Recognising Team Activities from Noisy Data

Alina Bialkowski[1,2], Patrick Lucey[1], Peter Carr[1], Simon Denman[2], Iain Matthews[1], Sridha Sridharan[2]
[1]Disney Research, Pittburgh, USA
[2]Image and Video Laboratory, Queensland University of Technology, Australia
{alina.bialkowski,s.denman,s.sridharan}@qut.edu.au,
{patrick.lucey,peter.carr,iainm}@disneyresearch.com

Abstract

*Recently, vision-based systems have been deployed in professional sports to track the ball and players to enhance analysis of matches. Due to their unobtrusive nature, vision-based approaches are preferred to wearable sensors (*e.g. *GPS or RFID sensors) as it does not require players or balls to be instrumented prior to matches. Unfortunately, in continuous team sports where players need to be tracked continuously over long-periods of time (*e.g. *35 minutes in field-hockey or 45 minutes in soccer), current vision-based tracking approaches are not reliable enough to provide fully automatic solutions. As such, human intervention is required to fix-up missed or false detections. However, in instances where a human can not intervene due to the sheer amount of data being generated - this data can not be used due to the missing/noisy data. In this paper, we investigate two representations based on raw player detections (and not tracking) which are immune to missed and false detections. Specifically, we show that both team occupancy maps and centroids can be used to detect team activities, while the occupancy maps can be used to retrieve specific team activities. An evaluation on over 8 hours of field hockey data captured at a recent international tournament demonstrates the validity of the proposed approach.*

1. Introduction

As the sophistication of analysis increases in professional sport, more organisations are looking at using player tracking data to obtain an advantage over their competitors. For sports like field-hockey, the dynamic and continuous nature makes analysis extremely challenging as game-events are not segmented into discrete plays, the speed of play is very quick (*e.g.* the ball can move at 125km/h), and the size of the field is very large, with each player free to occupy any area at any time. A common approach to this problem is to use each player's trajectory path (e.g. linear or polynomial) and learn a combined model which can anticipate the future location of each player [12, 13, 15]. However, as reliably tracking players in this environment over relatively long periods of time (i.e. > 1 min) remains an unsolved computer vision problem, humans are required to manually correct tracks so that a continuous track of each player is obtained[1].

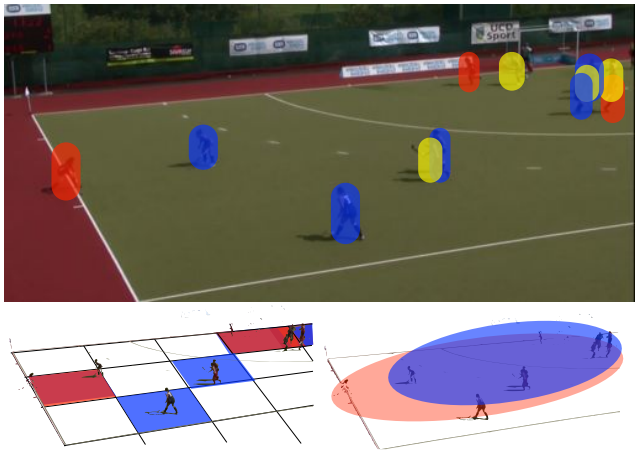


Figure 1. (Top) Detecting and tracking players over long-periods of time is challenging and often results in missed detections (highlighted in yellow) and false detections. (Bottom) In this paper, we compare two representations which are robust to false and missed detections: (left) an occupancy map, which is formed by quantising the field into discrete areas), and (right) team-centroid, which is formed by finding the mean, covariance of the detected players. We show that while both representations can detect team activities quite well, the occupancy map can be used to retrieve high-level activities to good effect.

