Director's Cut: Analysis and Annotation of Soccer Matches

Manuel Stein, Halldór Janetzko, Thorsten Breitkreutz, and Daniel Seebacher ▪ *University of Konstanz*

Tobias Schreck ▪ *Graz University of Technology*

Michael Grossniklaus, Iain D. Couzin, and Daniel A. Keim ▪ *University of Konstanz*

In an effort to improve soccer player and match analysis, a visual-interactive and data-analysis support system focuses on key situations by using rule-based filtering and automatically annotating key types of soccer match elements. The proposed methods lay the foundation for innumerable further and deeper analysis tasks.

Soccer has evolved from a minor sport into popular worldwide phenomenon. Today, the FIFA (Fédération Internationale de Football Association) has more member countries than the United Nations. As a result of this increased popularity and mass marketing, the structures of professional soccer clubs have in turn evolved over the decades. Modern soccer clubs are akin to corporate entities that center around a respective soccer team and its successful operation. Many auxiliary and infrastructure departments in clubs provide supportive functions such as the promotion of young players (club development), medical treatment of players (performance maintenance and optimization), and game analysis (development and alignment of tactics and strategies). The game analysis department employs video analysts and is directly connected with the coaching team. The task of these experts is to identify the strengths and weaknesses of their teams and those of their opponents, both before and after matches. Their findings are used to adjust training and thereby raise the team's awareness of dangerous situations.

For decades, video analysts used video recordings of matches that were manually processed, annotated, and edited for analysis and presentation. With advances in sensor technology, it has become possible to actively track player position and event data with high temporal and spatial resolution. Furthermore, video cameras installed on stadium roofs additionally allow passive tracking of players. Depending on availability and regulations (for example, FIFA disallows active tracking), either or both of these modalities can be used to capture match data. Most recently, the increasing availability of automatically recorded motion and event data enables the development of automatic data-analysis methods to help in the soccer analysis process. Many of the traditional video analyst tasks involve manual inspection and transcription of video material to identify scenes of interest and to gather descriptive data, for example, on player performance. With motion and event data now readily available, it is time to ask how we can make the traditional video analysis tasks more efficient and what novel analyses are now possible.

Specifically, this article looks at how we can be appropriately transform and visually represent soccer motion data to help analysts detect interesting match patterns and assess players and groups of players. To address this goal, we focus on the workflow of professional soccer clubs depicted in Figure 1. Typically, a coach has hypotheses about how to improve the performance of his team during a soccer match. He tells the video analysts which kind of situations to look for and how they should visually prepare those situations for instructive playback. Based on the demonstration videos, the coach can either look for further insights or present the findings to his team. Our work sup-
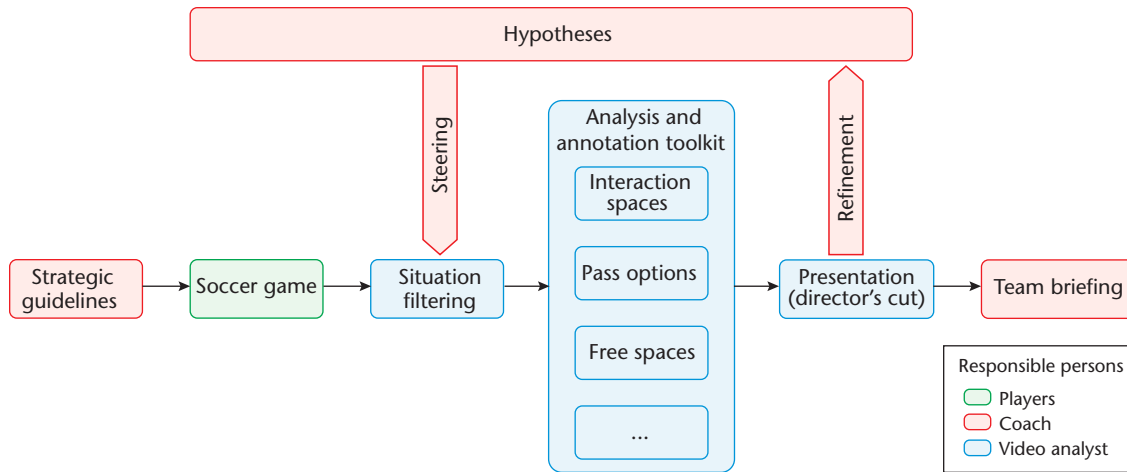
Figure 1. Typical analysis workflow in professional soccer clubs. Different colors show the person responsible for each step. The video analyst's work is driven by the needs and hypotheses of the team coach.

ports the video analysts during their analysis of specific aspects of soccer matches by employing visual data-analysis methods and semiautomatic techniques. We reduce the analysts' workload by speeding up the filtering for specific situations using an adjustable rule-based selection engine. This results in a director's cut with automatic annotations that the video analyst can further edit. (Our use of the term director's cut refers to movies, where a director's cut represents the director's viewpoint of the story.)

In this work, we contribute effective automatic annotation methods addressing, in a heuristic and domain-dependent way, three essential pillars of soccer match analysis: interaction spaces, free spaces, and pass options. Together with rule-based annotation, they enhance and speed up the otherwise manual interactive analysis and annotation process. Hence, the methods lay the foundation for innumerable further and deeper, more insightful analysis tasks. (See the "Related Work in Interactive Data Analysis" sidebar for previous research.)

Annotation of Key Game Elements

We conducted several informal expert interviews with soccer video analysts to determine which visual annotation functions would help their work. As Figure 1 shows, analysts pay particular attention to players' interaction spaces, arising free spaces, and possible pass options while assessing a soccer situation. We begin by describing our proposed advanced annotation capabilities in detail, which were designed and implemented as a result of the expert design process. As an added feature, we have provided several videos of our results online (see http://files.dbvis.de/stein/Interaction_Space, http://files.dbvis.de/stein/Free_Space, and http://files.dbvis.de/stein/Pass_Alternatives).

Interaction Space

Soccer players in our expert's traditional working environment are visualized as colored points on a soccer pitch. But players are characterized by more than their x and y location at a specific moment in time. During a soccer match, each player has a surrounding area, which he aims to control. This area is called the *interaction space*. Interaction spaces are especially interesting when analyzing passes or their reception. Furthermore, player duels are directly related to interaction spaces.

