Director's Cut: Analysis and Annotation of Soccer Matches

Manuel Stein*, Halldór Janetzko*, Thorsten Breitkreutz*, Daniel Seebacher*, Tobias Schreck**, Michael Grossniklaus*, Iain Couzin*, Daniel A. Keim*

**University of Konstanz. Email: firstname.lastname@uni-konstanz.de*
***Graz University of Technology. Email: Tobias.Schreck@cgv.tugraz.at*

Abstract

For development and alignment of tactics and strategies, professional soccer analysts spend up to three working days manually analyzing and annotating professional soccer matches. Our aim is to improve soccer analysis by providing visual-interactive and data analysis support for annotating important types of soccer match elements, such as player interaction spaces, free spaces, and pass options. This article explores an enhanced workflow annotating key situations and features in soccer matches following the idea of a director's cut as known from movies. Tailored to the respective analysis task, our system allows to focus on key situations by rule-based filtering and pre-annotation. We evaluate our approach by analyzing real-world soccer matches and several expert studies. Quantitative measures show that our annotation methods for pass options and free spaces perform significantly better than naïve solutions. As a result, our methods lay the foundation for innumerable further and deeper analysis tasks.

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

1. Introduction

Soccer evolved from a minority sport towards world-wide mass sport. Today, the Fédération Internationale de Football Association (FIFA) contains more countries than the United Nations. With ongoing attraction and mass marketing, the structures of professional soccer clubs evolved over the decades. Modern soccer clubs can be regarded as corporate entities, with the soccer team and its successful operation at the center. Many auxiliary and infrastructure departments in clubs provide supportive functions such as promotion of young players (club development), medical treatment of players (performance maintenance and optimization), and game analysis (for development and alignment of tactics and strategies). The game analysis department is directly connected with the coaching team and employs video analysts. The task of these experts is to identify strengths and weaknesses of their own team and of opponents, both in retrospective of historic matches, and in anticipation of upcoming matches. Their findings are used to adjust the training and thereby raise the team's awareness for dangerous situations, preparing for matches.

For decades, video analysts used video recordings of matches, which are manually processed, annotated, and edited for analysis and presentation. With advances in sensor technology, it has recently become possible to actively track player position and event data of soccer matches with high temporal and spatial resolution. Furthermore, video cameras installed on the stadium roof additionally allow passive tracking of players. Depending on availability and regulation (e.g., FIFA disallows active tracking), either or both of these modalities can be used to capture match data. The now 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, e.g., on player perfor-

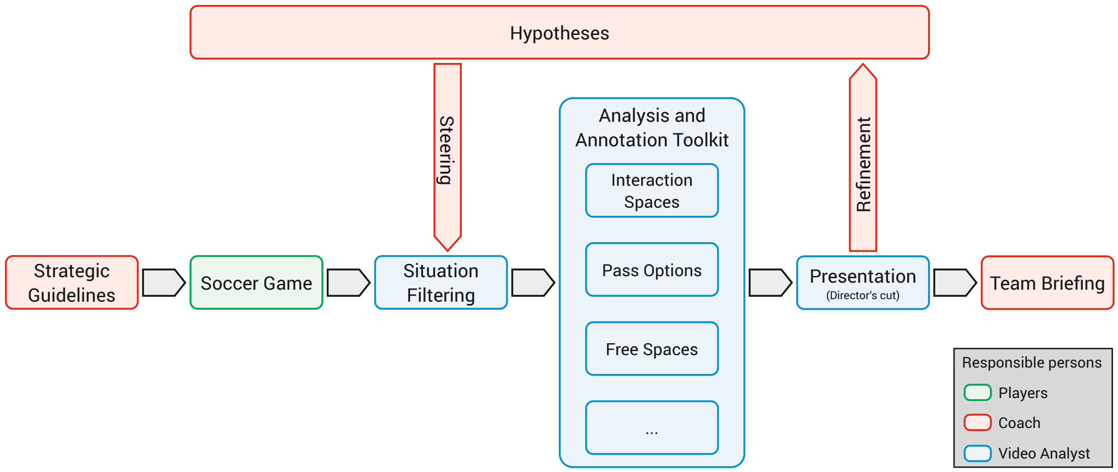


Figure 1: Typical analysis workflow in professional soccer clubs. Different colors show the respective responsible person for each step. The work of video analysts is driven by the needs and hypotheses of the team's coach.

mance. With motion and event data being readily available, one can ask how the traditional task of video analysts can be performed more efficiently, and which novel analyses can be supported.

