data

MDPI

Article

How to Make Sense of Team Sport Data: From Acquisition to Data Modeling and Research Aspects

Manuel Stein [1,*], Halldór Janetzko [2], Daniel Seebacher [1], Alexander Jäger [1], Manuel Nagel [3], Jürgen Hölsch [1], Sven Kosub [1], Tobias Schreck [4], Daniel A. Keim [1] and Michael Grossniklaus [1]

1 Department of Computer and Information Science, University of Konstanz, 78457 Konstanz, Germany; Daniel.Seebacher@uni-konstanz.de (D.S.); Alexander.Jaeger@uni-konstanz.de (A.J.); Juergen.Hoelsch@uni-konstanz.de (J.H.); Sven.Kosub@uni-konstanz.de (S.K.); Daniel.Keim@uni-konstanz.de (D.A.K.); Michael.Grossniklaus@uni-konstanz.de (M.G.)

2 Department of Geography, University of Zurich, 8057 Zurich, Switzerland; Halldor.Janetzko@geo.uzh.ch

3 Department of Biology, University of Copenhagen, 2100 Copenhagen, Denmark; Manuel.Nagel@bio.ku.dk

4 Institute for Computer Graphics and Knowledge Visualization, Graz University of Technology, 8010 Graz, Austria; Tobias.Schreck@cgv.tugraz.at

* Correspondence: Manuel.Stein@uni-konstanz.de; Tel.: +49-07531-88-3076

Academic Editors: Jamal Jokar Arsanjani, Marco Helbich, Amin Tayyebi and Amit Birenboim
Received: 24 September 2016; Accepted: 23 December 2016; Published: 1 January 2017

Abstract: Automatic and interactive data analysis is instrumental in making use of increasing amounts of complex data. Owing to novel sensor modalities, analysis of data generated in professional team sport leagues such as soccer, baseball, and basketball has recently become of concern, with potentially high commercial and research interest. The analysis of team ball games can serve many goals, e.g., in coaching to understand effects of strategies and tactics, or to derive insights improving performance. Also, it is often decisive to trainers and analysts to understand why a certain movement of a player or groups of players happened, and what the respective influencing factors are. We consider team sport as group movement including collaboration and competition of individuals following specific rule sets. Analyzing team sports is a challenging problem as it involves joint understanding of heterogeneous data perspectives, including high-dimensional, video, and movement data, as well as considering team behavior and rules (constraints) given in the particular team sport. We identify important components of team sport data, exemplified by the soccer case, and explain how to analyze team sport data in general. We identify challenges arising when facing these data sets and we propose a multi-facet view and analysis including pattern detection, context-aware analysis, and visual explanation. We also present applicable methods and technologies covering the heterogeneous aspects in team sport data.

Keywords: sport analytics; visual analytics; high frequency spatio-temporal data

1. Introduction

Recent progress in sensor development results in increasing interest in recording and analyzing movement in team sports. In this article, we focus on team sports than can be classified as invasive team ball games with two opposing teams competing against each other and trying to score more points than the opponent to win a game. We have chosen this specific focus for two reasons. First, while the interaction of opposing teams in invasive team sports makes the analysis of sports data more challenging, it also opens up more opportunities for findings. Second, many of the world's most popular team sports, e.g., soccer, football, basketball, hockey, rugby, handball, etc. are invasive. Due to the popularity of these sports, the availability of corresponding data sets and the interest

in their analysis are currently on the rise. Professional team sport companies invest substantial resources to analyze the own team's performance as well as the performance of future opposing teams. Various aspects and several data sources are important descriptors for the performance of a team. In practice, some of these data sets are kept confidential by respective stakeholders, e.g., when they contain exact movement trajectories. Other data sets, e.g., basic statistics, are publicly available for analysis purposes (see Section 2.1 for several examples).

Depending on the available data, different analysis tasks can be executed. Analysts usually do not only want to have information about the *what* (e.g., "Team A won against Team B" or "Player X passed more often than player Y") but instead want to investigate the *why* behind these facts. There is a need to understand why a certain movement happened and what the influencing factors were. For example, why did a player decide to move to Point A instead of Point B and what influence did this movement decision have on members of the own and opposing teams. The results of such analyses will help, e.g., in scouting or training. However, analysis often focuses on pure statistical approaches. For decades, movement and tactical analysis has been done manually by inspecting video recordings of past matches.

In this article, we give an overview of how to work with team sport data in general. Therefore, we introduce the various data types that are available and relevant for team sport analytics. Furthermore, we highlight the challenges that need to be overcome when gathering and working with team sport data. We focus on the different research aspects arising as displayed in Figure 1 with respect to the set of heterogeneous data. Figure 1 shows that the arising research aspects can be grouped into specific domains. Data acquisition describes what needs to be done at first to get the data, e.g., through video processing. The context domain allows us to enrich the data (e.g., through data fusion) after the acquisition with useful additional information. After data acquisition and enrichment the analysis domain allows us to search for patterns. The resulting team sport analysis is on the basis of high-dimensional data that contains time series as well as trajectory data.

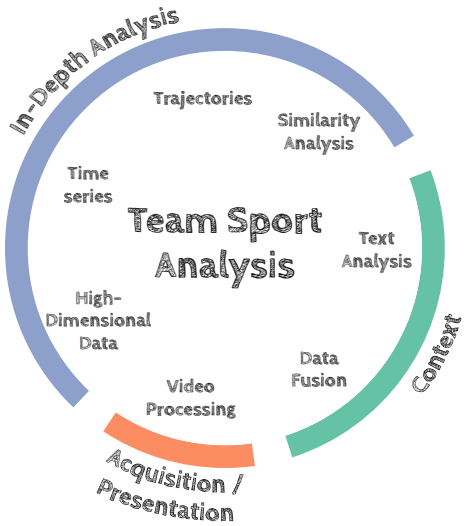


Figure 1. From acquisition (e.g., video processing) and data enrichment (e.g., data fusion) through context information to in-depth analysis tasks (e.g., trajectory analysis) on the raw data, many research fields are covered when analyzing team sport data.

We show which general computer science problems can be addressed while working on team sport data and propose our methodology to handle these challenges. Ultimately, research on team sport analytics not only influences the sport domain but other fields as well. We will describe the impacts on other sciences such as biology with a focus on collective behavior analysis. Throughout this article, we will use soccer as a prime example for our proposed methodology since it is a highly popular team

sport. Nevertheless, we present a general overview of data, methods, and tasks that are applicable to all invasive team ball games. We contribute a concise description of the enablers for data-driven sports analysis. We will start with sensing the team sport by relevant sensors and data sources. By abstracting the data space, we can reveal general research aspects being related to data mining and visualization. We will identify Visual Analytics as an important analysis methodology when dealing with such complex and heterogeneous data analysis questions by analyst users in interactive systems.

2. Team Sport Data

Data acquisition for analysis in team sport can be achieved by various ways related to the variety of data types and sources. Characterizing team sports data is either possible by describing different technical aspects as for instance the used acquisition method (e.g., optical recognition, local positioning systems, triangulation, or manual recording) or by discussing the different data types arising from different data sources. In the following sections, we introduce various data sources in detail as depicted in Figure 2 and describe them in detail highlighting important technical aspects. Practically, most data in invasive team sports (like player movement, events, and descriptive statistics) are extracted from video and sensor data. The available information can be enriched by context data such as the location of the match and weather information. Relevant contextual information can also be obtained from live streams in social media channels, which may reveal interesting facets of the game from an audience perspective. The latter is again presented in heterogeneous data formats; depending on the social media channel, this may assume the form of text, images, or video feeds. In the following, we will elaborate each data type with information about data characteristics such as size, accuracy, and resolution. Furthermore, we will give an overview with respect to where such data can potentially be obtained from (i.e., which company offers which service). Whenever applicable, we will introduce possible architectural requirements, e.g., information about hardware or analytics supporting databases and index structures. We believe that the most comprehensive analysis covers and combines all of the mentioned data types below.

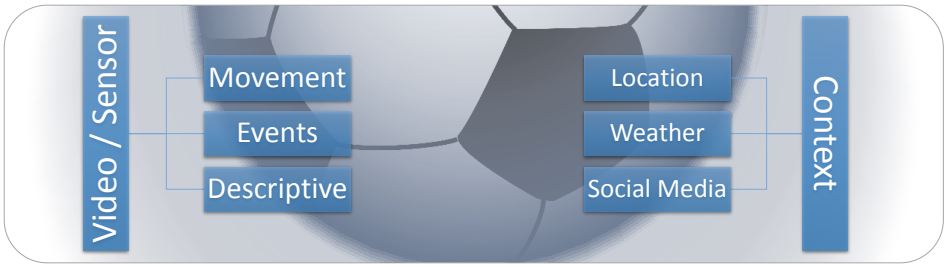


Figure 2. Relations and hierarchies between different data types in team sport. Most data can be extracted either from video or sensor data. Additional information is provided by supplemental context data.

2.1. Video and Sensor Data

Video recordings are ubiquitous in invasive team sports and there is an increasing demand for professional analyses. Video recordings range from television recordings from mass media with various perspectives (e.g., Sky TV has 24 cameras on the ground for soccer matches [1] while for the NFL up to 70 cameras are used during a super bowl match [2]) to professional recordings carried out by the teams themselves. Consequently, video data generally can be considered as the most available data source and for example can be obtained by capturing television recordings of professional matches. Companies like STATS [3] or Opta [4] offer the service of extracting movement, event, and statistic data based on their own recordings. Working directly with video data is much cheaper than assigning a professional company to track a team's players, events, and statistics. However, the extraction of player movement from video recordings is a non-trivial task. Nevertheless, due to recent progress

in the field of image and video processing, recent publications showed the feasibility of extracting movement data based on video sources [5–7].

Another possibility recording team sports is capturing the players' movement with sensors directly attached to players or game objects (e.g., ball, sidelines, targets, etc.). Practically, the applicability of this acquisition modality is depending on the legislation adopted by the sports associations. This data acquisition may be partially restricted in some invasive team sports (as it was for example in soccer until 2015) while allowed for others. For instance, the NFL [8] allows active tracking by sensors placed on the player shoulders in cooperation with Zebra Technologies [9]. Sensors may allow real-time capturing of data via wireless data transfer. The dataset of the ACM DEBS 2013 Grand Challenge is a perfect example for these kinds of sensor data [10]. Video data is often the basis for movement extraction. Nevertheless, videos can be enriched with analysis results visualized in the raw video data respecting perspectives and angles.

