

Article

Recognition of Repetitive Movement Patterns—The Case of Football Analysis

Udo Feuerhake

Institute for Cartography and Geoinformatics, Leibniz University Hanover, Appelstraße 9a, Hannover 30167, Germany; feuerhake@ikg.uni-hannover.de; Tel.: +49-511-762-19369

Academic Editor: Wolfgang Kainz

Received: 9 September 2016; Accepted: 7 November 2016; Published: 9 November 2016

Abstract: Analyzing sports like football is interesting not only for the sports team itself, but also for the public and the media. Both have recognized that using more detailed analyses of the teams' behavior increases their attractiveness and also their performance. For this reason, the games and the individual players are recorded using specially developed tracking systems. The tracking solution usually comes with elementary analysis software allowing for basic statistical information extraction. Going beyond these simple statistics is a challenging task. However, it is worthwhile when it provides a better view into the tactics of team or the typical movements of an individual player. In this paper an approach for the recognition of movement patterns as an advanced analysis method is presented, which uses the players' trajectories as input data. Besides individual movement patterns it is also able to detect patterns in relation to group movements. A detailed description is followed by a discussion of the approach, where different experiments on real trajectory datasets, even from other contexts than football, show the method's benefits and features.

Keywords: spatio-temporal analysis; pattern recognition; trajectory; football analysis

1. Introduction

Just like the ever-increasing importance of football in the media and the fast growth of the related market, the analysis of football games is becoming more and more important. It is utilized in different domains with different purposes. For instance, the television broadcasters fade in analysis results via overlays or split screen.

However, it is not only the media that is interested in such facts and statistics. The football clubs also want to examine the performance of the team during matches or training. Further use cases could be the automated analysis of one's own team during training or the next opponent to discover its tactics, strengths or weaknesses. At the moment, the latter is done manually via video inspection. Often, some details, like the velocity during a sprint, the recognition of inconspicuous movement patterns or repetitive pass sequences, cannot be determined or will not be noticed even by a large coach team since they are hard to recognize with the naked eye. For this purpose, several systems have been developed to analyze football matches in real time. Those systems consist of a hardware (object tracking) component, where the positions of the players, balls and referees are tracked via video, radio or GPS tracking, and a software component, in which the recorded data is semi- or fully automatically evaluated.

The evaluation usually consists of different analyses, which provide information about the players' or teams' performances. They increase the knowledge about the players' or teams' behavior during the game. This knowledge is important as it enables deep insight into players' characteristically behavior or team tactics. The amount of knowledge gained ranges from a single number, which, for example, describes a parameter of the running performance, to complex movement patterns, which contain detailed information about repetitive or characteristic movements of a player or a whole team.

There is quite a wide range of different analyses, which differ in terms of gained knowledge and algorithmic complexity.

The scheme in Figure 1 shows different levels of complexity, in which the common football analysis tasks can be classified. They are arranged in an ascending order from “basic” to “advanced”. The knowledge gain increases as well. The “basic” analyses are of the lowest complexity, but they also provide the least knowledge about the players’ movement behavior or tactics. They mainly consist of pure measurements or simple aggregations. Popular examples for these tasks are heat maps (Figure 2a), which provide an overview over the players’ actual locations preferences and ranges during the game and the covered distance. The latter may reflect the fitness level of the evaluated player, even if a high running performance does not always mean a good overall performance. The behavioral or tactical knowledge gain is relatively low. The “medium” level consists of analyses, which require more than simple aggregation algorithms. A representative for this group is the automated detection of team parts and distances in between (Figure 2b). It enables the coach to evaluate the behavior of the team in respect to the a priori given tactical rules. The most complex analyses with the highest knowledge output are located in the “advanced” level. Solving those tasks requires sophisticated algorithms, which, for example, are able to find pass sequence patterns (Figure 2c), to determine passing possibilities (Figure 2d) or to recognize movement patterns.

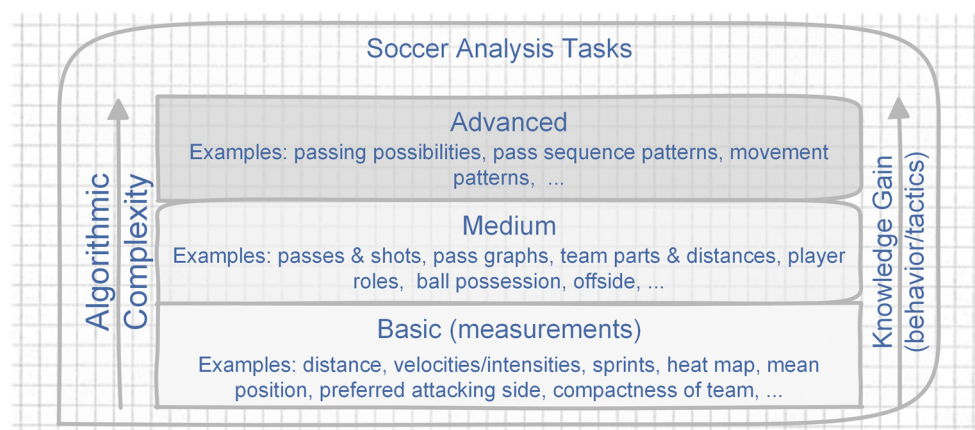


Figure 1. The typical football analysis tasks arranged in three levels of complexity.

Most of the current analysis approaches are located in the lower complexity levels and thus mainly consist of collecting statistical data or performing medium movement analyses. We are going to present an approach which enables an advanced movement analysis in terms of recognizing movement patterns. One advanced task consists in finding repetitive and a priori not known movement behavior of moving individuals or groups. Examples are typical player paths or behaviors of a team in certain play situations. Those patterns may provide deep insights into the movement behavior of the players individually as well as into the tactical movements of a whole team. They can be utilized to either characterize a player by typical movements or to predict movements.

However, the recognition of those patterns is a challenging task, since we do not know what they look like or what to search for. In contrast to a priori known patterns, e.g., predefined geometric paths like circles or group patterns like flocks, we cannot use any matching strategy to identify the pattern instances in the data. A brute force approach would still be possible, but the computational complexity combined with the size of the datasets would not lead to a reasonable solution. In order to identify them, we propose an approach which is based on the transformation of a trajectory pattern recognition problem to a sequence mining problem. To this end, we transform the trajectories to sequences of movements, which are later searched for repetitive subsequences forming the requested patterns. Furthermore, the proposed method is designed in that way that it can be used in other scenarios besides football analysis as well.

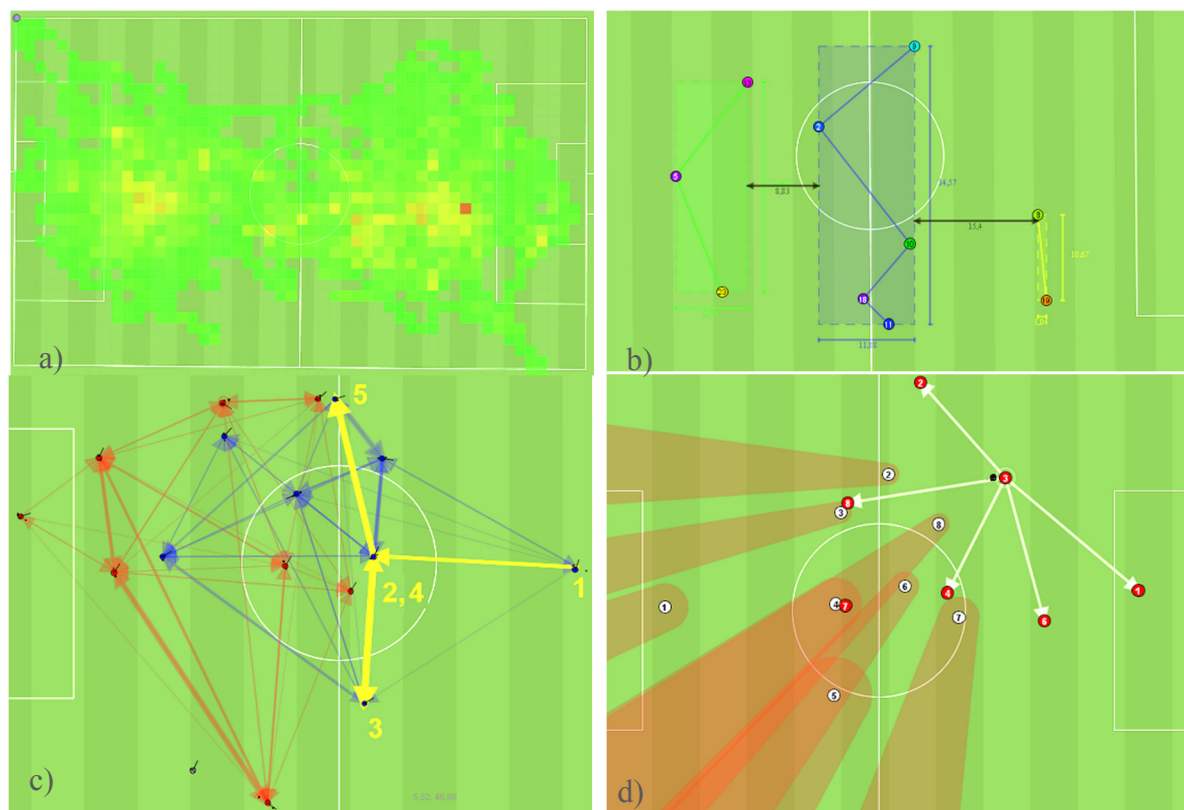


Figure 2. Four analysis tasks from different complexity levels: (a) a player's heat map; (b) the team parts and distances in between; (c) pass sequence patterns (the yellow numbers represent the order of passing players) and (d) the passing possibilities of a ball possessing player (the black dot is the ball, possible passes are marked via white arrows).

