Computers & Graphics 68 (2017) 84–95

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on SIBGRAPI 2017

Visual soccer match analysis using spatiotemporal positions of players

Vinicius Machado[a], Roger Leite[b], Felipe Moura[c], Sergio Cunha[d], Filip Sadlo[e], João L.D. Comba[a,*]

[a] Instituto de Informática – UFRGS, Brazil
[b] Vienna University of Technology – TU Wien, Austria
[c] Universidade Estadual de Londrina – UEL, Brazil
[d] Universidade Estadual de Campinas (Unicamp), Brazil
[e] Heidelberg University, Germany

ARTICLE INFO

Article history:
Received 28 March 2017
Revised 9 August 2017
Accepted 10 August 2017
Available online 18 August 2017

Keywords:
Sports analytics
Visual analytics

ABSTRACT

Soccer is a fascinating sport that captures the attention of millions of people in the world. Professional soccer teams, as well as the broadcasting media, have a deep interest in the analysis of soccer matches. Statistical summaries are the most widely used approach to describe a soccer match. However, they often fail to capture the evolution and changes of strategies that happen during a game. In this work, we present visual designs to help understanding a soccer match from the spatiotemporal position of players. We receive as input the coordinates of each player throughout the match, as well as the associated events. We present a pixel-oriented layout that summarizes the changes of player positions and tactical schemes during the match. Also, we revisit a technique used for flow analysis to help us identify where does a player move from a given region in the field. We developed our approach in conjunction with colleagues from physical education with experience in soccer analysis, as well as experts on soccer data extraction. We demonstrate the utility of our approach in several match situations, and provide the feedback given by the experts.

1. Introduction

Soccer is a very passionate team sport that attracts the attention of many fans in the world. Soccer has an important economic aspect since billions of dollars are spent to construct successful teams that win titles. However, forming winning teams is not a trivial matter, and soccer managers are looking for any information and analysis tools that can support decisions that impact the success of a team. One example is the extraction of data from the video footage of soccer matches, such as the position of players during the match, possession of the ball, actions performed (e.g., shots, crosses, tackles, fouls, etc.). This data is not easy to evaluate, due to the high volume of information and the high degree of interrelation.

Recent approaches to the scientific analysis of soccer footage data showed success to assist coaches in their decisions on team strategy, opponent analysis, and scouting prospection of players [1,2]. Most of the approaches described today rely on statistics that span the entire soccer match (e.g., heatmap of a given player on the pitch, possession time, etc.). Only recently we are seeing more sophisticated approaches for this analysis, such as the analysis of soccer player trajectories presented by Shao et al. [3]. We agree that the position of players on the field play an important role in this analysis, and describe alternate ways to visually present this information.

Our proposal relies on two basic visual designs. The first one is the pixel-oriented display proposed by Keim et al. [4]. We use this approach to create a Player Attribute Heatmap (PAH), which is a matrix display of the positions of players during a portion of the match (often a half-time). This visual display allows creating an image that tells a story of how the match developed itself from the perspective of the position of the players in the field. Our PAH can also be parameterized to use different ordering strategies while showing this information (e.g., ordering by vertical or horizontal positions of players), which can review trends such as preferred sides of the field at specific time intervals of the match. We also apply this visual design to display preferred tactical schemes used throughout the match in a Tactical Scheme Heatmap (TSH).

The second approach relies on the pathline glyphs [5] designed for the visualization of unsteady 2D flow. We address with this design the visual clutter of trajectories of players and their inherent complexity. This technique allows creating miniature glyphs that

* Corresponding author.
E-mail addresses: comba@inf.ufrgs.br, joao.comba@gmail.com (J.L.D. Comba).

http://dx.doi.org/10.1016/j.cag.2017.08.006
0097-8493/

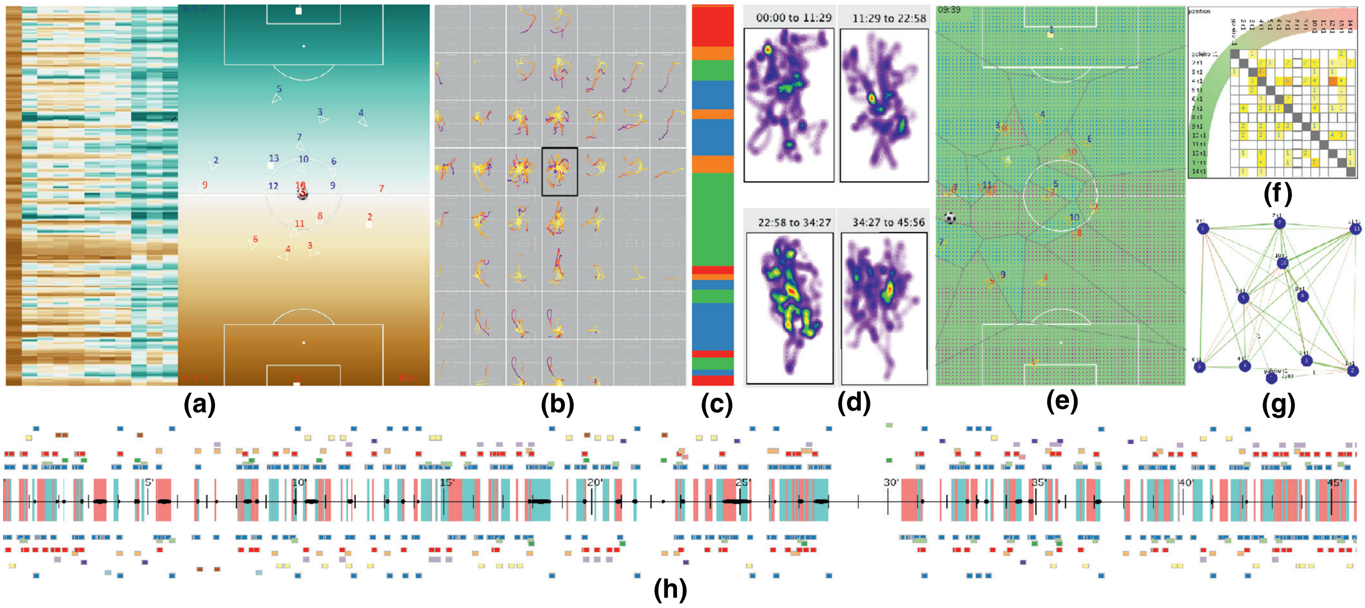


Fig. 1. Main components of our visual soccer analysis system: (a) Player positions during the match are color-coded and displayed in a matrix called Player Attribute Heatmap. (b) Trajectories of players are analyzed using a variation of pathline glyphs. (c) Tactical schemes are represented in time in a matrix called Tactical Scheme Heatmap. (d) Multiple heatmaps enable the analysis of player occupation in different time intervals of the match. (e) Voronoi diagram representation shows player proximity at a specific time of the match. Pass matrix (f) and pass graph (g) summarize passes statistics among players. (h) Event window displays all events in the match and enables interaction to constrain the analysis to a specific time interval of the match. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

represent player trajectories starting from any part of the field. For example, it is possible to identify for an attacker the preferred directions of movement when starting from a central part of the offensive field.

In summary, the contributions of this paper include:

- a system for the visual analysis of soccer matches based on the spatiotemporal position of players;
- a parametrizable matrix-oriented design that allows displaying the position of players and tactical schemes throughout the match;
- a visual summarization using pathline glyphs of the player trajectories that started from a given region in the field;
- an evaluation study of the visual design and feedback from soccer experts.

Fig. 1 gives an overview of the main components of the system, including additional functionality. The data used in this work was generated by our soccer expert collaborators.

2. Related work

Visualization of sports data is an active and comparably young field of research [6]. MatchPad [7] presents a visual timeline with glyphs to analyze the performance of players in a rugby match, while Chung et al. [8] proposed a system to visualize and interact with sports data in general. There are several systems designed for the visualization of soccer data, and we refer to Gudmundsson et al. [9] for a comprehensive survey of these aspects, and to Anderson and Sally [10] for a historical perspective on soccer over the years. Perin et al. [11] describe ways to navigate in soccer ranking tables. The first visualizations of soccer data were typically drawn by hand during the match [12]. Kim et al. [13] propose a model to specify the organization of the line of defenders, thus supporting the understanding of their strategy as well as changes over time. Fonseca et al. [14] use the Voronoi diagram to define metrics about the occupied area in defensive and offensive moments. Still analyzing the geometry of match configurations, Duarte et al. [15] uses other geometric statistics, like the convex hull and circumference to analyze football teams. Peñas et al. [16] use statistical measures to demonstrate the differences between the winning and losing teams. Salvo et al. [17] also rely on statistics to evaluate the demands placed on elite soccer players according to their positional role at different work intensities. Nowadays, companies such as OptaPro [1] and STATS [2] sell detailed data about soccer matches, requiring dedicated visualization techniques. Nevertheless, most of these data are still purely statistical, and there are only a few visualization tools that allow for a proper investigation. Match Insight [18] is a powerful real-time tool that allows statistical and visual analysis, but does not offer analysis of matches based on the position of players over time, which is key to improving this analysis.