In this article, we address the problem of how soccer motion data can be appropriately transformed and visually represented to support detection of interesting match patterns, and to assess players and groups of players. To address this goal, we focus on the analysis workflow of professional soccer clubs as depicted in Figure 1. Typically, the strategic guidelines of the coach are to be implemented by the players during the respective soccer match. The coach has hypotheses of how to improve the performance of his own team being reflected from the analysis process. Video analysts are told which kind of situations they should look for, and how they should visually prepare those situations for instructive playback. Based on such demonstration videos, the coach can either look for further insights or present the findings to his team. In our work, we aim to support video analysts during their analysis of specific aspects of soccer matches employing visual data analysis methods. We develop semi-automatic techniques integrating expert knowledge, which are used to cope with individual game annotation tasks. We reduce the workload by speeding up the filtering for specific situations by a rule-based selection engine. Rules are easily adjustable and result in a director's cut with automatic annotations still being editable by the video analyst. We use the term *director's cut* in analogy 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, namely 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.

2. Related Work

Our work relates to several strands of research. First, we relate to a body of works in visual and interactive data analysis. As the size and complexity of data is increasing, there is a growing need for effective tools to understand, search, and analyze such large data sets. Visual Analytics [7] is concerned with researching effective approaches for data analysis, relying on interactive visualization to present and navigate large data sets by appropriate visual structures. At the same time, the goal is to incorporate appropriate data analysis algorithms, ranging from data preprocessing and data reduction to pattern detection. Analysis of large data is a common problem in many areas, including business and finance, social media data, science, engineering, and many more. Owing to the advances in data sensor technologies, recently also in the sports domain, large amounts of data can be acquired, e.g., in form of player motion data or match event data.

Also recently, there has been an interest in visualization for analyzing sports data, as discussed, e.g., during the 2014 IEEE VIS Workshop on Sports Data Visualization [1]. Several systems have to date been proposed for visual analysis of sports data. Legg et al. [9] developed a visual analytics system to visually explore Rugby match data based on user searches and visualized relevant scenes on a rugby pitch diagram. A carefully crafted system for interactive visualization of soccer matches, based on annotated events and movement data was presented by Perin et al. [10]. There exists a large body of work on the definition of match- and sport-specific performance measures and their automatic computation. For example, Kim et al. [8] analyzed the defenders of a soccer team for their spatial formations during a game. Measures of player influence in a soccer game were explored by Fonseca et al. [2].

In this work, we are concerned with the visual annotation of players, movements, and formations in a soccer match. In that respect, we relate to works on movement analysis. Visual movement analysis was applied in many areas, some of which include indoor movement patterns of persons [5] as well as outdoor movement of pedestrians in a street network [11]. A visual analysis of large amounts of GPS trajectories of taxis was undertaken by Huang et al. [4]. For additional background on motion anlaysis, visualization and applications, please see the sidebar.

In previous work, we have investigated the analysis of soccer games based on low-level features extracted from motion data [6, 12]. Specifically, we formed feature vectors of player properties like speed, distance to ball, passing relationships between players, etc. These features were applied to segment different situations in a match, compare players for similarity, and classify interesting situations. However, the features used there were generic low-level movement features. In this work, we define a set of domain-specific mid- and high-level features for soccer analysis, based on concepts of area of influence and spatial relationships. Together with appropriate visualization techniques, soccer matches can be explored from multiple perspectives, and can be readily integrated into the above mentioned systems.

3. Annotation of Key Game Elements

We conducted several informal expert interviews with soccer video analysts to derive requirements as to which visual annotation functions would help. As listed in Figure 1, analysts pay particular attention to players' *interaction spaces*, arising *free spaces*, and possible *pass options* while assessing a soccer situation. In the next paragraphs, we will describe the proposed advanced annotation capabilities in detail, which were designed and implemented as an outcome regarding the expert design process. We additionally provide several videos of our results (`http://files.dbvis.de/stein/Interaction_Space`, `http://files.dbvis.de/stein/Free_Space`, `http://files.dbvis.de/stein/Pass_Alternatives`).

3.1. 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 interaction space. We introduce a model consisting of several factors determining the notion of interaction space in Section 3.1.1.

Interaction spaces are especially interesting when analyzing passes or their reception. Furthermore, player duels are directly related to interaction spaces. Other indepth analyses such as the detection and assessment of free spaces in Section 3.2 or the calculation of pass options in Section 3.3 build on the calculation of interaction spaces.