Determining interaction space. The foundation for our calculation of interaction spaces was conceptualized by Jean-Francis Grehaigne and his colleagues.[1] By performing several experiments, they measured how a player's interaction space morphs according to different velocities. Interestingly, small changes in speed are enough to drastically change the interaction space from a circle to a circular sector. (See the "Applications of Movement Analysis" sidebar for related work in movement data analysis.)

We extend this discrete model and derive a continuous model by using polynomial interpolation, allowing a smooth transition of interaction spaces:

$$\text{angle}(v) = -0.0038\pi \cdot v^3 + 0.0793\pi \cdot v^2 - 0.6108\pi \cdot v + 2\pi.$$

Furthermore, we enhance their model by integrating the distance between the player and ball, reflecting that a player who is further away from the ball has more time to react, which is described by R as follows:

$$\text{radius}(v) = \begin{cases} R \cdot \overline{v} & \text{if } v \approx 0 \\ R \cdot v & \text{otherwise} \end{cases}$$

Related Work in Interactive Data Analysis

Our work relates to several strands of research. First, in visual and interactive data analysis, there is a growing need for effective tools to understand, search, and analyze large datasets as the size and complexity of data increases. Visual analytics is concerned with researching effective approaches for data analysis, relying on interactive visualization to present and navigate large datasets by appropriate visual structures.[1] At the same time, the goal is to incorporate appropriate data-analysis algorithms, ranging from data preprocessing and data reduction to pattern detection. The analysis of large data is a common problem in many areas, including business and finance, social media, science, and engineering.

Owing to the advances in data-sensor technologies, large amounts of sports domain data can now be acquired, such as player motion or match event data. Thus, researchers have proposed several systems that are specific to the visual analysis of sports data.[2] For example, Philip Legg and his colleagues developed a visual analytics system to visually explore rugby match data based on user searches and visualized relevant scenes on a rugby pitch diagram.[3] Charles Perin and his colleagues also presented a carefully crafted system for interactive visualization of soccer matches based on annotated events and movement data.[4] A large body of work seeks to define match- and sport-specific performance measures and their automatic computation. For example, Ho-Chul Kim and his colleagues analyzed the defenders of a soccer team for their spatial formations during a game.[5] Measures of player influence in a soccer game have also been explored.[6]

In this work, we are concerned with the visual annotation of players, movements, and formations in a soccer match. In that respect, our research also relates to works in movement analysis. Visual movement analysis has been applied in many areas, some of which include people's indoor movement patterns[7] and pedestrians' outdoor movement in a street network.[8] Researchers have also undertaken the visual analysis of large amounts of GPS trajectories from taxis.[9]

In a previous work, we analyzed soccer games based on low-level features extracted from motion data.[10,11] Specifically, we formed feature vectors of player properties such as speed, distance to ball, and passing relationships between players. We used these features to segment different situations in a match, compare players to identify similarities, and classify interesting situations. However, the features used were generic, low-level movement features. In the main text, we define a set of domain-specific mid- and high-level features for soccer analysis based on the area of influence and spatial relationships.

References

1. D. Keim et al., eds., *Mastering the Information Age: Solving Problems with Visual Analytics*, Eurographics Assoc., 2010.
2. R. Basole et al., organizers, *Proc. 1st IEEE Workshop on Sports Data Visualization*, 2013.
3. P.A. Legg et al., "Transformation of an Uncertain Video Search Pipeline to a Sketch-Based Visual Analytics Loop," *IEEE Trans. Visualization and Computer Graphics*, vol. 19, no. 12, 2013, pp. 2109–2118.
4. C. Perin et al., "SoccerStories: A Kick-Off for Visual Soccer Analysis," *IEEE Trans. Visualization and Computer Graphics*, vol. 19, no. 12, 2013, pp. 2506–2515.
5. H.-C. Kim, O. Kwon, and K.-J. Li, "Spatial and Spatiotemporal Analysis of Soccer," *Proc. 19th ACM Sigspatial Int'l Conf. Advances in Geographic Information Systems*, 2011, pp. 385–388.
6. S. Fonseca et al., "Measuring Spatial Interaction Behavior in Team Sports Using Superimposed Voronoi Diagrams," *Int'l J. Performance Analysis in Sport*, vol. 13, no. 1, 2013, pp. 179–189.
7. Y. Ivanov et al., "Visualizing the History of Living Spaces," *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 6, 2007, pp. 1153–1160.
8. T. Schreck et al., "A Visual Analytics Approach for Assessing Pedestrian Friendliness of Urban Environments," *Geographic Information Science at the Heart of Europe*, Springer, 2013, pp. 353–368.
9. X. Huang et al., "Trajgraph: A Graph-Based Visual Analytics Approach to Studying Urban Network Centrality Using Taxi Trajectory Data," *IEEE Trans. Visualization and Computer Graphics*, vol. 22, no. 1, 2015, pp. 160–169.
10. H. Janetzko et al., "Feature-Driven Visual Analytics of Soccer Data," *Proc. IEEE Conf. Visual Analytics Science and Technology (VAST)*, 2014, pp. 13–22.
11. M. Stein et al., "Visual Soccer Analytics: Understanding the Characteristics of Collective Team Movement Based on Feature-Driven Analysis and Abstraction," *Int'l J. Geo-Information*, vol. 4, no. 4, 2015, pp. 2159–2184.

where the radius is the radius of the interaction space, v is the player's speed, R is the time until the ball reaches the player, and $\bar{v}$ is the player's average speed.

Figure 2 illustrates our approach with different velocities and distances between the ball and player. The maximum time of 10 seconds will occur when the ball and player are located in opposite corners of the soccer pitch.

When a player is next to players on the opposing team, his movement is restricted and he needs to pass them. Furthermore, opposing players will try to block him. Interaction spaces must reflect these interdependencies, as Figure 3 shows. A player's

interaction space is restricted to the area that he can reach before opposing players. The intersection in the interaction spaces is computed and used to visualize the corresponding restricted interaction spaces.

