Real-Time Crowdsourcing of Detailed Soccer Data

Charles Perin, Romain Vuillemot, and Jean-Daniel Fekete, *Member, IEEE*

Abstract— This article explores how spectators of a live soccer game can collect detailed data while watching the game. Our motivation arouse from the lack of free detailed sport data, contrasting with the large amount of simple statistics collected for every popular games and available on the web. Assuming many spectators carry a smart phone during a game, we implemented a series of input interfaces for collecting data in real time. In a user study, we asked participants to use those interfaces to perform tracking tasks such as *locating players in the field*, *qualifying ball passes*, and *naming the player with ball* while watching a video clip of a real soccer game. Our two main results are 1) the crowd can collect detailed–and fairly complex–data in real-time with reasonable quality while each participant is assigned a simple task, and 2) a set of design implications for crowd-powered interfaces to collect live sport data. We also discuss the use of such data into a system we developed to visualize soccer phases, and the design implications coming with the visual communication of missing and uncertain detailed data.

Index Terms—Sports, Crowdsourcing.

1 Introduction

Sport, and soccer in particular, is a universal language that is understood by nearly everyone, regardless age, culture and country. Spectators usually carefully observe a game and then talk about actions, give general comments and predict outcomes. Spectators may also have strong opinions on players or teams, that may on one side bias their perception and analysis, but on the other reinforce their engagement to watch games and further pay attention to better support their position.

However, as visualization practitioners, we constantly struggle to find *detailed data* from actual observations of the game, such as player positions, type of ball kicks, etc. Even professional data used in our system SoccerStories [15] was missing details (such as the players' position at every instant) and sometimes low accuracy (positions are only provided by regions of the field). An example of such data is illustrated on Figure 1. Above all, this data is very expensive because collected by paid workers, and can only scale up to more games by adding more workers. As far as we know, manual annotation remains the single, yet most accurate and reliable way to collect this kind of data. Amateurs already collect general sport data (using tracking boards or spreadsheet) augmenting databases of specific websites [6, 21] or Wikipedia, but with a limited level of detail.

Regarding free data, we observed only the availability of simple statistics or opinions in summary articles and forums. This can be explained by the complexity of automatic collection of factual data. While humans are good at detecting players position and interpret their behavior, automatic recording by RFID sensors or GPS on players is costly and involves technical constraints. The analysis of existing live video feed with multiple points of view [17] and object tracking with occlusions is still a difficult problem [7]. Twitter produces interesting quantitative insights on volumes [19] and even identifies highlights [18] of sport events, but those data are not structured around particular game events and does not cover every single action.

As far as we know, manual annotation remains the single, yet most accurate and reliable way to collect this kind of data. Amateurs already collect general sport data (using tracking boards or spreadsheet) augmenting databases of specific websites [6, 21] or Wikipedia, but with a limited level of detail.

Having detailed data would foster research in the field of Information Visualization, but would also have a broader impact in the same way sport data-driven decision-making have been successful for estimating players market value [13]. Live and detailed data could assist *Coaches* for player substitution and tactic adjustments; *Journalists* to support live comments with statistics; *Players* for personalized reflection on their own performance; *Supporters* for game preparation and betting. With such data, advanced visualizations could also provide new analytics [4, 15] and support communication, such as with Real-Time Glyphs [12] for Rugby players' performance summaries and CourtVision [8] for Basketball players' shooting performances.

This quasi non-availability of detailed data contrasts with the important audience for sport. Last FIFA 2010 Soccer World Cup had a cumulative world television audience of one third of the planet's inhabitants. Spectators usually either stay at home, sit in a bar, or go to the stadium to watch a game. Can a fraction of those 3 billion hours of game watching be turned into a collaborative effort to collect detailed soccer data? With the advent of crowd-powered interfaces, the growing availability of smart phones and increased broadcasting of events, we explore how spectators can be involved into a global effort for collecting detailed data at low cost.

There are two main challenges we believe are critical to collect detailed soccer data: asking simple tasks with low cognitive effort, while preserving the pleasant experience of watching a game. The latter challenge is particularly difficult because users have to watch an external scene (on a TV screen on in real settings in the stadium) while entering some input. Because data input can be extremely complex, we propose to crowdsource the different tasks through adaptive input interfaces. We expect to achieve the same level of quality and expressiveness as experts, near real-time, and for free.

In this work, we explore how real-time crowdsourcing of detailed soccer data can be achieved by an adaptation of human computation and crowdsourcing techniques. In a user study, we asked participants to perform tracking tasks such as *locating players in the field*, *qualifying ball passes*, and *naming the player with ball* while watching a video clip of a real soccer game. Our main results are that 1) a crowd can collect detailed–and fairly complex– data in real-time with reasonable quality while each participant is assigned a simple task, and 2) a set of design implications for crowd-powered interfaces to collect live sport data. We also discuss the use of such data into a system we developed to visualize soccer phases, and the design implications coming with the visual communication of missing and uncertain detailed data.

2 Related Work

A crowd-powered interface is at the intersection of two research areas: *human computation* and *crowdsourcing*. To the best of our knowledge, there has never been any application of *crowd-powered* interfaces to

- *Charles Perin is with INRIA and Université Paris-Sud. E-mail: charles.perin@inria.fr.*
- *Romain Vuillemot is with INRIA. E-mail: romain.vuillemot@gmail.com.*
- *Jean-Daniel Fekete is with INRIA. E-mail: jean-daniel.fekete@inria.fr.*

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 4 October 2013.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