Due to the enormous volume of data that vision-based tracking systems generate, coupled with the cost and time required to clean up the tracking data by a human, often large amounts of tracking data is rendered unusable as it contains "holes" or "gaps". To counter these issues, recent success in the area of multi-agent tracking has been gained by the use of additional "contextual" features to improve

[1]This is either done live during a match or cleaned-up post match.

tracking performance. As player motion and position (*i.e*. proximity to teammates and opponents) is heavily linked to where the action on the field is taking place and the game-context (*i.e*. is one team attacking or defending), these contextual features can be used to fill in the gaps of missed tracks (due to missed or false detections). Most notably, Liu *et al*. [17] used a coarse player occupancy map to get an indication of the game-state (*i.e*. is one team attacking or defending) to improve player tracking, while Lucey *et al*. [19] used team centroid as a contextual feature to approximate player role in conjunction with a spatiotemporal bilinear model to clean-up noisy data. In this paper, we compare: 1) the occupancy map representation – which is formed by quantising the field into a series of areas and counting the players in each area; to 2) a team-centroid representation – which is formed by calculating the mean and covariance of the detections (see Figure 1).

To enable this research we used player detection data captured via 8 fixed high-definition (HD) cameras, across seven complete field-hockey matches (over 8 hours of match data for each camera). We utilise a state-of-the-art real-time player detector [3] to give player positions at every frame, affiliate detection results into teams using a colour histogram model, and compare both approaches across a series of labelled team activities. Additionally, we show the utility of these representations for the task of play retrieval. We evaluate the performance relative to ground truth annotations, and demonstrate that our descriptor is able to quickly and accurately locate activities similar to a query without any tracking information.

2. Related work

Due to the host of military, surveillance and sport applications, research into recognising group behaviour has recently increased dramatically. Outside of the sports realm, most of this work has focussed on dynamic teams (i.e. where individual agents can leave and join teams over the period of the observations). An initial approach was to recognise the activities of individual agents and then combine these to infer group activities [1]. Sukthankar and Sycara recognised group activities as a whole but pruned the size of possible activities by using temporal ordering constraints and agent resource dependencies [26, 27]. Sadilek and Kautz [23] used GPS locations of multiple agents in a "capture the flag" game to recognise low-level activities such as approaching and being at the same location. All of these works assume that the position and movements of all agents are known, and that all behaviours can be mapped to an activity within the library. Recently, Zhang *et al*. [29] used a "bag of words" and Support Vector Machine (SVM) approach to recognise group activities on the Mock Prison dataset [4].

Sport related research mostly centres on low-level activity detection with the majority conducted on American Football. In the seminal work by Intille and Bobick [11], they recognised a single football play *pCurl51*, using a Bayesian network to model the interactions between the players trajectories. Li *et al*. [16], modelled and classified five offensive football plays (dropback, combo dropback, middle run, left run, right run). Siddiquie *et al*. [24], performed automated experiments to classify seven offensive football plays using a shape (HoG) and motion (HoF) based spatio-temporal features. Instead of recognising football plays, Li and Chellapa [15] used a spatio-temporal driving force model to segment the two groups/teams using their trajectories. Researchers at Oregon State University have also done substantial research in the football space [9, 8, 25] with the goal of automatically detecting offensive plays from a raw video source and transferring this knowledge to a simulator. For soccer, Kim *et al*. [13] used the global motion of all players in a soccer match to predict where the play will evolve in the short-term. Beetz *et al*. [2] developed the *automated sport game models* (ASPOGAMO) system which can automatically track player and ball positions via a vision system. Using soccer as an example, the system was used to create a heat-map of player positions (i.e. which area of the field did a player mostly spend time in) and also has the capability of clustering passes into low-level classes (i.e. long, short etc.), although no thorough analysis was conducted due to a lack of data. In basketball, Perse *et al*. [22] used trajectories of player movement to recognise three type of team offensive patterns. Morariu and Davis [21] integrated interval-based temporal reasoning with probabilistic logical inference to recognise events in one-on-one basketball. Hervieu *et al*. [7] also used player trajectories to recognise low-level team activities using a hierarchical parallel semi-Markov model.

It is worth noting that an enormous amount of research interest has used broadcast sports footage for video summarisation in addition to action, activity and highlight detection [14, 18, 10, 28, 20, 6, 2, 5], but given that these approaches are not automatic (i.e. the broadcast footage is generated by humans) and that the telecasted view captures only a portion of the field, analysing groups has been impossible because some individuals are normally out of frame. Although similar in spirit to the research mentioned above, our work differs as: 1) we rely only on player detections rather than tracking, and 2) we compare across many matches (7 compared to 1).

