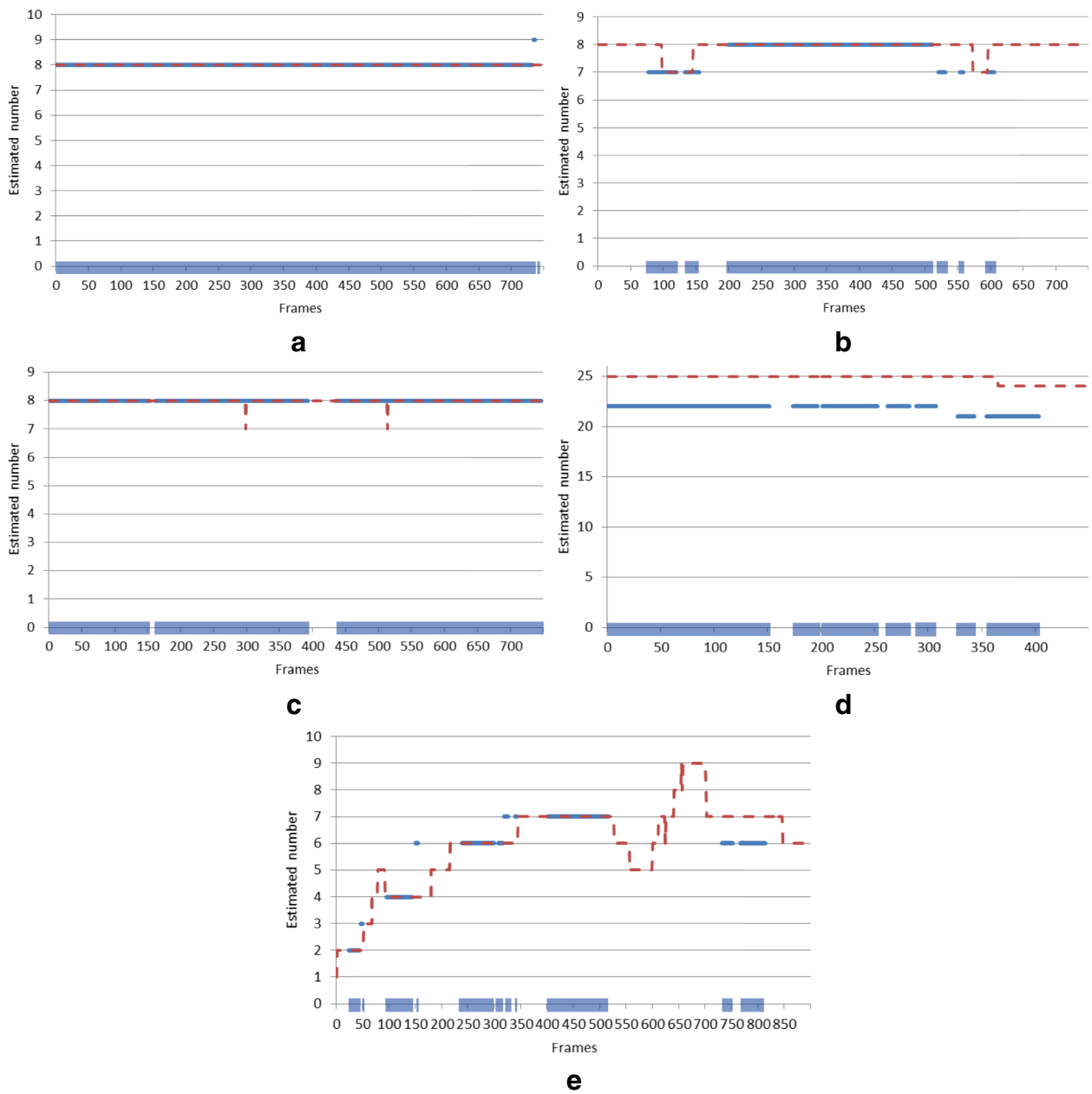**Fig. 5** Frame from the courtyard sequence