Potential duel areas. When two interaction spaces overlap, it is unclear which two players will dominate the intersection area. Individual skills, reaction times, and chance all influence ball possession or loss in these ambiguous cases. We visualize these areas using hatching, as Figure 4 depicts. Hatched areas are locations that both players can potentially reach simultaneously. Consequently, we first intersect both players' interaction spaces and determine the potential duel areas in the middle of both players.

In soccer, it is only natural that many players are close to each other and consequently several potential duel areas will arise. We optionally avoid the visual clutter of presenting too many nearby duel areas by unifying them, showing a contour line. Figure 5 illustrates our aggregation-based approach. In these surrounded areas, we use transparent colors and depict the density of players. The intensity of the coloring depicts a higher possibility of one team winning the ball. We apply an adjusted density-based clustering method to detect players that are close to each other. As input parameters, the clustering method needs the number of players forming a cluster and a neighborhood radius to accumulate players to a cluster. In our case, the radius is dependent on the size of the respective interaction space. Found clusters are visualized by their convex hull.

Free Space
Free space is an important feature of soccer, but its automatic estimation is not straightforward. We start by observing that each player covers a certain amount of the soccer pitch. If we simplify and assume that each player covers a circular region of 4 meters around him and the soccer pitch is 68 meters wide and 105 meters long, then all 22 players will cover approximately only 15 percent of the soccer pitch, so it is naturally sparse. However, actions in soccer are usually centered around the ball and free spaces are of different importance. Thus, the definition of free space as being the regions not covered by any player is consequently too simple. However, there exists only an intuition among soccer experts and not a precise definition of free space. We conducted several interviews with subject matter experts and identified two approaches to assess free space.
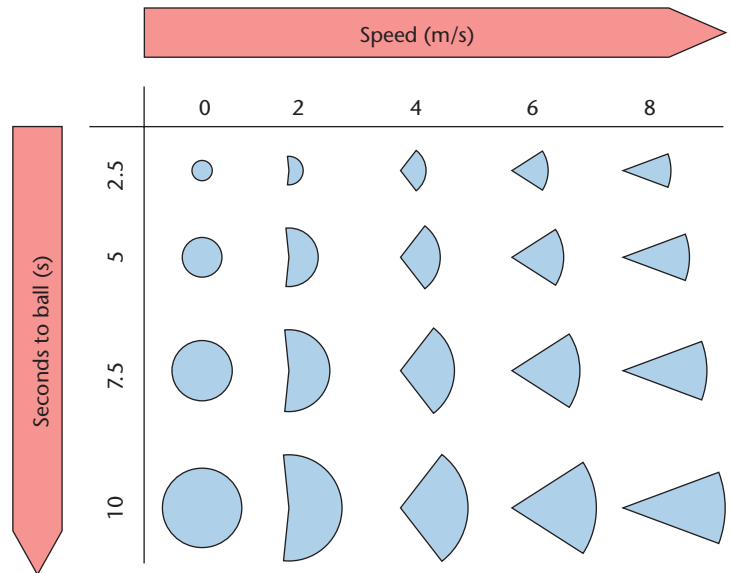

Figure 2. Speed and distance between the player and ball influence the calculation of interaction spaces. The faster a player moves forward (to the right of this figure), the less possible it is for sudden changes in heading.
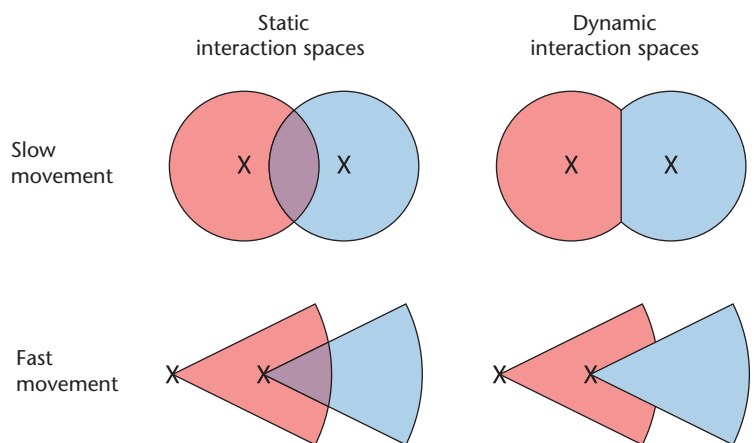

Figure 3. Interaction spaces are influenced by adjacent players. Because the opposing team's players restrict a player's movement, interaction spaces thus must reflect these interdependencies.
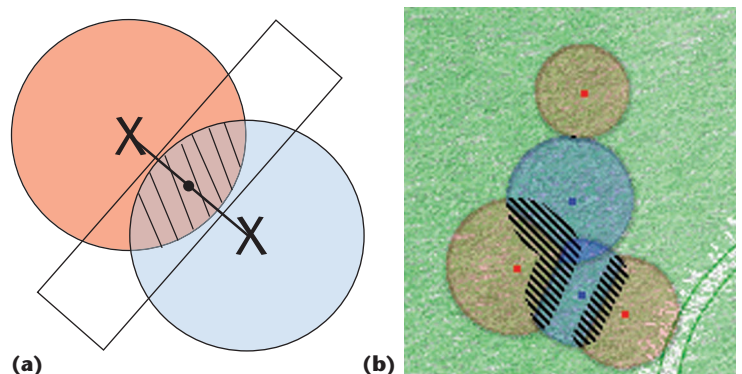

(a)   **(b)**

Figure 4. Potential duel area of two players visualized using hatching. Hatched areas are locations that can potentially be reached by (a) two or (b) more players simultaneously.

Applications of Movement Analysis

Movement data is a ubiquitously relevant type of data arising in many application domains. Apart from sports, it is extremely relevant to any sort of flow analysis, such as traffic and logistics. Motion can occur on many scales. For example, the expansion of the universe is studied on the astronomic scale. The middle scale might look at the movement of animals or pedestrians, either in isolation or within a group of objects (such as a herd or swarm). Complex interrelationships may be found and studied across time and space. Motion also occurs on micro scales, such as Brownian molecular motion.