3. Detection Data

3.1. Field-Hockey Test-Bed

To enable this research we used player detection data captured via 8 fixed HD cameras, and over seven complete field-hockey matches (over 8 hours of match data for each

Figure 2. View of the field-hockey pitch from the 8 fixed HD cameras.

Match code	Activities	Frames Annotated	
		1st Half	2nd Half
1-JPN-USA	✓	-	-
2-RSA-SCO	✓	-	-
5-USA-SCO	✓	-	-
9-JPN-SCO	✓	-	-
10-USA-RSA	✓	14352	-
22-RSA-IRL	-	17861	-
23-ESP-SCO	✓	-	-
24-JPN-USA	✓	20904	7447

Table 1. Itemised list of analysed field hockey data.

camera). Each camera is connected to a computer which extracts the player positions and their team from the video feed. This is then relayed to a central hub via optic fibre where the detections are merged, and can be analysed online to perform activity recognition and other analysis. The cameras are attached to light-pole structures at a height of 18m, and provide complete coverage of the field. Example images from the eight cameras are displayed in Figure 2. From this test-bed, we collected and analysed over 8 hours of video footage from a recent field hockey tournament. The analysed matches are listed in Table 1, along with the number of frames annotated with players' team and field position (x,y). Due to the enormous amount of time it takes to manually label player tracks during a match, we limited our labelling effort to three matches (four halves). However, match statistics or team activities were labelled for seven complete matches as can be seen in this table.

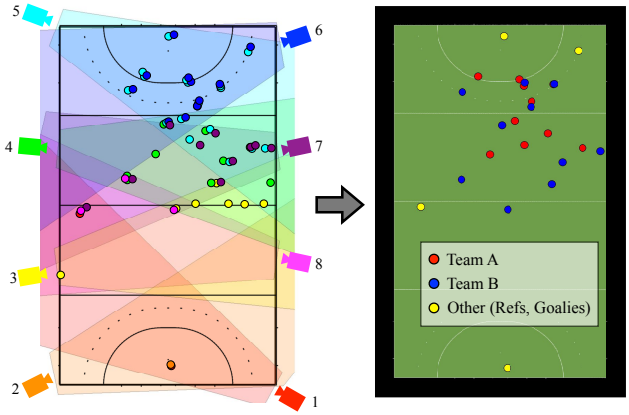


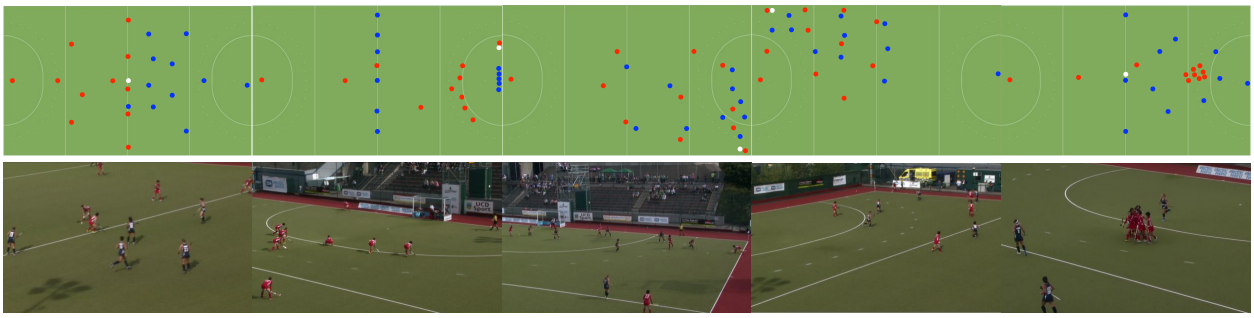
Figure 3. (Left) We detect players in each camera using a real-time person detector. (Right) We then classify the detections into one of the two teams and aggregate the information from each camera to extract the state of the game at each time instant.