The paper is organized as follows: in the next section an overview of related work is presented. In Section 3 our approach is described in detail. Section 4 contains descriptions of different experiments where the approach is applied to real trajectory data from football games and car traffic. There, the extracted movement patterns are shown and evaluated. Finally, we conclude this paper by summing up the main achievements and giving an overview of possible extensions and future tasks.

2. Related Work

2.1. Commercial Football Analysis Systems

There are several companies which have developed football analysis systems consisting of the tracking hardware and the corresponding analysis software. While there is a lot of research on the data generation component, the analysis component shows potential for improvement. For a better view we concentrate on catapult [1], the Deltatre AG [2], Prozone Sports Ltd [3] and Chyronhego [4]. While first three are focused on football, Chyronhego also offers solutions for other types of sports. All of them analyze the player and ball movements and create player and team statistics. They present the results using tables, charts and heat maps as means of visualization. Deltatre further provides analyses like team border strips, the offside line and player connection lines, which help to visualize relative movements of team parts. They also have developed a goal line technology, which is based on an additional magnetic field tracking system and operates parallel to the player tracking system. In this way they achieve high accuracies when determining whether the ball is behind the line. Prozone offers several applications, which have different analytical purposes. Besides a database analysis tool, they provide apps for analyzing the current match, the referee and also the next opposing team

based on its last recorded matches. Since they also track the ball, they are able to carry out ball related analyses, e.g., for passes or goal kicks. Each of those tools offers a variety of basic movement and ball related analyses. To the best of our knowledge, more sophisticated analyses like movement pattern recognition or sequence analyses, which would belong to the “advanced” level of the task classification presented in Figure 1, are not provided.

When looking at individual sports like running or fitness in general, companies like Adidas (micoach) [5], Nike (nike+) [6] or Garmin [7] and platforms like runtastic [8] provide solutions to evaluate sporting activities. In most cases movement data collected with accelerometers or GPS receivers is evaluated. Statistics on the covered distance, velocities and accelerations are created, possibly with the possibility for visual inspection and analysis, as well as a plot of the trajectories on a map. More detailed analyses on the trajectories are not supported.

2.2. Movement Pattern Recognition

Besides these professional tools there are some scientific approaches to analyze football and, in particular, to recognize movement patterns. They have been developed in the context of moving point analyses and have been tackled both from a computational geometry perspective and a decentralized computing point of view. Further, we distinguish between the recognition of a priori known and unknown patterns. When the patterns are known, the recognition is similar to a pattern matching task, whereas the search for unknown patterns is rather a mining process.

In relation to pattern matching, a lot of approaches exist to identify defined group movement patterns, e.g., flock, leadership or encounter patterns (i.e., [9–11]). Those patterns are clearly described in [12]. Another algorithm which is able to detect group patterns as well as individual movement patterns is proposed by [13]. They analyze discretized and relative motions (*REMO*) of the observed objects. To this end, they create a matrix representation (rows: objects, columns: time steps), which is then searched for patterns using spatially extended regular expressions.

There is also related work concerning pattern mining. In the context of analyzing football, there are several approaches to recognize patterns. A previous study [14] developed a comprehensive toolbox, which provides some tools to analyze the player trajectories and passes. When analyzing the movements, they look for subtrajectory clusters, such as repetitive player movements. In order to find those clusters, they use clustering techniques proposed by [15]. The passing analysis also contains a type of pattern recognition in terms of frequent pass sequences. They are extracted by traversing each branch of a generated suffix tree. There are several stand-alone approaches aiming at the extraction or classification of movement (or tactical) patterns. In [16,17], attacks are categorized by their starting location and an a priori defined scheme. Several approaches deal with the extraction of team or group movement patterns in general. In [18] a learned “Spatio-Temporal Driving Force Model” to characterize group motion patterns is used. In [19] a framework is introduced using a feature model and the features’ morphological properties to analyze football tactics. Reference [20] uses a hierarchical architecture of artificial neural networks to find the tactical patterns. Reference [21] describes a way to not find movement but passing patterns. They applied a multi-scale matching method based on contour comparisons. To extract ball movement patterns, which may occur during sequences of passes, [22] proposes a step-wise mining method, which uses different similarity measures to compare the ball’s trajectory and encounters translation, scaling and rotation invariance.

Looking besides the football analysis, further approaches can be found in other domains, e.g., traffic or animal movement. Those can also be transferred to our context. A couple of methods (i.e., [23–27]) use clustering algorithms in combination with distance measures, e.g., edit distances, Dynamic Time Warping, or Longest Common Subsequence, to identify similar trajectories and derive typical object movements. A related method based on the transformation of the trajectories into sequences of class symbols is presented by [28]). This symbolic representation is then used to compare the sequences with the help of a normalized weighted edit distance as distance measure. In [29] an algorithm is described that enables the detection of patterns in terms of object groups which have

the same movement behavior. They use a mining algorithm to detect local movement patterns which afterwards are clustered using a similarity measure to identify group relations. Furthermore, there is a group of approaches which mine periodic patterns in 1-dimensional symbolic sequences. Their key challenge is the transformation of the 2-dimensional movement data into 1D-sequence data. To this end, they generate sequences of rectangular [30], frequently visited [31] or predefined regions [32], which are visited by the trajectories. In this way, they simultaneously reduce the dimension of the data and the high number of trajectory points to more meaningful aggregations. The patterns are then extracted by using existing sequence analysis methods.

To sum up, there are a lot of sophisticated approaches dealing with the extraction of patterns in movement data. However, they do not really fit with our use case. On the one hand, we do not want to match predefined patterns, as the patterns we are looking for are a priori unknown. On the other hand, the methods, which also mine unknown patterns, often work on either whole trajectories or on segments and thus require some kind of segmentation as preprocessing. Since we assume that in our use case patterns only extend over some parts of a trajectory, we cannot work on whole trajectories. However, we also deliberately avoid a segmentation, because we are not able to identify the relevant trajectory parts in advance and do not want to cut possible patterns. Besides that, a reasonable and not arbitrary partitioning of trajectories from a football game without any additional information, e.g., ball possession, play situations, game interruptions, is a quite challenging task. The most likely fitting methods are the approaches that are based on sequence mining. However, a determination of spatial regions, which would be the sequence items, is not applicable in this context as we are dealing with unconstrained player movements.

3. The Movement Pattern Recognition Approach

In this work we present an analysis method which recognizes movement patterns of the players or the team, respectively. The process of our algorithm consists of three stages and is schematically presented in Figure 3. It starts with an input, which is the trajectory data provided by some tracking solution. A following preprocessing prepares the trajectories for being analyzed. In the sequence-based pattern recognition stage, the movement patterns are extracted. In the following sections each of those stages is described in detail.

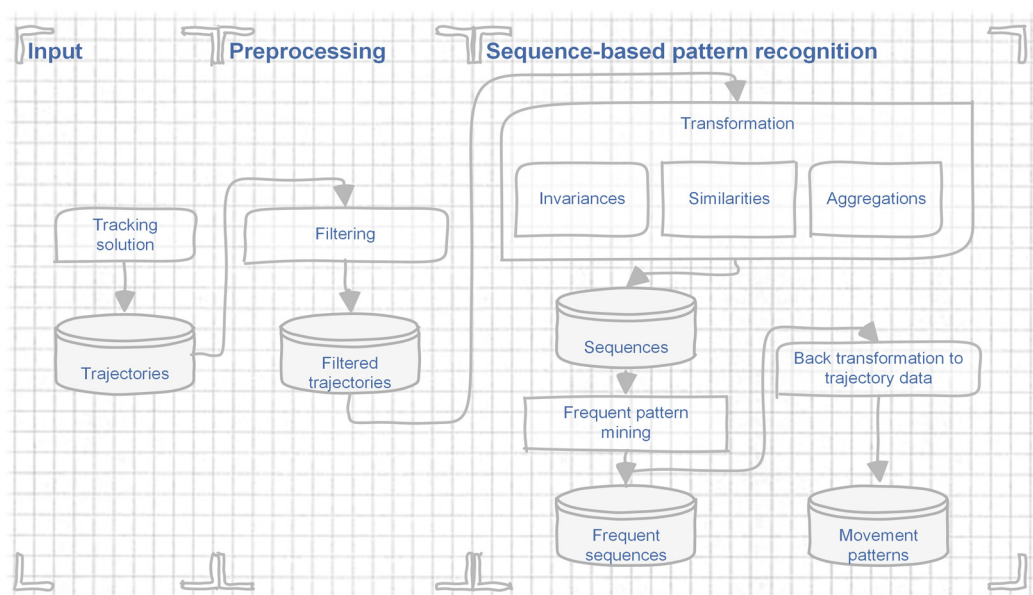


Figure 3. Scheme of our pattern recognition approach.

