An Analysis of Football Players Shooting Habit by Using Data Visualization Technology

Hafez Eslami Manoochehri, Runze Zhang
School of Computer Science
University of Texas at Dallas
Dallas, Texas
Email: hafez.eslamimanoochehri@utdallas, runze.zhang@utdallas.edu

***Abstract*—In this paper, we study the shoot habit of a soccer player through visualization techniques. We proposed a new visualization model for a soccer game in which we can analyse the likelihood of shot from the field and the likelihood of shot target. We categorizes each shot as: 1) off target, 2) saved and 3) scored. We further demonstrate a visualization method for analysis of player's habit, specially for the balls in target.**

Keywords: Soccer Visualization, Player Shooting Habit

I. Introduction

Association Football, more commonly known as football or soccer, is a team sport played between two teams of eleven players with a spherical ball. It is played by 250 million players in over 200 countries and dependencies, making it the world's most popular sport. The game is played on a rectangular field with a goal at each end. The object of the game is to score by getting the ball into the opposing goal [1]. A soccer field is divided into areas, visible with white landmarks: a center circle, penalty areas, corner quarter-circles. Games last 90 minutes divided in two half periods, and the winning team is the one that scores more goals than the other. Players have roles (goalkeeper, fullback, midfielder and forward) indicating their theoretical position on the field, but the actual position is adjusted according to tactics (e. g., the most commonly used tactical line-up is called $4-4-2$: 4 defenders, 4 midfielders, and 2 forwards).

The use of professional data in soccer has been rapid growth recently. Start from the late 1990s, the use of video and data to inform performance began to gain traction in the sport. Pioneered by Prozone, systems to support more detailed pre and post-match analysis gradually emerged, Derby County being the first UK club to adopt the technology in 1998. Within a few short years there had been a paradigm shift, with the majority of clubs at the top level of European football adopting structured analysis processes and dedicated performance analysts. Where once clubs had to rely on grainy VHS footage to conduct even the most basic analysis, they were now able to access high-quality footage and a wealth of accompanying statistics at the click of a button. In the modern game, the use of data is all about finding the extra 1%, the detail that can exploit even the slightest weakness in the opposition and make the difference between winning and losing. Rather than simply applying data to tactical performance, objective information is now used throughout the clubs to enhance efficiency and develop processes that enable the organization to be as well-prepared as possible from the pitch to the boardroom. Today, the data can be used to better define playing styles, quantify player abilities, to measure passing vision and reduce injury risk [2].

Goals is the most important part in football matches, many data types will be factors to goal. Squawka is one of the worlds fastest growing digital media and entertainment businesses in football. Squawka's proprietary technology visualises millions of data points beautifully to help fans understand the game they love. The data and graphics then also fuel cutting edge editorial and social media content - a unique proposition in a sport that is consumed by billions around the world. According to the research of related works. There are still not have property football data analysis team have produce the graphic to show the relationship with shoot location and goals results. What we try to do is to use data visualization technology to make some data graphics to analysis player's habits about shoot location. This results will help football club improve their training levels and make specific plan before meet different teams.

One of the challenges to visualize shot data is to how present different analysis in only one 3D image. In this paper, we tried to use different object colors, line colors and shapes and colormap surfaces to distinguish different analysis. Our contributions can be summarized as follows:

- We used some scaling methods to make it able to present both soccer field data and target space data in one image.
- We present the likelihood of shots made in the field by a color map. Also the likelihood of the target (end location of the ball) is shown by a surface in Z space. In that space, we illustrate different results of the shots by different colors.
- To better analyse the shoot habit, we visualize the connection of the start location of the ball shot by a player to the end location of the ball. Further, we distinguished the end location of the ball in six regions, if the ball is on target.