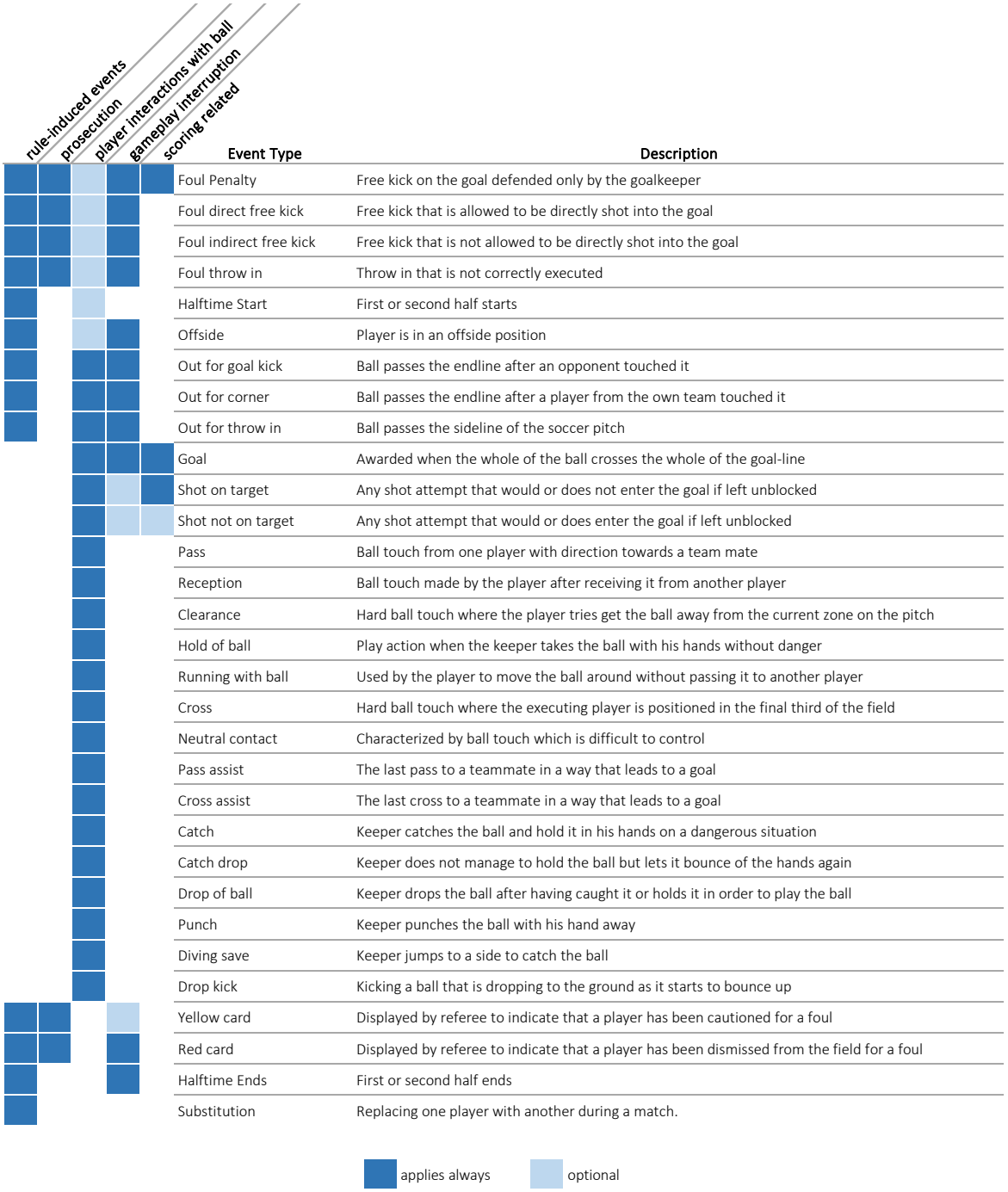
2.1.1. Movement Data

Movement data describes where an actor or game object (e.g., a player or the ball) is located at a specific point in time. Locations are usually measured by local coordinate systems with reference to the game pitch. These measurements contain as a minimum the x- and y-coordinates and sometimes also the z-coordinate. Depending on the acquisition technique, the positions are usually sampled around 10 to 25 times per second (Hz). In the sample dataset of the ACM DEBS 2013 Grand Challenge, each player has two sensors (one in each shoe; the keeper additionally has sensors in gloves) that each transmit position reports at 200 Hz. The ball contains one sensor that transmits at 2 kHz. For each sensor, a timestamp (X, Y, Z)-position, as well as both overall and component-based velocity and acceleration is retrievable. Storing a full game results in approximately 10 GB of data. Body postures are not recorded in the sample dataset due to the large amount of active sensors needed for robust posture recognition. Nevertheless, we expect these data arising in the foreseeable future especially with passive recognition from video data [11,12].

2.1.2. Event Data

Sport games can be described by an ordered sequence of events. We define events as actions being match-relevant and happening during the match. Events typically are derived from movement data by automatic video analysis; also manual annotation is possible and done professionally by some data providers. From a technical perspective, events are timestamped occurrences of previously known and defined categories optionally annotated with spatial coordinates or additional information as involved players. Most events will be directly ball related and corresponding to actions with the ball (for instance passes or dribbling). Other events may be time-dependent (e.g., start and end of a play period) or not directly dependent on the ball (e.g., a foul situation during a free kick). A multitude of methods to detect such events exist (in fact too many to list them here) ranging from manual annotation to fully automated systems [13–17]. The resulting streams of semantic data are already widely used in industry and scientific communities [18–20].

In practice, events might lack in accuracy, as they are usually annotated manually or as fully automatic recognition may produce false positive and negative events. The company Opta, for example, is commercially providing event data. As event data contains mostly information about players interacting with the ball, event data enables us to conduct overall game statistics (e.g., passing networks, pass accuracy, or time between gaining the ball and shot on target).



Event Type	Description
Foul Penalty	Free kick on the goal defended only by the goalkeeper
Foul direct free kick	Free kick that is allowed to be directly shot into the goal
Foul indirect free kick	Free kick that is not allowed to be directly shot into the goal
Foul throw in	Throw in that is not correctly executed
Halftime Start	First or second half starts
Offside	Player is in an offside position
Out for goal kick	Ball passes the endline after an opponent touched it
Out for corner	Ball passes the endline after a player from the own team touched it
Out for throw in	Ball passes the sideline of the soccer pitch
Goal	Awarded when the whole of the ball crosses the whole of the goal-line
Shot on target	Any shot attempt that would or does not enter the goal if left unblocked
Shot not on target	Any shot attempt that would or does enter the goal if left unblocked
Pass	Ball touch from one player with direction towards a team mate
Reception	Ball touch made by the player after receiving it from another player
Clearance	Hard ball touch where the player tries get the ball away from the current zone on the pitch
Hold of ball	Play action when the keeper takes the ball with his hands without danger
Running with ball	Used by the player to move the ball around without passing it to another player
Cross	Hard ball touch where the executing player is positioned in the final third of the field
Neutral contact	Characterized by ball touch which is difficult to control
Pass assist	The last pass to a teammate in a way that leads to a goal
Cross assist	The last cross to a teammate in a way that leads to a goal
Catch	Keeper catches the ball and hold it in his hands on a dangerous situation
Catch drop	Keeper does not manage to hold the ball but lets it bounce of the hands again
Drop of ball	Keeper drops the ball after having caught it or holds it in order to play the ball
Punch	Keeper punches the ball with his hand away
Diving save	Keeper jumps to a side to catch the ball
Drop kick	Kicking a ball that is dropping to the ground as it starts to bounce up
Yellow card	Displayed by referee to indicate that a player has been cautioned for a foul
Red card	Displayed by referee to indicate that a player has been dismissed from the field for a foul
Halftime Ends	First or second half ends
Substitution	Replacing one player with another during a match.

Figure 3. Various different kinds of events categorized by their characteristics.

We give an example of a set of potentially relevant events for soccer analysis in Figure 3. This list of events derived from our previous works on a continuously expanded system for feature- and event-based soccer analysis [21–23]. Although arguably it may be extensible, and is not tailored towards a specific model from Sport Science, we believe it is a practical starting point for reasoning about the types of events potentially useful for analysis. The table contains the type of event, a short description of when the event is recognized, as well as a proposed categorization of events that share similar characteristics. We distinguish the following event categories:

- *Rule-induced events* are events that occur as a result of the match rules. For example, if the ball passes the sideline of the soccer pitch, it has to be thrown in again by the opposite team.

- Events tagged with *prosecution* indicate that there was a foul behavior of the related player(s) which is penalized.
- *Player interactions with ball* is about events that happen when a player is touching the ball. Observable, almost every event that gets tagged falls under this category besides yellow and red cards, the end of a halftime and a substitution.
- Events that interrupt the match gets marked as *gameplay interruption*.
- If an event has a direct relation to scoring (e.g., a shot on the goal) we mark it as *scoring related*.

Events from different categories may lend themselves to different analysis tasks. Also, they are subject to possible detection or description inaccuracies and may require parameters to work. For example, *running with ball* requires appropriate threshold setting for distance between player to ball to be recognized. We observe that events may belong to different categories mandatory or optionally. In our table, we mark events dark blue if it always belongs to the respective category. If an event is marked in lighter blue instead, the event may falls under this category, but does not have to.

2.1.3. Descriptive (Statistical) Data/Derived Data

Properties of a player or a team can be characterized by descriptive (statistical) data. Descriptive data include everything that can be counted or measured during one or several matches, for example, how often a player passes or the maximum speed and acceleration. Today, some of these descriptors can be measured automatically by tracking devices. However, most of this data is collected manually by analysts. Due to historical reasons, descriptive statistics are the most common data sources for team sport analyses, as automatic movement recording is a relatively new technical achievement.