Across the different scales, a number of fundamental analysis tasks exist. These include segmentation (identifying subsets of motion), abstraction (aggregation and simplification of motion), correlation and comparison of movements, and classification of movement patterns. Researchers have studied data-analysis techniques for many of these tasks, and visual representations can help provide effective user analysis for large trajectory data.[1]

In addition to sports analysis, many other applications use motion analysis. In many data-analysis scenarios, data may be represented as points in a diagram space. For example, the Gapminder system represents time-dependent data in animated scatterplots (www.gapminder.org). In another work, features of trajectories in animated scat-terplots are computed and used to detect and segment interesting subintervals, based on outlying motion features.[2] In general, it is interesting to analyze group movement data to distinguish different roles of the moving member elements[3] or to correlate movement patterns with location properties.[4] It is also interesting to ask if we can come up with a taxonomy of movement patterns.[5]

References

1. G. Andrienko et al., *Visual Analytics of Movement*, Springer, 2013.
2. T. von Landesberger et al., "Feature-Based Automatic Identification of Interesting Data Segments in Group Movement Data," *Information Visualization*, vol. 13, no. 3, 2017, pp. 190–212.
3. M. Andersson et al., "Reporting Leaders and Followers among Trajectories of Moving Point Objects," *GeoInformatica*, vol. 12, no. 4, 2008, pp. 497–528.
4. G. Andrienko et al., "Discovering Bits of Place Histories from People's Activity Traces," *Proc. IEEE Symp. Visual Analytics Science and Technology (VAST)*, 2010, pp. 59–66.
5. S. Dodge, R. Weibel, and A. Lautenschuetz, "Towards a Taxonomy of Movement Patterns," *Information Visualization*, vol. 7, nos. 3–4, 2008, pp. 240–252.
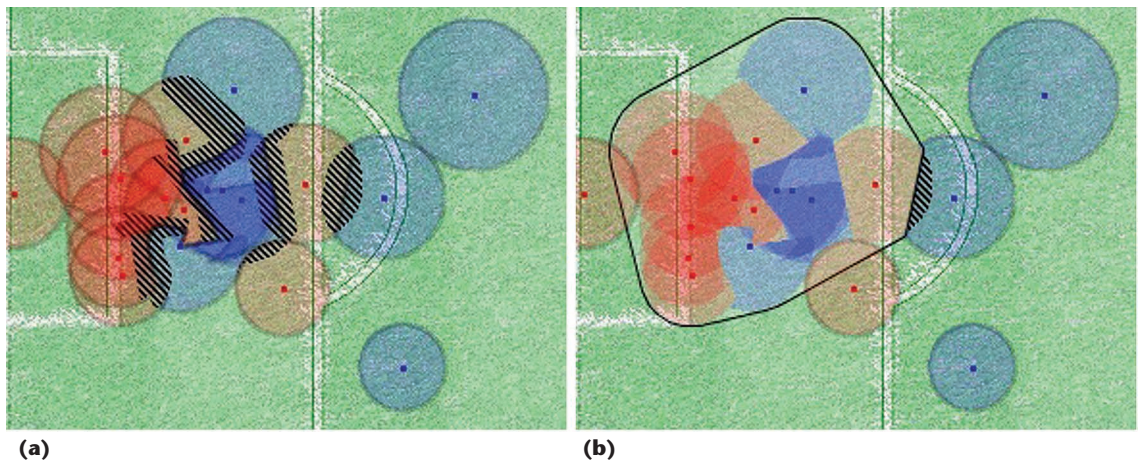
Figure 5. Optional aggregation-based visualization. To avoid the visual clutter in dense player regions, we use transparent colors to indicate (a) many overlapping duel areas. (b) In the hull-based visualization, color intensity depicts a higher possibility of one team winning the ball.

The first approach judges the notion of relevance of free space. The basic assumption is that the complete soccer pitch can be seen as free space with 22 exceptions. Employing domain knowledge, analysts manually partition the soccer pitch into several areas and decide which free spaces are relevant and irrelevant. Unfortunately, the resulting free spaces are not necessarily reproducible by asking several analysts.

The second approach describes free space as a measure of how much a player is put under pressure. Such pressure will hinder a player's ability to move freely and interact with the ball. We categorize pressure into three classes:

■ *No pressure.* The player is able to move around freely.

- *Weak pressure*. The player is being targeted by an opponent moving toward him. The player has less time to act.
- *Strong pressure*. The opponent is close to the player and trying to get the ball. The player has nearly no opportunity to act proactively because his priority is to defend the ball. Chances for errors increase severely.

Based on our interviews and findings, we developed several methods for the exact determination of free space. With our approach, we use both approaches for defining free space. We assess detected free spaces by the respective size, the number of opposing players, and the distance to the opposing goal.

In detail, our method works as follows. We segment the soccer pitch into grid cells of 1 square meter. We assign to each cell the player with the highest probability of arriving there first with respect to distance, speed, and heading. Consequently, free space can be defined as the region a player can reach before opposing players.

We visualize the resulting free space by drawing a colored grid on the soccer pitch. The analyst can see an overview about the spatial distribution of both teams. The analysis can additionally be steered into the direction of player behavior, for example, when inspecting zonal defense. Figure 6 illustrates our free space visualization.

Because the grid-based free space is possibly too detailed for our experts and furthermore might provide too high a level of accuracy, we introduce a rectangular abstraction of free spaces. To calculate the abstract free space, we adapted an existing algorithm, which detects the largest rectangle within a character matrix and has a runtime of $O(M \cdot N)$. We adjusted the algorithm to accept a number matrix as input.

