Large-Scale Analysis of Formations in Soccer

***Abstract*—Due to the demand for better and deeper analysis in sports, organizations (both professional teams and broadcasters) are looking to use spatiotemporal data in the form of player tracking information to obtain an advantage over their competitors. However, due to the large volume of data, its unstructured nature, and lack of associated team activity labels (e.g. strategic/tactical), effective and efficient strategies to deal with such data have yet to be deployed. A bottleneck restricting such solutions is the lack of a suitable representation (i.e. ordering of players) which is immune to the potentially infinite number of possible permutations of player orderings, in addition to the high dimensionality of temporal signal (e.g. a game of soccer last for 90 mins). Leveraging a recent method which utilizes a ”role-representation”, as well as a feature reduction strategy that uses a spatiotemporal bilinear basis model to form a compact spatiotemporal representation. Using this representation, we find the most likely formation patterns of a team associated with match events across nearly 14 hours of continuous player and ball tracking data in soccer. Additionally, we show that we can accurately segment a match into distinct game phases and detect highlights. (i.e. shots, corners, free-kicks, etc) completely automatically using a decision-tree formulation.**

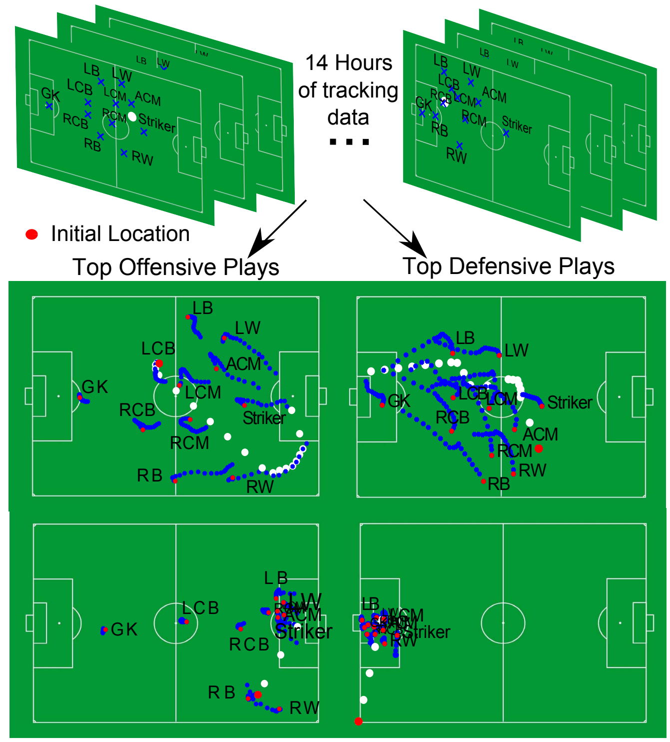

Fig. 1. (Top) Given a large amount of player and ball tracking data of a particular team, in this paper we devise a method which can discover (bottom) the top offensive and defensive plays a team utilizes based on a ”role-representation” (e.g. left-wing (LW), right-back (RB), Striker (ST)).

I. Introduction

To go beyond current analysis and gain an advantage over their competitors, many sporting organizations have recently looked to use tracking technologies which can locate the position of the ball and players at each time instant in professional leagues [10], [25], [28], [32]. Even though there is potentially an enormous amount of hidden team behavioral information to mine from such sources, due to the sheer volume as well as the noisy and variable length of the data, methods which can adequately represent team behaviors are yet to be developed. The most troubling issue is that of ”permutations”, and forming a representation which is immune to this problem. For example, given 11 players on a soccer[1] team, there exists 11! (or over 39 million) different possible player orderings. If we include the opposition, then there exists $(11!)^2$ (or $> 1.5 \times 10^{15}$). Obviously, in practice, players normally maintain their positions for most of the match but even with a couple of positional swaps, the number of permutations explodes making feature comparisons and team modeling prohibitive. Additionally, the representation has to deal with issue of player substitution and generalize across different teams. To overcome this bottleneck, recent research has looked to using a “role-representation” which essentially finds the permutation matrix at each time-step to maintain feature correspondences over time [21].

Once the issue of permutations have been resolved, the next challenge is to find meaningful patterns from the large volumes of data which can uncover common patterns of a team’s play. However for sports like soccer, this is very challenging as the low-scoring and continuous nature make it very hard to associate segments of play with high-level behaviors (i.e. annotating tactics/strategy/style/systems of play is very subjective and unreliable). However, as sports have very clear goals and objectives, we can condition our analysis on these objectives to do meaningful analysis. For example in soccer, the clear objective of a team is to score more goals than the opposition. Even though other latent variables are present (i.e. passing patterns, defensive assignments etc.), events such as shots on goal, corners, free-kicks – both from an offensive and defensive point-of-view – are probably the most important events to analyze. Consequently, from a planning perspective,

[1]Even though it is most commonly called “football” around the world, we refer to “Association Football” as soccer in this paper to avoid confusion with other football codes.

having an automated tool which can cluster similar team behaviors associated with important events such as shots would be a very useful tool for a coach/commentator/analyst.

In this paper, to incorporate the temporal evolution of the play, we segment goal-scoring opportunities by going back T seconds from a shot. Given these temporal sequences, we employ a feature reduction strategy to find a low dimensional approximation of the data. Using the compact representation based on a bilinear spatiotemporal basis model, we then use an $L\infty$ matching-pursuit exemplar-based clustering algorithm to find the most likely methods a team scores and concedes across 9 whole matches, or 14 hours of tracking data (see Figure 1). To enable our large-scale analysis of team behaviors, we first need to automatically detect these events of interest. Using just the player and ball information we show that we can achieve close to perfect detection using a decision forest formation.