Additionally, datasets are available to access historical data such as prior match results between the teams, squad-memberships, final standings of leagues and career records. Beside several commercial providers and unstructured text websites, freely available and machine readable sources can be found online [24–28]. Bergmann et al. [29] show how such data sources can be matched and stored.

2.2. Context

External and Environmental factors can be obtained by matching, location, date of a match with external data sources, such as characteristics of the stadium (capacity, open/closed roof) [30]. Historical and current weather records [31] can be used to heuristically estimate the quality of the playing field and possible effects on player performance due to high or low temperatures and air humidity. Furthermore, Ekin et al. [14,32] showed that weather and stadium context can influence the effectiveness of video based tracking and needs to be taken into account.

2.2.1. News and Social Media

Community-generated reports about games can be gathered from social media platforms such as Twitter [33], reddit [34], or Wikipedia [35]. Yucesoy et al. [36] investigate for example the relation of tennis players' popularities and their performance based on visits to their Wikipedia entries. Users write about context and progress before, during and after the event. Many of these services provide APIs to gather at least partial datasets for matches from the past and in real time.

Works from traditional news sources are collected in several projects similar to the European Media Monitor [37]. A real time stream of world wide journalistic articles can be retrieved with additional meta data and event detection added by the data provider.

3. Abstracting the Data Space

We described the multitude of facets and aspects of data in team sports in the sections above and pointed to the heterogeneity of recorded team sport events. However, what we still need to discuss are the abstract ingredients of team sport. We investigate the nested abstraction levels depicted in Figure 4

following a bottom-up approach starting with team sport and going all the way to the basic data types, namely geospatial and temporal data.

In team sport data, there are two competing groups that have opposed predefined objectives, meaning that (if the match did not end in a tie) only one of the two groups can achieve their objective and the other group loses the game. The challenge of analyzing team sport data is that movement is restricted by a pitch and rules, driven by the predetermined objective, and influenced by the movement of own and opposing team players. For illustration purpose, we exemplify these properties in American football: the movement of players and teams is limited to the pitch. The movement of the two opposing groups is clearly driven by a predetermined goal. The group possessing the ball wants to cross "the opposition's goal line with the ball, or catch or collect the ball in the end zone" [38]. The counter-objective of the opposing team is to prevent this from happening and to gain possession of the ball. American football is a very good example to illustrate how groups and individuals influence their movement mutually. Examples include the defensive line trying to block the running back (group influencing an individual and vice versa), the offensive line pushing against the defensive line (group influencing group) and corner backs covering the receivers (one individual influencing one individual). Rugby is another example for team sport being analyzed nowadays as shown by Cintia et al. [39]. Gudmundsson et al. [40] published recently a comprehensive survey covering several aspects in team sports data. Our paper is aimed at both a survey and a concept paper, and our main goal is to widen the scope of considerations for team sports analysis as compared to previous surveys. We build on top of this survey by discussing the possible manifold data sources and their influences on the analysis process, hence widening traditional sensor data-based analysis schemes. As we point out, existing techniques often stem from the machine learning side, and may underweight the importance of visual-interactive interfaces for understanding team sport data and bringing background knowledge to the analysis. Therefore, we show several examples for Visual Analytics interfaces dealing with team sport data, to inspire thinking about possible user-oriented sport data analysis. Figures 5–7 exemplify novel visual analytics interfaces which allow interactive analysis for aspects including spatial movement patterns, feature-based analysis and space-time segmentation, and visual comparison of events. Team Sport Analysis requires, in our view, and interdisciplinary approach where Sport Science, Behavioral Science, and Data Science including Data Visualization can all benefit to better using and understanding team sports data. By outlining this big picture, we indicate potential for interdisciplinary advance of research in the area.

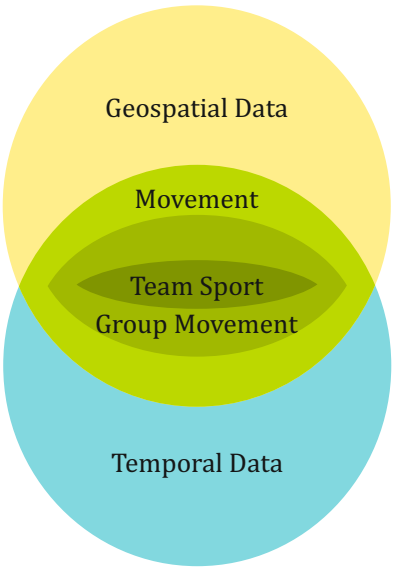
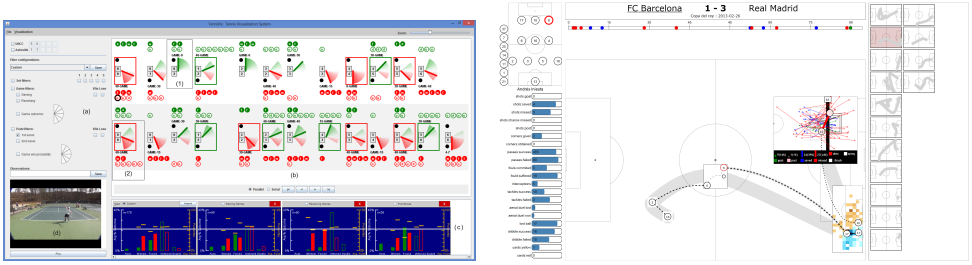


Figure 4. The abstract ingredients of team sport.

One abstraction level higher, team sport can be seen as a specialization of group movement as shown in Figure 4. Group movement can contain either cooperative or competitive behavior or any combinations of both being exactly what can be observed in team sport. Group movement is also studied in biology and behavioral science. For example, there exists the so-called Movebank Project [41], a database about animal movement on a global scale. In nature, group movement occurs on very different scales, ranging from two (courtship or mating) to a thousand (social insects) individuals. The technical challenges of tracking individuals in animal groups still exists: even under laboratory conditions only parts of collectives can be marked and tracked. Consequently, the field of behavioral biology focuses on animal movement based on recent advances in data acquisition enabling a high temporal and spatial resolution of animal movement. GPS-tracking devices and classical tracking are used to record animal movements in the wild on a more global scale while video tracking analysis are used for rather locally restricted movements in the wild or under laboratory conditions. The increase in resolution led to an increased data volume to be analyzed. One of our visions is to research similarities and differences of team sport and animal collectives. The question is which patterns occur in both domains and which techniques are suitable in which analysis scenario.

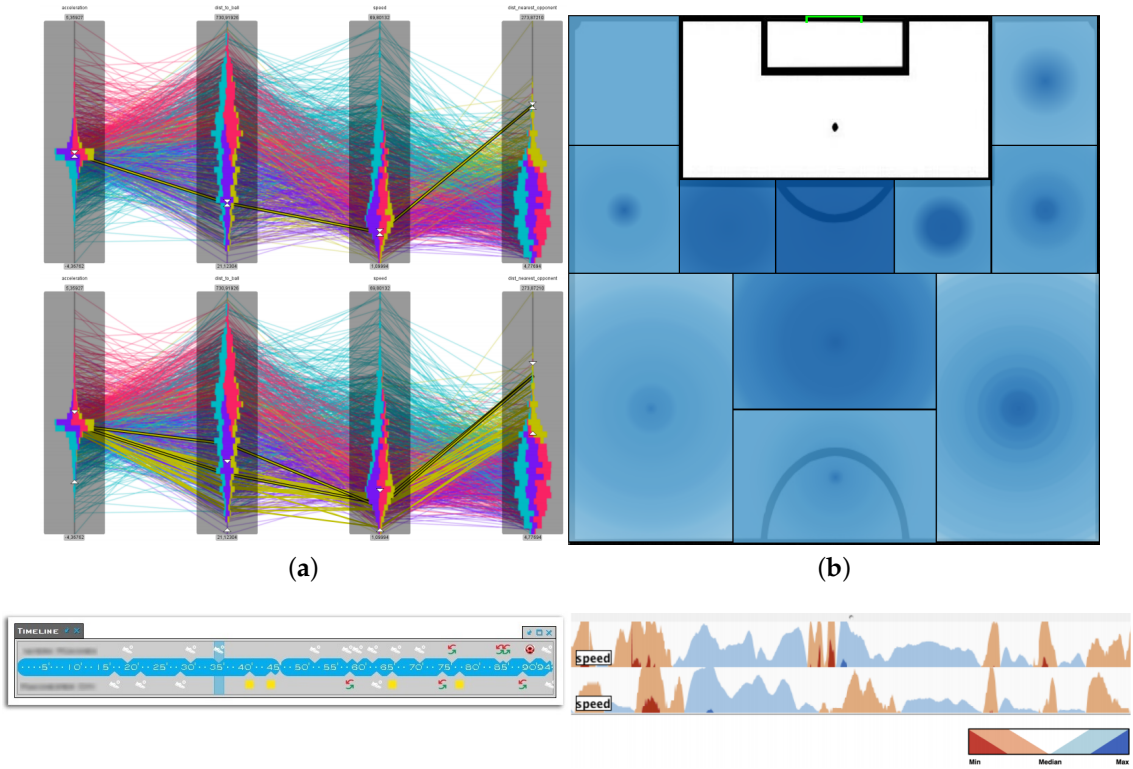
On the highest abstraction level, movement in general is defined according to Andrienko et al. [42] as the path of moving entities through space and time. Andrienko et al. [43] suggest analyzing movement at two different granularities, of the individuals and of the group as a whole. Nevertheless, we also have to deal with pure time-series or time-stamped event data and with spatial topology data. There exists many works dealing with pure time series analysis mostly focusing on similarity calculation and pattern analysis. A good overview over state of the art methods in temporal data mining is given by Fu [44] and visualization methods for time series are discussed for instance by Aigner et al. [45]. Temporal analyses alone cannot explain all the behavior observed as important spatial aspects are neglected. Nevertheless, temporal visualizations are crucial to convey temporal patterns. From a pure spatial perspective, computational geometry can be the starting point to analyze trajectories in team sports data as described by De Berg et al. in [46]. However, pure geometric approaches do not cover the temporal aspects of movement. Consequently, spatio-temporal analysis are key for a successful analysis. For instance, Kang et al. [47] represent the movement of soccer players as trajectories and propose a model which quantitatively expresses the performance of players based on the relationships between the player trajectories and the ball. Another player network based analysis with respect to performance is performed by Cintia et al. [48,49]. Football strategies are investigated based on network theory analysis by Pena et al. in [50]. A different spatial topology based approach would be the use of graph theory in sports. Bourbousson et al. [51] analyze basketball matches and Clemente et al. [52] use graph theory to analyze soccer matches.

