



**HAL**  
open science

## Visualizing and Analyzing Disputed Areas in Soccer

Jules Allegre, Romain Vuillemot

► **To cite this version:**

Jules Allegre, Romain Vuillemot. Visualizing and Analyzing Disputed Areas in Soccer. Visualization in Data Science, Oct 2020, Salt Lake City (Virtual Conference), United States. 10.1109/VDS51726.2020.00010 . hal-02951454

**HAL Id: hal-02951454**

**<https://hal.science/hal-02951454>**

Submitted on 28 Sep 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Visualizing and Analyzing Disputed Areas in Soccer

Jules Allegré\*

LIRIS, École Centrale de Lyon

Romain Vuillemot†

LIRIS, École Centrale de Lyon

## ABSTRACT

Space ownership models assign 2D areas to individuals, based on their ability to reach locations according to their direction and speed. In this paper, we investigate the case where two or more individuals can reach a given location *simultaneously*. We refer to those locations as *disputed areas*, as there are tension and uncertainty on ownership, which is an important spatial analysis tool, e. g., in sports where players share a space with adversaries. We present the process to calculate those disputed areas from existing space ownership models and introduce several visualizations and analysis of those areas using sport tracking data from Liverpool 2019’s goals. Those areas have been particularly insightful to understand assists, the ultimate pass that is critical for a team to score. We also report on feedback from experts both on the relevance of those areas as well as their visual design.

**Index Terms:** Human-centered computing—Visualization—Visualization techniques—Treemaps;

## 1 INTRODUCTION

In many domains, space is not shared: countries have borders; animals have hunting areas; soccer players control certain areas of the pitch. Thus, it is important to characterize those *disputed areas*, as they are the focus of attention for economic, political, or survival reasons. In this paper, we focus on soccer, where space ownership is crucial to receive, control, and shoot the ball into the opponent team’s goal. To understand disputed areas, one first needs to understand non-disputed or *controlled* areas. Such areas result from space-time models, which key concept is a *first arrived, first served* approach, but can greatly vary according to the complexity of the model. The most simple models do not account for players’ physical characteristics (as illustrated Figure 1 (a)): players move towards all directions, at an equal and constant speed. More advanced models (e. g., [2, 6]) account players’ initial orientation and running speed: as a result, some areas around them—even the close ones—are not accessible immediately (e. g., the ones in their back).

The contribution of this work is a formal definition and representation of disputed areas, which are overlaps of controlled areas of two or more players. Such overlaps indicate there is uncertainty on who controls it (Figure 1 (b)). While such areas appear for every pair of relations between individuals, they do not hold the same importance: some areas have been voluntarily created to upset the stability of a region, e. g., the ones controlled by defenders so that the opposing team can pass the ball where a teammate can receive it.

In this paper, we first present the step-by-step process to calculate such *disputed areas* and its adaptation to the specificity of soccer. Section 2 presents the occupation models we rely upon to calculate such areas. Section 3 details the features of our calculated areas and how to efficiently estimate its parameters to optimize the relevance of the areas (e. g., not too large, not too close to the players). In Section 4 we apply our model to a case study: Liverpool 2019’s goals, to

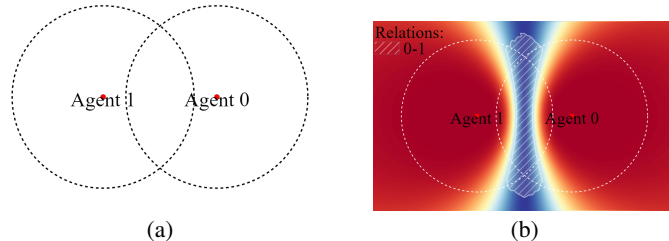


Figure 1: (a) two agents with their controlled area (dotted circles). (b) Example of disputed area (diagonal hatching) where ownership is disputed by two or more people. We also plotted a red blue color map of agents simultaneity (blue for synchronisation) which characterize the overlapping of controlled areas.

understand the role of disputed areas in the players’ decisions. We present in Section 5 several visual designs to better emphasize the disputed areas and discuss models optimization and improvements. Finally in Section 6 we report on feedback from experts to which we presented our examples and for which they suggested both models and visual design improvements.

## 2 DEFINITION AND CALCULATION

We define *disputed areas* as follows:

*“An overlap of controlled areas by two or more agents.”*

This definition aims at covering a broad scope of applications, not just for soccer. However, in this paper, we will narrow it down to soccer, and show how the model can be customized to take into account this sport’s singularities. The first one being that areas should be *relevant* according to a goal: for instance areas close to the goalkeeper is more relevant than the ones close to the corners. Also, areas that are too far from players will have a lower level of relevance as occupation models usually are relevant close to players. The non-relevance level will be filtered out according to a threshold we estimate. We also expect areas to have *stability*, i. e. they cannot suddenly appear/disappear as human motions usually are continuous and predictable. Finally, those areas should be *understandable* by analysis so their construction should remain simple with a visual design that conveys its properties, e. g., spatial, and relational properties.

Our approach to calculating disputed areas relies on a layered approach, where each location is assigned a quantity that represents the degree of occupation which we use to define closed areas. We will then refine those areas to comply with the singularities we introduced above (relevance, stability, and understandability).

### 2.1 Step 1: Picking an Occupation Model

There exists multiple occupation models [2, 6, 8, 10], and our work relies on two of them. The first one is the *temporal distance model* from Rolland et al. [6]. It takes into account the speed, orientation, and force of each agent to compute time between agents and

