Discovering Team Structures in Soccer from Spatiotemporal Data

Alina Bialkowski, *Member, IEEE,* Patrick Lucey, Peter Carr, Iain Matthews, *Member, IEEE,* Sridha Sridharan, *Senior Member, IEEE,* and Clinton Fookes, *Senior Member, IEEE*

Abstract—In team sports like soccer, utilising tracking data for analysis is challenging due to the dynamic and multi-agent nature of the data. The biggest issue surrounds the changing of positions or "roles" between players on a frame-to-frame basis, which causes misalignment of the data and makes it difficult to perform team analysis. In this paper, we present an unsupervised method to learn a formation template which allows us to "align" the tracking data at the frame level. Not only does this approach give important contextual information to facilitate large-scale analysis (e.g. we know when a player is in the left-wing position compared to left-back), it also yields the team structure or "formation" which serves as a strong descriptor for identifying a team's style. The utility of the approach is demonstrated on a full season of player and ball tracking data from a professional soccer league consisting of over 21.5 million frames of player tracking data.

Index Terms—Formation, team analysis, multi-agent, sports analytics, soccer, role, alignment, group behaviour, spatio-temporal data.

1 Introduction

In many professional team sports such as soccer, tracking systems are providing large amounts of player and ball tracking data for post-match analysis and reporting of statistics. Despite this, large-scale mining of such data has been limited due to the difficulty in representing dynamic multi-agent trajectories. One of the main issues centres around the lack of spatial alignment in the tracking data and this is apparent when observing the long-term player distributions. In Fig. 1(a) the distribution of each player's position across half a match (45 mins) is shown, demonstrating how the continuous interchanging of player positions results in significant overlap in the distributions. Although the team structure and role of each player is usually decided before a match by the coach, the formation or structure executed can differ a lot from the initial plan. Even after accounting for translation variation as in Fig. 1(b), there is still overlap in their spatial distributions, highlighting variations in players' relative positions on a frame-to-frame basis. This misalignment of the tracking data must be overcome to discover the true structure of a team and to perform large-scale spatiotemporal team analysis.

In this paper, we present a role-alignment method to learn a team's formation within a match directly from player tracking data, based on the minimum entropy data partitioning method [1], [2]. This disentangles players into distinct roles (such as in Fig. 1(c)), providing a representation of the actual formation a team played over a match, and brings spatial structure to tracking data to enable individual and team analysis to be performed. Compared to existing analysis methods which simply plot locations of a particular player for an event or their mean position over time, our role-alignment method adds important contextual information to player analysis with regards to their teammates. For example in soccer (see Fig. 1(d)), given we have a player who starts on the right-wing but then switches to the left wing, we get two distinct types of behaviours (i.e. left and right wing play). Current analysis conducts

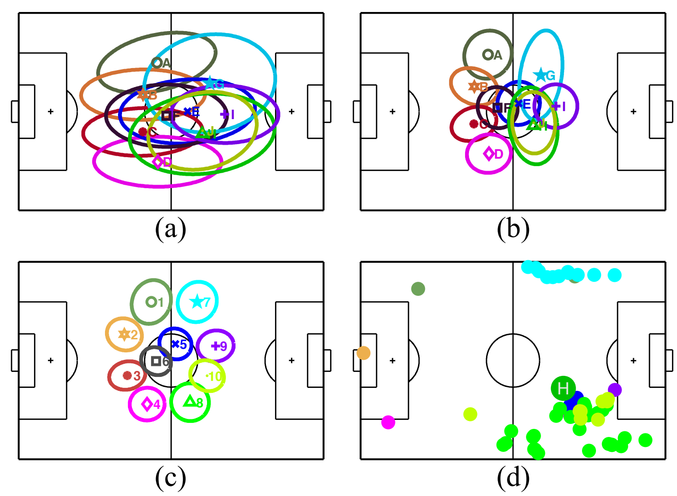


Fig. 1. Player position swaps throughout a match cause misalignment and high overlap in tracking data which needs to be overcome to perform large scale-analysis. In (a), the mean and covariance of each player's position across half a match is shown. After normalising for translation variation as in (b), there is still high overlap in the player distributions. Using our role-assignment procedure, these overlapping distributions can be disambiguated and the underlying structure or *formation* of the team can be extracted and visualised as in (c). This formation provides context for in-game analysis, giving the relative position or "role" of each player at each frame of the match. Compared to the mean ball-touch location which is often used in match analysis, our method provides context as in (d), where Player H's role during each ball touch has been coloured relative to the discovered roles in (c), highlighting the distinct roles of right-wing (green) and left-wing (cyan) which are missed when simply taking the mean.

- *A. Bialkowski, S. Sridharan and C. Fookes are with the Image and Video Laboratory, Queensland University of Technology, Australia, QLD, 4000. E-mail: a.bialkowski@connect.qut.edu.au*
- *P. Lucey is with STATS, Chicago, USA.*
- *P. Carr and I. Matthews are with Disney Research, Pittsburgh, USA.*

Manuscript received May XX, 2015; revised March XX, 2016

the analysis based on his original position or "role" (right-wing). Our approach provides a contextual label noting the player's role at that specific moment. The role-alignment also enables visualisation of team structure and clustering which can be used to find teams which play similarly, or to find the different structures a team adopts in different circumstances (e.g. a team may play one style at home and another away, or one style against a top-team and another against a bottom team). The utility of our approach is demonstrated using a full season of player and ball tracking data from a professional soccer league (> 400,000,000 data points).

2 Related Work

With the recent deployment of player tracking systems in professional sports, a recent influx of research has been conducted on how to use such data sources (See [3] for a review of recent methods of spatio-temporal data analysis in sports). Although all team sports are instantiations of multi-agent trajectories, most current work using spatiotemporal data has focussed on individual behaviours thus avoiding the issue of alignment. Examples of this include work done in basketball where individual shooting, rebounding and decision-making characteristics are analysed [4], [5], [6]. Miller et al. [7] used non-negative matrix factorisation to characterise different types of shooters in basketball by modelling shot attempts as a point-process. In soccer, Lucey et al. [8], [9], detected a team's style of play using an occupancy map of team's ball movement. Gudmundsson and Wolle [10] clustered the passes and movement of individual players. Pena and Touchette [11] used network theory to characterise team patterns by fixing players in their nominal position and quantifying importance based on the number of passes between players. In tennis, Wei et al. [12], [13] used Hawk-Eye data to predict the type and location of the next shot based on the behaviour of the opponent.

In multi-agent domains, the common thread of aligning trajectories has centred on using a predefined quantised representation or codebook of the environment. The seminal work of Intille and Bobick [14] used pre-aligned trajectories to recognise a single American football play. Zhu et al. [15] combined the movements of the players and the ball in soccer into a single "aggregate trajectory" to classify goal scoring events into categories. Jiang et al. [16] detect the game state (attacking/defending) in soccer from broadcast video using a finite state machine based on scene analysis. Perse et al. [17] recognised activities in basketball by converting player trajectories into a string of symbols based on key player positions and actions using a quantised court. Bricola [18] recognised activities in basketball from player trajectories by segmented the trajectories into tracklets and matching them to codewords using dynamic time warping. Stracuzzi et al. [19] recognised group activities in American Football using a labelled dataset of actions and the trajectories were labelled by matching them to the closest in the labelled dataset. Dynamic time warping was used to compare the signals and the features of each aligned point. Atmosukarto et al. [20] used visual features consisting of the spatial distribution of gradient intensity for the offensive side of the line of scrimmage to classify offensive formations. Kim et al. [21] used motion fields to predict the future location of the ball in soccer. Carr et al. [22] estimated the centroid of team motion using real-time player detection data to predict the future location of play for automatic broadcasting purposes.

