

# Visual-Interactive Search for Soccer Trajectories to Identify Interesting Game Situations

*Lin Shao<sup>1</sup>, Dominik Sacha<sup>1</sup>, Benjamin Neldner<sup>1</sup>, Manuel Stein<sup>1</sup> and Tobias Schreck<sup>2</sup>*

<sup>1</sup>*University of Konstanz, Germany*

<sup>2</sup>*Graz University of Technology, Austria*

## Abstract

*Recently, sports analytics has turned into an important research area of visual analytics and may provide interesting findings, such as the best player of the season, for various kinds of sports. Soccer is a very popular and tactical game, which also attracted great attention in the last few years. However, the search for complex game movements is a very crucial and challenging task. We present a system for searching trajectory data in soccer matches by means of an interactive search interface that enables the user to sketch a situation of interest. Furthermore, we apply a domain specific prefiltering process to extract a set of local movement segments, which are similar to a given sketch. Our approach comprises single-trajectory, multi-trajectory, and event-specific search functions based on two different similarity measures. To demonstrate the usefulness of our approach, we define a domain specific task analysis and conduct a case study together with a domain expert from FC Bayern München by investigating a real-world soccer match. Finally, we show that multi-trajectory search in combination with event-specific filtering is needed to describe and retrieve complex moves in soccer matches.*

## Introduction

Nowadays, most mobile devices, such as smartphones, smart-watches, or tablets, are supporting Global Positioning System (GPS) tracking functions, and enable the possibility to let users log their activities during sport or leisure time. Often, spatial movement data is captured, giving rise to trajectory data. Due to the rapid development of GPS tracking devices and the continuous growth of data repositories that provide such trajectories, more and more application areas in geospatial data analysis emerge. For example, one upcoming research field is trajectory search for sports analytics. In recent years, there has been increasing interest in methods for analyzing statistics and tactical game plays in sports. Therefore, players get equipped with GPS sensors or are tracked by high-resolution cameras to extract the raw movement data. Also, plenty of approaches exist, which analyze whole matches, individual player statistics, or provide hypothesis prospects for sports, such as tennis, basketball or soccer [9, 12, 19].

In the field of soccer analytics, most of the existing approaches focus on statistical feature extraction, like speed, acceleration or running distance, which are visualized by a variety of techniques to discover insights in matches [12]. The drawback of these approaches is that they only utilize the statistical information and manually annotated events (e.g., goals, free kicks, fouls, etc.) to analyze interesting game situations. However, in a soccer game there may exist many different and important match-winning

phases, which are time-consuming and expensive for manual annotations. One unexplored research field is the direct trajectory search for untagged movement patterns, such as individual player movements, tactical team movements or rehearsed ball passing paths in matches. The search of such trajectories may help coaches, managers, scouts, and other decision makers to find crucial game situations, and thus analyze the performance of individual player or the whole team. The search complexity of these large movement data is challenging and also referring to general geospatial analysis problems, since we are using interdependent trajectories on a small limited area. Soccer is a very tactical and strategic game, and thus the search space can quickly become overcrowded by 23 simultaneously interacting trajectories of players and the ball.

In this paper, we present an approach for sketch-based visual search and exploration of soccer trajectories to discover crucial game situations based on single-player, multi-player and event-specific trajectories. Therefore, we apply two different similarity search approaches based on spatio-temporal point distributions and movement directions of a trajectory. For the purpose of comparing locally similar trajectory segments, we introduce a domain specific filtering process, which extracts relevant trajectory segments. Our search interface is inspired by tactical drawings from soccer coaches and should reflect a search modality closely related to coaching and analysis practice in this domain. Furthermore, we design a task analysis for movement search in the context of soccer analytics and specify the needs for an easy and intuitive search interface. We demonstrate the usefulness of our approach by conducting a case study together with a professional youth coach of one of the top soccer clubs, FC Bayern München.

## Related Work

In the following, we discuss a selection of related work in the context of our approach. Three main research areas are sketch-based search, movement analysis and general visual analytics of soccer data.

## Sketch-based Search Techniques

In the image and video retrieval domain, sketch-based search has developed into an important retrieval technique over the past few years. It allows the user to search for specific content of a video or image, and deals with the basic problem of the textual search that merely searches for tagged annotations, which are related to the video or image. Recently, a lot of research has been done in this area. In [6, 7], novel approaches for indexing and evaluating sketch-based systems were proposed. A recent work in sports analytics [15] uses sketch-based search techniques to analyze player

movements in rugby matches. It uses multiple distance-based similarity measures, which compares the user sketch against video scenes. Beside the traditional image and video search, sketch-based search techniques can also be applied to other complex data sets. In [8, 13], 2D sketches are used to search for 3D objects. Lee et al. [14] make use of synthesized background shadows to help users in sketching objects. Further applications of sketch-based search include retrieval in graphs [27], or for bivariate data patterns in scatter plots [22, 24]. Regarding data generation, in [28] 2D shape sketches are used to create and manipulate high-dimensional data spaces.

### Movement Analysis

Due to the growth of mobile computing and GPS supported devices, the analysis of trajectories has become an interesting task. For instance, by using visual analytics tools and mining trajectories of animals, it is possible to discover movement and migration patterns of certain groups of animals [25, 26]. In [23], textual descriptions are combined with trajectory data of a traveler for trip planning and recommendations. Nowadays, many applications for visual analytics of movement data exists, including the analysis of vessel movements [30], traffic jams [29], or sports data [15, 32, 33]. Moreover, a survey of existing visual analytics approaches concerning movement data is given in [1, 3].

Important data analysis methods in this area include the similarity search, segmentation, or clustering of trajectories. Recent research on trajectory search approaches focus on optimizing the retrieval process, which is an interesting problem in itself, but beyond the scope of this paper.

### Soccer Analytics

Recently, knowledge discovery methods have successfully been applied to analyze large amounts of soccer and sports data. A recent approach was introduced by Bialkowski et al. [4], in which the authors detect player formations within a whole match including potential positional interchange. The authors approach this by applying a role-based representation that dynamically updates each player's relative role at each frame. Another recent work on analysis in sport has been presented by Lucey et al. [17], where a novel technique is presented to estimate the probability of a successful chance in soccer. The authors discovered several factors, which contribute to the probability of a chance. These factors were determined and evaluated by analyzing a whole season of player and ball tracking data. In another related work by Janetzko et al. [12], a system for analyzing high-frequency position-based soccer data at various levels of detail covering player and event-based analytical views is presented. Among other things the authors performed single and multiple player analysis, for example, for the detection of player phases. Furthermore, constellations and formations were analyzed to, for example, evaluate the performance of the back four formation. Another system for the analysis of soccer data has been introduced by Perin et al. [18]. They developed a tool combining different perspectives on soccer data after segmenting the data into meaningful units. The segmented units can be analyzed in different visualizations. Other presented systems focus on the detection of events of interest. For instance, Rathod et al. [20] retrieve these events by a fuzzy inference system with an input of features and grass percentages extracted from key frames. Xiong et al. [31] extend this idea by extracting audio sources, such as

applause. Other mentionable approaches for analyzing soccer players and teams have been presented by Gudmundsson et al. [10, 11]. They presented several approaches based on position data to extract basic events, such as kick-offs, corner kicks or throw ins, and applied a cluster analysis on single player's subtrajectories to identify frequent movements during the match.

