Investigating the Sketchplan: A Novel Way of Identifying Tactical Behavior in Massive Soccer Datasets

Daniel Seebacher, Tom Polk, Halldor Janetzko, Daniel A. Keim, Tobias Schreck, and Manuel Stein

Abstract—Coaches and analysts prepare for upcoming matches by identifying common patterns in the positioning and movement of the competing teams in specific situations. Existing approaches in this domain typically rely on manual video analysis and formation discussion using whiteboards; or expert systems that rely on state-of-the-art video and trajectory visualization techniques and advanced user interaction. We bridge the gap between these approaches by contributing a light-weight, simplified interaction and visualization system, which we conceptualized in an iterative design study with the coaching team of a European first league soccer team. Our approach is walk-up usable by all domain stakeholders, and at the same time, can leverage advanced data retrieval and analysis techniques: a virtual magnetic tactic-board. Users place and move digital magnets on a virtual tactic-board, and these interactions get translated to spatio-temporal queries, used to retrieve relevant situations from massive team movement data. Despite such seemingly imprecise query input, our approach is highly usable, supports quick user exploration, and retrieval of relevant results via query relaxation. Appropriate simplified result visualization supports in-depth analyses to explore team behavior, such as formation detection, movement analysis, and what-if analysis. We evaluated our approach with several experts from European first league soccer clubs. The results show that our approach makes the complex analytical processes needed for the identification of tactical behavior directly accessible to domain experts for the first time, demonstrating our support of coaches in preparation for future encounters.

Index Terms—Sport analytics, soccer analytics, visual analytics

✦

1 INTRODUCTION

IN data-driven soccer match analysis and coaching, analysts need to examine massive amounts of movement data collected with the help of GPS sensors or optical tracking. Their main objective is to prepare their team and develop tactics targeted to the game-playing style and tactical behavior of opposing teams. In this process, they face two major challenges. First, they have to search through large amounts of collected data to find relevant and repeating key situations of their and the opposing team, such as build-up play. This includes identifying reoccurring patterns in both the cooperative and competitive movement behavior in the detected situations, which includes positions of the players or the direction in which the ball is most likely to be kicked. Identifying these patterns helps coaches and analysts understand the tactics of the opposing teams in key situations. The second challenge lies in how this information and these insights can be effectively and efficiently communicated to coaching assistants and players, enabling optimal preparation in training sessions before an upcoming match.

Traditional approaches in analysis of soccer match data typically rely on tedious manual video analysis [1], [2] and formation discussions using whiteboards. To some extent, interactive7 visual expert systems are available to teams, based on collected movement and event data. However, these systems often only use data from single matches or, in some cases, even only partial matches. Consequently, tedious manual work is required to extrapolate the results to a holistic analysis of teams. Furthermore, for any single match, many factors could account for its specific results, such as the daily condition of individual players, the tactics of the opposing team, or even weather conditions.

To understand a teams' tactics and playing style objectively, it is crucial to not only consider the behavior of a team in a single match, but also, to analyze it in an aggregated form over several matches. An additional drawback of existing expert systems is that they often rely on complex and advanced user interfaces and interactions. The extensive training required to use these complex systems puts them out of the reach of key users: the coaches.

We solve both challenges by bridging the gap between tools and procedures familiar to professionals in the sports domain and sophisticated soccer analysis techniques used in research, as shown in Fig. 1. We present a light-weight,

- Daniel Seebacher, Tom Polk, and Daniel A. Keim are with the University of Konstanz, 78464 Konstanz, Germany. E-mail: {Daniel.Seebacher, thomas.polk, keim}@uni-konstanz.de.
- Halldor Janetzko is with the Lucerne University of Applied Sciences and Arts, 6002 Luzern, Switzerland. E-mail: halldor.janetzko@hslu.ch.
- Tobias Schreck is with the Graz University of Technology, 8010 Graz, Austria. E-mail: tobias.schreck@cgv.tugraz.at.
- Manuel Stein is with Subsequent GmbH, 78464 Konstanz, Germany. E-mail: manuel.stein@subsequent.ai.
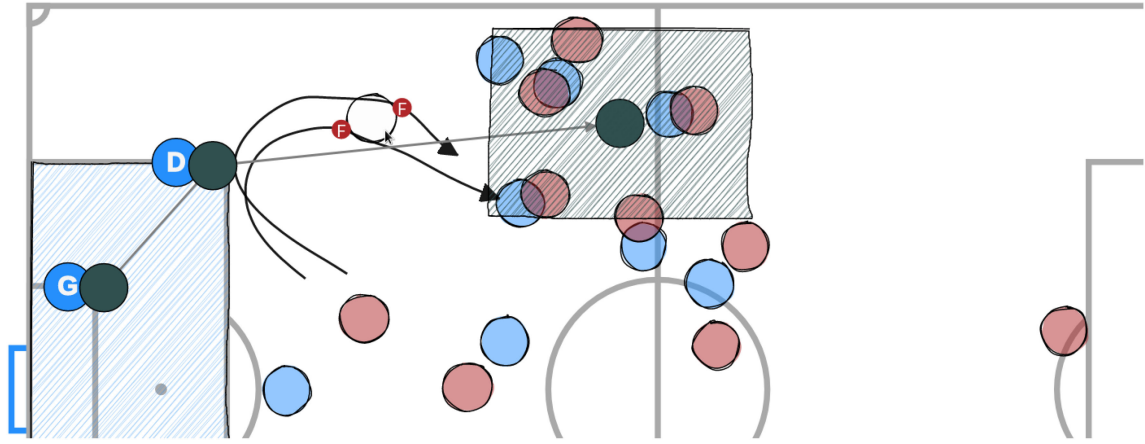
Fig. 1. We introduce a novel visual analysis design for searching and comparing team situations in soccer multi-match data. Our design is based on the magnetic tactic board, widely used in practice for coaching and analysis in team sports including soccer. Users efficiently and effectively specify team situations by dragging & dropping magnets representing players, such as **G**oalkeeper or **D**efender, or the ball to pitch positions and simplified sketch interactions. Appropriate visual data abstractions enable users to understand, compare, and explain situations of interest for coaching and presentation tasks. The above example shows two slow build-up play situations retrieved with our approach. Aggressive playing behavior of the **F**orward to immediately pressure the defender is clearly visible in both.

simplified and effective interaction and visualization system, which we conceptualized in an iterative design study with the coaching team of a European first division soccer club. Our simplified, straightforward design is immediately (or walkup-) usable by all domain stakeholders while, at the same time, can leverage advanced data retrieval and analysis techniques that support multi-game analysis. The metaphor we employ is a *magnetic tactic-board* that is highly responsive to interactions based on magnet movements, such as the one shown in the real-world example in Fig. 2.

Our contributions include a novel approach for the efficient and effective search and analysis of situations in massive soccer movement data. We identified the need for our solution in an iterative design study with the coaching team of a European first division soccer club and reviewing relevant related work in Section 2. Our unique insight into the work and problems of analysts, coaches, and decision-makers in the soccer domain enabled us to identify four high-level requirements, from which we derived a detailed set of ⌕ search, Σ analysis, and ⌁ visualization requirements as described in detail in Section 3. Based on these requirements, we propose a visual analytics system tightly following the metaphor of a *magnetic tactic-board*, which is already very familiar to our targeted end-users of soccer coaches, analysts, and players. In our system, we make technical contributions in Section 4 to effectively retrieve relevant situations in multi-match data using domain-inspired sketch search operations. We also solve challenges such as high input inaccuracy, and large data set processing. In addition, our visual designs allows to effectively compare, understand, and explain to others information from a potentially large number of found match situations. Our visual contributions solve the identified search and analysis requirements resulting in an intuitive system presented in Section 5. We evaluate our approach using data from 250 soccer matches with domain experts from European first league soccer clubs in Section 6, demonstrating its usability and efficacy to provide useful insights. In Section 7, we discuss our results and suggest directions for future work.

2 RELATED WORK

Our work relates to a larger body of previous work in visual analysis of sports data, which in turn often relies on visual analysis of different data types obtained from sport activities. Also, visual sports data analysis often relies on approaches for searching and comparing data, including data clustering, classification, and query-based data exploration including sketch-based interaction.

2.1 Visual Analysis of Sports Data

Analysis of sports data is a challenging and relevant problem, with many application opportunities e.g., for coaching, scouting, and performance assessment tasks [3]. Due to recent advances in e.g., sensor technology [4], [5] or computer vision [6], [7], [8], [9], it is possible to collect many different data types from sports activities, e.g., movement data [10], body postures [11], [12], [13], [14], and detecting events [15], [16], [17]. Visual analysis of sport data has recently attracted significant research in visual data analysis. In recent surveys [18], [19], [20], [21], [22], encompassing overviews of techniques are given for different types of sports and have identified three broad categories of data: statistical or in-sport tracking data, and metadata.

Approaches have been introduced to data for major sport types [23], [24], [25], [26], [27]. For example, there are several solutions supporting tennis match data analysis. The *TenniVis* system focuses on point outcomes relative to key match factors, such as who is serving, the serve number, and the game score [28]. The *CourtTime* system allows to visually compare points made by players during tennis matches, based on setting up spatial situations involving the players which allow to find similar situations across many games [29]. The *iTTVis* system provided visual analysis of table tennis matches from multiple perspectives, including temporal, statistical, and tactical perspectives [30]. Analysts are able to search for frequent stroke patterns of players. The system supports data from individual matches, but does not support querying the

1) *Initial situation* **2)** *Moving one player* **3)** *Reaction of the other players*

Fig. 2. Over the last two years, we closely collaborated with the head coach of an European first league soccer club. In weekly meetings, we held interviews to characterize and formalize the domain problems focusing on the day-to-day work. In this figure, snippets of a video conference with the head coach can be seen where he uses a tactic-board to explain player formation changes to the authors.

data based on input player positions or situation. Although existing approaches that do not fully incorporate spatial ball and player movement data can address the *what* and *when*, we need spatial data to truly get deeper insights that help us answer *why*. An approach that illustrates the potential of incorporating spatial data in analysis and visualization is ShuttleSpace [31]. 3D trajectory data is immersively visualized in a 1:1 virtual reality reconstruction of a badminton pitch, enabling the exploration and analysis of data from the players' actual point of view.

2.2 Visual Analysis of Soccer Data

Besides tennis, another very popular sport type is soccer. Due to a wide interest in soccer, numerous papers and whole books presenting a large variety of methods focusing on its analysis have been published [32], [33], [34]. In the following, we restrict ourselves to the most similar techniques to the one proposed in this paper. A few years ago, *SoccerStories* was proposed and represented the first major contribution to the visualization of soccer match data [35]. It allows to interactively explore movement and event data from soccer matches in a rich visualization design, introducing appropriate visual representation of various soccer-specific events like ball passes, dribbles, or shots on goal. Soccer data analysis benefits from integration of analytical components for automatic derivation of features of interest from the underlying movement or event data. Janetzko *et al.* propose the feature-oriented analysis of soccer matches based on generic spatio-temporal features, including player speed, acceleration or player/ball distance measures as a function of time[36]. Many more features specific to the soccer domain have to date been leveraged for visual soccer analysis, including notions such as game pressure [37], interaction and free spaces [38], team formations over time [39], and more [40], [41]. Domain-specific and generic features can be used, among others, to cluster players and game situations [42], segment meaningful match sequences [43], and identify interesting and outlying match situations or player properties [44].