A key ingredient to conduct path analysis is to find semantically-correct representations of such trajectories [19]. On the other hand, small multiples of pathline glyphs [5] provide an overview of important sets of trajectories by reducing occlusion and allowing for detailed inspection. The challenging problem is how to aggregate and represent different members of a team in a single image. We extend the notion of pathline glyphs in this work to handle multiple trajectories within a cell. Misue [20] presented an approach to analyze the skills of individual players. For the representation of relationships, matrices are a prominent choice. Henry and Fekete [21] present a matrix-based visualization to analyze social networks, while Dinkla et al. [22] represents groups with a dual adjacency matrix and a graph. Similar to the heatmap approach we use in this work, Oliveira et al. [23] visualize a set of temporal series related to heart rate collected during running races. We propose a matrix-oriented design that creates a visual representation of a soccer match that can lead to insights on player placement during the match (individual players or teams). Our design is peculiar because it is composed by a small number of columns (the number of players) and a large number of rows (the time instances). In addition, as we demonstrated, this design can lead to further insights if the rows or the columns of this matrix are reordered following another criterion. Another matrix-oriented technique was proposed by Andrienko and Andrienko [24] to represent traffic. Vehlow et al. [25] describe the evolution of communities in dynamic graphs inspired by pathline glyphs [5], we describe a multi-scale visualization of player trajectories in a soccer match starting at different places in the pitch. A sketch-based interface is given by Shao et al. [3] to search for specific trajectories, while Janetzko et al. [26] use parallel coordinates to compare trajectories of different players.

Spatiotemporal data was used to examine how player tracking information can be used to analyze team formation. Lucey et al. [27] propose an entropy map to show the characteristics of ball movement of each team. Bialkowski et al. [28] propose an analysis in short contexts for players and teams covering situations where players change their strategy and formation. Another work by Lucey et al. [28] analyzes situations where players change their strategy and formation. Other recent works [29,30] present analysis based on phases defined by sequential moves of one team, both presenting different techniques to represent these phases. Perin et al. [29] associate glyphs to events, and create a visual flow for different parts of the match. They also propose an automated commentary texting that associates actions with players. Janetzko et al. [30] apply different statistics metrics related to the match, player, teams, and time intervals, allowing comparison of different segments of the game.

In the design of various heatmap-based visualizations, as well as the use of pathline glyphs, we believe our approach offers novel ideas to support the analysis of the *evolution* of soccer matches. Our approach provides a global visualization that allows for a better understanding of player behavior and their trajectories during a match. For tactical analysis, we propose ways to follow the evolution of tactical schemes used in the match. We extended pathline glyphs to the context of soccer data analysis, representing trajectories of players in different spatial locations and time intervals. Using this technique in a focus+context approach allows inspection of movements in parts or at the entire field.

3. Data representation and analysis requirements

In this section, we describe the data format captured from soccer matches. Also, we list a set of questions used as guidelines for the development of our approach and its evaluation. We formulated these questions with the collaborator soccer experts that provided the data and helped evaluate our proposal.

3.1. Data

The dataset of each soccer match comprises spatiotemporal tracking data of the two-dimensional coordinates of each player on the pitch for each second of the match. Separate files store the tracking data, each representing a half-time. The first file contains the *position of players* at every second for around 45 min (2700 positions in total). The player's position was obtained from automatic tracking of video data of the respective matches, with a subsequent labeling of the players. In addition to this information, there is a second file of manually annotated *match events*, such as control, pass, dribble, shots on goal, tackle, defense, goal kick, throw-in, corner kick, offside, foul, goal, and running with the ball. One example of event is: “there is a pass from player one to player two at second 12”.

It is important to observe that the spatiotemporal data does not include the position of the ball at every instant, but only when a pass occurs. The main reason for the absence of this data is the fact that automatic extraction of the position of the ball from video footage is challenging due to occlusions that hide the ball location. However, it is possible to assert when a team has the ball possession from the passing events.

In total, we obtained data for six matches from the first division of the Brazilian soccer league. Four of these matches include the same team. However, there is no identification of the names of individual players. Therefore, it is not possible to compare the performance of a player over different matches. Nevertheless, each player has a unique identifier in a match.

3.2. Analysis requirements

Soccer is a complex game, with many complex analysis demands. To drive our development, we formed a team of three soccer collaborators. The first collaborator is a professor in physical education, with research on biomechanics analysis and mathematical models for soccer. The second expert is also a professor in physical education and biomechanics, with practical experience in the development of computer vision systems to extract soccer data from video. The third expert has a Ph.D. degree in computer science, with emphasis on data mining and has a position in a company that produces soccer data from video and elaborate soccer analysis reports.

In cooperation with the soccer experts, we elaborated a set of requirements to be supported by our analysis tool. One difficulty raised by the experts was the fact that most analysis tools offered only a statistical summary of a match, with no or little insight over how the game evolved over time. Finding appropriate approaches to enable such analysis was the main motivation in the development of our tool, and involved an approach for a global space-time representation of a match, that gives a quick grasp of events through time, and provides context and guidance for in-depth visualization components.

The outlined requirements include:

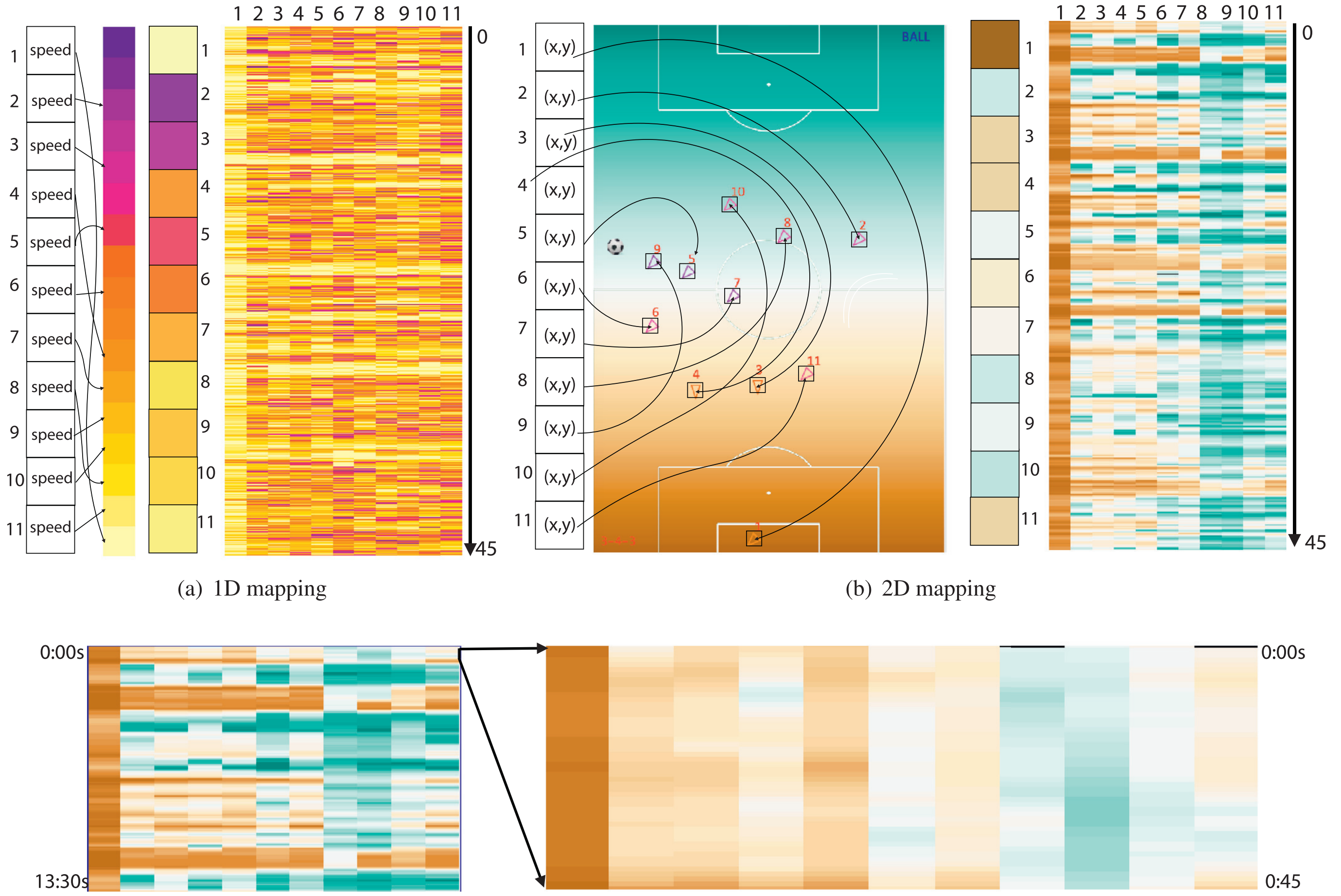
- *R1*: a compact view of the evolution of the behavior of both teams, to reveal global patterns in the match.
- *R2*: individual and team evolution for different directions, such as vertical movements (attack–defense) and horizontal movements (left–center–right), or at specific parts of the pitch (such as the middle-field, box, among others).
- *R3*: automatic identification of the tactic formation at a given instant of time, and visualization of the evolution of the tactical formation.
- *R4*: visualization of the trajectories of players in given parts of the match and regions of the pitch.
- *R5*: comparative analysis of player's trajectories to identify marking patterns.
- *R6*: ability to filter the data to narrow the analysis in main events (goals, counter-attacks, etc.), at specific time intervals, or for specific players.