Working with team sport data, we can apply methods of and contribute to different areas of computer science. One obvious example could be the research field of big data analytics [53] proposed by Russom. According to Russom *Big Data* is defined by the three "*Vs*", namely *Volume*, *Velocity* and *Variety*. Often a fourth "*V*" is mentioned, namely the *Veracity*, which is the uncertainty in the data, see e.g., Buhl et al. [54]. These are all attributes which we are facing when working with team sport data. For instance, we have to work with large volumes of data when analyzing video data of several leagues for multiple seasons or we have to consider the veracity of social media data.



(**a**) (**b**)

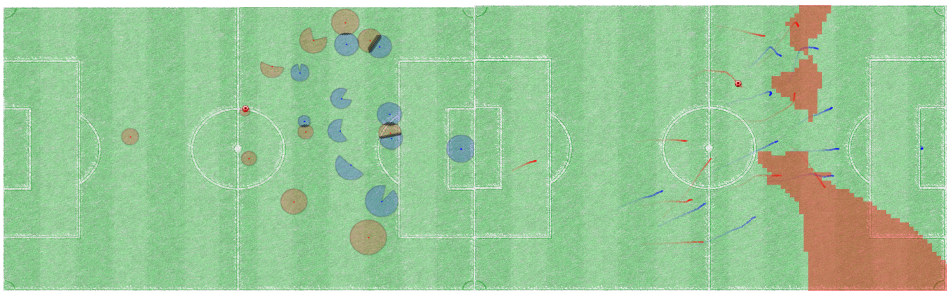
Figure 5. Two recent systems that aim to improve the understanding of sport data by several visualization techniques. (**a**) TenniVis [55]; (**b**) SoccerStories [56].



(**a**) (**b**)

(**c**) (**d**)

Figure 6. Example uses of different information visualization techniques in recent publications in the field of visual sport analytics. (**a**) Enhanced parallel coordinates [57] displaying, for a single player, the average values of four features within a phase. Each line represents a phase, which is defined as a time-interval, in which player behavior doesn't change; (**b**) Spatial visualization [58] highlighting the dangerousness of set plays executed in various regions of the soccer pitch. The dangerousness for each region is mapped to the blue hue. White meaning safe, dark blue meaning dangerousnes; (**c**) Temporal visualization [22] displaying the occurence of user-selected events, like fouls, goals or exchanges; (**d**) Horizon graph [21] showing the feature *speed* for two defense players within a set time interval.



(**a**) (**b**)

Figure 7. Interaction Spaces (**a**) [23] are designed to visualize the surrounding area each player aims to control; Free Spaces (**b**) [23] describe how much players are put under pressure.

4. Research Aspects

In the previous sections, we motivated why it is interesting to spend time on the analysis of team sports data. We gave an overview to the available sources and types in team sport data. By abstracting the data space, we identified connected research fields in computer science. In this section, we envision desired outcomes of team sport analyses and discuss pressing research challenges.

4.1. Definitions

Before we go into detail about the arising research challenges, we here provide required definitions.

4.1.1. Behavior

Behavior is an (re)action of an organism with a desired goal provoked by external stimuli or internal states such as motivation and arousal. For example, hunger provokes the goal "find food" and would be an internal motivational state that elicits foraging behavior. During foraging, external stimuli are used to fine-tune the foraging behavior and adapt behavioral patterns to instantaneous conditions. If a forager encounters an obstacle in its way the movement will be modified to overcome the challenge (temporary goal: pass the obstacle and go on with primary goal) or if any danger occurs, foraging behavior might be replaced by a behavior like fighting or escaping as the goal has changed essentially (goal: "I will survive!"). The definition of movement is highly context specific. In general, movement can be differentiated into movements in space like for migrating animals or movements referring to the movement ability of an individual restricted by anatomical and physiological capabilities. Movement behavior in space often is built up by distinct behavioral patterns. Orientation behavior elicited by the goal to know the position relative to the position that should be reached, walking/flying/crawling behavior etc. that describes the kind of movement an organism is operating with the goal to reach the destination. In Drosophila larvae these two phases of spatial movement are not simultaneously performed and therefore distinguishable by the observer [59]. In other organisms orientation and moving occur at the same time and orientation behavior is often less visible to the observer. These are only two aspects that are crucial throughout spatial moving and further aspects (arousal, motivation etc.) should be taken into consideration when the causes for a spatial movement are in question.

4.1.2. Movement Pattern

Formalizing movement patterns needs to reflect the variability of real-world movement. Today, we are able to automatically track groups of moving entities and need techniques to describe not only individual movement but also collective movement. In this paper, we extend the definition of movement patterns as "concise description of frequent behaviors, in terms of both space (i.e., the regions of space visited during movements) and time (i.e., the duration of movements)"

by Gianotti et al. [60]. Movement patterns in team sport need to reflect collective group actions or reactions. Consequently, movement patterns are not necessarily frequently observed behaviors but coordinated in order to achieve a certain goal. A certain goal can often be achieved by alternative behavioral patterns and with experience and higher cognitive abilities the alternatives can be anticipated and planned. The ability to plan and estimate different strategies to reach a certain goal will lead to tactical considerations, especially in team sports, but might also be applied to animals with higher cognitive abilities. Formalized movement patterns can help to semantically annotate or interpret relevant intentions of the moving entities.

4.1.3. Group Behavior and Movement

Analysis of movement patterns or, more general, of behavior of individuals can nowadays be combined with simultaneous recordings of diverse parameters. 3D accelerometer data (behavior and energy use), remote sensing data (e.g., weather, habitat) and/or data about interactions with other tagged individuals can be assessed additionally to traditional tracking methods/data (reviewed by Kays et al. [61]). The increasing amount of data about crucial impacts on individual behavior can then be used to explain or predict behavioral responses of an individual on the different levels of behavior (see the works of Bergner [62], Hogan [63] and Sumpter [64]. The investigation of behavior becomes even more complex when (social) interactions of individuals are added like in, for example, team sports or collective behavior in animals.

The aggregation of individuals often leads to "collective behavior [that] is defined as the behavior of aggregates whose interaction is 'affected by some sense that they constitute a group' but who do not have procedures for selecting or identifying leaders or members" [65]. Flocks of birds and shoals of fish are the textbook example for collective behavior in animals. The aggregate of individuals moves as a whole structure based on the individual behavioral patterns that are again influenced by the fact that the individual behavior is affected by the aggregate. This interdependency of individual and group behavior causes that no (obvious) leader can be identified by the observer. Different mathematical models can be applied on collective behavior in animals (e.g., by Sumpter [64]) and synergies that arise in team sports [66] might also apply to synergies arising in superorganisms like social insects. Although the 'final cause' (consequences of behavior, definitions by Hogan [63]) are very different between team sports (e.g., winning a match) and animals (e.g., passing genes to the next generation), certain structures of behavior might be shared across team sports and superorganisms [67]. The integration of different research fields might allow to apply principles revealed in animal superorganisms to team sports performance analysis and vice versa [67]. The analysis of movement patterns of individuals therefore will benefit from a close collaboration by researchers of different fields in which new movement analysis and visualization methodologies can arise [68] and help to understand collective behavior in the wild or/and under restricted conditions like team sports or lab experiments.

4.2. Research Challenges

Analysts inspecting team sport data usually pursue various goals such as identifying strengths and weaknesses of their own team and of opposing teams. The insights are used to improve training and raise the team's awareness when preparing for upcoming matches. To identify strengths and weaknesses, analysts, for example, want to understand why a team won a past match. Widespread statistical approaches, however, typically provide only a basic overview (aggregation) over the characteristics of a match. Therefore, one of the most challenging tasks for statistical analysis is the identification of expressive statistical features that help to gain more insight into factors that could influence the outcome of a match. However, in order to get a better understanding of the outcome of a match, we have to analyze our data on a more fine-grained level (e.g., movement of single players). One example for the inherent complexity is shown in Table 1 with a real world example. Possession, shots and duel quota, among other statistics, are usually used to statistically compare

teams. They indicate that both teams are equally strong. Considering that Germany beat Brazil 7–1 this is quite surprising. A different sight on the game outcome comes with two new statistics called Packing and IMPECT, both developed by the Impect GmbH [69]. Packing is the number of outplayed players and IMPECT is the number of outplayed defenders. A detailed explanation of both statistics is given by Regenhuber [70]. These statistics are better suited to explain the outcome of this match, with Germany having outplayed approximately 60% more defenders than Brazil.

Table 1. Statistics of a real-world game: Brazil vs. Germany (2014 FIFA World Cup).

Statistic	Brazil	Germany
Goals	1	7
Possession	52%	48%
Shots	18	14
Duel Quota	51%	49%
Packing	341	402
IMPECT	53	84

However, even improved statistical measures cannot provide more than an indication about why a team won in a match. Features like packing in our example above show that Germany did a better job in outplaying Brazilian players, but we cannot see how these attacks were played in particular. Analysts want to find out why and how something important happened in a match as well as filter out noise obscuring interesting patterns. There are potentially three different ways how analysts identify patterns. The first way is to detect the pattern completely manually supported with appropriate visualizations. The second way are semi-supervised approaches, where analysts—at the beginning of the analysis—define a template of how the pattern should look like. Based on the template the system starts to look for patterns matching the given criteria. Lastly, analysts might have no clue what to look for and therefore the system should identify everything of general interest employing high-dimensional data descriptors.

Explaining movement patterns, however, is very complex, as the movement of a single actor is depending on the movement of all other actors. Because of these interdependencies, almost every action causes a reaction. Adding to the complexity is that typically, there are certain role definition in a sports team that influence movement. Particularly, team captain and goal keeper may steer or influence the movement patterns of team players, e.g., following a tactical decision to defend or press certain players. The role of such leadership effects is prominent not only in soccer analytics, but also a question of leader-follower relationships in domains like expeditions and explorations, military operations etc. This complexity in soccer analytic is added to by the fact that we do not only have to look at the interdependencies within a team but also the interdependencies between two opposing teams.