Our specific contributions of this work are: 1) Using a role representation, we employ a feature reduction strategy to form a compact representation using a DCT-based temporal basis, 2) Devise a method which uses decision-forests to automatically segment a soccer game into distinct game-events, and 3) Discover a team's most probable methods of scoring and conceding across 9 entire games of soccer.

II. Literature Review

With the recent improvement in vision-based tracking technologies, spatiotemporal data has been used extensively in the visualization of sports action. Examples include systems which track baseball pitches in Major League Baseball [31], and ball and players in basketball and soccer [28], [32]. Hawk-Eye deploy vision-based systems which track the ball in tennis and cricket, and is often used to aid in the officiating of these matches in addition to providing visualizations for the television broadcasters [10]. Partial data sources normally generated by human annotators such as shot-charts in basketball and ice-hockey are often used for analysis, as well as passing and shot charts in soccer [25], and field hockey [33].

As the problem of fully automatic multi-agent tracking from vision-based systems is still an open one, most academic research is still centered on the tracking problem [3], [7], [26]. The lack of fully automated tracking approaches has limited team behavioral research to works on limited size datasets. The first work which looked at using spatiotemporal data for team behavior analysis was conducted over 10 years ago by Intille and Bobick [12], [13]. In this seminal work, the authors used a probabilistic model to recognize a single football play from hand annotated player trajectories. Since then, multiple approaches have centered on recognizing football plays [18], [19], [30], [34], but only on a very small number of plays (i.e. 50-100). For soccer, Kim et al. [16] used the global motion of all players in a soccer match to predict where the play will evolve in the short-term. Zhu et al. [38] analyzed tactics in soccer matches by building multiple trajectories using analysis of spatiotemporal interactions. In basketball, Perse et al. [27] used trajectories of player movement to recognize three types of team offensive patterns. Masheswaran et al. [23] use a data-driven approach to predict the location of rebounds given the incoming shot. Morariu and Davis [24] integrated interval-based temporal reasoning with probabilistic logical inference to recognize events in one-on-one basketball. Hervieu et al. [11] also used player trajectories to recognize low-level team activities using a hierarchical parallel semi-Markov model. More recently, Wei et al. [36] looked at predicting tennis shots using Hawk-Eye tracking data. Lucey et al. [20], [22] used ball-tracking data to discriminate team's playing style in soccer. Atmosukarto et al. [5] were able to detect the line of scrimmage for plays in American Football, and the type of player formation the offensive team takes on. Bialkowski et al. [6] recognized activities from noisy player detections. Wang et al. [35] addressed the problem of ball tracking in team sports by formulating the tracking in terms of deciding which player, if any, owns the ball at any given time. In a recent paper [21], they tackle the problem of infinite number of possible "permutations" by finding the permutation matrix at each time-step to maintain feature correspondence dependent on the formation of a team, which can greatly reduce the possible decision space.

In the vision community, the area of large-scale retrieval and recognition has gained a lot of attention recently. Most of the efforts have centered on retrieving/recognizing specific objects from millions of images. As comparing against every training example is prohibitive, a common method of circumventing this issue is to use hash-tables, where given the initial hash-key a local sub-space or neighborhood of potential candidate is searched. Techniques such as Locality Sensitive Hashing (LSH) have worked reasonably well for this problem [29], [37]. The key idea behind this method is to use different hash functions to ensure the probability of collision is much higher for objects that are close to each other. Compression techniques such as Product Quantization (PQ) [8], [15] and hybrid approaches have also worked well [14].

In this paper, we leverage the recent "role-representation" utilized in [21] and the idea of hash-tables used in large-scale analysis. We do this by first detecting the context of the match by automatically segmenting the match into discrete game-states, and using these as our hash-keys we do large-scale analysis of team formations over time based on these game-states.

III. Automatic Match Segmentation

A. Data

To enable our research, we utilized the (x, y) positions of both players and ball across 9 complete matches from a top-tier European soccer league (See Figure 2 for an example). The fidelity of the data is at the centimeter level, and was sampled at 10 fps. In each of these 9 matches, one team was constant while the opposition was different. These matches also had associated event-level data (see Table I). For our analysis, we focussed our analysis on the one constant team (i.e. our analysis was independent of the opponent).

Fig. 2. An example of the tracking data for both player and ball. Player positions are shown with their assigned role (discussed in Section IV.A.). In this work, we focussed on a team which used a 4-3-3 formation.

B. Methodology

Given the spatiotemporal tracking data, our goal was to automatically segment a match into distinct game phases (i.e. in-play, stoppages etc.) as well as important events (i.e. shots, free-kicks, corners). As these segments and events are normally annotated by a human, having an automatic method of doing this can alleviate this burden from a human, who can then focus on higher-level tasks such as strategy analysis. Additionally, it allows for deeper automatic analysis such as play clustering as this requires events to be initially segmented/detected.