### Sketch-based Search for Soccer Trajectories

In this section, we describe the relevant search space and tasks for the soccer domain. In addition, this section defines the scope of the work presented in this paper.

#### Soccer Search Space

In order to describe the sketch-based search space, we make use of the types of movement analysis tasks proposed by Andrienko et al. [1]. Movement analysis can be described using four different foci: Movers  $M$ , spatial events  $E$ , space  $S$ , and time  $T$ . With that respect, in our soccer case movers are all moving (spatio-temporal) objects on the soccer pitch ( $M \in \{Player, Ball\}$ ), spatial events can be described as movement or rule based events ( $E \in \{Movement, Rule\}$ ), space can be referred to moving areas of interest, such as empty/free or occupied areas, or fixed areas such as dangerous areas near a goal ( $S \in \{Moving, Fixed\}$ ). Finally, time is related as the temporal constraints and units given by a soccer game (e.g., half-time) or the duration of soccer moves ( $T \in \{Unit, Duration\}$ ). In addition, especially in soccer there are several further dimensions that can be combined with the focal set ( $\{M, E, S, T\}$ ) for searching. In order to describe the soccer search space as a whole we propose the following dimensions:

**D1-What to Search:** All elements of the focal set ( $\{M, E, S, T\}$ ) and their combination and the movements attributes (e.g., speed).

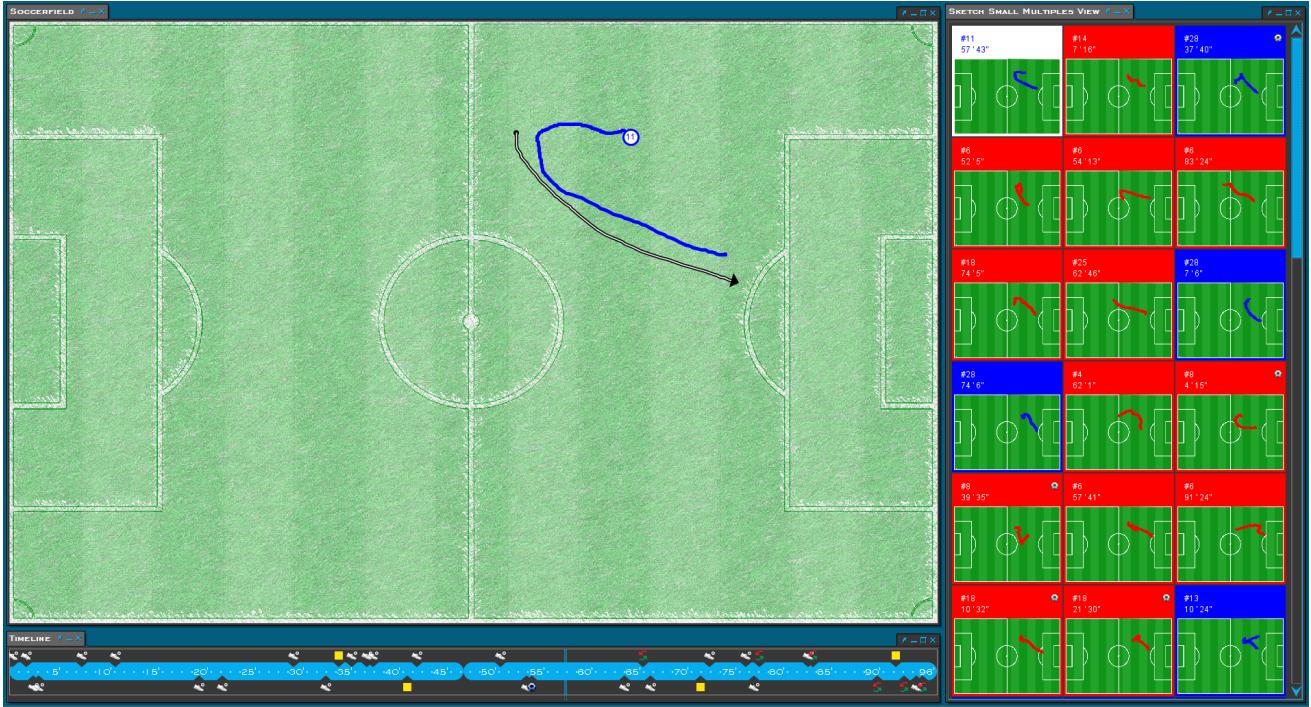
**D2-Invariances:** Having specified a spatial query this can be *translated*, *scaled*, or *rotated* in order to search for similar elements.

**D3-Filtering/Constraints:** Filtering can be done by adding constraints to all elements of the focal set. In soccer, relevant constraints are temporal (e.g., only first half), event-specific (e.g., only shot-events), or object-attribute-specific (e.g., tactical position of a player or only movements including high speed values), and spatial (e.g., only in a specific region). Furthermore, these constraints can be applied for specific or all elements (specific vs. global).

**D4-Cardinality:** Furthermore, we distinguish if the analyst is seeking for one or several elements as well as their temporal order. Synchronous searching refers to a search where all elements occur together at the same time, whereas asynchronous or sequence searching considers time shifts between the objects.

#### Our Scope

**Search Space:** Our approach is inspired by board sketches typically known from soccer coaches or analysts. Therefore, our approach enables the analyst to intuitively define a query object directly on the soccer pitch by applying a sketch of the desired trajectory at the desired place. That is why our search space is limited to moving objects  $M$  at first (D1). In this case, sketching is the creation of a spatial query object [2]. Furthermore, our similarity search is designed to find the exact position, rotation, and scaling of the trajectory (D2). The filtering and constraints can be added to the similarity search for individual trajectories by specifying the occurrence of specific movers, events or time intervals (D3).



**Figure 1.** Our implemented prototype showing a single-trajectory search by applying the spatial distribution approach presented in Section 4. In this illustration, we are searching for movement patterns that start from the outer midfield and run into the direction of right penalty area. The best matching results are shown in the small multiple view on the right hand side. By clicking on a thumbnail (highlighted in white), the particular trajectory will be directly displayed on the soccer pitch. Additional information about the scene like jersey number, player name, time or trajectory type (displayed in different colors) are shown in the small multiple thumbnail.

Finally, the similarity search can iteratively be refined by adding further sketched trajectories. The similarity search for these multiple spatial query objects is designed for synchronous occurrence that identifies situations with all the desired movements happening within a short time span (*D4*).