We can abstract this research challenge by generalizing our moving players to moving entities. This abstracted view allows us to widen our analysis possibilities to other domains, because conflicts of interest about where to go and what to do are a primary challenge of living in a collective. Groups do occur in nature hinting to correlating biological fitness. In the analysis of collective behavior, researchers are interested in detecting the way animals achieve consensus in stable groups with stratified social relationships [71]. When analyzing collective behavior biologists want to determine factors directing the movement of, for example, birds flocks or fish shoals. In both cases the group formation decreases the risk of predation and thus increases the fitness of a single individual. Consequently, the required analysis possibilities overlap and allow the transfer of developed techniques from one domain to the other helping to solve challenges.

Context-aware analysis is another important aspect when exploring team sport data. Analysts might have identified interesting patterns and want to explore them in-depth enriched with context information. Movement is more than just x- and y-coordinates: entities are motivated to

move by intrinsic and extrinsic contextual factors. Context-aware analysis means that insights into group dynamic behavior are enabled by incorporating collective movement models in the analysis process. We thereby even incorporate research from the field of collective animal movement [72–77].

Finally, experts want to explain why, when and how specific movement behavior is expressed because of tactical behavior. Tactical behavior, in our understanding, represents the overall effort practiced on the field to eliminate the factor of luck as much as possible. Analysts want to retrieve explanations of observed cooperative movement patterns in reaction to competitive movement patterns by enhanced visualizations. Another open issue is tracking external influences as coach advices and the corresponding team reactions.

The challenges arising with team sport data are not only concerning the analysis of movement itself. Technical challenges as efficient data storage, querying, and processing arise as well. Improved tracking techniques and devices led to an increase in data volume gathered from team sports as well for wildlife observations in behavioral biology. Large datasets and in some cases vast amount of individuals tracked require new and efficient handling methods. The need for novel streaming technologies that allow real time analysis enabling analysts and coaches to gain information even, for example, during the halftime break is increasing. Consequently, creating a encompassing view of team sport data, integrating trajectories, time series, similarity search, high dimensional data, text, and video image processing/analysis will be the major challenge promising large synergy effects and real insights.

5. Methodology

Analyzing team sports data requires a toolbox of methods from different domains including sport domain knowledge. From our previous work we consider the following areas non-exclusively important. *Models* help in deriving patterns and features describing observed behavior in the domain. Models can stem from various perspectives. For example, in a data-driven way models can be obtained from statistical analysis and mathematical formulations; or from concepts developed in sport science. *Data mining* is the corresponding domain in computer science and uses mathematical and statistical approaches. The communication of modeling results is often achieved by methods from the *Information Visualization* domain. Information visualization is part of computer science and strongly connected to computer graphics and cognitive psychology. Recent research proposed a synergetic combination of data mining and information visualization by a *Visual Analytics* [78]. The core idea here is to implement steerable data mining methods and immediate visual feedback of the analysis results. Highly interactive analysis systems are the outcome supporting the integration of domain knowledge. We next detail aspects in the main areas of data modeling, data mining, interactive visualization and visual analytics.

5.1. Data Modeling

Generally speaking, modeling is about giving structure to the problem of sport analysis, by prescribing which aspects are of importance to the analysis. Ultimately, this needs to be informed by the task of the analysis. Different examples are, e.g., the short-term performance analysis of a single team member, versus the long-term analysis of the team performance or its evolution over time. The modeling process will eventually identify a set of variables and/or events to observe, and a quantitative scheme to aggregate and/or compare the measures. We distinguish two main approaches to guide the modeling: Domain-specific modeling is based on theories and concepts from *Sport Science* [79,80], which generalize relationships between actions and outcomes in the respective sport domain. On the other hand, *data-driven* or *explorative* modeling does typically not assume previous knowledge about the domain, but is guided by dependencies found in the data directly. In practice, both approaches often go hand-in-hand: There are expectation and theoretical models about typical dependencies in the sports events, and the analysis measures are derived to validate and quantify the expected dependencies. On the other hand, patterns that one may observe inconsistently

in larger amounts of measurement data may lead to new domain-specific theories and hence guide the modeling approach conceptually.

Team sport is fundamentally about space-time interaction of players—understood as a combination of operation, communication, and strategy—with the most influential interaction being the interference of the opposing team which excludes in many respects absolute scales of measurement. The ongoing lack of reliable key performance indicators for soccer teams points in this direction (see, e.g., the illuminative discussion in [81,82]). A deeper understanding of the interaction structure and dynamics between and within teams during a match is meaningful for making progress in team sports analysis. Relative phases [83,84], couplings [85], invasion profiles [86], or centralities in passing networks [87] represent exemplary initial concepts.

As the data collected for a match shows only one realization of a contingent situation ("what happened"), the challenge is to determine the (path-dependent) set of *possible* actions executable in that situation ("what not happened"). Operationally, the presence of the opposing team first and foremost restricts the space-time regions on the pitch accessible for meaningful action of a focal actor. On the communication level, selecting an action from the set of accessible actions ("decision") depends furthermore on the intra-team movement patterns, space cognition, and information processing. Here, coming up with realistic and feasible action models is very demanding [88–90], but worthwhile for answering to questions like "Which pass is best to execute given a set of possible passes?" Sole ball-oriented data is very limiting in respect thereof, as no information on players without the ball is available; the situation becomes much better with additional positional data of all players (see, e.g., [90]). Moreover, technology is available (at least in basketball) to track head positions and head orientations of the players providing data for a feasible inference of mental maps [91], e.g., which players are seen by a player. On the strategic level of interaction, expectations, style of play, tactical orientations, etc. determine a *normal* behavior of the teams. Data-based analyses of these aspects are thus required for an unbiased evaluation of action sets. However, corresponding techniques are currently less developed. Methods of game theory could be applied, e.g., building strategic games from multi-parameter interaction models based on tactical action-reaction schemes available from expert's domain knowledge.

In principle, there are two approaches to analyze the interaction of two teams in competitive dyads (matches). First, characterize matches by the behavioral characteristics shown by the teams, i.e., observed events are directly assigned to individuals, groups, or teams with positive or negative evaluation. This is the standard bottom-up approach. Second, characterize the behavior of the teams by the characteristics of matches, i.e., observed events add up to global information on the match before they are projected onto the teams. This can be classified as top-down. Next to context information (match type/competition, venue, audience, weather, etc.), typical global characteristics are parameters related to match speed. Speed can be hardly assigned to one side only. Though there is no standard definition of match speed, the growing popularity of packing rates (which relate the space occupied by a number of opposing players to time represented by pass duration) can be seen as a need in this global information.

5.2. Data Mining

Data mining is the automatic or semi-automatic discovery of patterns in data sets, which are too large to analyze manually. It is the analysis stage of the knowledge discovery in databases process, which also encompasses *Selection, Pre-Processing, Transformation and Interpretation* [92]. Data mining itself is an umbrella term, which encompasses various methods from different computer science fields. It uses, inter alia, methods from machine learning, artificial intelligence, statistics and database systems [93]. Fayyad et al. [92] group data mining techniques into six general categories. In the remainder of this section, we use these categories in order to review existing data mining techniques and their possible application to sports data. Apart from example use cases, we will also survey

quantitative findings that have been obtained in previous work from applying data mining techniques to sports data.

5.2.1. Clustering

Clustering is the grouping of similar objects into clusters. The concept of a cluster, however, is not clearly defined, which is why so many different clustering algorithms exist [94]. Generally speaking, a cluster is a set of objects which are more similar to each other, than to those in other clusters, however similar is defined. One could use one of the clustering algorithm by Lee et al. [95] to identify common movement patterns of individuals or groups. In American Football this could be used to identify the most often used passing patterns of a team. An example where clustering is used is the work of Janetzko et al. [21]. Here clustering is used to find common behavioral patterns of individual players. For this the *k-means* [96] algorithm is used on feature vectors consisting of various features like player speed or distance to ball.

5.2.2. Classification

"Classification is learning a function that maps (classifies) a data item into one of several predefined classes" [92]. These classes can either be defined manually or alternatively generated automatically using clustering, which is described in Section 5.2.1. This data mining method can, for example, be used—to stay with the American Football example—to automatically identify in which passing pattern a new pass can be classified. We successfully used classification in one of our previous works [22] to detect dangerous situations. A dangerous situation in our case was a shot on the goal of a team. First we trained classifiers with various features which were collected shortly before the event "shot on goal". Afterwards the classifiers were applied on the data, to identify periods with similar feature values, but where the event "shot on goal" did not occur. Thus detecting potentially dangerous situations, which were previously unknown. Users then judged these situations on their dangerousness. This information was used to iteratively retrain the classifiers. The effectiveness of this approach was shown in a quantitative evaluation, in which two experts were tasked with finding dangerous situations. There the experts could find dangerous situations with a F_1-*Score* of up to 66%.

5.2.3. Regression

Regression or regression analysis is the process of estimating the relationship between dependent and independent variables of an experiment [97]. Its goal is the determination of parameters of a function (such as α_1, α_2 and ϵ in $f(x) = \alpha_1 \times x + \alpha_2 \times x + \epsilon$). If a relationship exist and if a good fitting can be archived the resulting function can be used to predict future observations. Regression analysis can be used to expose statistical correlations between and model the behavior of players [98] and teams in tournaments [99].

5.2.4. Summarization

"Summarization is a key data mining concept which involves techniques for finding a compact description of a dataset" [100]. Methods like calculating the mean or standard deviation or dimensionality reduction are often used to analyze and visualize large and complex data sets. Clustering (see Section 5.2.1) can also be used as a summarization method, with the centroid used as a representative for the whole cluster. This method has beneficial use to display the complex and large datasets, which are common for team sport analysis. For instance, techniques like the well-known PCA [101] or the recent t-SNE [102] are used to visualize high-dimensional data in lower-dimensional space, for example a soccer pitch representation. Perin et al. [56] show this by proposing a visual abstraction and summarization system of, for example, certain attack paths.