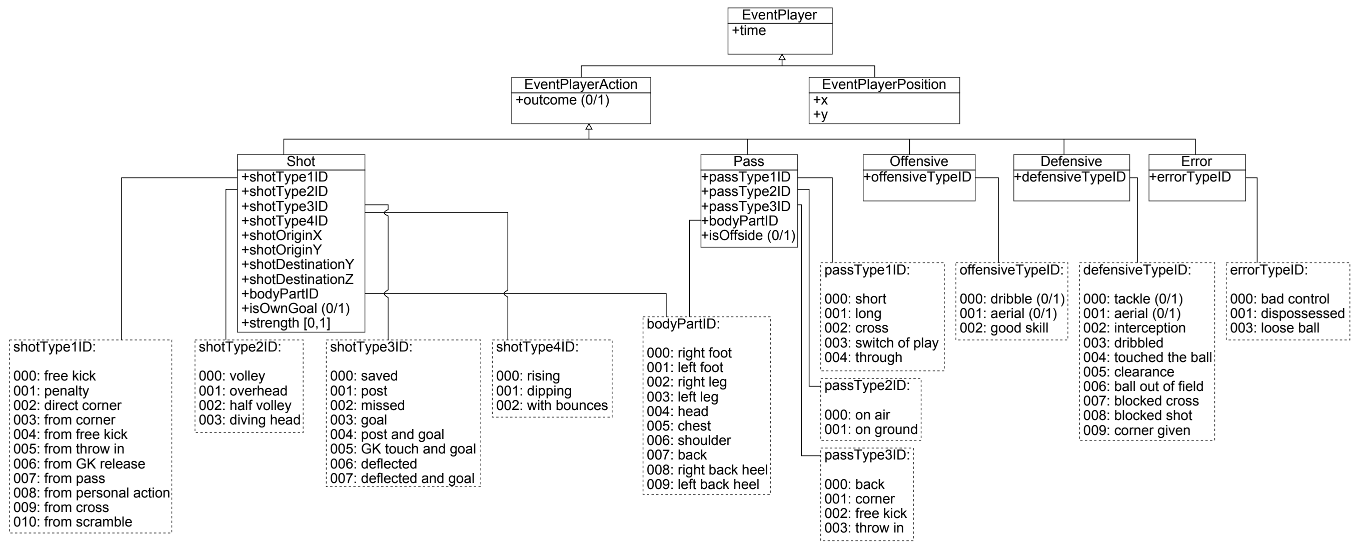

Fig. 1. Example of detailed actions in soccer, and their organization as a hierarchy.

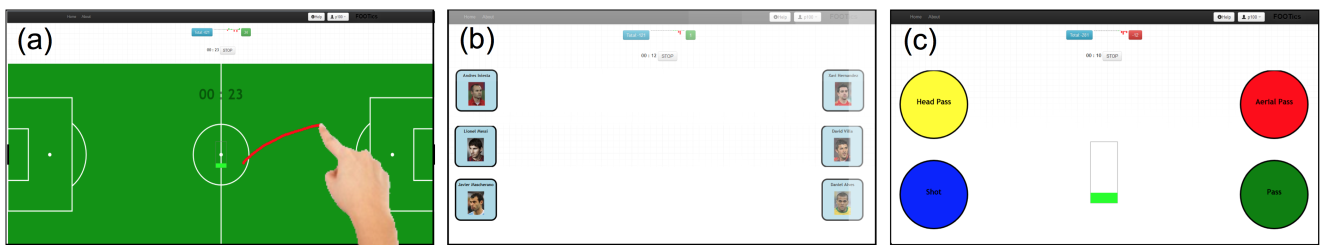

Fig. 2. Three tasks we implemented into three tracking interfaces: (a) *locating players in the field* by finger-tracking (WHERE), (b) *naming the player with ball* by pressing his icon (WHO), and (c) *qualifying ball passes* by selecting the type of the event (HOW).

collect detailed sport data, and more generally any data at a low level of details and for different types of data structures.

Human computation [11, 16] has been a topic of much research for solving problems too difficult for computers, by asking humans to solve simple tasks. It originates with simple and enjoyable tasks, but became applicable to real life problems such as collective text writing or trip planning [22]. Training users is performed under the explicit assignment of the task, or by qualification using an indirect task. The validation is often performed by votes or once a threshold (*i. e.,* once a number of players gave the same answer) is reached [11].

Crowdsourcing [16] is the use of an unknown public to perform tasks, often simple enough to be executed by anyone. It is generally a faster and cheaper manner than traditional ways (*e. g.,* recruiting experts), and workers remain paid. Available workers are recruited through online markets (*e. g.,* Amazon's Mechanical Turk). The overall process can be divided into three steps: *break down complex tasks into micro-tasks*, *dependencies management*, and *quality control* [9]. At the intersection of human computation and crowdsourcing advanced mechanisms have been developed to reduce crowd contact latency [1] and advanced task partitioning to consider global constraints [9]. It results in additional tasks or incentives to pay to obtain such a quality of service. Crowd-powered interfaces dedicated to sport have been applied to identifying football highlights [18], by processing public Twitter data as an implicit source of crowdsourcing.

Despite previous wokr already investigated video annotation of sport events [20] using Amazon's Mechanical Turk, it is valid under certain constraints for the workers: first, the tracking is not performed in real time; second, the video is embed in a dedicated interface, which is not faithfull to most of the spectators' habits, *i. e.,* watching the game full screen on a separate device.

3 Crowdsourcing Detailed Soccer Data

3.1 Characterizing Detailed Soccer Data

Soccer is a team-based contact sport between objects of interest that are players interacting with each others and resulting in *continuous* sequences of actions. Other objects of interest can be included such as the ball or referees. Objects of interest and their interactions are highly multidimensional data, always including a timestamp and a position on the field.

We identified the different qualifiers for characterizing soccer actions. Figure 1 illustrates the diversity of data inputs neededed to reach a low level of detail. For example, a player (categorical data type) catches the ball in the middle of the field, performs various types of ball touch (ordinal type): control, dribble, pass and shoot, and kicks the ball to another player in different ways: soft, high or long pass. Additional qualifiers resulting from the interpretation of the spectator can be added, such as success of an event (binary) and its quality (ordinal value). Once a game is over, measures such as ball possession or number of attempts can be computed. Generally, a soccer game is made of long sequences, but sequences can also be short, visually complex and subject to high variation of interpretation by users[1].