Current approaches to this task, have employed rules that model the typical pattern of features within particular sport events [9]. These rules are mainly based on manual observation and heuristic knowledge. Such methods may be able to solve the problem in one specific sport but can not be generically applied to others. In our approach, we wish to use a machine-learning approach which learn a set of classifiers to automatically detect events of interest, making this method less subjective. To tackle the problem, we propose a two-layer hierarchical approach. In the top layer, a classifier is trained to break a match into the following segments: 1) *in-play*, 2) *stoppages*. *In-play* refers to when the match is being continuously played, while *stoppages* refers to when the game has stopped due to various reasons. These include times when the ball is out, fouls, player-injury, substitution, etc.

In the second layer, we further split the *stoppage* phase into: 1) Out-for-corners, 2) Out-for-goal-kicks, 3) Foul – Free-kicks, and 4) Out-for-throw-in. Other game phases such as player substitution and player injury are ignored since player substitution looks very similar to *out-for-throw-in* and there are not enough examples for player injury. *In-play* phases are further classified into 1) Highlights, and 2) Non-highlights. Highlights refer to all goal opportunities (both offensive and defensive). The complete segmentation scheme can be found in Figure 3.

We based these labels on the likely motion patterns of teams during these periods. For example, instead of annotating corners/free kicks as part of an *in-play* phase, we labeled it as a *stoppage* as it represents the period in a game where a team breaks from their normal formation into a "set-piece" formation. As this change occurs when the ball goes out or the referee calls for a free-kick, we label these phases when this occurs until ball is next played.

C. Classifier Training

After defining game phases, we need to train these classifiers. The first step is to collect training samples. By leveraging event labels from the match-data, we are able to extract play segments from 9 matches for each game phase. The number of plays for each game phase is listed in Table II.

Training the classifier is still a challenging task since each play has a variable length. Many approaches have employed a Dynamic Time Warping [27] technique to find an optimal match between given sequences (e.g. time series). Although such method can be used to normalize the length of sequences, it may affect/change the property of original sequence (e.g. the speed is an important attribute of a play). Alternatively, one can divide sample into small equal length chunks and use these chunks as training data. We choose the latter since we think speed is an important discriminative feature for our task especially for separating in-play and stoppages.

TABLE I
FREQUENCY OF MATCH EVENTS.

Event	Occurence
Pass	4397
Cross	192
Goal	9
Out for Corner	112
Out for Goal-kick	78
Out for Throw-in	435
Clearance	450
Offside	25
Substitution	42
Foul-free-kick	188
Shot on target	73
Shot not on target	34

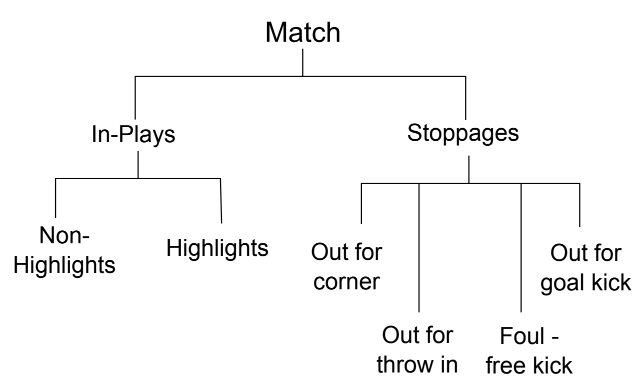

Fig. 3. The complete segmentation scheme. In the top layer, a match is segmented into *in-play* and *stoppages*. In the second layer, stoppages are further split into four sub-phases and in-play is segmented into highlights and non-highlights. Highlights refer to all goal opportunities (both offensive and defensive)

TABLE II
NUMBER OF PLAYS AND SEGMENTS FOR EACH GAME PHASE

Game Phase	Num of Plays	Num of Segments
Stoppages	965	8315
In-Plays	199	9946
Out for Corners	112	1426
Out for Free-Kicks	185	2377
Out for Goal-Kicks	169	1923
Out for Throw-in	435	2518
Highlights	106	106
Non-Highlights	235	1784

After normalizing the sample length, we randomly select 4000 samples from each class to train the top layer classifier. Random selection can avoid bias of the data. For example, if training data is all from the first half or same opposition team, the result could be inaccurate. In the second layer, 1000 samples are selected for training for each class. Finally, a decision-forest is used to classify game phases. We test its performance over different features: 1) ball position only, 2) centroid of team position, and 3) all player positions.

D. Parameter Selection

To find training samples for a highlight detector, we segment goal-scoring opportunity by going back 10 seconds from the shot. This length is decided by manually watching 50 examples of play and taking the average. Choosing a suitable chunk length is critical for classifier learning. If the length is too small, classifier may lose useful temporal information. If the length is too long, it may include irrelevant information. For Non-Highlight Vs Highlight detector, since highlights all have a same length of 10 seconds, 10 seconds is a suitable chunk length to select. For in-play vs stoppage and sub-phases within stoppage, the selection is more challenging since they all have a variable length. As most events from stoppage phase is ranging from 3 to 9 seconds, we test our classifiers using chunk length from 1 seconds to 5 seconds. Average percentage accuracy is used to measure the performance (See Figure 6). After experiments, 4 seconds are chose as the final length as it gives the best result.