Figure 7a shows how the biggest rectangle within a matrix of numbers can be calculated applying the Vandevoorde technique.[2] Our abstraction algorithm follows these steps:

1. It starts with a grid-based representation of a particular player's available free space. We determine the minimum bounding rectangle containing this representation.
2. A numerical, zero-initialized matrix is filled with ones that correspond to the free space. Consequently, the result is an array approximating the shape of the free space.
3. The Vandervoorde algorithm works in a scanline fashion and processes the matrix row by row. In each row, the start and end point (and
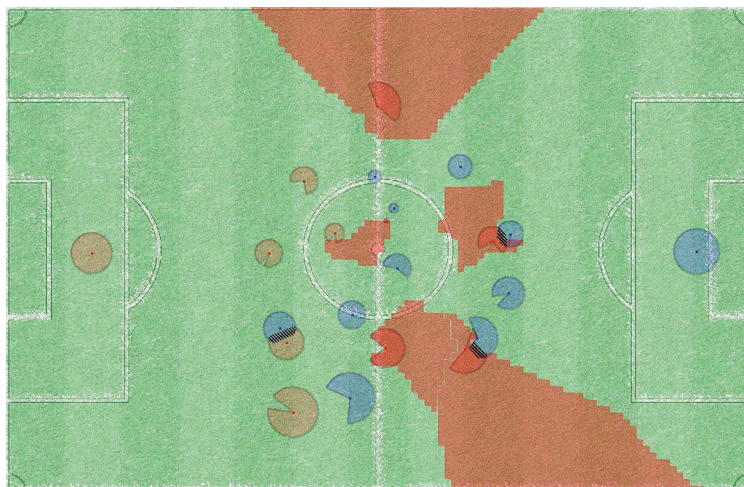
Figure 6. Grid-based free space visualization for five players on the red team. The red portions of the soccer pitch indicate the red team's areas of free space.

therefore also the length) of a connected row of ones is stored. Merging and comparing the stored values of each row, the largest contained rectangle is determined.
4. In the last step, we transform the grid-based rectangle back to the image space (see Figure 7b).

Figure 7c shows the result of this example. We enhance the notion of abstraction and the visibility using hatching.

Pass Options

In soccer, passes are an effective way to circumvent opposing players and advance to the opposing goal. However, the choice of a wrong pass option may inhibit the team's performance or lead to a loss of the ball. Passing is a complex decision-making process, often accompanied by opposing pressure. In a usually short time span, players must decide which other player to pass to, resulting either in a low-risk pass or a more insecure gaining space. Passes are influenced by many factors such as the correct moment in time with respect to the movements of the overall playing field as well as the strength, distance, precisions, and risk preferences of the involved players.

To improve overall game play, analysts explore potential pass options to reveal alternative tactics. Thus, we developed an analysis technique for pass alternatives for a given player in a given game situation. We then visualize possible pass alternatives with arrows pointing toward potential receiving players.

In theory, a player can pass to any of his 10 team mates at any time, but in practice the actual number is typically much lower, depending on the game situation. Based on our expert interviews, our assessment of pass alternatives includes several
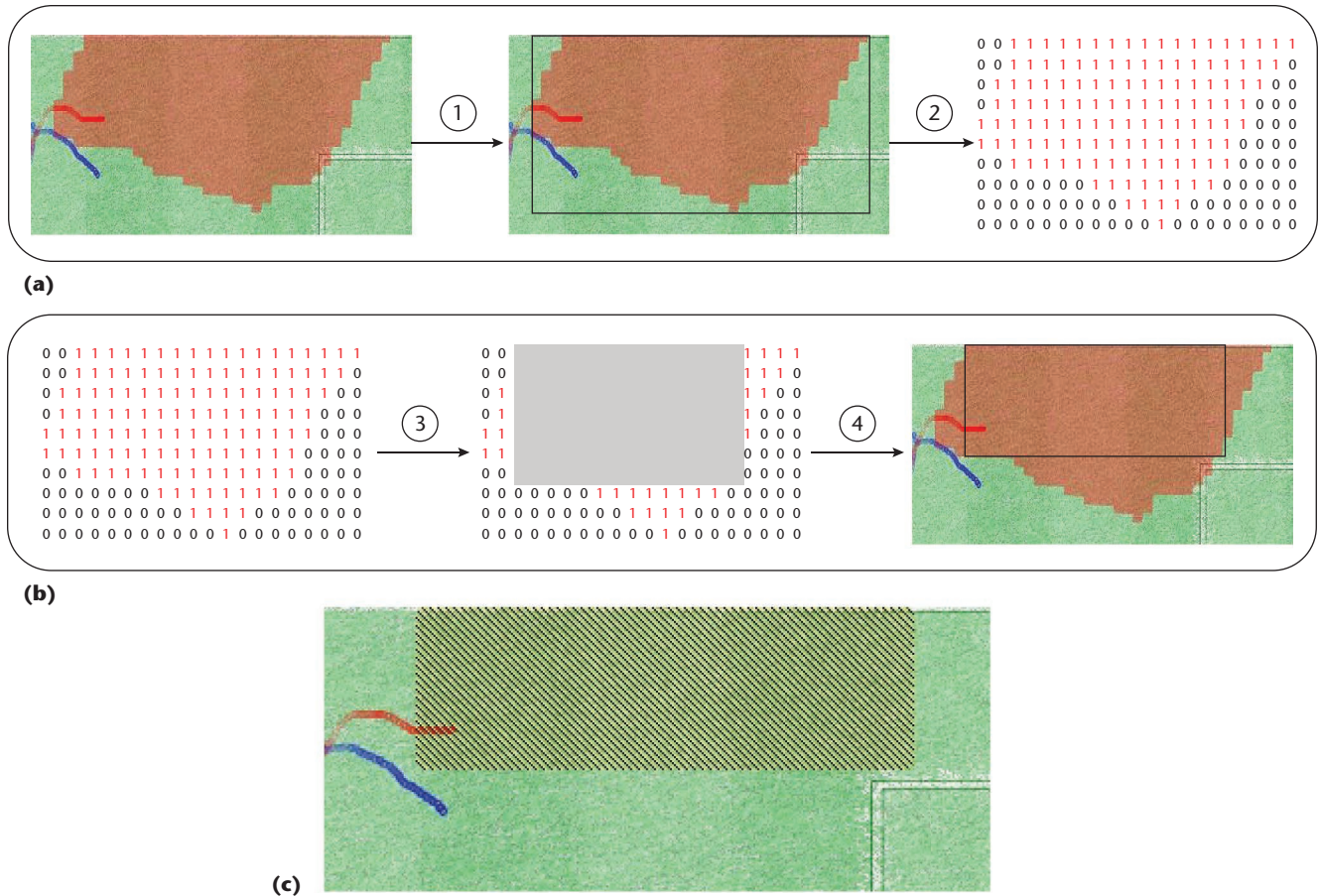
(a)

(b)

(c)

Figure 7. Abstraction algorithm example. (a) We first calculate the biggest rectangle within a matrix of numbers. (b) After following the steps of our abstraction algorithm, we transform the grid-based rectangle back to the image space. (c) In the final result, we enhance the notion of abstraction and the visibility using hatching.