3.1. Input

Our analyses are based on trajectories for each individual player. It can be generated by a video-, radio- or GPS-based tracking solution. Those systems obtain the positions at discrete time steps. The sampling interval ranges from few nanoseconds to few seconds. The geometric accuracy depends on the used system: while video and radio tracking provides accuracies within few centimeters, the GPS tracking depends on the sensors used and generally is within few meters, absolutely, and within one meter, relatively. Besides the quite expensive professional tracking systems offered by Deltatre AG, Prozone Sports Ltd or Chyronhago, which provide high quality movement data, a novel approach by [33] fuses GPS- and video-based tracking in order to exploit their individual advantages. It aims to combine the reliability of GPS tracking with the high geometric accuracy of camera detections. In this way, systems, which may consist of a smartphone (working as a camera) and a set of GPS-devices (carried by the players) can be used to obtain trajectory data of a similarly high quality. Combined with such a low-cost system, an automatic football analysis may also become attractive to financially weaker users.

3.2. Preprocessing

Depending on the used tracking solution, the input trajectories contain systematic as well as random errors. Systematic errors are caused by inaccurate calibrations or bad measurement conditions, e.g., non-calibrated cameras or bad viewing angles during video tracking. They are predictable and constant during the observation and, for example, can therefore be removed by transforming the data. Random errors are unpredictable and inconstant. Averaging methods can use multiple observations in order to reduce them.

In our approach we use trajectory data from GPS and video tracking. While the GPS trajectories show the typical GPS errors, the video tracking contains more random errors. Those are often caused by player occlusions or erroneous bounding box detections which lead to player location errors (Figure 4). We use transformations to correct the systematic inaccuracies. For the reduction of the random errors we apply a filtering technique. Besides Kalman filtering, which certainly is often a good choice for this task, a simpler (centered) mean filter also provides sufficiently good results (Figure 4, right). The filtered trajectories are then input for the next stage of our algorithm.



Figure 4. Inaccuracies in trajectories due to erroneous bounding box detections. **(Left)** The calculated bounding box does not match the object exactly. **(Right)** The resulting raw trajectory (red line) is afflicted with significant inaccuracies, which are reduced after having applied a filtering technique (blue line).

3.3. Sequence-Based Pattern Recognition

The pattern recognition approach has been introduced in [34]. Originally, it was a generic method to detect movement patterns of object groups of constant group size. In this work we adapt it to the football scenario and create a generalized version, which is also capable of analyzing individual movement patterns. As shown in Figure 5, the algorithm consists of consecutive processing steps which are described in detail in this section. The right hand side of this scheme has already been treated in an original paper.

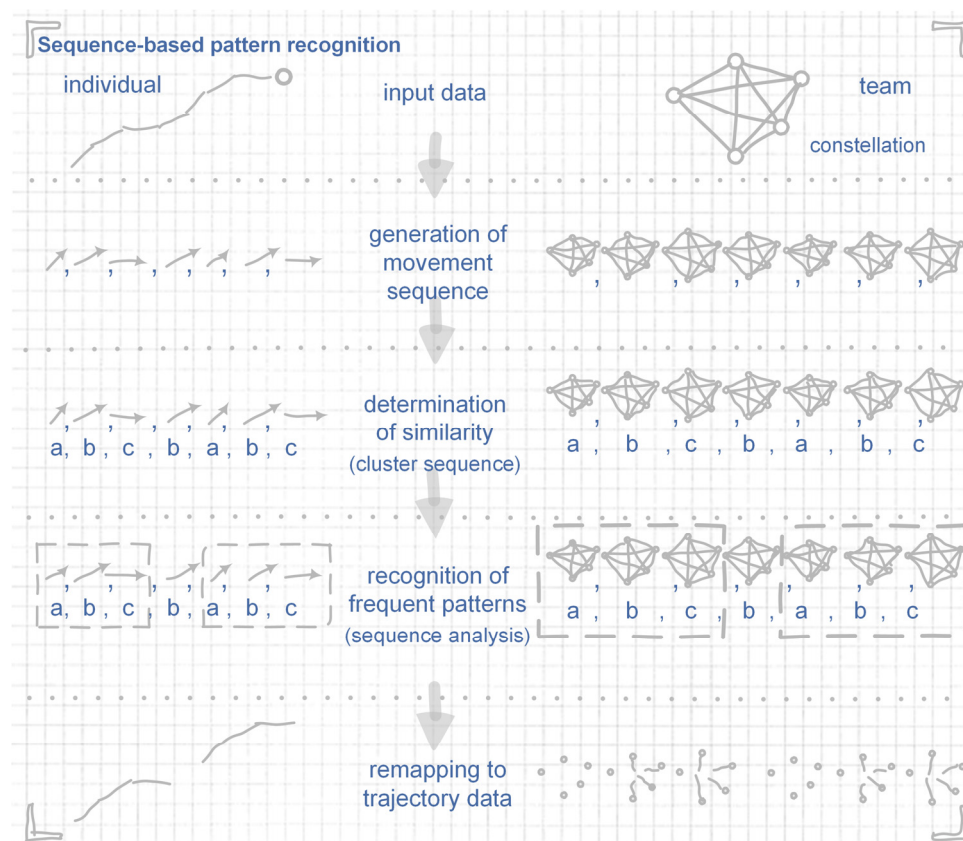


Figure 5. A detailed scheme of the recognition process for individual and team movement patterns, as it is contained as “sequence-based pattern recognition”—stage in Figure 3.

3.3.1. Input to the Algorithm

In general, our pattern recognition approach uses the filtered trajectories as input. Depending on the use case, in which either individual or team movement patterns are searched, the input of the algorithm differs. The input for the individual analysis is the single trajectory of the observed player. Its pendent to detect team patterns takes into account the trajectories of all team members simultaneously. To this end, object constellations are used, which describe the positions of objects relative to each other by position relations, e.g., coordinates, distances, angles, etc. Depending on the selection, there are $\binom{n}{2}$ position relations stored in a constellation, where n is the number of observed objects. In our case a constellation represents a formation of a team at a certain point in time. Due to the fact that we also want to detect transformed (translated or rotated) patterns, we have to choose the suitable position relations to be stored in a constellation. The constellations are described by a vector of these relation values. Using the distances between the positions makes a constellation invariant regarding rotation and translation. Figure 6 visualizes a constellation example and the requirements concerning transformation invariance. Depending on the application, the three scenarios are considered equal.

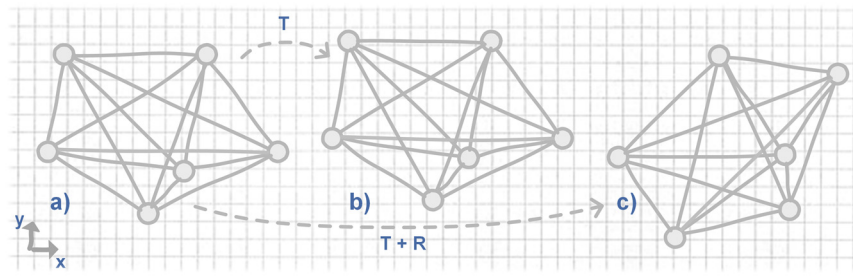


Figure 6. Depending on the desired invariance regarding translation T (b) and additionally rotation $T + R$ (c) the shown constellations will be treated to be equal to (a).

3.3.2. Generation of the Movement Sequence

Our approach is based on the transformation of the whole trajectory data into a sequence of movements S_{tot} , which can be fed into sequence mining methods to extract movement patterns. This sequence lasts the whole observation period and consists of sequence elements I_i which contain information about the object movements at each time step.

$$S_{tot} = \{I_0, I_1, \dots, I_{n_{tot}}\} \quad (1)$$

For a case in which we are analyzing single-object movements and we do not demand that the patterns be invariant in any way, the elements are the player's position in x and y (for the 2D case). If we are looking for team movement patterns, each element will be a constellation that gives information about the players' locations at one point in time. If we allow pattern subsequences to be translated, we have to change the content of the sequence elements to movement vectors for the individual analysis and to vectors containing the distances in x and y between each pair of players for the team analysis. In Table 1 other possible scenarios are listed.