\*e-mail: jules.allegre@ec119.ec-lyon.fr

†e-mail: romain.vuillemot@ec-lyon.fr

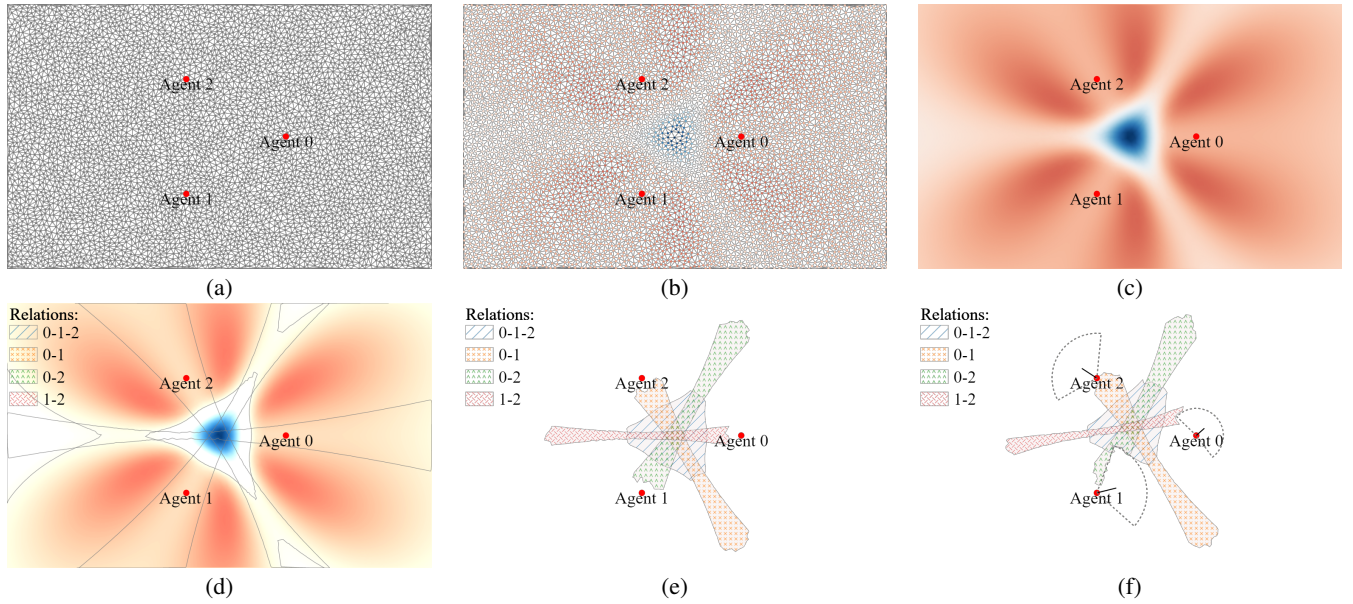


Figure 2: Our process to calculate the *disputed areas* is organized as follows: (a) a continuous 2D space is discretized into nodes (which location is random but dense enough to uniformly cover the whole space); (b) each node is assigned a weight according to the occupation model (blue representing a strong synchronization (weight close to 1) and red a weak one (weight close to 0) as represented in a continuous way on (c)); (d) overlaps from the occupation model are drawn using contours according to a threshold. (e) The *disputed areas* are filtered based on their relevance, and based on the speed of each agent (f).

points of the space. We based the *disputed area* on it to generate a space partition. The second occupation model is the *influenced zones model* developed by Stein [8] and we will discuss it later in Section 3. This second model will be used to remove some parts of the occupation model, mostly areas very close to the agents.

We picked the *temporal distance model* [6] as it provides a sufficiently good estimation of human motions while remaining sufficiently easy to implement and quick to calculate. The model output is quantitative values for each location, referring to the time an agent needs to reach it from its starting point.

## 2.2 Step 2: Discretizing Regions Based on the Model

The occupation model now needs to be discretized otherwise too many locations will be calculated. We proceeded by placing nodes to partially cover the 2D space using a random (Poisson) distribution. This choice is motivated by the fact that we want random points distributed heterogeneously to get zones with smoother edges and avoid a pixelization of our zones. The distributed points are illustrated Figure 2 (a). Now, to determine if two agents reach a point of the space at the same time we need to look at the difference between their temporal distance (e. g., time to reach the point). If more than two agents are implied, we just average each difference. Then, we use this value to compute a weight which determines if the weighted point is part of the *disputed area*. This weight goes from 0 to 1 where 1 means a perfect synchronization of all agents. Figure 2 (b) shows the weighted dots for the 3 agents interaction and Section 3 details the calculations. Then, Figure 2 (c) shows the corresponding continuous color map by interpolating the values between nodes.

The color map provides us with the discretized space points but does not define clear groups of similar points separated by *boundaries*. Moreover, other information is provided by this continuous scale which makes it difficult to identify patterns. Therefore, we chose to select nodes with a certain weight to draw the boundaries contours, as illustrated Figure 2 (d) for all interaction available with 3 agents. Then, we only keep the bounded areas which are our *disputed area*, with a threshold value given in the next Section 3.

We will also apply further filters, e. g., agents properties such as the speed (Figure 2 (f)).

## 2.3 Step 3: Visualizing Disputed Areas

Table 1 is a summary of our visual mapping to convey calculated disputed areas. Those decisions aimed at making those areas easy to identify based on the relations they belong to. We thus filled the area with different textures and colors, along with a legend (top left part of the visualization). Textures enable to combine multiple overlapping categories and are simple to draw using SVG. Each area is transparent so that the pitch below remains visible. Gray contours have been added to make those areas perimeter easy to identify.

## 3 PARAMETERS AND FILTERS OPTIMISATION

Figure 3 (a) shows four *disputed areas* resulting from three agents. As we can see space is rapidly fully covered. To prevent visual overload as well as to focus on the relevant areas, we added extra calculation steps of threshold parameters to filter the areas based on the properties from the previous section (weight and temporal distance), as well as a *vicinity* criteria to remove areas too close to the agents, based on the *influenced zones model* developed by Stein et al. [8].

### 3.1 Assigning Reachability Weight to Grid Nodes

As seen in Section 2, the first step of the boundaries contour drawing is to assign a weight to each node of the grid (Figure 2 (b)). This weight will be modified according to the other parameters we introduce below. Then, we define a threshold to build the contours connecting the nodes with similar weight than the threshold value.