5.2.5. Change and Deviation Detection

Change and deviation detection, which is often also called outlier detection, refers to the detection of observations, which do not correspond to the already existing patterns. On method to find such observations is "Grubbs' test for outliers" [103]. One possible use case in the research field of team sport analysis could be the detection of players which perform extremely better or worse than all other players.

5.2.6. Dependency Modeling

Dependency modeling or also called association rule learning is defined as the identification of significant relations between variables in the data. To identify interesting rules the measurements of confidence and support are often used [104]. These rules are often used for market basket analysis, to identify which items are often bought together. They could also be used in the field of team sport analysis, to identify which events frequently occur if, for example, a goal was scored.

5.2.7. Summary

Team sport analysis is a growing research field using complex, high-dimensional data to model the behavior of groups and individuals. There exist already many potential uses of data mining methods in the field of team sport analysis. However, it is necessary to either adapt existing methods from various research fields or to develop new algorithms to allow an effective analysis of team sport data.

5.3. Information Visualization

Information visualization is a growing research field and became quite prominent in the 1990's. There are three main tasks of visualizations being directly transferable to the sports domain: exploration, hypotheses validation, and hypotheses generation. Exploration, here introduced as an example, is usually the first step when dealing with a previously unknown data set. Overview visualizations help to identify the descriptive features or to detect interesting patterns. The most important visualization techniques for sports data are statistical visualizations as scatter plots or parallel coordinate plots [105] as well as more specialized spatial and temporal visualizations (such as displayed in Figure 5). Spatial visualizations help to investigate distributions and individual or group movement patterns. Temporal visualizations [45] show how features change over time and effect each other. Horizon graphs [106] are a good example for a space efficient visualization of time-dependent data. Pixel-based visualizations [107] are even more space efficient, encoding data values by color-coding single pixels of the display. The challenge in pixel-based visualization is an effective layout of the data items on the screen. For instance, temporal data can be laid out hierarchically with the help of a technique called Recursive Patterns [108]. However, the high complexity and multi-dimensional aspects in sports data require novel visualization techniques. Abstracting the collected data in a suitable way and pointing experts to interesting aspects is key for a successful analysis. A selection of visualizations techniques used in recent publications can be seen in Figure 6.

5.4. Visual Analytics

Combining data mining with information visualization is creating synergistic effects of human and machine. This combination is called visual analytics and enables experts to include their domain knowledge during the analysis process by interactive and steerable data mining methods and immediate visual feedback of the results. We believe that visual analytics is a very effective way to cope with the challenging data properties in the team sport domain. The research challenges described in the section above showed that although the desired outcome is clear, we need to deal with ill-defined data analysis problems. For example, detecting movement patterns being of interest to the analyst requires a semi-formal description of interests of the analyst. However, this transfer

from the sports domain to the data domain being understandable by machines is very difficult. Visual analytics proposes a transparent analysis process, where the manifold parameter choices of data mining algorithms are as comprehensible as possible. By interactive exploration of the data and parameter space, domain experts should get a feeling for their data. However, there is also translation and communication needed between sport experts and visual analytics experts. One possibility to enable a productive communication between experts of both domains could be the use of a *Liaison* as suggested by Simon et al. [109]. A very first example for a resulting system can be seen in Figure 7, where we provided methods to interactively add context to the given data.

In Figure 7a we developed so called *interaction spaces* to indicate the surrounding area each player aims to control [23]. Depending on speed and distance to the ball we determined a continuous model based on the conceptualization from the sport scientists around Grehaigne et al. [110]. We extended our model to show potential duel areas when two interaction spaces overlap. A player's interaction space is restricted to the area that can be reached before opposing players. In Figure 7b we visualize the free spaces, which we defined as the regions players from one team can reach before the players of the opposing team can. In order to verify that our definition corresponds with the ideas from domain expert, we conducted a qualitative evaluation. In this evaluation experts were asked to sketch the free spaces for a given game situation. In ~80% of the cases, at least one of the free spaces, drawn by experts, overlapped with one of our computed free spaces. Thus confirming that our method to calculate free spaces is valid and valuable. The selection of proper data analysis methods is strongly dependent on the analysis task and usually done by visual analytics experts. Whereas, the choice of the right parameters is a joint effort, as judging results can be usually only achieved with domain knowledge. Up to now, there exists no extensive visual analytics framework for all possible analysis tasks as each application domain needs its very own approach.

6. Discussion and Conclusions

The analysis of team sport data is a potentially very useful, yet inherently difficult problem. In this paper, we gave a high-level overview to this fields as well as to identify future research aspects to contribute on. Team sport analytics involves many challenges and problems. First, the choice of data acquisition is difficult, as the type and quality of data available determines the potential analysis that can be done. Current acquisition includes video analysis, using position sensors or manual encoding. There is much additional data one can think of to include in a meaningful analysis, e.g., biomedical and physiologic measurements of players could be taken into account. In the wider sense, data acquisition also needs to consider data modeling, integrating, and cleaning, each representing significant work steps.

Being aware of data quality and comparisons of data sources is of critical importance. With regards to this work quality observations are especially needed when it comes to the trajectories of players and balls from video image sources. While we recognize the need for detailed benchmarks, such analysis has not yet been done in related literature and was beyond the scope of this article. In the future we will strive to fill this gap.

For meaningful analysis, domain requirements coming from external experts are important and must be taken into account. Specifically, analysts nowadays are, for example, interested in the analysis, prediction and performance monitoring of their team. Each problem, however, has many sub problems. The problem of data acquisition, for example, translates into video processing, classification and annotation problems. Nevertheless, team sport analysis is a highly interesting research field where experts from many different computer science subjects (computer vision, graph theory, network analysis, visual analytics, ...) can bring in their expert knowledge.

An important problem of doing analysis relates to modeling—what are the aspects in the sports events/matches which are of importance? How can we qualitatively and quantitatively assess individual player performance or the development of tactic and strategic capabilities of a team? Answers to these questions need to consider both data-driven approaches from Data Science, but also,

models and concepts developed in Sport Science. Eventually, we see that both can go hand in hand to design appropriate analysis systems. Future work needs to better characterize the role and influence that both fundamental approaches have and how to combine them. In surveying existing analysis techniques, it may be instructive to distinguish these according to the level of detail on which they operate. Single events and situations (e.g., success rate of corner kicks) can be easily statistically computed. However, it is a difficult problem to assess strategic factors than can *explain* why a specific success rate is observed, or how it could be influenced. In particular, the development of explanatory models for different temporal scales from short- to long term, and considering involved dimensions, is seen an important problem for future work.

We also discussed visualization technology to help analysts interactively explore relevant sports data and help with interpreting patterns. We note that many analysis goals cannot be statically defined once and for all, but depend highly on the context of the analysis. For example, to prepare a team for a match against an upcoming other team, one coach needs to analyze or predict the strength of ones team in context of the other team. To this end, the analyst needs ways to define the context in which to do the analysis, for example, on the offensive of defensive sides. Interactive visualization is a key technology to provide adaptive analysis systems.

In Section 2.1.2 we provided a soccer event taxonomy based on previous data-driven soccer analysis systems we implemented so far. While it may serve as a starting point to guide the definition of event detectors and descriptors, the taxonomy is a starting point and may be extended. Specifically, we may align it with existing taxonomies from Sports Science. Also, while these events are obviously interesting for analysis, not all of them can be detected or quantified with the same prevision, and the latter may also depend on the data acquisition modality at hand. Also, the given taxonomy table considers situations within a game only. For strategic analysis, more high-level or long-term events may be recorded as well, e.g., like when a team moves up a league, acquires new players by transfer, or having the coach change a training system. The latter may be particularly interesting for strategic analysis and correlation of strategic decisions with short-term performance measures.

Also, it would be interesting to assess how representative it is for other sports like basketball or ice hockey. Open questions also pertain to the predictability of behavior in sports. As a first step to this end, we need to define a notion of behavior, and ways to formalize it, and then by experiments one may research if behavior is predictable and to which extent. We note that different kinds of individual and collective movement behaviors can be defined. Collaboration with professional sport analysts in requirement definition and eventually, case studying will be helpful to this end. A German journal recently interviewed [111] the German Football Association chief analyst, Christofer Clemens. In this interview, Mr. Clemens stressed the importance of data-driven sports analytics. Particularly, he stated that answers need to be found to questions such as what constellations of players are more successful in scoring goals, how important it is to quickly pass between the opponents defenders, and where it is important to outnumber players of the opposite team. Such requirements by domain experts can well serve to guide future research efforts in sports analytics.

Author Contributions: Manuel Stein is the leading author of this work. He performed the research and designed the structure of this article. Furthermore, Manuel Stein wrote the paper together with Halldór Janetzko, Daniel Seebacher and Alexander Jäger, Sven Kosub provided the section about Data Modeling. Manuel Nagel contributed in the discussion of transferability to the fields of biology. Together with all authors of this work, Jürgen Hölsch, Tobias Schreck, Daniel A. Keim and Michael Grossniklaus further revised the paper and gave substantial contributions to the design and analysis of this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ryan, M. The Impossible Job: Sky TV Have 24 Cameras but Referees Can Only See So Much. 2010. Available online: http://www.dailymail.co.uk/sport/football/article-1301170/The-impossible-job-Sky-TV-24-cameras-referees-much.html (accessed on 24 June 2016).