Table 1. Different requirements for the invariances of the patterns lead to different contents of the sequence elements (x, y : coordinates; Φ : heading; r : length of movement vector).

Invariances	Content of Sequence Elements	
	Individual	Team of n Players ($j, k = 1, \dots, n$)
None	$I_i = \begin{bmatrix} x & y \end{bmatrix}^T$	$I_i = \begin{bmatrix} x_j & y_j \end{bmatrix}^T$
Translation	$I_i = \begin{bmatrix} dx & dy \end{bmatrix}^T$ or $I_i = \begin{bmatrix} \varphi & r \end{bmatrix}^T$	$I_i = \begin{bmatrix} dx_{j,k} & dy_{j,k} \end{bmatrix}^T, j \neq k$
Translation + Rotation	$I_i = \begin{bmatrix} d\varphi & r \end{bmatrix}^T$	$I_i = \begin{bmatrix} d_{j,k} \end{bmatrix}^T, j \neq k$

The sampling rate, which determines the length of the time steps, has to be chosen reasonably and greatly depends on the use case. For instance, when analyzing player trajectories during a football game, we have to deal with unconstrained movements including a high number of changes in direction and speed. In order to preserve the details of the movements we have to choose a correspondingly high sampling rate. This means, in general, that a higher sampling rate enables a more precise capturing of the player movements and results in more detailed patterns. However, it also causes a longer sequence of movements and thus requires a greater computational effort.

3.3.3. Determination of Similarity

Since a football field is a Euclidean movement space [35] with few limitations on movement, we have to deal with free and continuous motion of the players. We further have to take the uncertainties concerning the movement sequence elements into account, which stem from the inherent

inaccuracies of the measurement devices (see Section 3.2). Because of that, we do not expect to find exactly matching subsequences of movements in a limited observation timespan. Therefore, we derive a sequence of elements containing discrete values from S_{tot} by discretizing and clustering the movements. In this way we search for similar elements instead of requiring them to be equal. The measure of similarity, which also depends on the use case, is the calculated distance in the space, which is set up by the vectors stored in the sequence elements. Similar sequence elements are assigned to the same cluster. To this end, we have two possibilities: on the one hand we can use predefined feature characteristics, e.g., movement directions like north, south, west, east for individual and translation invariant movements. On the other hand we can use a density-based clustering, like *DBSCAN* [36], or a centroid-based clustering method, like k-means [37], if we want to identify a priori unknown clusters. To control the degree of similarity which is required that sequence elements are assigned to the same cluster, we can use the required clustering parameters, e.g., the k-means algorithm needs the number of clusters. The higher the number of resulting or predefined clusters is, the higher the similarity of equally clustered elements will be. The result of this step is a sequence of elements $S_{tot,cluster}$ with corresponding cluster-names and is used as input for the next step.

3.3.4. Recognition of Frequent Patterns

As we are interested in repetitive patterns, we assume a pattern P to be an at least $suppc_{min}$ (minimum support count) times repeating subsequence S_i with a minimum length of l_{min} in the total sequence $S_{tot,cluster}$:

$$P = \{S_0, S_1, \dots, S_{suppc}\} \quad (2)$$

$$suppc(P) \geq suppc_{min} \quad (3)$$

$$l(P) = l_{min} \quad (4)$$

By following this assumption, we are able to apply an existing frequent pattern mining method. Those methods usually require as input parameters the minimal subsequence length l_{min} and the minimum support count $suppc_{min}$. If we further allow slight deviations within frequent subsequences and thus apply an approximate sequence mining, we need another parameter d which quantifies the allowed deviation. The approximate mining methods use certain distance metrics to determine the deviation between two subsequences. Common are edit distances, like the Levenshtein distance, which provide the minimum number of required single-element edits (i.e., substitution, insertion or deletion) to transform one sequence into another. Depending on whether we search for exact or approximate frequent subsequences, we have to apply a corresponding mining algorithm. While for the exact case methods like the Apriori [38] or the FP-Growth [39] algorithm is appropriate, for the approximate case the Baeza-Yates-Gonnet algorithm [40], amongst others, is possible. In both cases all frequent subsequences are determined which meet the requirements given by the corresponding parameter set. Each of those frequent subsequences containing all its occurrences forms one pattern in the cluster sequence.

3.3.5. Remapping to Trajectory Data

As we are not only interested in the pattern itself, but also in their instantiations in the data, we have to remap the found cluster patterns to the original movement subsequences in S_{tot} and to the trajectories, respectively, to get the actual movement patterns. To this end, we use a suitable data structure, which is described in [34] and links the corresponding sequence elements of $S_{tot,cluster}$ and S_{tot} and thus allows a mapping in both directions. The mapping from the element in the movement sequence to the actual object movement is done via timestamp and object-id.

4. Experiment and Discussion

4.1. Datasets

In order to evaluate our developed pattern recognition approach, we applied it to three datasets with different characteristics. While the first two datasets contain movement information of players during football matches, the third contains car trajectories and thus is from a completely different context.

In the first experiment we processed an available football dataset, which was generated at the Fraunhofer IIS [41] and contains a small field football game with one referee and 8 players in each team (Figure 7, left). It was recorded with their own Real-Time Locating System. During a 60 minutes match the players and the referee were equipped with two sensors close to their feet. The balls were equipped with a sensor. For this experiment we merged both foot trajectories to get one representative player trajectory. The positions of the objects were recorded at a sampling rate of about 200 Hz (the balls even at 2000 Hz) with an accuracy of few centimeters, so it is a very accurate and temporally highly resolved dataset.

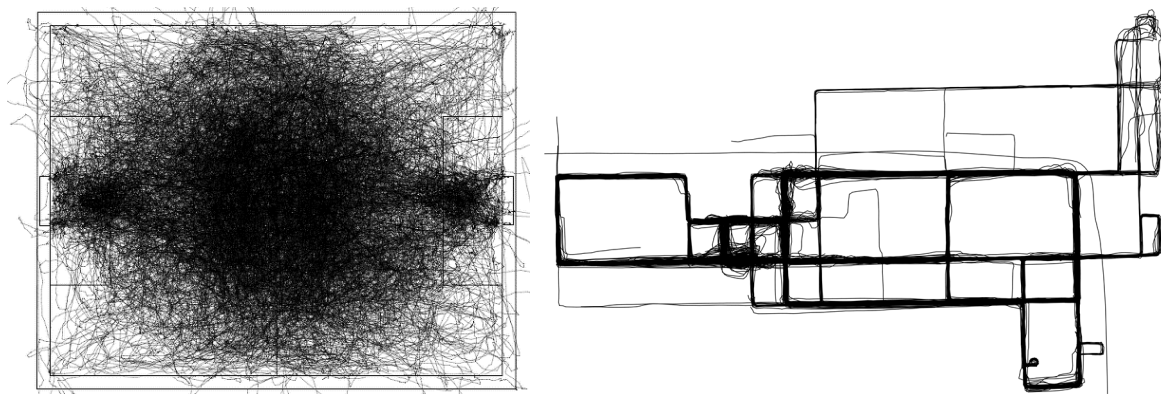


Figure 7. **Left:** The high accurate trajectories contained in the football dataset of the first experiment.
Right: The car trajectories processed in the third experiment.

For the second experiment we use a large football dataset of GPS trajectories. It contains the movement information of a whole team (11–14 players) in more than 20 complete games. The GPS devices which have been used to record the trajectories provide data with a sampling rate of 5 Hz with an accuracy of about 5 m on average. Thus, we have to process about 27 k trajectory points per player per game (in total about 7 million points). In contrast to the previous one, this dataset contains much more movement information although the spatial accuracy and temporal resolution are lower.

The third and last dataset contains data from the traffic context. In the Chicago data from [42], the trajectories have been recorded via GPS with an average sampling rate of about 4 Hz. The dataset contains 889 trajectories with in total about 118k points. In this scenario the observed objects move in a network space [35] which leads to fewer degrees of freedom in movement and thus to a high density of trajectories, although the observation space is much larger than in both previous experiments. We assume that this high density also favors the chance of finding patterns even if no invariances are allowed. In Figure 7right the complete dataset is shown.

In Table 2 an overview of the used datasets, their characteristics and the considered invariances is given.

Table 2. An overview of the different datasets used for the experiments in this section.

Dataset	Experiment	Characteristics	Used Invariances
FRAUNHOFER FOOTBALL	1	Spatial res.: high accurate (few cm) Sampling: 200 Hz Euclidean movement space 11.5 m points (1 game)	None, translation, translation & rotation
GPS FOOTBALL	2	Spatial res.: 5–10 m (GPS) Sampling: 5 Hz Euclidean movement space ~7 m points (>20 games)	Translation
MAPCONSTRUCTION (CHICAGO)	3	Spatial res.: 5–10 m (GPS) Sampling: ~4 Hz Network movement space 118 k points	None

4.2. Result Verification and Pattern Interestingness

The verification of basic movement analyses is straightforward when having ground truth data or using alternative methods to compare. However, the verification of movement patterns is quite challenging, since there is no ground truth data in general. The comparison to results of pattern mining approaches, which are based on a clustering of trajectory segments, is possible; however, we do not expect to find comparable patterns due to the very different way of approaching the problem. By comparing the elements of the individual sequences, we can evaluate the correctness of the resulting patterns, which, however, is ensured by using a tested and correctly working sequence mining algorithm. We cannot evaluate the completeness of the results, since no ground truth data are available for our experiments.