Fig. 6 Blue solid lines represent the estimated number of people. The estimated number is only available during frames in stable periods, which are also marked with blue on the *x*-axis. The red broken line is the ground truth annotated number of targets. **a** Indoor thermal sequence 1. **b** Indoor thermal sequence 2. **c** Indoor thermal sequence 3. **d** Outdoor RGB sequence. **e** Courtyard thermal sequence

Table 1 Results—indoor thermal sequence 1

	TP (%)	FP (%)	FN (%)	ID switch	MOTA (%)
Kalman	80.18	0.83	10.83	539	79.35
SMOT	92.35	8.88	1.30	381	83.47
Original CEM	91.43	2.22	5.70	172	89.22
Ours	93.95	1.38	2.80	195	*92.57*

The main dataset we use for both test and training is the thermal data captured in an indoor sports arena. In order to cover the entire field of 20×40 m, three images are captured simultaneously and stitched horizontally to a total image size of 1920×480 pixels. This dataset is captured during an indoor soccer game with six to eight players on the field in all frames. Two minutes of video are captured and manually annotated for tracking. The coverage is separated into four sequences of 30 s each in

Table 2 Results—indoor thermal sequence 2

	TP (%)	FP (%)	FN (%)	ID switch	MOTA (%)
Kalman	73.46	0.37	14.27	728	73.09
SMOT	95.03	17.74	0.47	267	77.29
Original CEM	89.31	0.62	8.40	136	88.68
Ours	92.82	0.30	5.03	128	*92.51*

Table 4 Results—outdoor RGB sequence

	TP (%)	FP (%)	FN (%)	ID switch	MOTA (%)
Kalman	29.32	0	33.98	4098	29.31
SMOT	60.93	0.04	15.98	2578	60.90
Original CEM	72.83	0.01	19.42	865	72.83
Ours	77.63	0.01	15.40	779	*77.62*

order to have a temporal window manageable for a global tracking algorithm. One sequence is used for training, and the remaining three are used for testing. This dataset is captured with an AXIS Q1922 LWIR sensor with approx. 25 fps. Figure 4 shows a frame from this thermal dataset.

The second dataset is 30 s of video captured at an outdoor soccer field. Twenty-five people are present in most frames, performing different exercises related to soccer. The images are captured with an RGB fisheye camera (Hikvision DS-2CD6362F-I(S)(V)) with a resolution of 2048×2048 pixels with 15 fps. The images are cropped to the region of interest at a final resolution of 876×827 pixels. The dimensions of the observed field area are 52×68 m, and the camera is mounted approx. 10 m above the ground. Figure 1 shows a frame from this dataset.

In addition, to show the transferability to applications other than sports, we test the tracking algorithm on a 30-s sequence captured in a courtyard environment with a thermal camera of type AXIS Q1922. This is a more general tracking scenario with small groups of pedestrians walking in a scene with few entrance/exit areas. However, because of the thermal sensor, the similar appearance of people is still a significant challenge. A frame from this dataset is shown in Fig. 5.

We find that there is a lack of publicly available team sports datasets suitable for multi-target tracking. We will contribute to building a wide purpose dataset by publishing the thermal soccer sequences along with annotations for tracking on our website[2].

7.2   Weight parameters
The parameters of the original energy function, Eq. 3, are adjusted to the sports scenario, where we discard the appearance term (α) and weigh the dynamic model (β) with 0.5, due to the erratic motion often observed

in sports. The remaining terms are weighted equally: $\alpha = 0, \beta = 0.5, \gamma = 1, \delta = 1, \epsilon = 1$.

The weight parameters w_1 and w_2 introduced in Section 5 are fitted experimentally in order to adjust the influence of each term. We use the 30 s training sequence, described in Section 7.1. Combinations of the following parameter values are tested for w_1 and w_2: {0, 0.1, 1, 10, 20, 100, 250, 500, 750, 1000}.

The results seem to be slightly more sensitive to w_1, where the accuracy is highest at $w_1 = 500$, while the accuracy varies less than 0.1% with w_2 values from 250 to 1000. We fix $w_2 = 500$. The high values are explained by non-normalised terms of the energy function.