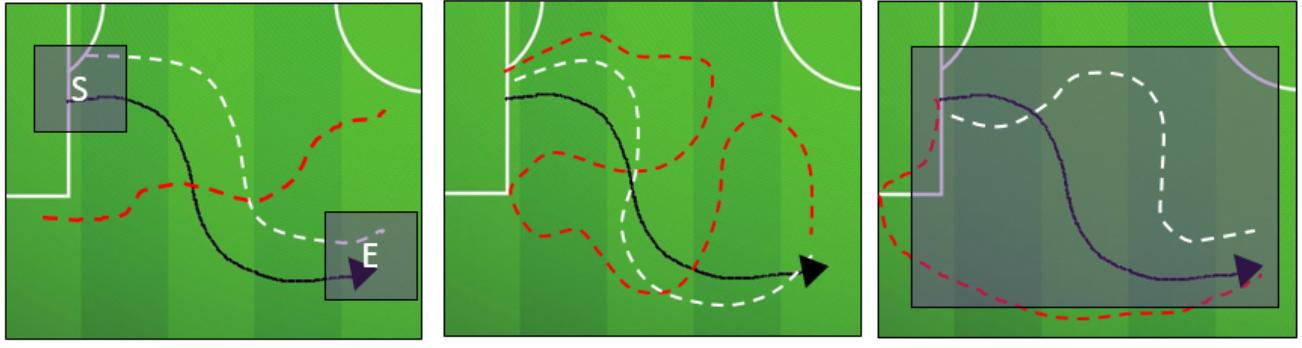
**Tasks:** Our approach is designed to support typical *relational-seeking tasks* where “items that are related in a specified way need to be detected” [2]. In our case the system presents the user trajectories that are “equal” or “similar” to the query object. Furthermore, our approach enables the analyst to inspect the result set and therefore to perform further *lookup* or *comparison* tasks. Our design is guided to support the following two higher level analysis tasks: (1) searching a specific movement/situation that happened at a specific time, and (2) identifying, comparing and relating a set of similar movements. Concrete examples for these higher level tasks are given in the next subsection. In order to support these tasks, we propose the following analysis workflow. First, the analyst creates a query object (trajectory sketch). Then the analyst may inspect the result set and to further refine the analysis he or she can add additional filters, constraints or trajectory sketches and even remove old ones. This way every search can be refined, improved and narrowed down to the specific task that is needed. Furthermore, the analyst will be able to switch between two different distance functions that are adapted for our two analysis tasks (see Section 5). Finally, in each step the analyst can inspect the situation of interest with additional visualizations and animation (by showing all players, ball trajectory, heatmaps of player movements, etc.).

## Domain Tasks

This subsection lists some soccer search tasks that emerged during our discussions, observations and requirements analysis with the domain expert. In most cases, the first step for an analyst is to sketch the ball trajectory (or a player trajectory where the player is in ball possession) because the ball movement determines all the other movements. As a second step further filters and trajectories are added in order to characterize specific situations. Analysts prefer to search for attack movements instead of defense situations, as the domain expert explains, in attacks trajectories and the players’ movements are characteristic and of interest, whereas in defense behavior spaces or groups are more crucial. That is why our approach focuses on searching for attack situations. We are able to distinguish two kinds of attack situations (where a specific team is in ball possession): *Middle*, or *Side* attacks.

**Middle Attacks:** Middle attacks are situations in which a player in the middle receives the ball in order to control it (typically done by a robust forward player) or situations where the ball is passed in the back of the opponents’ defense line (also known as a “through ball”).

**Side Attacks:** Side attacks can be distinguished between attacks where the player tries to reach the baseline in order to pass (or cross) the ball in the penalty area of the opponents or attacks where the players try to reach the goal directly from the wings. A third class of side attacks cover crosses that are impacted from the half court towards the goal. The identification of these attack types determine the domain tasks. Further, the domain expert describes *Preparation Phases* for attacks.



(a) Similar start and end position.

(b) Similar trajectory length.

(c) Similar marginal restriction.

**Figure 2.** Demonstration of our preprocessing steps to filter out dissimilar trajectory segments (illustrated in red). First, we only consider trajectory segments, which are running through a similar start and end area of the user sketch (a); then we compare the length of user sketch and remaining candidates (b) and finally, we are using a bounding box to check for deviations within their path (c). The black arrows illustrate the user sketches and the trajectories that fulfill the filtering constraints are colored in white.

**Preparation Phases:** Preparation phases are situations in which the team in ball possession traverses and switches the ball from one side to another, zig-zagging towards to the opponents' goal, forcing the opposing team to open up and produce free spaces. However, the preparation is an optional phase that can be included in an attack followed by a finish (e.g., shot on goal).

A concrete domain task for the identification of a specific situation could be that the coach wants to find an attack over the right hand side at the beginning of the match where the side player reaches the baseline and passes the ball to his forward who then performs a direct shot on target. Another concrete domain task could be to identify how many attacks have been performed via the left hand side and thus derive new tactics or analyze strengths and weaknesses of the opponent or own team. This query could be refined by adding sketches of the forwards that run into the penalty area in order to receive a cross. Further examples and solutions to solve these tasks are illustrated in Section 5.

## System

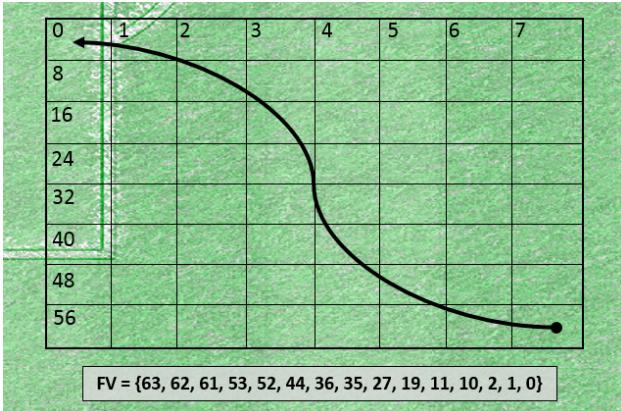
In this section, we describe our sketch-based search techniques to find similar movements and provide an overview of the developed components of our system. The focus of our work is on the development of an easy and intuitive search technique for finding complex movements in soccer matches. Our prototype system for discovering interesting game situations is depicted in Figure 1.