Inherent to team sports is the interaction of the team members, hence interaction analysis is a specifically important topic for soccer and sports analysis. Tools that focus on passing analysis are proposed, including calculating passing alternatives and computing frequent pass sequences [45], or

analyzing dynamic changes in a team's passing tactics [46]. Other important analytical tasks in sports and soccer analysis include data reduction and classification/prediction. Sacha *et al.* [43] introduce data reduction approaches based on simplification and aggregation of movement patterns. The reduction may be done interactively, or automatically based on data features of interest. Data aggregations can help in analyzing general underlying patterns. Andrienko *et al.* propose a framework for querying and aggregating data, using a new time-homomorphic reference systems, enabling a subsequent comparative visual exploration coordinated movement of soccer players in different classes of situations as they develop [47]. Regarding classification/prediction, machine learning tools including a Support Vector Machine (SVM) can be applied for interactively defining what-if scenarios [48]. Specifically, a SVM classifier is trained on passing data to predict the likelihood of success from a pass from a user-specified location on the field. Resulting probabilities are displayed as a heatmap. Furthermore, a technique to predict optimal goal shooting positions is proposed.

2.3 Sketch-Based Interaction and Searching

Visual data exploration systems often include facilities for interactive specification of patterns to search and compare. For sports data exploration, learning from historic movement is a central data aspect. Hence, similarity search methods for movements (trajectories) are required. To date, many methods exist for similarity search and comparison of movement data [49], [50]. Specification of movement patterns to search for can be done by numeric specification of movement parameters, e.g., length, speed or direction features, which may however be cumbersome and non-intuitive for non-expert users. An intuitive specification of movement can be done via sketch interfaces, which allow users to draw patterns they are interested in directly on the display. Sketch interfaces are valuable tools in modeling and retrieval of visual information, e.g., in shape construction [51], [52] or image retrieval [53]. Sketch interfaces have also been used in visual data analysis, e.g., in analysis of patterns in traffic movement [54], time series [55], or scatter plot diagrams [56]. Several works in visual soccer analysis include sketch-based searching [57], [58] or combined sketch- and feature-based searching [5]. Sha *et al.* present

another related approach for multi-agent trajectory search by example or sketch in basketball [59]. In their follow-up work, they present a unqiue feature, namely, to directly sketch search queries into broadcast videos of matches to find similar situations [60]. Typically, such sketch interfaces allow searching with a more or less precise specification of trajectories. Our sketch interface is particularly light-weight and allows query specification by as little as dragging & dropping one player on the pitch (which is why we regard it as a walk-up usable interface). Furthermore, users can add subsequent actions and context via sketches, but our query interface requires a minimum of user input.

2.4 Positioning of Our Work

Our work supports visual analysis of multi-match soccer data, and employs methods from the state-of-the-art for trajectory analysis, including feature-based data clustering and similarity search. We extend the current state of the art by providing a novel way to identify tactical behavior, based on the magnetic tactic-board metaphor. Our interface supports sketch- and drag-and-drop-based specification of queries for soccer situations, allowing coaches and non-experts to find match situations of interest. Key difference to previous works is the level of abstraction of our user search interface and result visualization, since users can describe situations as a sequence of states that need to occur instead of specifying the intermediate steps, as would be the case when sketching a trajectory. Additionally, it enables users to identify and retrieve situations in which no movement occurred, e.g., when a team is standing in a particular formation, which would not be possible with the aforementioned trajectory-based retrieval systems. Furthermore, a distinctive feature that our approach offers is the temporal alignment of multiple movement sequences, as highlighted in Section 4.2. Our result visualization shows an aggregate overview of many situations retrieved from a multi-game match database. The latter is implemented via a query relaxation approach, tailored to fit specifically the abstracted visual query interface. Like other works, our design is inspired by domain expert requirements, however, it delineates itself from other works by specifically addresses coach-team instruction situations, requiring abstracted situation search and visualization for effective communication.

3 REQUIREMENT ANALYSIS

In July 2018, we were contacted by the coach of an European premier league soccer club. The coach is also a former professional soccer player as well as certified video analyst. He approached us during the club's search for a novel visual analysis system that would enable the coaching team to process their match data enabling multi-match analysis. The club had collected massive amounts of soccer movement data in high resolution but was lacking adequate analysis possibilities. Their analysis capabilities were limited to basic statistical approaches concentrating on physiological data of each player, for example, how many sprints or the overall distance a player performed in various time intervals during a match. However, these traditional physiological-focused methods do not allow to draw conclusions on the tactical behavior or the strategies of teams on the pitch. This includes

the detection as well as the analysis of player behavior during specific match situations, e.g., how the formation of an opposing team evolves during counter-attacks or set play situations. In addition, coaches and analysts lack appropriate communication media to convey identified strategies to their players. Hence, identifying, observing and communicating the *match plan* of an upcoming opposing team remained a manual and exhausting task.

In this design study, our goal is to develop a novel interactive visual analysis system enabling new perspectives for multi-match analysis in massive soccer datasets. Over the last two years, we collaborated closely with the club: head coach as well as all assistant coaches and video analysts participated in weekly meetings, following the established nine-stage methodology by Sedlmaier *et al.* [61]. During these weekly meetings over the course of two years, we initially held interviews with each coach and analyst to characterize and formalize the domain problems. Especially, we focused on their day-to-day work to strengthen our understanding of what information is currently being used in match analysis, how analyses are being performed, and how insights are communicated with the first team. Afterwards, we used the weekly meetings to discuss analytical requirements of the coaching team, present our iteratively improving prototype and collect feedback for further progress. With the consent of the coaching team the meetings were recorded on video and distributed among the authors for further discussion. The knowledge gained from these meetings lead us to the following four high-level design requirements:

- *Focus on individual opponents.* Coaches are usually focused on their next opponent in order to tailor training preparation and strategies to best defend against their opponent's strengths and exploit their vulnerabilities. This includes past head-to-head match-ups with the opponent as well as matches against this opponent from teams with similar play style as their own.

- *Analyze specific situations.* Coaches are not just looking for overall summary statistics of their matches. However, they want to look more closely at specific types of situations, including set plays like corner kicks and goal kicks. Across all phases of play, coaches view offensive and defensive formations in reaction to specific situations as situational problems to be solved. They are not only looking for their own solutions, but also for ways to anticipate and counteract their opponent's solutions.

- *Support efficient analysis.* Most teams cannot afford a large number of analysts to provide coaches with timely, strategic input. In worst case, coaches must work alone or with little help to quickly perform highly targeted analyses that are most likely to yield actionable results. This includes being able to quickly setup and evaluate specific match situations through direct interaction techniques that are easy to learn and be realized.

- *Easily communicate results.* Coaches do not perform their analyses in isolation or simply for their own good. The analyses they perform are only as valuable

as their ability to clearly communicate their findings to key members of the team. Accordingly, coaches need tools and techniques that help them get their players to understand situational problems and the role of each player in the proposed solutions. This includes both the ability to communicate abstract ideas as well as to demonstrate these concepts through readily accessed video samples from prior matches.

Although these requirements have been derived specifically for analyzing soccer matches, they are generalizable to numerous other invasive sports where opponents can directly impact each other. These requirements serve as both the foundation for our design concepts, as well as the evaluation metric for making trade-offs between alternative designs. All identified requirements are condensed as follows:

S 🔍 Search Requirements are crucial to enable the analyst finding proper results effectively and efficiently.

S1 *Searching in selection of games* allows focusing the query on relevant results. A typical query would be to search the last n games or games against opponents with similar play style.

S2 *Finding specific situations* in matches is important in order to externalize and verify internal hypothesis of coaches. Instead of vague memories of situations, historic recordings help to show the true movements and locations of all players.

S3 *Describing situations appropriately* allows the coach interacting with the search system to accurately specify desired situations though a visual language.

S4 *Iterative querying* allows altering and refining queries in the course of the analysis. This means, for example, that a coach does not have to place all magnets in advance; a few are enough to find specific situations.

S5 *Search by subsequent actions* supports to specify subsequent actions directly happening after the specified situation. For instance, analysts might be interested in only those situations where the ball moved to the left side after a goalie kick-off.

S6 *Query relaxation* is crucial to allow users to be less exact in specifying input location parameters needed by the search algorithms. A threshold around location parameters needs to be supported since coaches are most interested in a combination of absolute and relative positioning of players.

A Σ Analysis Requirements correspond to the questions tackled in strategic analyses.

A1 *How often did a situation occur?* Knowing whether situations are rare exceptions or common patterns is important to prioritize and tailor training for an immediate effect.

A2 *Which players and teams are involved?* It is crucial to identify the players or roles of players and select the proper teams for a successful analysis. The identification of individual players or roles helps to directly derive strategic directives for single players.

A3 *Pattern detection* identifies common patterns in positioning and movement and is tightly connected to

strengths and weaknesses of teams. Situational training helps players to master detected patterns.

A4 *Investigate individual situations in detail.* Although the detected positioning and movement patterns can give an overview of the resulting situations, it is still crucial to investigate individual situations in detail to understand the individual actions of the players involved.

V 〰️ Visualization Requirements aim for a clear communication of selected situations and a historical record of movements.

V1 *Intuitiveness* results in easy and understandable visualizations using same visualization concepts and metaphors of all information recipients. The designed system needs to be as user-friendly and understandable as possible and, in the best case, even allow for usage by the players themselves.

V2 *Resolving of situation* conveys the movement alternatives and development of situations. This requirement allows the static player locations to evolve into a semi-static contextualized visualization communicating temporal aspects.

V3 *Details-on-demand* serves two purposes: we do not overload the display with unnecessary details, but rather allow interested analysts to dig deeper into the dataset whenever needed. Additional details are, for instance, the movement path of players, roles of players, and the exact locations of the players.

V4 *Communicating findings* allows all involved parties to deepen their understanding of the strategic consequences their own movements have. Coaches and analysts can easily share and discuss their findings in online and offline meetings.

4 ENHANCING THE TRADITIONAL TACTIC-BOARD

As outlined in our design requirements, crucial tasks during match preparation can be divided into two major challenges: Finding specific match situations in a set of recorded matches and performing a thorough analysis of these situations to identify tactical patterns in the collective movement behavior. Our proposed system solves both of these challenges following a real-world concept based on a simple tool widely employed by coaches and players: the *magnetic tactic-board*. This tool is not only used in soccer, but also in a wide variety of invasive and competitive team-sports world-wide. Tactic-boards are currently being deployed in numerous sports environments (see our interview with a former professional international soccer player and now head coach of an Austrian first league team in Section 3) and used to manually analyze and visualize match situations and their development, for example, during counter-attack situations to communicate strategies to their players.

One major drawback of tactic-boards is the extensive manual effort required to setup tactical situations for analysis. Since every analysis on the tactic-board is performed manually, no metadata or additional information, such as which players are usually involved in these situations, can be linked or visualized on the tactic-board. For instance, to illustrate a kick from one player to the other, the coach would position

them at their ideal position and move the ball from one to the other. However, this exact situation may not occur in the real game. Thus, each of these illustrations also reflects the individual biases of the coach and analyst.

By creating a virtual and enhanced representation of a magnetic tactic-board, we can address all of the drawbacks of its traditional counter-part without sacrificing its advantage of being an intuitive and well-known yet powerful visualization of team positioning and movement data. In the following, we first introduce a set of required definitions and notations. Afterward, we discuss how a virtual tactic-board can be utilized as a visual query interface to identify situations of interest in a large set of multi-match data, and contrast our approach against conventional trajectory-based retrieval techniques. Finally, we illustrate how identified situations of interest can be summarized to be used for detecting patterns in the retrieved movement data.