The weight of a node depends on the time  $t_i$  needed by an agent  $i$  to reach this node. More specifically it is the difference of the two agent times to reach the point,  $\Delta t_{ij} = t_i - t_j$ , taken into account. This time is calculated with the *temporal distances model* and assume that an agent moves towards a node with a constant force,  $F = 10 \text{ m s}^{-1}$ , developed by Rolland et al. [6]. Then to obtain the weight,  $w$ , we use the following equation:

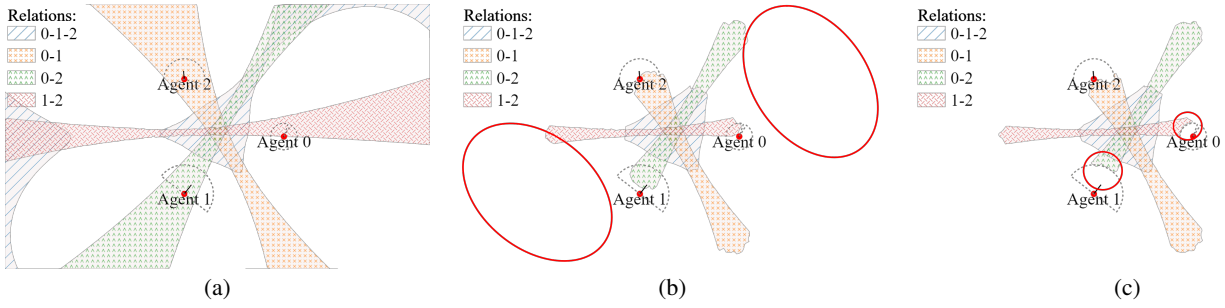


Figure 3: The *disputed areas* filtered step by step for each combinations available with three agents. The dotted areas around the agents are the *influenced zones* developed by Stein et al. [8]. The black lines represent the speed vector of each agent and the red circles indicate where is the relevant information. We add a filter to each sub-figure from left to right and we begin with the weight filter,  $w_{thresh}$ , (a), then the temporal distance filter,  $t_{thresh}$ , (b) and finally the agent proximity thresholds,  $r_{infl}$ , (c).

Symbols	Signification
	Agent 0
	Speed
	Influenced control area, [8]
	Disputed area with the relation between the agent numbers who build the area. If the text is red, the area does not exist.
	Barycenter fills with the color of its associated area.
	Indicator to highlight the relevant information of the figure.
Soccer Symbols	
	Player with its index number and its team color in background. If the contour is highlighted in yellow, the player has the ball.
	Ball
	Pitch lines
	Look of a player to another location.
	Ball trajectory.

Table 1: Symbols used to design the *disputed areas*.

$$w = \frac{\sum_{i=0}^n \sum_{j=i+1}^n e^{-\frac{\Delta t_{ij}^2}{2}}}{\frac{n(n-1)}{2}} \quad (1)$$

with  $n$  the number of interacting agents. This equation represents a mean of normal distribution taking into account each time difference between each pair combination of agents. Thus, the weight is within  $[0, 1]$ . If we plot the weight for each node and interpolate it with a color from red to blue where blue represents the synchronization of all agents we obtain a result similar to Figure 3.

Henceforth, to draw the contour of an area we just need to add a threshold based on the weight. The choice of this threshold is not obvious because for a large number of agents interacting we do not have the same maximum weight value than for a small number of agents interacting. Thus, we rely upon the percentage of nodes in the area within the threshold value. After several visualizations we set the ratio  $R = 0.1$  which is consistent with our vision of those areas. Then, we select the  $R \times \#nodes$  nodes with the highest weight value. The selected node with the lowest value will determine the weight threshold value,  $w_{thresh}$ , red line on Figure 4. We can also

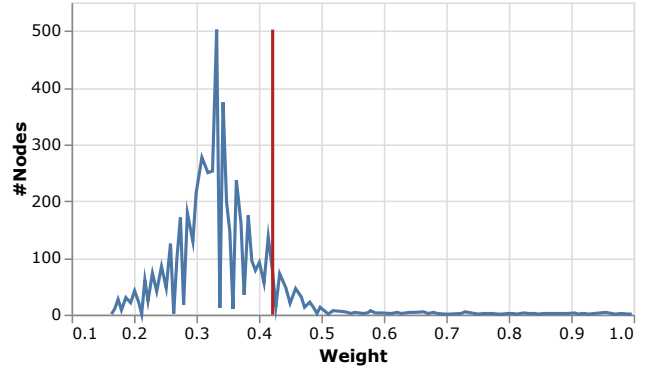


Figure 4: Number of nodes in function of weight value. The red bar splits 90% of the nodes on the left side and the remaining 10% on the right side, it is adjusted with  $R = 0.1$  in this case. We can read  $w_{thresh}$  on the x axis thanks to the red bar.

visualize the distribution of nodes in function of the weight value on this figure.

Once we have  $w_{thresh}$  we can draw our primary area as illustrated on Figure 3 (a).

### 3.2 Reachability Reduction by Time

Figure 3 (a) shows the expanded areas from agents. Therefore we can question the relevance of such an area because the model is not taking into account the time needed to reach a point of the area but times differences. Currently, the model could also draw a *disputed area* far away from the agents and this area will consequently be not interesting because we assume that the uncertainty increases in function of the temporal distances. Thus, we need to define a maximum temporal distance to discard such areas.

To achieve this, a threshold needs to be set relative to the position of the agents because a fixed small threshold value will cut the *disputed area* for distant agents, and inversely a large value will not filter the *disputed area* for close agents. So, we simply take the maximum temporal distance among all possible temporal distances between two agents and we set it as the threshold value,  $t_{thresh}$ , Equation 2. We can see Figure 3 (b) the areas filtered by the weight and the time threshold.

$$t_{thresh} = \max(\{t_{ij} \mid i \neq j, 0 < i < N \text{ and } 0 < j < N\}) \quad (2)$$

with  $t_{ij}$  the time for agent  $i$  to reach agent  $j$ . If the time needed by

an agent to reach a node is less than  $t_{thresh}$ , then we assign a weight of 0 to this node to discard the node.

### 3.3 Reachability Reduction by Agent Proximity

Our model draws areas with only a subset of all relations with other agents. For instance, on Figure 2 (c) we drew the area for the relation  $0-1$  though there are three agents in the plane. This feature allows seeing where we will have the interactions between the selected agents but it is not taking into account the other agents at all, even though they are here and might act as obstacles. Thus, we use a model of space occupation called *influenced zones model* which determines a region near an agent and considers this region as inaccessible for the other agents because it clearly owns it (Figure 3 (a)). This model is taken from Stein et al. [8]. It computes an angle,  $\alpha$  as a function of the velocity  $v$ , assuming that the maximum speed is  $10\text{ms}^{-1}$  and the angle will decrease when the speed increase:

$$\alpha = -0.0038\pi \cdot v^3 + 0.0793\pi \cdot v^2 - 0.61081\pi \cdot v + 2\pi \quad (3)$$

Along with a radius of influence  $r_{influ}$  (in meter) in function of a parameter  $T$  (in second) set to a fix value and the agent's velocity  $v$ :

$$r_{influ} = \begin{cases} T \cdot \langle v \rangle & \text{if } v < 1\text{ms}^{-1} \\ T \cdot v & \text{else if } 1\text{ms}^{-1} < v < 10\text{ms}^{-1} \\ T \cdot 10 & \text{else.} \end{cases} \quad (4)$$

Then, the threshold is done by affecting a weight of 0 to all nodes in the area of influence of each player as if no dispute can occur, Figure 3(b) to (c).