3.1.1. A Model for Determining Interaction Space

The foundation for our calculation of interaction spaces was conceptualized by Grehaigne et al. [3]. By performing several experiments they measured how the interaction space of a player morphs according to different velocities. Interestingly, small changes in speed are enough to drastically change the interaction space from a circle

Applications of Movement Analysis

Movement data is a ubiquitously relevant type of data arising in many application domains. Apart from sports, as studied in this paper, it is extremely relevant to any sort of flow analysis including traffic, logistics, etc. Motion can occur on many scales. For example, on the astronomic scale the expansion of the universe is studied. On a middle-scale, e.g., the movement of animals or pedestrians, either in isolation or within a group of objects (e.g., herd, swarm) is of interest. Complex interrelationships may be found and studied across time and space. Also, motion happens 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 (e.g., aggregation and simplification of motion), correlation and comparison of movements, and classification of movement patterns. Data analysis techniques have been studied for many of these tasks, and visual representations can help to provide for effective user analysis for large trajectory data[a].

In addition to sports analysis, many further 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 scatter plots[b]. In another work, features of trajectories in animated scatter plots are computed and used for detection and segmentation of interesting sub intervals, based on outlying motion features[c]. In general, it is interesting to analyze group movement data to distinguish different roles of the moving member elements[d], or to correlate movement patterns with location properties[e]. It is also interesting to ask if one can come up with a taxonomy of movement patterns[f].

[a] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel: *Visual Analytics of Movement.* Springer, 2013

[b] Gapminder. http://www.gapminder.org/

[c] T. von Landesberger, S. Bremm, T. Schreck, and D. Fellner. *Feature-based automatic identication of interesting data segments in group movement data.* Information Visualization, 13(3):190-212, 2014.

[d] M. Andersson, J. Gudmundsson, P. Laube, and T. Wolle. *Reporting leaders and followers among trajectories of moving point objects. GeoInformatica 2008; 12(4): 497-528.*

[e] G. Andrienko, N. Andrienko, M. Mladenov, M. Mock and C. Poelitz. *Discovering bits of place histories from people's activity traces*. In: Proceedings of IEEE Symposium on Visual Analytics Science and Technology, pp. 59-66.

[f] S. Dodge, R. Weibel A. Lautenschuetz. *Towards a taxonomy of movement patterns*. Information Visualization 2008; 7(3-4): 240-252

Sidebar on Analysis Methods for Movement Data to a circular sector. We extend this discrete model and derive a continuous model by using polynomial interpolation allowing a smooth transition of interaction spaces as depicted in Equation 1.

$$angle(v) = -0.0038\pi \cdot v^3 + 0.0793\pi \cdot v^2 - 0.6108\pi \cdot v + 2\pi \quad (1)$$

Furthermore, we enhance their model by integrating the distance between player and ball reflecting that a player who is further away from the ball has more time to react described by R in Equation 2. Figure 2 illustrates our approach with different velocities and distances between ball and player. The maximum time of ten seconds will occur when ball and player are located in opposite corners of the soccer pitch.

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

$$\begin{aligned} \text{with } radius &: \text{radius of interaction space} \\ v &: \text{speed of player} \\ R &: \text{time until ball reaches player} \\ \bar{v} &: \text{average speed of player} \end{aligned} \quad (2)$$

When a player is next to players of the opposing team, his movement is restricted and he needs to pass them by. Furthermore, opposing players will try to block him. Interaction spaces have to reflect these interdependencies as shown in Figure 3. The interaction space of a player is restricted to the area that he can reach before opposing players. The intersection of interaction spaces is computed and used to visualize the corresponding restricted interaction spaces.

3.1.2. *Potential Duel Areas*

When two interaction spaces overlap, it is not clear who of the 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 by hatching as depicted in Figure 4. Hatched areas can be described as locations being potentially reached by both players simultaneously. Consequently, we first intersect both players' interaction spaces

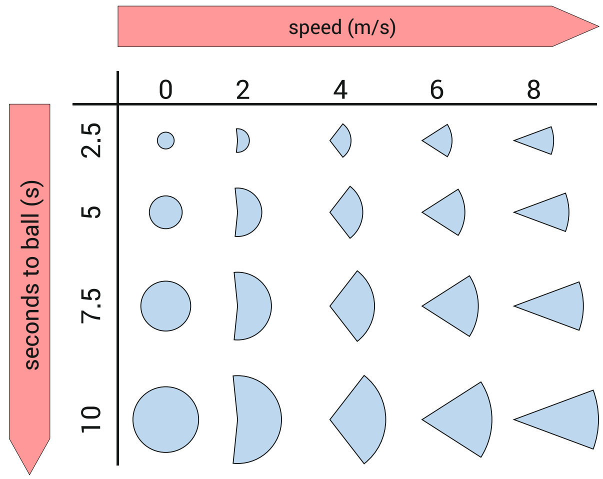


Figure 2: Speed and distance between player and ball influence the calculation of interaction spaces. The faster a player moves forward (to the right of this figure) the less possible are sudden changes in heading.

