Automatic Event Extraction and Video Summaries From Soccer Games

Asgeir Mortensen[1], Vamsidhar Reddy Gaddam[1], Håkon Kvale Stensland[1], Carsten Griwodz[1], Dag Johansen[2], Pål Halvorsen[1]

[1]Simula Research Laboratory, Norway & University of Oslo, Norway
{asgeirom, vamsidhg, haakonks, griff, paalh}@ifi.uio.no

[2]University of Tromsø, Norway
dag@cs.uit.no

ABSTRACT

Bagadus is a prototype of a soccer analysis application which integrates a sensor system, a video camera array and soccer analytics annotations. The current prototype is installed at Alfheim Stadium in Norway, and provides a large set of new functions compared to existing solutions. One important feature is to automatically extract video events and summaries from the games, i.e., an operation that traditionally consumes a huge amount of time. In this demo, we demonstrate how our integration of subsystems enable several types of summaries to be generated automatically, and we show that the video summaries are displayed with a response time around one second.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Search; H.5.1 [**Multimedia Information Systems**]: Video

General Terms

Experimentation, Performance

Keywords

tracking player positions, panorama video, video summaries

1. INTRODUCTION

An athlete's performance has for a long time been carefully analyzed, and in order to improve and learn how to become better, both during exercises and game scenarios, video recordings are often made. In addition, a huge amount of statistics are captured. For example, in arena sports such as soccer, there exists a large number of systems for game analysis. Some allow coaches to list events, some make video recordings, and some capture player movements automatically. In this respect, our Bagadus system [7, 12, 13] allows users to mark and follow one or more players live during a game, and play out expert annotations in real-time. Nevertheless, some events end up without a tag, and identifying these often requires reviwing larger portions of the game footage. A post analysis of a game often includes manual generation of video summaries and game highlights, and often the coaches generate individual summaries to different players.

Making this type of video summaries obviously consimes a lot of time. Video summarization is therefore a challenge computer scientists, and video retrieval researchers in particular, have been trying to automatize and solve for many years. There exist many approaches presented in the literature [5, 4, 10], and in our context of the soccer stadium, techniques based on motion, color, clustering, event or objects should be useful [2]. However, the accuracy with respect to precision and recall have been too low, the feature extraction from the videos are too slow, and the events that may be searched for are too limited. In this respect, we have earlier used external sources for metadata found in live-text-commentaries web-pages written by journalists [9, 8], and later, similar approaches have been proposed [3]. Such an approach is good for extracting semantic events, but there are still types of events that is hard to find, e.g., analyzing players positions (both to the field and to other players) and players movements (like speed and directions).

In Bagadus, we have therefore in the context of stadium sports proposed to integrate and synchronize position and movements sensors with video data for certain types of event extractions. Based on real datasets from Norwegian Elite division games, we have successfully added functionally to the system to make database queries and extract video frames based on a synchronized clock. Our proposed demo will show how our integration of subsystems enable several types of summaries to be generated automatically, and we show that the video summaries are streamed and displayed on the user device with a response time in the second range.

MMSys'14, March 19-21 2014, Singapore, Singapore
ACM 978-1-4503-2705-3/14/03.
http://dx.doi.org/10.1145/2557642.2579374

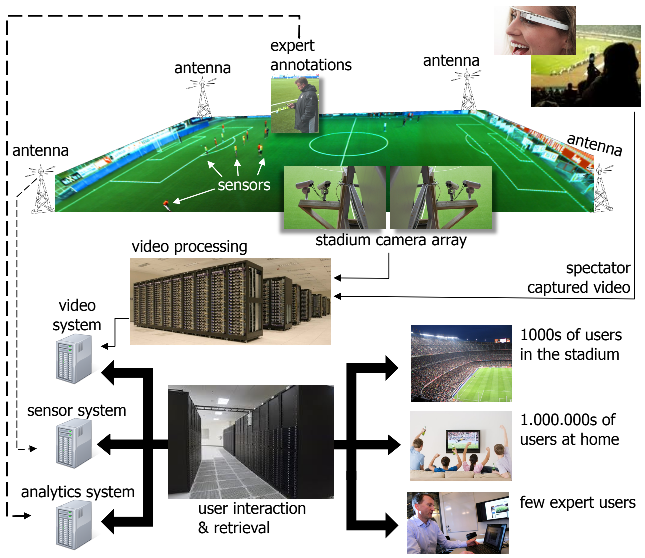

Figure 1: System setup at Alfheim stadium (see [7]).