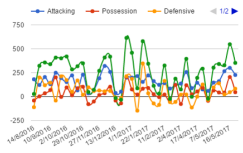
The rest of the paper is organized as follows. Section II discuss some related work. In section III, we state how we

gather the data for this work. Our approach is discussed in section IV. Section V present the outcome of our approach and finally conclusion is presented in section VI.

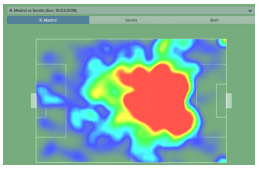
II. Related Work

Here in this section, we briefly present the related works in the field of soccer data visualization. There are some data visualization technology implemented in football field. By using these technology, coach can see the results directly instead of finding the corresponding in the huge statistic data. Here are some recent data visualization works in soccer area.

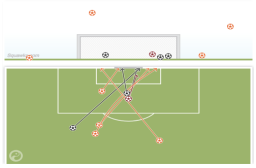
(a) Shots Analysis.



(b) Performance Score



(c) Match Heat Map



(d) Match Statistics

Fig. 1: Illustration of various images. (a) The graphic to show the goals location in each match. White means the home team made, red means the away team made. (b) By evaluating the performance of the whole match, the data point about the attacking, possession, defensive will be added separately. (c) The match heat map. Action Heat map displays the areas of the pitch where on-ball events most frequently happen. Action Areas Highlights where on-ball events happen on the pitch according to Squawkas 13 action areas. (d) Some data statistical comparison between two teams. Possessions, passing success percentage, shoot numbers are always the data which coach will be interested in.

Soccer Stories, a system for the visual exploration of soccer phases [1]. It uses a series of soccer-related visualizations, called faceted views, that they selected-after a thorough review of current visualizations for soccer data-for each group of actions within a phase, connected and ordered on a soccer field. It provide design guidelines for each visualization and their layout in a spatio-temporal flow. The team evaluated Soccer Stories for exploring and communicating findings about games that would enrich experts in tactics current reports on blogs or newspapers. they also explored design alternatives for embedding soccer phases into word-sized graphics. In these experiments they received a very enthusiastic feedback and participants consider further use of Soccer Stories to enhance their current workflow. Soccer Stories tackles the challenge of using visualization for complex data in an application domain which so far has been dominated by game statistics (i.e. numbers). Its contributions are the following: A review of existing soccer visualizations and their matching with soccer's most important actions and phases; Soccer Stories, a system that combines several visualizations to explore and communicate game phases; Two qualitative experiments that validated Soccer Stories for exploring and reporting on tactics analyses, and the design alternatives of compact representation of games as word-sized graphics to embed in articles [1]. The dimensionality of soccer statistics relating to goalkeeper specific data can be difficult to interpret [2]. Leveraging our existing statistical analytics tool, Soccer Scoop, they developed a goalkeeper visualization add-on that can assist a team manager. With the newly developed goalkeeper visualization tool, a team manager can compare a single goalkeeper between two games, measure the overall performance of the goalkeeper both for games played at home or away, as well as to devise the appropriate training exercises needed to strengthen any visible weakness. To keep with the continuity of the visualization styles of our first tool, Soccer Scoop, the goalkeeper visualization applies similar techniques, such as glyphs, details on demand, color, and Gestalt principles. They present a system for analyzing high-frequency position-based soccer data at various levels of detail, allowing to interactively explore and analyze for movement features and game events [3]. Our Visual Analytics method covers single-player, multi-player and event-based analytical views. Depending on the task the most promising features are semi-automatically selected, processed, and visualized. Their aim is to help soccer analysts in finding the most important and interesting events in a match. We present a flexible, modular, and expandable layer-based system allowing in-depth analysis. The integration of Visual Analytics techniques into the analysis process enables the analyst to find interesting events based on classification and allows, by a set of custom views, to communicate the found results. The feedback loop in the Visual Analytics pipeline helps to further improve the classification results. They evaluate their approach by investigating real-world soccer matches and collecting additional expert feedback. Several use cases and findings illustrate the capabilities of our approach [4] [5].