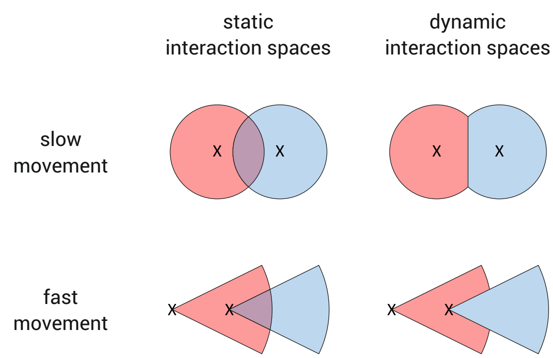


Figure 3: Interaction spaces are influenced by adjacent players.

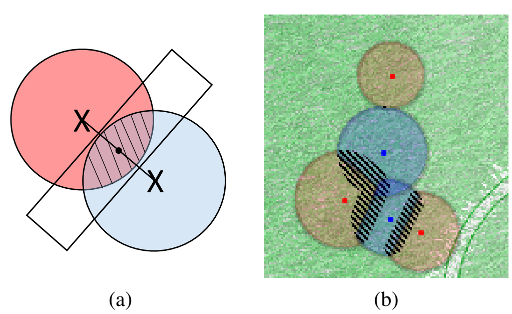
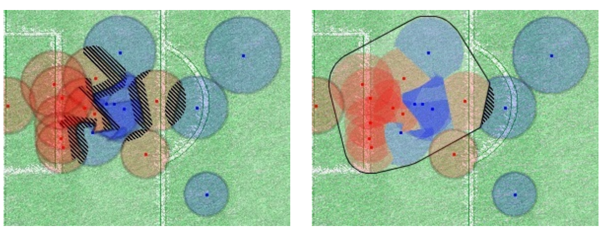


Figure 4: Potential duel area of two players visualized by hatching.

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 visual clutter of presenting too many nearby duel areas by unifying these, showing a contour line. We show our aggregation-based approach in Figure 5. In these surrounded areas, we use transparent colors and depict the density of players. We assume that intense coloring depicts a higher possibility for one team winning the ball. We apply an adjusted density-based clustering method (DBSCAN) to detect players being close to each other. The clustering method needs as input parameters 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.



(a) Many overlapping duel areas (b) Hull-based visualization

Figure 5: We provide an optional aggregation-based visualization avoiding visual clutter in dense regions.

3.2. Free Space

Free space is a very interesting, important feature of soccer. 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. Furthermore, actions in soccer are usually centered around the ball and free spaces are of different importance. 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.

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 for how much a player is put under pressure. Pressing will hinder a player to move freely around and interact with the ball. We categorize pressing into three classes:

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

Weak pressure. The player is already being targeted by an opponent moving towards him. The player has less time to act.

Strong pressure. The opponent is close to the player and tries to get the ball. The player has nearly no opportunity to act proactively as 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 follow both ideas of the free space relevance as well as the amount of pressure a player is experiencing. We assess detected free spaces by the respective size, the amount 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 one 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 is able to perceive an overview about the spatial distribution of both teams. The analysis can additionally be steered into the direction of player behavior, e.g., when inspecting zonal defense. Figure 6 illustrates our free space visualization.

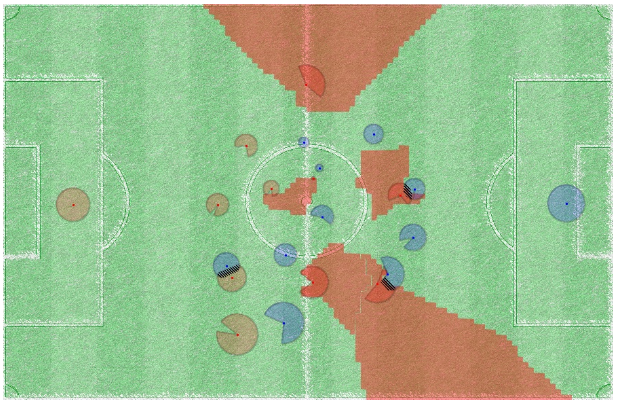


Figure 6: Grid-based free space visualization for five players of the red team.