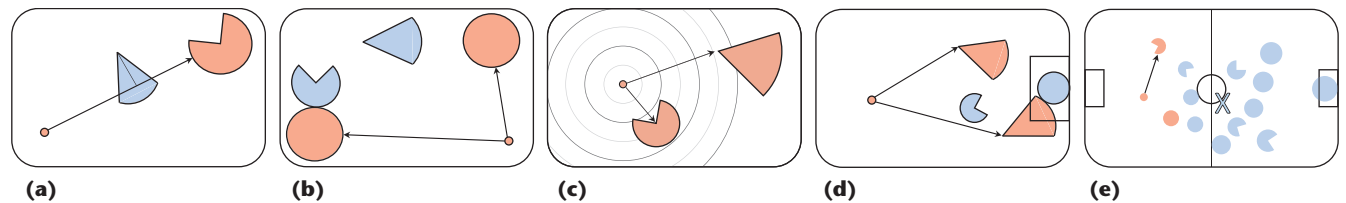**(a)**   **(b)**   **(c)**   **(d)**   **(e)**

Figure 8. Five criteria for the automatic assessment of passes based on interviews with subject matter experts: (a) pass intersecting opposing interaction space, (b) pressure for potential pass receiver, (c) pass distance, (d) pass in the direction of the opposing goal, and (e) pass away from a cluster of opposing team players. Our criteria focus on the risk involved with a certain pass alternative rather than on the quality of the pass.

criteria, which are displayed in Figure 8. We do not assess the quality of a pass but rather determine the risk involved for various alternatives. We enable the analyst to visualize the k highest rated pass options, where k is set by the user.

A low pass will usually be of higher risk if the pass intersects an opponent's interaction space (see Figure 8a). Lofted passes however are independent of intersecting interaction spaces because opposing players are unable to reach the ball. The closer an opposing player (described by his interaction space) is to the path of a low pass, the worse we

assess this pass. In the last section, we introduced levels of pressure that are the foundation for our second criterion. We calculate the pressure a potential pass receiver is experiencing and prefer pass alternatives with no or weak experienced pressure (see Figure 8b). Additionally, low passes should not cover too large a distance; longer passes increase the time the opposing team has to react. Furthermore, long-distance passes are usually less accurate. Consequently, we rate short-distance passes as more secure than long-distance low passes (see Figure 8c). Passes ending close to the opposing goals are

ranked higher than passes ending further away (see Figure 8d). The last criterion, which is reflected in Figure 8e, indicates that a pass should avoid clusters of opposing players. If there are, for example, more players in the right half of the soccer pitch, it would be beneficial to pass to the left half of the soccer pitch. The chances of losing the ball are higher in areas with many opposing players. If a player avoids these areas, he will force the opposing team to react, allowing his team to gain space.

Rule-Based Annotation

We developed a rule-based annotation feature to simplify match annotation and reduce the amount of time that is needed by automatically highlighting specific features. A rule specifies potentially interesting situations that are preselected by the system and presented to the user for adoption and/or refinement. Our goal was to simplify the manual annotation process. We enable analysts to define rules by selecting time ranges and corresponding annotation features. The results can be used for presentation or to illustrate findings to the coach and the team. The coach can revise or confirm his hypotheses and potentially reiterate through the annotation process. Our rule-based approach has the following advantages:

- *Modularity*. Our annotation system can be easily extended with further functionality by defining new rules.
- *Integration of expert knowledge*. Rules can be seen as externalized expert knowledge, directly integrated into our analysis system.
- *Adaptability*. Rules can be adjusted to the respective properties of a single match as matches differ greatly.

We support rule definition by analysts and provide a natural-language GUI. Analysts are able to define and modify their own rules without needing programming skills. In our system, a rule consists of three main components:

- **If**. In the first step, the video analyst specifies circumstances under which a rule should be applied. If one or all of the circumstances are fulfilled, the second (Then) part is automatically executed.
- **Then**. The second step defines which annotations are to be used on the soccer pitch. All examples introduced in the "Annotation of Key Game Elements" section can be used and adjusted to the analyst's needs. Visual feedback is used to foster the analyst's understanding of the new rule.

- **For**. The analyst decides for which team the newly defined rule shall be used.

Evaluation

In this section, we demonstrate the applicability of our proposed techniques by giving insight into the results of several of our quantitative and qualitative evaluations.

We invited two soccer experts to evaluate our designed free spaces. One expert has been an active soccer player for 24 years and has been working as a coach for 10 years. He currently works for the German soccer club FC Bayern München as a certified coach in the youth sector. A certified coach needs to be experienced in the theory and practice of video analysis. The other expert has been an active soccer player for 19 years and is now serving as an accredited referee. As an active soccer player, he regularly participates in briefings for his team, where video analysis is used to improve team performance. Hence, he is a well-suited candidate to evaluate our approach, given his encompassing professional experience.

We concluded our evaluation with a structured interview regarding the understanding and usefulness of our proposed system. We asked in detail about the expected impact that our proposed methods will have on the work of a professional soccer analyst and how they experienced the various techniques (interaction spaces, free spaces, pass options, and rule-based annotation). Lastly, we asked for suggested improvements.

Data

The data analyzed was provided from a collaboration with the sports analytics provider Prozone (http://prozonesports.stats.com). The dataset consists of 66 professional soccer matches. Time-stamped, 2D position data is available for each of the 22 players with a temporal resolution of 100 milliseconds. Furthermore, the data includes manually annotated events (such as fouls, passes, and crosses) with information about position, time, and event-specific information for the involved player. These events are less frequent and lack accuracy because they were manually tagged.