Because of this, we introduce the pattern interestingness as a metric, which makes the results comparable and also describes the information gain the patterns can provide, which is our original intention. The information gain strongly depends on the applications scenario. In common scenarios, in which movement patterns are used, namely movement prediction and characterization, the gain is determined by the pattern extent and the similarity of the contained subsequences. For example, the longer and more similar the movement patterns are, the better the movement of an object can be predicted. The extent of the pattern is determined by the number of trajectory segments $suppc$ and their lengths $l(P)$. When analyzing individual movements, $l(P)$ is the mean spatial distance s the player has covered in the contained subsequences (Equation (5)). In the case of analyzing group movements, $s(S_i)$ is the mean of all involved player distances D_j (Equation (6)).

$$l(P) = \text{mean}_{1 \leq i \leq suppc} s(S_i), \quad (5)$$

$$s(S_i) = \text{mean}_{1 \leq j \leq n} D_j, \quad (6)$$

The similarity $sim(P)$ is calculated by

$$sim(P) = 1 / (1 + \max_{1 \leq i, j \leq suppc} d(S_i, S_j)) \quad (7)$$

For the determination of the distance between two sequences $d(S_i, S_j)$ we calculate the distance between their corresponding trajectories by using the Fréchet metric. For the individual case we simply consider each combination of sequences. However, since there is one trajectory per player for each sequence for the group motions, we first have to determine the corresponding trajectories across the sequences, before we can calculate the sequence distances. Therefore, the interestingness I of a pattern is

$$I(P) = suppc(P) \cdot l(P) \cdot sim(P). \quad (8)$$

This score enables the usage of our algorithm in a visual analytics context, since it points the user to interesting patterns. The final interpretation of the patterns has to be done by user himself or herself. Figure 8 shows some resulting patterns ordered by their interestingness scores. Please note that the score increases from left to right. In the following section, we will use this measure to evaluate the results of our approach. We further visually inspect the patterns concerning their reasonability based on our sufficient football expertise.



Figure 8. Illustration of the interestingness score based on individual movement patterns: it increases from left to right. The different colors represent different cluster assignments of each single movement/sequence element.

4.3. Movement Pattern Recognition Results

4.3.1. Experiment 1

Since our approach requires a set of parameters during the different steps, we first did a parameter study to find out the most reasonable setting for each kind of invariance given in Table 1. The parameters were the number of clusters k (k -means) and the input $suppc_{min}$ and l_{min} to the Apriori mining algorithm. For both latter we assumed fix values, namely $suppc_{min} = 2$ as minimal possible support count and $l_{min} = 10 m$ for sufficiently long patterns.

In this way, we investigated the influence of the number of clusters, which control the degree of similarity, on the resulting patterns. We evaluated each setting by summing up the resulting interestingness scores to obtain an overall interestingness (Table 3).

The parameter study shows the following trends: when the number of clusters increases, the mean similarity also increases, whereas the number of the resulting patterns, the mean support count and pattern length decrease. The total interestingness has a maximum as it depends on all factors. In this way, the study provides the most promising parameter settings for an individual and group analysis (red/yellow highlights in Table 3), which will be used in the following experiment. Further, looking only at the maximum values for each invariance combination, the number of resulting patterns increases with each additional invariance (individual: 983, 1895, 5609 patterns, group: 188, 207, 243 patterns). In Figure 9 we further show some pattern examples, which were recognized with those settings.

In the same dataset, 207 team movement patterns were found by our extraction method (same algorithms as before) when we use the parameter setting (k -means: $k = 128$ clusters) provided by the study (Table 3) for translation invariance. Figure 10 shows one group movement pattern (constellations 408–413), which lasts 6 seconds and consists of a sequence of 3 different constellation types (2 green, 2 dark blue, 2 blue clusters); it shows how five (red encircled) players in the group move top down. It occurs twice during the observation time (i.e., also in the cluster sequence 831–836). This pattern shows a typical defense behavior, in which the team shifts from right to left field (the lonely right player is the goal keeper of the team) to put pressure on the ball-possessing players of the opposing team.

Table 3. Influence of the number of clusters on the resulting (individual/group) patterns. Maximum interesting values for individuals are highlighted in red, for groups in yellow.

# Clusters	# Patterns (ind. / group)	Ø Suppc [-]	Ø Length ([m])	Ø Similarity ([1/m])	Σ Interestingness [-]
No invariance					
4	99/83	2.04/2.00	44.06/84.20	0.07/0.008	622.9/111.8
8	180/126	2.08/2.05	33.87/44.93	0.08/0.011	1014.5/127.7
16	439/188	2.14/2.15	25.22/28.23	0.10/0.020	2369.3/228.2
32	813/209	2.21/2.15	18.39/21.64	0.13/0.020	4295.4/194.5
64	983/192	2.19/2.15	15.22/16.59	0.15/0.021	4914.8/143.8
128	787/234	2.16/2.12	13.30/11.69	0.19/0.026	4295.7/150.8
Translation invariance					
4	1895/83	2.78/2.0	19.88/84.20	0.20/0.008	20946.0/111.8
8	1411/39	2.46/2.03	15.66/119.08	0.31/0.010	16850.6/94.3
16	817/96	2.27/2.05	13.92/48.95	0.37/0.011	9551.9/106.0
32	391/123	2.16/2.04	13.03/32.42	0.41/0.014	4511.9/113.9
64	155/173	2.06/2.06	12.43/19.44	0.42/0.018	1667.0/124.7
128	34/207	2.0/2.10	11.91/13.12	0.40/0.023	324.0/131.2
Translation + rotation invariance					
4	5609/88	2.26/2.02	14.38/53.73	0.02/0.018	3645.7/171.9
8	2521/154	2.66/2.14	22.06/33.51	0.02/0.017	2958.6/187.7
16	2245/197	2.79/2.14	15.95/23.83	0.02/0.018	1998.1/180.8
32	1348/209	2.80/2.14	14.57/19.97	0.02/0.016	1175.3/142.9
64	587/243	2.34/2.23	13.48/18.02	0.02/0.020	370.3/195.3
128	172/200	2.06/2.17	12.90/11.09	0.03/0.023	137.1/110.7

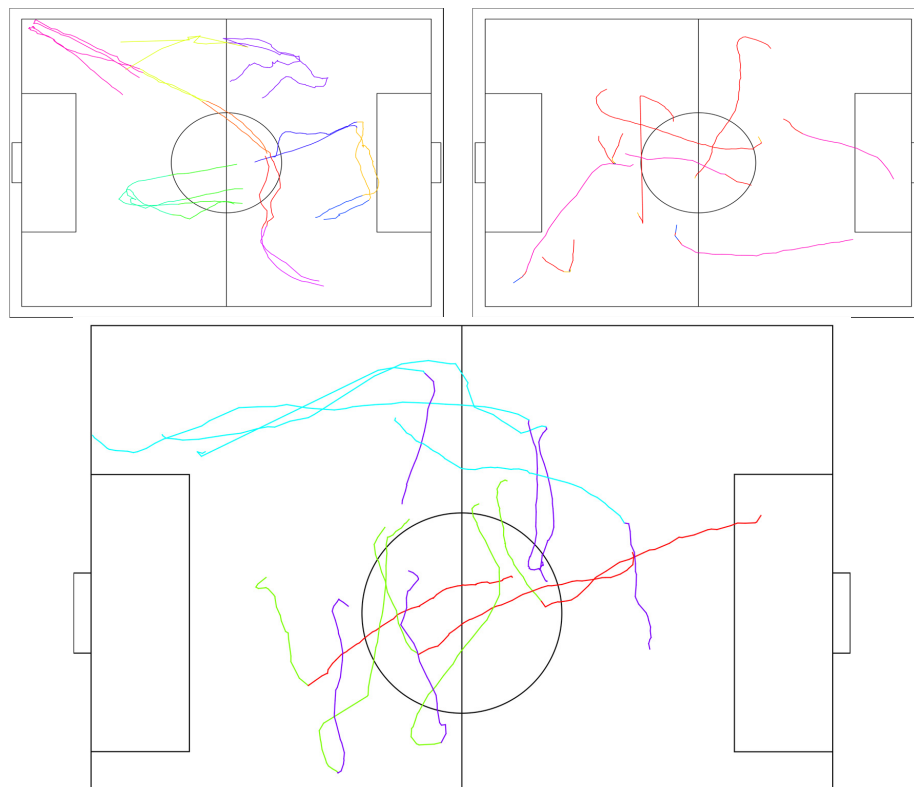


Figure 9. Some resulting movement patterns: **(Top left)** no invariances: spatially overlapping trajectories are found. **(Top right)** Translation and rotation invariance: the trajectories belonging to the same pattern are shifted and rotated. **(Bottom)** Translation invariance: in this case only shifts are allowed. The colors symbolize different cluster assignments of each single movement/sequence element.