Figure 1 illustrate some visualizations for soccer sport. For instance, Fig. 1a shows the graphic to show the goals location in each match. White means the home team made, red means the away team made. Fig. 1b shows the performance score in which a whole match performance is evaluated by the data point about the attacking, possession, defensive will be added separately. By putting the link of the whole matches performance score, we can trend of teams status. Fig. 1c is the match heat map. Action Heat map displays the areas of the pitch where on-ball events most frequently happen. Action Areas Highlights where on-ball events happen on the pitch according to Squawkas 13 action areas. After match, Fig. 1d will show some important data statistical comparison between two teams. Possessions, passing success percentage, shoot numbers are always the data which coach will be interested in.

III. Data Collections

To analysis the shooting habit in football field, the most important data we need to get are the shot location and the

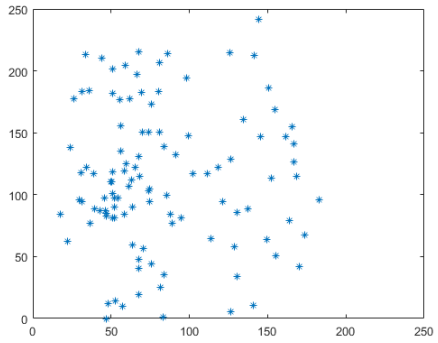


Fig. 2: The illustration of shots made by a playe from top.

target location. However, unfortunately, most of the soccer data are not publicly available. In this work, we gather the original data from Opta Sports [1]. We identified match information to define the corresponding between shot location data and target location data by Squawka and then extract the raw data from Opta sport. The original data format is look like as follow (written in the html file): $<$ $circlecx$ $=$ "186.24"cy $=$ "86.02059999999994"r $=$ "6.5"$fill$ $=$ "$\#ffffff$"$stroke$ $=$ "$\#ffffff$"$style$ $=$ "$stroke-width$: 2.4; $stroke-opacity$: 0.5;" $><$ $/circle>$ In this this paper, we only collected Cristiano Ronaldo shoot history in 2016/2017 season in Spanish La Liga league [6] [7] [**?**].

A. Data Standard

Unfortunately, the data extracted by the Opt sport website is not easy to visualize. The reasons are: 1) the scale of coordinates of the scooer field space is different than the coordinates of the gate space. There after, we call the space of inside and outside of the gate, gate space. 2) the soccer fields coordinates are also not meaningful. To solve the first issue, we tried to identify the exact scaling for each space. To address the second issue, we used the following formula to expand the points in the field space to match the real ball locations in the soccer field with shots in the field space. The formula is given by:

$$x' = x + \lambda * (x - m) \tag{1}$$

where x' is the new coordinate, x is the shot location in field space, λ is the scaling factor which controls the expansion and m is the middle point in the field space.

IV. Model Shot Location Statistic

In this section, we present our contributions in soccer visualization in three different parts. First, we show we can visualize the shot locations. Second, we show the visualization of target locations and finally we show our 3D model.

[1]Opta Sports is a sports data company with headquarters in London and other offices in Leeds, Munich, Bassano del Grappa, Milan, Paris, Madrid, Montevideo and Amsterdam.

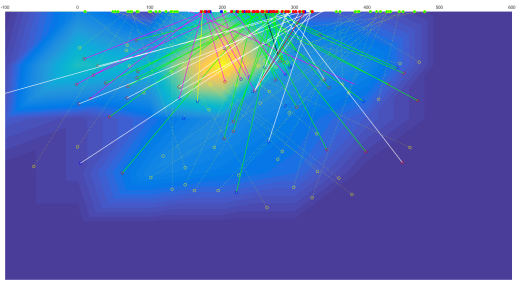


Fig. 3: Illustration of lilelihood of shots made ba player from different areas of the field.