4.1 Formal Definitions and Notations

The basis for our proposed system is spatio-temporal multi-match data G. For each match $g \in G$, the data contains spatio-temporal information describing the position of all moving objects $m \in M$ on the pitch for the whole duration of the game. This includes the players $p \subset M$ of both teams and the ball $b \subset M$. Both of these entities are moving objects as described by the conceptual movement framework of Andrienko et al. [49].

A match situation $\rho \in P$ can now be described as the movement of a set of individual entities during a specific time interval $[t_1, t_2] \subset T$, e.g., as the function $\mu : M \times T \rightarrow S$. These situations can commonly be described as a set of positions that individual players and the ball must occupy at specific times $t_i \in T$. For instance, during a corner-kick, the ball and at least one player must first be at one corner of the pitch and then the ball usually lands in the penalty area a short time later. Following this conceptual movement framework, we describe situations as a set of movements. Movement μ consists of a sequence of spatial events, i.e., location visits (m, t, s) of a set of movers m, where each mover m_i must appear at an assigned location s_i at specific time t_i. For example, a corner situation can occur when the following location visit events for at least one player p_i and the ball b are present: $(p_i, t_1, (52.5, 34))$ and $(b, t_1, (52.5, 34)), (b, t_2, (47.5, 0))$, with $(52.5, 34)$ being the location of the bottom right corner of the soccer pitch and $(41.5, 0)$ the location of the penalty mark.

Additional event data, as described by Stein et al. [20], are also available. These can be used to implement further desired features that enable a more in-depth analysis of the game. For instance, a query to find back-passes from the defender to the goalkeeper can be created by dragging a ball magnet from the defender's position into the penalty area. However, without further information, this query also includes crosses from the opposing team into the penalty area. To better differentiate between such situations, we offer two additional variants of the ball magnet, which explicitly denote possession or a change in possession. On the one hand, the ball possession magnet, requiring that the ball is in possession of the specified team for each location visit. On the other hand, the ball conquest magnet, indicating that the possession of the ball must change from one team to the other at the given location. By combining both data sources,

movement and event data, detailed cooperative as well as competitive tactical movement behavior can be described by using our system. Existing articles provide detailed overviews about tactical movement behavior as well as their developments in soccer [62].

4.2 Magnet-Based Situation Retrieval

In our proposed system, we apply the previously described formal model for the interactive identification of situations of interest (Section 4.1). In order to realize an intuitive query creation interface, analysts and coaches can describe situations by placing magnets representing the players and the ball at key positions on a digital, two-dimensional tactic-board. We can separate our process for identifiying situations into two steps: movement detection and situation retrieval, which we describe in detail in the following sections. For all of our geographic computations, we employ the established and well-known spatial database *PostGIS* [63].

4.2.1 Movement Detection

Each placed magnet is converted into a *location visit* event and can be represented as a spatial range or nearest neighbor query. However, using either query type alone has drawbacks. Spatial range queries have the problem of trying to select the right parameter distance threshold ϵ, especially in our use-case, since the distribution of players is heavily situation-dependent. During open play, the distribution of players is considerably more varied than during a corner-kick, where almost all players are located inside the penalty area. Thus, choosing a suitable ϵ that fits both situations is nearly impossible. Similarly, nearest neighbor queries have the problem of trying to choose an appropriate number of neighbors. This problem becomes apparent when placing a magnet on the pitch and dragging it to a second location resembling movement μ of a mover m from s_1 to s_2. Although we can find the n nearest neighbors of both magnets, the massive amount of movement data makes the chance of finding a pair of players of the same situation or even the same game nearly impossible. However, since the user does not know the exact location of players and cannot be expected to make pixel-perfect inputs, an additional complicating factor is that the manual placement of magnets results in a high degree of input uncertainty when creating spatial queries. To tackle these problems in order to satisfy ⚲ S6, we present a technique we call *query relaxation*, which is a three-step process where we combine spatial range and nearest neighbor queries to find relevant movements despite the aforementioned adversities. An intuitive use case that illustrates this approach is finding movement μ of a player m moving from one location to another, as shown in Fig. 3.

As a first step, we create a candidate set for each placed magnet $m_1, \ldots, m_{n-1}, m_n$, depicted in blue in Fig. 3. For this purpose, we perform a spatial range query around each placed magnet, represented by the white circle with the dashed border, to find players who have been near the placed magnet in the selected games. The distribution of players, can vary extremely around the placed magnets, depending on the position and situation. For example,
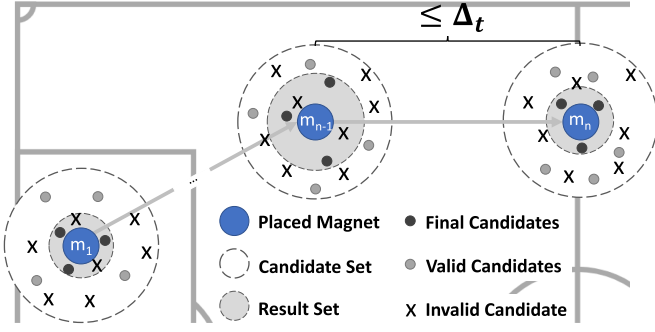
Fig. 3. Our proposed query relaxation approach to detect movement despite high user input uncertainty of a player from the left location to the right location 🔍 S6. A list of candidates around each placed blue magnet is selected. Using task-dependent join conditions, such as a maximum time threshold Δ_t between *location visits*, we identify valid candidates. Finally, we can relax the query depending on the results as well as the selection algorithm. In this instance we want to find at least three movements.

during a corner kick situation the density of players in the penalty area is extremely high, while extremely low in the midfield at the same time. Therefore, we deliberately select a large radius (ten meters in our experiments) for the initial range queries in order to find the necessary players in a wide range of situations. As a result of these range queries, we obtain a set of candidates $C_i, \ldots, C_{n-1}, C_n$ for each of the placed magnets. Each candidate set consists of a set of players $c \in C$ which are located within the defined radius around the placed magnet in the selected games. For each player c we have the following information: id, timestamp, match, and position.

Second, in order to extract movements of individual players, it is necessary to identify players that are present in all candidate sets. Furthermore, it has to be taken into account that these are not only the same players in all candidate sets, but also that they are present during the same games and at approximately the same time. To obtain this set of valid candidates V, highlighted in grey in Fig. 3, the individual candidate sets C are combined in order by a sequence of θ-joins: $V = C_1 \bowtie_\theta \ldots C_{n-1} \bowtie_\theta C_n$. In our case, the predicate θ corresponds to equality of player identity, equality of match identity, and temporal ordering of candidates: $c_i.\text{id} = c_j.\text{id} \wedge c_i.\text{match} = c_j.\text{match} \wedge 0 < c_j.\text{timestamp} - c_i.\text{timestamp} < \Delta_t; c_i \in C_i, c_j \in C_j$. Here, we allow for a maximum time difference Δ_t, which was experimentally determined to be ten seconds. At smaller values, too few movement sequences were detected, as only very rapid and linear movements between two magnets are considered. For larger values, on the other hand, irrelevant movement sequences were detected too frequently.

Finally, for each tuple in the set of valid candidates, an error coefficient is computed, which is the sum of the squared distances of the found players to the respective placed magnets in each candidate set C (see Eq. (1)). This error coefficient can then be used to relax our query on the fly to broaden or narrow down the search, since the full list of valid candidates and error coefficients only needs to be precomputed once. Here we can either employ a top-n query or calculate a suitable cutoff value based on a predefined absolute value or, for example, based on the mean error to obtain the final result set, highlighted in black in Fig. 3.

$$\epsilon_\mu = \sum_{i=1}^{n} d(c_i.\text{position}, m_i)^2 \qquad (1)$$

Eq. (1): Error calculation of movement μ using the sum of squared distances from the magnets played by the user m_i to the found players in the corresponding candidate set $c_i \in C_i$. To penalize large errors more severely, we employ the squared distance to calculate the error coefficient.

Following these three steps allows us to find meaningful results without the need for fine-tuning the ϵ threshold of the spatial range queries for each movement in each situation and avoids the issue of not finding relevant results when using nearest neighbor queries. An major benefit of movement detection with the magnet metaphor is that we enable a more abstract search than is the case with conventional trajectory search systems. In our case, only the definition of the start condition is necessary. We support an optional definition of an end and as well as an arbitrary number of intermediate conditions. In a trajectory search system, all intermediate steps also need to be defined implicitly and are taken into account for the similarity computation, although they might be irrelevant for some situations.

4.2.2   Situation Retrieval

Our query relaxation approach is also applicable to retrieve situations defined by one or more movement sequences, in order to satisfy our search requirements (Section 3). We support an arbitrary number of movement sequences, described by placed magnets, as well other constraints, such as restricted areas, which are important, for example, during kick-offs or penalties. For instance, we place a player near a corner μ_p as well as a ball μ_b, which is subsequently dragged into the penalty area to define a corner-kick situation. We now need to adapt the previously defined join condition for movement detection, to not use the player identity, but the match identity and time instead $S = \mu_b \bowtie_\theta \mu_p$, where the predicate θ corresponds to $\mu_b.\text{match} = \mu_p.\text{match} \wedge \mu_b.\text{timestamp} = \mu_p.\text{timestamp}$. Since each of the movements can consist of a tuple of an arbitrary number of placed magnets, we need to specify which timestamps to use for the join predicate θ. In this case, μ_p consists of one placed magnet $m_{p,1}$ and μ_b consists of two placed magnets $m_{b,1}, m_{b,2}$. Since both movements start at the corner, the timestamps of $m_{p,1}$ and $m_{b,1}$ need to be equal.

We can support an arbitrary number movements in a query, equivalent to the creation of the valid candidate set during the individual movement detection, by a sequence of θ-joins to include the results of additional *movements*.

Finally, a situation occurs when the result set of the join of all the spatial subqueries for all *movements* is non-empty. The result of this spatial query is the set of all found situations identified by match identity and time. For each situation we can also calculate the combined error coefficient, which is the sum of the error coefficient of all combined invidiual movements $\sum_{i=1}^{n} \epsilon_{\mu_i}$

A distinctive feature that our situation-retrieval method offers is the temporal alignment of movement sequences, as presented in schematic form in Fig. 4, showing slow build-up play over a defender into the midfield. Actual examples of this schema are shown in Figs. 1 or 7 (Q_1). Our query consists of three movements, as well as a restriction, sketched
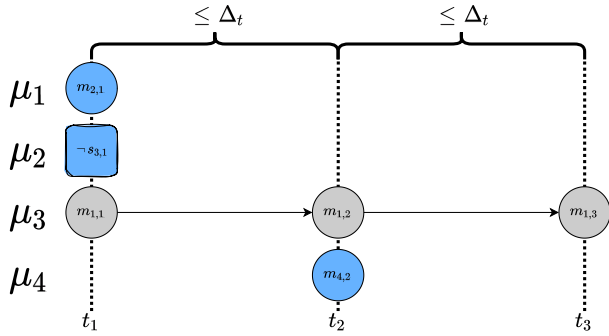
Fig. 4. Situation retrieval based on a set of user-defined movement sequences μ_{1-4}. Using a temporal alignment of the individual location visit events that the movements consist of, we are able to describe the desired situation in more detail. To narrow down our search results to only find specific situations, we limit the time between location visits of a movement to Δ_t.

by the user, that there should not be any players of the opposite team in the penalty area: μ_1, the goalkeeper, μ_2, the sketched restriction, μ_3, the ball, and μ_4, a defender. These movements can then be aligned temporally according to the individual conditions of the found movement sequences. For example, at $t1$, ball, goalkeeper, and the restriction that no red players are allowed in the penalty area must be present. Afterward, the ball and a defender are to be located at the left edge of the penalty area at time $t2$. Finally, the ball needs to be in the left midfield at time $t3$. Since our way of defining movement sequences consists of atomic events, the temporal alignment is trivial since only the temporal condition has to be adjusted when joining the individual movements.