2. Glaser, A. The Cameras That'll Make the Super Bowl Way More Interesting This Year. 2016. Available online: http://www.wired.com/2016/01/the-cameras-thatll-make-the-super-bowl-way-more-interesting-this-year/ (accessed on 24 June 2016).
3. STATS. Available online: http://www.stats.com/ (accessed on 8 August 2016).
4. Opta. Available online: http://www.optasports.com/ (accessed on 8 August 2016).
5. Seo, Y.; Choi, S.; Kim, H.; Hong, K.S. Where are the ball and players? Soccer game analysis with color-based tracking and image mosaick. In Proceedings of the International Conference on Image Analysis and Processing, Florence, Italy, 17–19 September 1997; Springer: Berlin/Heidelberg, Germany, 1997; pp. 196–203.
6. Liu, J.; Tong, X.; Li, W.; Wang, T.; Zhang, Y.; Wang, H. Automatic player detection, labeling and tracking in broadcast soccer video. *Pattern Recognit. Lett.* **2009**, *30*, 103–113.
7. Pérez, P.; Hue, C.; Vermaak, J.; Gangnet, M. Color-based probabilistic tracking. In Proceedings of the European Conference on Computer Vision, Copenhagen, Denmark, 28–31 May 2002; Springer: Berlin/Heidelberg, Germany, 2002; pp. 661–675.
8. Pelissero, T. Player-Tracking System Will Let NFL Fans Go Deeper Than Ever. 2014. Available online: http://www.usatoday.com/story/sports/nfl/2014/07/30/metrics-sensor-shoulder-pads-zebra-speed-tracking/13382443/ (accessed on 24 June 2016).
9. Zebra Technologies. Available online: https://www.zebra.com/us/en/nfl.html (accessed on 8 August 2016).
10. ACM DEBS 2013 Grand Challenge. Available online: http://www.orgs.ttu.edu/debs2013/index.php?goto=cfchallengedetails (accessed on 8 August 2016).
11. Pishchulin, L.; Insafutdinov, E.; Tang, S.; Andres, B.; Andriluka, M.; Gehler, P.V.; Schiele, B. DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation. *CoRR arXiv* **2015**, arXiv:1511.06645.
12. Insafutdinov, E.; Pishchulin, L.; Andres, B.; Andriluka, M.; Schiele, B. DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model. In Proceedings of the 14th European Conference on Computer Vision (ECCV 2016), Amsterdam, The Netherlands, 11–14 October 2016; pp. 34–50.
13. Tovinkere, V.; Qian, R.J. Detecting Semantic Events in Soccer Games: Towards A Complete Solution. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Tokyo, Japan, 25 August 2001.
14. Ekin, A.; Tekalp, A.M.; Mehrotra, R. Automatic soccer video analysis and summarization. *IEEE Trans. Image Process.* **2003**, *12*, 796–807.
15. Xie, L.; Chang, S.F.; Divakaran, A.; Sun, H. Structure analysis of soccer video with hidden Markov models. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, FL, USA, 13 May 2002; Volume 4.
16. Assfalg, J.; Bertini, M.; Colombo, C.; Del Bimbo, A.; Nunziati, W. Semantic annotation of soccer videos: Automatic highlights identification. *Comput. Vis. Image Underst.* **2003**, *92*, 285–305.
17. Xu, C.; Zhang, Y.F.; Zhu, G.; Rui, Y.; Lu, H.; Huang, Q. Using webcast text for semantic event detection in broadcast sports video. *IEEE Trans. Multimed.* **2008**, *10*, 1342–1355.
18. Newell, C.D.; Wood, M.D.; Costello, K.M.; Poetker, R.B. Automatic Story Creation Using Semantic Classifiers for Images and Associated Meta Data. U.S. Patent 200,803,069,95 A1, 11 December 2008.
19. Radelet, M.A.; Lephart, S.M.; Rubinstein, E.N.; Myers, J.B. Survey of the injury rate for children in community sports. *Pediatrics* **2002**, *110*, e28.
20. Kujala, U.M.; Taimela, S.; Antti-Poika, I.; Orava, S.; Tuominen, R.; Myllynen, P. Acute injuries in soccer, ice hockey, volleyball, basketball, judo, and karate: Analysis of national registry data. *BMJ* **1995**, *311*, 1465–1468.
21. Janetzko, H.; Sacha, D.; Stein, M.; Schreck, T.; Keim, D.A.; Deussen, O. Feature-driven visual analytics of soccer data. In Proceedings of the 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), Paris, France, 25 October 2014; pp. 13–22.
22. Stein, M.; Häußler, J.; Jäckle, D.; Janetzko, H.; Schreck, T.; Keim, D.A. Visual Soccer Analytics: Understanding the Characteristics of Collective Team Movement Based on Feature-Driven Analysis and Abstraction. *ISPRS Int. J. Geo-Inform.* **2015**, *4*, 2159–2184.
23. Stein, M.; Janetzko, H.; Breitkreutz, T.; Seebacher, D.; Schreck, T.; Grossniklaus, M.; Couzin, I.D.; Keim, D.A. Director's Cut: Analysis and Annotation of Soccer Matches. *IEEE Comput. Graph. Appl.* **2016**, *36*, 50–60.
24. Football.db—Free Open Public Domain Football Data. Available online: http://openfootball.github.io/ (accessed on 8 August 2016).
25. SoccerStats.us. Available online: http://soccerstats.us/ (accessed on 8 August 2016).

26. Football-data.org—RESTful Football Data. Available online: http://api.football-data.org/index (accessed on 8 August 2016).
27. FootballSquads. Available online: http://www.footballsquads.co.uk/ (accessed on 8 August 2016).
28. Football Data Dump from football-data.co.uk. Available online: https://github.com/jokecamp/FootballData/tree/master/football-data.co.uk (accessed on 8 August 2016).
29. Bergmann, T.; Bunk, S.; Eschrig, J.; Hentschel, C.; Knuth, M.; Sack, H.; Schüler, R. *Linked Soccer Data*; I-SEMANTICS (Posters & Demos); Citeseer: Gaithersburg, MD, USA, 2013; pp. 25–29.
30. StadiumDB—Stadium Database. Available online: http://stadiumdb.com (accessed on 8 August 2016).
31. NNDC—Climate Data Online. Available online: http://www7.ncdc.noaa.gov/CDO/cdo (accessed on 8 August 2016).
32. Ekin, A.; Tekalp, A.M. Robust dominant color region detection and color-based applications for sports video. In Proceedings of the 2003 International Conference on Image Processing (ICIP 2003), Barcelona, Spain, 14 September 2003; Volume 1.
33. Twitter Developers. Available online: https://dev.twitter.com/ (accessed on 9 August 2016).
34. Reddit API Documentation. Available online: https://www.reddit.com/dev/api/ (accessed on 9 August 2016).
35. Wikipedia. Available online: https://www.wikipedia.org/ (accessed on 7 November 2016).
36. Yucesoy, B.; Barabási, A. Untangling performance from success. *EPJ Data Sci.* **2016**, *5*, 17.
37. European Media Monitor. Available online: https://emm.newsbrief.eu (accessed on 9 August 2016).
38. BBC Sport—American Football—NFL in a Nutshell. Available online: http://news.bbc.co.uk/sport2/hi/other_sports/american_football/3192002.stm (accessed on 21 July 2016).
39. Paolo Cintia, M.C.; Pappalardo, L. The Haka Network: Evaluating Rugby Team Performance with Dynamic Graph Analysis. In Proceedings of the DyNo, 2nd International Workshop on Dynamics in Networks, San Francisco, CA, USA, 18 August 2016.
40. Gudmundsson, J.; Horton, M. Spatio-Temporal Analysis of Team Sports—A Survey. *CoRR arXiv* **2016**, arXiv:1602.06994.
41. Movebank. Available online: http://movebank.org (accessed on 8 August 2016).
42. Andrienko, G.; Andrienko, N.; Bak, P.; Keim, D.; Wrobel, S. *Visual Analytics of Movement*; Springer: Berlin/Heidelberg, Germany, 2013.
43. Andrienko, N.; Andrienko, G.; Barrett, L.; Dostie, M.; Henzi, P. Space transformation for understanding group movement. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 2169–2178.
44. Fu, T.-C. A review on time series data mining. *Eng. Appl. Artif. Intell.* **2011**, *24*, 164–181.
45. Aigner, W.; Miksch, S.; Schumann, H.; Tominski, C. *Visualization of Time-Oriented Data*; Springer: Berlin/Heidelberg, Germany, 2011.
46. De Berg, M.; Van Kreveld, M.; Overmars, M.; Schwarzkopf, O.C. Computational geometry. In *Computational Geometry*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 1–17.
47. Kang, C.H.; Hwang, J.R.; Li, K.J. Trajectory analysis for soccer players. In Proceedings of the 2006 Sixth IEEE International Conference on Data Mining Workshops (ICDM Workshops), Hong Kong, China, 18 December 2006; pp. 377–381.
48. Cintia, P.; Giannotti, F.; Pappalardo, L.; Pedreschi, D.; Malvaldi, M. The harsh rule of the goals: Data-driven performance indicators for football teams. In Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris, France, 19 October 2015; pp. 1–10.
49. Cintia, P.; Rinzivillo, S.; Pappalardo, L. A network-based approach to evaluate the performance of football teams. In Proceedings of the Machine Learning and Data Mining for Sports Analytics Workshop, Porto, Portugal, 11 September 2015.
50. Pena, J.L.; Touchette, H. A network theory analysis of football strategies. *arXiv* **2012**, arXiv:1206.6904.
51. Bourbousson, J.; Poizat, G.; Saury, J.; Seve, C. Team coordination in basketball: Description of the cognitive connections among teammates. *J. Appl. Sport Psychol.* **2010**, *22*, 150–166.
52. Clemente, F.M.; Couceiro, M.S.; Martins, F.M.L.; Mendes, R.S. Using network metrics in soccer: A macro-analysis. *J. Hum. Kinet.* **2015**, *45*, 123–134.
53. Russom, P. *Big Data Analytics–TDWI Best Practices Report, 4th Quarter*; The Data Warehousing Institute: Renton, WA, USA, 2011; pp. 1–35.
54. Buhl, H.U.; Röglinger, M.; Moser, F.; Heidemann, J. Big data. *Bus. Inform. Syst. Eng.* **2013**, *5*, 65–69.