7.3   Counting
The first iteration consists of the counting algorithm, described in Section 4, which estimates stable and unstable periods as well as the number of people present during stable periods. The counting algorithm is thoroughly evaluated in [29], but we compare here the ground truth number of targets with the estimated number for each test sequence. Figure 6 presents the results of the counting algorithm for all test sequences. The number of people is only estimated during stable periods and plotted with solid blue. Stable periods are also marked with blue on the x-axis. The ground truth is plotted with a broken red line.

Figure 6 shows that sports sequences 6a–d are dominated especially by stable periods, which is one of the main reasons we propose this method for team sports applications.

7.4   Comparison
We compare the results of our method to the original implementation of the tracking algorithm presented in [13]. Furthermore, we compare it to two different tracking algorithms suitable for multi-target tracking with

Table 3 Results—indoor thermal sequence 3

	TP (%)	FP (%)	FN (%)	ID switch	MOTA (%)
Kalman	66.42	0.05	20.41	790	66.37
SMOT	94.66	18.29	0.53	288	76.38
Original CEM	86.15	0.02	10.24	217	86.13
Ours	93.78	0.07	2.32	234	*93.71*

Table 5 Results—courtyard thermal sequence

	TP (%)	FP (%)	FN (%)	ID switch	MOTA (%)
Kalman	87.04	0.93	4.92	432	86.11
SMOT	92.45	12.24	2.34	280	80.21
Original CEM	88.13	1.38	3.72	438	86.76
Ours	92.99	1.38	3.83	171	*91.61*

objects of similar appearance. The first is an online tracking algorithm based on the Kalman filter, as described and implemented in [32]. The second algorithm, called SMOT, is a recent algorithm showing state-of-the-art results and chosen because it is specifically aimed at tracking objects of similar appearance [33]. We apply the

publicly available implementation of this tracker using the IHTLS similarity method. Three parameters should be fitted in order to adjust to the specific tracking scenario. We use our 30 s training sequence for experimentally fitting these parameters, given the following parameter values: $min\_s = 0.02, hor = 5$ and $eta\_max = 1$.
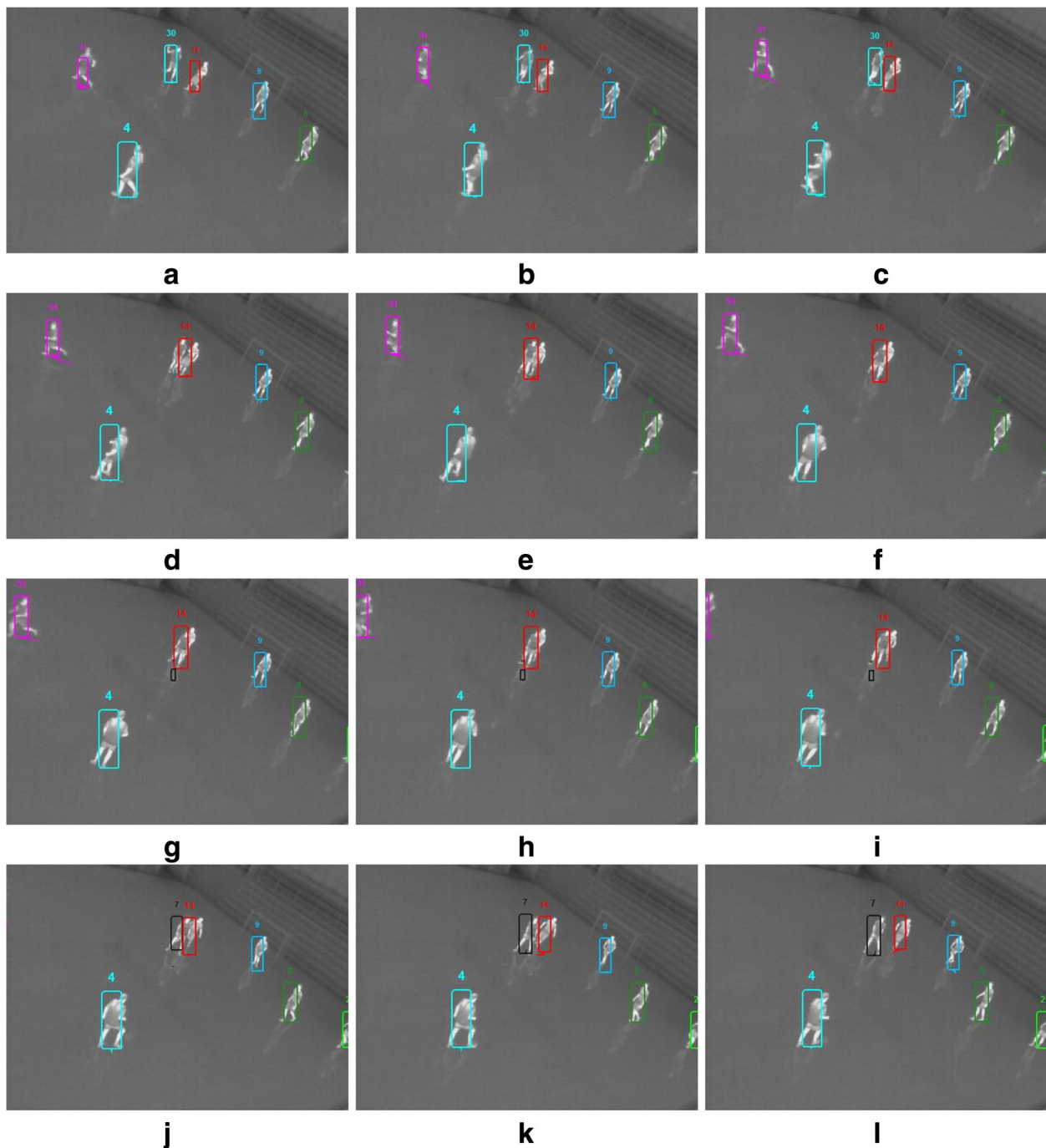