The initial idea of aligning player trajectories based on role was proposed by Lucey et al., [23] who used a codebook of manually labelled roles. This type of approach was used to discover how teams achieved open three-point shots in basketball [24]. Bialkowski et al. [25] also used a similar approach to investigate the home advantage in soccer, and Wei et al. [26] used it to cluster different methods of how teams scored a goal. Although these works all align the multi-agent data is some form, our work differs as we learn this alignment directly from the data.

TABLE 1
Inventory of the soccer dataset used for this work.

Statistic	Frequency
Teams	20
Matches	374
Frames	21.5M
Data Points	480M
Ball Events	981K

TABLE 2
List of events annotated throughout each match.

Pass	Foul - Direct FK	Cross	Catch Drop Save
Pass Assist	Foul - Indirect FK	Cross Assist	Catch Save
Corners	Foul - Penalty	Reception	Punch
Shot on Target	Foul - Throw-in	Reception Assist	Punch Save
Shot off Target	Offside	Reception Save	Diving
Goal	Yellow Card	Catch	Diving Save
Own Goal	Red Card	Catch Drop	Drop of Ball
Neutral Clear Save	Running with Ball	Chance	Substitution
Block	Drop Kick	Pass Save	Hold of Ball
Clearance Uncontrolled	Neutral Clearance	Player Out	Clearance

3 Data: Player Tracking in Soccer

For this work, an entire season of soccer player tracking data from Prozone [27] was utilised. The data consists of 20 teams who played home and away, totalling 38 matches for each team or 380 matches overall. Six of these matches were omitted due to missing data. The 20 teams are referred to using arbitrary labels $\{A, B, \ldots, T\}$. Each match consists of two halves, with each half containing the (x, y) position of every player at 10 frames-per-second. This results in over 1 million data-points per match, in addition to the 43 possible annotated match events (e.g. passes, shots, crosses, tackles etc.). Each of these events contains the time-stamp as well as location and players involved. An inventory of the tracking data is given in Table 1, and a list of events annotated in each match is given in Table 2.

4 Discovering Formations from Data

In team sports like soccer, there is an inherent global structure that a team adheres to termed a *formation*. This is effectively a strategic concept which defines how a team distributes its players across the field in an aim to maximise their chances of winning. A formation is usually labelled in terms of defensive, midfield and attacking lines (e.g. a 4-4-2 formation has four defenders, four midfielders and two strikers) and even though the formation is usually decided before a match by the team coach, players can actively change roles during a match and how the formation is played can differ a lot from the initial plan. Detecting the played formation gives useful insight into the strategy adopted by the team and provides a template to align player tracking data, enabling clustering and role-based player analysis.

Given all the player tracking data across a season, our aim is to automatically find the formation that characterises how each team played in each match-half. Mathematically, a *formation*, $\mathcal{F}$, can be defined as an arbitrarily ordered set of N *roles*, $\{R_1,\ R_2,\dots,\ R_N\}$, which describes the spatial arrangement of N players. In this work, a "heat-map" approach in which each role is represented by a probability density function of expected location is used.

Estimating the underlying formation the team played over a match-half from player tracking data $\mathbf{D}$, is equivalent to finding the most probable set $\mathcal{F}^*$ of 2D probability density functions,

$$\mathcal{F}^* = \arg\max_{\mathcal{F}} P(\mathcal{F}|\mathbf{D}). \tag{1}$$

To begin, the 2D probability density function $P(\mathbf{X}=\mathbf{x})$ which models the tracking data $\mathbf{D}$ is considered. In other words, $P(\mathbf{x})$ represents the heat-map for an entire team. The heat-map of the entire team can be modelled as a linear combination of the heat maps for each role,

$$\begin{aligned} P(\mathbf{x}) &= \sum_{n=1}^{N} P(\mathbf{x}|n)P(n) \\ &= \frac{1}{N}\sum_{n=1}^{N} P_n(\mathbf{x}). \end{aligned} \tag{2}$$

Strategically, a team needs to spread out its players so that the entire field is adequately covered. As a result, the probability density functions of each role should exhibit minimal overlap with one another. Equivalently, each role probability density function should exhibit minimal overlap with the team's probability density function. Following the ideas of minimum entropy data partitioning [1], [2], Kullback-Lieber divergence can be employed to measure the overlap between two probability functions $P(x)$ and $Q(x)$,

$$KL(P(x)\|Q(x)) = \int P(x)\log\Big(\frac{P(x)}{Q(x)}\Big)dx. \tag{3}$$

Since divergence is a strictly positive quantity (and completely overlapping probability density functions have zero divergence), a penalty V_n is employed based on the negative divergence value between the heat map $P_n(\mathbf{x})$ of an individual role and that of the team $P(\mathbf{x})$,

$$V_n = -KL\big(P_n(\mathbf{x})\|P(\mathbf{x})\big). \tag{4}$$

Computing the optimal formation $\mathcal{F}^*$ is equivalent to determining the optimal set $\mathcal{F}^* = \{P_1(\mathbf{x}),\dots,P_N(\mathbf{x})\}^*$ of per-role probability density functions $P_n(\mathbf{x})$ that minimise the total overlap,

$$\mathcal{F}^* = \arg\min_{\mathcal{F}} V. \tag{5}$$

Substituting the expressions for KL divergence into the total overlap cost illustrates the dependence on each role-specific 2D probability density function

$$\begin{aligned} V &= \sum_{n=1}^{N} P(n)\Big(-KL\big(P_n(\mathbf{x})\|P(\mathbf{x})\big)\Big) && (6)\\ &= -\sum_{n=1}^{N} P(n)\int P_n(\mathbf{x})\log\Big(\frac{P_n(\mathbf{x})}{P(\mathbf{x})}\Big)dx && (7)\\ &= -\sum_{n=1}^{N} P(n)\int P(\mathbf{x}|n)\log P(\mathbf{x}|n)dx \\ &\quad +\sum_{n=1}^{N} P(n)\int P(\mathbf{x}|n)\log P(\mathbf{x})dx. && (8) \end{aligned}$$

The expression for V is drastically simplified when put in terms of entropy

$$H(x) = -\int_{-\infty}^{+\infty} P(x)\log(P(x))dx. \tag{9}$$

The total overlap cost, in terms of entropy, becomes

$$\begin{aligned} V &= -H(\mathbf{x}) + \sum_{n=1}^{N} P(n)H(\mathbf{x}|n) && (10)\\ &= -H(\mathbf{x}) + \frac{1}{N}\sum_{n=1}^{N} H(\mathbf{x}|n). && (11) \end{aligned}$$