### Movement Search

Our approach to search for movement patterns is based on a three-step process including trajectory filtering, feature extraction and similarity measure. Since a soccer match lasts at least 90 minutes and involves 23 individual trajectories of  $M \in \{Player, Ball\}$ , the search space can quickly become too large for manual search and hinder the search process for a particular movement pattern. Furthermore, a trajectory over 90 minutes can be very long and has to be investigated in terms of their local similarity. We therefore use a preprocessing step to eliminate trajectories that have no locally similar movements to a given user search based on a

coarse filtering step. The preprocessing step comprises three further filter constraints, which are depicted in Figure 2. In the first step, all trajectories will be checked if they contain a local path that runs through a similar start and end position as the user sketch. After that, the path length of all remaining trajectory segments are compared to the length of the drawn sketch. All segments which exceed or do not fulfill a certain threshold length will be eliminated. The last filter constraint refers to the overall spatial similarity of the trajectory area. If a trajectory segment exceeds a marginal area of the user sketch it will also be filtered out. These filtering steps utilize the original geospatial positions on the soccer pitch and have the benefit that the computation is very fast and does not need any trajectory segmentation procedures in advance. It quickly returns a rough set of similar local paths for the further search process. Another advantage against a pre-segmentation is that our approach is capable of searching trajectories without interruptions and thus even player movements during breaks (e.g., throw-in) can be found. The parameter settings for the three filtering thresholds influence the resulting number of similar trajectories and can be adjusted by the user. We propose to use at least a tolerance range of 20% on the user sketch length and a range of 10% on the sketch location (bounding boxes), to find enough candidates for the similarity ranking.

In order to identify the most similar movement patterns and to rank the result set, we are using a feature-based similarity function. Hence, we need to describe the characteristics of the not filtered segments as well as the user sketch by a suitable feature vector. Therefore, we provide two different descriptors based on the spatial distribution and the movement directions of a trajectory, each tailored to one of our two analysis tasks from Section 3. The descriptor for measuring the spatial distribution is used to accomplish task 1 - searching a specific movement/situation that happened in the game. For this task, we assume that the analyst watched a match and wants to retrieve a specific situation. In this case, the spatial orientation and the trajectory path are the most important feature that describes the happened scene (e.g., corner kick). Our descriptor uses a flexible grid approach, which adapts the grid size according to the user sketch size on the soccer pitch and thus limits the space of interest ( $S \in \{Moving, Fixed\}$ ). This



**Figure 3.** Composition of a feature vector by using our grid-based approach (spatial distribution). The trajectory starts at the bottom right corner (index 63) and traverses the grid towards the upper left corner (index 0).

ensures that small sketches (e.g., a dribble action on a specific area) will be considered more precisely than a rough sketch that runs over the whole soccer pitch (e.g., a long diagonal pass). The benefit of this flexible grid approach is that imprecise sketches on the right place are sufficient to detect a existing situation that happened in the game. By default, we superimpose an  $8 \times 8$  grid on the user sketch and apply the same grid size to the comparing trajectories. Each cell corresponds to an unique index and the feature vector is composed of the particular indices where the trajectories pass through. Figure 3 demonstrates the composition of a feature vector.

The second descriptor focuses on the characterization of the general structure and is therefore better suited for our second analysis task - identifying and comparing similar movements. In contrast to the first descriptor, our aim here is to ensure that similar trajectories with slight shifts are rated significantly better than dissimilar trajectories, which pass the same cells. This is especially relevant when it comes to complex analysis tasks that require more generic queries for searching typical movements or novel tactics of the opponents. To this end, we defined eight different direction types ( $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$ ) to describe the structure of a given trajectory. The direction types are encoded as numerical values and the feature vector represents the direction types from start to the end position of the user sketch. This has the advantage that we can prioritize the similarity search by the characteristics of a movement ( $M \in \{Player, Ball\}$ ). Finally, the last step is to compute the distance between the feature vectors in order to rank the trajectories. Here we use an edit distance metric [16] for measuring the minimum distance between the user sketch and the trajectory candidates.

This applies to a one-to-one movement search (single-trajectory), but in the event of a multi-trajectory search, we perform individual queries for each drawn sketch and aggregate the computed distances for the overall similarity. Furthermore, we verify whether the matched trajectories correspond to a joint action with respect to the time. This means that we are only searching for movements that took place during the same period of time. At this point, we are considering two different temporal constraints to define synchronously occurring movements. The analyst can choose between a chronological sequence and a simultaneous occurrence

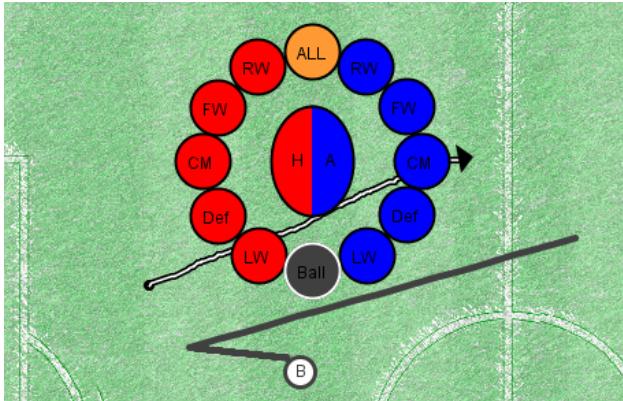
of trajectories (D4). For sketching general attack situations like middle attacks or side attacks, we recommend to use the search for simultaneous occurring trajectories, which is more flexible and only considers whether the found movements occur within a short time span. In contrast, to fulfill the constraints of the chronological sequence, the found movements have to occur within a short time span and also in the same order as the drawn trajectories. Thus, this search setting can be used to sketch situations including after-effects (e.g., a counterattack after a failed side attack). Eventually, we aggregate the distances of the trajectories that belong together and calculate the overall similarity for ranking.

A last important aspect that needs to be considered is the temporal constraint of the players positioning, which depends on the two half-times of the game. By that we mean that players will change their positions after the half-time break, and thus makes the search for spatial queries of individual players more challenging. In order to perform spatial- and player-dependent queries, we provide a position mirroring function that maps the player and ball of the second half formation to their origin (first half positioning). Depending on the analysis task, this feature can be enabled or disabled.

### Interaction

For searching soccer movement trajectories, we provide a sketch-based interface that allows analysts to sketch their queries. Since soccer coaches often tend to sketch tactical movements on boards, we utilize an interactive soccer pitch that enables users to directly sketch movements of  $M \in \{Player, Ball\}$  on the field (see Figure 1). This helps the user to precisely draw spatial movement queries such as middle attacks or side attacks. Alternatively, a query can also be created by sketching two rectangular areas that describes the start and end position of a trajectory, as shown in Figure 8 and 9. The benefit of this alternative trajectory search is that more general queries may be drawn and a greater proportion of deviating but also similar directed trajectories may be found. The rectangular areas will replace the spatial constraints of start and end area in our preprocessing steps (cf. Section 4.1) and thus increase the number of potentially interesting movements. The original path of the query runs from center to center of both drawn rectangles and will be added automatically to the search. All sketching interactions are implemented with mouse-events and thus allow query drawings by mouse, digital pens or even fingers on touch tables.