4. Spatiotemporal visual analysis of players

In this section, we describe the visual designs to analyze the spatiotemporal position of players according to the requirements *R1*–*R6* formulated in the previous section

4.1. Player Attribute Heatmap

Our first visual design is called the Player Attribute Heatmap. The building block for the PAH is the encoding of speed or position of a player into a color. The color mapping is different for the one-dimensional (speed) and two-dimensional (position) attributes. The first mapping takes the speed of the player and maps into a 1D color table. The second mapping takes the coordinates of the player and maps to a 2D color table defined over the entire soccer pitch or at a specific region of interest. Fig. 2 illustrates both mappings.



(a) 1D mapping

(b) 2D mapping

(c) zoom into a specific interval of time

Fig. 2. PAH mapping of 1D and 2D attributes: (a) 1D color mapping of the speed at a given time of each of the 11 players. 1D-mappings are stacked in chronological order to generate a speed PAH. (b) 2D color mapping of the position at a given time of each of the 11 players. In this example, we use a divergent color scale in the vertical direction to allow the analysis of vertical movements. 2D mappings are stacked in chronological order to generate a position PAH. (c) Resolution of the PAH can be increased using a zoom into a specific interval of time to reduce the impact of pixel aggregation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The mapping of the p players of the same team in a given time instance is used to color a $1 \times p$ matrix of rectangular boxes. By default, this matrix is filled from left to rigth using the player number (from 1 to p). The width and height of this rectangle can be changed depending on the screen resolution. We *stack* the result of the $1 \times p$ matrix from top to bottom for all time instances t (around 2700). The resulting matrix has dimensions $t \times p$ and is called the PAH for a given team. By default, rows are ordered by time, from top to bottom, and the columns are ordered by the number of the player, from left to right. Fig. 2 illustrates possible color mappings and examples of PAHs generated with a given match data. The PAH of both teams is displayed side-by-side to allow comparison of movement patterns. For this comparison to occur, we must reflect the coordinates of one team in such way that both teams are always facing the same direction. The reflection of coordinates allows a comparison of attack–defense or left–right movement patterns.

An important aspect to observe is the fact that the vertical resolution of the display used to show the PAH often does not have the required 2700 pixels (or more) in height. To overcome this problem, we perform an aggregation over the PAH to be able to display the entire PAH within the allocated resolution. However, this aggregation might lead to coarse patterns depending on the resolution used. To overcome this problem, we use a lens metaphor to zoom into a smaller window of time in the PAH. We create a second view of the PAH for this zoom interval with enough resolution to show each different time instance, as presented in Fig. 2(c). Besides the choice of the player attribute and color mapping, the PAH provides another level of configuration, regarding the ordering of its rows and columns. The default ordering displays the information of individual players in separate columns ordered by the player number, while the rows are ordered from top to bottom in increasing time ordering. Consider now the possibility of changing, for each time step individually, the ordering of the columns to use the y-coordinate of the player (y-column ordering). Instead of the default ordering, we use the y-coordinate to order the column entries from left (smaller y) to the right (larger y). This ordering has the effect of shuffling the ordering of the players for each time step, while the rows remain the same. This ordering generates a PAH to analyze the vertical movements in the pitch, which might suggest if a given team is playing more on attack or defense. Therefore, in this ordering, the color mapping has shading changes in the y-direction. One example of the default and y-column ordering is given in Figs. 2(b) and 3(a), respectively. Similarly, a x-column ordering can be defined (Fig. 3(c)), but this ordering requires a color mapping with changes in x-direction. This ordering allows one to identify horizontal movements that can indicate if a team plays more on the left, center, or right side of the pitch.

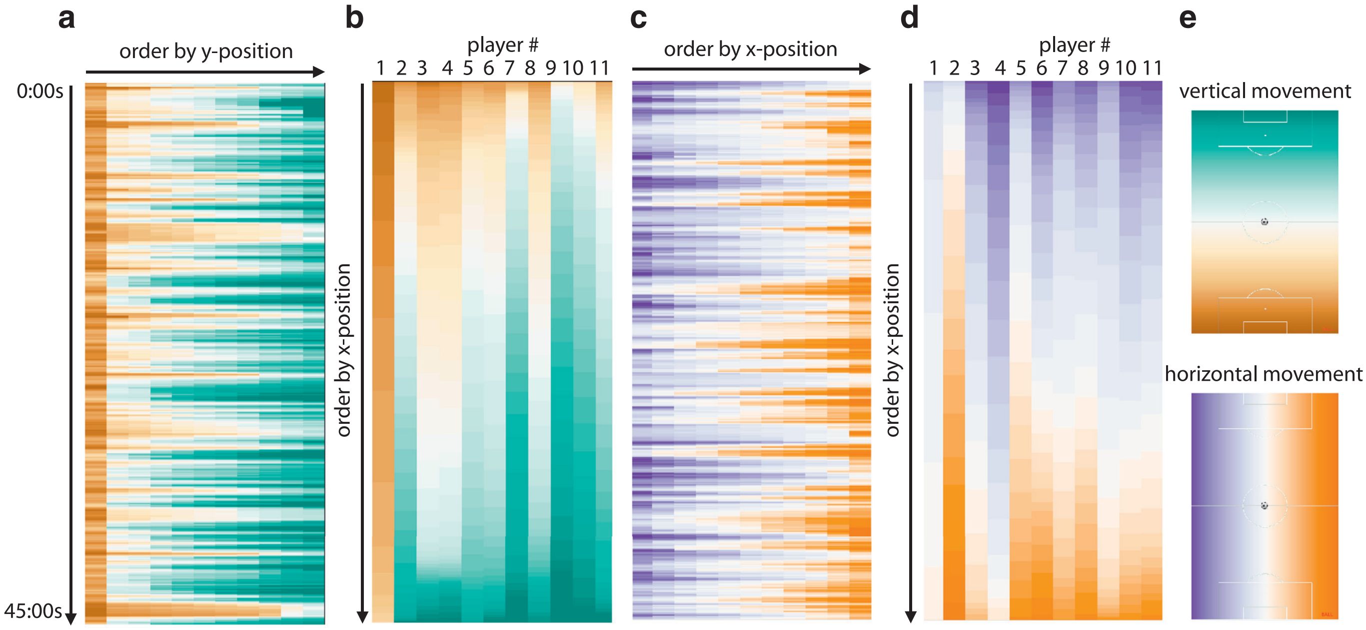


Fig. 3. PAH ordering. The *y*-column ordering (a) shows the evolution of a team with respect to attack or defense. The *y*-row ordering (b) shows in each column the distribution of vertical locations of each player. We observe that the team was more on offense (green) than on defense (brown). Similar analysis is done for horizontal movements (c, d), showing the left (purple) and right (orange) sides of the pitch. (e) 2D color mapping for vertical and horizontal movements. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In summary, column orderings are helpful to the analysis of global movements of each team.

Another degree of freedom is to change the ordering of rows, which is useful for the analysis of individual players. This new ordering keeps the columns fixed for each player. We order the time dimension by either the *x*- or *y*-row positions of players to understand vertical or horizontal movement patterns of a given player. Finally, it is possible to have a combined row and column ordering for each dimension. Similarly, it is possible to define column and row orderings for the speed attribute. Fig. 3 illustrates different orderings for the position attribute.