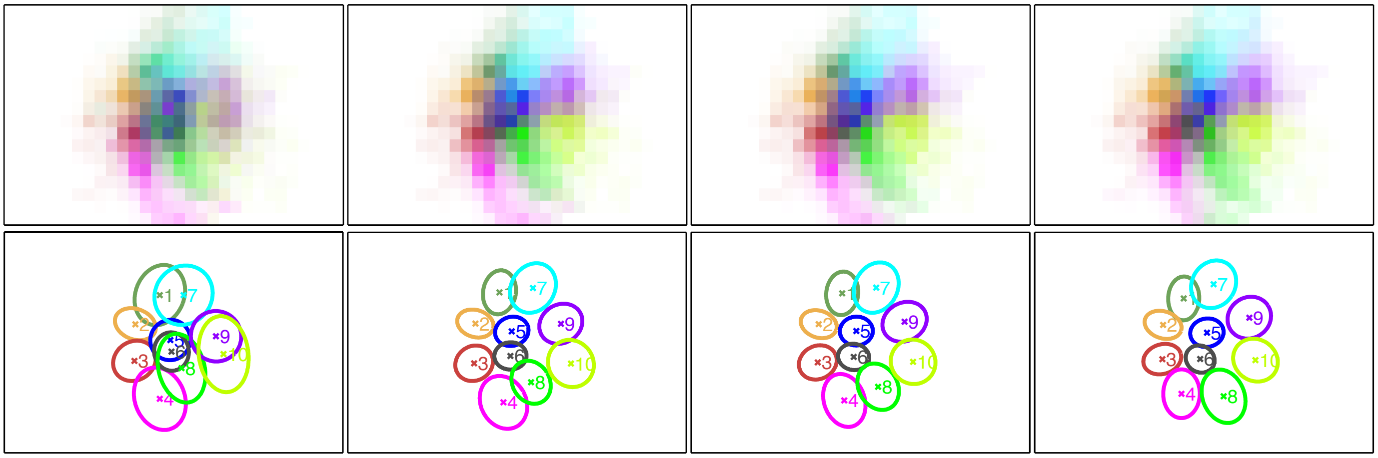
Substituting Equation 11 into Equation 5 and ignoring the constant term $H(\mathbf{x})$, the optimal formation is the set of role-specific probability density functions that minimise the total entropy

$$\mathcal{F}^* = \arg\min_{\mathcal{F}} \sum_{n=1}^{N} H(\mathbf{x}|n). \tag{12}$$

4.1 Procedure

As there is no way to solve this problem efficiently, an approximate solution can be achieved using the expectation maximisation (EM) algorithm [28]. The proposed procedure is similar to k-means clustering except with the constraint that at each frame, each player must be assigned to a unique role. Instead of assigning each data point to its closest cluster, the linear assignment cost of assigning roles to players is minimised per frame, to ensure a one-to-one assignment of roles to players.

The procedure is visually presented in Fig. 2. Firstly, the data is normalised so that teams are attacking from left to right and the effects of translation are negated by normalising the tracking data to have zero mean in each frame. This results in a formation being represented as the spatial distribution of each role relative to the team's centroid. The scale is not normalised as this can provide important information about the strategy of a team. The



(a) Initial Roles (b) Iteration 1 (c) Iteration 2 ... (d) Final Roles

Fig. 2. Example of the role discovery procedure, displaying the role distributions ($P_n(\mathbf{x})$) at each iteration, drawn relative to the team centroid in the centre of each bounding box. Each colour/number combination represents a role distribution, and these are drawn as heat-maps in the top row and as 2D Gaussians (showing mean and covariance in position) in the bottom row. The initial role distributions (a), are calculated by assuming each player is assigned a single role over all frames and taking their distribution over the half. Taking (a) as the template, each frame is assigned to these roles and the updated distributions are shown in (b). This is then used as the template for the next iteration and the procedure is repeated until convergence, resulting in well separated role distributions as in (d), which appears to be a 4-4-2 formation (four defenders, four midfielders and two attackers). The data is drawn with the team attacking left to right.

initial formation is set by arbitrarily assigning each player a unique role label at the start of the match and maintaining these roles throughout the entire duration of the tracking data. Even though there is overlap between the distributions of some players, this provides a reasonable estimate of the formation as players tend to play one role for the majority of the time. An example of the initial occupancy maps for each role are shown in Fig. 2 (a). Role labels are then assigned to the players at each frame of the tracking data by formulating a cost matrix based on the log probability of each position being assigned a particular role label. The Hungarian algorithm [29] is used to compute the optimal assignment of role labels at each frame based on the current formation template. Once role labels have been assigned to all frames of the tracking data, the probability density functions of each role are recomputed, giving an updated formation template. The process is repeated until convergence, resulting in well separated probability density functions as in Fig. 2 (d). In this way, each player is assigned to a role at each frame of the tracking data and the role probability distributions ($P_n(\mathbf{x})$) are discovered, providing the formation that the team played over the match-half.

5 Visualising and Clustering Team Formations

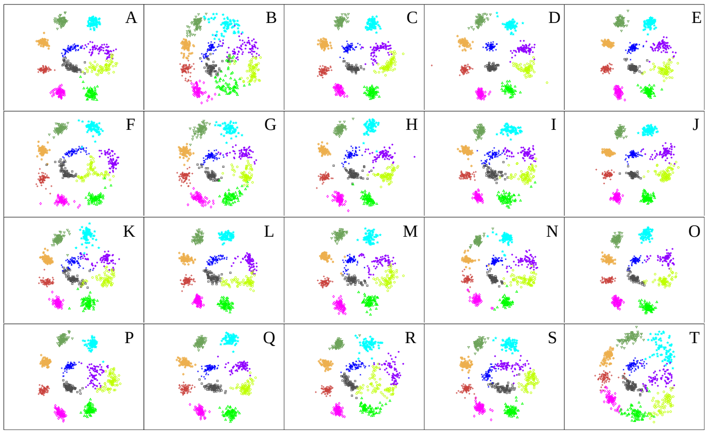
The proposed formation discovery procedure was performed for each team and match-half of the dataset in Table 1 excluding formations where players were sent off, resulting in the detection of 1411 formations. Each formation consists of a set of ten distinct role probability distributions, representing the structural arrangement of the team over a match-half.

The formations for each of the 20 teams (A-T) for every match-half are shown in Fig. 3 (a). As can be seen in this figure, most of the teams tend to play the same formation across the season with only a slight variation occurring in some of the positions. For example, only teams B and T seem to have some variation across the course of a season, while others like teams A, F, P and R only have a minor change in the midfield (i.e. one holding midfielder vs two, or playing with one striker vs two). Other than that, most teams tend to be rather staunch in what they play. The most dominant formation appears to be a 4-4-2, with some teams varying the midfield as described above. Only one team appeared to play with three defenders (team T).