3.2 Guidelines for Tracking Detailed Soccer Data

To support the tracking of detailed data, we derive the five following guidelines inspired from human computation and crowdsourcing:

R1 *Familiarity with soccer:* since tracking has to be performed quickly when watching a game, viewers must have some previous experiences (*e. g.,* from watching previous games, playing

[1]Controversies around referees' decision is an example of such variation of interpretation in soccer.

in real life or playing video games) to allow players detection and action understanding. Consequently, a minimum amount of familiarity is required.

R2 *Split tasks into simple and independent ones:* soccer data collection has to be split into simple tasks, requiring low cognitive load and minimum ambiguity. Also, tasks must be independent from previous tasks or should not required the game to be finished (*e. g.,* such task as finding highlights in games cannot be done in real time because it requires all the actions of the game to be available).

R3 *Quick recruitment:* any player watching a game can be recruited and start a task very quickly, without any qualification stage. Data validation will be done after data have been collected (See *Validation:*).

R4 *Sustaining attention and engagement:* by providing tasks requiring permanent attention with a balance between too simple and too complex tasks in order to keep participants aware and prevent drop outs.

R5 *Validation:* while data is collected, some mechanism can be used for validation using such strategies as vote or using a ground truth like an existing dataset of detailed data.

3.3 Implementation of Simple Tracking Tasks and Validation

We picked up three simple tracking tasks (Figure 2), focusing on players and the ball, which represent the main objects of interest in soccer. We implemented those tasks in three tracking interfaces using the previously explained principles. We decided to test them with a validation dataset (**R5**), enabling a visual feedback on accuracy and to progressively adapt complexity of the task (from simple to complex **R1**, **R2**) to participants' performance for engagement (**R4**). We detail quality measures because a ground truth exists, if it does not exist then no adaptation or validation can be made. Using those three interfaces detailed below, we expect to reconstruct full soccer game sequences.

Where is the player? (WHERE): *continuous* tracking of a player's position on the soccer field. The participant tracks one player and inputs its position into the interface representing a soccer field. The task begins with low accuracy requirement, but progressively levels up as the participant becomes more accurate. The tracking area radius, which represents a circle of tolerance around what should be recorded, decreases when the participant performs well and increases when he has wrong results. We measure the tracking error, *i. e.,* the distance between the finger and the expected target to be recorded.

Who has the ball? (WHO): naming the player with the ball. The task consists of pressing a player's icon during the time he has the ball. The task begins with one single player to be tracked, and more players are added or removed according to sustained accuracy: one player is added if the accuracy is higher than 50% and one player is removed if the accuracy is lower than 30%. The minimum number of players to track is one and the maximum is six.

What type of ball kick? (HOW): identifying the different ball kicks. The task is to qualify ball kicks among a list (pass, shoot, aerial pass and head) every time a player from the tracked team kicks the ball. Similar to the player's location, we change the expected accuracy over time, proportionally to the participant's performance. However, the accuracy is not spatial for this task but temporal using a threshold for tracking reactivity. The threshold to do a good tracking evolves between $100ms$ and $1000ms$ according to the participant's errors.

The philosophy is to start with an interface requiring few input or accuracy, which then becomes progressively complex if the user performs well. This is a process that is well established in video games where the performance is rewarded with more complexity, over *levels*.

3.4 User Engagement

To maintain participant's engagement, we designed the interfaces to provide feedback according to his performances. Feedback addresses the validity issue when a participant is currently performing a task. We extended the interfaces to inform the participant on his own tracking performance and progress across levels. A barchart informs the user on his performances, where red and green bars indicate bad and good inputs, respectively (Figure 3(a)). A stack is also filled according to the current level (Figure 3(b)). The player then receives a score depending on its level and accuracy ((Figure 3(a), rectangle on the right)), colored red or green for bad and good inputs, and his total score is displayed ((Figure 3(a), blue rectangle)). The homepage of the application also shows an history of the highscores.

Fig. 3. Visual feedbacks on user's performance: (a) total score, barchart of recent inputs' results, and last input score; (b) stack indicating the current level.

More complex feedback can be embedded into the buttons, such as to display each player statistics for notational analysis [12], but we discarded those as it added complexity to an interface the user was not permanently looking at.

3.5 Implementation

We implemented the three interfaces in JavaScript using the D^3 library [2]. They run in any modern web browser and are mobile friendly, which means we did not have to implement the interfaces as native applications. To start collecting data, participants are required to sign in and pick up a task for a live game, and data are sent to a MongoDB [14] hosted on a remote server. The server compares the inputs of each participant to the ground truth, either annotated by experts, either computed according to all participant's results aggregated. Participants receive their score on their interface, but also the history of their previous tracking.

4 EXPERIMENT

We conducted an experiment to assess data collection using the previous three interfaces (WHERE, WHO, and HOW, Figure 2). We recruited 12 (all male) unpaid volunteers of two categories: 6 never-to-rarely watch soccer (G1) and 6 regularly watch soccer (G2). Average age is 29.8 years (min: 23, max 39). 16% use a touch screen daily.

As dataset, we selected two videos: a 4-minute one for the training session, and a 10-minute one for the experiment. Both are from the UEFA 2011 Champion's League final between FC Barcelona and Manchester United. We chose this particular game because it is already known by football fans, and the Barcelona team offers a pleasant and technical game with long sequences of events (reducing the replays and interruptions). Each sequence was manually annotated by two experts who processed the video clips frame by frame, and double-checked to prevent errors. During the sequences, the overview camera was used nearly 90%, and the remaining 10% were close-up views.

Those video annotations serve as ground truth for the controlled experiment. However, in real settings, the truth does not exist as no real-time detailed data exist. Our approach is then to consider participant's results as ground truth, and to evaluate individual accuracy by its distance to it.

Participants sat in front of a 19 inch LCD monitor displaying the game with sound and sat at the distance they felt the most comfortable with. They did not interact with the video, which was launched by the operator. All the tracking was performed with a Samsung Galaxy Tab 10.1 tablet with touch interaction. Participants were free to hold the tablet and interact with it as they wished, to get closer to a real-life condition. The order of tasks was counterbalanced within each group (G1 and G2) to minimize learning effects. All the participants filled out a post-experiment questionnaire designed to capture their feedback on the tasks, game understanding and engagement. The experimental session lasted in average 60 minutes. The participants were provided general instructions and details for each task, such as screen-shots of the levels, and task-dependant informations: a visual characteristic of the players they have to track (for WHERE and WHO) and differences between ball kicks (for HOW).