In summary, the PAH addresses requirements *R1* and *R2*.

4.2. Tactical scheme heatmap

The analysis of the player's positions in the PAH gives a fine detail over the positions of players, but in some situations, we need a higher level of abstraction. For example, soccer is a strategy game and coaches organize their teams into three or four regions. The number of players in each section define a given strategy, also called a tactical scheme. For instance, a 4-3-3 tactical scheme refers to placing four players on defense (D), three in the middle (M), and three in the attack (A). We use schemes defined by three sections, but our tool also supports schemes with four or more sections. Inspired by the visual design used in the PAH, we used similar ideas to display the changes in tactical schemes.

First, we need to discuss how to compute the tactical scheme of a team. It is common to use a fixed tactical scheme for the analysis of the complete match, and investigate how the team deviates from this formation. However, coaches often perform tactical variations during a match to confuse the opponent. To be able to analyze the change of schemes, we did not use a fixed tactical assignment but instead computed the formation from the player positions at each time step. We used a *k*-means clustering to find three distinct groups of players along the *y*-direction. More involved algorithms [31,32] for tactical scheme detection would likewise be possible.

One issue we had to address was the rapid change of formations in consecutive time steps, which result in outlier schemes that could confuse the visualization. To reduce outliers, we applied a filtering taking into account a sliding window centered at each time step, and selected the most common scheme used. In our tests, a window of 10 adjacent schemes provided adequate results, but any other window size is possible. A second issue to take into consideration was related to the request of the experts. They suggested considering tactical schemes from a pre-defined list, and not all possible schemes. If the clustering algorithm computes a scheme that is not on the list, we use the closest scheme based on a similarity distance between schemes. Our distance function sums the differences in each section and uses the number of players in the middle, defense, and attack as tie-breakers. For example, suppose the list of supported schemes only consists of the 4-4-2 and 3-5-2 schemes, and we detected the 4-5-1 scheme. The difference between 4-5-1 and these two schemes is two since they differ by a total of two players. The 3-5-2 is chosen using the tie-breaker since it has the same number of players in the middle as 4-5-1.

We use the similar stacking construction of the PAH to display the number of players in each of the sections in the field. For convenience, we call the matrix we obtain by the name of Tactical Scheme Heatmap. Our first idea was to display the three-section tactical scheme using three columns (D, M, A), each defined by a sequential color mapping associated with the number of players in the region. As requested by the experts, we added a second option that associates a single color to each scheme using a categorical colormap. Histograms of the used schemes are displayed along the TSH to help the analysis. Fig. 4 illustrates the TSH of the first 10 min of a match.

4.3. Pathline trajectories

Another way to analyze the spatiotemporal position of players is to consider the individual trajectories of players. Since a soccer match represents a complex spatiotemporal interaction of players, the trajectories of players provide a valuable basis for advanced analysis. To address the visualization of trajectories, we deviate

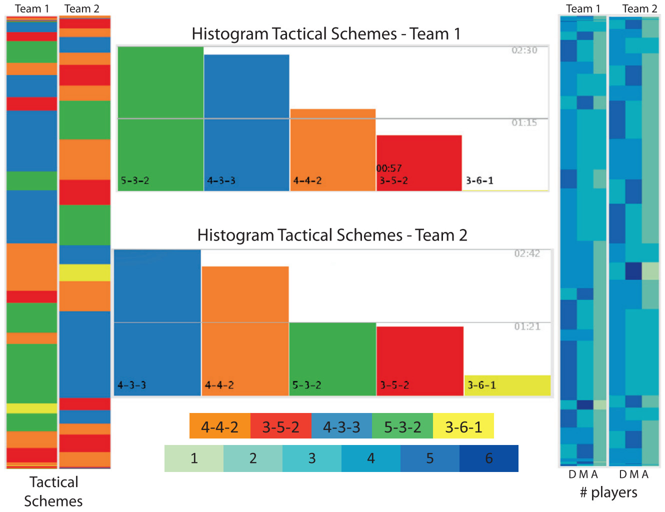


Fig. 4. Tactical scheme heatmap and histograms for two teams for 10 min of a match. The first heatmap (left) associates a single color with each tactical scheme using a qualitative colormap. We display in the center the tactical scheme histograms. The second heatmap (right) uses a sequential colormap to map the number of players on defense (D), middle (M), and attack (A) to a given color. We display the tactical scheme at each time using three columns, stacking the schemes from top to bottom. Darker (lighter) columns mean more (less) players in a given region. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

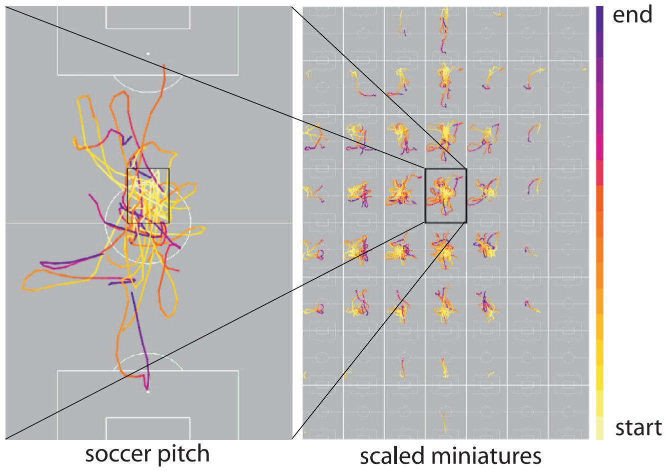


Fig. 5. Pathline construction. The soccer pitch is discretized into miniatures. For a given miniature cell, all trajectories that start in that cell are drawn over the pitch using the illustrated color scale. The resulting trajectories are scaled down and drawn inside the respective miniature. The miniatures offer an overview of all trajectories. The user can browse through the miniatures to select an interesting one, and the corresponding enlarged pitch view is shown.

from the two visual designs based on heatmaps discussed before, but we keep the goal of producing a visual design that can produce an insightful summary of the data and how it changes in time.

One problem to overcome is the fact that all trajectories of a given player over time may generate a substantial clutter due to the confined space and overplotting of trajectories. Unlike other trajectory clustering approaches that aim at finding similar trajectories, we want to answer queries such as *"where does a player move from a given region in the field?"*. In other words, where does a forward goes if it starts running in a region closer to the midfield, for example. The design we propose to answer this question is based on pathline glyphs [5], which were originally designed for the visualization of unsteady 2D flow.

The main idea behind pathline glyphs is the ability to analyze data at multiple scales, providing focus and context. The idea is to partition the domain into non-overlapping regions and draw inside each region a single downscaled version of a pathline, i.e., a trajectory. Inside each such cell, a seed point is found that is invariant to downscaling. Once found, the pathline is integrated from this seed point. This concept was shown to be very useful to understand unsteady 2D flow because there is a single trajectory inside each cell. The extension of pathline glyphs for the visualization of trajectories of a player in soccer has an important problem to overcome: the fact that it can have not one but multiple trajectories, each representing different movements a soccer player started in a given region of the field. Since we want to capture all trajectories that pass a given cell within a certain time interval, we need to handle multiple "seeds". In other words, this means that we need to consider multiple locations where trajectories might start within a cell.

A constraint given by our data is that trajectories must not consider ball stoppage times. Also, we limit the length of trajectories to a user-specified length (e.g., 30 s), but this value can be changed interactively. Fig. 5 describes the pathline glyph construction. We divide the pitch into miniatures according to a selected scale. We used different scales, starting from a single cell for the entire field, to finer levels of resolution. For each player, we divide the trajectory into intervals of specified length. Subsequent trajectories are processed using a sliding window. For each trajectory, the coordinates of its first point are used to locate the miniature cell. The trajectory is drawn over the field using a color scale to create a distinction between the start and end of the trajectory, as shown in Fig. 5. The scaled version is drawn inside each miniature of the field.

5. Results

In this section, we present the main results obtained using our approach and evaluate the visual designs proposed.

5.1. Player Attribute Heatmap

The purpose of the PAH design is to support the analysis of the evolution of the position of players throughout a match. In Fig. 6, we use PAH to understand a full game. Two divergent color mappings allow us to investigate the vertical and horizontal movements in the field (green–brown for vertical movements, and purple–orange for horizontal movements). We estimate the preferred direction of horizontal and vertical movement by comparing shading changes in the PAHs side-by-side.