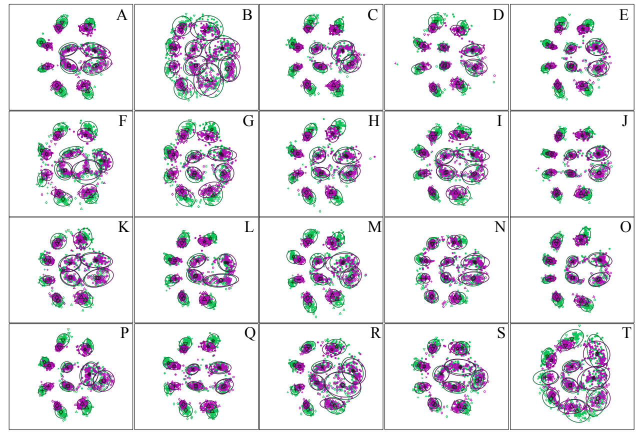
The aligned data also enables the visualisation of formations in different game states such as attacking and defending as in Fig. 3 (b). It can be seen that teams tend to keep similar structures in attacking and defending, but there is an evident spreading of the team in attacking compared to defending across all teams. We do not normalise for scale so that we are able to see such tactical variations.

5.1 Short-Term Formations

In addition to representing the long-term behaviour of the team in terms of formation or team structure, the role-aligned player tracking data can be visualised over shorter durations, to dynamically represent how a team plays throughout a match. Compared to existing statistics which only contain sparse team information (e.g. # corners, # shots, % possession), the proposed approach can represent the spatio-temporal characteristics of the match in terms of formations and position. One of the statistics which broadcasters present during a live-broadcast is the possession duration of both teams over the past 5 minutes which gives an indication of which team is dominating. While this is insightful, it does not give any information about where this is happening. Using a sliding window of 5 minutes on the role assigned player positions, the play progression can be visualised in terms of team formations using 2D Gaussians to represent the role distributions over the time window. A film-strip of this approach is shown in Fig. 4.



(a) Formations computed from the whole match-half



(b) Formations computed during attacking and defending

Fig. 3. The formations for every match-half within the season, organised by team (labelled A to T). Figure (a) represents formations computed over the whole match-half with colours representing different roles and (b) was computed for when the team was attacking (green) and defending (purple) based on ball possession. The formations are drawn so that teams are attacking from left to right. For clarity of visualisation, only the mean for each role for each match is shown instead of displaying the full distribution.

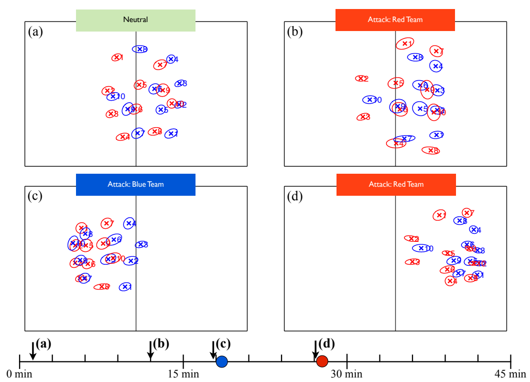


Fig. 4. Film strip representing a 45 min match-half in terms of formation (with the home team in red attacking from left to right, and circles on the timeline indicating goals). Using a sliding window of 5 mins, the progression of the match in terms of team structure and location on the field can be seen. (a) During a neutral portion of the game, it can be seen that both teams are playing a `4-2-3-1` formation. (b) Next, the red team can be seen to make an attack by spreading out and advancing its players forward. (c) Before the blue team scores, the centre midfielder (role 9) moves forward to aid in the attack. (d) In the final example, the red team scores, with the whole team positioned close to the goal.

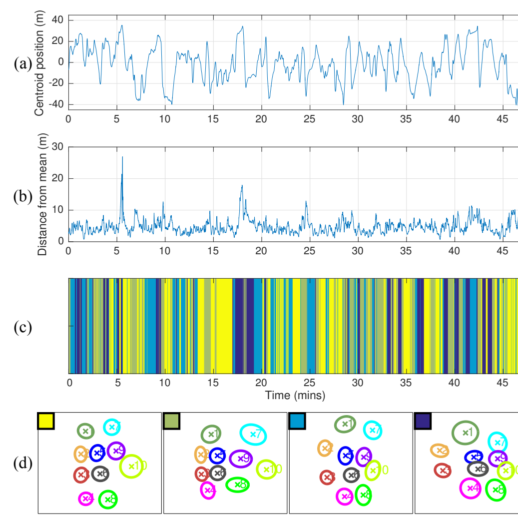


Fig. 5. Detecting variation in formation within a match. (a) The team's x-centroid (positive values indicate closer proximity to the opponent's goal), (b) The distance between the player tracking data at each frame relative to the mean formation indicating deviations from the team's mean formation, (c) The assigned role clusters at each frame relative to (d), the within-match-half formation clusters.

5.2 Within-Match Formation Variations

The proposed procedure finds the formation that best describes the *whole* match-half's tracking data. This provides a single formation template to provide a consistent spatial ordering of the player tracking data across the match-half. Given the fluid nature of team sports, the position of players and their formation will vary continuously throughout a match. Detecting specific formations is challenging due to the unsupervised nature of our approach (there aren't pre-computed templates for different types of formations). We propose two approaches to detect formation variations within a match: 1) clustering the role-aligned player positions and 2) calculating the distance of each frame to a template.

We define the mean formation, $\overline{\mathcal{F}^*}$, as the mean (x, y) location of each role distribution in the formation template,

$$\begin{aligned}\overline{\mathcal{F}^*} &= \left[\overline{P_1}(\mathbf{x}), \overline{P_2}(\mathbf{x}), \ldots, \overline{P_N}(\mathbf{x})\right]^T, \\ &= \left[(x_1, y_1), (x_2, y_2), \ldots, (x_{10}, y_N)\right]^T.\end{aligned} \tag{13}$$

Since the tracking data is aligned to the formation, similarities between different frames of the tracking data $(\mathbf{x}_t)$ can be gauged using standard distance functions such as the mean Euclidean distance between corresponding roles,

$$d(\mathbf{x}_{t_1}, \mathbf{x}_{t_2}) = \frac{1}{N}\sum_{n=1}^{N} \|\mathbf{x}_{t_1}(n) - \mathbf{x}_{t_2}(n)\|^2 . \tag{14}$$

An example of detecting formation variations within a match using the two approaches is shown in Fig. 5. We used k-means clustering on the role-ordered tracking data to detect four formation clusters within the match. Note that because formations are continuously changing and differences in formation are subjective, 4 clusters were chosen arbitrarily with k-means clustering as a proof-of-concept. We also show the deviation of each frame relative to the mean formation. Interestingly, deviations in formation coincide with close proximity to either team's goals, and especially when rapidly moving from one side of the field to the other. It can be seen that when the team moves forward to attack, the formation often changes to a more attacking formation with players moving from the defensive to the attacking line in what appears to be a 2-4-4 formation.