As the grid-based free space is possibly too detailed for our experts and furthermore might suggests too high a level of accuracy, we introduce a rectangular abstraction of free spaces. For the calculation of the abstract free space, we adapt an existing algorithm, which detects the largest rectangle within a character matrix and has a runtime of $O(M \cdot N)$. We adjust the algorithm to accept a number matrix as input. See the sidebar for more information about how we calculate the abstraction.

3.3. Pass Options

In soccer, passes are an effective way to circumvent opposing players and advancing to the opposing goal. However, the choice of wrong pass options may inhibit the team's performance or lead to a loss of the ball, e.g., directly in front of the own goal. To improve overall gameplay, analysts explore potential pass options revealing alternative tactics. Passing is a complex decision making process, often accompanied by opposing pressure. In a usually short time span, players have to decide to which other player to pass, 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 playfield as well as the strength, distance, precisions, and risk preferences of the involved players. We developed an analysis technique for pass alternatives, detailed below, for a given player in a given game situation. Possible pass alternatives are then visualized with arrows pointing towards the potential receiving players.

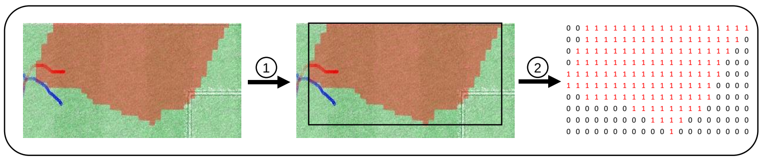
In theory, a player can pass to any of the ten team mates at any time whereas in practice the actual number is typically much lower, depending on the game situation. Our assessment of pass alternatives includes several criteria displayed in Figure 7 resulting from expert interviews. We do not assess the quality of a pass but rather the risk involved as players presumably choose safer alternatives. We enable the analyst to visualize the k highest rated pass options with k being set by the user.

In the following paragraph, we will describe all assessment criteria in detail. A low pass will usually be of higher risk if the pass intersects the interaction space of an opponent as depicted in Figure 7 (a). Lofted passes however are independent of intersecting interaction spaces as 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 Section 3.2, we introduced levels of pressure being 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 as shown in Figure 7 (b). Additionally, low passes should not cover too large a distance, because otherwise the time the opposing team has to react increases. Furthermore, long distance passes are usually less accurate. Consequently, we rate short distance passes more secure than long distance low passes as illustrated in Figure 7 (c). Passes ending close to the opposing goals are ranked higher than passes ending further away as displayed in Figure 7 (d). The last criterion is reflected in Figure 7 (e) and describes 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 will be beneficial to pass to the left half of the soccer pitch. 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.

4. Rule-Based Annotation

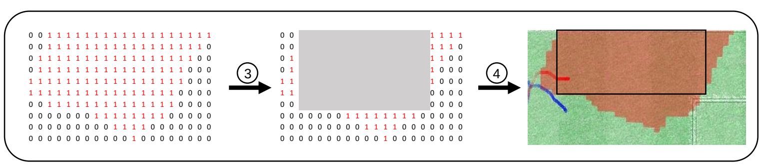
We developed a rule-based annotation feature to simplify match annotation and reduce the amount of time that is needed by automatically highlighting specific features. A rule specifies potentially interesting situations

Abstracting free spaces

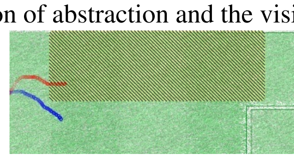


This example shows how the biggest rectangle within a matrix of numbers can be calculated applying the technique of Vandevoorde[a].

1. Our abstraction starts with a grid-based representation of the available free space of a particular player. We determine the minimum bounding rectangle containing this representation.
2. A numerical, zero-initialized matrix is filled with ones in correspondence to the free space. Consequently, the result is an array approximating the shape of the free space.
3. The algorithm by Vandervoorde works in a scanline fashion and processes the matrix row by row. In each row, the start and end point (and 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. As a last step, we transform the grid-based rectangle back to the image space.

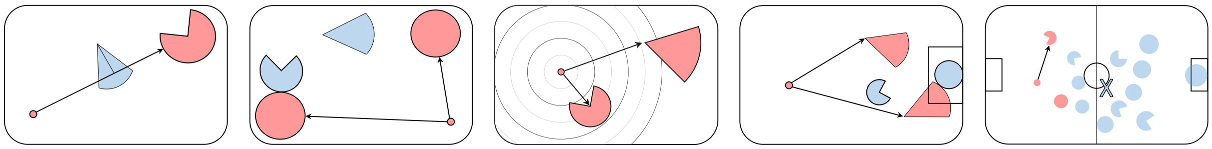


Below we see the result of the previously introduced example. We enhance the notion of abstraction and the visibility by hatching.