### 3.4 Agents Relations

We draw multiple *disputed areas* but we did not detail the concept of relations. The simplest form of relations are between two agents as seen on Figure 1, we name it a *pair* relation. But we could add more agents such as illustrated in Figure 3 where four areas are drawn: three pairs relation and one triple relation (e. g., triple agents interaction). Therefore, for a three-agent model, we can build four *disputed areas*. So, if we have  $N$  agents, the equation 5 tell us the maximum number of areas we can build,  $P_{relation}$  as follows:

$$P_{relation} = \sum_{k=2}^N \binom{N}{k} \quad (5)$$

Each time we add an agent to interact for an area, it modifies the weight of the nodes based on the equation 1.

## 4 CASE STUDY: LIVERPOOL 2019 GOALS

We applied our model on a publicly available soccer dataset, Liverpool 2019's goals provided by Friends of Tracking<sup>1</sup>. This dataset is composed of players' names and positions on the pitch, as well as the position of the ball before each goal, which are the most important aspects of soccer games analysis [4,5]. We provide the context of the dataset as well as the detailed analysis of two goals using our *disputed areas* calculation.

**Context.** Liverpool—alias *the Reds*—are an English soccer team led by Jürgen Klopp since 2015. In 2019, the team won the first places of the Champions League, the UEFA Supercup, and the Club World Cup. They are known for their offensive style of play thanks to their three forwards players: Mohamed Salah, Roberto Firmino, and Sadio Mané. We picked two of their 2019 goals based on the diversity of the situation they offer.

**Liverpool vs Newcastle.** We started with the second goal of Liverpool against Newcastle United from the 14th of September

of 2019 (both the video footage<sup>2</sup> and animated visualization<sup>3</sup> are available). According to Harshal Patel analysis [3], knowing the offensive power of Liverpool, the Newcastle strategy is based on solid defense lines and counter-attacks. Thus, Liverpool needs to increase the speed of its passes and the movement of its players to drag opponent players out of their positions and to release free spaces. The beginning of the match was difficult for Liverpool and the strategy of Newcastle paid off rapidly by a goal from Jetro Willems at the 8th minute. It wakes up Liverpool which speeds up their play to create open spaces for shoots. Several attempts are done until the 28th minute when Sadio Mané scored thanks to Robertson's movement which drags Mané's defender away. The two teams are aware of the opposing strategy and the second goal from Liverpool results from an offensive pass behind the defensive line by Roberto Firmino to Sadio Mané who scores after his run and a missed catch of the ball by Martin Dúbravka, the opposing goalkeeper.

This second goal is our focus of analysis. Figure 5(a) shows the instant Firmino decides to make a (too deep) pass to Mané, who needs to run fast to catch it. Dúbravka (from the opposite team) runs as well very fast to try to catch the ball. The confrontation between the players is highlighted in Figure 5 (b). Figure 6 confirms this area is of interest as using an occupation model, it results in a white zone meaning that there is not an obvious dominance by a team to control this part of the pitch. Therefore, it is where we expect our *disputed areas*.

As our *disputed area* calculations can be generated for any pair of players, we first need to define which relation is of interest from this situation. First, we know there will be a confrontation between Mané and Dúbravka so we add the relation  $1-21$ , then Mané is surrounded by Schär and Lascelles and all are running towards the same location which might be a *disputed area*, we add this relation too,  $1-4-5$ . On the right wing we also have Salah and Dummett who are running and seem to dispute the same area so we add the  $3-6$  relation. Eventually, Firmino has a little pressure on his left side by Hayden and we want to plot their *disputed area*, relation  $2-9$ .

Resulting *disputed areas* are illustrated on Figure 7. As expected, the areas  $1-21$  and  $1-4-5$  are visually important as they capture where the ball will arrive.  $2-9$  has little interest but shows a *disputed area* close to Lascelles predicting where it would have a dispute if Firmino does not pass. Eventually,  $3-6$  shows us that Salah needs to move aside if he wants to have a chance to get the ball. Roberto chose to pass forward in the *disputed areas* to let a chance to Mané to have the ball. The intersection of the gray dotted lines in Figure 7 indicates where Mané will cross the ball trajectory. We see this intersection is behind the *disputed area*  $1-21$  and so we expect that the goalkeeper will catch it first. It was the case but he missed catching the ball and Mané took advantage of the situation and scored.

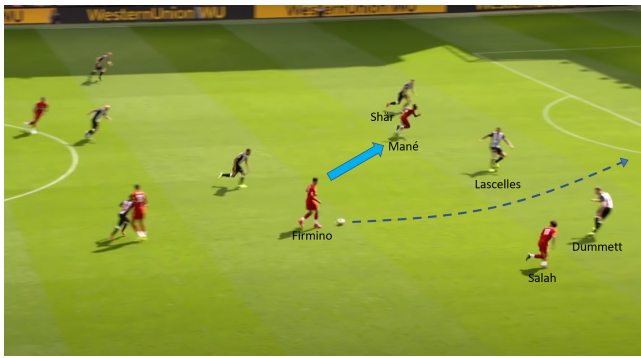
Figure 8 provides a temporal visualization of the evolution of the area of each *disputed area* to assess the stability of our model (Section 2). A visual analysis of this chart shows the model is quite stable as the variations are not too high even though there are some locals peaks.

**Liverpool vs Porto.** We then picked the second goal of Liverpool against FC Porto on the 17th of April of 2019. The match was the quarter-final of the champions league. During the first leg, The Reds won 2-0, thus for this second leg the tension for Porto is high, also, the winner will go to the semi-final. Despite that, Liverpool was under pressure during the beginning of the match and Porto missed a lot of occasions to score. After the first controversial goal from Mané, the second goal started with a counter-attacked. Trent Alexander-Arnold did a long passed forward to Mohamed Salah. Felipe Monteiro did not succeed to intercept the pass and Salah scored using his left foot.