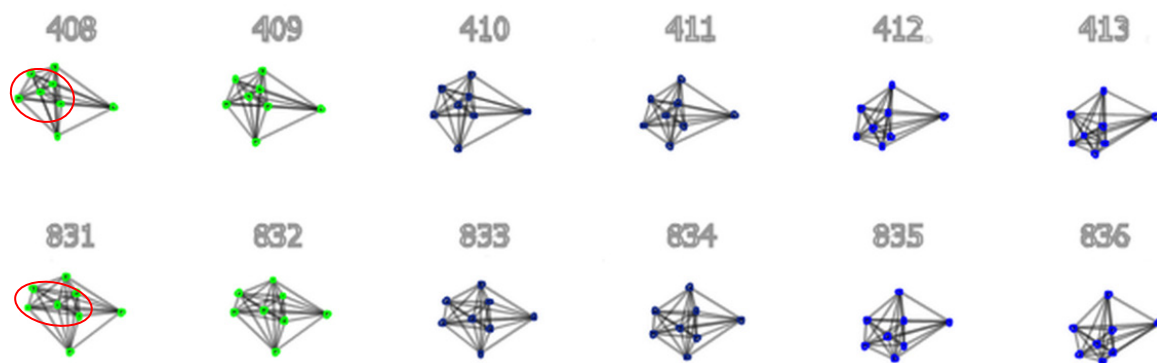


Figure 10. Movement pattern of the whole team which occurs twice during the game.

4.3.2. Experiment 2

In contrast to the first experiment, where we processed a dataset containing only one football game, we used the second dataset containing several games and thus much more movement information for each player for this experiment. Because of that we could process each player individually to obtain player specific patterns, which also contain more player specific knowledge. Therefore, in this experiment we selected players with different roles to check whether there are specific patterns. We compare a central midfielder and a wing player, whose heat maps (Figure 11) show their typical positions during the games. If we require the patterns to be neither translation- nor rotation-invariant, the distribution of the patterns will correspond to the heat maps. Further, the orientation of the patterns will be different for both players: the patterns of the wing player will be mainly horizontally orientated, while the central midfielder's patterns will be orientated both horizontally and vertically.

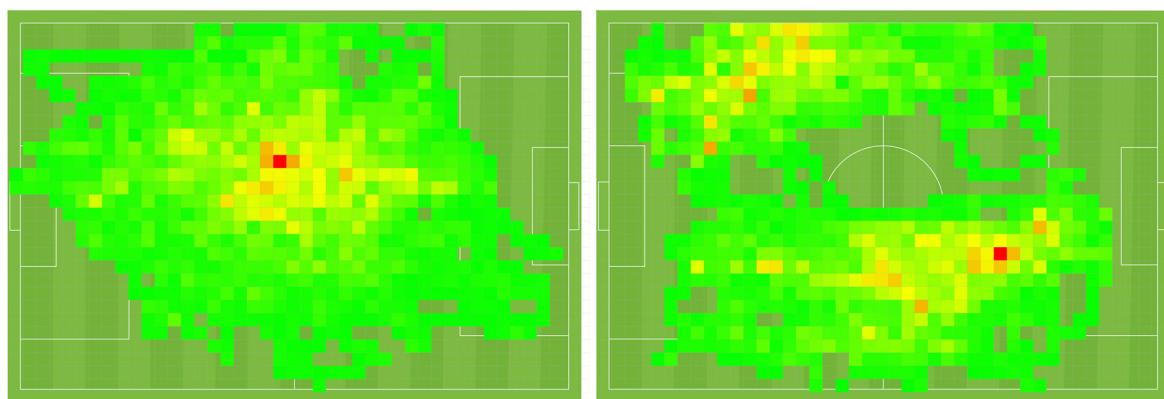


Figure 11. Typical heat maps of a central midfielder (left) and wing player (right) during the games.

In this experiment we allow the patterns to be translated. Further, we used the k-means and Apriori algorithm. The parameter setting was the same we identified in the previous experiment. We evaluated 12 games, in which both players participated. In Figure 12 we show some patterns for both players. Using this setting we can observe the same orientation behavior of the patterns. Further, there is a difference in the shapes of the patterns. The patterns of the wing player (Figure 12 right) are mainly straight runs with only little change in direction. Contrary to that, the patterns of the central midfielder (left) contain significantly more turns. This also fits to the usual behavior of both player roles. While the midfielder in general has more freedom of movement in most of the play situations, the wingman has to stick to his position/side most of the time. It is also possible to look at the results from a tactical point of view, which certainly is interesting for the coaches or scouts. For instance, it is possible to observe different behaviors of the wing player in both match periods. Please note that

the lower three (blue, yellow, pink) stem from the first, the upper three patterns (cyan, red, green) from the second half of the match. Especially, the cyan and blue colored patterns illustrate different movement behaviors. While in the first half (blue pattern) the player stayed longer on the wing and thus moved straight along the outer line, in the second half (cyan pattern) he left his position quite early (few meters before the centerline) and moved directly towards the opponent's goal. The reason why he changed his behavior cannot be found by this analysis. Nevertheless, it is an important piece of information which can be used by either the player's coaches to judge his performance or by the opponent coaches to prepare their teams.

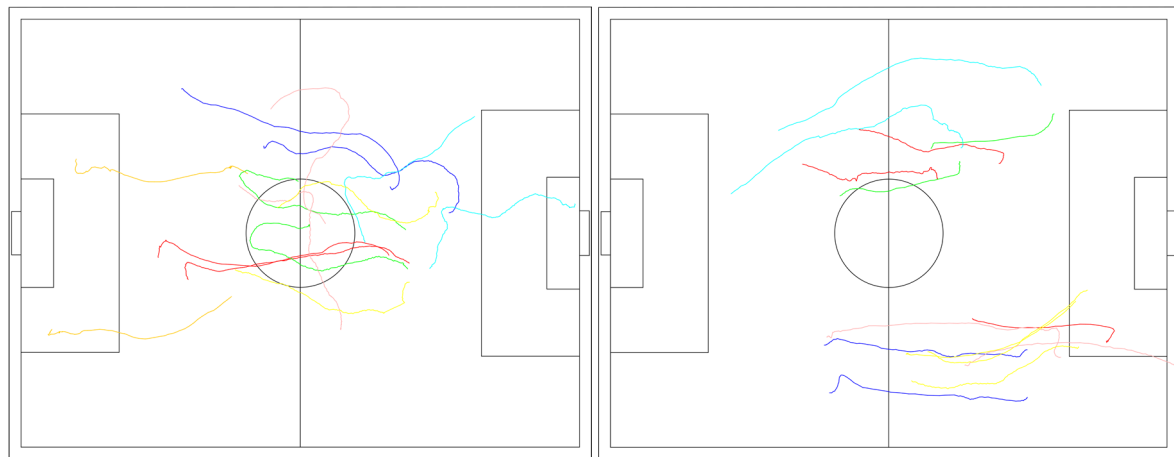


Figure 12. A comparison of movement patterns for players with different roles. **Left:** center midfielder. **Right:** wing player.

4.3.3. Experiment 3

With this experiment we wanted to prove the portability of our approach to other data and contexts. However, we had to consider the characteristics of the dataset used. Since the movements of the objects were strongly influenced by the underlying street network and we were not interested in repetitive network structures but in repetitive movements, we required the patterns to be neither translated nor rotated this time. We further used 64 clusters (k-means) to determine the similarity of the movements. In order to mine the patterns, we again used the Apriori algorithm with the same parameter setting as in both previous experiments. We only looked for patterns of individuals and skipped the group pattern analysis because the dataset does not contain any information about meaningful groups of cars, e.g., convoys, which travel together.

In total, 350 patterns were found. Compared to the previous parameter study (Table 3) the mean support count (2.04) and similarity (0.08 1/m) are a little lower. However, the mean length of a pattern (406.21 m) and the total interestingness score (23,203) are much higher, which can be explained by the less degrees of freedom in movement in this movement space. Please note that the application of the interestingness score is also valid for this scenario, since there are no context-dependent factors in Equation (8). In Figure 13 some of the most interesting patterns, according to our interestingness score, and the underlying street network (grey color) are shown. The presented patterns (each pattern has its own color) consist of two or more overlapping trajectory segments and represent frequently taken routes through the city. Those routes are often on the main roads and the feeder roads, respectively. However, the patterns can also show frequently used rat runs or shortcuts those popular routes can be used for different fields: urban planning or management can use this information to optimize the street network design as well as the traffic management. Navigation systems can make use of this information to predict the user's next trajectory and, in a collaborative way, when the information is shared among the road users, to predict traffic in the corresponding area. Further, it will be possible to

extract public transport, such as bus lines, as they usually use the same routes regularly, if they are included in the data.