Further filters are used to explore and navigate the result set (D3). To indicate a drawn sketch as a specific movement type of  $M \in \{Player, Ball\}$ , we equipped the search interface with a radial layout menu, which shortens the distance and time to navigate to an item [21]. By means of this menu users can filter for trajectories of a particular team (home, away), player groupings (defender, center / midfielder, striker, right winger, left winger), or the ball. Figure 4 shows the radial menu and a chosen filter constraint on ball trajectories. In order to select a certain player, we also provide an tabular view of all players in a additional option panel. For an event specific search of  $E \in \{Movement, Rule\}$  the query can also be connected to a particular event, such as shot, corner kick, free kick, cross, or chance. This ensures that the found scenes are initiated or produce such an event in a short time. For instance, drawn side attacks may be filtered for shot events to receive only the completed attempts. Moreover, by selecting a time range



**Figure 4.** Our radial menu for filtering trajectories based on team, player groupings and the ball. By selecting a particular filter criterion the result set gets updated and the best ranked trajectory will be displayed on the soccer pitch. This view demonstrates the selection of the ball trajectory.

$T \in \{Unit, Duration\}$  on the timeline, which is placed under the soccer pitch (see Figure 1), results can also be restricted by time. Consequently, analysts may search for situations that happened in the second half.

### Visual Representation

When a query has been performed, the most similar trajectory will be displayed on the soccer pitch next to the user sketch (as shown in Figure 1). The found trajectories are colored according to the trajectory type. Red trajectories refer to movements of the home team, whereas blue trajectories refer to the away team and gray trajectories to the ball. We indicate the movement direction by showing a circular starting position, which also reveals additional information about the trajectory, like jersey number, or a ball label.

The temporal information ( $T \in \{Unit, Duration\}$ ) of the found scene can also be highlighted as interval on the timeline (illustrated in Figure 1). This allows analysts to jump to the exact game phase and analyze the found situation. For browsing the remaining result set a small multiple view is used that represents the same information for all individual trajectories. The visual encoding of the thumbnails in the small multiple view is as follows: The background and trajectory color corresponds to the trajectory type (home team, away team or ball) and additional information like player name, jersey number and time are shown on top. In case of a multi-trajectory search, the background color encodes the percentage of ball possession of the respective time interval, since several sketches of different teams could be drawn. Our small multiple view is shown on the right hand side of Figure 1. By clicking on a particular thumbnail, the trajectories will be enlarged, displayed on the soccer pitch and the selected time interval will be updated. This helps the user to accomplish our tasks from Section 3.

### Application and Evaluation

In this section, we demonstrate the usefulness of our search system by performing our domain tasks that we defined in Section 3. To evaluate both of our descriptors, we structured the domain tasks into two similarity areas: (1) searching for the exact spatial position of a movement and (2) searching for the main

directions of the movement. The search tasks for the spatial distribution descriptor focuses on searching *specific situations* (1) and the direction-based descriptor on identifying/comparing *similar movements* (2) in a match.

Therefore, we used professional soccer match data provided within a collaboration with the sports analytics provider Prozone<sup>1</sup>. The data set is not publicly available and has to be anonymized as it was a professional game. For each player of both teams two-dimensional position data are available with a temporal resolution of 100 milliseconds. The data are enhanced by manually user annotated events containing information about position, time, involved players and event-specific details. We designed a qualitative case study to evaluate our approach together with a soccer expert. The invited expert has been an active soccer player since 24 years and has been working as a coach since 10 years; he is currently employed by the German soccer club FC Bayern München. The evaluation was performed on a notebook, connected to a 24 inch Full HD monitor, where the system was running. The program was executed in full screen mode. The domain expert was asked to express his thoughts and opinions while using the system, according to the *Thinking Aloud* [5] method. During the evaluation we took notes of comments and own impressions for subsequent analysis.

### Task 1: Specific Situations

A very important task in soccer analysis is to retrieve situations of interest that are already known to the user. This might be important when the analyst identified an interesting situation during the match and later wants to re-watch it to gain knowledge. Hence it is important, to retrieve movement patterns based on exact spatial position on the soccer pitch. The task in this section is designed to retrieve a match situation with our system after watching the video recording. Therefore, we investigated a professional soccer match from a first-class league in Europe. We selected a number of match highlights from a video summary and tried to retrieve the respective scene together with our domain expert. The selected scenes were:

#### Scene 1: Side Attack (Corner Kick)

The first highlight of the match was a corner kick in 1:30 minute of the game that also leads to the first opportunity for the home team (colored in red). As our domain expert recommend to start sketching with ball trajectories and then include additional movement trajectories to specify the query, we start the search by sketching a trajectory from the respective corner position into the penalty area. The sketch and corresponding results are shown in Figure 5. After the initial query, we found the only two corner kick situations (from the respective position) in the game and two player movements with similar trajectories. Since a total of only four trajectories have passed this way, all other trajectories were filtered out by our prefiltering conditions (cf. Section 4). The searched corner kick scene at 1:30 minute can be found in second place (highlighted in white - starting a little bit earlier). In the event of a larger result set the matched trajectories can also be restricted in terms of time (e.g. only first half) or moving object (e.g. only ball) to detect such set-plays (D3).

#### Scene 2: Middle Attack

<sup>1</sup><http://www.prozonesports.com/>



(a) Query for Scene 1. (b) Result set of query 4 (a).

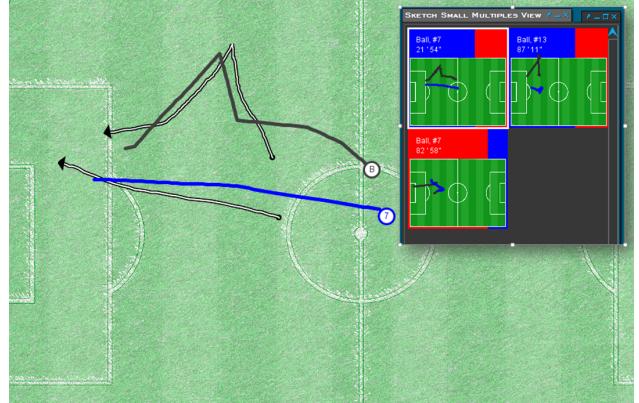
**Figure 5.** User sketch and the corresponding results for Scene 1. In total the search system found four similar trajectory paths including two corner kick situation. All found trajectories start at the upper right corner and go through the penalty area.

The second scene depicts a more complex move from the blue team in 22:00 minute, which includes a rehearsed ball passing and multiple player shifts to create an opportunity on goal. The selected scene shows a counter attack that starts in the midfield of the soccer pitch by controlling and distributing the ball to perform the "through ball" (a pass in the back of the opponents' defense line). Therefore, the central player passes the ball to the wing player on the right hand side and runs to the right baseline in order to create a new "open man" (undefended player) and opens the space for a secondary striker at the same time. To retrieve such a scene, an advanced query with additional information about secondary trajectories and trajectory types are needed. We roughly sketched the whole ball passing path and a second running path of the striker that runs into the open space and has the opportunity on goal. To specify a trajectory type the sketches can be additional assigned by our radial menu (see Figure 4).