5 RESULTS

Our experiment resulted in a large collection of temporal records of events and positions. We quantitatively measure the performance of each group for each task. We also drill down into the aggregated values to explore individual performances and effect of the time and position of the players to track on the participant's accuracy.

Data were analysed using a Student's t-tests on our measures. Figure 4 shows the results for G1, G2, and all participants combined for each task. We also show the results for three different aggregations, considering G1, G2, and all participants. The aggregations were performed using a majority-vote and the DBSCAN clustering method [5].

5.1 WHERE

We did not find significant differences between G1 and G2 for this task. However, the standard deviation is higher for G2, implying a more heterogeneous tracking. The aggregation of all data is better than for the best participant (99.9% of good tracking, and a low distance error). Nevertheless, 23.4% of the expert annotations are missing (3006 records over 3600 annotations): the aggregation process removes the false-positives and increases the percentage of good tracking, but does not recover the missed events.

Figure 5 illustrates the results for WHERE for a 3-minute sequence: blue areas represent the player positions that have been missed; red areas the player positions that were tracked by mistake. Clearly, patterns appear: nearly all the participants have blue areas at the same time, *i. e.,* need some time to realise that the player appeared on the screen. By watching the video, we observed this is generally due to a camera view change. Another clear pattern is the similar absolute distance error for all participants when tracking and this is explained by the tracked player's position on the field: tracking was more accurate when the player was close to field landmarks (*e. g.,* the field lines) than when he was far away from any landmark. The last observation is the impact of the soccer field's projection. We found that the error was 10 times higher in x (field shorter length) than in y (field longer length). We hypothesize that this is due to the lateral position of the camera which makes the tracking of far away players less accurate. It is also explained because the longer length of the field is less distorted while participants have to mentally inverse the projection of the shorter length of the field.

Figure 6 shows and explains the error rates at the beginning and end of a typical tracking time span (from the player's appearance to his disappearance) as well as the landmarks importance. Calling P the player to be tracked, when P appears in the camera view (0), participants started the tracking after a varying reaction time (blue areas or high error rates). Then, the error is low because P is close to the central field line and circle, providing useful landmarks to minimise the error (1). Between (1) and (2), P progressively goes far away from the landmarks and the error increases. Between (2) and (3), P goes back to the center of the field and the error decreases. Between (3) and (4), P is close to landmarks but accelerates strongly, resulting in an increase of the absolute error due to participant's reaction time. Between (4) and (5), P is far from any landmark but walks very slowly, making the tracking easier. Finally, when P leaves the camera view (6), there is usually a short time during which participants keep tracking him for a while, resulting in wrong tracking (red areas) until they realize the player left the camera viewport.

5.2 WHO

We found a significant difference between G1 and G2 for the number of tracked events ($p < .05, m_{G1} = 70.8, m_{G2} = 106.8$). The percentage of good tracking was better after aggregating all the results than only one of the two groups. Also, when between 8% and 35% of

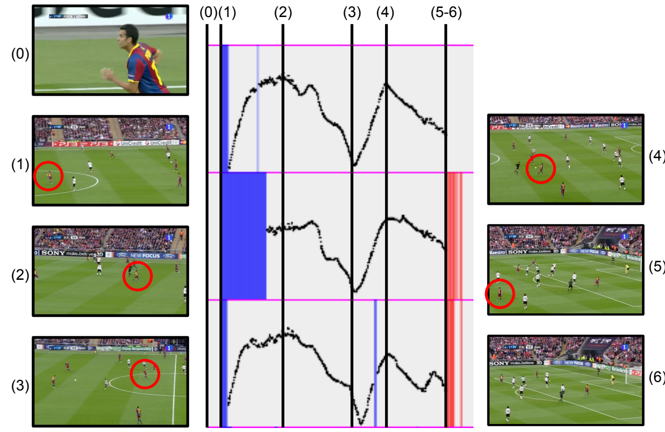

Fig. 6. A typical interaction during a time interval, extracted from Figure 5, showing how the camera view, the player position on the field, his proximity to visual landmarks, and his speed can affect the tracking for WHERE, similarly for almost all participants.

the events were missed by the participants, only 10% were still missing after aggregating all the results. Finally, the error for G1 had high standard deviation when G2 was more homogeneous.

Among the tasks, WHO was judged the most difficult for non fans. Indeed, it involved the participants to recognize the players (**R1** is not supported). Having better knowledge of the players identity and so better answers, participants in G2 increased in level and had more players to track, which explains the significant difference in terms of number of tracked events.

Non-soccer fans can be recruited for this kind of tracking tasks: fans have higher standard deviation but the mean is very close between G1 and G2 for the % of good tracking and the error, supporting (**R1**). Overall, participants responded with an average $200ms$ supporting the request for quick responses (**R3**).

5.3 HOW

There was no significant difference between G1 and G2, with a % of good tracking above 80% for both groups. The aggregation of participants results gave a better result for G2 and all participants confounded than for G1. 4% of the events were missing after the aggregation of all participants results.

We observed two types of error for this task: temporal errors (the delay to track the event), and qualification errors (wrong event recorded). We think that results may be more accurate by desynchronizing the two types of input: first, the worker inputs the time at which the event occurs, then he qualifies it if he has time. Errors are also due to events frequency: too many events in a short time span increase the difficulty to track all, and too less events don't maintain the user's attention.

Users enjoyed task HOW the most (ranked 7 times first and never last). Questionnaire's answers confirmed our intuition that the similarity between this interface and game console devices really engaged the participants. Also, contrary to WHO, the increasing difficulty of the task modifies only the requested reaction time and does not modify the interface layout, keeping its spatial consistency.