3.2. Player Detector and Team Affiliation

An illustration of our player location and team affiliation method is shown in Figure 3. For each camera, we extract player image patches using a real-time person detector [3], which detects players by interpreting background subtraction results in terms of 3D geometry, where players are coarsely modelled as cylinders of height 1.8 m. The overlapping camera arrangement allows players missed in one view to potentially be detected in another camera view. However, players may still be missed when they stand close to one another. The detected player image patch bounding boxes represent a height of 1.8×0.9 meters and vary in size depending on the distance from the camera (40-100 pixels height). Detected patches are normalised to 90×45 pixels, and classified into teams using colour histograms of the foreground pixels.

The A and B channels of the LAB colour space are used (the luminance channel is ignored as it is affected by illumination changes), with nine bins for each dimension, and the histograms are normalised to sum to one. The team models are learnt from a training set of approximately 4000 training histograms, using k-means clustering, and we compare histograms using the Bhattacharyya coefficient. An image patch is classified to the closer of the two models, or if it falls outside a threshold, it is put into the "others" team (i.e. noise, referees, goalies). In our dataset, teams are always dressed in contrasting colours, so colour histograms are suitable for distinguishing between the two teams. The detections are then aggregated by projecting the player positions to field co-ordinates using each camera's homography and utilising the covariance matrix of each player position (representing the player position error) to merge detections.

The performance of the detector and team classification compared to ground truth annotated frames using precision



(a) Face-off (b) Penalty corner (c) Long corner (d) Defensive corner (e) Goal scored

Figure 4. Diagrams and examples of structured plays that occur in field-hockey

Match code	Detector	Team A	Team B
10-USA-RSA-1	81.1%	67.2%	77.7%
22-RSA-IRL-1	80.3%	84.9%	70.2%
24-JPN-USA-1	89.5%	91.7%	90.0%
24-JPN-USA-2	85.8%	72.4%	79.7%

Table 2. Precision values after aggregating all cameras

Match code	Detector	Team A	Team B
10-USA-RSA-1	89.0%	98.3%	98.4%
22-RSA-IRL-1	88.4%	97.6%	98.2%
24-JPN-USA-1	87.5%	95.2%	97.4%
24-JPN-USA-2	90.0%	97.6%	97.0%

Table 3. Recall values after aggregating all cameras (Team A and Team B are relative to what was recalled by the detector)

and recall metrics is shown in Tables 2 and 3 respectively. In these tables, it is evident that while recall is high, the team classification has quite low precision in some matches. The poor performance is mainly attributed to non-team-players (referees, goalies, and false-positive player detections) being misclassified into one of the teams, since the image patches contain a combination of team colours, especially in the varying lighting conditions. A better model representation could be used for colour modelling, and online learning of the colour models to adapt with changes in illumination would further improve results. From these results, it is evident that our team behaviour representation must be able to deal with a high degree of noise.

3.3. Team Activity Labels

Seven complete matches were annotated with common activities that occur in field hockey, which are listed along with their frequency count in Table 4 for each match half. Both pictorial and broadcast examples of these five activities are shown in Figure 4. To quantify our approach, we look at classification and retrieval of these activities. Each of these five activities correspond to activities or statistics that an analyst would label during a game. As all of them have distinctive spatial locations and motion patterns, the reliability of labelling these activities are very high. In this work, we try to automatically detect these activities based solely on noisy player detections (*i.e.* there is no ball or player identity information).

	Face off	Pen. Cnr	Goal	Long Cnr (L)	Long Cnr (R)	Def Cnr (L)	Def Cnr (R)
1-JPN-USA-1	3	2	2	11	5	4	4
1-JPN-USA-2	2	6	1	4	10	7	3
2-RSA-SCO-1	2	4	2	11	4	3	3
2-RSA-SCO-2	3	9	2	3	12	4	3
5-USA-SCO-1	3	4	2	7	4	1	7
5-USA-SCO-2	3	8	2	3	3	2	2
9-JPN-SCO-1	2	4	2	8	7	5	2
9-JPN-SCO-2	1	1	0	10	10	6	0
10-USA-RSA-1	5	9	5	5	5	8	0
10-USA-RSA-2	6	4	5	6	7	4	1
23-ESP-SCO-1	3	4	2	7	6	1	1
23-ESP-SCO-2	3	7	2	9	5	2	1
24-JPN-USA-1	4	3	3	9	6	5	1
24-JPN-USA-2	2	2	1	5	9	7	6
Total	42	67	31	98	93	59	34