In each PAH, a single row represents one instant of time. We select the y-column re-ordering to better evaluate the movements. In each row of the PAH, we order by position and not by player number. For example, if a row of the PAH is mostly green in the y-column PAH, it means that most players were on the attack. Consider, for example, the three rightmost columns of the y-column PAH in Fig. 6. We observe that team A has more players on attack than team B since there are more people with green shades in team A, and more people with brown shades in team B. This pattern continues until minute 28 when team B regains control of the match. Similarly, the x-column PAH allows to examine the preferred side a team plays. The darker shades of orange suggest that team A prefers playing on the right side of the field. On the other hand, team B prefers the left side, even when it controls the match.

Design evaluation: The benefit of the PAH depends on the color mapping and the ordering selected. The choice of colormaps in the PAHs plays a central role in this process. In one of our first experiments, we adopted a 2D colormap that interpolated four perceptually distant colors as proposed by Steiger et al. [33]. This colormap showed to be confusing since there was no clear intuition

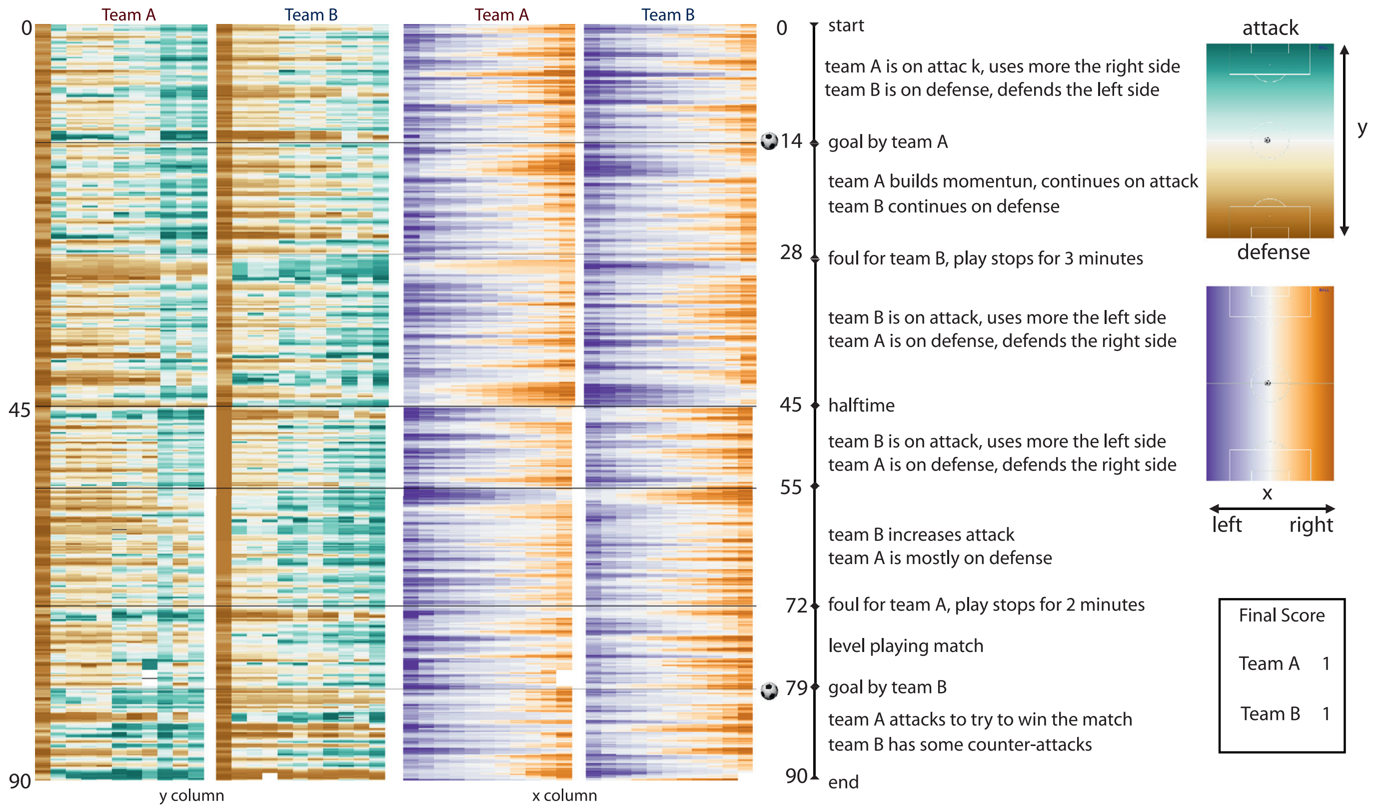


Fig. 6. PAH of an entire match between teams A and B. The PAH using *y*-column ordering is used to identify attack (green) or defense (brown) patterns, while *x*-column ordering allows the analysis of left (purple) and right (orange) patterns. We divided the match into seven interesting time intervals based on the *y*-column PAH, and describe each part to the right of the timeline above. We noticed that team A, playing at home, dominates early the match, scores at minute 14 and keeps control until minute 28. Team B regains control until around minute 72, where the match becomes more leveled. Team B scores a goal in minute 79. In the remaining minutes, there are other attacks from both teams. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the direction of movement of the players. After several attempts, two separate divergent colormaps were preferred to analyze vertical and horizontal movements in the game. Nevertheless, the fundamental piece to make the PAH more useful was the reordering. As illustrated in Fig. 6, our approach provides a clear picture on the horizontal and vertical movements.

5.2. Tactical scheme heatmap

In similar fashion to the PAH, we can observe the evolution of tactical schemes in a match in the TSH. In Fig. 7 we review the two segments of the game discussed in Fig. 6 from the perspective of the tactical scheme. We display two TSHs using the categorical and sequential colormaps described previously.

We show the time spent in each formation with the categorical colormap. We added histograms that include the time when a team has or not the possession of the ball. As mentioned before, we derive this information from the passing sequences and other events. In the user interface, it is possible to select a given scheme. The TSH only shows the times where the team applies the selected scheme. Fig. 7 shows the results of this analysis.

The sequential colormap, on the other hand, is useful to identify a concentration of players in a given region of the field during the match. We map the number of players in each section to a different color. In this example, we select shades of blue, with darker blue shades meaning more players. In summary, this design helps to evaluate each section separately by looking at the individual columns.

Design evaluation: This design depends on the algorithm to detect tactical schemes. In our first attempt, rapid changes in tactical of schemes made the analysis difficult, and only effective for shorter periods of time. By including a time window to make the algorithm more robust, the TSH became more conclusive. The results we obtained show how a certain team prefers a tactical scheme, and how it changes during the match. The separate analysis when a team has the possession of the ball was also helpful to understand attack and defense strategies. Usually, a team attacks and defends with a different number of players in each section. Also, the highlight in the TSH of the moments a given scheme occurred in time showed to be very useful. Both categorical and sequential were useful. In some situations, the experts prefer to see when which scheme was used. In others, they want to evaluate the number of players at a given section (e.g., the midfield). For this purpose, the sequential colormap was the preferred design.

5.3. Pathline trajectories

The pathline trajectories are essential to investigate the complex trajectories of players during a match. Following a focus+context approach, the pitch is divided into a grid of rectangular regions, each encoding the set of trajectories that start at the corresponding region. Fig. 8 compares the trajectories of two specific players. The visualization of trajectories that start at any place in the field offers a more informative information than a traditional heatmap of the locations of the player. The collection of all miniatures indicates that player 10 has most sub-trajectories starting in midfield, while number 9 plays more on offense. Also, the grid of miniature trajectories suggests interesting regions to be inspected next. For each player, we selected four regions, and the resulting visualization shows the entire field with trajectories that started in the corresponding region. This analysis can show different