2. SOCCER SCENARIO

In the area of sports, analytic athlete performance applications have recently had an huge increase in interest, and "statistics keep changing the way sports are played — and changing minds in the industry" [6]. In this respect, we have earlier developed and described *Bagadus* [7]. This system aims to fully integrate various sub-systems and enable an automatized real-time presentation of sport events. A running prototype is deployed at Alfheim Stadium, i.e., the home arena for Tromsø IL soccer club.

A brief overview of the architecture and interaction of the different components is given in figure 1. Here, a *video* subsystem [**?**] consisting of a video camera array cover the full field with sufficient overlap to identify common features necessary for camera calibration and image stitching. A *tracking* subsystem [1] uses sensor to provide player position information and movement statistics. Finally, an *analytics* subsystem [**?**] enables coaches to register predefined events quickly with the press of a button on modern devices rather than using pen and paper. In summary, Bagadus provides an integrated system where operations are automatically managed and all data, i.e., video frames, database position records and expert annotations, are time-synchronized. Additionally, Bagadus has a large potential to provide users at home with personalized content and interactive services, but this is however out of the scope of this demo and is subject to ongoing work.

3. VIDEO SUMMARIES BY COMBINING SENSOR AND VIDEO DATA

There exists several systems that provide game statistics, player movements, video highlights etc. However, to a large degree, the existing systems are offline systems, and they require a large portion of manual work to integrate information from various computer systems and expert sport analytics. In the context of the proposed demo, the existing systems miss is the possibility to automatically extract video events and summaries.

In the context of Bagadus, the combination of video with sensor data can be used to make such video highlights automatically. Our sensor system from ZXY sport tracking [1] delivers information for each player such as field coordinates, heading and direction of the player, heart rate, estimated energy used since last sample, current speed and total distance moved. Thus, various summaries can be retrieved based on different combinations of these. For example, composing a video for "all the events where defender X is in the other team's 18-yard box in the second half" is, according to discussions with many Norwegian coaches, a manual time-consuming operation of at least the duration of "the second half" (45 minutes). On the other hand, by using Bagadus, such a summary can start within milliseconds.

Figure 2 gives a brief overview of the operation. The chief analytic expert would like to make a summary of all situations for the situation defined above (opponents 18-yard box). Using an interface (under construction[1]) that translates the situation to a standard SQL database query, the system retrieves all database records that match the conditions. As described in section 2, all data in Bagadus is time synchronized, and the time information retrieved from the database is used to retrieve the corresponding video frames from the video sub-system. The video frames are then combined into a video summary, encoded and sent to the client display device.

4. SOME EXAMPLE QUERIES

In this section, we give some examples of queries that can be executed and the response time for presenting the corresponding video summaries to the client device. We have used a publicly availalbe dataset [11] to perform the experiments, and the system is currently using a PostgreSQL database. The below examples are tested and executed on a laptop, i.e., a Zenbook with a Core i7 1,9 GHz CPU running the entire system in order to isolate the system response time. For remote users, network latencies will be added. Each measurement have been repeated 20 times, and in table 1, we present the average, maximum and minimum response times from the command is issued until the first frame has been shown on the monitor.

	Average	Maximum	Minimum
18-yard box	0.79	1.00	0.74
10 mps sprints	0.78	0.82	0.75
3-meter distance	1.05	1.10	0.99

Table 1: Video summary response times (seconds).

[1]At the time of the writing, we do not have any logical interface, i.e., the users must write traditional SQL queried. Our target users will probably not have such knowledge, and more intuitive, user-friendly interfaces are currently being researched.