Pass Options

We implemented two heuristics to compute a player's pass options. The first heuristic h_1 assumes that the player always passes the ball to the nearest teammate and will be used as a lower boundary for our evaluation. Our second heuristic h_2 is the one we described earlier (in the first
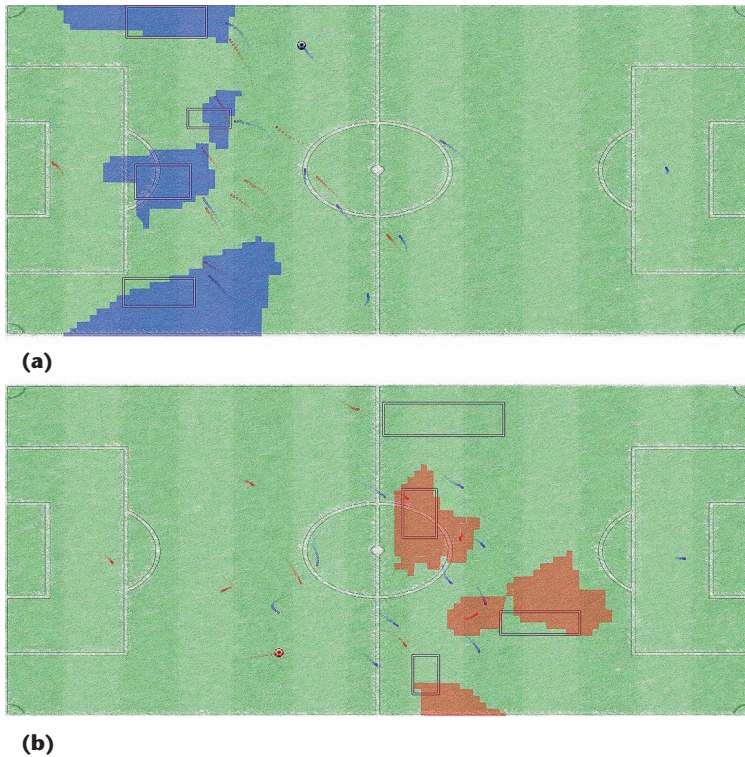
(a)

(b)

Figure 9. Evaluation of two example situations. The experts drew in the free space (rectangles). We then computed the best free spaces with our technique and overlaid them (blue and red shapes). We show both (a) the results that match the manual and automatic annotations as well as (b) the frequently missed free space pattern (top rectangle). In these cases, a player runs from the back freely over the sides.

Table 1. Number of true and false predictions for h_1 and h_2.

Heuristic		h_2	
	Prediction	True	False
h_1	True	302	70
	False	99	188

"Pass Options" section). Obviously, a player has up to 10 pass options, but many of them are not plausible or rational.

In this evaluation, as a default, we restricted ourselves to visualizing only the two most probable passes according to our model. Considering more pass options will automatically result in higher accuracy. We consequently reduce our accuracy of the predicted pass option, but simultaneously produce less visual clutter and thus increase the readability of our visualization.

In this evaluation, we analyzed 659 passes and compared the predictions of our heuristics with the players' actual passes. If the actual pass was one of our two most probable options, we counted this pass as a correct prediction. With this setup, we got the following results: a true positive rate

of 56.4 percent for h_1 and 60.8 percent for h_2. Because we propose only low-pass options and ignore lofted-pass options, we consequently achieve rates of around 60 percent. Both types of pass options are much better than random guessing, which would have a true positive rate of 1/10 + 9/10 · 1/9 = 20 percent.

We additionally can show that our heuristic h_2 is better than the simple heuristic h_1. For each pass, we assessed whether both h_1 and h_2 were correct, whether either one of them was correct, or whether both of them were wrong. Table 1 gives the number of occurrences for each case. Our null hypothesis is that there is no difference between h_1 and h_2. We calculated a p-value of 0.03125 with the McNemar Chi-Squared Test for Count Data. This p-value provides strong evidence that heuristic h_2 is significantly better than h_1.

Free Space

For our quantitative study, we conducted an evaluation where we showed the same 52 situations containing 204 free spaces from several soccer matches to each of our subject experts. We visualized each situation by showing the spatial locations of the 22 players and the ball. We asked our experts to draw the four most important free spaces for each situation. After the expert finished annotating each situation manually based on his knowledge, we automatically computed the free spaces using our method to assess our technique's accuracy.

We were also interested in understanding the characteristics of the manually detected free spaces that our method did not find. Figure 9 shows example results for situations with manually annotated free spaces combined with an overlay of our calculated free spaces.

The results of our expert study are promising. The first expert found 171 of the 204 free spaces, and the second expert identified 154. In 140 of the 204 cases, both experts detected a free space found by our technique as well. During our evaluation study, the domain experts drew 236 distinct free spaces overall. In 167 of these 236 cases, both experts marked the same free spaces. Consequently, the experts agreed for approximately 70 percent of the free spaces.

Evaluating our automatic free-space-detection algorithm, we counted 326 matches of manually drawn free spaces with the ones proposed algorithmically. As both experts together could match their annotation with ours in 408 cases, this results in an accuracy of 79.65 percent. As a consequence, we can see that our proposed method calculating and visualizing a team's free spaces

is both valid and valuable. Confirming our viewpoint with respect to the complexity of free-space detection, the experts agreed in only two-thirds of the cases. Nevertheless, we found free spaces in around 80 percent of the cases confirmed by at least one expert.

Discussion

Our work aims to support video analysts during their analysis of soccer matches by employing visual analytics techniques. The experts judged our system as very helpful for the detection, exploration, and comparison of interesting game situations. The greatest impact is expected in the area of processing and presenting findings within a match. According to the experts, our approach is based on the real position and movement data of each player and enables coaches to confirm or reject hypotheses using facts instead of intuition. Furthermore, the amount of time needed to analyze a single match is expected to decrease drastically. Our experts believe professional video analysts would make extensive use of a system designed with respect to their workflow. Our research is an important step toward the efficient analysis and explanation of game situations.

Our qualitative evaluation of interaction spaces showed that the computed interaction spaces correspond with the experts expectations. Nevertheless, they proposed improving our system by enabling a deeper analysis that allows a user to zoom into a single duel area and inspect particular movement behavior. Specifically, the experts wished for a visualization encoding where a player should move to minimize an opposing player's interaction space.