5.3 Clustering Team Formations

To get an indication of the types of formations used by teams across the league, agglomerative clustering was employed on the formations. In agglomerative clustering, each observation starts in its own cluster and pairs of clusters are merged based on distance, forming a cluster hierarchy. The Earth Mover's Distance (EMD) [30] was used to compute the distance between corresponding role probability densities, and the distance between formations was calculated as the sum of the distances between corresponding roles. Agglomerative clustering was chosen as it provides a flexible and non-parametric approach to discover the types of formations used across the dataset. Different clustering thresholds of the hierarchy can be observed, and a cut-off of six clusters is shown in Fig. 6, with the mean role positions of each formation assigned to the cluster overlaid over one another. Six clusters were chosen as this allows the coarse categories of formations to be visualised. Segregating further resulted in clusters that look very similar, while a smaller number had too much variation within the clusters. It can be seen that clustering resulted in the discovery of distinct formation classes - e.g. Cluster 2 and 3 have only one striker in the front, Cluster 1 and 5 have two strikers, while Cluster 4 and 6 appear to have three. Cluster 4 is the only cluster with three defenders at the back with the remainder all having four.

By observing the clustering assignment frequency (top right of each cluster in Fig. 6), we can see which formations are more commonly adopted by teams. Cluster 1, which appears to be a 4-4-2, is the most common formation with approximately 54.11% of formations being assigned to this cluster, followed by Cluster 2 (22.30%), which appears to be a 4-2-3-1. This gives insight into the strategies adopted by teams (e.g. having 2 strikers instead of 1 may be considered a more attacking strategy).

To evaluate the clustering results, the cluster groups were quantitatively comparing against ground truth formation labels. The ground truth labels were annotated by a soccer expert who annotated the most frequently observed formation for each match-half and each team according to the arrangement of players in defensive, midfield and attacking lines (4-4-2, 4-2-3-1, 4-3-3, 3-4-3, 4-1-4-1, or 'other' where the team either did not display a dominant formation or was not one of the given labels). To evaluate the results, the label of each cluster was estimated as the most frequent ground truth label within the cluster and the results are presented as a confusion matrix in Fig. 7.

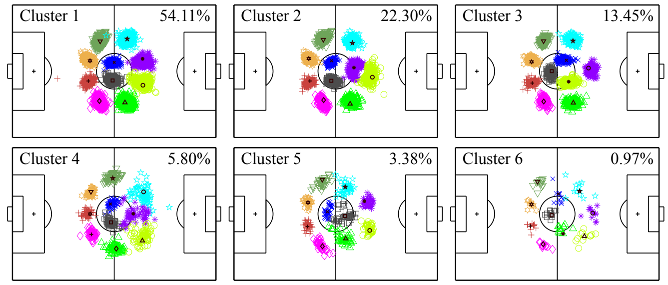


Fig. 6. Clustering results across the league of data, showing the grouping of the mean formations from every match-half into six clusters. Each of the ≈1400 formations is drawn in its corresponding cluster and the median of each cluster is overlaid in black. Each dot point represents the mean role position of a formation, with each role assigned a different colour. The percentages refer to the proportion of examples assigned to each cluster, giving an indication of which formation types are favoured by teams across the league. A preference for what appears to be a 4-4-2 formation is apparent with 54% of the data belonging to this cluster. All formations are normalised so that the team is attacking from left to right.

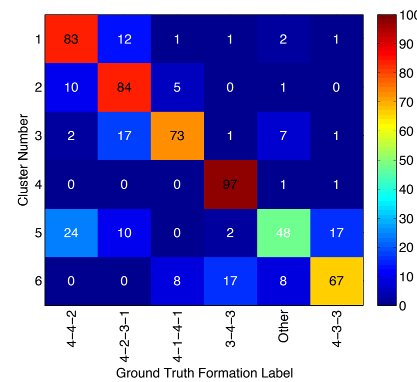


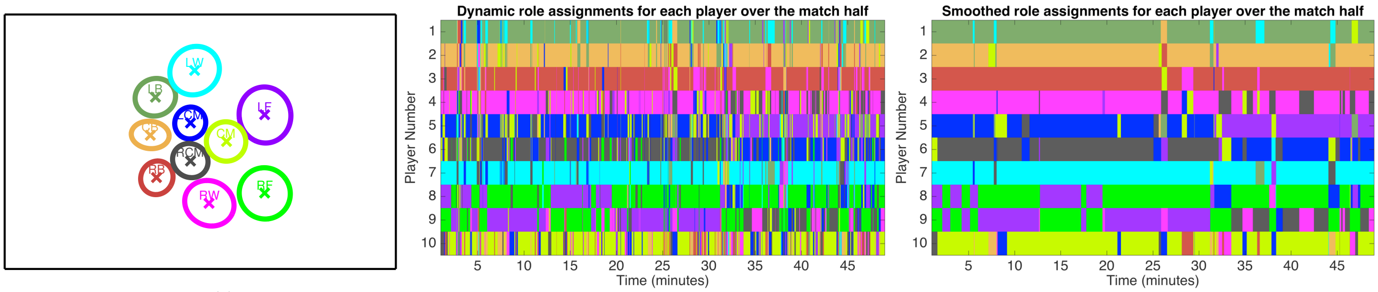
Fig. 7. Formation clustering results presented as a confusion matrix, showing the proportion of each cluster belonging to each ground truth formation label.

It can be seen from Fig. 7 that the discovered formation clusters match the ground truth annotations well, with high within-cluster label agreement and an overall correct classification rate of 75.33%. The most confusion is in Cluster 5 often being classified as a 4-4-2 and 4-3-3. On visual inspection of the misclassified examples, sometimes the formation appears in between two clusters, e.g. there is some confusion between the 4-4-2 and 4-2-3-1 formations when the second striker is positioned slightly behind the other.

5.4 Individual Player Analysis

Compared to existing analysis which often only looks at the mean behaviours of each player, the role assignment method dynamically assigns players to roles throughout a match and therefore allows the different characteristic behaviours of each player to be analysed and visualised (either across time as in Fig. 8 or by ball event as in Fig. 9).

An example of the roles of each player over a match-half relative to the discovered formation are shown in Fig. 8. This example highlights how frequently players alternate positions throughout a match and how versatile they are within the formation. In plot (b) it can be seen that role swaps on a frame-to-frame basis are very frequent. Plot (c) represents a 1 min smoothed version of the role assignments

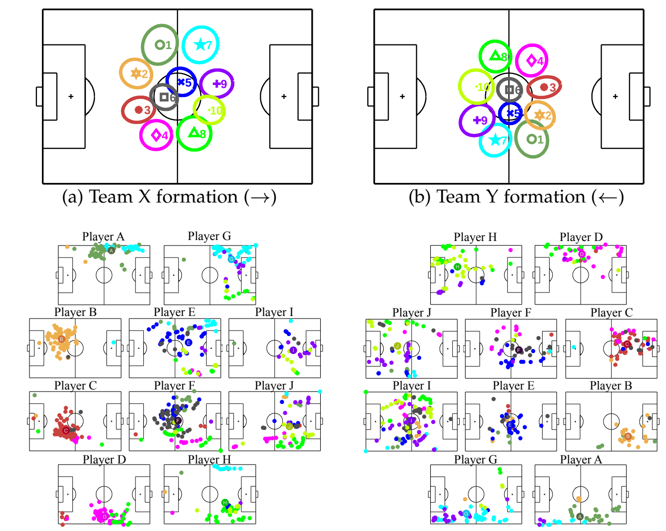


(a) (b) (c)

Fig. 8. The behaviour of a team over half a match is shown, demonstrating: (a) Their overall formation found using the proposed formation discovery procedure (with roles represented as 2D Gaussians). (b) A timeline showing the role assigned to each player at each frame, coloured by role. (c) A 1 min smoothed version of the role assignments (ignores temporary role swaps).The roles are labelled as {left-back(LB), centre-back(CB), right-back(RB), left-centre-midfield(LCM), right-centre-midfield(RCM), centre-midfield(CM), left wing(LW), right-wing(RW), left-forward(LF), right-forward(RF)}

(to ignore temporary role swaps) and shows the dominant roles taken by each player. From this, it can be seen that there are longer-term formation swaps especially between the two forwards (player 8 and 9) who alternate positions throughout the match, and there is also a large change in roles around the 32nd minute, perhaps indicating a last minute strategic variation.