E. Result

To evaluate our method, we randomly select 800 testing samples for each class. Testing samples are selected with no overlap of training data. The result shows that a decision tree classifier can achieves an excellent performance (See Figure 4). Ball position turns out to be the best feature compared to player positions and centroid of the team. In the task of classifying in-plays and stoppage, ball position can achieve almost perfect detection. Player positions is the second best feature which achieves an average percentage accuracy around 90%

In second layer, when segmenting stoppage state, ball position is still the best discriminative feature especially for free-kick's and throw-in's (See Figure 5 for detail). Without the ball, player positions can achieve an average percentage accuracy of 90.75%. Highlight detection is the most challenging task (See Figure 7) which archives 80.5% using only ball position, 75% using player positions and 60.5% using the centroid.

IV. LARGE-SCALE ANALYSIS OF TEAM BEHAVIOR

In order to enable large-scale analysis, we first find a compact representation of the spatiotemporal data which allows efficient and meaningful feature comparison. Then, using the representation, we present a $L\infty$ matching-pursuit exemplar-based clustering method which can find a team's most probable method of scoring, conceding, corners and free-kicks over a large scale of data.

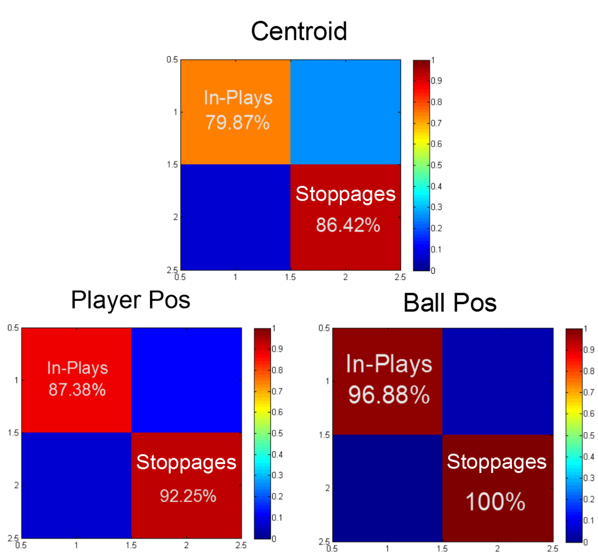

Fig. 4. Plot showing the *in-play* VS *stoppages* segmentation result using different features. Row represents actual class while column is refereed to actual class

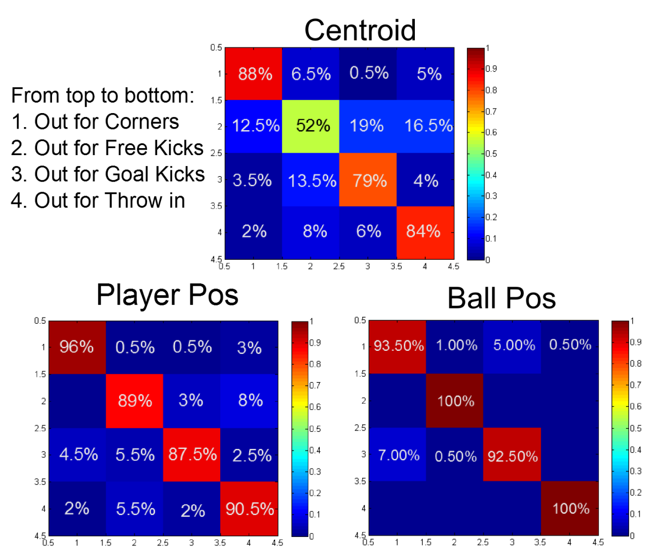

Fig. 5. Plot showing the segmentation result of sub phases within *Stoppages* phase using different features.

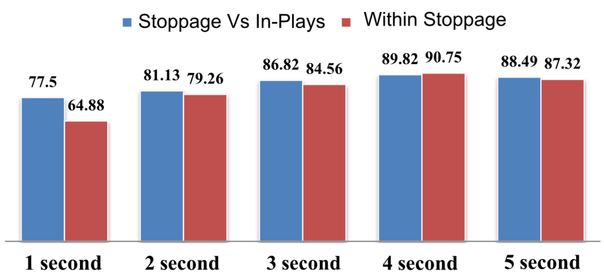

Fig. 6. Averaged Classification result with different chunk length using team position feature.

A. Role Assignment

Recently, researchers have looked to using a "role-representation" [21] to analyze team behavior. Such representation is immune to player substitution and the explosion of permutations which allows meaningful feature comparisons. A role representation is more meaningful since formations are defined by roles and individual responsibilities, not identities (e.g. names or jersey numbers). For example, in a 4-3-3 formation, roles are defined as $\mathcal{R} = \{$left back, right back, left center back, right center back, left center midfield, right center midfield, left wing, right wing, attacking center midfield, striker, goal keeper$\}$

Given player positions $\mathbf{p}_t^\tau = [x_1, y_1, x_2, y_2, ..., x_P, y_P]^T$ of P players in team τ at each time instant t. The goal here is to find the permutation matrix, $\mathbf{x}_t^\tau$, which give us a vector of role positions: $\mathbf{r}_t^\tau = \mathbf{x}_t^\tau \mathbf{p}_t^\tau$. Note here, each element $x_t(i,j)$ is a binary variable, and every column and row in x_t must sum to one. If $x_t(i,j) = 1$ then player i is assigned role j.

In [21], a four stage[2] approach is proposed to tackle the problem. First of all, a-state-of-the-art player detector is employed to detect player positions at each timestamp. For each observation, it provides player position (x, y), a timestamp t and a team affiliation estimate $\tau \in (\alpha, \beta)$. If any player is miss detected, an algorithm will be used to infer the position of that player based on spatiotemporal correlations.