Our experts regarded the detection of interesting free spaces as very effective. Even hard to detect and not obvious free spaces were identified correctly in various sample situations. The experts approved both visualization approaches representing free spaces. The grid-based free-space visualization supports a more accurate way to investigate the free spaces as it reflects the computation. However, the rectangular abstraction indicates inherent uncertainties of the free-space computation. Consequently, the experts consider the abstract rectangular visualization as the more intuitive.

Our experts also appreciated the visual representation of pass alternatives. In addition, they wished it was possible to focus on specific kinds of passes, such as from the wings. In this work, we partially omitted movement while computing the potential pass options and assumed players would stop and wait for a pass. As a next step, we will in-tegrate movement heading and velocity to improve our computation.

The GUI for the rule-based annotation enabled the experts to define their hypotheses as a rule in our annotation system without the need to learn programming. They especially appreciated the ability to fine-tune rules with respect to conditions.

The quantitative evaluation of our pass options showed that our advanced heuristic achieves statistically significant better results than the simple heuristic proposing passes to the nearest player. Quantitative evaluating our free space detection showed that we are able to detect and rank free spaces with a precision of approximately 80 percent. We analyzed the evaluation results and found that a common pattern exists in most of our missed free spaces. In these cases, a player runs from the back freely over the sides (as Figure 9b illustrates) and the free space in front of him is not rated high enough to be displayed. Consequently, we plan to adjust our ranking algorithm to assign a higher weight to such patterns.

Our approach is based on the real position and movement data of each player and enables coaches to confirm or reject hypotheses using facts instead of intuition.

Our approach is a step toward semiautomatic annotation and support in the video analysis process for soccer experts. It can be extended in many directions, including interesting research questions. Future work includes transferring our methods from single to multimatch analysis. This would, for example, enable us to analyze set-plays in greater detail. In this scenario, we can apply our developed methods on many matches while, for example, comparing how players create free spaces. Following this line, we might define hypotheses about whether a player always takes advantage of existing free spaces or instead focuses on distracting the opposing defenders or determine in which situations the behavior would change. Currently, our system is based on a number of assumptions and models for computational assessment of interaction space, free space, and pass alternatives. By comparing players in a series of matches, we could personalize the used models for individual players, achieving higher accuracies for the predictions made. It would also be interesting to assess whether the models depend on the combinations

of players confronting each other or whether they are more static per player.

Furthermore, we see potential in looking for dissimilarities among different situations to help discover new and formerly unknown situational behaviors. Eventually, we want to integrate a learning phase that allows our system users to integrate their expert knowledge into the calculation and ranking used by our analysis techniques. An expert will manually mark what he perceives as a good pass and then our visual analytics system will extract all characteristic features and create the ranking computation accordingly.

So far, we have focused on soccer video analysis. However, our methodologies may be applicable to other application domains, including the analysis and annotation of animal movements (such as formulating hypotheses about swarm movement) or traffic situations (such as accident reconstruction). These domain-specific computational functions for assessing movement situations will help to greatly improve analysis in areas that in many cases still take place manually today.

References
1. J.-F. Grehaigne, D. Bouthier, and B. David, "Dynamic-System Analysis of Opponent Relationships in Collective Actions in Soccer," *J. Sports Sciences*, vol. 15, no. 2, 1997, pp. 137–149.
2. D. Vandevoorde, "The Maximal Rectangle Problem," *Dr. Dobbs J.*, 1 Apr. 1998; www.drdobbs.com/database/the-maximal-rectangleproblem/184410529.

Manuel Stein *is a PhD student in the Data Analysis and Visualization Group at the University of Konstanz. His research interests include movement analysis and team sports analytics, in particular pattern detection, decision-making processes, and context-aware visual analysis. Stein has an MS in computer science from the University of Konstanz. Contact him at manuel.stein@uni-konstanz.de.*

Halldór Janetzko *is a postdoctoral researcher in the Data Analysis and Visualization Group at the University of Konstanz. His research interests include temporal and geospatial phenomena with a special focus on collective movement. Janetzko has a PhD in computer science from the University of Konstanz. Contact him at halldor.janetzko@uni-konstanz.de.*

Thorsten Breitkreutz *is a master's student in the Data Analysis and Visualization Group at the University of Konstanz. His research interests include spatiotemporal data analysis and the effective integration of user interaction in visualization techniques. Breitkreutz has a BS in computer science from the University of Konstanz. Contact him at thorsten.breitkreutz@uni-konstanz.de.*

Daniel Seebacher *is a research associate and PhD student in the Data Analysis and Visualization Group at the University of Konstanz. His research interests include similarity search in heterogeneous high-dimensional data. Seebacher has an MS in computer science from the University of Konstanz. Contact him at daniel.seebacher@uni-konstanz.de.*

Tobias Schreck *is a professor with the Institute for Computer Graphics and Knowledge Visualization at the Graz University of Technology. His research interests include visual analytics and 3D object retrieval. Schreck has a PhD in computer science from the University of Konstanz. Contact him at tobias.schreck@cgv.tugraz.at.*

Michael Grossniklaus *is a professor in the Computer and Information Science Department at the University of Konstanz. His research interests include query processing and optimization techniques for graph databases and data streams. Grossniklaus has a PhD in computer science from ETH Zurich. Contact him at michael.grossniklaus @uni-konstanz.de.*

Iain D. Couzin *is the director of the Max Planck Institute for Ornithology, Department of Collective Behavior, and the chair of biodiversity and collective behavior at the University of Konstanz. His research aims to reveal fundamental principles that underlie evolved collective behavior, and consequently he works on a range of biological systems from insect swarms to fish schools and human crowds. Contact him at iain.couzin@uni-konstanz.de.*

Daniel A. Keim *is full professor and head of the Data Analysis and Visualization Research Group at the University of Konstanz's Computer Science Department. His research combines automated data analysis and interactive visualization techniques for the purpose of exploring large data in various applications domains. Keim has a PhD in computer science from the University of Munich. Contact him at daniel.keim@uni-konstanz.de.*