Table 4. Activity frequency in each match half

4. Representing Team Behaviors

4.1. Team Occupancy Maps

Team sports like field-hockey are played over a very large spatial area. An intuitive representation of sports would be to track all players (maintaining their identity) and the ball, which would result in a 46 dimensional signal (i.e. 23 objects in x and y coordinates – 11x2 players, 1 ball). However, since we cannot reliably and accurately track the player and ball over long durations (*e.g.* 35mins), an alternative is to represent the match via player detections. By using detections, we overcome the issue of tracking but as a consequence we remove the player identity component of the signal, and need another method to maintain feature correspondences. We propose to employ an occupancy descriptor, which is formed by breaking the field into a series

of spatial bins and counting the number of players that occupy each of the bins.

The team occupancy descriptor $\mathbf{x}_i^o$ is a quantised occupancy map of the player positions on the field for each team represented at time i. Given we have the locations of players from the player detector system and have assigned team affiliation, for each frame, an occupancy map is calculated by quantising the field into P bins, and counting how many player detections for that team fall within each location. The dimensionality of the formation descriptor is equal to twice the number of bins (i.e. $P \times 2$) so that both teams A and B are accounted for, resulting in $\mathbf{x}_i^o = [a_1, \ldots, a_P; b_1, \ldots, b_P]$, where a_l and b_l are the player counts for teams A and B in each bin l. Depending on the level of complexity of the activity that we want to recognise, we can use varying descriptors (coarse to fine). In this paper we evaluate five different descriptor sizes: $P = 2(2 \times 1)$, $P = 8(4 \times 2)$, $P = 32(8 \times 4)$, $P = 135(15 \times 9)$, and $P = 540(30 \times 18)$. The different quantisations represent how much tolerance there is in player's positions within an activity and can be thought of as the space each player occupies in the activity (e.g. in 15×9 quantisation, each player occupies an area of 6 m^2).

Since an activity can occur for either team, we compare the template descriptors in both orientations ($\mathbf{x}^o = [\mathbf{a}, \mathbf{b}]^T$, and $\mathbf{x}^o = [\mathbf{b}_{rot}, \mathbf{a}_{rot}]^T$, where $\mathbf{a}_{rot}$ represents a rotation of the field by 180° for team a's formation descriptor, so that the new descriptor is given by $\mathbf{a}_{rot}[i] = \mathbf{a}[P+1-i]$, for $i = 1, 2, \ldots, P$). We take the minimum of the two orientations as the distance measure. Examples of the Team Occupancy Maps are displayed in Figure 5.

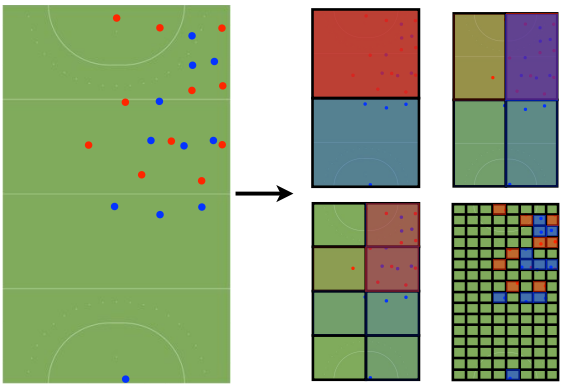


Figure 5. Example Team Occupancy Maps using 2×1, 4×2, 8×4 and 15×9 descriptor sizes.

4.2. Team Centroid Representation

Given the player detections and their team affiliations, the centroid representation, $\mathbf{x}_i^c$ is found by calculating the mean and covariance of the player positions for each team. As with the team occupancy representation, we compare centroid features in both orientations, and the rotated positions are given by $x_{rot} = 91.4 - x$ and $y_{rot} = 55.0 - y$ (where 91.4 m $\times$ 55.0 m are the dimensions of the field and x and y are the positions on the field). An example of the team centroid representation is depicted in Figure 6.