A special case occurs when we retrieve two situations in immediately successive time instants. This case can occur relatively often in the real data, for example, when a player waits half a second before executing a free-kick. Instead of considering such successive situations as independent, we merge them into a situation interval. Here we consider only the situation with the smallest error coefficient as the representative of the merged situations. However, we could also focus on the first, last or any other arbitrary selected time point in the interval of merged situations, depending on the task. This way, we detect a set of time instants per match that represent all the found situations of interest and the involved players. For each situation, we retrieve the tracking data of all moving objects on the pitch as well as additional thematic attributes, such as the role and team-association of the found players, which will be the foundation for the following in-depth analysis.

4.3 Situation Aggregation

The set of found situations returned by our query serves as a basis for further analysis and visualizations. An important feature needed by the domain experts is the detection of patterns $\Sigma A3$ in the positioning and movement of players and the ball on the field. To meet this requirement, we proceed in two stages. First, we perform a clustering of the positions of all players to find the most probable formation within the situation as described by the analyst. Afterwards, in the second stage, for each cluster of players found in the first stage, we perform a clustering of their subsequent movements to identify any common movement patterns among the players within a cluster.

4.3.1 Formation Detection

To detect formations in the retrieved situation (see Section 4.2.2), we use the position data of the players and the ball. Since each soccer team consists of 11 players, except when a player is sent off after a red card, formation detection requires finding 11 clusters for each team and 1 cluster for the ball, 23 clusters in total.

Standard clustering approaches, which are often used for spatio-temporal data, such as DBSCAN or k-Means [64], cannot be used in our case. An illustrative example against both methods are corner kick situations, where most players, across all retrieved situations, are located within the penalty area. Density-based methods would only detect a single cluster in this case, instead of showing us the 11 most likely positions per team. With k-Means, we can specify the exact number of clusters we want to find. However, the problem arises that the number of assigned players to each cluster can vary considerably. Here again, when examining corner kick situations, we would find one or more clusters in the penalty area that are assigned a disproportionate number of players while the remaining clusters are distributed to outliers outside the penalty area, which would result in an extremely unequal cluster size distribution. Thus we would detect a formation, however, it would not reflect reality since not only one, two, or three players are located in the penalty area, but usually the majority of both teams.

To overcome this problem, we need a clustering algorithm that lets us specify the number of clusters while simultaneously yielding equal-sized clusters. In particular, we use the *k-means-constrained library* [65], based on Bradley *et al.*'s *Constrained k-means* clustering algorithm [66], which models the cluster assignment as a minimum cost flow problem based on defined constraints. We make use of these constraints to enforce equal-sized clusters. Thus we can achieve a more fine-grained creation of clusters in areas with a high density of soccer players while not missing clusters in areas with a low density of soccer players.

4.3.2 Movement Pattern Detection

Ultimately, we analyze the most frequent movement trajectories of the players assigned to the clusters computed in the previous step. For this, we perform a trajectory clustering of their subsequent movements. Since the overall direction of the movement is more important than its exact end position, we first translate all trajectories to have the same starting position, that is, the centroid of the cluster. Afterward, we follow standard approaches for trajectory clustering, as described by Andrienko *et al.* [64]. First, we calculate the pairwise distances between all trajectories using dynamic time warping [67]. Second, we perform density-based clustering. However, instead of DBSCAN, we employ the HDBSCAN algorithm [68]— in particular the implementation by McInnes *et al.* [69] — a density-based, hierarchical clustering method, since compared to other density-based clustering algorithms, it is quite robust to parameter selection.
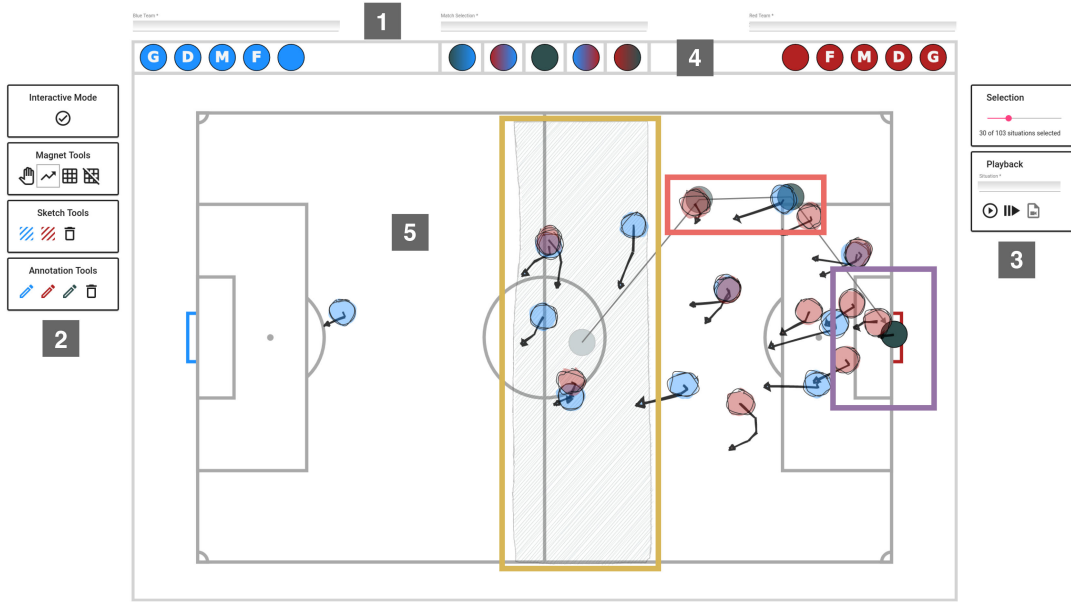
Fig. 5. Overview of our virtual tactic-board. The match selection component (*1*) allows the quick selection of relevant teams and matches. The tool palette (*2*) provides the necessary functionalities to construct more complex situation queries as well as annotation tools. The magnets bar (*4*) in combination with the pitch (*5*) is our recreation of the analog magnetic tactic-board and enables the user to create queries in a familiar way, namely by placing and moving magnets around as well as providing contextual information using drawings. The situation selection and inspection tools (*3*) allow to filter all found situation as well as inspect individual situations in detail. This example shows plays from anywhere in the **offensive midfield** over the **right side** that end with a **shot on goal**.

5   VISUAL EXPLORATION OF MULTIPLE MATCHES

The tactic-board is a readily understood concept to visualize and communicate movement data for a majority of people in the soccer domain and most other invasive team sports. This familiarity is the reason why the tactic-board was chosen as the basis for our visual design. It includes all the functionality needed to satisfy the search and analysis requirements ⚲ S1–S6, Σ A1–A4, while fulfilling the visualization requirements ⌁ V1–V4. Fig. 5 provides an overview of our visual interface and its major components, namely: (*1*) Team and Match Selection, (*2*), Tool Palette, (*3*) Situation Palette, (*4*) Magnet Bar, and (*5*) Pitch. In the following sections, we will present the supplementary visualizations (*1–4*) as well as the primary visualization (*5*) in detail and describe how this combination can enable a seamless visual analysis of tactical behavior in massive soccer datasets.

5.1   Supplementary Visualizations, Dialogs and Tools

Although the supplementary visualizations are not part of the classical magnetic tactic-board, they serve an essential purpose in combination with our primary visualization: They allow us to show the additional features necessary to fulfill search, analysis and visualization requirements at a glance instead of hiding them behind dialogs, menus, and shortcuts. This design choice makes the application entirely usable with just one mouse button or finger on a tablet.

The team and match selection dialogs (*1*) allow the user to assign teams to correspond with the blue and red team in the subsequent analysis steps. It is not only possible to assign a single team to each color, but a set of teams, to represent not only 1:1 relationships but also 1:n,

n:1 or n:m relationships. The only restrictions are that at least one team must be assigned to each color and that one team can only be assigned to one color at a time since a team can never play against itself. The result of this selection is a list of matches in which the blue teams play against the red teams. This selection can be further filtered down using the match selection dialog. This way, the user is able to create arbitrary selections of matches, as required by ⚲ S1.

The tool palette (*2*) design is inspired by commonly used applications, such as graphics editors like Paint or Photoshop. Currently, we support five different groups of tools. First, the users can toggle the interactivity mode of the application. If the interactive mode is activated, each inter-action with the tactic-board produces a new query to find similar situations that are immediately visualized on the tactic-board. This immediate feedback can be regarded as an auto-completion of queries, which allow the user to fine-tune the remaining aspects of the query and effortlessly drill-down to specific situations. Hence, we enable an itera-tive alteration and refinement of the search query as required by ⚲ S4. If the interactive mode is deactivated, the user can place multiple magnets before a query is issued, or can even use the application as a classical tactic-board with annotation capabilities, as long as the interactive mode remains deactivated the whole time.

The situation palette (*3*) allows for the filtering of found situations and to select individual situations for further investigation, as required by Σ A4. Users can either view this playback directly in our application via animation or they have the option to watch a video of real the situation.

The last group of tools are the magnet tools used with the magnets in the magnet bar (*4*) and the pitch. The magnet

move tool, depicted by the hand icon, allows uses to move the magnet from one position to another. The move tool is required to satisfy search requirements ⚲ S2–S3. The magnet trajectory tool, depicted by the symbolic trajectory, allows the user to model movement of the ball or players by dragging a magnet that is currently on the pitch and dropping it at its desired target location in order to satisfy requirement ⚲ S5. Finally, the magnet region query tools are used to issue even more broad queries that can't be supported by placing individual magnets, such as when plays from somewhere in the offensive midfield to the goal should be investigated, as shown in the example in Fig. 5.

The fourth set of tools are the sketch tools that allow users to provide further contextual information to specific situations. At the moment, we support sketching regions in which only players of one team are allowed. This is a powerful supporting tool since many situations in soccer have clear restrictions where players are allowed to be. For instance, during a goalie kick-off, there are no players of the opposing team allowed in the penalty area. Similarly, during open play, we can filter out situations in which a player would be in an offside position.

Finally, the annotation tools that allow free-form drawing and have no effect on the search and analysis of situations but are required in order to satisfy 〰 V4 to communicate additional information or give suggestions, for instance, alternative movement possibilities of players.