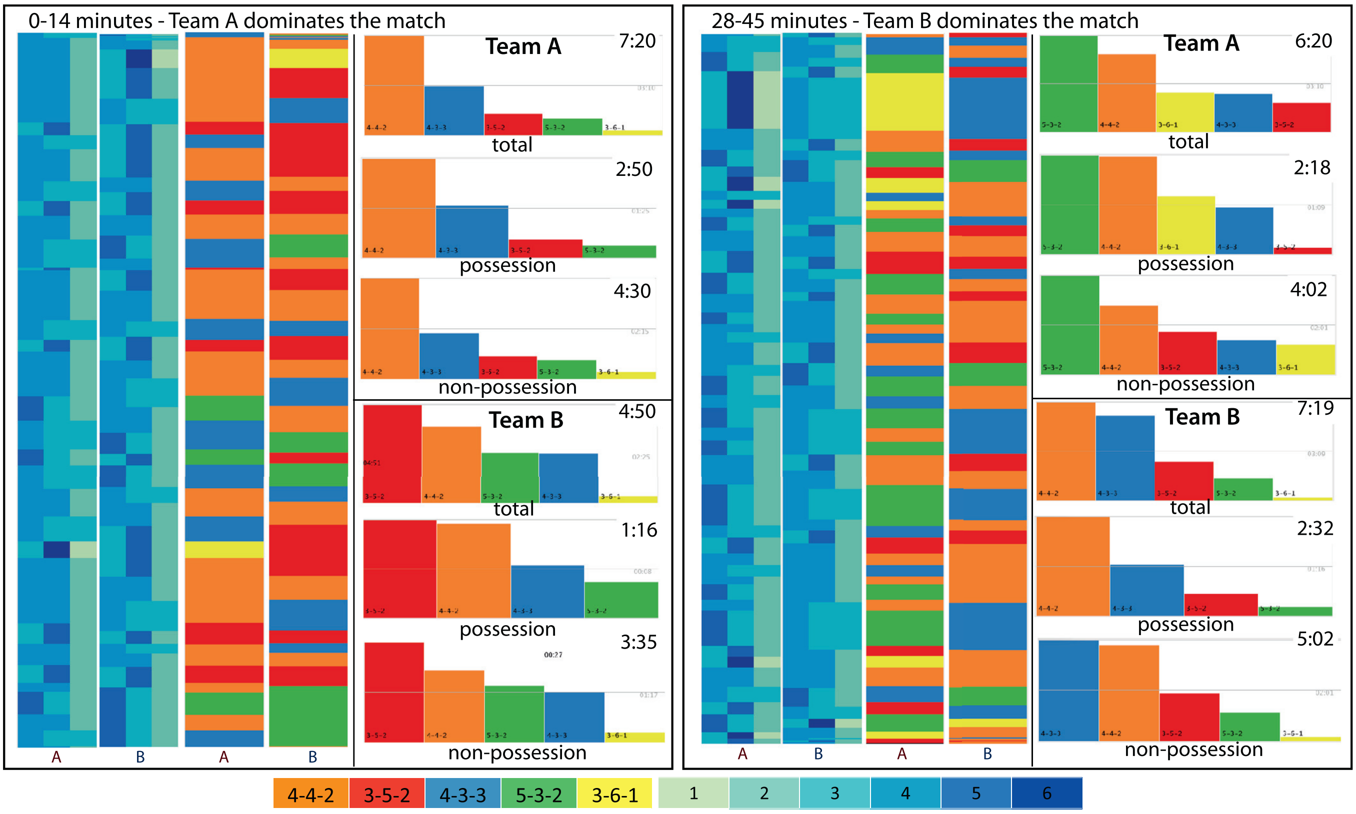


Fig. 7. TSHs for two periods of the match discussed in Fig. 6. We show two TSH alternatives: mapping schemes to colors, and mapping the number of players on defense, middle, and attack to colors. We display for each team the histograms of tactical schemes, including the histograms for when a team has or not the possession of the ball. In the first 14 minutes, team A attacks and prefers the 4-4-2 and 4-3-3 formations, while team B defends with a 5-3-2 and a 4-4-2 formation. The TSH allows us to observe when each scheme occurred. We observe that team B concentrates players in the middle when playing defense, does not use a 3-6-1 when do not have possession. On the other hand, from minutes 28–45, team B attacks using a 4-4-2 and 4-3-3, and team defends in a 5-3-2 and 4-4-2. Team A now has many players in the middle, especially in the beginning, when the team uses a 3-6-1 scheme. Team B has fewer players in the middle in this time interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

trajectory patterns (circular, vertical, or horizontal patterns). Fig 8 details some of the conclusions reached in this comparison.

Design evaluation: Our design works well with a linked view representation that offers a focus+context analysis of the trajectories. Also, additional filtering such as changing the start and end time of analysis is helpful. This design was useful to evaluate the different trajectories a player exhibits in various parts of the field, and at different time intervals. As used in other works, the colormap applied to the trajectory was beneficial to identify the direction of movement. Different sizes of the sliding function allows a finer control to inspect the trajectories. Finally, the 3D pathline glyph was helpful to visualize when two players follow similar trajectories (e.g., a forward from one team and a defense player from the other). However, it would be more useful if done automatically, which was not implemented.

5.4. Additional functionality

In the accompanying video, we include additional functionality of our system. Most additional results are related to the user interface and the linked views supported by the system. We highlight the additional visualization where one can animate the position of players at any portion of the match, the coupled views for the PAH, TSH, pathline trajectories, and others such as the ability to show multiple heatmaps, Voronoi diagrams, and passing data (Fig. 1).

6. Evaluation

We perform two different evaluations of the system. The first evaluation comprised of a demonstration of the system to soccer experts, which gave us feedback on the visual designs proposed and the situations that they might be used for. The second evaluation comprised of an evaluation study where specific tasks were given to subjects, in their majority not soccer experts. We detail the evaluation results obtained below.

6.1. Experts evaluation

The experts found that our approach allowed a comprehension of two important features of the soccer match, one related to physical performance, and another related to tactical aspects. To understand the physical demands, soccer experts analyze the percentage of time or distance covered in a given range of velocities. The player velocity (PAH speed) allows comparing the players that sustained greater periods of high intensity running and, mainly, when they performed as a function of the time. If such analysis happens during a match, it is possible to identify when a given player exhibits a performance decrease that may connect with fatigue. From a tactical perspective, the 2D mapping yields data about where players (and, consequently, the team) are concentrated during the match, and it allows one to visualize when teams predominate in attacking and defending the pitch. The ability to find movement patterns offers valuable insight to elaborate physical and tactical drills that simulate real match situations. The experts also commented that at first, the positional PAH was hard to interpret,

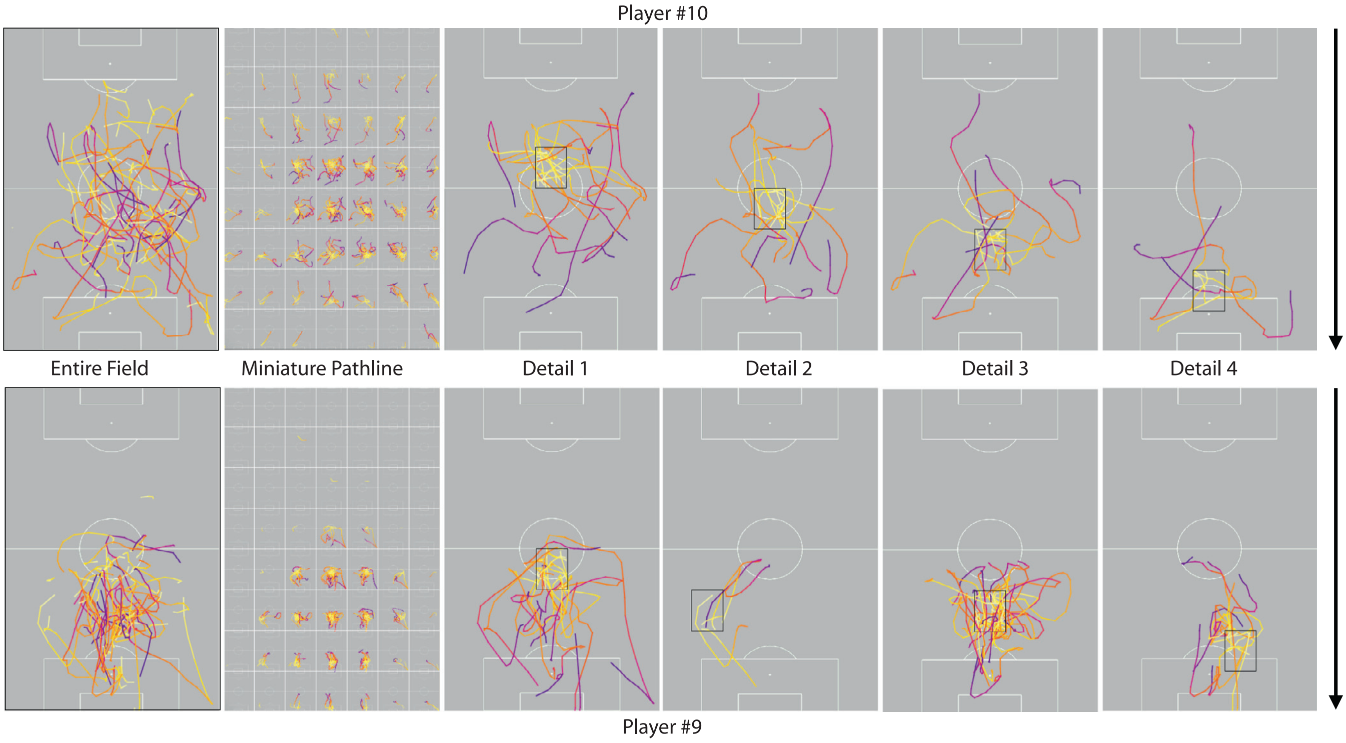


Fig. 8. The pathline glyphs for all players provide a quick summary of the preferred trajectories a given player takes from a grid of possible regions. We compare the movements of players 10 (top) and 9 (bottom). First, a summary of their movements in the entire field is shown to illustrate a general view of occupancy of each player, but this representation is nearly equivalent to a traditional heatmap visualization. On the other hand, the miniatures using pathline glyphs give an overall idea of the movements that started in specific locations of the pitch. We selected four interesting locations for both players, and evaluate the corresponding set of movements in the entire pitch. Player 10 has many actions in the midfield and large movements. On the other hand, player 9 has more action in the attack field, with most actions at the center of the field. The movements are shorter near the center of the field but more vertical towards the sides of the field.

especially with the options of different colormaps. After the choice of a specific divergent colormap that uses distinct colors for the horizontal and vertical movements, it became more natural and intuitive.