5.4 Questionnaire results

Participant's answers to the questionnaire revealed similarities and differences between the two groups.

Tasks difficulty: there is a clear trend in task difficulty estimation by the participants, G1 and G2 confounded: HOW is the easiest task, followed by WHERE, and WHO is the most difficult one.

Interest in data representations (Figure 8): half and more of the participants in G1 are not interested in getting additional informations about the game, the players, and statistics. On the opposite, all the participants in G2 are interested in such data. Most of them would like

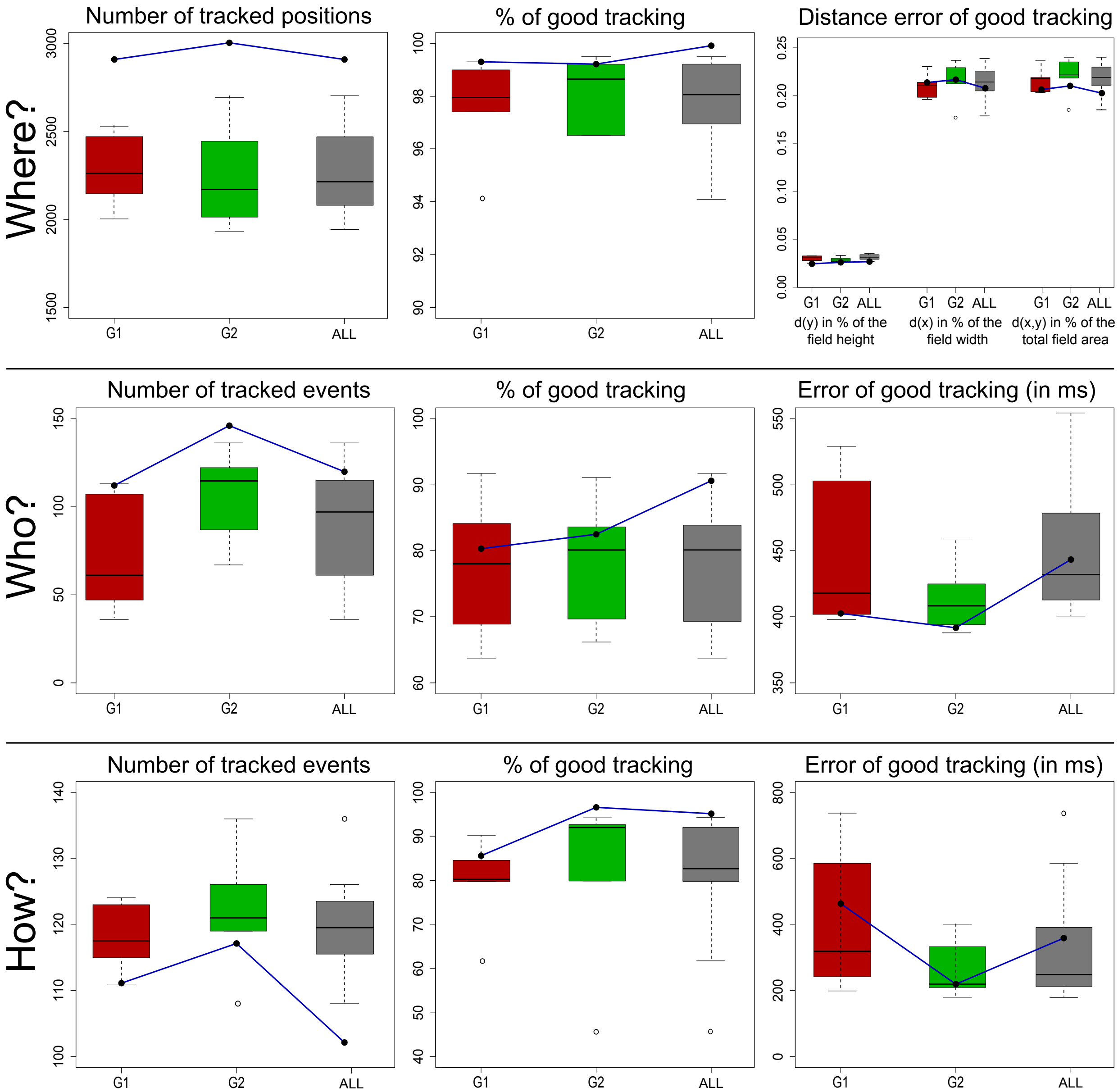

Fig. 4. Results for the 3 tasks for G1, G2, and all participants. The % of good tracking represents the quality of the recorded data, without considering the missed events. The blue lines are the values after aggregating G1, G2, and all.

to access written, statistical, and graphical representations of the data during the game or after (only one would like to access these before).

Why would they collect data? (Figure 7, left): the mean values for all questions for G1 and G2 are close. The main difference is that non soccer fans (G2) are more likely to participate if they can choose their task and connect to other people (social game).

Which incentives? (Figure 7, right): unsurprisingly, paying workers would be a good incentive, especially for soccer fans (G1). We also observe that G1 participants are more interested in collecting data for fun and for free than G2 participants, who are more interested in getting a non-money reward such as a free access to game broadcasts.

How long would they collect data?: on average, participants would collect data between 5 and 9 minutes, half the participants saying that they would do it between 10 and 15 minutes. Collection times are similar for both groups.

Interest in consulting the data?: none of the participants in G1 answered yes to this question while 5 out of the 6 participants in G2 answered yes. Again, this is not a surprise that soccer fan are more interested in soccer data than other peoples.

5.5 General observations

Most participants looked at their score and visual feedback during the game, but did not find it relevant. However, they are eager to compare their performance once the game is finished, most with their social network and real-time leader boards.

From the informal feedback we received, the vast majority of subjects enjoyed the tracking. Some in G1 who did really hate soccer were positively surprised and happy to do the experiment: it was the first time they understood some soccer strategies and learn a lot of things about soccer. The participants compared this experience to playing a video game, having a lot of fun and not considering the task as tedious.