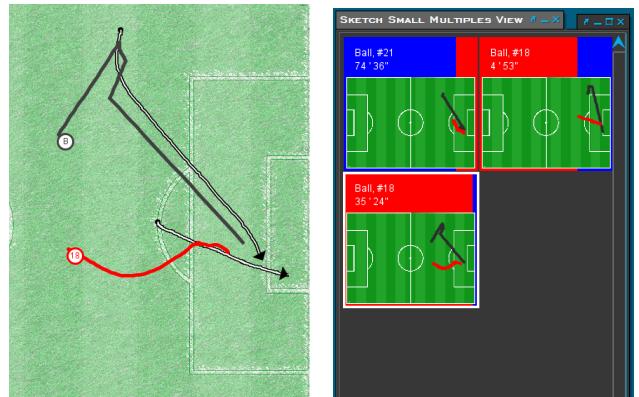
Figure 6 illustrates the two drawn user sketches and the best matching results of our spatial distribution descriptors in a small multiple view. This time, we explicitly indicate the first drawn trajectory as ball trajectory (upper sketch) and the second (lower sketch) as any player trajectory of the away team (blue) to keep the results clear and reasonable. Without specifying the types of both trajectories the result set would contain more than hundreds of possible trajectory combinations that happened during the match. By means of our filter constraints, we guarantee that all matching results include a zigzag kind pattern of the ball trajectory and a straight running path of any player of the away team that is located below. The result set reveals that such a situation has occurred three times during the match, namely at 21:54, 82:59 and 87:11 minute of the match. Moreover, we rediscover the trajectories of the desired scene and ranked the situation on first place.

### Scene 3: Side Attack

The third and last scene that we want to retrieve is the last chance



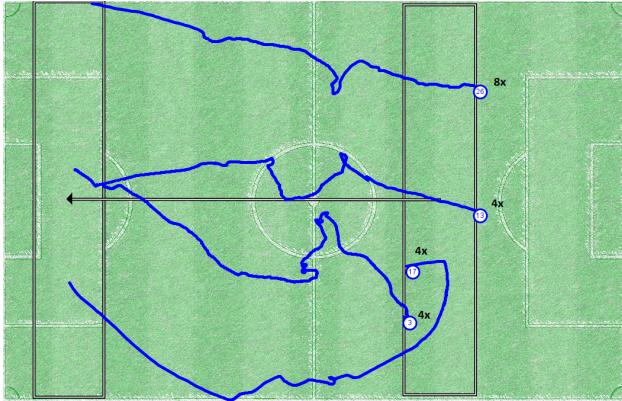
**Figure 6.** Illustration of a multi-trajectory query for Scene 2 including the best matching result (visualized on soccer pitch) and further results in the small multiple view. In order to rediscover the situation we determined the first sketch (upper arrow) as ball trajectory and the sketch arranged below as player trajectory of the blue team (away team).



(a) Query for Scene 3. (b) Result set of sketch 6 (a).

**Figure 7.** Multi-trajectory sketch and the corresponding results for Scene 3. The sketched trajectories are determined as ball (upper sketch) and player (lower sketch) trajectory (a). The background color of the small multiple thumbnails reveals additional information about the game phase (b).

for the home team before the half-time break in 35:30 minute. In this scene a player of the red team is crossing the ball from the upper outside area into the penalty area while a teammate and striker is running into the box to complete the attack. To sketch this query, we again start with sketching the ball trajectory from corresponding position into the penalty area and include another trajectory that indicates the running path of a player of the home team. This time, we again found the desired scene (ranked on third place) and two other very similar situations with homogeneous movements of ball and player, shown in Figure 7. This might be a rehearsed movement of the team that occurs three time during the game, but on closer inspection the small multiple thumbnail reveals that the situation ranked on first place happened in the second half of the game and the ball possession of the blue team is more than 80% (background color) at this situation. Consequently, the situation ranked on first place is a defensive play of the home team, which involves similar movement patters like the attempted attacks at 4:53 minute and 35:24 minute.



**Figure 8.** Overview of the offensive attempts of defenders from the away team. The two rectangular boxes determine the start and end position of the searched trajectories. This query reveals that the right defender participated eight times in offensive situations.

### Task 2: Similar Situations

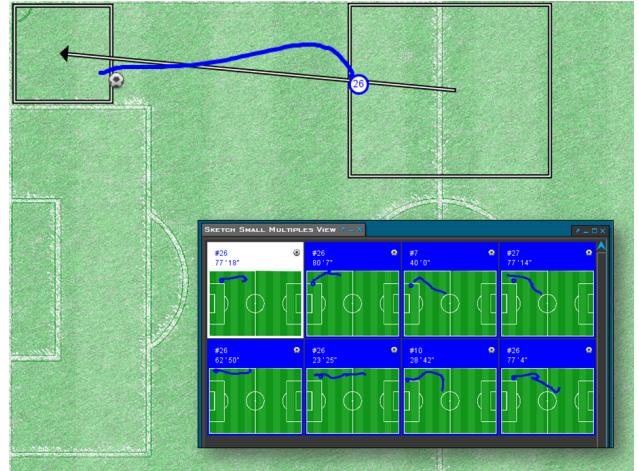
Another task is to identify and explore several similar situations with respect to a given search pattern. This is important when an analyst wants to find out, for example, how upcoming opponents structure their offensive play. In this respect, it is also important to formalize a more generic query that covers typical movements and identifies novel tactics of the opponents, which are not yet known. Hence, for this task, we apply our position mirroring function that maps the playing direction of both teams into a uniform direction (i.e., no changeover after the half-time break) and include our spatial area sketches combined with the direction-based descriptor to make the search more flexible. For the experiment, we again investigated the same match from Task 1 together with our invited expert and let him answer a few predefined questions by using our tool. Our analysis questions were:

#### Question 1: Which defender of the away team participates in offensive situations at most?

By means of the mirroring function, the players of the away team are always playing from the right to the left side and spatial queries can be performed on both half-times simultaneously. In order to answer this question, the expert sketched two rectangular areas that describe start and end points of a trajectory. Moreover, he limited the search space by selecting only the defender of the away teams in our radial menu. An illustration of the result is presented in Figure 8. We found in total twenty movements of the away team defenders that start in their own half of the court and run into the danger zone (near opponents goal). The three defenders with the jersey number 3, 13 and 17 participated four times in offensive situations whereas right defender with the jersey number 26 (the defender who puts in the most effort in offensive play) attempts eight times. Also interesting to consider are the simultaneous attempts of the defenders. For instance, in 1:56 minute all four defenders are participated in offensive play, which also led to an opportunity for the away team.