Coaches find it useful that we can identify clusters of players during the match, as a collective behavior. Even though our tactical scheme detection was simple, in combination with the PAH, they were helpful to detect a change of strategies according to the scoreline during the match. They found the sequential colormap with three colors not as intuitive as the categorical colormap. However, they found it useful to use it to detect concentrations at the middle of the field, which usually happens when a team is defending. The pathline glyph was found intriguing and much more expressive than traditional heatmaps. In particular, they like the ability identify trajectory patterns of a given player or group of teammates. It provided insights about where players move when they are in specific regions of the pitch. For example, they use the pathline glyphs to identify the behavior of players when they are attacking or defending. They find it particularly useful that the analysis can be done in specific regions of the pitch, with selected players and periods of the match. They also requested that unusual movements could be found automatically by the system. This automatic selection was not implemented in our system and would require a clustering algorithm. Shao et al. [3] offer an approach for this problem, and this will be the subject of future work.

6.2. User-study evaluation

We performed an evaluation study to assess our visual designs. The main components of this study are detailed below:

- *Design and procedure:* The evaluation was performed using an online form with accompanying explanation of the designs (included as supplemental material).
- *Hypotheses:* We formulated three main hypotheses. *H1:* The PAH is useful for identifying movement and evolution patterns. *H2*: TSH is useful for or identifying tactical schemes evolutions. *H3*: Pathline glyphs are a useful tool for identifying movement and marking patterns.
- *Subjects:* 18 subjects participated voluntarily in the study, 5 of them with experience in soccer data analysis, the remaining with experience in computer science.
- *Tasks:* We formulated a questionnaire with a total of 17 questions. For example, given a PAH, one question is to identify the part of the field (e.g., defense, middle or attack) a given player prefers. Similarly, we formulated questions for the other visual designs. The list of all analysis and evaluation questions are shown in Fig. 9. Eleven questions (Q1–Q11) showed a visual design (PAH, TSH, or pathline glyphs) and asked to answer analysis questions about soccer situations. The remaining six questions (E1–E6) posed issues related to the usefulness of the designs to meet the analysis questions using a Likert scale (“1-strongly disagree”, “2-disagree”, “3-neutral”, “4-agree”, “5-strongly agree”).

The evaluation results for questions E1–E3 show that the subjects agree that the PAH design, in its original standard format or the variations using row or column-ordering, is useful to identify preferred field locations for a team or a specific player. The answers to questions 1 and 2 agree on the locations of player 9 in the field in 5-min intervals, especially when this player is in the attack. When the player is closer to midfield, there are more divided opinions, which might be related to the color scale used.

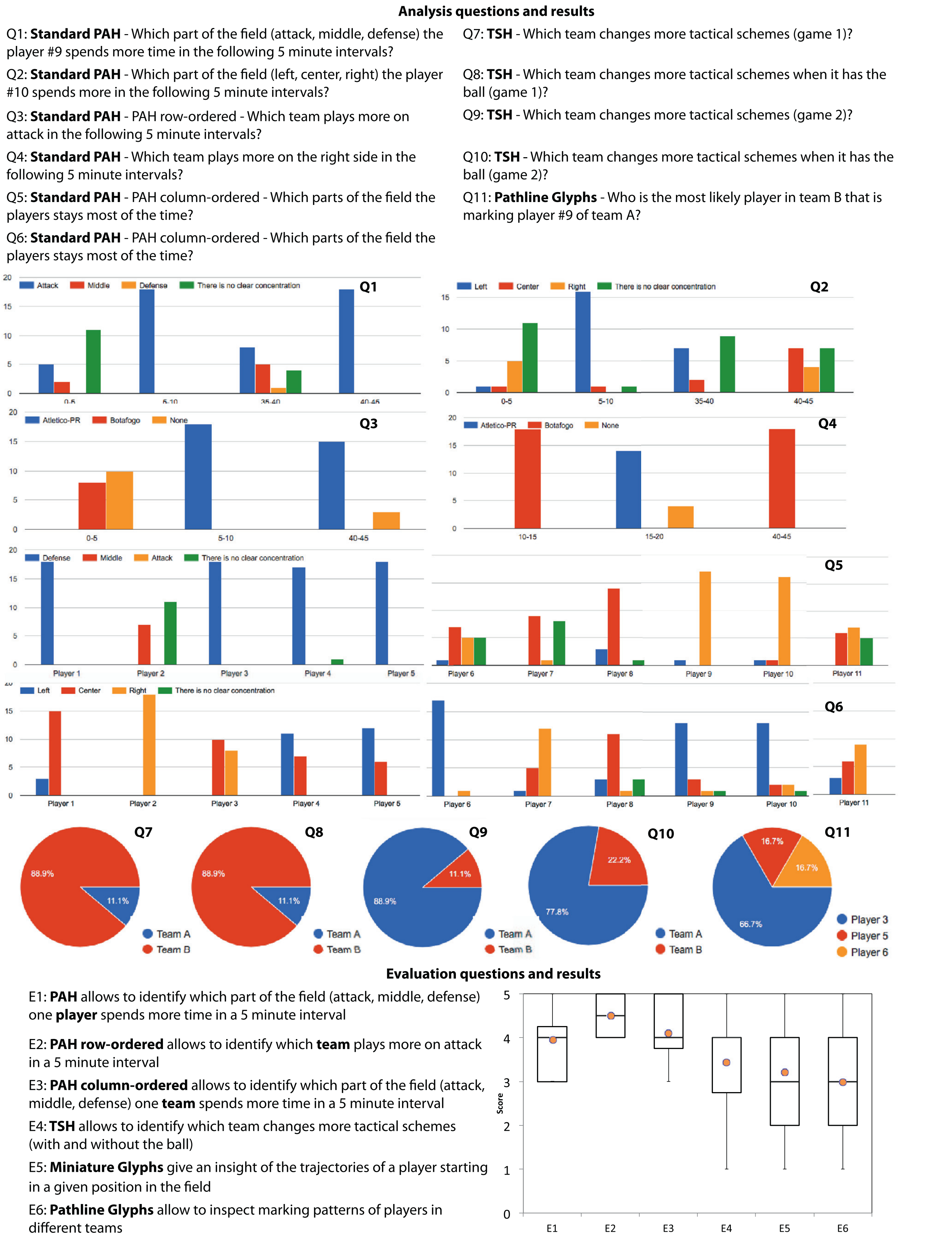


Fig. 9. Evaluation study of the visual designs. The study is composed of eleven analysis questions and six evaluation designs. We show the result of all analysis questions, as well as the results for the six evaluation questions and a box plot with results for the 18 subjects that participated in the study.

We suspect that in these situations, another color scale that covers smaller regions can better illustrate the change of positions. The concentration patterns are more easily identified when using the ordered PAHs, but again there is some disagreement when players are closer to the midfield (for vertical color scales) or central in the field (for horizontal color scales) due to the divergent color scale used. In Fig. 9, we illustrate the results for question Q5. Overall, the users were very positive about the PAH designs, and believe they satisfy requirements *R1* and *R2*, and consequently hypothesis *H1*. The results for questions Q1–Q6 show that there are situations that subjects agree consistently, but also that there are some uncertainty to answer questions in some situations.

The evaluation result for question E4 is related to requirements *R3*. The answers to questions Q7–Q10 show that near 90% of the subjects agree on the team that has more tactical changes during a period of the match. However, when asked about the ability to answer these questions using the TSH, ten subjects agree the TSH be useful, four are neutral, and four do not agree. An important point to be studied in the future is how these answers change when we consider smaller time intervals (the tests we used showed changes for 14-min intervals). Our conclusion is that hypothesis *H2* is satisfied in most situations, in particular to identify general trending tactical schemes.