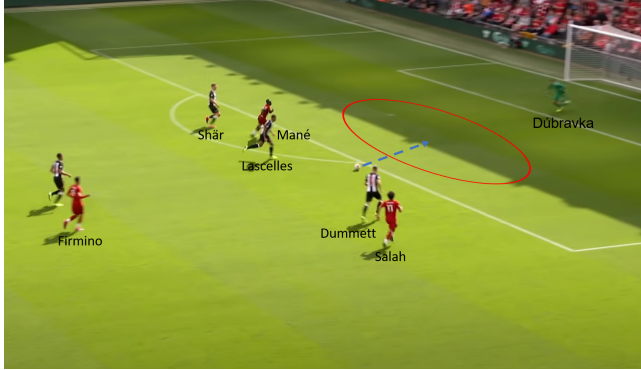
<sup>2</sup><https://www.youtube.com/watch?v=RGFaJDAselI&t=254s>

<sup>3</sup><https://observablehq.com/@julesallegre/zones-data-from-conflict-zone-model-soccer-application>

<sup>1</sup><https://github.com/Friends-of-Tracking-Data-FoTD/>



(a)



(b)

Figure 5: TV screen captures of the match between Liverpool and Newcastle at the 40th minute of the game. **(a)** We see Roberto Firmino watching Sadio Mané (blue filled arrow) and the situation before he passes the ball. The dashed arrows indicate the ball trajectory. **(b)** Sadio Mané, Martin Dúbravka, the other defenders (Paul Dummett, Jamaal Lascelles, and Fabian Schär) followed by Mohamed Salah are running toward the next ball position where we expect a *disputed area* (red circle).

We picked this example as it provides a different situation than the previous one, but also to illustrate our model because it shows an abnormal feature. We can see in Figure 9 that the area for player 1-7 (Salah and Monteiro) is very thin even though the space in front of them is free and the speed of the two players nearly have the same orientation and value. We expected to have a very large *disputed area* due to the uncertainty of where the ball will land and the proximity of the two players. It demonstrates that our intuition is not a good tool for this situation. But the more trouble aspect is seen when we look at the sequence of images<sup>5</sup> (frame 116 to frame 152 for the match *Porto 0 - [2] Liverpool* in the notebook). During the run of Salah and Monteiro we observe a size oscillation of the area in Figure 10 though the players' speed is not oscillating, Figure 11. Therefore, we assume that our model is very sensitive to small orientation and position variations which might bring problems for stability requirements. We also plotted the 5-6-10 relation to show that Alexander-Arnold did not have the option to pass the ball to Sadio Mané (5) because the *disputed area* is too far away from the goal.

## 5 DISCUSSION

In this section we discuss the current features of our model and how to optimize both its calculation and visual representation

<sup>5</sup><https://observablehq.com/@julesallegre/zones-data-from-conflict-zone-model-soccer-application>

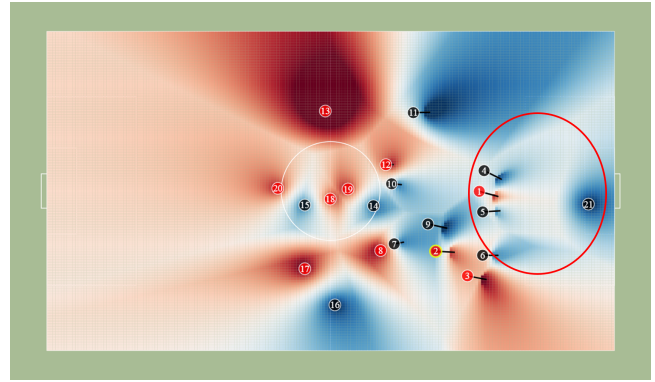


Figure 6: Color map of the pitch at the frame 48 based on the temporal distances model from Rolland et al. [6], we re-implemented as a notebook<sup>4</sup>. The blue represent the Newcastle control and the red the Liverpool one. We observe blue zones and white zones in the penalty area which means a dominance for defenders (e.g. Newcastle) represented by the red circle. Note that numbers are not related to the player number of the season 2019 but are indexes. A part of the following list (red: **Liverpool**, black: **Newcastle**): 1: **Sadio Mané**, 2: **Roberto Firmino**, 3: **Mohamed Salah**, 4: **Fabian Schär**, 5: **Jamaal Lascelles**, 6: **Paul Dummett**, 9: **Isaac Hayden**, 21: **Martin Dúbravka**.

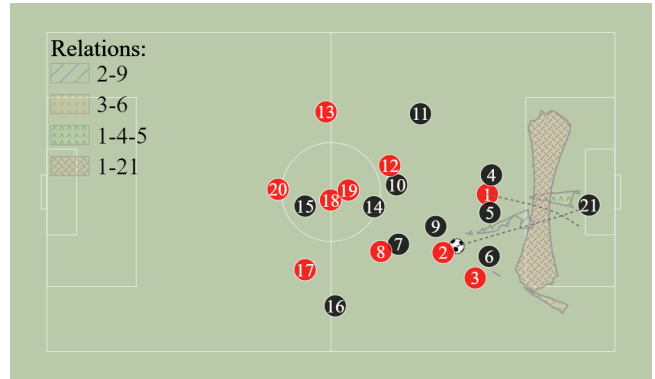


Figure 7: Plot of the *disputed area model* on the 52<sup>nd</sup> frame, for the four relations 1-21, 1-4-5, 3-6 and 2-9. We see a huge area between Mané (1) and Dúbravka (21). The two dotted lines indicate Mané's trajectory and ball's trajectory.

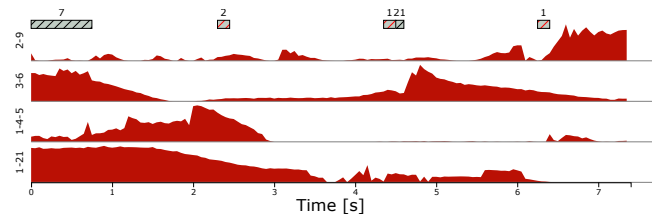


Figure 8: Plot of the areas evolution of each *disputed areas*, 1-21, 1-4-5, 3-6 and 2-9. The Y-scale of each subplot is independent from each other and thus value of the area is not comparable between two *disputed areas*. The hatched rectangles on the top of the figure show which player possesses the ball, the stroke colors match with the player background colors.