[a] D. Vandevoorde: *The maximal rectangle problem.* `http://www.drdobbs.com/database/the-maximal-rectangle-problem/184410529`, 1998 [accessed April 26, 2016]

Sidebar on the Abstraction of Free Spaces



(a) Pass intersecting opposing interaction space (b) Pressure for potential pass receiver (c) Distance of a pass (d) Pass in direction of opposing goal (e) Pass away from cluster of opposing team

Figure 7: Five criteria for the automatic assessment of passes based on interviews with subject matter experts. Our criteria focus on the risk involved with a certain pass alternative and not on the quality.

which are pre-selected by the system and presented to the user for adoption and/or refinement, meant to simplify the manual annotation process. We enable analysts to define rules for 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 easily be 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 highly.

We support rule definition by analysts and provide a graphical user interface based on natural language. Analysts are able to define and modify their own rules without the need of programming skills. In our system, a rule consists of three main components:

If. In the first step, the video analyst specifies under which circumstances the rule should be applied. If one or all of the circumstances are fulfilled the **Then** part is automatically executed. An example is “If there exists a very well rated free space near the opponents goal...”.

Then. This part defines which annotations are to be used on the soccer pitch. All examples introduced in Section 3 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.

5. Evaluation

In this section, we demonstrate the applicability of our proposed techniques by giving insight into the results of several quantitative and qualitative evaluations we conducted. For the evaluation of our designed free spaces, we invited two soccer experts.

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 theory and practice of video analysis. The other expert has been an active soccer player for 19 years and is now serving as accredited referee. As an active soccer player, he regularly participated in briefings of his team, where video analysis was used to improve team performance. Hence, he too was 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. In detail, we asked for the expected impact that our introduced methods will have on the work of a professional soccer analyst and how the various techniques (interaction spaces, free spaces, pass options, and rule-based annotation) were experienced. At last, we asked about suggestions for improvement. We will describe the conclusions we drew from these interviews in Section 6.

5.1. Data

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

5.2. Pass Options

We implemented two heuristics to compute the pass options of a player. 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 described in Section 3.3. Obviously, a player has up to ten pass options with many of them being not plausible/rational.

In this evaluation, we restricted ourselves per default to visualize only the two most probable 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 increase the readability of our visualization with less visual clutter.

In this evaluation, we analyze 659 passes and compare the predictions of our heuristics with the actual passes of the players. If the actual pass is one of our two most probable options, we will count this pass as a correct prediction and otherwise, as a wrong one. With this setup, we get the following results: a true positive rate of 56.4% for h_1 and 60.8% for h_2. As we propose only low pass options and ignore lofted pass options, we consequently achieve these rates around 60%. Both are much better than random guessing which would have a true positive rate of $1/10 + 9/10 \cdot 1/9 = 20\%$.

We additionally can show that our heuristic h_2 is better than the simple heuristic h_1. For each pass, we asses if both h_1 and h_2 were correct, whether either one of them was correct, or whether both of them were wrong. The number of occurrences for each case are displayed in the contingency Table 1. Our null hypothesis is that there exists 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.

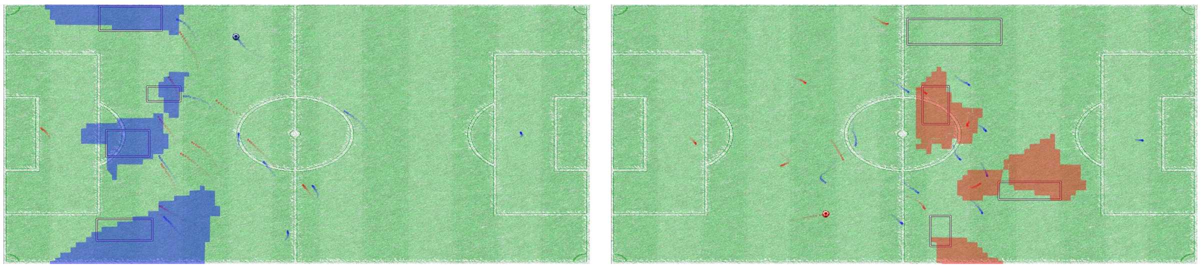
		h_2 true	h_2 false
h_1	true	302	70
	false	99	188

Table 1: Contingency table of the number of true and false predictions of h_1 and h_2