6 Discussion

Our early results led use to identify design implications and opportunities to improve detailed data collection. While we discuss those with Soccer, we think they may have broader application such as to other sports or domains with spatio-temporal data.

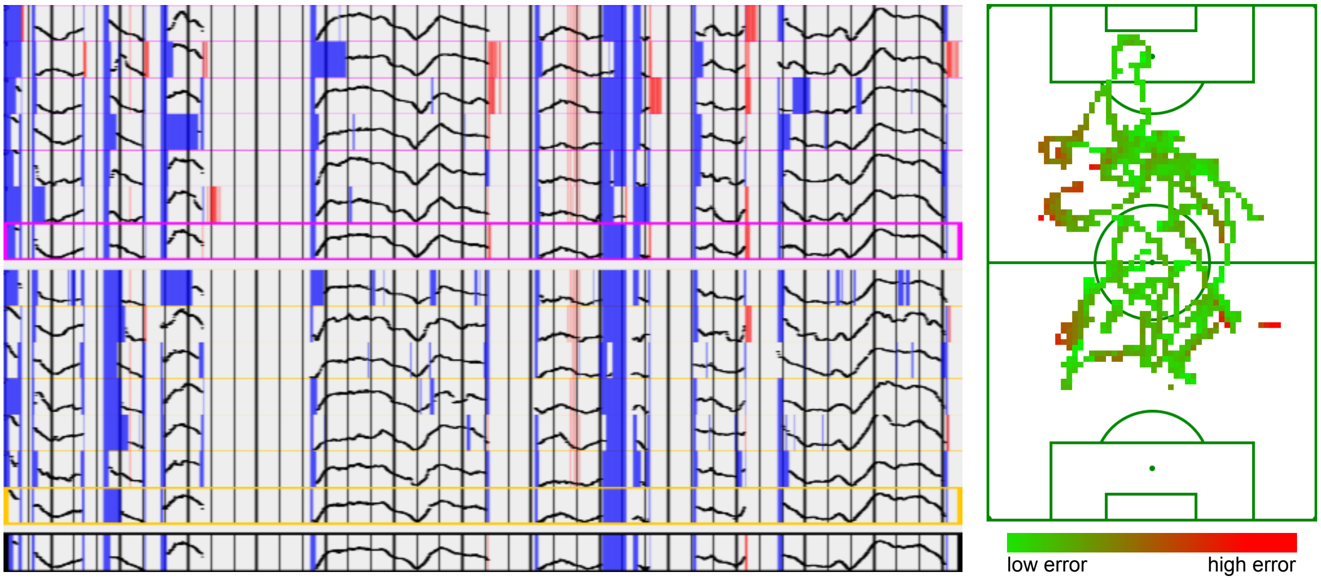

Fig. 5. Results for all the participants for a 3-minute sequence performing the task WHERE for one player. G1 is pink and G2 is yellow, with corresponding aggregations in bold rectangles. The black rectangle at the very bottom is the aggregation for all participants. Blue and red areas are missed and wrong tracking, respectively, and black line charts are plots of the absolute distance error between the user's input and the actual player's position. On the right is the heatmap of absolute distance error on the soccer field (light green is low error, light red is high error).

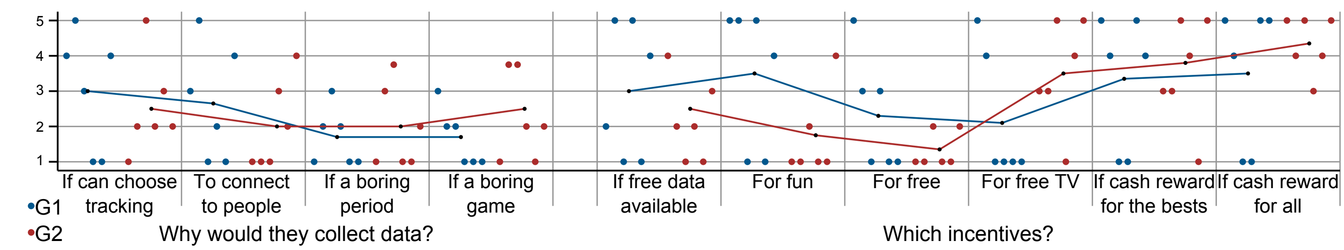

Fig. 7. Participant's answers to the questionnaire: for which reasons would they collect data and which incentives would encourage them to do so. Each colored dot is a participant's answer on a likert scale from 1 (not at all) to 5 (a lot) and the blue and red lines are mean values for G1 and G2, respectively.

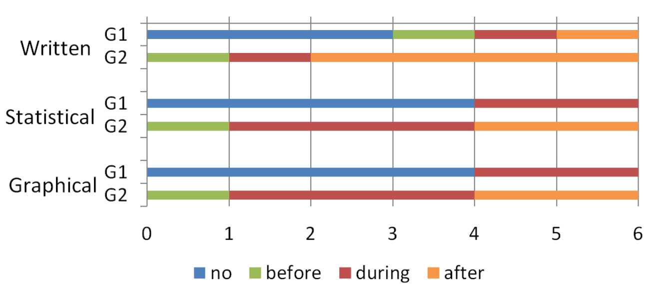

Fig. 8. Participant's answers to the questionnaire: Type of data representation they would be interested in, and when.

6.1 Design Implications

The following design implications are issued from our results, from our experience in the design of our interfaces, and from the study:

- Non-soccer fans can be recruited for tracking simple tasks, under the condition of a short preliminary training.
- Non-soccer fans and fans can both perform tasks that do not require any prerequisite about soccer.
- Tracking time and data qualifiers may be separated tasks, unless there are few qualifiers or enough time to set the qualifiers.
- Landmarks (*e. g.,* lines on the field) or any visual/temporal tracking aid are useful for participants to calibrate themselves for position tracking.
- Events (with duration) that appear or disappear in the view port of the video lead to missing data at the beginning and interpretation at the end of the tracking.
- Using change of difficulty levels and keeping the score visible can help maintain attention and may also support more complex tasks for participants with good results.
- Scene complexity increases the error rate, but low complexity may lead to loss of attention and needs a special treatment.
- Social reward is not as important as money but may advertise friends of the participants to join the process, and lead to better performance for the participant and retainment.