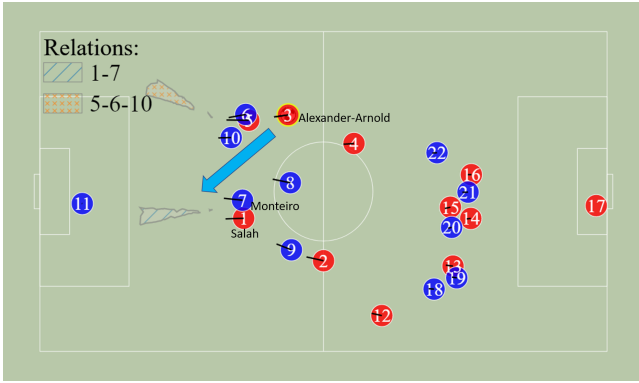


Figure 9: Frame of the match Porto (in blue) against Liverpool (in red) around the 65th, frame 172 of the Friends of Tracking dataset. We can see Alexander-Arnold (3) ready to pass the ball to Salah (1) indicated by the blue filled arrow. Monteiro (7) is between them.

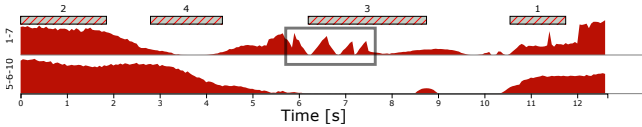


Figure 10: Rapid size variation for *disputed areas* 1-7 and 5-6-10 during the second goal of Liverpool against FC Porto. In the grey square we can see an oscillation of the area 1-7 during the run of Salah and Monteiro, the limit of this square correspond approximately to 6s to 7.5s after the beginning of the action.

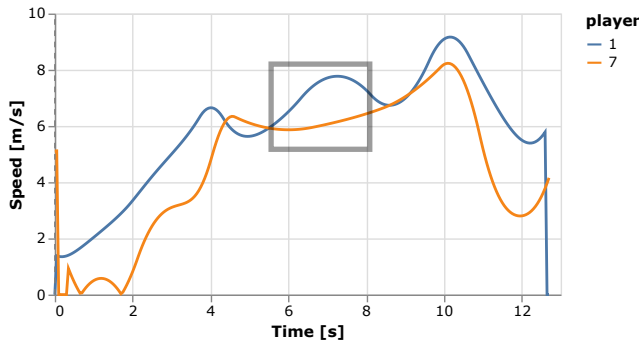


Figure 11: Speed variation of the player 1 Salah and the player 7 Monteiro during the second goal of Liverpool against FC Porto. The grey square covers the time laps between 6s to 7.5s where the size area 1-7 is oscillating.

### 5.1 Minimum Time Threshold

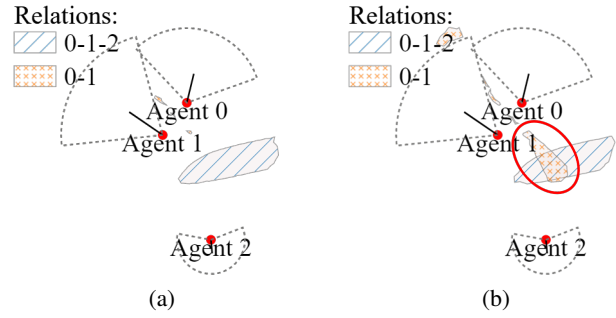


Figure 12: Plot of two *disputed areas* to reveal the time threshold issue for close agents. The red circle highlights the apparition of the *disputed areas*.

The time threshold is based on a temporal distance calculated by the position and the speed of each agent in interaction. It means that a single value is calculated for all agents and the threshold is calculated with the same value for each of them. It has the advantage to be simple and coherent with the intuition to have *disputed areas* not too far away from the agent which works well for the medium and large temporal distance between agents, but when we are looking to two very close agents (relative to the global scale) their *disputed area* will be too thin or not defined such as on Figure 12 (a). This effect is amplified if we add the proximity threshold because we filter the areas close to the agent. Thus, we chose to set a minimum threshold value chosen by the user. In our case we use  $t_{thresh}^{min} = 5$  s. We see the result in the Figure 12b.

### 5.2 Simpler and Smoother Area Shapes

We draw our contour by selecting nodes on a random grid, thus some parts of the area shape may be sharp and thus may result in complex polygons difficult to track. There are multiple solutions to address this issue. We could refine the nodes grid to make transitions between two points more continuous but it will increase the calculation cost, we could adapt our threshold parameters to select points in larger bands and let the contour function (we used the *d3.tricontour()* from the D3 [1] library) has more options to draw the contours but we will loose boundary accuracy, or we could include the threshold parameters in the weight function (Equation 1) to obtain smoother variations of the weight and so smoother contours but it will delete the option to add layers of filters.

Also, we could smooth the edges by post-processing of the contour, e. g., by approximating the shape of our zones using ellipses (Figure 13). It greatly simplifies the contours but we loose accuracy on boundaries. Moreover, some areas (e. g., concave ones) could not be approximate by an ellipse and we lose the information. This problem goes deeper than we thought and we need further study to obtain a correct simple form.

### 5.3 Zones of Influence Overlap

In our proximity threshold, we considered that a zone of influence is not uncertain because it is obviously possessed by the owner of this zone. But those regions overlaps mean that multiple agents possess this zone, thus it is by definition a *disputed area*. We can see on Figure 14 (a) and 14 (b) the effect of adding this filter. This is achieved with a little modification of our precedent threshold: instead of removing all points from influence zones we just remove the points which are in one influence zone only.

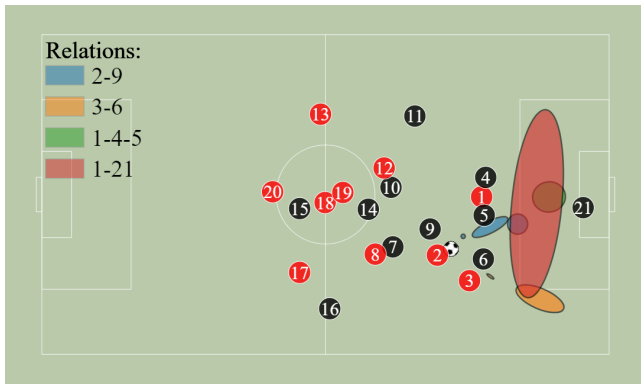


Figure 13: Simplification of the *disputed areas* for the 52<sup>nd</sup> frame of Liverpool-Newcastle. Those simplifications approximate the previous areas by ellipses.