Roles can also be used to provide context in analysing player events throughout a match. That is, we can know what position a player was relative to their team mates for every action they performed. In Fig. 9, all the ball events within a match-half are displayed, segmented by player identity and coloured by their role at the time of each event. On the left are the events for the team attacking left to right, and their opposition is shown on the right of the figure. The capital letter indicates the mean ball touch location for that player. For Team X (on the left), interesting behaviour can be observed for the players playing left wing and right wing who swap roles for part of the match and the role representation is able to detect these characteristic behaviours (coloured in green and cyan). If the mean of each player's actions were simply taken, this important tactical variation would be missed. The variability in roles can also be observed for each player (e.g. Team Y has much more variability in the roles that each player adopts compared to Team X).



(a) Team X formation (→) (b) Team Y formation (←)

(c) Events for each player, Team X (→) (d) Events for each player, Team Y (←)

Fig. 9. Role-context player analysis for two opposing teams, Team X (attacking →) and Team Y (attacking ←). After discovering the formation for team X and Y, shown in (a) and (b) respectively, the ball events for each player across the match can be analysed using role context. In the bottom row, each field represents a player and the position of all their ball touches throughout the match. The colour of the dots indicates their role at the time of the event relative to the team's formation shown in the top of the figure. Rather than just knowing where a player touched the ball, we know where the player was relative to their teammates which provides important contextual information.

6 Predicting Team Identity

To determine how to best represent the playing style and characteristics of a team, a series of team identity experiments were conducted on the full season of soccer data described in Section 3. The challenge was, *given only player tracking data and ball events, how can the identity of each team best be predicted?* To do this, three types of match descriptors which describe team behaviour were generated: 1) match statistics, 2) ball occupancy, and 3) team formation, as shown in Fig. 10

6.1 Match Descriptors

Match Statistics: During a match, various statistics that capture team and individual behaviour are annotated. Table 2 lists the statistics annotated in our dataset. While the number of these match statistics is quite large, the majority are quite sparse with only a couple of these events labelled per match. A few of the most important match statistics is what is traditionally reported in summarisation of matches (i.e. goals, shots on target, shots off target, passes, corners, yellow and red-cards). In the match statistics descriptor for this analysis, we compute the frequency counts of each match statistic to represent team behaviour as a vector.

Ball Occupancy: Associated with the match statistics/events are the time and location for each occurrence. To form a representation of this information, the approach used in [8], [9] was adopted which consists of estimating the continuous ball trajectory at each time-stamp by linearly interpolated between events, as well as which team had possession (ignoring stoppages). The field is then split into a

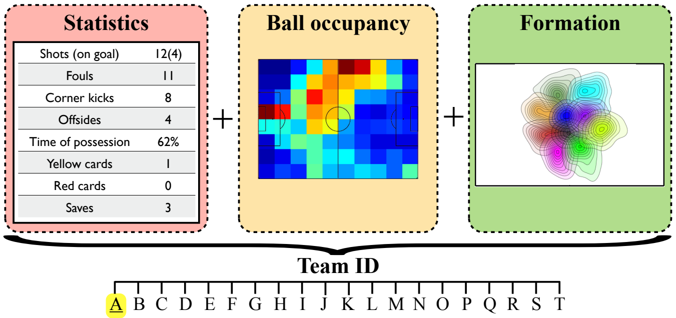


Fig. 10. Based solely on match statistics, ball movement patterns, and the formation descriptor, the identity of a soccer team can be predicted with high accuracy.

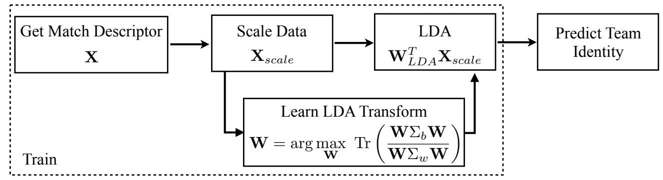


Fig. 11. Block diagram for learning the discriminative feature vector and predicting team identity. Given a match descriptor, the data is first scaled then multiplied by $\mathbf{W}^T$, found using LDA, to yield a discriminative feature vector. The LDA matrix is learnt using the team identity labels and their match descriptors in the training set. Team identity is then predicted using k-NN.

10×8 spatial grid and ball occupancy of each of these grids for each team were calculated (i.e. a vector of how often the team was in possession of the ball in each location over the match). A visualisation of a ball occupancy example is shown in Fig. 10 (centre).

Formation Descriptor: For each match-half, the formation descriptor $\mathcal{F}^*$ was found from the player tracking data using the method described in Section 4.1. This gives an $M \times N$ matrix where M refers to the number of cells in the field and N is the number of roles (set to 10, as the goal-keeper was omitted, as well as games which had a player sent off). A depiction of the formation descriptors for each team for all match-halves was presented in Fig. 3. As teams are rather rigid in the way they play across a season, it suggests that this is a useful feature in discriminating between different teams. Another interesting point is, as teams vary little in terms of playing style throughout the season, this could be used as a powerful prior for preparing against an opposition in upcoming matches.

6.2 Team Identity Experiments

The team identity experiments were performed using a "leave-one-match-out" cross-validation strategy where one match was left out to test against, and the remaining matches were used as the train set. The block-diagram in Fig. 11 summarises the procedure. Firstly, the three descriptors described above were generated and the features were linearly scaled to be in the range $[0, 1]$. To obtain a compact but discriminative representation, linear discriminant analysis (LDA) was used. LDA was selected as it explicitly models the difference between classes and helps to determine the distinguishing features of a team. The transformation matrix $\mathbf{W}$ was learnt from the training set using the team identity as the class labels (i.e. $C = 20$). Then at testing time, the features were multiplied by $\mathbf{W}^T$ to yield a lower dimensionality discriminant feature vector of dimensionality $C - 1$. To predict the identity label of the teams in the test match, a k-nearest-neighbour classifier was used with the Euclidean norm as the distance metric. A neighbourhood of $k = 10$ was chosen as this provided the best results for most descriptors, however, the order in performance of the different descriptors was consistent across various k.

The results for each descriptor is shown in Fig. 12. In the first experiment (Fig. 12(a)), it can be seen that using only match statistics is a poor indication of team identity with an overall accuracy of 17% (chance is 5%). This is expected as match statistics only contain coarse event information without any spatial or temporal information about the ball or the players. Using ball occupancy gives marginally improved performance over match statistics with an accuracy of 19% (Fig. 12(b)). This is well below the 33% obtained in previous works [8], [9]. A possible explanation of the performance difference could be due to the coarse estimation of the possession strings and the ball occupancy maps from the event data.