[2]The first two steps are skipped in this research since we are dealing with tracking data

The role assignment task is formed as an optimization problem where the goal is to minimizes the square L_2 reconstruction error.

$$\mathbf{x}_t^{\tau*} = \underset{\mathbf{x}_t^\tau}{\text{argmin}} ||\hat{\mathbf{r}}^\tau - \mathbf{x}_t^\tau \mathbf{p}_t^\tau||_2^2 \tag{1}$$

This is a linear assignment problem where the cost of each entry is:

$$\mathbf{C}(i,j) = ||\hat{\mathbf{r}} - \mathbf{p}_t(j)||_2 \tag{2}$$

To solve the assignment problem, the mean formation is used for initialization as the team should maintain this basic formation in most circumstances. In [21], the mean formation is found by taking the mean value of $200,000$ frames of human annotated data. Finally, the optimal solution can be found using the Hungarian algorithm [17].

B. Data Compression

In team sports like soccer, a team's formation is guided by tactics designed by the coach/manager. Due to synchrony of motion between players (i.e. moving forward and backward with respect to when they are attacking and defending), we anticipated that there was heavy redundancy in the signal. In the spatial domain, a player's position is conditioned on their team-mates, opponents and the ball, as well as the dimensions of the field and rules (i.e. offside). In the temporal domain, player movements are governed by physical limits, such as acceleration.

Given n frames of role data, each frame has x, y positions for 11 players and a ball. The total dimension is $12 \times 2 \times n = 24 \times n$. The task here is to find a low dimensional approximation which can captures and exploits the dependencies across both the spatial and temporal dimensions . Recently, Akhter et al. [1] has proposed a method named bilinear spatiotemporal basis model. In his method, a time-varying signal can be factorized/modeled using two orthogonal basis functions. The idea is, given a time-varying signal, spatially, the 2D formation or shape at each time instance can be represented the as a linear combination of a small number of shape basis vectors b_j weighted by coefficients $\mathbf{s}^i = \sum_j \omega_j^i b_j^T$. Temporally, the trajectory of a particular point can also be represented as a linear combination of trajectory basis vectors, θ_i as $\mathbf{s}_j = \sum_i \alpha_i^j \theta_i$, where α_i^j is the coefficient weighting each trajectory basis vector. Then the spatiotemporal signal can be factorized using both shape and trajectory bases linked together by a common set of coefficients.

In our problem domain, spatially, we have positions of 11 players and a ball at each time instant. However, not all players are involving in every event. For example, a goalkeeper may not play an important role on an offensive play. To find the

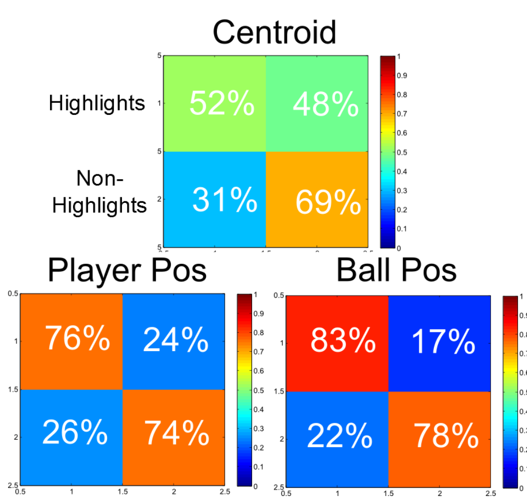

Fig. 7. *Highlights* VS *Non-Highlights* segmentation result using different features

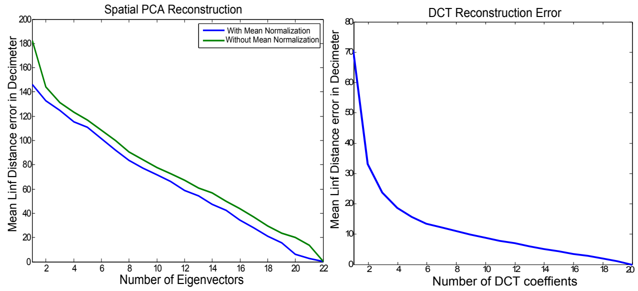

Fig. 8. Left plot showing the reconstruction error as a function of the number of eigenvectors used to reconstruct the signal using the $L\infty$ norm for non-mean and mean-normalized features. Right plot showing the reconstruction error as a function of the number of DCT coefficient.

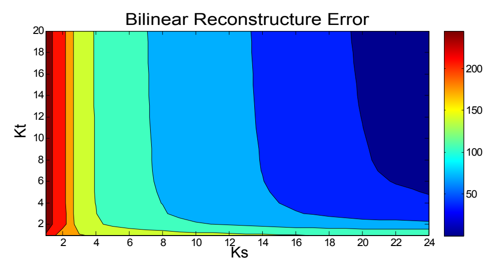

Fig. 9. The mean $L\infty$ distance error of the bilinear spatiotemporal basis model. Y axis refers to number of DCT coefficient while X axis refers to number of eigenvectors.

low dimensional manifold, we first select PCA as the basis function in the spatially domain. Figure 8(left) illustrates the compressibility of the signal in the spatial domain using PCA. We test its compressibility on both mean normalized formation and unnormalized formation.