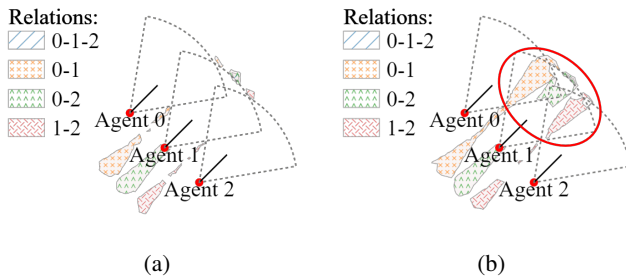


Figure 14: *Disputed areas* for 3 agents close to each other with overlap of their influence zone (dotted areas), [8]. The model filters all points in influence zones in (a) and filters only points in a single influence zone in (b). The red circle highlights the apparition of *disputed areas*.

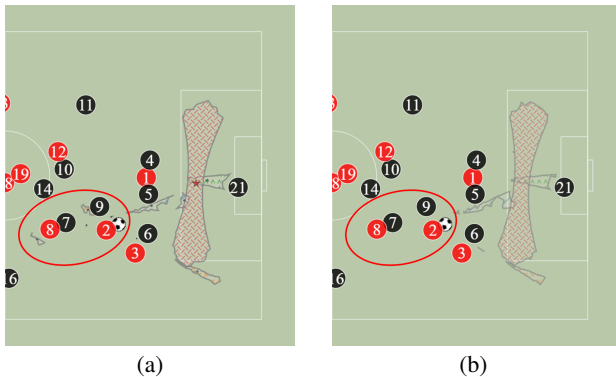


Figure 15: *Disputed area model* applies on a soccer pitch to highlight the filtering of the areas away from the objective: the goal. It is a frame from the match Liverpool-Newcastle played on September 14<sup>th</sup>, 2019. The red circle indicates where the areas disappear.

#### 5.4 Filtering by Specific Locations

In Section 4, we pointed out that some *disputed areas* are not interesting because they do not capture the purpose of the game: to put the ball into the goal. Thereby, if we add a filter that selects areas according to their position and relatively to an objective and the position of the agents, we would have a better approach for soccer. Figure 15(a) shows that players 2, 7, 8 and 9 are close to areas. Those areas are behind the agents who create them if we take the objective

as the front position. Therefore, we consider them irrelevant and we filter them out, and the result is illustrated in Figure 15 (b). The criteria to determine if an area is behind or in front of an agent is based on the barycenter and it is showed by a star.

#### 5.5 Limits

To initialize our model on soccer data, we first visually analyzed the action watching the video footage and animated visualizations, and then we selected the relations after multiple tests. Nevertheless, we could miss interactions between players because we only rest on our intuition, perception of the action, and understanding of the soccer strategies. Thereby, a good optimization would be to develop a function available to search and to return the relevant relations and validate it with soccer experts.

To configure and validate our calculations of disputed areas, we mostly relied upon visual analysis of the results. We plan to compare our calculations with other models to understand its relevance and its defects. Those comparison models could either be with different occupation models, but we also plan to conduct a more systematic parameter space exploration to identify significant variations that may result in more relevant models.

#### 6 FEEDBACK FROM EXPERTS

We presented the *disputed areas* visualizations to two experts in data visualization and soccer analysis. We sent a threefold questionnaire to the experts: we first introduce the Liverpool vs Newcastle second goal situation (Figure 5) with open questions on how they would analyze the action, then questions on our model in this situation, finally different variations of visualization. The questionnaire is available as supplemental material.

**Expert #1** is the co-author of an occupation model we used in this article. For him, Figure 5(a), it seems obvious that Firmino will pass forward and try this risky pass because he has enough space between the defenders, and Mané is already running. Salah is not a pass option but he drags Dummett and makes the pass easier for Firmino. Expert #1 co-developed the occupation model [6] of Figure 6 for another sport (basketball game), we used it with the little modification that we draw the zone of each player and not the zone per team. We could see little white areas between red zones which show us that the zones are individualized. He remarked that the model is too deterministic for a soccer application because it is not taking into account the multitude of uncertainty that can occur, whereas the pitch control model from Fernandez et al. [2] does. Eventually, we presented to him our model plotted in Figure 7. The area 1-21 is relevant and gives a clear area of dispute. He assumes that we might use it to optimize the goalkeeper placement according to the forward run. The area 1-4-5 is interesting because it shows where the dispute is, which is relevant, but we do not know where are the zones that the players occupy and it makes the area's understanding more difficult. Also, the relation averages the sub weight of each pair relation, thus the pair 4-5 too, even though they are two defenders! This remark reveals that our model needs to be deeply modified to be correctly adapt to soccer applications. The area 2-9 could be more relevant if we take the players 5 and 6 into account and the area 3-6 has little interest because Firmino (2) will not pass the ball to Salah (3) and his role is only to drag Dummett (6) out of the defensive line.

He also made several remarks about the application and improvements in the model. He imagined that our model could predict the free spaces if we plot all areas with interaction pairs between players of different teams and if we select only the spaces not drawn. We could also improve its usefulness by coupling our model with the *off ball scoring opportunity model* which informs us about the zones suitable to score [7]. In the same way, we could add the *pitch control model* to our visualization to observe the occupied areas and the disputed ones. The last remark was to integrate the weight of each



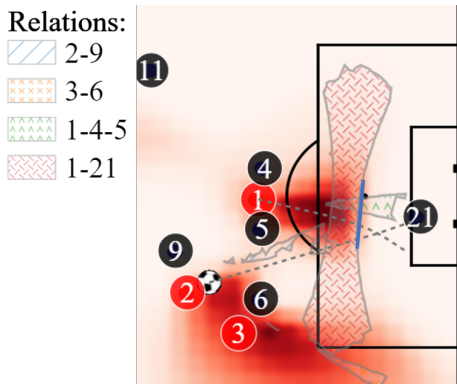


Figure 16: *Off Ball Scoring Opportunity model* plotted on the 52<sup>nd</sup> frame of Liverpool-Newcastle from the Friends of Tracking dataset (in red gradient) with the overlap of the *disputed area model* (textures). Grey dotted lines are Mané (1) trajectory and ball trajectory. The blue line is a part of the 1-21 area boundary and highlight the limit of the *OBSO area*.

node along the match time to see which areas of the pitch are the most disputed.