The most impressive performance by far is the formation descriptor which obtains over 67% accuracy, showing that teams have an underlying signal which can be encapsulated in the formation descriptor (Fig. 12(c)). While it may be obvious that using spatio-temporal data to quantify a team should be much better than using match statistics or ball occupancy information, this is not possible without alignment of the data, and no existing representations exist for summarising team's formations over matches other than simply mean player positions which doesn't reflect the true structure. Our role-alignment enables the spatio-temporal data to be utilised in this way. Combined the descriptors by concatenating all the scaled features further improves the performance to over 70% which shows there is complimentary information within the other descriptors.

6.3 Team Behaviour Across the Season

The high classification rate of 70% indicates that teams do have a characteristic "style" or match behaviour, and the given match features provide useful information for comparing and characterising teams. Here, we explore how we can use this information to observe the similarities between different teams and the variation of each team across the season.

Given a set of team behaviour descriptors, a discrete set of styles (match behaviours) can be observed using k-means clustering. We cluster the lower dimensionality feature vectors of each match (computed used LDA as in the team identity experiments) and the variation in style for each team using $k = 5$ clusters is shown in Fig. 13. Team T stands out, being in a style cluster of its own, which could be explained by the distinctly different formation from all other teams, with 3 defenders at the back (as was observed in Fig. 3). Most teams play a single style, while teams E and R vary their playing styles more frequently than other

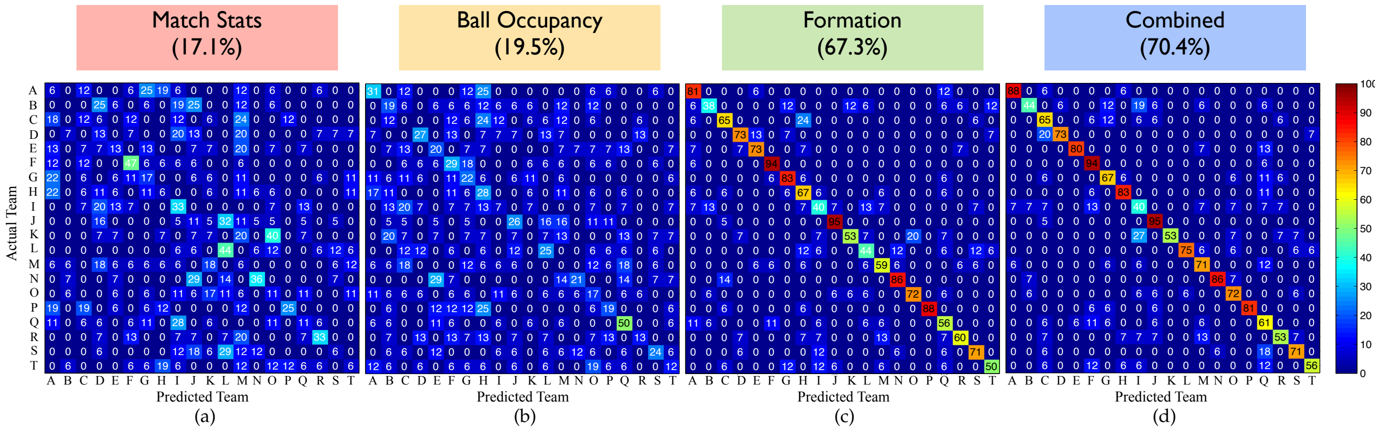


Fig. 12. Team identity results for the various descriptors: (a) match statistics, (b) ball occupancy, (c) formation descriptor and (d) fused all descriptors.

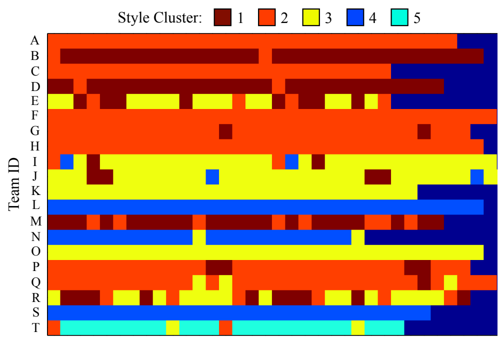


Fig. 13. Shows the variation in style each team has across a season when 5 style clusters are used. Each coloured block represents the formation style the team played for a match half and they are concatenated chronologically, excluding match halves that were missing data or had a player sent off (i.e. < 10 field players).

teams. Knowing what behaviour a team adopts in different situations can be useful in preparing for upcoming matches.

7 Summary

In this paper, a *formation descriptor* was proposed to align multi-agent data and discover team structures automatically from data. This was done by minimising the entropy of a set of player role distributions, disentangling the player tracking data into distinct role distributions to allow the discovery of the underlying team structure. This was efficiently solved using an expectation maximisation approach that simultaneously assigns players to roles throughout a match, and discovers the team's overall formation (set of role distributions). The proposed approach is completely unsupervised, and learns the spatial structure of a team directly from data. The role-alignment provides a consistent spatial ordering across the tracking data to enable a host of new group behaviour analysis tasks to be performed such as formation visualisation, large-scale formation clustering and role-based player analysis. It was shown that the method can visually summarise a game, giving an indication of dominance and tactics. Additionally, the formation descriptor was shown to represent the characteristic style of teams significantly better (3 times more) than other match descriptors typically used to describe team behaviour. The utility of the approach was demonstrated in performing large-scale individual and team analysis using a full season of data from men's professional soccer, consisting of over 21.5 million frames of player tracking data, spanning 20 teams and 374 matches.