The magnet bar, which can be found on some analog tactics boards, serves as a repository for the magnets. In order to add a magnet to the pitch, it can be dragged from the magnets bar and placed at its desired target location. To remove a magnet, it can be dropped over the magnet bar. As can be seen in Fig. 5, we support several types of magnets for each team, which correspond to the different roles a soccer player can have. Usually, magnets on a tactic-board are either blank and their role is either communicated verbally or is determined by the spatial context, or the role is specified using a numeric identifier [70], for instance, 1 is the Goalkeeper. However, both of these schemes have drawbacks. Using no identifiers can be too ambigious and the numeric identifiers can be too restrictive, especially when comparing multiple teams with different formations. In discussion with our expert, we therefore decided on a middle course and use broader role descriptions, namely: **G**oalkeeper, **D**efender, **M**idfielder, **F**orward as well as an unnamed wildcard magnet that matches every role type. Additionally, we include the ball magnet (black) and the ball possession (black to blue or red to black gradient), and ball conquest magnets (red to blue or blue to red gradient) discussed in Section 4.1.

5.2 Soccer Pitch

The soccer pitch (5) is the core of our system and serves both as the interaction component to create search queries by modeling situations of interest as well as the visualizaton component for the detailed inspection of the found situations. To achieve this seamless visual analysis approach, the supplementary visualizations, dialogs and tools (1–4), described above are combined with the enhanced query creation and situation aggregation algorithms, described in Section 4, to satisfy all requirements identified in Section 3.

5.2.1 Visualization and Interaction Design

In the visual design of our system, be it in the form of placed magnets, trajectory or region-based queries, or additional sketches, we have deliberately limited ourselves to visualizations that would also be possible on an analog magnetic tactic-board. Thus, players are visualized with the help of magnets and the movements of players as well as the remaining queries with the help of arrows or hatched areas, which would usually be created using whiteboard markers. Similarly, we aim to base our interaction design as much as possible on the interactions of the conventional tactic-board. Thus, a situation request in which a player should stand at a certain position is created by dragging a magnet from the magnets bar (4) to the corresponding position on the pitch. Likewise, a trajectory request is constructed by selecting the trajectory tool and dragging a magnet from its old position to a new position. This exact process can be observed in the example in Fig. 2. However, we additionally draw the old position of the magnet as well to ensure an overview of the initial and end situation after several such actions. Region-based queries and sketches are produced by drawing a certain region on the pitch with either a magnet or a pen. These visual and interaction design choices and restrictions were chosen to provide an interface that would be intuitively understandable to users from the domain to meet 〰 V1.

5.2.2 Visual Situation Aggregation

Once a user has defined a situation with the help of the tool palette (2), it is sent to our backend. There, it is converted into a request using the methods we describe in Section 4 and searched for in all the matches selected by the user. All found situations are aggregated: First, using the constraint-based k-means to detect the formation of the players. Second, using the HDBSCAN algorithm as described in Section 4.3 to detect the most likely subsequent movements of the clustered players. We then visualize an overview of the most likely formation and most probable aggregated movement in the retrieved situations. The clusters are added as new glyphs to the pitch visualization. However, we need to have a distinction between user-placed magnets and formation clusters, as found in the aggregation step. Since hue and position are already used, we need an additional visual variable. Studies show that sketchiness is a suitable choice for the visualization of uncertain data and visualizations for non-expert audiences [71], which both apply in our case. Hence, we decided to use sketchiness as the visual variable for the glyph to distinguish between formation clusters and user-placed magnets. Furthermore, to encode the variance of the distribution of the underlying clustered data, we can vary the drawn clusters' degree of sketchiness to illustrate this without incorporating any further visual variables, helping us avoid information overload caused by too many different used visual variables. If all moving objects assigned to a cluster are positioned in the same location, we draw a near perfect circle, as shown left in Fig. 6. On the other hand, if all moving objects are distributed all over the soccer pitch, we draw a very rough sketch, as shown right in Fig. 6. Finally, we visualize each the most probable movements of the clustered players as found by our trajectory clustering algorithm using arrows. We are encoding the size of the trajectory cluster additionally by the thickness of the drawn arrow. If the trajectory
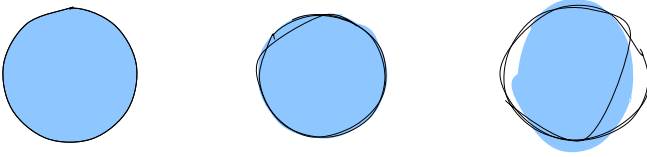
Fig. 6. Clusters drawn in various levels of sketchiness to encode the variance of the distribution of the locations of the underlying moving objects.

clustering does not identify any cluster in the subsequent movements of the players, no arrow with the most probable movement will be drawn.

5.2.3 Interactive Situation Exploration

After the situation query has been submitted and processed using our relaxed query method satisfying 🔍 S6, the user can start with the exploration of the found situations. Here, we support several options in order to fulfill the analysis requirements while respecting the visualization requirements, apart from showing how often such a situation occurs to satisfy Σ A1. First of all, the user can get an overview of the average distribution of his or her players and the players of the opposing team during the found situations. For this, the clusters drawn in the previous step are used, since they enable the user to quickly gain insight into the positioning of the teams, which can provide conclusions about the tactical approach of the teams, satisfying Σ A3.

We also offer details-on-demand 〰 V3, when the user hovers over a magnet or a drawn cluster. As can be seen in Fig. 1, we show the actual positions of the players in individual situations as well as their individual subsequent movements 〰 V2. We draw a trajectory from the start of each situation until several seconds after the situation has ended. To indicate the position of the player at the end of the current situation, we use a small circle that also shows the actual role of the player, thus enabling a detailed examination of individual players throughout the situation and afterward. Additionally, we offer a tooltip that provides information about the most common roles in a cluster as well as the most likely players (Σ A2). For data privacy reasons, we do not give an example of this. However, this information can be important for coaches, as they can identify players who are often involved in certain moves, which can help them to adjust their team line-up and tactics.

In order to provide even more information about individual situations and to satisfy Σ A4, 〰 V2, we also offer playback functionality. The user selects one situation and can play it either from start to the end for either the current sketched situation or the match phase in which the situation took place. Thus, we can offer information not only about the situation itself, but also about how the individual situations emerged. Additionally, the user can also view the original video recording of a situation to obtain information not contained in the 2D data and visualizations. We have deliberately decided not to offer an animation of aggregated situations. During our experiments, we realized that this leads to several problems. For instance, the average position of the players and the ball does not always overlap entirely, which can lead to ghost movement. This means the ball cluster changes direction or speed even though no player

cluster is directly positioned at the ball. This can lead to confusion when watching the animation since it shows behavior that cannot physically occur during regular play.

Finally, after the exploration of aggregated and individual situations and identification of tactical behavior by the users, they can also use our proposed system to communicate this information with others. For this purpose, they can rely on the overview visualization, as well as the playback functionality. Additionally, they can provide further information using our annotation tools, which are available in both case. In this way, they can highlight certain key aspects of the situations or certain plays or even individual players, thus guiding the attention to these critical aspects during the presentation of their found results to help them communicate their findings as required by 〰 V4.

5.3 Implementation Details

The system presented in this section is built as an interactive website, allowing us to use the same application with the same interaction design on desktop computers, laptops, and touch-oriented interfaces such as tablets. The datasets are stored in a PostgreSQL database [72], including the PostGIS extension [63]. The communication to the database was realized as a RESTful API using Flask [73], and the frontend was implemented as an Angular web application [74]. Two additional Javascript libraries are heavily used, namely, d3. js [75] for the interactions and visualizations and rough. js [76] for the sketchy drawing of visualizations.

6 EXPERT EVALUATION

In order to continually evaluate and iteratively improve our proposed approach for novel multi-match analysis, we performed weekly meetings with the coach (in the following referred to as *Expert A*) of a Austrian first league soccer club that contacted us two years ago as described in Section 3. Additionally, *Expert A* underwent training to become a certified video analyst. In order to get insights into the generalizability of our proposed approach, we also invited five additional soccer experts that were not involved in the design and development process of our system (in the following referred to as *Expert 1–6*) for user studies of our final system as well as additional interviews. These participants were selected to cover, together with *Expert A*, all potential user groups in the soccer domain, namely, coaches, players, and analysts. *Expert 1* and *Expert 2* are both soccer players since 2008 and 2011, respectively, and since 2018 both play for the Austrian first league club TSV Hartberg. *Experts 3–5* are professional soccer match and video analysts. *Expert 3* is a video analyst since 2008 and currently the head of analysis for the German first league club SV Werder Bremen since 2016. *Expert 4* is match and video analyst for the German first league club TSG 1899 Hoffenheim since 2013. *Expert 5* is working in scouting, recruitment, and performance analysis for the English first league soccer club Leicester City since 2019. *Expert 6* is a match and performance analyst for the Spanish first league soccer club FC Barcelona. The interviews with *Expert A* and *Experts 1–4* were conducted in German, thus quotes and comments are translated into English in the following sections. The interviews with *Expert 5–6* were conducted in English, therefore their quotes are verbatim.

(Q_1) *Slow build up play over the left side*   **(Q_2)** *Ball conquest followed by counter*   **(Q_3)** *Crosses into the penalty area*
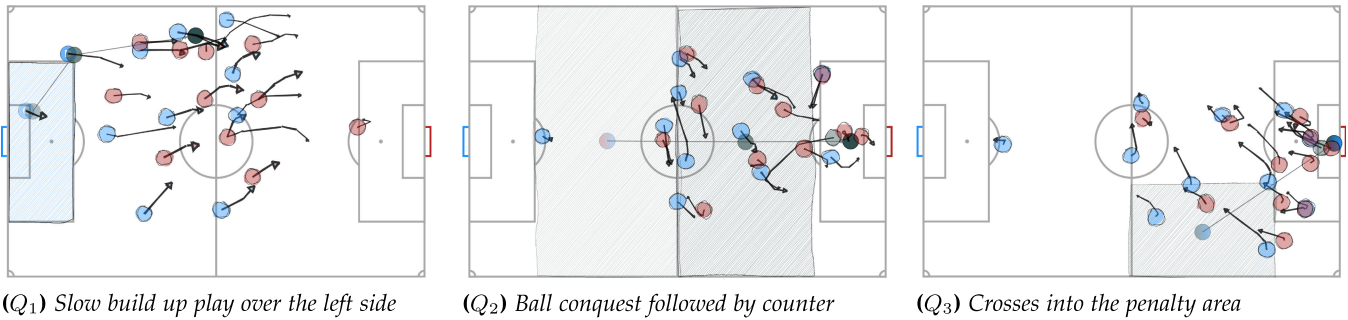
Fig. 7. Prepared queries for the evaluation. For each query, the participants were asked to judge the relevance of the first five retrieved results. The results in Table 1 show that using only a few magnets, we can achieve excellent retrieval performance, with an overall nDCG@5 of 0.88 over the three queries with our five study participants.

6.1 Study and Interview Design

In our expert studies, we investigated the multi-match analysis capabilities of our proposed approach with each expert individually. The basis for the expert studies was spatio-temporal multi-match data of 250 individual games from which movement data was extracted using optical tracking. In our case, the spatial data is specified with a resolution of 10 centimeters using coordinate-based referencing (x and y position on the pitch) with the center spot of the pitch being the origin. The temporal data is a discrete, linearly ordered set of time instants with a resolution of 100 milliseconds. In total our dataset consists of around 330 million tracked locations of the ball and players. We began each session by introducing each expert to the system and its available interaction and analysis features. We first described how our design was motivated by the familiar magnetic tactic-board, using examples from our initial interview with Expert A (see Fig. 2). Afterward, we introduced them to our digital magnetic tactic-board by first using it to emulate a physical, analog tactic board to demonstrate the visualizations and introduce them to the interaction possibilities. We then turned on the query mode to show how their interactions can be used as the basis for queries to find suitable situations in our multi-match data set and how the additional visualizations help to provide an overview of the retrieved results. To further familiarize them with our digital representation, we performed several typical queries together with the experts. After these introductory steps, we presented them three predefined queries, which were chosen to include a wide range of different sketch tools and magnets, shown in Fig. 7:

Q_1) Slow build-up game, starting at the goalkeeper into the left midfield via a defender on the left side (Fig. 7 (Q_1)).