Temporally, due to the high temporal regularity present in almost all human motion, the Discrete Cosine Transform (DCT) has been found as a suitable temporal basis for trajectories of faces [1], [2] and bodies [4]. A significant dimensionality reduction can be obtained since the human motion in a short period is quite smooth. In Figure 8(right), we see that in an offensive play, a 20 frame tracking data can be effectively represented using coefficient of a fourth order DCT with an maximum error less than 2 meters. The figure also suggests that team formation in temporal domain is more compressible than spatial domain and mean normalization can only slightly improve the compressibility.

When combining PCA and DCT in bilinear model, figure 9 shows the reconstruction error with various K_t and K_s (K_t refers to the number of coefficient in temporal domain). If we select $K_t = 5$ and $K_s = 20$ with an maximum error around 5 meters, we can reduce the dimension of the data from 480 to 100. This means a reduction of over 4 times and better compressibility can be achieved in longer plays.

Alternatively, because in spatial domain, formation data is not very compressible, one can apply the basis model only to temporal domain. Instead of using PCA, we can simply pick the centroid of a team and the ball position. This will make the data more compressible ($K_s = 2$) which allows efficient clustering. But original data can not be reconstructed back.

C. L-Infinity clustering

Once the representation is resolved, the next challenge is to find meaningful patterns from the large volumes of data which can uncover common patterns of a team's play. For soccer, the clear objective of a team is to score more goals than the opposition. Even though other latent variables are present (i.e. passing patterns), event such as shots on target, shot off target, corner, free-kick (both offensive and defensive) are probably the four most important events to analyze. Particularly, we wish to find a team's most likely patterns in those four events.

To tackle the problem, we first use our proposed event detector (see previous section) to find the frame index of these event. After that, we segment goal-scoring opportunities by going back 10 seconds from the shot and we segment corners/free-kicks by taking 10 seconds after a corner/free kick is triggered. After the data has been prepared, we employ to an unsupervised clustering method to uncover patterns of a team's play. K-Means algorithm is a widely used method for unsupervised clustering. In [21], K-means is used to find the top N formations or tactics in Hockey. However, K-means clustering requires a good initialization. If K is not chosen properly, dissimilar data may be clustered into one cluster. To avoid this situation, we use a hierarchical clustering method which is similar to matching-pursuit but based on examples and not a linear combination of examples[3].

As the dimensionality of the 10 second spatiotemporal feature is large, which can effect the clustering algorithm, we first compress the feature. After that, we treat each compressed feature as a cluster. We then conduct the cluster as follows:

1) Randomly select a cluster;
2) Compute the distance between each clusters;
3) If there exist at least one cluster with a distance less than the threshold t, go to step 4; else go to step 7;
4) Find the nearest cluster from the selected cluster;
5) Merge two clusters and update the centroid k, this merged cluster become new selected cluster;
6) Repeat step 3-6;
7) Store the merged cluster, remove this merged cluster from the set;
8) If there is any cluster left, repeat step 3-7; else finish clustering;

D. Parameter Selection

When clustering, parameters are specified as following:

1) Centroid of a cluster is calculated using the median value;
2) The length of each play is selected as 10 seconds long (100 frames);
3) $L\infty$ norm is used to calculate the distance between clusters. We chose the $L\infty$ norm instead of the L_2

[3] Due to rules of the game, a linear combination may result in situations which can not exist so specific match examples are preferred.

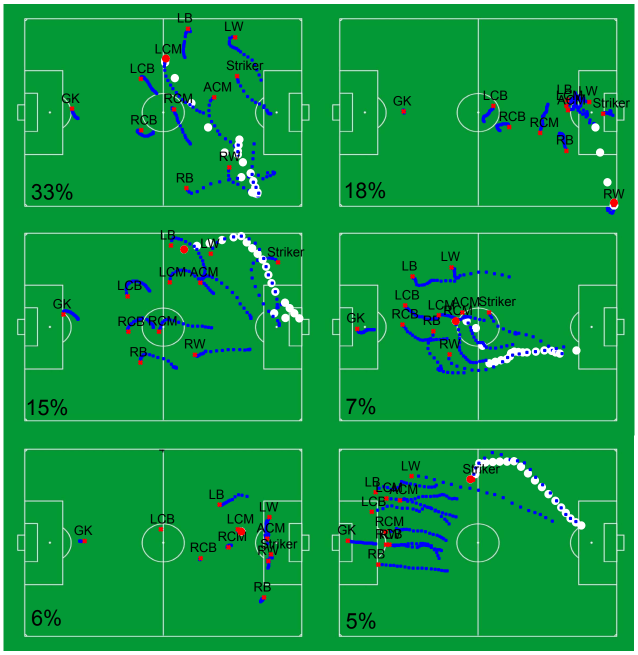

Fig. 10. Plot showing top six scoring patterns of a team. Percentage on the left bottom indicates the how often a team score in this way. Red dot indicates the starting location of each player/ball.