The evaluation results for question E5 and E6 are related to requirements *R4* and *R5*. To answer question E5, we ask participants to consider the miniature glyphs displayed in Fig. 8(a) to assess if they give insights of the trajectories of all players. The answers were divided, eight responses were positive, and six were more negative, while four were neutral. We obtained an even more divided pattern in the answers of question E6. We conclude that this design still has room for improvements and will be the subject of future work. We plan to use other partitions of the plane, as well as adaptive subdivisions. We concluded that *H3* is partially satisfied and more work needs to be done in this design, as wells as more evaluation.

One point not directly assessed in this study is the ability to filter the analysis to specific events and time intervals (requirement *R6*). However, such features of the interface were fundamental in the design of the questions used in the evaluation study, and can also be observed in the accompanying video.

7. Conclusion and future work

In this work, we described visual designs to support the analysis of the spatiotemporal positions of players during a soccer match. The team behind the design of this work included soccer experts, described before, that supplied the soccer data used in the analysis, as well as helped in all stages of development and evaluation. Central to our ideas was the analysis of the evolution of the match from a team or individual perspective. We proposed PAH, a heatmap-based approach to summarize player attributes such as position and speed, and offer an at-a-glance view of the match. Similarly, the TSH is also a heatmap approach to enable the analysis of the changes in tactical formation during a match. To further investigate player trajectories, we used the concept of small multiples to improve the analysis of heatmaps. Finally, we extended an approach originally proposed to analyze 2D flows, the pathline glyphs, to understand the intricate trajectories taken by players during a match. Results were presented revealing interesting insights, and a discussion including feedback from expert users was given.

The system has more functionality than described here, and there are many avenues of future work. We would like to test more matches (e.g., all matches of a premium soccer league). Also, we would like to incorporate machine learning and data mining algorithms to support automated analysis of further aspects of matches, in particular related to pathline glyphs.

Acknowledgments

The authors wish to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. This work was supported in part by CNPq process 308851/2015-3 and FAPESP process 2016/05205-1.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.cag.2017.08.006.

References

[1] OptaPro. http://optasportspro.com; 2017 [accessed 06.05.17].
[2] Stats. http://www.stats.com; 2017 [accessed 06.05.17].
[3] Shao L, Sacha D, Neldner B, Stein M, Schreck T. Visual-interactive search for soccer trajectories to identify interesting game situations. Electron Imaging 2016;2016(1):1–10.
[4] Keim DA. Designing pixel-oriented visualization techniques: theory and applications. IEEE Trans Vis Comput Graph 2000;6(1):59–78.
[5] Hlawatsch M, Sadlo F, Jang H, Weiskopf D. Pathline glyphs. Comput Graph Forum 2014;23(2):497–506.
[6] Page M, Moere AV. Towards classifying visualization in team sports. In: Proceedings of international conference on computer graphics, imaging and visualisation; 2006. p. 24–9.
[7] Legg PA, Chung DHS, Parry ML, Jones MW, Long R, Griffiths IW, et al. MatchPad: interactive glyph-based visualization for real-time sports performance analysis. Comput Graph Forum 2012;31(3pt4):1255–64.
[8] Chung DHS, Parry ML, Griffiths IW, Laramee RS, Bown R, Legg PA, et al. Knowledge-assisted ranking: a visual analytic application for sport event data. IEEE Comput Graph Appl 2016;36(3):72–82.
[9] J. Gudmundsson, M. Horton, Spatio-temporal analysis of team sports – a survey, CoRR; abs/1602.06994. http://arxiv.org/abs/1602.06994; 2016 [accessed 06.05.17].
[10] Anderson C, Sally D. The numbers game: Why everything you know about football is wrong. Penguin; 2013.
[11] Perin C, Vuillemot R, Fekete J-D. A table!: improving temporal navigation in soccer ranking tables. In: Proceedings of ACM conference on human factors in computing systems, CHI'14. New York, NY, USA: ACM; 2014. p. 887–96.
[12] FootballDraw. http://www.susken-rosenthal.de/fussballbilder/; 2017 [accessed 06.05.17].
[13] Kim H-C, Kwon O, Li K-J. Spatial and spatiotemporal analysis of soccer. In: Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems, GIS'11. New York, NY, USA: ACM; 2011. p. 385–8.
[14] Fonseca S, Milho J, Travassos B, Araújo D. Spatial dynamics of team sports exposed by Voronoi diagrams. Hum Mov Sci 2012;31:1652–9.
[15] Duarte R, Araújo D, Folgado H, Esteves P, Marques P, Davids K. Capturing complex, non-linear team behaviours during competitive football performance. J Syst Sci Complex 2013;26(1):62–72.
[16] Lago-Peñas C, Lago-Ballesteros J, Dellal A, Gómez M. Game-related statistics that discriminated winning, drawing and losing teams from the Spanish Soccer League. J Sports Sci Med 2010;9(2)):288–93.
[17] Salvo VD, Baron R, Tschan H, Montero FJC, Bachl N, Pigozzi F. Performance characteristics according to playing position in elite soccer. Int J Sports Med 2007;28(3):222–7.
[18] SAP. http://go.sap.com; 2017 [accessed 06.05.17].
[19] Andrienko GL, Barrett L, Dostie M, Henzi SP. Space transformation for understanding group movement. IEEE Trans Vis Comput Graph 2013;19(12):2169–78.
[20] Misue K. Visualization of team skills in project-based learning. In: Proceedings of international conference on digital information processing, e-business and cloud computing (DIPECC2013); 2013. p. 8–15.
[21] Henry N, Fekete J-D. MatLink: enhanced matrix visualization for analyzing social networks. In: Proceedings of international conference on human–computer interaction - INTERACT; 2007. p. 288–302.
[22] Dinkla K, Riche NH, Westenberg MA. Dual adjacency matrix: exploring link groups in dense networks. Comput Graph Forum 2015;34(3):311–20.
[23] Oliveira G, Comba J, Torchelsen R, Padilha M, Silva C. Visualizing running races through the multivariate time-series of multiple runners. Graph Patterns Images 2013;26:99–106.
[24] Andrienko G, Andrienko N. Spatio-temporal aggregation for visual analysis of movements. In: Proceedings of IEEE Symposium on visual analytics science and technology, 2008, VAST'08; 2008. p. 51–8.
[25] Vehlow C, Beck F, Auwrter P, Weiskopf D. Visualizing the evolution of communities in dynamic graphs. Comput Graph Forum 2015;34(1):277–88.
[26] Janetzko H, Stein M, Sacha D, Schreck T. Enhancing parallel coordinates: statistical visualizations for analyzing soccer data. Electron Imaging 2016;2016(1):1–8.

[27] Lucey P, Oliver D, Carr P, Roth J, Matthews I. Assessing team strategy using spatiotemporal data. In: Proceedings of ACM SIGKDD international conference on knowledge discovery and data mining, KDD'13. New York, NY, USA: ACM; 2013. p. 1366–74.
[28] Bialkowski A, Lucey P, Carr P, Yue Y, Sridharan S, Matthews I. Large-scale analysis of soccer matches using spatiotemporal tracking data. In: Proceedings of IEEE international conference on data mining (ICDM); 2014. p. 725–30.
[29] Perin C, Vuillemot R, Fekete J-D. Soccerstories: A kick-off for visual soccer analysis. IEEE Trans Vis Comput Graph 2013;19(12):2506–15.
[30] Janetzko H, Sacha D, Stein M, Schreck T, Keim D, Deussen O. Feature-driven visual analytics of soccer data. In: Proceedings of 2014 IEEE conference on visual analytics science and technology (VAST); 2014. p. 13–22.
[31] Wei X, Sha L, Lucey P, Morgan S, Sridharan S. Large-scale analysis of formations in soccer. In: Proceedings of 2013 international conference on digital image computing: techniques and applications (DICTA); 2013. p. 1–8.
[32] Wang Q, Zhu H, Hu W, Shen Z, Yao Y. Discerning tactical patterns for professional soccer teams: An enhanced topic model with applications. In: Proceedings of ACM SIGKDD international conference on knowledge discovery and data mining, KDD'15. New York, NY, USA: ACM; 2015. p. 2197–206.
[33] Steiger M, Bernard J, Mittelstädt S, Lücke-Tieke H, Keim D, May T, et al. Visual analysis of time-series similarities for anomaly detection in sensor networks. Comput Graph Forum 2014;33(3):401–10.