5.3. Free Space

The evaluation of our introduced free spaces was performed with the help of our previously introduced subject matter experts. 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 the experts. Each situation was visualized by showing the spatial locations of the 22 players and the ball. We gave our experts the task 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 with our method from Section 3.2. Consequently, we can assess the accuracy of our technique. We are furthermore interested in understanding the characteristics of manually detected free spaces that our method did not find. Figure 8 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 very promising. The first expert found 171 of the 204 free spaces while the second expert identified 154. In 140 of 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, both experts agreed among themselves in approximately 70% 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



(a) Matching manual and automatic annotation

(b) Missed free space (top)

Figure 8: Evaluation result of two example situations where the expert had to draw in the free space (white rectangles). Afterwards, we computed the best free spaces with our technique and overlaid them (blue and red shapes). The left figure shows a complete match of manual and automatic annotation. The right figure characterizes a frequently missed free space pattern. In these cases, someone runs from the back freely over the sides.

79.65%. As a consequence, our introduced method calculating and visualizing free spaces of a team can be seen as valid and valuable. Confirming our viewpoint with respect to the complexity of free space detection, experts do agree among each other only in two thirds of the cases. Nevertheless, we find free spaces in around 80% of the cases confirmed by at least one expert.

6. Discussion

Our work aims to support video analysts during their analysis of soccer matches by employing Visual Analytics techniques. The experts judge 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 by facts instead of intuitions. Furthermore, the amount of time that is needed for the analysis of 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 as depicted in Figure 1. Our research is seen as an important step towards 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 to improve our system by enabling a deeper analysis by zooming into a single duel area and inspect particular movement behavior. Specifically, the experts wished for a visualization encoding where a player should move to minimizing the interaction space of an opposing player.

The detection of interesting free spaces was regarded very effective by our experts. Even hard to detect and not obvious free spaces have been 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 the more intuitive one.

The visual representation of pass alternatives was appreciated by our experts. As an addition, they wished for the possibility to focus on specific kinds of passes, for example from wings. In this work, we partially omitted movement while computing the potential pass options and assumed players to stop and wait for the pass. As a next step, we will integrate movement heading and velocity to improve our computation.

The graphical user interface for the rule-based annotation enabled the experts to define their hypotheses as a rule in our annotation system without the need to learn to program. Especially, the possibility to fine-tune rules with respect to conditions was positively mentioned.

We conducted a quantitative evaluation for the pass options and the free spaces. The 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. Evaluating our free space detection showed that we are able to detect and rank free spaces with a precision of approximately 80%. We analyzed the evaluation results and found that there exists a common pattern in most of our missed free spaces. In these cases, someone runs from the back freely over the sides as displayed in Figure 8 (b) and the free space in front of him is not rated high enough to be displayed. Consequently, we plan to adjust our ranking algorithm in order to assign a higher weight to such patterns.

Our approach is a step towards semi-automatic 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 multi-match 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 free spaces are created by players. Following these lines, we might define hypotheses whether a player always takes advantage of existing free spaces, or instead rather focuses on distracting the opposing defenders, or 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. With player comparison along a series of matches, we may personalize the used models per player, achieving higher accuracies for the predictions made. It would also be interesting to assess whether the models also 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 support the discovery of new and formerly unknown situational behaviors. Eventually, we want to integrate a learning phase allowing users of our system 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 afterwards our Visual Analytics system extracts all characteristic features and creates the ranking computation accordingly.

So far, we focused on soccer video analysis, an important field of application. However, our methodologies may be applicable to other application domains such as analysis and annotation of animal movements (e.g., formulating hypotheses about swarm movement) or traffic situations (e.g., accident reconstruction). There, domain-specific computational functions for assessing movement situations will help to greatly improve analysis, which in many cases still happens manually today.

7. Conclusion

We presented a Visual Analytics system supporting common analysis tasks of video analysts in professional soccer clubs. Therefore, we enhanced the way analysts are looking at data of soccer matches instead of a pure manual and time-consuming video editing routine. In this work, we introduce our system laying analysis foundations by providing concrete measures for the most essential pillars of match analysis: interaction spaces, free spaces, and pass options. Additionally, we provided insights how such a system can be controlled by domain experts using a rule-based annotation framework. We evaluated our approach in several quantitative and qualitative studies, with and without subject matter experts, and discussed a range of possible extensions as well as interesting future work directions.

Acknowledgment

The authors wish to thank the experts for their valuable feedback and discussions. The soccer data used in this publication were generously provided by Prozone.

References

[1] R. Basole, E. Clarkson, A. Cox, C. Healey, J. Stasko, and C. Stolper (Organizers). First IEEE visworkshop on sports data visualization, Oct. 14, 2013.