References

[1] S. Roberts, R. Everson, and I. Rezek, "Minimum entropy data partitioning," *IET*, pp. 844–849, 1999.
[2] Y. Lee and S. Choi, "Minimum entropy, k-means, spectral clustering," in *International Joint Conference on Neural Networks*, 2004.
[3] J. Gudmundsson and M. Horton, "Spatio-temporal analysis of team sports - A survey," *CoRR*, vol. abs/1602.06994, 2016. [Online]. Available: http://arxiv.org/abs/1602.06994
[4] K. Goldsberry, "CourtVision: New visual and spatial analytics for the NBA," in *MIT Sloan Sports Analytics Conference*, 2012.
[5] R. Masheswaran, Y. Chang, J. Su, S. Kwok, T. Levy, A. Wexler, and N. Hollingsworth, "The three dimensions of rebounding," in *MIT Sloan Sports Analytics Conference*, 2014.
[6] D. Cervone, A. D'Amour, L. Bornn, and K. Goldsberry, "POINTWISE: Predicting points and valuing decisions in real time with NBA optical tracking data," in *MIT Sloan Sports Analytics Conference*, 2014.
[7] A. Miller, L. Bornn, R. Adams, and K. Goldsberry, "Factorized point process intensities: A spatial analysis of professional basketball," in *ICML*, 2014.
[8] P. Lucey, A. Bialkowski, P. Carr, E. Foote, and I. Matthews, "Characterizing multi-agent team behavior from partial team tracings: Evidence from the English Premier League," in *AAAI Conference on Artificial Intelligence*, 2012.
[9] P. Lucey, D. Oliver, P. Carr, J. Roth, and I. Matthews, "Assessing team strategy using spatiotemporal data," in *ACM SIGKDD*, 2013.
[10] J. Gudmundsson and T. Wolle, "Football analysis using spatio-temporal tools," *Computers, Environment and Urban Systems*, 2013.
[11] J. Peña and H. Touchette, "A network theory analysis of football strategies," *arXiv preprint arXiv:1206.6904*, 2012.
[12] X. Wei, P. Lucey, S. Morgan, and S. Sridharan, "Sweet-spot: Using spatiotemporal data to discover and predict shots in tennis," in *MIT Sloan Sports Analytics Conference*, 2013.
[13] X. Wei, P. Lucey, S. Morgan, and S. Sridharan, "Predicting shot locations in tennis using spatiotemporal data," in *DICTA*, 2013.
[14] S. Intille and A. Bobick, "Recognizing planned, multi-person action," *Computer Vision and Image Understanding*, vol. 81, pp. 414–445, 2001.
[15] G. Zhu, Q. Huang, C. Xu, Y. Rui, S. Jiang, W. Gao, and H. Yao, "Trajectory based event tactics analysis in broadcast sports video," in *International Conference on Multimedia*. ACM, 2007.
[16] S. Jiang, Q. Huang, and W. Gao, "Mining information of attack-defense status from soccer video based on scene analysis." in *International Conference on Multimedia and Expo (ICME)*. IEEE, 2007.

[17] 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.

[18] J.-C. Bricola, "Classification of multi-agent trajectories," Master's thesis, EPFL, 2012.

[19] 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*, vol. 32, no. 2, 2011.

[20] I. Atmosukarto, B. Ghanem, S. Ahuja, K. Muthuswamy, and N. Ahuja, "Automatic recognition of offensive team formation in American Football plays," in *CVPRW*, 2013.

[21] 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.

[22] P. Carr, M. Mistry, and I. Matthews, "Hybrid robotic/virtual pan-tilt-zoom cameras for autonomous event recording," in *ACM Multimedia*, 2013.

[23] 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.

[24] P. Lucey, A. Bialkowski, P. Carr, Y. Yue, and I. Matthews, "How to get an open shot: Analyzing team movement in basketball using tracking data," in *MIT Sloan Sports Analytics Conference*, 2014.

[25] A. Bialkowski, P. Lucey, P. Carr, Y. Yue, and I. Matthews, "Win at home and draw away: Automatic formation analysis highlighting the differences in home and away team behaviors," in *MIT Sloan Sports Analytics Conference*, 2014.

[26] X. Wei, L. Sha, P. Lucey, S. Morgan, and S. Sridharan, "Large-scale analysis of formations in soccer," in *DICTA*, 2013.

[27] Prozone, www.prozonesports.com.

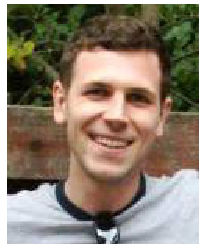
[28] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[29] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

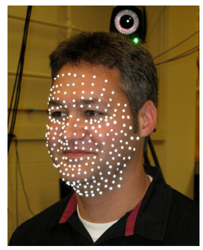
[30] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *IJCV*, 2000.

Alina Bialkowski is a computer vision researcher at the University College London. She received her PhD in the field of computer vision in 2015 at the Queensland University of Technology (QUT), Australia, where she also received a BEng (Electrical Engineering). During her doctoral studies, she worked with Disney Research Pittsburgh where she developed algorithms and tools to automatically monitor and analyse team sports. Her research interests are in feature learning, modelling and visualising large sets of visual and spatio-temporal data.

Patrick Lucey is currently the Director of Data Science at STATS. His charter is to maximize the value of fine-grained player tracking data currently captured in high-performance sports. Previously, Patrick was at Disney Research for 5 years, where he conducted research into automatic sports broadcasting using large amounts of spatiotemporal tracking data. Previous to that, he was a Postdoctoral Researcher at the Robotics Institute at Carnegie Mellon University conducting research on automatic facial expression recognition. Patrick received his BEng(EE) from USQ and his PhD from QUT, Australia in 2003 and 2008 respectively. He has won best paper awards at INTERSPEECH (2007) and WACV (2014) international conferences. His main research interests are in artificial intelligence and interactive machine learning in sporting domains.



Peter Carr is a Senior Research Engineer at Disney Research, Pittsburgh. His current work focuses on realtime computer vision algorithms with applications in robotics, machine learning and multi-object tracking. He received his PhD from the Australian National University in 2010, under the supervision of Prof. Richard Hartley. Peter received a Master's Degree in Physics from the Centre for Vision Research at York University in Toronto, Canada, and a Bachelor's of Applied Science (Engineering Physics) from Queen's University in Kingston, Canada.



Iain Matthews is a Principal Research Scientist and the Associate Research Director at Disney Research in Pittsburgh. He received a BEng degree in electronic engineering in 1994, and a PhD in computer vision in 1998, from the University of East Anglia, UK. He was Systems Faculty in the Robotics Institute at Carnegie Mellon University until 2006 working on face modelling and vision based tracking. He spent two years at Weta Digital, NZ, as part of the team that developed the facial motion capture system for the movies Avatar and Tintin. In 2008 he joined the newly formed Disney Research in Pittsburgh as a Senior Research Scientist leading the computer vision group. He holds an adjunct faculty appointment in the Robotics Institute at CMU. Iain is a member of the IET, IEEE and ACM.

Sridha Sridharan Professor Sridha Sridharan has a BSc (Electrical Engineering) degree and obtained a MSc (Communication Engineering) degree from the University of Manchester, UK and a PhD degree from University of New South Wales, Australia. He is currently with the Queensland University of Technology (QUT) where he is a Professor in the School of Electrical Engineering and Computer Science. Professor Sridharan is the Leader of the Research Program in Speech, Audio, Image and Video Technologies (SAIVT) at QUT, with strong focus in the areas of computer vision and machine learning. He has published over 500 papers consisting of publications in journals and in refereed international conferences in the areas of Image and Speech technologies during the period 1990-2016. During this period he has also graduated 60 PhD students in the areas of Image and Speech technologies.

Clinton Fookes is a Professor in the School of Electrical Engineering & Computer Science in the Science & Engineering Faculty of the Queensland University of Technology in Brisbane, Australia. He holds a BEng (Aerospace/Avionics), an MBA and a PhD in the field of computer vision. Clinton actively researches in the fields of computer vision and pattern recognition including video surveillance, biometrics, human-computer interaction, airport security and operations, and complex systems. Clinton has attracted over $15M of cash funding for fundamental and applied research from external competitive sources and has published over 140 internationally peer-reviewed articles. He has been the Director of Research for the School of Electrical Engineering & Computer Science. He is currently the Discipline Leader for Vision & Signal Processing. He is a Senior Member of the IEEE. He is also an Australian Institute of Policy and Science Young Tall Poppy and an Australian Museum Eureka Prize winner.