A. Shot location visualization

First off, we drive the statistics for a player shooting habit. In order to that, we assume a field as a grid of 5×7. First we count how many shots are made from each square in the grid. The numbers are as follows:

$$M = \begin{bmatrix} 0 & 5 & 1 & 13 & 2 & 2 & 0 & 0 \\ 1 & 12 & 10 & 27 & 7 & 5 & 0 & 0 \\ 2 & 1 & 3 & 6 & 4 & 2 & 0 & 0 \\ 0 & 1 & 5 & 4 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Then we formed the colormap of this data gathered in Fig. 2 in shown in Fig. 3 to show the likelihood of a shot from each point in the field. The darker blue areas indicate less and brighter yellow areas indicate more likelihood of a shot by a player.

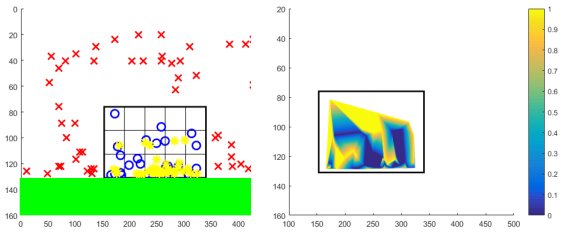
B. Target location analysis

The second analysis is simply illustrating the end position of the ball. Based on our data, we could categorize end position of the ball as: 1) *Off target*, 2) $Saved$ and 3) $Scored$. Here in Fig. 4a, different types of shot result is showing by different colors. Respectively, red $*$ is used for *off target*, blue o presents saved balls and yellow $\star$ shows the scored balls. Fig. 4b, shows the corresponding heatmap of in target balls. The blue areas represent the likelihood of score and the yellow areas represent the likelihood of ball being saved by a goal keeper.

C. 3D Model

In this section, we tried to put everything together using Matlab. The goal is to visualize field shots, target shots, connections of the corresponding start and end location of the shots and also distinguish different regions of the target. In the 3D Model, the lines link the shot start point and its target point. This model is shown in Fig. 5 filled red, blue and yellow circles represent scored, saved and off target shots. Similarly, in the field, the corresponding start location of the shots are show by the same colors. To avoid confusion, these balls are not filled.

In this model, the field is coloured based on the likelihood of shot of a particular player, from dark blue to bright yellow.



(a) (b)

Fig. 4: a) Shows the target balls (end location of shots). Yellow represents off target, blue saved and red shows scored balls. b) shows the heatmap of the target. The bluish areas indicates the more scored balls and the yellowish areas represent the more saved balls.

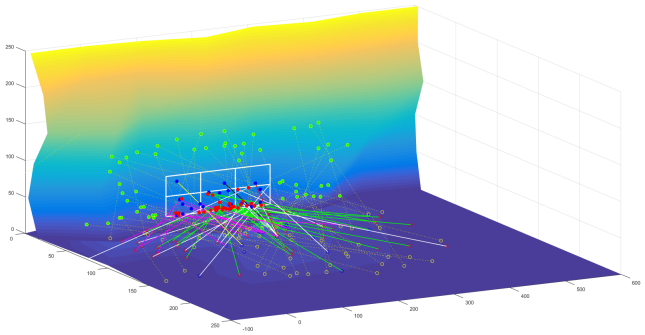


Fig. 5: A 3D model of a player habit. The X surface is shown by a colormap, indicating the lilkihood of shots. The Z surface represent the likelihood of end location of the balls.

The former states the less likelihood and latter the more. For each goalkeeper, it is very important to know what is the shot habit of a player from each area of the field. For example, if a player is in the right corner side of the box, what is his habit to shoot? More precisely, which area of the target is more likely he shoot? To address this question, we made, connect balls from the field to the target area. We assume a target is a grid (3×2) and assign a unique color to each connection line for each square in the target grid. For instance, in Fig. 5, all the the shots went to the top right corner of the target are shown by black connection line.