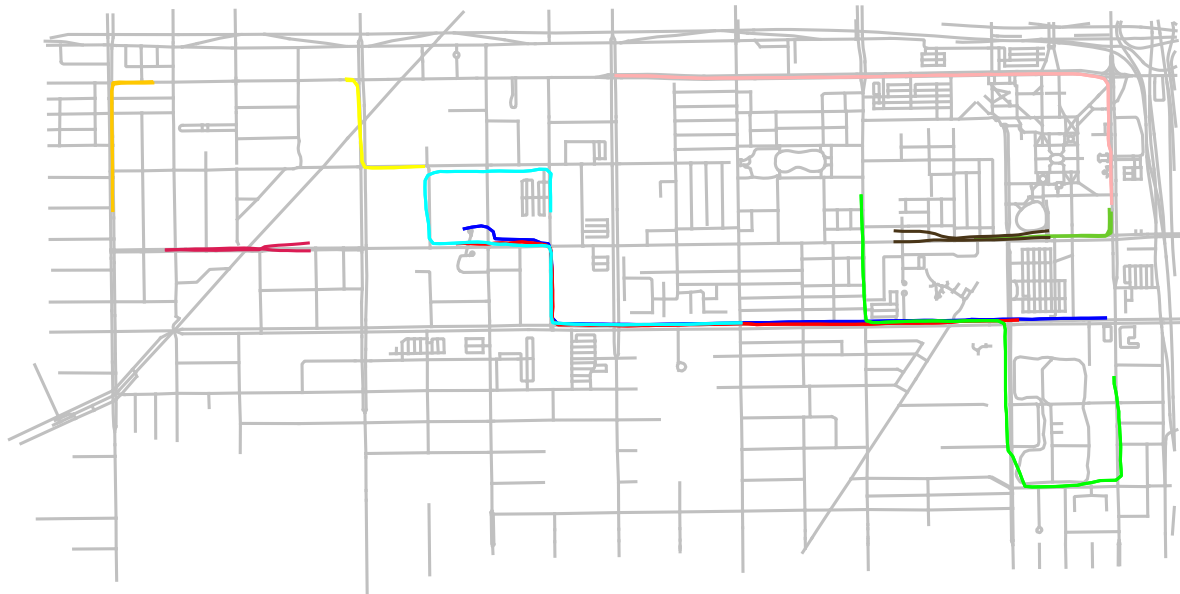


Figure 13. Some of the most interesting movement patterns extracted from the traffic dataset. Each color represents a pattern.

5. Conclusions and Outlook

5.1. Summary

5.1.1. Motivation and Approach

The intention of this work was to go beyond the common football analysis tasks to reveal more meaningful knowledge about the movement behavior of individual players as well as the tactics of whole teams. To this end, we have identified movement patterns as an important provider of that kind of knowledge. We have presented an approach to recognize a priori unknown patterns in individual and group movements based on trajectory data.

The approach consists of a preprocessing, in which errors in the movement data are reduced, and a sequence-based pattern recognition. The latter uses a discretization of the trajectory data to sequence data, which then are analyzed by sequence mining methods to identify repetitive subsequences. Finally, those subsequences are transformed back to the original trajectory data to obtain the trajectory segments, which are part of a pattern.

We applied this approach in three different experiments to trajectory data from football games as well as from a traffic scenario. In each of those experiments we were able to identify movement patterns of the observed objects. These patterns usually provide more knowledge about the movement behavior of the players, or objects in general, than other basic or medium analysis methods (Figure 1) which are often based on basic statistics.

5.1.2. Features of the Approach

The structure of our whole approach is modular, so it is possible to exchange the proposed methods in the different steps as long as the alternatives are able to provide results of the same structure. For instance, we have already proposed different alternative clustering methods for generating $S_{tot,cluster}$ or mining methods to identify frequent subsequences.

In current state of this approach the patterns cannot be extracted in real-time. Due to their high computational complexities the proposed clustering (DBSCAN: $O(n \log n)$, k-means: $O(nkdi)$) and sequence mining (candidate generation in Apriori: $O(2^d)$, FP-Growth: $O(n^2)$) methods do not allow a real-time analysis. In order to identify patterns at runtime our approach has to be customized in the way that incrementally working algorithms for clustering, e.g., IncrementalDBSCAN [43] or I-k-means ([44], or sequence mining, e.g., IncSpan ([45], are used. However, due to the modular structure, this exchange of subparts of the overall algorithm is supported.

Besides its internal flexibility, our approach also offers an application to other analysis scenarios than to the football analysis. Since it has no further requirements, it is applicable to trajectories of any kind of moving point objects from various research domains, e.g., animal behavior analysis, traffic management, surveillance, etc. For instance, this is described in the last experiment of this work, and in the previous work [34], where an experiment on animal (sea gull) movement was conducted. The set of parameters for the different methods that are used in this approach as well as the possibility to define the degrees of freedom of the patterns, which show up in different invariances (Table 1), makes it portable to other use cases, which has been proved by the third experiment of the experimental section.

5.2. Outlook

5.2.1. Extension of the Approach

There are several possibilities to extend the presented work in future. As already mentioned before, one remaining item is to enable real-time capability, which is useful for certain applications, with the given means.

Moreover, different context information can be included to identify even more meaningful patterns as well as to speed up the algorithm to address real-time processing. At the moment only the movement of the observed players is considered. For instance, using the player role (e.g., left or right wing player, forward, etc.) as prior information could help to reduce the pattern search space. In order to do this, movement possibilities, which are very unlikely for the corresponding position, will not be considered during the recognition process as the chance to recognize a pattern in these movements is very low, too. Further, since this movement is often influenced by the positions of the opponents and team members as well as the position of the ball, their trajectories could also be included. This could be achieved by integrating the corresponding position information in the vectors, which are contained in the sequence elements of $S_{tot,cluster}$.

Further, different play situations could be analyzed separately to obtain, for instance, offense and defense patterns of individual players or the whole teams. The same applies for the execution of standard situations like corner or free kicks. Even there, typical rehearsed player or team movements could be identified. For this purpose, the time periods, which contain those play situations, have to be determined first, for example by using supervised learning methods. In this way the search space is reduced to only relevant trajectory segments. This reduction leads to the identification of fewer but more meaningful patterns.

5.2.2. Utilization of Movement Patterns

Another open item is the utilization of the resulting patterns. In [46] two typical usages are named: characterization and prediction. In terms of characterization, the recognized patterns are used to describe the observed players or teams, respectively. The description can either be used as simple information gain for the coaches or viewers or to recognize players purely based on their movements. The latter is useful when this additional player information is fed to the tracking system to improve its performance by an automatically reassignment of lost players in complex scenes. A first step towards the characterization of the players is done in the second experiment, where we compare patterns of two different players with different roles. This knowledge can then be used as described above.

The prediction provides knowledge about future movements of the players. In Figure 14, two instances, one for the football and another for the traffic scenario, are shown where movement patterns describe different possibilities of what the future trajectory may look like, assumed that the object (green) located at the beginning of those patterns. For this purpose, we have to identify the current movement pattern of a player. With the help of this pattern we are able to make prediction regarding the future trajectories. Thus, we need a sufficiently high number of patterns for each player to cover as many situations as possible and to become more precise. In the traffic context, so-called user-aware navigation systems, which automatically predict the user's usual destinations, e.g., the daily way to work and back home or the way to regularly visited free-time activities, can make use of the extracted patterns. The knowledge for the predictions can be retrieved from the currently travelled path and the user's route history. This history is mined for patterns which then are matched to the current path to derive the possible destination.

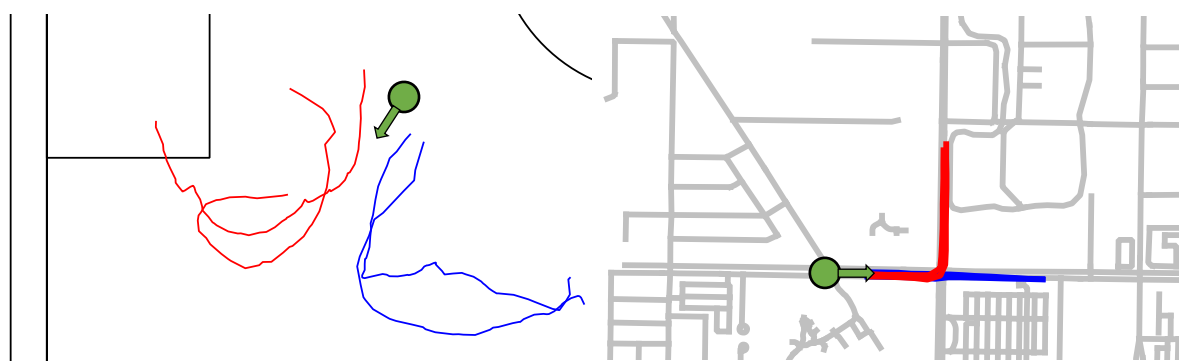


Figure 14. Movement patterns can be used to predict future object movements. In both cases (**left:** football, **right:** traffic) two patterns (red and blue) describe the possible future trajectories of the green object (player/car).