Q_2) Ball conquest by a team in their half of the pitch, followed by a fast counter-attack ending in the opponent's penalty area (Fig. 7 (Q_2)).

Q_3) Crosses from the blue team over the right side into the penalty area of the opossing team (Fig. 7 (Q_3)).

For each of these queries, we had *Experts 1 – 5* rate the relevance of the first five retrieved results. They could rate the results as either not relevant, moderately relevant, or highly relevant. Afterward, the study participants were free to follow their own analysis ideas and perform their own queries. For both the predefined queries and their own analysis ideas, the experts had access to the whole multi-match soccer dataset of 250 matches. Over the whole study, we encouraged the participants to provide comments and

feedback via a Thinking-Aloud Protocol [77], [78]. Lastly, we had an interview with our study participants. We conducted a questionnaire with yes/no questions, but the participants were encouraged to provide short explanations or justifications for their answers. We encouraged an open discussion in which the study participants could give us their ideas for possible applications of our system, suggestions for improvement, or other comments. Ultimately, we were interested in determining if and how our proposed system would find a place in their day-to-day analysis routine.

6.2 Findings

We summarize the results of our interviews with our study participants in the following section. First, we conduct a quantitative analysis of the relevance ratings of the three queries from Fig. 7. Afterward, we present qualitative results from two use cases, which were developed by the experts during our evaluation. Finally, we highlight feedback and remarks from the participants that emerged during the Thinking-Aloud process and subsequent discussion.

6.2.1 Quantitative Assessment

We evaluated the retrieved situations for three different queries with all newly invited experts 1-5. Each expert was asked to assess the relevance of the first five situations found for each of the queries shown in Fig. 7. They could rate the relevance as either not relevant, moderately relevant, or highly relevant. To take these relevance ratings into account and to punish showing irrelevant or only moderately relevant results before highly relevant results, we have chosen the normalized discounted cumulative gain (nDCG@5) [79] as our quality metric for retrieval performance. An advantage of this metric is that it does not require all situations to be labeled already, which is unfeasible for such a large dataset on the one hand. On the other hand, labeling a situation is entirely dependent on the analyst and varies between users, meaning that there is no definable gold standard that could be used to calculate metrics such as recall. The results of this evaluation are summarized in Table 1. We additionally averaged the nDCG per expert, per query and overall. The overall score of 0.89 shows that our tool is capable of retrieving relevant results for a multitude of situations, requiring only a few user inputs. There is a small decrease in performance for Q_2, with an nDCG of 0.81 compared to 0.92 for Q_1 and 0.94 for Q_3. However, this decrease can be explained in a small discrepancy in our definition of the ball conquest, compared

TABLE 1
Results of Our Retrieval Performance Evaluation Showing
the nDCG@5 for Each Query Described in Fig. 7
for Our Five Study Participants

Participant	Q_1	Q_2	Q_3	Average Expert Score
Expert 1	1.00	0.90	0.77	0.89
Expert 2	1.00	1.00	1.00	1.00
Expert 3	0.77	0.77	1.00	0.85
Expert 4	1.00	0.76	0.87	0.88
Expert 5	0.77	0.55	1.00	0.78
Expert 6	1.00	0.86	1.00	0.95
Average Query Score	0.92	0.81	0.94	**0.89**

The overall results show the capability of our approach to find relevant results, with a slight decrease in performance for Q_2, which can be explained by how we define ball conquest.

to the expectations of the study participants. For most experts, a ball conquest occurs, when the team that conquers the ball gains a lasting control over the ball. Our current definition was, that this occurs, as soon as the one team loses control of the ball. This short difference of a few seconds was the main issue that lead most experts to rate queries as *not relevant* or *moderately relevant* instead of *moderately relevant* or *highly relevant*.

6.2.2 Qualitative Evaluation

The experts found numerous interesting patterns and gained valuable insights while using our system. Due to the limited space, we restrict ourselves in the following to the three most interesting use cases.

The first use case, shown in Fig. 8, was found during the exploration by *Expert 1*. In this figure, the kick-offs of the opposing team's goalkeeper are to be examined in general. The goalkeeper and the ball were dragged into the penalty area to create this query. Additionally, the sketch tool was used to indicate that no red team players are to be in the penalty area. Afterward, the first 40 situations were selected and summarized. In summary Fig. 8 (1), it can be clearly seen that the most likely option is a long pass into the left midfield. The visualization of all movements Fig. 8 (2) corroborates the summarization. Some individual deviating options exist, but in most cases, the ball was mostly shot directly into the midfield.
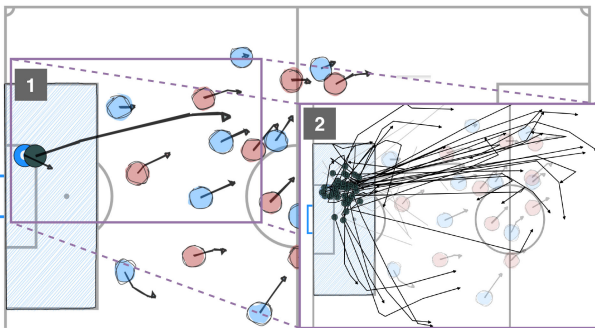

Fig. 8. *Expert 1* wanted to investigate the kick-off behavior of a goalkeeper of an opposing team. Our summary visualization (1) and detail view (2) show, that in most situations the goalkeeper plays a long ball into the midfield. This was previously unknown, yet interesting behavior for *Expert 1*.
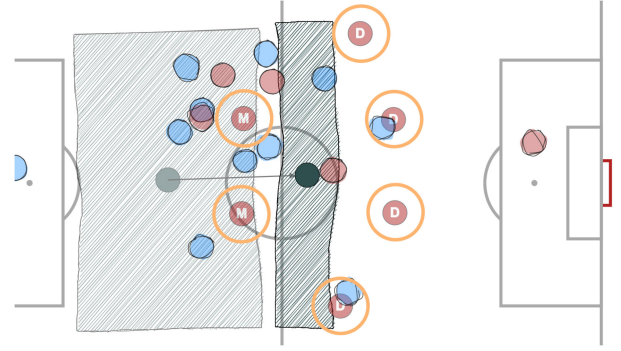

Fig. 9. *Expert 3* investigated situations in which teams successfully overplayed the line of midfielders of the opposing team. He placed six player magnets, highlighted in orange, two midfielders and four defenders of the read team, and used the ball magnet and the drag functionality to simulate a pass over the line of midfielders.

According to *Expert 1*, this is was new insight for him. He already knew beforehand that the defenders of this team often play long balls as well, hoping to conquer the second ball in the opponent's half. However, he had not known beforehand that the goalkeeper shoots it directly into the midfield that often compared to other teams, and this was an interesting new insight for him.

The second use case of note was explored by Expert 3. He wanted to find situations where teams successfully overplayed the line of midfielders of the opposing team. To construct this query, he first placed the ball in the blue team's half and additionally placed two midfielders and four defenders of the red team around the midfield line, as shown in Fig. 9. He then used the drag & drop functionality to move the ball into the area between the midfielders and the defenders. The resulting situation, which he examined in detail, was "in the direction he envisioned it to go" and that this is the sort of "key query" that an analyst faces.

The third use case was found by *Expert A* when investigating ball conquests in front of the opposing team's penalty area. The specific situation of interest is shown in Fig. 10. Here, a mispass of the red team's right defender is shown, after being put under pressure by the blue team's forward, which resulted in a ball conquest of the blue team, leading to a dangerous shot on goal. According to *Expert A*, this particular defender often has problems with such long passes to the other side; thus, putting him under pressure may force him to make disproportionately more mispasses in such situations. This way, his team can create more options close to the opposing team's penalty area, leading to more and better chances for shot on goals. To him, such situations are of great interest to show to his team, as they can serve as examples of which tactics to employ in their next match against specific opponents.

6.2.3 Expert Interview Feedback

According to all invited experts, the tactics board plays a vital role in preparation and matches, for example, during half-time meetings, and is seen here as a suitable medium to facilitate communication between analysts, coaches, and players. All experts agree that linking the tactics board to massive soccer data sets is useful for searching situations. Especially that there is the possibility to analyze several
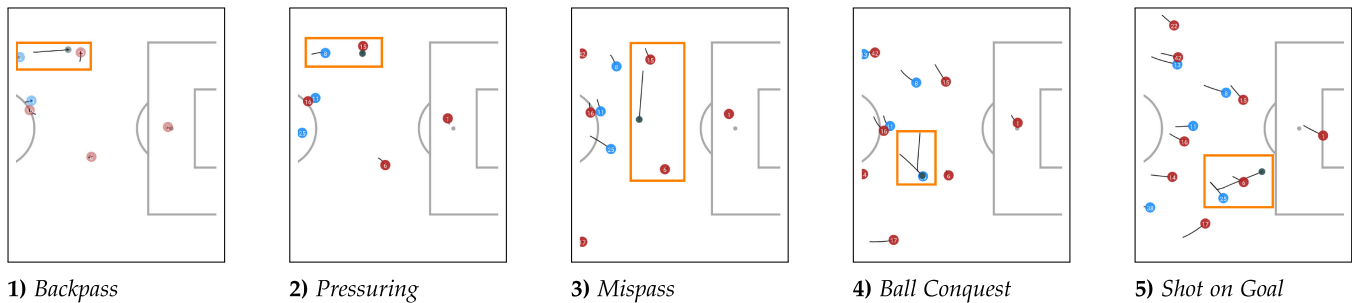
| 1) *Backpass* | 2) *Pressuring* | 3) *Mispass* | 4) *Ball Conquest* | 5) *Shot on Goal* |

Fig. 10. Ball conquest by blue team followed by a shot on goal after a mispass of the red team's right defender, as found by *Expert A*. According to him, the defending player is prone to mispasses when being put under pressure. Thus, this is a valuable situation to show to his team, to teach them to use the same tactic during their next games against the red team.

games, especially from different teams, at once. According to *Expert 4*, "This is crucial to identify recurring patterns."

The visual design was judged to be sound and understandable, apart from *Expert 1*, and it was clearly distinguishable for them, through the sketchy design, which magnets were placed by themselves, and which were shown by the computer as aggregations. *Expert 1* stated, "I can't imagine much about it if you only see red and blue circles and arrows." For him, the inspection of individual situations in detail was of much greater importance. Here, the other experts also agreed with him that a subsequent analysis of particular situations with the help of the animation and the video is instrumental. *Expert 3* wished for even closer integration with the video. Not only should the corresponding situation be shown, but additional information, for example, passes, the defensive line, or annotations from the tactic-board, should be directly visualized in the video.

The experts were satisfied with the visualization and query construction options offered, especially in view of the fact that the tactic-board is to serve as a communication medium between the different user groups in soccer. *Experts 3, 4* and *5* would have liked to see an even more comprehensive range of interaction and analysis possibilities, especially for analysts. For example, the inclusion of the game context, such as the current score, or more options to make the queries more or less specific.