6.2 Limits in Input Devices and Interfaces

We identified the main bottleneck, in the quantity of collected data and its quality, as being the design of the input interface and the type of device we used. As we think the interfaces can easily be improved, the specific use of a tablet during our experiment is a hard limit. As a reminder, we used a tablet because they become extremely popular. However, they require the user to look where to put his finger, even once familiar with the interfaces. A solution to prevent such a split of attention would be the use of video games gamepads as input devices. This is indeed very common for advanced video games players to know exactly where the input buttons are, and they do not need to look at both the screen and the gamepad. The mapping of buttons with the task would follow the same one as with soccer video games. How-

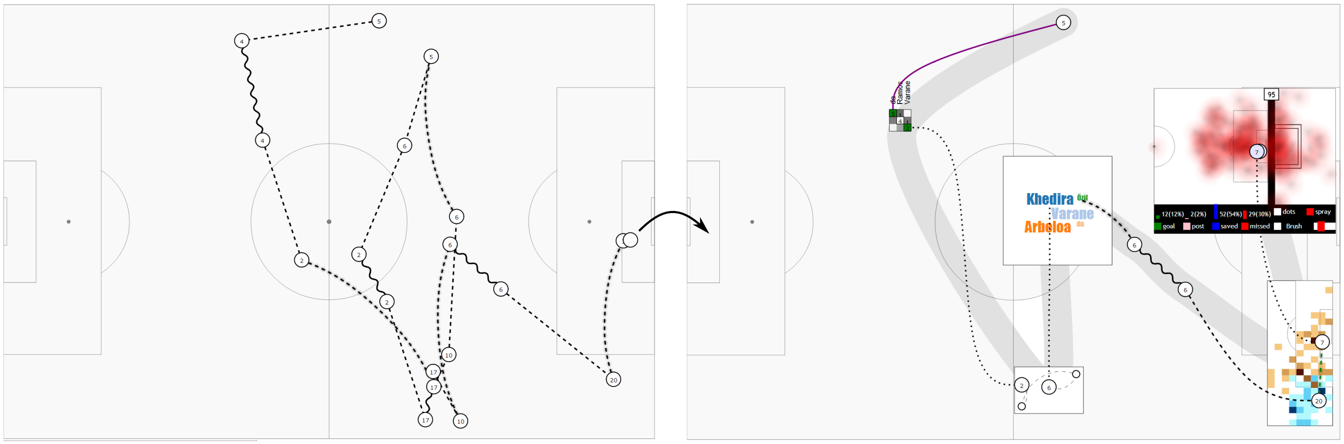

Fig. 9. SoccerStories [15] shows soccer phases either as node-link (Left) or as a series of connected visualizations (Right).

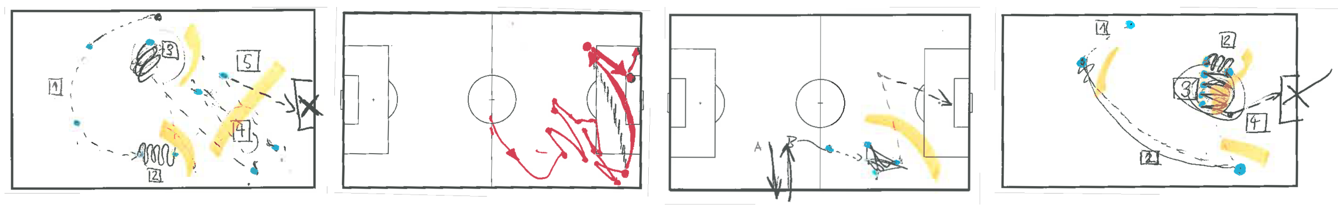

Fig. 10. Hand drawings of soccer phases during a participatory design session. Each drawing shows the design variations of visualizations for soccer data, and the use of many visual variables in this context.

ever gamepads are not as widespread as tablets and mobiles, and few people are very familiar with them. Further investigations are required to study their use.

A second bottleneck relies in the user lookup time for qualifiers, such as for the HOW task. Indeed, qualifiers are organized as a tree (Figure 1) which cannot fully be displayed on a small screen making it difficult to browse rapidly. Browsing such a tree to lookup for a pass requires to explore *player event*, the type of *pass*, its *outcome* (success or not), and other *pass qualifiers* (*e. g.,* short, long, cross), *ball qualifiers* (*e. g.,* on air, on ground), *origin qualifiers* (*e. g.,* corner, free kick), and *body qualifiers* (*e. g.,* right foot, head, chest, shoulder). Selecting the right value requiring breadth and depth search. A strategy to reduce the lookup time, among others, is to split the tree in multiple trees, resulting in specific tasks similar to the HOW interface. For example, detailed shots data can be tracked by a pool of participant indicating that the event is a pass and five other pools of participants indicating the `outcome`, the `passType1ID`, `passType2ID`, `passType3ID`, and `bodyPartID`. Another approach might be to use other input modalities such as voice recognition or a gesture dictionary to browse such a data structure while not being consuming too much visual attention.

7 OPPORTUNITIES: ENHANCING SOCCERSTORIES

We now explore how a specific visualization system can be enhanced with detailed data: SoccerStories [15] (Figure 9) which represents soccer phases as series of visualizations. A *Phase* is a series of actions by one team, and is separated from another phase by transitions that occur whenever the other team touches the ball. For each phase, SoccerStories groups actions as series of connected visualization, on a soccer field. It thus provides a visual signature of a subpart of a game, that has been highly appreciated by experts for discovery and communication.

Phases detection. SoccerStories was originally built using the ball's position on the field and the type of actions players performed with it. Phases are immediately detected once the other team touches the ball. However, experts often mentioned that phases do not always start or end when a team touches or loses the ball. It rather depends on a global team domination over certain period of time, even if the team loses the ball during that period. By using a HOW task with the two teams as attributes, phase detection can be crowdsourced and be added over existing data as an extra descriptor for actions.