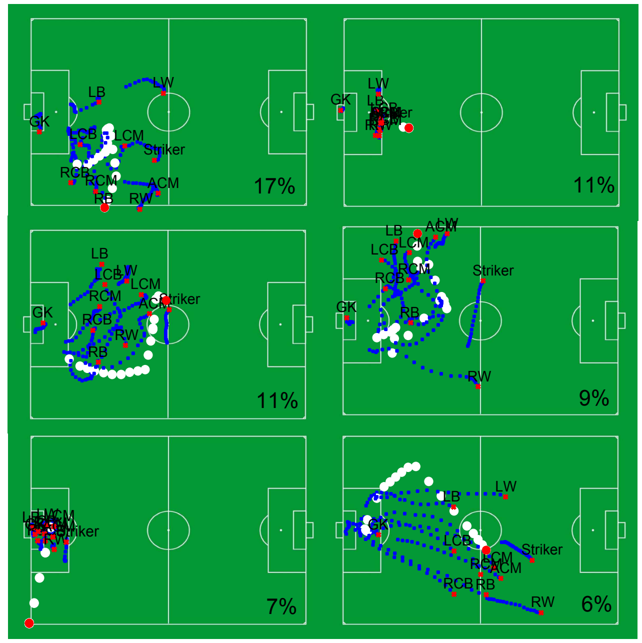

Fig. 11. Plot showing top 6 conceding patterns of a team.

norm because large deviations may signify very different formations

4) When compressing, we use *centroid of all player positions + ball position* as the spatial feature. DCT is selected as the temporal basis function and we select $K_t = 5$;

E. Clustering Results

1) Offensive Play Analysis: To find a team's most probable methods of scoring, our clustering method is performed on all goal-scoring opportunities for this team which includes normal shots, corners and free-kicks. Figure 10 shows the top six scoring methods for a team. The red dot indicates the starting location of each player/ball. As can be seen in this figure, in the first example shown in the top-left corner, 33% of the goal scoring opportunities occur in this fashion where the ball starts on the left-hand-side on the half-way line and then moves to the right-wing who cuts back to the top of the box for a shot on goal. The second top method of getting a shot on goal is via a corner-kick from the right hand side (18%). The third top method starts again from the left-back starting from the half-way line. The fourth and sixth methods seems to be counter attacks, while the fifth is via a free-kick.

2) Defensive Plan Analysis: A similar analysis was conducted on the defensive plays. The top 6 plays of conceding a shot are shown in Figure 11. Unlike the offensive plays, the top method of conceding a shot was only around 17% which came from intricate play on the right-hand side. The second top method appears to be a free-kick, the third again comes from the left-hand side. The fifth method comes from a corner while the sixth comes from a counter attack.

3) Corner Analysis and Free-Kicks: Not only can continuous plays be analyzed, but also set-pieces such as corner-kicks. Figure 12 shows clustering result for all corners. In 1-4, the ball is kicked directly to a striker, who is trying to deflect the ball by head or foot into the goal. 5 and 6 both look like set plays, designed to draw out the defense away from the goal mouth, possibly to create more space for an eventual goal shot. As can be seen in most situations, the team of interest has their defenders around the center of the field which means the flank are potential outlets for a quick counter attack. Similar analysis can be done for free-kicks as shown in Figure 13.

V. Summary and Future Work

In this paper, we presented a method which allows for large-scale analysis of team behavior across large volumes of player and ball tracking data. By using a role-representation which has been recently proposed, in addition to a spatiotemporal bilinear basis model which can form a compact representation of the signal we can cluster plays of a team which can describe their most likely motion patterns associated with a particular event (such as shots, corners, free-kicks). Additionally, we propose a two-layer hierarchical approach to automatically segment a match. Using a decision-tree formulation, we can accurately retrieve events or detect highlights.

In terms of future work, many aspects can be improved. Current model/analysis only takes role data from one team into account. However, team sports like soccer are adversarial so a joint representation is required as a team's behavior is heavily dependent on what the other team is doing. We hope to include such features into a prediction model which can accurately predict the short-term action (i.e. where is the next shot coming from) and ultimately match prediction using team behavioral features. Additionally, we plan to investigate a M-best type

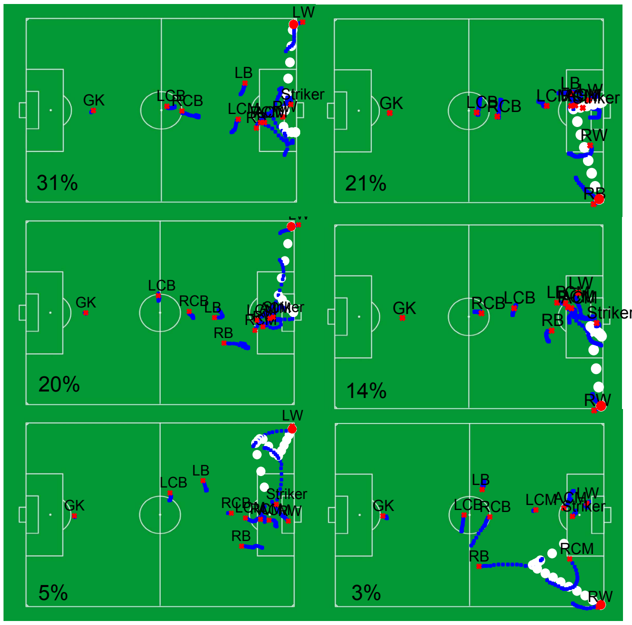

Fig. 12. Corner Analysis.

