UNIVERSITY OF TROMSØ UIT

FACULTY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE

Ruoksat

A system for capturing, storing and presenting the digital footprints of coach knowledge and execution

Kim-Edgar Sørensen

INF-3981 Master's Thesis in Computer Science
June 2013

Abstract

Acknowledgements

Table of Contents

List of Acronyms

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Soccer teams and athletes are constantly looking for possibilities to gain advantages over opponents. In soccer your next opponent is analyzed down to the smallest details to find weaknesses and strengths. All this to be able to take advantage of your opponents weak points and handle their strengths. An example of a typical weakness is a team playing 4-4-2 and gives up space between the lines (back-four and midfield). An strength can be finding this space, typical with a number 10 player.

There are several ways to gather information about your opponent. From looking through whole matches to advanced tools, which can highlight key information for you.There are two main aspect of the analysis process; First you need to gather the information and secondly is how to present the information, usually for the coaching staff.

1.2 Problem definition

This thesis will develop a system complementing the Muithu and Bagadus systems. Focus will be on soccer opponent analytics, where a data repository need to be developed capturing important events relevant for this type of analytics. Specially we want to identify the breakthrough players. A user interface component providing the core information about the opponent also has to be developed.

1.3 Interpretation

Our thesis is that providing an infrastructure for capturing events like attacks leading to an attempt on the opponent goal, and presenting this information through an user interface with graphs and illustrations. We are interested in how long it typically takes to manually capture all relevant events from a single match. For this we need to develop a data model that reflects what we want to capture from each attack. This for being able to do relevant queries on the data afterwards and gain insight.

Focus will however mainly be on developing a domain model

The design and development processes will be performed in collaboration with staff of the Norwegian soccer club Trms IL. An end-user comparison of the system against currently available tools will perform a final evaluation, and the result of this evaluation will be used to conclude the thesis.

1.4 Methodology

The final report of the ACM Task Force on the Core of Computer Science [1] divides the discipline of computing into three major paradigms:

- *Theory*: Theory: Rooted in mathematics, the approach is to define problems, propose theorems and seek to prove them in order to determine new relationships and progress in computing.
- *Abstraction*: Rooted in the experimental scientific method, the approach is to investigate a phenomenon by stating hypothesis, constructing models and simulations, and analyzing the results.
- *Design*: Rooted in engineering, the approach is to state requirements and specifications, design and implement systems that solve the problem, and test the systems to systematically find the best solution to the given problem.

For this thesis the design process seems to be the most suitable out of the three paradigms. The design process consist of 4 steps and is expected to be iterated when tests reveal that the latest version of the system does not satisfactorily meet the requirements.

- *State requirements and specification*: A need or problem is identified, researched, and defined.

- *Design and implement the system*: Data models and a system architecture are designed. Prototypes are implemented.
- *Test the system*: Assessment and testing of the aforementioned prototypes.

1.5 Outline

- *Item 1*: Presents some background information related to the thesis
- *Item 2*: Describes the requirement specification
- *Item 3*: Describes the general system model

Chapter 2

Background

2.1 Related work

In this chapter we present some background related to creating a multi-media data repository. There are two main approaches: capturing data manually often by human annotating it, and automatic capturing. Automatic may or may not have a substep where a human has to filter/annotate the data.

2.1.1 Manually capturing

2.1.1.1 Opta

One of the big players Opta uses manual input to create their data repository. They have editorial teams across the world that captures data manually for the most popular soccer leagues. For example to capture statistics for one match, 3 humans have to be involved to be able to annotate all data. The data is captured via an application specifically created for the purpose of capturing data as quick and easily as possible. The editorial teams of Opta need to be able to identify a player, registrate a pass his made or a tackle, in a very short time to be able to keep up with the pace of the game. They study things like which shoe color a player has to be able to quickly identify the player.

Opta capture all types of actions like passes, type of pass, attacks, and interceptions. For each action they log they add a series of description tags like pitch coordinate, player, team and time-stamp. For every single pass they registrate if it was a through ball, normal ball or even a headed flick on from a long ball. For shots they registrate the foot it was kicked with, if it was a volley and so on . All this is done while the match is playing. About 1600 individual events are recorded in a standard match.

<Event id="290575408" event_id="5" type_id="1" period_id="1" min="0" sec="5" player_id="20856" team_id="810" outcome="1" x="44.6"y="61.1" timestamp="2007−08−12T13:00:24.827" last_modified="2007−08−12T13:00:25">
<Q id="1774596260" qualifier_id="141" value="91.6"/>
<Q id="1429253465" qualifier_id="140" value="49.9"/>
<Q id="1084400575" qualifier_id="56" value="Back"/>
</Event>

Here is an example of an event registrated in the Opta database. The event has a series of qualifiers describing it. Except from the obvious as timestamp and last modified dates we see that the player id, team id, time of event, x and y coordinates and the outcome of the event are registrated. Also we see that some extra details are included. In this example it maps to a pass from [44.6, 61.1] to [91.6, 49.9].