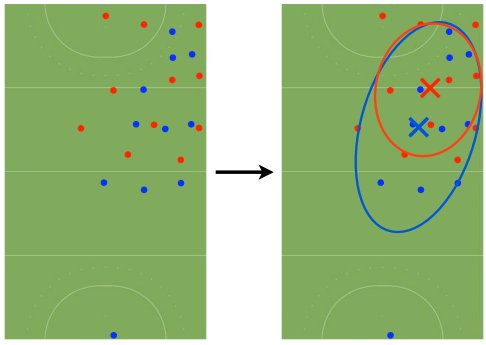


Figure 6. The team centroid representation is overlaid on the player detections. The 'x' represents the mean position for each team (i.e. the centroids), and the ellipses represent the covariances of their positions.

5. Experiments

5.1. Isolated Team Activity Recognition

To compare the different representations, we conducted a series of isolated activity recognition experiments. As these activities coincide with a single event (i.e. the ball crossing the outline, or a goal being scored), they do not have distinct onset and offset times. To account for this, we used the event as the start of the activity and went forward 10 seconds as the offset time, which gave us a series of 10 second play clips. We split the annotated activities of Table 4 into testing and training sets using a leave-one-out cross-validation strategy, where one half was used for testing and the remaining match halves for training. We used a k-Nearest Neighbour classification approach, taking the mode activity label of the closest k examples in the training set, using $L2$ as our distance measure. Confusion matrices using $k = 10$ are presented in Figure 7.

We achieve the best accuracy using the centroid descriptor, with an accuracy of 79.3%, followed closely by an 8×4 descriptor, with an accuracy of 78.2%. Most activities are well recognised, however goals are most often misclassified, being randomly classified as the other activities, as they are less structured, with a lot of variability possible. Defensive corners and long corners are sometimes confused as the main difference is the team which maintains possession, which is not discernible from the occupancy or centroid descriptors.

The centroids outperform the team formation descriptors, which may be attributed to the fact that these activities can be described on a macroscopic scale (i.e. by the

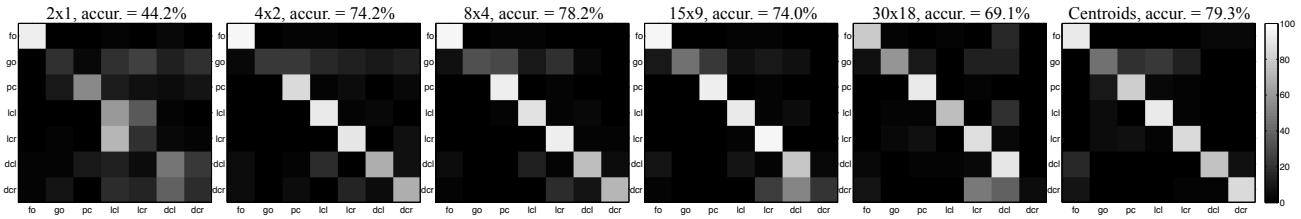


Figure 7. Confusion matrices for isolated activity recognition using different descriptor sizes, and centroids (far right)

global distribution of players, which is captured by the centroid, rather than individual positions as represented by the more fine descriptors). The 8×4 descriptor captures similar granularity to the centroid and is also very effective in distinguishing between the labelled activities, and both representations are able to accurately recognise the game state, despite their simplicity, in the presence of noise and without any tracking information.

5.2. Continuous Team Activity Recognition

Recognising team activities in a continuous sense is a more challenging task than isolated recognition, as events are not separated and a lot of movements and formations can appear very similar to labelled activities (particularly without knowledge of where the ball is, and in the presence of noise). In this section, we qualitatively demonstrate how our representations can be used to retrieve team activities in a continuous domain.