Experts 2 and *6* both made the interesting statement that, so far, the tactic-board is only used to visualize idealized movements and scenarios, which might not happen that way in the real game. Both said that by linking it to real data, it can show what happens in actual play, which is, much more dynamic and unpredictable than idealized scenarios, thus providing feedback useful and helpful for players.

Many individual suggestions for improvement were also mentioned, which we will address in the next version of our system, e.g., *Expert 5* would like to see a different design for the ball magnets, inspired by the black and white hexagon pattern. Apart from that, the experts are "*enthusiastic*" (quoting *Expert A*) about the new possibilities combining the metaphor of an analogous magnetic tactic-board with real underlying soccer movement data of multiple matches.

7 DISCUSSION AND CONCLUSION

We introduced a novel approach for the visual exploration and communication of tactics in soccer. The most important question to discuss is whether we are able to assess tactical behavior with our proposed design. We are well aware that we neglect important aspects of tactics as for example the current game context (scores), involved body-parts or further advanced features like the expressed body-language. Scores will be included in future as an additional filtering step which is easy to include in our system. Body posture detection is still expensive with respect to time and computation and more error-prone than pure location-tracking. We intentionally focused on the location data as absolute and relative positioning and the resulting movement is the strongest indicator for tactics in soccer and other invasive team sports. We enhance the location-based queries with role context semantically meaningful to the soccer domain. We chose the most common roles in soccer with goalkeeper, defender, midfielder, and forward. We show in our evaluation that domain experts do believe and agree with us that we are covering the exploration of tactics and foster the communication of tactics in their day-to-day work. In order to further support soccer coaches and analysts in the future, we plan to publish detailed descriptions of state-of-the-art tactical behaviors that can be identified by our system, what patterns can be found and how these can be interpreted. We are convinced, we presented a decent progress towards a key goal of visual analytics in sport: researching a useful design combining analytics with visualizations enabling domain experts to holistically understand their object-of-study and to communicate the results of their analysis.

In our design, we need to cope with uncertainties resulting from user input and our visual result communication based on clustering and abstraction. The graphical input for the location query is a transformation of the vague human thinking intermixing absolute and relative positioning into exact positions the computer is able to process. The exact absolute positioning of players is usually not interesting to domain experts when understanding tactics. We treat the imperfect user input by the introduced relaxation queries ensuring that locations of players can vary from the query by some extent. Additionally, we proposed and included region-queries making the involved query uncertainties explicit. The visual representation of the query result is based on clustering and aggregations of player positions and therefore introduces uncertainties into the visualizations. For a successful visual design, we need to address these uncertainties and make them transparent to the user. We include visual clues to denote the uncertainties with sketchy rendering of cluster representatives and give detailed information of the clustered

players on mouse-over. The expert study and the query results show our valid transformation of vague input into computer readable queries and experts understand that they look at aggregated query results.

We will address in our future work the inclusion of situational parameters as the current yellow and red cards or the scores as described above. These parameters are affecting the tactics of the teams and leading to other observable movement patterns. The movement patterns being displayed as query results could be communicated by further means of simplification and generalization. We plan as a first step to investigate the methods proposed by Sacha et al. [43]. The interactive abstraction would allow us to selectively include more detailed movement information maybe even enhanced with metadata or statistics. The goal is to achieve an interactive hybrid between abstracted and high-detail display. Another direction is the longitudinal analysis of our design influencing the day-to-day business of soccer coaches and analysts. We are especially interested in the commonalities and differences in search patterns of different entities: coaches, analysts, and even teams. Additionally, we would like to investigate the collaborative potential our approach offers. Currently, we only focused on individuals employing the visual query system and not on collaborative setting in a coach/co-coach or coach/analyst situation. Last but not least, an even more quantitative evaluation would strengthen our design research. Unfortunately, highly trained and well-known soccer coaches, video analysts and players are very hard to recruit for such evaluations. Consequently, in-depth analyses with few (in terms of statistics) experts are our envisioned way for an expressive evaluation. From a computational performance perspective, we are already able to compute efficiency scores of the situation querying in our data set of 250 games consisting of around 330 million individual tracked locations. For instance, the situation shown in Fig. 5, consisting of four subsequent actions of one entity, is searched for in five selected matches. Over ten tries, the query to identify those situations takes around 0.95 seconds on average. The situation in Fig. 7 (Q_1) consists of a region constraint query, two entities without subsequent actions, and one entity with three subsequent actions. Here, searching in 26 selected matches takes around 1.8 seconds on average. These results already look promising and enable an interactive visual exploration of tactical behavior. However, we still see room for improvements to enable near real-time querying of situations. This would allow for additional features, such as instant feedback, while a user is still in the process of dragging a magnet.

In conclusion, we proposed a visual analytics application supporting coaches visually exploring, externalizing, and communicating their knowledge about tactical behavior in soccer. The visual query interface allows soccer coaches to interactively and intuitively express their information need. Our quantitative results demonstrate that our relaxed query algorithm incorporating region queries copes with uncertainties and can provide meaningful situations in massive soccer datasets, requiring only a few interactions. In addition, the qualitative evaluation with the experts and their feedback shows that the found results can provide them

with new insights about the behavior of players and teams during soccer matches. Hence, at the time of writing this paper, the experts are enthusiastic about our novel approach and look forward to actively use it in every day match analysis.

REFERENCES

[1] Agile Sports Technologies, Inc., "Sportscode performance analysis systems," 2020. Last Accessed: Dec. 08, 2020. [Online]. Available: https://www.hudl.com/products/sportscode
[2] Dartfish, "Dartfish video and data analysis systems," 2020. Last Accessed: Dec. 08, 2020. [Online]. Available: http://www.dartfish.com
[3] D. Memmert and D. Raabe, Revolution im Profifußball: Mit Big Data zur Spielanalyse 4.0. Berlin Germany: Springer-Verlag, 2017.
[4] L. Probst, F. Brix, H. Schuldt, and M. Rumo, "Real-time football analysis with StreamTeam," in Proc. 11th ACM Int. Conf. Distrib. Event-Based Syst., 2017, pp. 319–322.
[5] L. Probst, F. Rauschenbach, H. Schuldt, P. Seidenschwarz, and M. Rumo, "Integrated real-time data stream analysis and sketch-based video retrieval in team sports," in Proc. IEEE Int. Conf. Big Data, 2018, pp. 548–555.
[6] M. Stein et al., "Bring it to the pitch: Combining video and movement data to enhance team sport analysis," IEEE Trans. Vis. Comput. Graph., vol. 24, no. 1, pp. 13–22, Jan. 2018.
[7] Y. Seo, S. Choi, H. Kim, and K.-S. Hong, "Where are the ball and players? Soccer game analysis with color-based tracking and image mosaick," in Proc. Int. Conf. Image Anal. Process., 1997, pp. 196–203.
[8] J. Liu, X. Tong, W. Li, T. Wang, Y. Zhang, and H. Wang, "Automatic player detection, labeling and tracking in broadcast soccer video," Pattern Recognit. Lett., vol. 30, no. 2, pp. 103–113, 2009.
[9] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in Proc. Eur. Conf. Comput. Vis., 2002, pp. 661–675.
[10] C. Mutschler, H. Ziekow, and Z. Jerzak, "The DEBS 2013 grand challenge," in Proc. 7th ACM Int. Conf. Distrib. Event-Based Syst., 2013, pp. 289–294.
[11] M. Stein et al., "Identifying the match plan: Context-oriented interactive visual analysis of football matches," in Football Analytics: Now and Beyond. A Deep Dive Into the Current State of Advanced Data Analytics, vol. 1. Barça Innovation Hub, 2019. sec. 10, pp. 146–173.
[12] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 7291–7299.
[13] A. Arbu és-Sangüesa, G. Haro, C. Ballester, and A. Martín, "Head, shoulders, hip and ball... hip and ball! Using pose data to leverage football player orientation," in Proc. 2nd Barça Sports Analytics Summit, 2019.
[14] L. Pishchulin et al., "DeepCut: Joint subset partition and labeling for multi-person pose estimation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 4929–4937.
[15] M. Stein, D. Seebacher, T. Karge, T. Polk, M. Grossniklaus, and D. A. Keim, "From movement to events: Improving soccer match annotations," in Proc. Int. Conf. Multimedia Model., 2019, pp. 130–142.
[16] F. Sulser, I. Giangreco, and H. Schuldt, "Crowd-based semantic event detection and video annotation for sports videos," in Proc. Int. ACM Workshop Crowdsourcing Multimedia, 2014, pp. 63–68.
[17] V. Tovinkere and R. J. Qian, "Detecting semantic events in soccer games: Towards a complete solution," in Proc. IEEE Int. Conf. Multimedia Expo, 2001, pp. 833–836.
[18] C. Perin, R. Vuillemot, C. D. Stolper, J. T. Stasko, J. Wood, and S. Carpendale, "State of the art of sports data visualization," Comput. Graph. Forum, vol. 37, no. 3, pp. 663–686, 2018.

[19] F. R. Goes et al., "Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review," Eur. J. Sport Sci., 2020. [Online]. Available: https://doi.org/10.1080/17461391.2020.1747552

[20] M. Stein et al., "How to make sense of team sport data: From acquisition to data modeling and research aspects," Data, vol. 2, no. 1, 2017, Art. no. 2.

[21] D. Memmert and R. Rein, "Match analysis, big data and tactics: Current trends in elite soccer," German J. Sports Med./Deutsche Zeitschrift fur Sportmedizin, vol. 69, no. 3, pp. 65–72, 2018.

[22] H.-C. Shih, "A survey of content-aware video analysis for sports," IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 5, pp. 1212–1231, May 2018.

[23] P. Cintia, M. Coscia, and L. Pappalardo, "The Haka network: Evaluating rugby team performance with dynamic graph analysis," in Proc. IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Mining, 2016, pp. 1095–1102.

[24] P. Lucey, A. Bialkowski, P. Carr, Y. Yue, and I. Matthews, "How to get an open shot: Analyzing team movement in basketball using tracking data," in Proc. 8th Annu. MIT SLOAN Sports Analytics Conf., 2014.

[25] L.-M. Xia, Q. Wang, and L.-S. Wu, "Vision-based behavior prediction of ball carrier in basketball matches," J. Central South Univ., vol. 19, no. 8, pp. 2142–2151, 2012.

[26] V. Bačík and M. Klobučník, "Possibilities of using selected visualization methods for historical analysis of sporting event–An example of stage cycling race tour de France," Quaestiones Geographicae, vol. 37, no. 3, pp. 5–22, 2019.

[27] M. Horton, "Algorithms for the analysis of spatio-temporal data from team sports," PhD thesis, School Inf. Technologies, Univ. Sydney, Sydney, Australia, 2018.

[28] T. Polk, J. Yang, Y. Hu, and Y. Zhao, "TenniVis: Visualization for tennis match analysis," IEEE Trans. Vis. Comput. Graph., vol. 20, no. 12, pp. 2339–2348, Dec. 2014.

[29] T. Polk, D. Jäckle, J. Häußler, and J. Yang, "CourtTime: Generating actionable insights into tennis matches using visual analytics," IEEE Trans. Vis. Comput. Graph., vol. 26, no. 1, pp. 397–406, Jan. 2020.

[30] Y. Wu et al., "iTTVis: Interactive visualization of table tennis data," IEEE Trans. Vis. Comput. Graph., vol. 24, no. 1, pp. 709–718, Jan. 2018.