#### Question 2: How often did the away team perform side attacks over the right hand side?

We assume that this question is addressed to side attacks where players try to reach the baseline and cross the ball into the penalty area. Figure 9 shows the suggested query from our domain ex-



**Figure 9.** Spatial area sketch to identify the side attacks from the away team over the right hand side. By including the event information it only returns real attack situations that involves a cross.

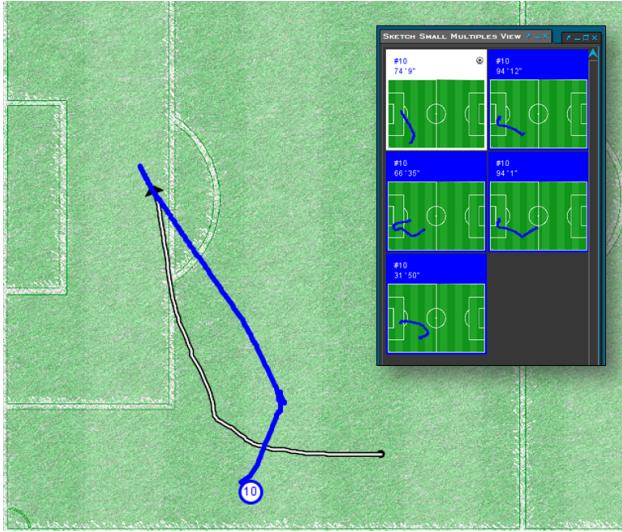
pert and the final result set of found side attacks. After applying the shown query, our visual search system identifies 21 trajectories that have a similar main direction, and similar start- and end points. However, the problem is that also defensive movements or movements without any ball possessions were found, which are not related to side attacks. To prevent such search results, we include another filtering constraint on the rule based events ( $E \in \{Movement, Rule\}$ ) to receive only movements, which contain a cross event. These results are marked with a ball symbol in the small multiple view and show the respective position on the soccer pitch where the cross happened. Finally, we found eight side attacks from the right hand side, as shown in Figure 9.

#### Question 3: How often has the right striker from the away team attacked over the left wing?

The right striker from the away team is a typical center forward player whose main function is to receive passes, win long balls and score goals. In the interest of receiving through balls these forward players have to continuously change their running paths and try to find open spaces at the right moment. Hence, it is appreciated when several forward players change their positions over time to create new attack situations and confuse the opponents defenders. However, our analysis task is now to find out how many times he changed the sides and attacked from the left wing side. Therefore, our expert sketched a typical left wing attack pattern that runs from the left wing into the penalty area (curved movement on the border) and specified this as a movement of the right striker. The small multiple view in Figure 10 reveals that such an situation occurs six times in the match and indicates that the player is following this scheme.

## Discussion

We presented a system that is capable of finding selected scenes from the video summary and can be used to answer analytical questions about the performance of individual players, team groupings or even the whole team. Moreover, we realized that a basic single-trajectory search is insufficient for complex movements in soccer data, and additional information about secondary trajec-



**Figure 10.** Search for a specific movement pattern of the right striker from the away team. It reveals that the player attacks the opponent six times from the left wing side.

tories, trajectory types, and associated events are needed. For Task 1, we might not always find the desired scene on first try, but this is due to the fact that we focused on retrieving the scenes by rough sketches, and hence did not optimize our user sketches. Of course, we could adapt our sketches to the original trajectories in order to get the best ranking, but that was not our aim and we decided to use the best results that we achieved after a few trials. Our aim was to develop an interactive search system that can be quickly and easily used. In our experiments, it was usually sufficient to perform a trajectory search with up to two movements including filter constraints. However, we also tested the performance of our system with a higher amount of trajectories and reached a average query time below 2 seconds for a trajectory search with up to six simultaneous movements.

Eventually, we also conducted an expert interview asking questions about the usefulness, limitations and suggestions for improvement. First of all, our domain expert was excited about the use of our visual search system and stated the contribution as substantial to soccer analytics. He enjoyed to play with the tool and confirmed our requirements of an easy and intuitive search interface. He also sees great potential for the search approach and indicated the usage of our tool for individual player analysis or to quickly search for interesting game situations, during half-time breaks (half-time analysis).

Furthermore, we identified together limitations of the search and provided a number of improvement ideas that we want to realize in the future. For instance, to enhance our search approach by adding translation-, rotation- or scale-invariances in order to enable a more accurate search for local movement pattern. It is also conceivable that an integration of a recommendation system, as presented in [24], could improve our search system by showing popular subsequent movements during the sketching process. Another important issue is to enrich the analysis possibilities of the system. Game phases could be automatically recognized by the system and integrated into the results. Also, it would be helpful to include derived performance measures about the players into the

system, eventually allowing to compare performance (e.g., distance covered, goals on average etc.) with play movements. The system could also be extended for multiple matches and thus, enable the possibility to search for typical behavior over a whole season. In addition, the system may annotate identified game phases of the opponent's previous matches to automatically learn their tactical strategies or build a summary report of favorite attacks. Finally, the search space of movers ( $M \in \{Player, Ball\}$ ) could be increased by adding the trajectories of the referees for decision analysis. Thus, questionable decisions can be analyzed by comparing the referee's point of view and the event position for the purpose of measuring the uncertainty of the referees.

## Conclusion

We presented a novel visual search approach including multi-trajectory and event-specific search options for identifying tactical movements and unannotated scenes in soccer matches. By means of our interactive search interface that supports area and path specific queries, sports analysts are able to conveniently sketch movement patterns based on individual, formation, or team dependent trajectories. We used a preprocessing step to extract a set of locally similar trajectory parts to a given sketch and applied two trajectory descriptors (similarity functions) to rank the similarity based on two different tasks. The two tasks include the search for specific situations that happened in a match and the identification of similar movement patterns. Moreover, we designed a domain specific task analysis for movement search in soccer data and conducted a case study together with a domain expert to show the applicability of the tool. Our system was designed as an expert tool that can reveal individual player behavior and tactical team movements in a match, which can be important for decision makers such as coaches, scouts or managers. However, our visual search system may also be applied to other application areas, such as animal movement analysis, or team sports where the search of simultaneous movements is needed. Thus, for instance, our methods could be used in sports like American football, basketball or rugby to identify important team movements, which creates open spaces and game deciding points.

## Acknowledgments

This work was partially funded by the Juniorprofessor Program of the Landesstiftung Baden-Württemberg within the research project *Visual Search and Analysis Methods for Time-Oriented Annotated Data*. We thank Andreas Riederer for his valuable feedback and helping us to evaluate the system. Furthermore, we thank Prozone for providing the data.