Visualization. Because each phase is divided into groups of actions to be visualized in a specific way, their segmentation is also an issue. SoccerStories only implements a simple grouping algorithm, which was considered good by experts we collaborated with, but can be improved in the same way as for phases detection. However, finding the semantic of the groups of actions (*e. g.,* series of pass, goal, etc.) is difficult and can be done by means of a WHO task that identifies groups instead of players. Since the visualization directly depends on the semantic of the group, if the latter is improved, then the former is immediately improved as well.

Providing context. The position of surrounding players on the field may provide context for each action. Indeed, they can be represented as points collected by a WHERE task. Their density and flow over time would indicate valuable information for an action. Eventually a new type of data can be collected: gaze direction of the players. Such data is particularly valuable for team sports, where players movements without the ball are crucial to understand a phase of a game. However, gaze direction is a vector with–above the position–a direction and a magnitude, requiring new interfaces that augments the WHERE one.

Uncertain data visualization. As collected data may be missing or erroneous, it is still unclear how to best convey those. We explored in a participatory design session (Figure 10) the visual variations that participants would naturally use when being asked to draw in details a soccer sequence they just watched. The drawings resulted in a similar flow on the field, across the participants. But surprisingly, they extensively used visual variables (*e. g.,*, lines types, thickness, ..) and symbols (*e. g.,* arrows, numbers, ..). It seems promising to investigate further the use visual variables to represent detailed data visualizations, such as blur [10] for focus+context and sketchiness [3] for uncertainty, in the context of SoccerStories, and sport visualization in general.

8 Conclusion and Perspectives

This paper presents an exploratory work that is a first step towards understanding what type of data and with which level of quality we can expect participants to collect in real-time. Based on initial requirements for this type of crowd-powered interface, we tested three simple tasks on a soccer game. It already emphasized the difficulty of game observation for amateurs, but also the impact of video distortion as a viewer of a projected scene and difficulties to identify players for most of the participants. We summarized our findings in design implications, that we think can be extended to other spatio-temporal sports. Our short term perspective is a follow-up experiment in real settings so as to cope with drop-outs, distractions (*i. e.,* talks, snacks, drinks), and assess different behavior depending on participant's favorite team [18]. Another perspective is asking participants their subjective rating on players and actions, and explore how existing systems or visualizations may be enhanced with such data, as well as the design of new ones.

Acknowledgments

We are grateful to Adrien Fekete (a.k.a. Junior) for helping us preparing the experiment. We also thank all the participants for their time and feedback.

References

[1] M. Bernstein, J. Brandt, R. Miller, and D. Karger. Crowds in two seconds: Enabling realtime crowd-powered interfaces. In *UIST '11*, pages 33–42. ACM, 2011.

[2] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, 2011.

[3] N. Boukhelifa, A. Bezerianos, T. Isenberg, and J.-D. Fekete. Evaluating Sketchiness as a Visual Variable for the Depiction of Qualitative Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2769–2778, Dec. 2012.

[4] A. Cox and J. Stasko. Sportsvis: Discovering meaning in sports statistics through information visualization. In *Compendium of Symposium on Information Visualization*, pages 114–115. Citeseer, 2006.

[5] M. Ester, H. peter Kriegel, J. S, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press, 1996.

[6] Football Outsiders. `http://www.footballoutsiders.com/`.

[7] P. Gabriel, J. Verly, J. Piater, and A. Genon. The state of the art in multiple object tracking under occlusion in video sequences. In *Advanced Concepts for Intelligent Vision Systems*, pages 166–173, 2003.

[8] K. Goldsberry. Courtvision: New visual and spatial analytics for the nba. *MIT Sloan Sports Analytics Conference 2012*.

[9] A. Kittur, B. Smus, S. Khamkar, and R. Kraut. Crowdforge: Crowdsourcing complex work. In *UIST '11*, pages 43–52. ACM, 2011.

[10] R. Kosara, S. Miksch, H. Hauser, J. Schrammel, V. Giller, and M. Tscheligi. Useful properties of semantic depth of field for better f+ c visualization. In *Proceedings of the symposium on Data Visualisation 2002*, pages 205–210. Eurographics Association, 2002.

[11] E. Law and L. Ahn. Human computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 5(3):1–121, 2011.

[12] P. Legg, D. Chung, M. Parry, M. Jones, R. Long, I. Griffiths, and M. Chen. Matchpad: Interactive glyph-based visualization for real-time sports performance analysis. In *Computer Graphics Forum*, volume 31, pages 1255–1264. Wiley Online Library, 2012.

[13] M. Lewis. *Moneyball: The art of winning an unfair game*. WW Norton, 2004.

[14] Mongo DB. `http://www.mongodb.org/`.

[15] C. Perin, R. Vuillemot, and J.-D. Fekete. SoccerStories: A Kick-off for Visual Soccer Analysis. *IEEE Transactions on Visualization and Computer Graphics*, Oct. 2013.

[16] A. Quinn and B. Bederson. Human computation: a survey and taxonomy of a growing field. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 1403–1412. ACM, 2011.

[17] H. Saito, N. Inamoto, and S. Iwase. Sports scene analysis and visualization from multiple-view video. In *Proc. of International Conference on Multimedia and Expo (ICME04)*, 2004.

[18] A. Tang and S. Boring. # epicplay: crowd-sourcing sports video highlights. In *CHI '12*, pages 1569–1572. ACM, 2012.

[19] Twitter #Euro2012. `https://euro2012.twitter.com/`.

[20] C. Vondrick, D. Ramanan, and D. Patterson. Efficiently scaling up video annotation with crowdsourced marketplaces. In *Proc. ECCV'10*, pages 610–623, Berlin, Heidelberg, 2010. Springer-Verlag.

[21] Yinzcam.com. `http://www.yinzcam.com/`.

[22] H. Zhang, E. Law, R. Miller, K. Gajos, D. Parkes, and E. Horvitz. Human computation tasks with global constraints. In *CHI '12*, pages 217–226. ACM, 2012.