In Figure 8, centroids for match half 24-USA-JPN-1 are displayed with ground truth labels for goals and penalty corners. It can be seen that goals correspond to regions where both teams are located close to the goals, followed by a movement to the centre of the field. A penalty corner ('PC') is characterised by the team centroids being separated for a duration of time (as they move into formation, and the attacking team plans their attack), followed by a convergence towards the goal when the ball is brought into play. This information can be used to quickly recognise the game state.

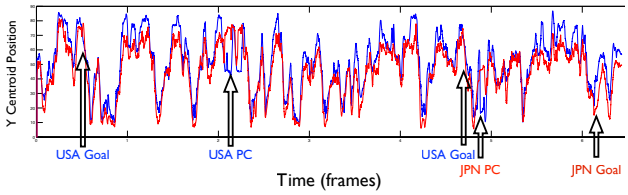


Figure 8. Team centroids (y-position) across a match half 24-USA-JPN-1. It can be seen that centroids provide important information for game state and can be used to assist in retrieving activities.

While centroids are very useful, many team behaviours will have similar centroids, and to pick up on more specific behaviours and activities, a finer descriptor is necessary. To demonstrate retrieval, we calculate the distances between the occupancy map descriptors extracted from a game and a template of the activity of interest using a sliding window. In Figure 9, we used a 15×9 descriptor to recognise two different activities in match half 2-RSA-SCO-2. A 15×9 descriptor was used as it was found that a smaller descriptor size was often confused with non-activities when compared in a continuous domain. It can be seen that the descriptor is able to effectively locate the ground truth activity regions for a penalty corner and a face off.

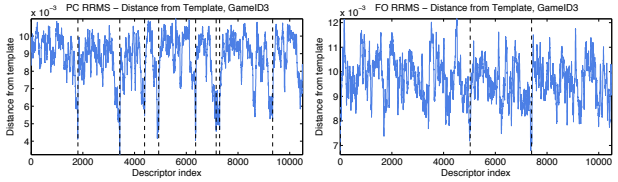


Figure 9. Retrieval distances for a Penalty Corner (left) and Face Off (right). The plots show the distances from an example activity template we wish to retrieve (in blue) vs ground truth activity onset (black dashed vertical line). A low distance (high similarity) is apparent at the ground-truth locations.

6. Summary and Future Work

Accurately tracking players over long durations of time is an unsolved computer vision problem, and prevents automated analysis of team sports using traditional representations based on player tracks. In this paper, we presented a fully automated method which is able to recognise team activities from raw player detections, without player tracking or ball information. We investigated two representations – a team centroid representation and team occupancy maps – which are robust to missed and false detections and demonstrated that both team occupancy maps and centroids can be used to accurately recognise team activities in the presence of noise. While both representations are able to detect team activities quite well, the occupancy map can be used to retrieve more specific team activities.

Future work will involve improving the team classification, and learning activities in an unsupervised fashion. We also seek to automatically predict future events/activities based on an observed sequence of play.

References

[1] D. Avrahami-Zilberbrand, B. Banerjee, L. Kraemer, and J. Lyle. Multi-Agent Plan Recognition: Formalization and Algorithms. In *AAAI*, 2010.

[2] M. Beetz, N. von Hoyningen-Huene, B. Kirchlechner, S. Gedikli, F. Siles, M. Durus, and M. Lames. ASPOGAMO: Automated Sports Game Analysis Models. *International Journal of Computer Science in Sport*, 8(1), 2009.

[3] P. Carr, Y. Sheikh, and I. Matthews. Monocular object detection using 3d geometric primitives. In *ECCV*. Springer, 2012.

[4] M. Chang, N. Krahnstoever, and W. Ge. Probabilistic Group-Level Motion Analysis and Scenario Recognition. In *ICCV*, 2011.

[5] T. D'Orazio and M. Leo. A Review of Vision-Based Systems for Soccer Video Analysis. *Pattern Recognition*, 43(8), 2010.

[6] A. Gupta, P. Srinivasan, J. Shi, and L. Davis. Understanding Videos, Constructing Plots: Learning a Visually Grounded Storyline Model from Annotated Videos. In *CVPR*, 2009.