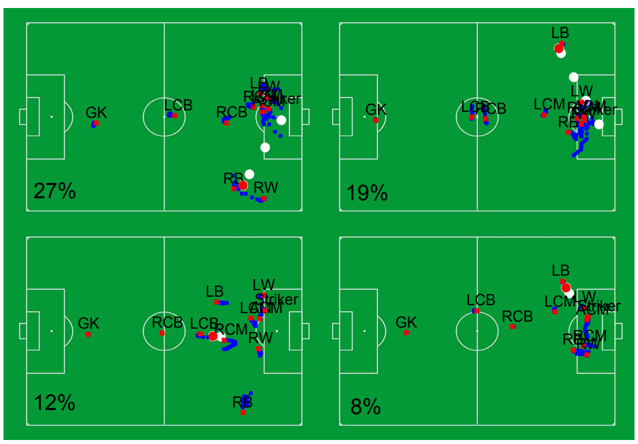

Fig. 13. Free Kick Analysis.

of prediction approach which will allow us to find a diverse solutions which needs to leverage probabilistic methods.

References

[1] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade. Trajectory space: A dual representation for nonrigid structure from motion. In *T. PAMI*, 2010.

[2] Ijaz Akhter, Tomas Simon, Sohaib Khan, Iain Matthews, and Yaser Sheikh. Bilinear spatiotemporal basis models. *ACM Trans. Graph.*, 31(2):17:1–17:12, April 2012.

[3] S. Ali and M. Shah. Floor fields for tracking in high density crowd scenes. In *ECCV*, 2008.

[4] O. Arikan. Compression of motion capture databases. In *ACM Transactions on Graphics, 25(3)*, 2006.

[5] I. Atmosukarto, B. Ghanem, S. Ahuja, K. Muthuswamy, and N. Ahuja. Automatic recognition of offensive team formation in american football plays. In *IEEE International Workshop on Computer Vision in Sports*, 2013.

[6] A. Bialkowski, P. Lucey, P. Carr, S. Denman, S. Sridharan, and I. Matthews. Recognizing team activities from noisy data. In *IEEE International Workshop on Computer Vision in Sports*, 2013.

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

[8] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, J. Singh, and B. Girod. Transform coding of image feature descriptors”. In *VCIP*, 2009.

[9] A. Ekin and M. Tekalp. Automatic soccer video analysis and summarization. *IEEE Trans. Image Process. ,*, 12(7):796–807, 2003.

[10] Hawk-Eye. www.hawkeyeinnovations.co.uk.

[11] A. Hervieu and P. Bouthemy. Understanding sports video using pplayer trajectories. In *Intelligent Video Event Analysis and Understanding. Springer Berlin/ Heidelberg*, 2010.

[12] S. Intille and A. Bobick. A framework for recognizing multi-agent action from visual evidence. In *AAAI*, 1999.

[13] S. Intille and A. Bobick. Recognizing planned, multiperson action. In *CVIU*, 2001.

[14] H. Jgou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large-scale image search. In *ECCV*, 2008.

[15] H. Jgou, M. Douze, and C. Schmid. Product quantization for nearest neighbor search. In *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2011.

[16] Kihwan Kim, M. Grundmann, A. Shamir, I. Matthews, J. Hodgins, and I. Essa. Motion fields to predict play evolution in dynamic sport scenes. In *CVPR*, pages 840–847, 2010.

[17] H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.

[18] R. Li and R. Chellappa. Group motion segmentation using a spatio-temporal driving force model. In *CVPR*, 2010.

[19] R. Li and S. Zhou. Learning multi-modal densities on discriminative temporal interaction manifold for group activity recognition. In *CVPR*, 2009.

[20] 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*, 2012.

[21] P. Lucey, A. Bialkowski, P. Carr, S. Morgan, S. Morgan, I. Matthews, and Y. Sheikh. Representing and discovering adversarial team behaviors using player roles. In *CVPR*, 2013.

[22] P. Lucey, D. Oliver, P. Carr, J. Roth, and I. Matthews. Assessing team strategy using spatiotemporal data. In *ACM SIGKDD Conference on Knowledge, Discovery and Data Mining (KDD)*, 2013.

[23] R. Maheswaran, Y. Chang, A. Henehan, and S. Danesis. Deconstruction the rebound with optical tracking data. In *MIT Sloan Sports Analytics Conference*, 2012.

[24] V. Morariu and L. Davis. Multi-agent event recognition in structured scenarios. In *CVPR*, 2011.

[25] Opta. http://www.optasports.com/.

[26] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool. You’ll never walk alone: Modeling social behavior for multi-target tracking. In *CVPR*, 2009.

[27] M. Perse, M. Kristan, S. Kovacic, and J. Pers. A trajectory-based analysis of coordinated team activity in basketball games. In *CVIU*, 2008.

[28] Prozone. http://www.prozonesports.com/index.html.

[29] M. Raginsky and S. Lazebnik. Locality-sensitive binary codes from shift-invariant kernels. In *NIPS*, 2009.

[30] B. Siddiquie, Y. Yacoob, and L. Davis. Recognizing plays in american football videos. In *Technical Report, University of Maryland*, 2010.

[31] SportsVision. http://www.sportsvision.com.au.

[32] Stats. http://www.stats.com/.

[33] M. Stockl and S Morgan. Visualization and analysis of spatial characteristics of attacks in field hockey. *International Journal of Performance Analysis in Sport*, 13:160–178, 2013.

[34] 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, 201.

[35] X. Wang, V. Ablavsky, H. B. Shitrit, and P. Fua. Take your eyes off the ball: Improving ball-tracking by focusing on team play. In *CVIU*, 2013.

[36] 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.

[37] Y. Weiss, A. Torralba, and R. Fergus. Spectral hashing. In *NIPS*, 2008.

[38] 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 *ACM Multimedia*, MULTIMEDIA ’07, pages 58–67, 2007.