[31] S. Ye et al., "ShuttleSpace: Exploring and analyzing movement trajectory in immersive visualization," IEEE Trans. Vis. Comput. Graph., vol. 27, no. 2, pp. 860–869, Feb. 2021.

[32] D. Sumpter, Soccermatics: Mathematical Adventures in the Beautiful Game. London, U.K.: Bloomsbury Publishing, May 2016.

[33] J. Castellano et al., Footbal Analytics: Now and Beyond. Barça Innovation Hub, 2019.

[34] C. Carling, A. M. Williams, and T. Reilly, Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance. Hove, U.K.: Psychology Press, 2005.

[35] C. Perin, R. Vuillemot, and J.-D. Fekete, "SoccerStories: A kick-off for visual soccer analysis," IEEE Trans. Vis. Comput. Graph., vol. 19, no. 12, pp. 2506–2515, Dec. 2013.

[36] H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. A. Keim, and O. Deussen, "Feature-driven visual analytics of soccer data," in Proc. IEEE Conf. Vis. Analytics Sci. Technol., 2014, pp. 13–22.

[37] G. Andrienko et al., "Visual analysis of pressure in football," Data Mining Knowl. Discov., vol. 31, no. 6, pp. 1793–1839, 2017.

[38] M. Stein et al., "Director's cut: Analysis and annotation of soccer matches," IEEE Comput. Graph. Appl., vol. 36, no. 5, pp. 50–60, Sep./Oct. 2016.

[39] Y. Wu et al., "ForVizor: Visualizing spatio-temporal team formations in soccer," IEEE Trans. Vis. Comput. Graph., vol. 25, no. 1, pp. 65–75, Jan. 2019.

[40] J. Fernandez and L. Bornn, "Wide open spaces: A statistical technique for measuring space creation in professional soccer," in Proc. Sloan Sports Analytics Conf., 2018.

[41] W. Spearman, A. Basye, G. Dick, R. Hotovy, and P. Pop, "Physics-based modeling of pass probabilities in soccer," in Proc. 11th MIT Sloan Sports Analytics Conf., 2017.

[42] M. Stein et al., "From game events to team tactics: Visual analysis of dangerous situations in multi-match data," in Proc. 1st Int. Conf. Technol. Innov. Sports Health Wellbeing, 2016, pp. 1–9.

[43] D. Sacha et al., "Dynamic visual abstraction of soccer movement," Comput. Graph. Forum, vol. 36, no. 3, pp. 305–315, 2017.

[44] M. Stein, J. Häußler, D. Jäckle, H. Janetzko, T. Schreck, and D. A. Keim, "Visual soccer analytics: Understanding the characteristics of collective team movement based on feature-driven analysis and abstraction," ISPRS Int. J. Geo-Inf., vol. 4, no. 4, pp. 2159–2184, 2015.

[45] J. Gudmundsson and T. Wolle, "Football analysis using spatio-temporal tools," Comput. Environ. Urban Syst., vol. 47, pp. 16–27, 2014.

[46] X. Xie et al., "PassVizor: Toward better understanding of the dynamics of soccer passes," IEEE Trans. Vis. Comput. Graph., vol. 27, no. 2, pp. 1322–1331, Feb. 2021.

[47] G. Andrienko et al., "Constructing spaces and times for tactical analysis in football," IEEE Trans. Vis. Comput. Graph., vol. 27, no. 4, pp. 2280–2297, Apr. 2021.

[48] E. Delibas, A. Uzun, M. F. Inan, O. Guzey, and A. Cakmak, "Interactive exploratory soccer data analytics," INFOR: Inf. Syst. Oper. Res., vol. 57, no. 2, pp. 141–164, 2019.

[49] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel, Visual Analytics of Movement. Berlin, Germany: Springer-Verlag, 2013.

[50] N. Pelekis, G. Andrienko, N. Andrienko, I. Kopanakis, G. Marketos, and Y. Theodoridis, "Visually exploring movement data via similarity-based analysis," J. Intell. Inf. Syst., vol. 38, pp. 343–391, Apr. 2012.

[51] L. Zeng, Y.-J. Liu, J. Wang, D.-L. Zhang, and M. M.-F. Yuen, "Sketch2Jewelry: Semantic feature modeling for sketch-based jewelry design," Comput. Graph., vol. 38, pp. 69–77, Feb. 2014.

[52] J. Lee and T. A. Funkhouser, "Sketch-based search and composition of 3D models," in Proc. 5th Eurographics Conf. Sketch-Based Interfaces Model., 2008, pp. 97–104.

[53] J. Saavedra and B. Bustos, "Sketch-based image retrieval using keyshapes," Multimedia Tools Appl., vol. 73, no. 3, pp. 2033–2062, 2014.

[54] Google, "SimpleMinCostFlow," 2020. Last Accessed: Dec. 08, 2020. [Online]. Available: https://developers.google.com/optimization/flow/mincostflow

[55] J. Bernard et al., "VisInfo: A digital library system for time series research data based on exploratory search - A user-centered design approach," Int. J. Digit. Libraries, vol. 16, no. 1, pp. 37–59, 2015.

[56] L. Shao et al., "Guided sketching for visual search and exploration in large scatter plot spaces," in Proc. EuroVA Int. Workshop Vis. Analytics, 2014, pp. 19–23.

[57] L. Shao, D. Sacha, B. Neldner, M. Stein, and T. Schreck, "Visual-interactive search for soccer trajectories to identify interesting game situations," in Proc. Conf. Vis. Data Anal., 2016, pp. 1–10.

[58] M. Stein, H. Janetzko, T. Schreck, and D. A. Keim, "Tackling similarity search for soccer match analysis: Multimodal distance measure and interactive query definition," IEEE Comput. Graph. Appl., vol. 39, no. 5, pp. 60–71, Sep./Oct. 2019.

[59] L. Sha, P. Lucey, Y. Yue, P. Carr, C. Rohlf, and I. Matthews, "Chalkboarding: A new spatiotemporal query paradigm for sports play retrieval," in Proc. 21st Int. Conf. Intell. User Interfaces, 2016, pp. 336–347.

[60] L. Sha et al., "Interactive sports analytics: An intelligent interface for utilizing trajectories for interactive sports play retrieval and analytics," ACM Trans. Comput.-Hum. Interact., vol. 25, pp. 13:1–13:32, Apr. 2018.

[61] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," IEEE Trans. Vis. Comput. Graph., vol. 18, no. 12, pp. 2431–2440, Dec. 2012.

[62] Wikipedia, "Association football tactics," 2021. Last Accessed: Nov. 17, 2021. [Online]. Available: https://en.wikipedia.org/wiki/Association_football_tactics#Moving_into_space_%E2%80%93_the_diagonal_run

[63] PostGIS Project Steering Committee, "PostGIS," 2020. Last Accessed: Dec. 08, 2020. [Online]. Available: https://postgis.net/

[64] G. L. Andrienko, N. V. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti, "Interactive visual clustering of large collections of trajectories," in Proc. 4th IEEE Symp. Vis. Analytics Sci. Technol., 2009, pp. 3–10.

[65] J. Levy-Kramer, "k-means-constrained," 2020. Last Accessed: Dec. 08, 2020. [Online]. Available: https://github.com/joshlk/k-mean-constrained

[66] P. S. Bradley, K. P. Bennett, and A. Demiriz, "Constrained k-means Clustering," Microsoft Research, Redmond, 2000.

[67] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. Acoust., Speech, Signal Process., vol. 26, no. 1, pp. 43–49, Feb. 1978.

[68] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Proc. Pacific-Asia Conf. Knowl. Discov. Data Mining*, 2013, pp. 160–172.

[69] L. McInnes, J. Healy, and S. Astels, "HDBSCAN: Hierarchical density based clustering," *J. Open Source Softw.*, vol. 2, no. 11, 2017, Art. no. 205.

[70] Deutsche Fußball Liga GmbH, "Soccer positions explained," 2020. Last Accessed: Dec. 08, 2020. [Online]. Available: https://www.bundesliga.com/en/bundesliga/news/soccer-positions-explained-names-numbers-what-they-do-2579–786

[71] J. Wood, P. Isenberg, T. Isenberg, J. Dykes, N. Boukhelifa, and A. Slingsby, "Sketchy rendering for information visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2749–2758, Dec. 2012.

[72] PostgreSQL Global Development Group, "PostgreSQL," 2020. Last Accessed: Dec. 08, 2020. [Online]. Available: https://www.postgresql.org/

[73] A. Ronacher, "Flask," 2020. Last Accessed: Dec. 08, 2020. [Online]. Available: https://palletsprojects.com/p/flask/

[74] Angular Team at Google, "Angular," 2020. Last Accessed: Dec. 08, 2020. [Online]. Available: https://angular.io/

[75] M. Bostock, V. Ogievetsky, and J. Heer, "D^3 data-driven documents," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2301–2309, Dec. 2011.

[76] P. Shihn, "rough.js," 2020. Last Accessed: Dec. 08, 2020. [Online]. Available: https://roughjs.com/

[77] T. Boren and J. Ramey, "Thinking aloud: Reconciling theory and practice," *IEEE Prof. Commun.*, vol. 43, no. 3, pp. 261–278, Sep. 2000.

[78] K. A. Ericsson and H. A. Simon, *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA, USA: MIT Press, 1984.

[79] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–446, 2002.

Daniel Seebacher is currently working toward the PhD degree with the Data Analysis and Visualization Group of Daniel Keim, University of Konstanz, Konstanz, Germany, since 2016. His research interests include the visual analysis of spatio-temporal events and their context. Concrete research domains include the analysis of the spread of invasive species, the study of intra-city meteorological phenomena, or the study of actions and interactions in sports such as soccer or tennis.

Tom Polk is currently a postdoctoral researcher with the Data Analysis and Visualization Group of Daniel Keim, University of Konstanz. His research interests center on visual analytics in applied domains utilizing spatio-temporal data. His current research efforts focus on tennis analytics with the goal of player improvement.

Halldor Janetzko received the MSc and PhD degrees in computer science from the University of Konstanz, Konstanz, Germany, in 2010 and 2015, respectively. He is a professor for computer science with the Lucerne University of Applied Sciences and Arts. His research interests include temporal, geospatial, and multi-dimensional phenomena with a special focus on visual data analytics.

Daniel A. Keim received the PhD and habilitation degrees in computer science from the University of Munich, Munich, Germany. He is currently a full professor and head of the Data Analysis and Visualization Research Group, University of Konstanz's Computer Science Department. His research combines automated data analysis and interactive visualization techniques for the purpose of exploring large data in various applications domains.

Tobias Schreck received the PhD degree in computer science from the University of Konstanz, Konstanz, Germany, in 2006. He is currently a professor and head with the Institute of Computer Graphics and Knowledge Visualization, Graz University of Technology. He previously was assistant professor with the University of Konstanz, Germany, and postdoctoral fellow with Technische Universität Darmstadt. His works are in the areas of visual analytics, information visualization, and applied 3D object retrieval.

Manuel Stein received the PhD degree from the University of Konstanz, Konstanz, Germany. He is CEO of the Subsequent GmbH, a novel company for real-time skeleton data extraction and innovative movement analysis. His research concentrates on the analysis of player movement with the goal of understanding cooperative and competitive behavior in team sports. He was invited keynote-speaker and chapter-author of the Barça Innovation Hub as well as mentor and keynote-speaker of the German Football Association DFB.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.