55. Polk, T.; Yang, J.; Hu, Y.; Zhao, Y. Tennivis: Visualization for tennis match analysis. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 2339–2348.
56. Perin, C.; Vuillemot, R.; Fekete, J.D. SoccerStories: A kick-off for visual soccer analysis. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 2506–2515.
57. Janetzko, H.; Stein, M.; Sacha, D.; Schreck, T. Enhancing Parallel Coordinates: Statistical Visualizations for Analyzing Soccer Data. In Proceedings of the IS&T Electronic Imaging Conference on Visualization and Data Analysis, San Francisco, CA, USA, 14 February 2016.
58. Stein, M.; Janetzko, H.; Lamprecht, A.; Seebacher, D.; Schreck, T.; Keim, D.A.; Grossniklaus, M. From Game Events to Team Tactics: Visual Analysis of Dangerous Situations in Multi-Match Data. In Proceedings of the International Conference on Technology and Innovation in Sports, Health and Wellbeing, Special Track "High level Sports in the XXI Century: Contribution From Industry and University to the Performance Optimization", Vila Real, Portugal, 1 December 2016.
59. Gomez-Marin, A.; Stephens, G.J.; Louis, M. Active sampling and decision making in Drosophila chemotaxis. *Nat. Commun.* **2011**, *2*, 441.
60. Giannotti, F.; Nanni, M.; Pinelli, F.; Pedreschi, D. Trajectory pattern mining. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 12 August 2007; pp. 330–339.
61. Kays, R.; Crofoot, M.C.; Jetz, W.; Wikelski, M. Terrestrial animal tracking as an eye on life and planet. *Science* **2015**, *348*, doi:10.1126/science.aaa2478.
62. Bergner, R.M. What is behavior? And so what? *New Ideas Psychol.* **2011**, *29*, 147–155.
63. Hogan, J.A. A framework for the study of behavior. *Behav. Process.* **2015**, *117*, 105–113.
64. Sumpter, D.J. *Collective Animal Behavior*; Princeton University Press: Princeton, NJ, USA, 2010.
65. Turner, R.H.; Killian, L.M. *Collective Behavior*; Prentice Hall: Upper Saddle River, NJ, USA, 1957.
66. Araújo, D.; Davids, K. Team synergies in sport: Theory and measures. *Front. Psychol.* **2016**, *7*, doi:10.3389/fpsyg.2016.01449.
67. Duarte, R.; Araújo, D.; Correia, V.; Davids, K. Sports teams as superorganisms. *Sports Med.* **2012**, *42*, 633–642.
68. Demšar, U.; Buchin, K.; Cagnacci, F.; Safi, K.; Speckmann, B.; van de Weghe, N.; Weiskopf, D.; Weibel, R. Analysis and visualisation of movement: An interdisciplinary review. *Mov. Ecol.* **2015**, *3*, 1.
69. IMPECT. Available online: http://www.impect.com (accessed on 7 November 2016).
70. Regenhuber, M. IMPECT & Packing: The Future of Football Analytics Is Here. Available online: http://bundesligafanatic.com/impect-packing-the-future-of-football-analytics-is-here/ (accessed on 7 November 2016).
71. Strandburg-Peshkin, A.; Twomey, C.R.; Bode, N.W.; Kao, A.B.; Katz, Y.; Ioannou, C.C.; Rosenthal, S.B.; Torney, C.J.; Wu, H.S.; Levin, S.A.; et al. Visual sensory networks and effective information transfer in animal groups. *Curr. Biol.* **2013**, *23*, R709–R711.
72. Niwa, H.S. Space-irrelevant scaling law for fish school sizes. *J. Theor. Biol.* **2004**, *228*, 347–357.
73. Vicsek, T.; Czirók, A.; Ben-Jacob, E.; Cohen, I.; Shochet, O. Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.* **1995**, *75*, 1226.
74. Biro, D.; Sumpter, D.J.; Meade, J.; Guilford, T. From compromise to leadership in pigeon homing. *Curr. Biol.* **2006**, *16*, 2123–2128.
75. Helbing, D.; Keltsch, J.; Molnar, P. Modelling the evolution of human trail systems. *Nature* **1997**, *388*, 47–50.
76. Helbing, D.; Schweitzer, F.; Keltsch, J.; Molnár, P. Active walker model for the formation of human and animal trail systems. *Phys. Rev. E* **1997**, *56*, 2527.
77. Schelling, T.C. Models of segregation. *Am. Econ. Rev.* **1969**, *59*, 488–493.
78. Keim, D.A.; Kohlhammer, J.; Ellis, G.; Mansmann, F. *Mastering the Information Age-Solving Problems with Visual Analytics*; Eurographics: Goslar, Germany, 2010.
79. Leser, R.; Moser, B.; Hoch, T.; Stoegerer, J.; Kellermayr, G.; Reinsch, S.; Baca, A. Expert-oriented modelling of a 1vs1-situation in football. *Int. J. Perform. Anal. Sport* **2015**, *15*, 949–966.
80. Schmidhofer, S.; Leser, R.; Ebert, M. A comparison between the structure in elite tennis and kids tennis on scaled courts (Tennis 10s). *Int. J. Perform. Anal. Sport* **2014**, *14*, 829–840.
81. MacKenzie, R.; Cushion, C. Performance analysis in football: A critical review and implications for future research. *J. Sports Sci.* **2013**, *31*, 639–676.
82. Carling, C.; Wright, C.; Nelson, L.J.; Bradley, P.S. Comment on 'Performance analysis in football: A critical review and implications for future research'. *J. Sports Sci.* **2014**, *32*, 2–7.

83. Bourbosson, J.; Seve, C.; McGarry, T. Space-time coordination dynamics in basketball: Part 1. Intra- and inter-couplings among player dyads. *J. Sports Sci.* **2010**, *28*, 339–347.
84. Bourbosson, J.; Seve, C.; McGarry, T. Space-time coordination dynamics in basketball: Part 2. The interaction between the two teams. *J. Sports Sci.* **2010**, *28*, 349–358.
85. Frencken, W.; de Poel, H.; Visscher, C.; Lemmink, K. Variablitiy of inter-team distances associated with match events in elite-standard soccer. *J. Sports Sci.* **2012**, *30*, 1207–1213.
86. Link, D. Using of Invasion Profiles as a Performance Indicator in Soccer. In Proceedings of the International Association of Computer Science in Sports Conference, Darwin, Australia, 22 June 2014.
87. Grund, T.U. Network structure and team performance: The case of English Premier League soccer teams. *Soc. Netw.* **2012**, *34*, 682–690.
88. Taki, T.; Hasegawa, J. Visualization of Dominant Region in Team Games and Its Application to Teamwork Analysis. In Proceedings of the International Conference on Computer Graphics (CGI'00), Hong Kong, 3 October 2000; p. 227.
89. Fujimura, A.; Sugihara, K. Geometric analysis and quantitative evaluation of sport teamwork. *Syst. Comput. Jpn.* **2005**, *36*, 49–58.
90. Gudmundsson, J.; Wolle, T. Football analysis using spatio-temporal tools. *Comput. Environ. Urban Syst.* **2014**, *47*, 16–27.
91. Xia, L.; Wang, Q.; Wu, L. Vision-based behavior prediction of ball carrier in basketball matches. *J. Cent. South Univ.* **2012**, *19*, 2142–2151.
92. Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. From data mining to knowledge discovery in databases. *AI Mag.* **1996**, *17*, 37.
93. Chakrabarti, S.; Ester, M.; Fayyad, U.; Gehrke, J.; Han, J.; Morishita, S.; Piatetsky-Shapiro, G.; Wang, W. *Data Mining Curriculum: A Proposal, Version 1.0*; Intensive Working Group of ACM SIGKDD Curriculum Committee: New York, NY, USA, 2006; p. 140.
94. Estivill-Castro, V. Why so many clustering algorithms: A position paper. *ACM SIGKDD Explor. Newsl.* **2002**, *4*, 65–75.
95. Lee, J.G.; Han, J.; Whang, K.Y. Trajectory clustering: A partition-and-group framework. In Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, Beijing, China, 11 June 2007; pp. 593–604.
96. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, 21 June 1967; Volume 1, pp. 281–297.
97. Lindley, D. Regression and correlation analysis. In *Time Series and Statistics*; Springer: Berlin/Heidelberg, Germany, 1990; pp. 237–243.
98. Dick, M.; Wellnitz, O.; Wolf, L. Analysis of factors affecting players' performance and perception in multiplayer games. In Proceedings of 4th ACM SIGCOMM Workshop on Network and System Support for Games, Hawthorne, NY, USA, 10 October 2005; pp. 1–7.
99. Carlin, B.P. Improved NCAA basketball tournament modeling via point spread and team strength information. *Am. Stat.* **1996**, *50*, 39–43.
100. Chandola, V.; Kumar, V. Summarization—Compressing data into an informative representation. *Knowl. Inform. Syst.* **2007**, *12*, 355–378.
101. Person, K. On Lines and Planes of Closest Fit to System of Points in Space. *Philios. Mag.* **1901**, *2*, 559–572.
102. Maaten, L.v.d.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
103. Grubbs, F.E. Sample criteria for testing outlying observations. *Ann. Math. Stat.* **1950**, *21*, 27–58.
104. Klemettinen, M.; Mannila, H.; Ronkainen, P.; Toivonen, H.; Verkamo, A.I. Finding interesting rules from large sets of discovered association rules. In Proceedings of the Third International Conference on Information and Knowledge Management, Gaithersburg, MD, USA, 29 November 1994; pp. 401–407.
105. Inselberg, A. The plane with parallel coordinates. *Vis. Comput.* **1985**, *1*, 69–91.
106. Reijner, H. The Development of the Horizon Graph. In Electronic Proceedings of the VisWeekWorkshop From Theory to Practice: Design, Vision and Visualization, 2008. Available online: http://www.stonesc.com/Vis08_Workshop/DVD/Reijner_submission.pdf (accessed on 29 December 2016).
107. Keim, D.A. Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Trans. Vis. Comput. Graph.* **2000**, *6*, 59–78.

108. Keim, D.A.; Ankerst, M.; Kriegel, H.P. Recursive pattern: A technique for visualizing very large amounts of data. In Proceedings of the 6th Conference on Visualization, Washington, DC, USA, 29 October 1995; p. 279.
109. Simon, S.; Mittelstädt, S.; Keim, D.A.; Sedlmair, M. Bridging the Gap of Domain and Visualization Experts with a Liaison. In *Eurographics Conference on Visualization (EuroVis)—Short Papers*; Bertini, E., Kennedy, J., Puppo, E., Eds.; The Eurographics Association: Cagliari, Italy, 2015.
110. Grehaigne, J.F.; Bouthier, D.; David, B. Dynamic-system analysis of opponent relationships in collective actions in soccer. *J. Sports Sci.* **1997**, *15*, 137–149.
111. Biermann, C. Wir Wollen Eine Revolution. 2015. Available online: http://www.11freunde.de/interview/ist-datensammeln-im-fussball-sinnlos (accessed on 7 September 2016).