[7] A. Hervieu and P. Bouthemy. Understanding sports video using players trajectories. In J. Zhang, L. Shao, L. Zhang, and G. Jones, editors, *Intelligent Video Event Analysis and Understanding*. Springer Berlin / Heidelberg, 2010.

[8] R. Hess and A. Fern. Discriminatively Trained Particle Filters for Complex Multi-Object Tracking. In *CVPR*, 2009.

[9] R. Hess, A. Fern, and E. Mortensen. Mixture-of-Parts Pictorial Structures for Objects with Variable Part Sets. In *ICCV*, 2007.

[10] C. Huang, H. Shih, and C. Chao. Semantic Analysis of Soccer Video using Dynamic Bayesian Networks. *T. Multimedia*, 8(4), 2006.

[11] S. Intille and A. Bobick. A Framework for Recognizing Multi-Agent Action from Visual Evidence. In *AAAI*, 1999.

[12] S. Intille and A. Bobick. Recognizing Planned, Multi-Person Action. *Computer Vision and Image Understanding*, 81:414–445, 2001.

[13] K. Kim, M. Grundmann, A. Shamir, I. Matthews, J. Hodgins, and I. Essa. Motion Fields to Predict Play Evolution in Dynamic Sports Scenes. In *CVPR*, 2010.

[14] M. Lazarescu and S. Venkatesh. Using Camera Motion to Identify Different Types of American Football Plays. In *ICME*, 2003.

[15] R. Li and R. Chellappa. Group Motion Segmentation Using a Spatio-Temporal Driving Force Model. In *CVPR*, 2010.

[16] R. Li, R. Chellappa, and S. Zhou. Learning Multi-Modal Densities on Discriminative Temporal Interaction Manifold for Group Activity Recognition. In *CVPR*, 2009.

[17] J. Liu, P. Carr, R. Collins, and Y. Liu. Tracking Sports Players with Context-Conditioned Motion Models. In *CVPR*, 2013.

[18] T. Liu, W. Ma, and H. Zhang. Effective Feature Extraction for Play Detection in American Football Video. In *MMM*, 2005.

[19] P. Lucey, A. Bialkowski, P. Carr, S. Morgan, I. Matthews, and Y. Sheikh. Representing and Discovering Adversarial Team Behaviors using Player Roles. In *CVPR*, 2013.

[20] A. Money and H. Agius. Video Summarisation: A Conceptual Framework and Survey of the State of the Art. *Journal of Visual Communication and Image Representation*, 19(2):121–143, 2008.

[21] V. Morariu and L. Davis. Multi-Agent Event Recognition in Structured Scenarios. In *CVPR*, 2011.

[22] M. Perse, M. Kristan, S. Kovacic, and J. Pers. A Trajectory-Based Analysis of Coordinated Team Activity in Basketball Game. *Computer Vision and Image Understanding*, 2008.

[23] A. Sadilek and H. Kautz. Recognizing Multi-Agent Activities from GPS Data. In *AAAI*, 2008.

[24] B. Siddiquie, Y. Yacoob, and L. Davis. Recognizing Plays in American Football Videos. Technical report, University of Maryland, 2009.

[25] D. Stracuzzi, A. Fern, K. Ali, R. Hess, J. Pinto, N. Li, T. Konik, and D. Shapiro. An Application of Transfer to American Football: From Observation of Raw Video to Control in a Simulated Environment. *AI Magazine*, 32(2), 2011.

[26] G. Sukthankar and K. Sycara. Hypothesis Pruning and Ranking for Large Plan Recognition Problems. In *AAAI*, 2008.

[27] G. Sukthankar and K. Sycara. Activity Recognition for Dynamic Multi-Agent Teams. *ACM Trans. Intell. Syst. Technol*, 2012.

[28] C. Xu, Y. Zhang, G. Zhu, Y. Rui, H. Lu, and Q. Huang. Using Webcast Text for Semantic Event Detection in Broadcast. *T. Multimedia*, 10(7), 2008.

[29] Y. Zhang, W. Ge, M. Chang, and X. Liu. Group Context Learning for Event Recognition. In *WACV*, 2012.