[2] Sofia Fonseca, João Milho, Bruno Travassos, Duarte Araújo, and António Lopes. Measuring spatial interaction behavior in team sports using superimposed voronoi diagrams. *International Journal of Performance Analysis in Sport*, 13(1):179–189, 2013.

[3] Jean-Francis Grehaigne, Daniel Bouthier, and Bernard David. Dynamic-system analysis of opponent relationships in collective actions in soccer. *Journal of Sports Sciences*, 15(2):137–149, 1997.

[4] Xiaoke Huang, Ye Zhao, Jing Yang, Chao Ma, Chong Zhang, and Xinyue Ye. Trajgraph: A graph-based visual analytics approach to studying urban network centrality using taxi trajectory data. *IEEE Comput. Graph. Appl.*, 22(1):160–169, 2015.

[5] Yuri Ivanov, Christopher Wren, Alexander Sorokin, and Ishwinder Kaur. Visualizing the history of living spaces. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1153–1160, 2007.

[6] H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. Keim, and O. Deussen. Feature-driven visual analytics of soccer data. In *Proc. IEEE Conference on Visual Analytics Science and Technology*, pages 13–22, 2014. Peer-reviewed full paper.

[7] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, editors. *Mastering The Information Age - Solving Problems with Visual Analytics*. Eurographics, 2010.

[8] Ho-Chul Kim, Oje Kwon, and Ki-Joune Li. Spatial and spatiotemporal analysis of soccer. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 385–388. ACM, 2011.

[9] Philip A. Legg, David H. S. Chung, Matthew L. Parry, Rhodri Bown, Mark W. Jones, Iwan W. Griffiths, and Min Chen. Transformation of an uncertain video search pipeline to a sketch-based visual analytics loop. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2109–2118, 2013.

[10] Charles Perin, Romain Vuillemot, Jean-Daniel Fekete, et al. Soccerstories: A kick-off for visual soccer analysis. *IEEE transactions on visualization and computer graphics*, 2013.

[11] T. Schreck, I. Omer, P. Bak, and Y. Lerman. A visual analytics approach for assessing pedestrian friendliness of urban environments. In *Springer Lecture Notes in Geoinformation and Cartography (Proc. AGILE International Conference on Geographic Information Science)*, pages 353–368, 2013. Peer-reviewed full paper.

[12] M. Stein, J. Häußler, D. Jäckle, H. Janetzko, T. Schreck, and D. Keim. Visual soccer analytics: Understanding the characteristics of collective team movement based on feature-driven analysis and abstraction. *Int. Journal of Geo-Information*, 4(4):2159–2184, 2015.

Manuel Stein received his M.Sc. in computer science from the University of Konstanz (2014). He is a PhD student at the Data Analysis and Visualization group at the University of Konstanz. His research interests are in the area of movement analysis and team sport analytics considering, for example, pattern detection, decision making processes as well as context aware visual analysis.

Halldór Janetzko received his B.Sc. (2008), M.Sc. (2010), and PhD (2015) in computer science from the University of Konstanz. Halldór Janetzko works currently as a post doctorate researcher in the Data Analysis and Visualization group at the University of Konstanz. His research interests are temporal and geospatial phenomena with special focus on collective movement.

Thorsten Breitkreutz received his B.Sc. (2015) from the University of Konstanz. He is currently working on his Master at the Data Analysis and Visualization group at the University of Konstanz. His research interests focus on spatio-temporal data analysis and the effective integration of user-interaction in visualization techniques.

Daniel Seebacher is a research associate and PhD student at the Data Analysis and Visualization group of Daniel Keim at the University of Konstanz since 2016. His research interests include similarity search in heterogeneous high-dimensional data. He received a M.Sc. in Computer Science from the University of Konstanz in 2015.

Tobias Schreck is a professor with the Institute for Computer Graphics and Knowledge Visualization at Graz University of Technology. Before, he was an Assistant Professor with the University of Konstanz and a Postdoc fellow with TU Darmstadt. He obtained a PhD in Computer Science in 2006 from the University of Konstanz. His research interests include Visual Analytics and 3D Object Retrieval.

Michael Grossniklaus is a professor for Databases and Information Systems at the Computer and Information Science Department of the University of Konstanz. He obtained his doctorate in computer science from ETH Zurich. His research focuses on query processing and optimization techniques for graph databases and data streams.

Iain D. Couzin is Director of the Max Planck Institute for Ornithology, Department of Collective Behaviour, and Chair of Biodiversity and Collective Behaviour, at the University of Konstanz, Germany. 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.

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