We took his remark into account and chose to compare our model with the *Off Ball Scoring Opportunity model* (e. g., *OBSO*). This model predicts the areas with a scoring potential for the attacking team. Expert #1 gave us the picture of the *OBSO* model plotted for the 52<sup>nd</sup> frame of Liverpool-Newcastle, the reddest a cell, the higher the probability to score is. We overlap the picture with our model to obtain the Figure 16. We observe that the highest probability to score is overlapping with the area 1-21 which confirms that this area is interpreted as a targeted area to score. We remark that the peak is not at the intersection between Mané trajectory (1) and the ball trajectory (dotted grey lines) which is another clue to say that the pass was too deep. Another feature shown by Figure 16 is that the area 1-21 draws the limit of the *OBSO* area (blue line). We also see that close to Salah (3) there is an opportunity to score but no correlation with *disputed area* seems evident. Therefore, the *OBSO* model can determine which area is relevant to the score and which one needs to be avoided.

**Expert #2**, is a data visualization designer whose primary client is an English Premier League team regularly in the top-5. For him, Mané has the advantage because his closest defender is in a wrong orientation but the Firmino's pass has to be perfect to not race through to the keeper. Salah could have been an option but it would have been difficult for Firmino to pass it to him. Moreover, this pass would probably not led to a score but to assist Mané by a cross. About the occupation model, it seems to underestimate the burst potential of Mané, Figure 6, which is coherent with the answer of Expert #1. For the area 1-21, our model *indicates the likely optimum weight of a pass available to Firmino to play the ball forward to Mane*. For 3-6, it seems very narrow such as our remarks about the area 1-7 for Liverpool vs Porto, Section 4. He also predicts the area 1-21 at the same place that we placed it. Eventually, he made several remarks about visualization improvements. It could be more readable to set the pitch and the players in black and white and color only the areas. Fill the areas with plain color, remove markers, and increase the opacity. We also change the size of the players if they are in relation to an area or not. The result can be seen in Figure 17.

## 7 CONCLUSION AND PERSPECTIVES

We presented a definition and a process to calculate *disputed areas*, and applied it to soccer. Through multiple scenarios, we presented the step-by-step process, as well as its application using real-world

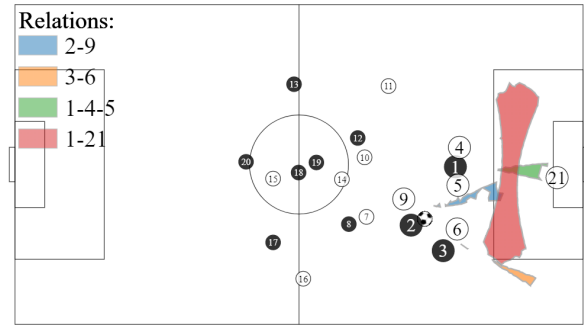


Figure 17: New visualisation of the *disputed areas* after feedback from Expert #2. It is the 52<sup>nd</sup> frame of Liverpool vs Newcastle from the Friends of Tracking dataset.

datasets. We also reported on feedback from sport analysis experts both on the relevance and visual design of those areas. As our work is independent of the underlying occupation models, and as it applies to a broad range of spatio-temporal situation, we expect further improvement and applications beyond sport. In particular, we discussed with urban analysts who are interested in visualizing *disputed areas* to understand redundancy in urban accessibility, i. e. the areas where too many public transportations serve some areas. This would help to prevent too much redundancy and thus provide better coverage of the city.

Furthermore, many improvements can be done on visualization and model implementation. As discussed with Expert #1 it could be more efficient to merge our model with another one or even providing more context by drawing it directly on the pitch [9]. Finally, another direction of improvement is to automatically detect players relations of interest as currently they are manually picked when creating the visualization.

## ACKNOWLEDGMENTS

The authors thank Gabin Rolland for providing the occupation models and code source. They thank Friends of Tracking for providing a public version of tracking data. They also thank the reviewers for the constructive feedback. Finally they thank Philippe Rivière for providing an early version of the prototype. This project was partially supported by the M2I project on Urban Mobility funded by the French Agency for Durable Development (ADEME).

## REFERENCES

- [1] M. Bostock, V. Ogievetsky, and J. Heer. D3: data-driven documents. 17(12):2301–2309. doi: 10.1109/TVCG.2011.185
- [2] J. Fernandez, F. C. Barcelona, and L. Bornn. Wide open spaces: A statistical technique for measuring space creation in professional soccer. p. 19.
- [3] H. Patel. Premier league 2019/20: Liverpool vs newcastle united - tactical analysis. Library Catalog: totalfootballanalysis.com Section: Match Analysis.
- [4] C. Perin, R. Vuillemot, and J.-D. Fekete. SoccerStories: A kick-off for visual soccer analysis. In *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, pp. 2506–2515. doi: 10.1109/TVCG.2013.192
- [5] C. Perin, R. Vuillemot, C. D. Stolper, J. T. Stasko, J. Wood, and S. Carpendale. State of the art of sports data visualization. 37(3):663–686. doi: 10.1111/cgf.13447
- [6] G. Rolland, R. Vuillemot, W. Bos, and N. Rivière. Characterization of space and time-dependence of 3-point shots in basketball. In *MIT Sloan Sports Analytics Conference*, p. 17.

- [7] W. Spearman. Beyond expected goals. In *MIT Sloan Sports Analytics Conference*, p. 17.
- [8] M. Stein, H. Janetzko, T. Breitzkreutz, D. Seebacher, T. Schreck, M. Grossniklaus, I. D. Couzin, and D. A. Keim. Director's cut: Analysis and annotation of soccer matches. In *IEEE Computer Graphics and Applications*, vol. 36, pp. 50–60. doi: 10.1109/MCG.2016.102
- [9] M. Stein, H. Janetzko, A. Lamprecht, T. Breitzkreutz, P. Zimmermann, B. Goldlücke, T. Schreck, G. Andrienko, M. Grossniklaus, and D. A. Keim. Bring it to the pitch: Combining video and movement data to enhance team sport analysis. In *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, pp. 13–22. doi: 10.1109/TVCG.2017.2745181
- [10] T. Taki and J. Hasegawa. Visualization of dominant region in team games and its application to teamwork analysis. In *Proceedings Computer Graphics International 2000*, pp. 227–235. IEEE Comput. Soc. doi: 10.1109/CGI.2000.852338