Acknowledgments: The funding of the “q-trajectories” project by DFG as the origin of this research is gratefully acknowledged. The publication of this article was funded by the Open Access Fund of the Leibniz Universität Hannover.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Catapult Australia | Wearable Technology for Elite Sports. Available online: <http://www.catapultsports.com/au/> (accessed on 8 September 2016).
2. Digitale Medien, TV-Übertragung, Backend-Services für den Sport. Available online: <http://www.deltatre.com/de/> (accessed on 8 September 2016).
3. Athlete Monitoring Software, Performance Analysis Software. Available online: <http://prozonesports.stats.com/> (accessed on 8 September 2016).
4. ChyronHego. Available online: <http://chyronhego.com/> (accessed on 8 September 2016).
5. Fitter, Schneller, Stärker | Wachs über dich hinaus | adidas miCoach. Available online: <http://www.adidas.de/micoach> (accessed on 8 September 2016).
6. NIKE+ Apps & Services. Available online: http://www.nike.com/us/en_us/c/nike-plus (accessed on 8 September 2016).
7. Sport | Garmin | Deutschland. Available online: <http://www.garmin.com/de-DE/explore/intosports> (accessed on 8 September 2016).
8. Runtastic: Laufen, Radfahren & Fitness GPS-Tracker. Available online: <https://www.runtastic.com/de/> (accessed on 8 September 2016).

9. Laube, P.; Duckham, M.; Wolle, T. Decentralized movement pattern detection amongst mobile Geosensor nodes. In *Geographic Information Science*; Cova, T.J., Miller, H.J., Beard, K., Frank, A.U., Goodchild, M.F., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 199–216.
10. Gudmundsson, J.; van Kreveld, M.; Speckmann, B. Efficient detection of patterns in 2D trajectories of moving points. *GeoInformatica* **2007**, *11*, 195–215. [[CrossRef](#)]
11. Benkert, M.; Gudmundsson, J.; Hübner, F.; Wolle, T. Reporting flock patterns. *Comput. Geom.* **2008**, *41*, 111–125. [[CrossRef](#)]
12. Dodge, S.; Weibel, R.; Lautenschütz, A.K. Towards a taxonomy of movement patterns. *Inf. Vis.* **2008**, *7*, 240–252. [[CrossRef](#)]
13. Laube, P.; Kreveld van, M.; Imfeld, S. Finding REMO—Detecting relative motion patterns in geospatial lifelines. In *Developments in Spatial Data Handling*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 201–215.
14. Gudmundsson, J.; Wolle, T. Football analysis using spatio-temporal tools. *Comput. Environ. Urban Syst.* **2014**, *47*, 16–27. [[CrossRef](#)]
15. Buchin, K.; Buchin, M.; Van Kreveld, M.; Luo, J. Finding long and similar parts of trajectories. *Comput. Geom.* **2011**, *44*, 465–476. [[CrossRef](#)]
16. Niu, Z.; Gao, X.; Tian, Q. Tactic analysis based on real-world ball trajectory in soccer video. *Pattern Recognit.* **2012**, *45*, 1937–1947. [[CrossRef](#)]
17. Zhu, G.; Huang, Q.; Xu, C.; Rui, Y.; Jiang, S.; Gao, W.; Yao, H. Trajectory based event tactics analysis in broadcast sports video. In Proceedings of the 15th ACM International Conference on Multimedia, Bavaria, Germany, 23–28 September 2007.
18. Li, R.; Chellappa, R. Group motion segmentation using a spatio-temporal driving force model. In Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13–18 June 2010.
19. Kim, H.C.; Kwon, O.; Li, K.J. Spatial and spatiotemporal analysis of soccer. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York, NY, USA, 1–4 November 2011.
20. Grunz, A.; Memmert, D.; Perl, J. Tactical pattern recognition in soccer games by means of special self-organizing maps. *Hum. Mov. Sci.* **2012**, *31*, 334–343. [[CrossRef](#)] [[PubMed](#)]
21. Hirano, S.; Tsumoto, S. Finding interesting pass patterns from soccer game records. In *European Conference on Principles of Data Mining and Knowledge Discovery*; Springer: Berlin, Germany, 2004; pp. 209–218.
22. Mutschler, C.; Kókai, G.; Edelhäusser, T. Online data stream mining on interactive trajectories in soccer games. *IEEE Trans. Intell. Transp. Syst.* **2015**. [[CrossRef](#)]
23. Pelekis, N.; Kopanakis, I.; Kotsifakos, E.E.; Frentzos, E.; Theodoridis, Y. Clustering trajectories of moving objects in an uncertain world. In Proceedings of the ICDM'09, Pisa, Italy, 15–19 December 2009.
24. Long, J.A.; Nelson, T.A. Measuring dynamic interaction in movement data. *Trans. GIS* **2013**, *17*, 62–77. [[CrossRef](#)]
25. Nanni, M.; Pedreschi, D. Time-focused clustering of trajectories of moving objects. *J. Intell. Inf. Syst.* **2006**, *27*, 267–289. [[CrossRef](#)]
26. Morris, B.; Trivedi, M. Learning trajectory patterns by clustering: Experimental studies and comparative evaluation. In Proceedings of the CVPR 2009, Miami, FL, USA, 22–24 June 2009.
27. Lee, J.G.; Han, J.; Li, X.; Gonzalez, H. TraClass: Trajectory classification using hierarchical region-based and trajectory-based clustering. *Proc VLDB Endow.* **2008**, *1*, 1081–1094. [[CrossRef](#)]
28. Dodge, S.; Laube, P.; Weibel, R. Movement similarity assessment using symbolic representation of trajectories. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1563–1588. [[CrossRef](#)]
29. Tsai, H.P.; Yang, D.N.; Chen, M.S. Mining group movement patterns for tracking moving objects efficiently. *IEEE Trans. Knowl. Data Eng.* **2011**, *23*, 266–281. [[CrossRef](#)]
30. Cao, H.; Mamoulis, N.; Cheung, D.W. Mining frequent spatio-temporal sequential patterns. In Proceedings of the Fifth IEEE International Conference on Data Mining, Houston, TX, USA, 27–30 November 2005.
31. Giannotti, F.; Nanni, M.; Pinelli, F.; Pedreschi, D. Trajectory pattern mining. In Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 12–15 August 2007.

32. Mamoulis, N.; Cao, H.; Kollios, G.; Hadjieleftheriou, M.; Tao, Y.; Cheung, D.W. Mining, indexing, and querying historical spatiotemporal data. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004.
33. Feuerhake, U.; Brenner, C.; Sester, M. GPS-aided video tracking. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 1317–1335. [[CrossRef](#)]
34. Feuerhake, U.; Sester, M. Mining group movement patterns. In Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Orlando, FL, USA, 5–8 November 2013.
35. Gudmundsson, J.; Laube, P.; Wolle, T. Computational movement analysis. In *Springer Handbook of Geographic Information*; Springer: Berlin, Germany, 2012; pp. 423–438.
36. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996.
37. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, 1–24 July 1967.
38. Agrawal, R.; Srikant, R. Mining sequential patterns. In Proceedings of the Eleventh International Conference on Data Engineering, Taipei, Taiwan, 6–10 March 1995.
39. Han, J.; Pei, J.; Yin, Y. Mining frequent patterns without candidate generation. *Knowl. Dis.* **2004**. [[CrossRef](#)]
40. Baeza-Yates, R.; Gonnet, G.H. A new approach to text searching. *Commun. ACM* **1992**, *35*, 74–82. [[CrossRef](#)]
41. DEBS 2013. Available online: <http://www.orgs.ttu.edu/debs2013/index.php?goto=cfchallengedetails> (accessed on 7 November 2016).
42. Map Construction. Available online: <http://www.mapconstruction.org> (accessed on 8 January 2016).
43. Ester, M.; Kriegel, H.P.; Sander, J.; Wimmer, M.; Xu, X. Incremental clustering for mining in a data warehousing environment. *VLDB* **1998**, *98*, 323–333.
44. Lin, J.; Vlachos, M.; Keogh, E.; Gunopulos, D. Iterative incremental clustering of time series. In *International Conference on Extending Database Technology*; Springer: Berlin, Germany, 2004; pp. 106–122.
45. Cheng, H.; Yan, X.; Han, J. IncSpan: Incremental mining of sequential patterns in large database. In Proceedings of the Tenth ACM SIGKDD, New York, NY, USA, 22–25 August 2004.
46. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*; Elsevier: Amsterdam, The Netherlands, 2011.



© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).