<Q id="1774596260" qualifier_id="coordX" value="91.6"/>
<Q id="1429253465" qualifier_id="coordY" value="49.9"/>
<Q id="1084400575" qualifier_id="56" value="Back"/>

2.1.1.2 PROZONE/AMISCO

PROZONE3 is another system that tries to track players. Their data capture system incorporates 8-12 cameras, which is strategically positioned throughout the stadium so it covers 100 percent of the ground with some redundancy in case of a faulty component. They also incoperate the TV-camera feed, which obivously always follows the ball. All cameras are hooked into one server and uploaded at the end of the game before sent to undertake the tracking process.

The coding and tracking process of the players movement and actions is a manual process at present time. The process of knowing where every players is and doing is still to be seen done by a software completely alone without human guiding. There is a large department dedicated for post processing the game. With the help of the many cameras with different camera angles the team track players and on the ball events. Events and players gets an

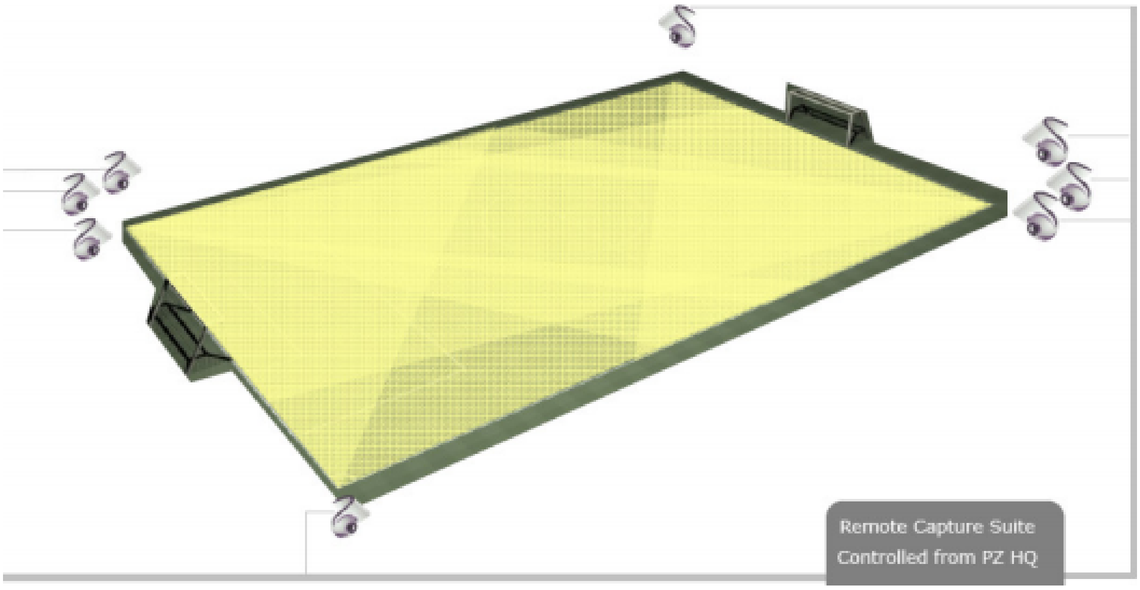

Figure 2.1: The PROZONE camera system illustrated

positional tag by using the pitch dimensions and coordinates. As different football grounds houses different pitch sizes the pitch dimensions has to be taken into account by calibrating the cameras (ASK SIMON). How do you know how far a player has run? This is calculated from the events and notating where players has run combined with the known pitch size and coordinates. The whole input process is done in a own software which helps you minimize the amount of manual work. The software follows rules created by basic machine learning algorithms to validate and verify the input. A simple example would be if the ball goes out of play the system will now that the next event will be a throw in, corner kick or goal kick [**?**].

They claim to be able to track every movements of every player on the pitch every 10th of a second without using 3 any physical equipment on the players. When it comes to accuracy Di Salvo et al. [7] conducted an empirical evaluation of deployed ProZone systems at Old Trafford in Manchester and Reebook Stadium in Bolton, and concluded that the video camera deployment gives an accurate and valid motion analysis. The data is after a match available through the PROZONE3 interface.

2.1.2 Automatic capturing

SAP places sensors in shin guards, clothing, and in footballs. The data can be applied to individual and team movement profiles to track distances, speed averages, ball possession, player tendencies and more .

A similar system is the ZXY sport trackingsystem. The system is in used by premier league soccer teams in Norway. Data captured is stored in Sybase databases with each match requiring about 500-700MB storage The players have to wear a belt around their waist to track their movement. The ZXY system is able to track the players movement very detailed with an accuracy of 0.5m. It relies on a radio-based signaling substrate to provide real-time high-precision positional tracking, including acceleration and heart rate. The home arena for Troms IL, Alfheim, is currently equipped with 10 receivers . A receiver tracks an specific area of the soccer field and combined they cover the whole pitch with some redundancy areas. To compute the positional data the stationary radio receiver uses an advance vector based processing of the received radio signal. The data is aggregated and stored into a relational database.