In this model, with a single glance, some statistics of a player can be grasped instantly. For example, likelihood of shoot and target of the shoot location can be seen. Further, it can be seen likely it is that the ball is in target or off target and if it is in target, it can be saved or scored. In additions, one can easily track each balls and see, for example, if the ball is scored, where the player shoot that ball. Besides, a goal keep can have a better knowledge of one player habit by analysing the in target statistics of the ball. In other words, how many of the balls are shoot to the upper right cornet of the target and how many of them are scored or saved.

V. Results and Discussion

In this paper, we present a visualization model to present and analyse a player shooting habit. We tried to use different shapes, colors and surfaces to visualize different factors of a player shooting habit.

In this project, we analyse the Cristiano Ronaldos one season (2016-2017) shot data in La Liga (Spanish league). The outcome of the model for this particular player can be summarized in the following.

- Ronaldo is used to shot from left. According to the color map in Fig. 3, through 2016/2017 season Ronaldos role is turning from a winger to the center forward, as his shot locations show he has shot most often in the penalty area. However, he is still used to shot and get score from a little bit right of the field.
- From Fig. 5, it can be concluded that we he shot from the left side of the target, most of his shots ended in right side of the target. On the other hand, interestingly, most of the shots made from the right side of the target ended in the middle or right side of the target. One can conclude, his habit is to shot to the right side of the target. However, some other factors like his running directions, or whether he has shot with his left or right foot should be taken into account to make a better conclusion.
- Another result we found is that Ronaldo s mostly scored when he shot new the gate. As he is getting old, he is becoming a center forward and therefore he scores from near the gate more than before.

VI. Conclusion and Future Work

In this mode, we tried to visualize a shooting statistics of a player to a single image. This image is able to capture different statistics and habit of a player such as likelihood of his shots from the field and also the end location of the balls. Besides, different regions of the target can be tracked and analysed.

This model can still be improved in many areas. Some extension of the model could be as follows: a) Larger dataset. Due to the time limitation, the dataset we made is only included one player for one season shot data. This can only represent his habit in this specific year. If the model can include more players and more years data, it would be more meaningful for all football club to analysis some teams or players shooting habit and training preference. b) UI for the model. In the future, this model can be written in an expert football data analysis program, with effective user interface. The user can select the specific time-line of data to show the specific part of visualization result and do analysis. c) More factors. end location of the balls is not the only important metric. Other metrics can be consider such as how many passes before a shot happened and the speed of shot. Many other reasons can also affect the percentage of goals/shots. In the future, this model can be extended with incorporating more data and more factors in better visualize a player shooting habit.

ACKNOWLEDGMENT

The authors would like to thank Dr. Zhang for his ideas and comments to make this project better.

REFERENCES

[1] C. Perin, R. Vuillemot, and J.-D. Fekete, "Soccerstories: A kick-off for visual soccer analysis," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 12, pp. 2506–2515, 2013.

[2] A. Rusu, D. Stoica, and E. Burns, "Analyzing soccer goalkeeper performance using a metaphor-based visualization," in *Information Visualisation (IV), 2011 15th International Conference on*. IEEE, 2011, pp. 194–199.

[3] D. Sacha, M. Stein, T. Schreck, D. A. Keim, O. Deussen *et al.*, "Feature-driven visual analytics of soccer data," in *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*. IEEE, 2014, pp. 13–22.

[4] B. Marr, "How big data and analytics are changing soccer," in *Linkedin*, 2015.

[5] "Squawka.com," in *Real-madrid — Football — Statistics — Form Guide — Squawka.com. [online] Available at: http://www.squawka.com/teams/real-madrid/stats*, 2017.

[6] "In a globalised world, the football world cup is a force for good," in *The Conversation*, 10 July 2014.

[7] "Association football," in *En.wikipedia.org - [Online]. Available: https://en.wikipedia.org/wiki/Association*$_f$*ootball*, 2017.