Fig. 7 Frames from thermal sequence 1 (cropped); every third frame is shown. Tracking results from the original CEM tracker are visualised. **a** Frame 1. **b** Frame 4. **c** Frame 7. **d** Frame 10. **e** Frame 13. **f** Frame 16. **g** Frame 19. **h** Frame 22. **i** Frame 25. **j** Frame 28. **k** Frame 31. **l** Frame 34

7.5 Results

For evaluating the performance, we use the multiple object tracking accuracy (MOTA) defined in the CLEAR MOT metrics [34]:

$$\text{MOTA} = 1 - \frac{\sum_t \left(\text{FN}_t + \text{FP}_t + \text{IDS}_t\right)}{\sum_t g_t} \tag{8}$$

where FN_t, FP_t and IDS_t are the number of false negatives, false positives and ID switches, respectively, for time t, while g_t is the true number of objects at time t.

The results are presented in Tables 1, 2, 3, 4, and 5. It is clear for all sequences that compared to the original CEM tracker, the number of true positives increases. For all sports sequences, the number of false positives and false
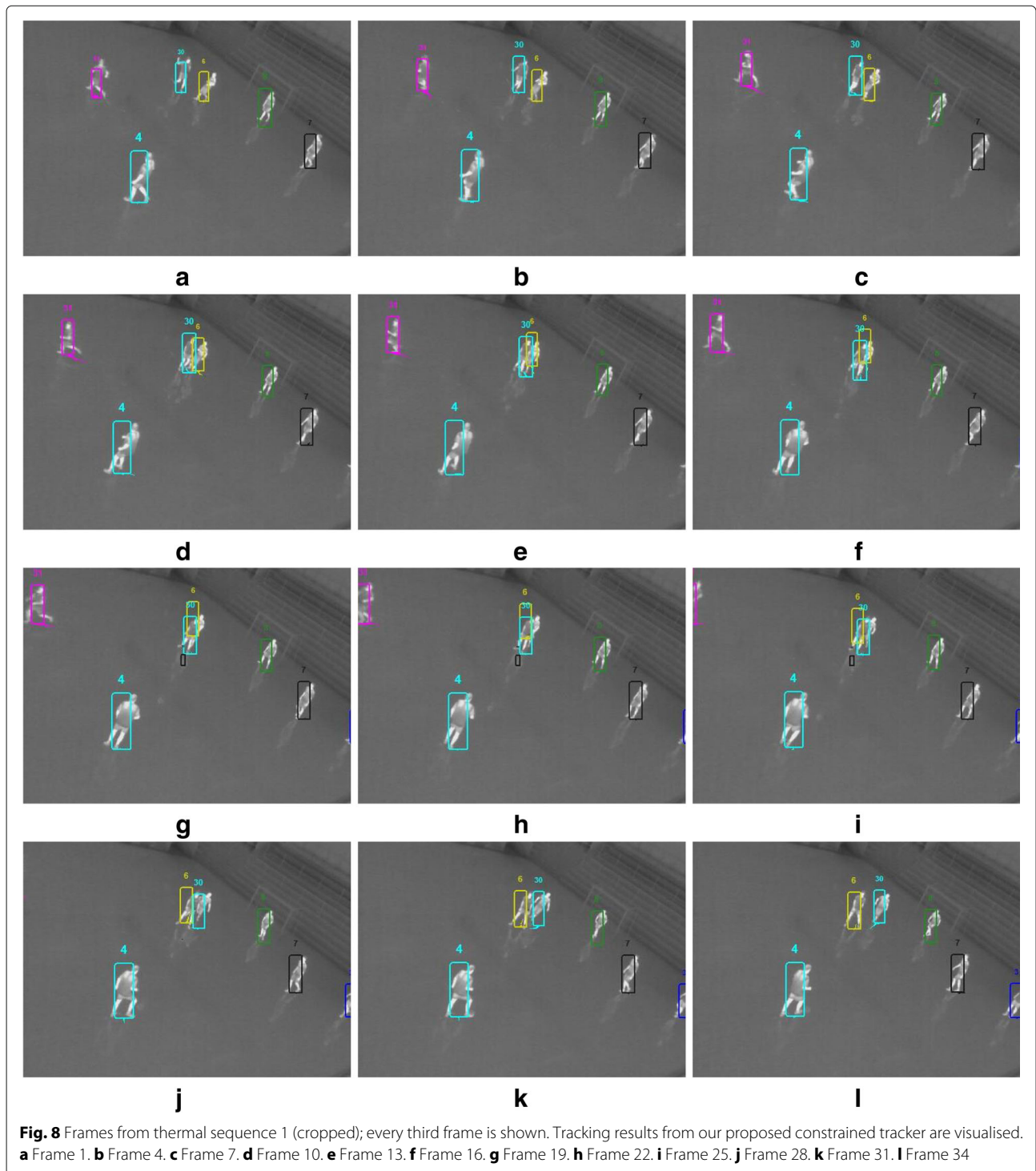

Fig. 8 Frames from thermal sequence 1 (cropped); every third frame is shown. Tracking results from our proposed constrained tracker are visualised. **a** Frame 1. **b** Frame 4. **c** Frame 7. **d** Frame 10. **e** Frame 13. **f** Frame 16. **g** Frame 19. **h** Frame 22. **i** Frame 25. **j** Frame 28. **k** Frame 31. **l** Frame 34

Table 6 Comparison between the MOTA results with automatic counting results and ground truth counting results as input

	Indoor thermal 1 (%)	Indoor thermal 2 (%)	Indoor thermal 3 (%)	Outdoor RGB (%)	Courtyard thermal (%)
Ours - aut. counting	92.57	92.51	93.71	77.62	91.61
Ours - GT counting	92.73	94.42	91.91	76.53	95.13

negatives also decreases. The number of ID switches are generally high due to unpredictable motion and a similar appearance. The final results show improvements on all sequences, with a 3–7% increase in MOTA compared to the original CEM tracker. The proposed constrained tracker also significantly outperforms both Kalman and SMOT trackers on all sequences. Figures 7 and 8 present a subsequence of frames with visualisation of tracking results from the original CEM tracker and our proposed constrained tracking algorithm, respectively.