Doing queries on the data can give you all sprints for a certain player. Players in certain areas of the field have more intensive sprints when they first are involved thus are more vulnerable to injuries. Knowing the actual physical load on players can prevent injuries by regulating the training intensity and amount of time on the workout.

2.1.2.1 Sensor based

2.1.3 Presenting data

PROZONE comes with software to present the data. Through the opposition analysis

2D animation Single player analysis Team analysis Pressing analysis Success/direction Player tempo Passing movements Receving the ball Player events

On individual level you can get basic tatistics like shots, total passes, passing success, crosses, shots on target, balls received, tackles, fouls.

[?]

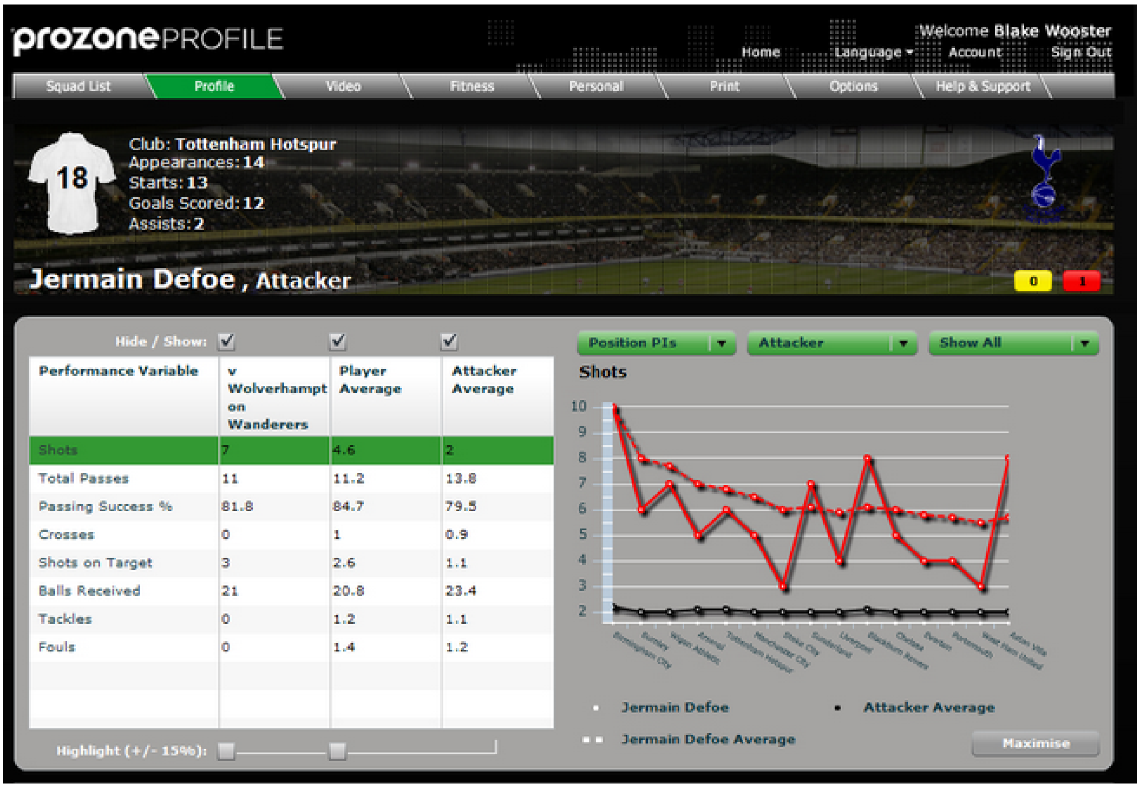

Figure 2.2: The PROZONE software - individual player analysis gives you statistics of performance over time

Chapter 3

Requirement Specification

Chapter 4

System Model

Chapter 5

Soccer Analytic toolkit

fjksdfjsafk asdjfl jadslk fjasldkfj sfiojslfjslkfj

skfljaslfjlakj

Chapter 6

Evaluation and Results

Chapter 7

Conclusion

This chapter presents our achievements, gives some concluding remarks and outlines possible future work.

7.1 Achievements

7.1.1 Concluding Remarks

7.2 Future Work

References

[1] D. E. Comer, David Gries, Michael C. Mulder, Allen Tucker, A. Joe Turner, and Paul R. Young. Computing as a discipline. *Commun. ACM*, 32(1):9–23, January 1989.

[2] Pl Haloversen Hvard D. Johansen, Svein Arne Pettersen and Dag Johansen. Combining video and player telemetry for evidence-based decisions in soccer.