## References

- [1] G. L. Andrienko, N. V. Andrienko, P. Bak, D. A. Keim, and S. Wrobel. *Visual Analytics of Movement*. Springer, 2013.
- [2] N. V. Andrienko and G. L. Andrienko. *Exploratory analysis of spatial and temporal data - a systematic approach*. Springer, 2006.
- [3] N. V. Andrienko and G. L. Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, 2013.
- [4] A. Bialkowski, P. Lucey, P. Carr, Y. Yue, S. Sridharan, and I. Matthews. Large-scale analysis of soccer matches using spa-

- tiotemporal tracking data. In *Data Mining (ICDM), 2014 IEEE International Conference on*, pages 725–730, Dec 2014.
- [5] T. Boren and J. Ramey. Thinking aloud: reconciling theory and practice. *Professional Communication, IEEE Transactions on*, 43(3):261–278, Sep 2000.
- [6] Y. Cao, H. Wang, C. Wang, Z. Li, L. Zhang, and L. Zhang. Mindfinder: Interactive sketch-based image search on millions of images. In *Proceedings of the International Conference on Multimedia*, pages 1605–1608. ACM, 2010.
- [7] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa. Sketch-based image retrieval: Benchmark and bag-of-features descriptors. *Visualization and Computer Graphics, IEEE Transactions on*, 17(11):1624–1636, Nov 2011.
- [8] M. Eitz, R. Richter, T. Boubekeur, K. Hildebrand, and M. Alexa. Sketch-based shape retrieval. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 31(4):31:1–31:10, 2012.
- [9] K. Goldberry. Courtvision: New visual and spatial analytics for the nba. In *2012 MIT Sloan Sports Analytics Conference*, 2012.
- [10] J. Gudmundsson and T. Wolle. Towards automated football analysis: Algorithms and data structures. In *Proc. 10th Australasian Conf. on Mathematics and Computers in Sport*. Citeseer, 2010.
- [11] J. Gudmundsson and T. Wolle. Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, (47):16–27, 2014.
- [12] H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. A. Keim, and O. Deussen. Feature-Driven Visual Analytics of Soccer Data. In *Proceedings of the 2014 IEEE Symposium on Visual Analytics Science and Technology (VAST '14)*, 2014.
- [13] J. Lee and T. Funkhouser. Sketch-based search and composition of 3D models. In *EUROGRAPHICS Workshop on Sketch-Based Interfaces and Modeling*, June 2008.
- [14] Y. J. Lee, C. L. Zitnick, and M. F. Cohen. Shadowdraw: Real-time user guidance for freehand drawing. In *ACM SIGGRAPH 2011 Papers*, pages 27:1–27:10, New York, NY, USA, 2011. ACM.
- [15] P. A. Legg, D. H. S. Chung, M. L. Parry, R. Bown, M. W. Jones, I. W. Griffiths, and M. Chen. Transformation of an uncertain video search pipeline to a sketch-based visual analytics loop. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2109–2118, 2013.
- [16] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710, 1966.
- [17] P. Lucey, A. Bialkowski, M. Monfort, P. Carr, and I. Matthews. “quality vs quantity”: Improved shot prediction in soccer using strategic features from spatiotemporal data. MIT Sloan Sports Analytics Conference, 2014.
- [18] C. Perin, R. Vuillemot, and J.-D. Fekete. Soccerstories: A kick-off for visual soccer analysis. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2506–2515, Dec 2013.
- [19] T. Polk, J. Yang, Y. Hu, and Y. Zhao. Tennivis: Visualization for tennis match analysis. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2339–2348, Dec 2014.
- [20] M. G. I. Rathod and M. D. A. Nikam. Review on event retrieval in soccer video. *International Journal of Computer Science & Information Technologies*, 5(4), 2014.
- [21] K. Samp and S. Decker. Supporting menu design with radial layouts. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 155–162. ACM, 2010.
- [22] M. Scherer, J. Bernard, and T. Schreck. Retrieval and exploratory search in multivariate research data repositories using regressional features. In *In Proc. of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, pages 363–372, 2011.
- [23] S. Shang, R. Ding, B. Yuan, K. Xie, K. Zheng, and P. Kalnis. User oriented trajectory search for trip recommendation. In *Proceedings of the 15th International Conference on Extending Database Technology*, EDBT '12, pages 156–167. ACM, 2012.
- [24] L. Shao, M. Behrisch, T. Schreck, T. von Landesberger, M. Scherer, S. Bremm, and D. A. Keim. Guided Sketching for Visual Search and Exploration in Large Scatter Plot Spaces. In *Proc. EuroVA International Workshop on Visual Analytics*. The Eurographics Association, 2014.
- [25] D. B. Shepard, A. R. Kuhns, M. J. Dreslik, and C. A. Phillips. Roads as barriers to animal movement in fragmented landscapes. *Animal Conservation*, 11(4):288–296, 2008.
- [26] D. Spretke, H. Janetzko, F. Mansmann, P. Bak, B. Kranstauber, S. Davidson, and M. Mueller. Exploration through Enrichment: A Visual Analytics Approach for Animal Movement. In *Proceedings of the 19th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '11, pages 421–424, New York, NY, USA, 2011. ACM.
- [27] T. von Landesberger, S. Bremm, J. Bernard, and T. Schreck. Smart query definition for content-based search in large sets of graphs. In *Proc. Int. Symposium on Visual Analytics Science and Technology*, pages 7–12. Eurographics Association, 2010.
- [28] B. Wang, P. Ruchikachorn, and K. Mueller. Sketchpadn-d: Wydiwyg sculpting and editing in high-dimensional space. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2060–2069, Dec 2013.
- [29] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. van de Wetering. Visual traffic jam analysis based on trajectory data. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2159–2168, 2013.
- [30] N. Willems, H. van de Wetering, and J. J. van Wijk. Visualization of vessel movements. In *Proceedings of the 11th Eurographics / IEEE - VGTC Conference on Visualization*, pages 959–966, 2009.
- [31] Z. Xiong, R. Radhakrishnan, A. Divakaran, and T. S. Huang. Audio events detection based highlights extraction from baseball, golf and soccer games in a unified framework. In *Acoustics, Speech, and Signal Processing. IEEE International Conference on*, volume 5, pages V–632. IEEE, 2003.
- [32] G. Zhu, Q. Huang, C. Xu, Y. Rui, S. Jiang, W. Gao, and H. Yao. Trajectory based event tactics analysis in broadcast sports video. In *Proceedings of the 15th International Conference on Multimedia*, pages 58–67. ACM, 2007.
- [33] G. Zhu, C. Xu, Q. Huang, W. Gao, and L. Xing. Player action recognition in broadcast tennis video with applications to semantic analysis of sports game. In *Proceedings of the 14th annual ACM international conference on Multimedia*, pages 431–440. ACM, 2006.

## Author Biography

*Lin Shao, born in 1986, studied Information Engineering at the University of Konstanz and obtained the Master's degree in 2013. His master thesis was about "sketch-based retrieval for bivariate data using image-based descriptors". Since then he has worked as a PhD candidate in the Data Analysis and Visualization Group at the University of Konstanz, Germany.*