This subsequence is a typical example of how an occlusion between two players is handled. As shown in Fig. 7, the original tracker loses one of the targets (light blue in the top right corner) between frame 10 and 25. From frame 28, a new ID is assigned to that person. The proposed constrained tracker tracks both targets throughout the subsequence. However, the IDs switch between these two targets once (yellow and light blue).

7.6 GT numbers

To analyse the influence of errors in the counting algorithm and the possibilities of the algorithm with a perfect counting result, we now compare the results from Section 7.5 with the results using ground truth numbers as input to the constrained algorithm. These results are presented in Table 6.

The results show that using a ground truth number as input to the tracking algorithm improves MOTA 0.16–3.52% on three sequences, while it gives a lower MOTA with 1.09–1.80% on the remaining two sequences. This indicates that errors in the counting algorithm do not have a large effect on the tracking result, as it is only implemented to guide the tracker. All results in Table 6 are better than the results produced by the original CEM tracker.

8 Conclusion

This work focuses on a robust tracking algorithm for team sports activities. We have shown how to combine an automatic counting algorithm with an offline tracking algorithm in order to constrain the number of tracks and improve reliability. The method is tested on four sports sequences from both indoor and outdoor scenes with 8 and 25 people, respectively, playing soccer and performing soccer-related exercises. Furthermore, we test a sequence of thermal video with pedestrians in a courtyard

to prove the applicability for other scenarios. All results show superior performance compared to three state-of-the-art trackers.

We plan to test the proposed method on several other types of team sports and refine the algorithms accordingly. For future work in this area, we will consider integrating an automatically recognised sports type, as prior context knowledge on the specific sports type may inform the tracker in ambiguous situations.

Endnote

[1] http://www.milanton.de/contracking/
[2] http://www.vap.aau.dk/dataset/

Authors' contributions
RG has designed the method, performed the experiments, and prepared this manuscript. TBM has been supervising the work and revising the paper. Both authors read and approved the final manuscript.

References
1. Gudmundsson J, Horton M (2016) Spatio-temporal analysis of team sports—a survey. arXiv:1602.06994 [cs.OH]
2. Luo W, Zhao X, Kim TK (2014) Multiple object tracking: a review. arXiv:1409.7618 [cs]
3. Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME–J Basic Eng 82(Series D):35–45
4. Doucet A, de Freitas N, Gordon N, (eds) (2001) Sequential Monte Carlo Methods in Practice. Springer, New York
5. Berclaz J, Fleuret F, Türetken E, Fua P (2011) Multiple object tracking using k-shortest paths optimization. IEEE Trans Pattern Anal Mach Intell (PAMI) 33(9):1806–1819
6. Jiang H, Fels S, Little JJ (2007) A linear programming approach for multiple object tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Minneapolis
7. Chari V, Lacoste-Julien S, Laptev I, Sivic J (2015) On pairwise costs for network flow multi-object tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Boston
8. Pirsiavash H, Ramanan D, Fowlkes CC (2011) Globally-optimal greedy algorithms for tracking a variable number of objects. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Colorado Springs
9. Zhang L, Li Y, Nevatia R (2008) Global data association for multi-object tracking using network flows. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Anchorage

10. Izadinia H, Saleemi I, Li W, Shah M (2012) (mp)2t: Multiple people multiple parts tracker. In: Proceedings of the 12th European Conference on Computer Vision - Volume Part VI, ECCV'12. Springer, Berlin. pp 100–114

11. Leibe B, Schindler K, Gool LV (2007) Coupled detection and trajectory estimation for multi-object tracking. In: Proceedings of IEEE International Conference on Computer Vision (ICCV). IEEE, Rio de Janeiro

12. Andriyenko A, Schindler K, Roth S (2012) Discrete-continuous optimization for multi-target tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Providence

13. Milan A, Roth S, Schindler K (2014) Continuous energy minimization for multitarget tracking. IEEE Trans Pattern Anal Mach Intell (PAMI) 36(1):58–72

14. Dehghan A, Assari SM, Shah M (2015) GMMCP-tracker: globally optimal generalized maximum multi clique problem for multiple object tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Boston

15. Zamir AR, Dehghan A, Shah M (2012) GMCP-tracker: global multi-object tracking using generalized minimum clique graphs. In: Proceedings of European Conference on Computer Vision (ECCV). Springer, Florence

16. Brendel W, Amer MR, Todorovic S (2011) Multiobject tracking as maximum weight independent set. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Colorado Springs

17. Wen L, Li W, Yan J, Lei Z, Yi D, Li SZ (2014) Multiple target tracking based on undirected hierarchical relation hypergraph. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus. pp 1282–1289

18. Wen L, Lei Z, Lyu S, Li SZ, Yang MH (2016) Exploiting hierarchical dense structures on hypergraphs for multi-object tracking. IEEE Trans Pattern Anal Mach Intell 38(10):1983–1996. https://doi.org/10.1109/TPAMI.2015.2509979

19. Huang C, Li Y, Nevatia R (2013) Multiple target tracking by learning-based hierarchical association of detection responses. IEEE Trans Pattern Anal Mach Intell 35(4):898–910. https://doi.org/10.1109/TPAMI.2012.159

20. Kuo CH, Huang C, Nevatia R (2010) Multi-target tracking by on-line learned discriminative appearance models. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, San Francisco. pp 685–692. https://doi.org/10.1109/CVPR.2010.5540148

21. Yang B, Nevatia R (2012) An online learned CRF model for multi-target tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, Providence. pp 2034–2041. https://doi.org/10.1109/CVPR.2012.6247907

22. Moeslund TB, Thomas G, Hilton A, (eds) (2014) Computer vision in sports. Springer, Switzerland

23. Santiago CB, Sousa A, Estriga ML, Reis LP, Lames M (2010) Survey on team tracking techniques applied to sports. In: Proceedings of International Conference on Autonomous and Intelligent Systems (AIS). IEEE, Povoa de Varzim

24. Liu J, Carr P, Collins RT, Liu Y (2013) Tracking sports players with context-conditioned motion models. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Portland

25. Zhang T, Ghanem B, Ahuja N (2012) Robust multi-object tracking via cross-domain contextual information for sports video analysis. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Kyoto

26. Xiao J, Stolkin R, Leonardis A (2014) Multi-target tracking in team-sports videos via multi-level context-conditioned latent behaviour models. In: Proceedings of the British Machine Vision Conference. BMVA Press, Nottingham

27. Lu WL, Okuma K, Little JJ (2009) Tracking and recognizing actions of multiple hockey players using the boosted particle filter. Image Vis Comput 27(1-2):189–205

28. Xing J, Ai H, Liu L, Lao S (2011) Multiple player tracking in sports video: a dual-mode two-way bayesian inference approach with progressive observation modeling. IEEE Trans Image Process 20(6):1652–1667

29. Gade R, Jørgensen A, Moeslund TB (2013) Long-term occupancy analysis using graph-based optimisation in thermal imagery. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Portland

30. Gade R, Jørgensen A, Moeslund TB (2012) Occupancy analysis of sports arenas using thermal imaging. In: Proceedings of the International Conference on Computer Vision and Applications. SCITEPRESS, Rome

31. Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1(1):269–271

32. Gade R, Moeslund TB (2014) Thermal tracking of sports players. Sensors 14:13679–13691

33. Dicle C, Camps OI, Sznaier M (2013) The way they move: tracking multiple targets with similar appearance. In: Proceedings of IEEE International Conference on Computer Vision (ICCV). IEEE, Sydney

34. Bernardin K, Stiefelhagen R (2008) Evaluating multiple object tracking performance: the CLEAR MOT metrics. EURASIP J Image Video Process 2008(1):246309