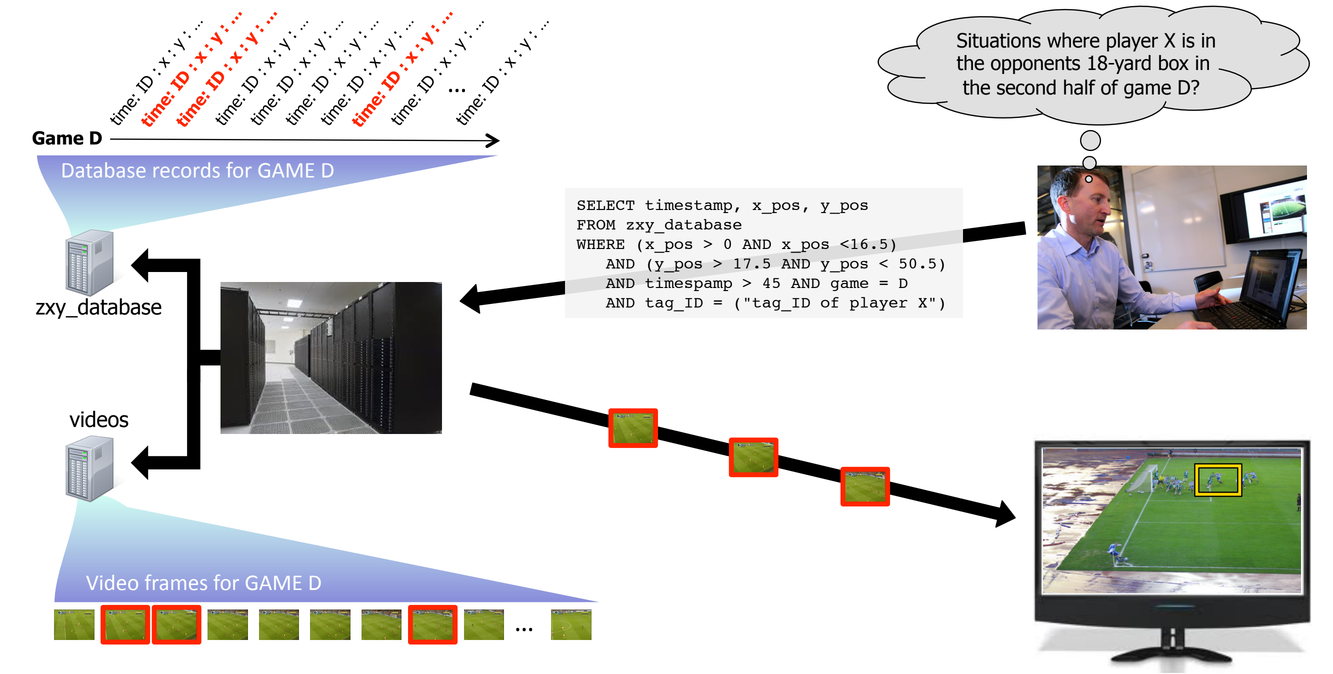

Figure 2: User interaction in Bagadus. The user sends a query, and the resulting database records are used to find corresponding video frames. These are again encoded into a video summary and presented to the user.

4.1 In opponents 18-yard box

The example from figure 2 finding all situations in game D where player "X" is in the opponents 18-yard box in the first half can be expressed in pseudo-code as

```
SELECT timestamp, x_pos, y_pos
FROM zxy_database
WHERE (x_pos > 0 AND x_pos <16.5)
  AND (y_pos > 17.5 AND y_pos < 50.5) AND timespamp < 45
  AND game = D AND tag_ID = X
ORDER BY timestamp
```

The x- and y-positions (fields coordinates) are also returned in order to mark the player as in figure 2 for example with a box. In the used data set, 9 players were in the opponents 18-yard box in the first half, and the respective times are marked in figure 3. Editing the video manually to generate a summary takes a lot of time, but using our system, the average startup latency over several runs for this video summary was 789 milliseconds (see table 1) using `tag_id = 10` as an example.

4.2 All sprints faster than 10 mps

A simple example of a query could be to retrieve all situations where a player (X) runs faster than 10 meters per second. It is just a small check of the player tag and that the speed is above 10. A simple pseudo-code could look like

```
SELECT timestamp
FROM zxy_database
WHERE speed > 10 AND tag_ID = X
ORDER BY timestamp
```

Note that this simple example is similar to queries like "heart rate above 150 heart-beats per minute" and "running in the direction of the penalty spot". The returned database

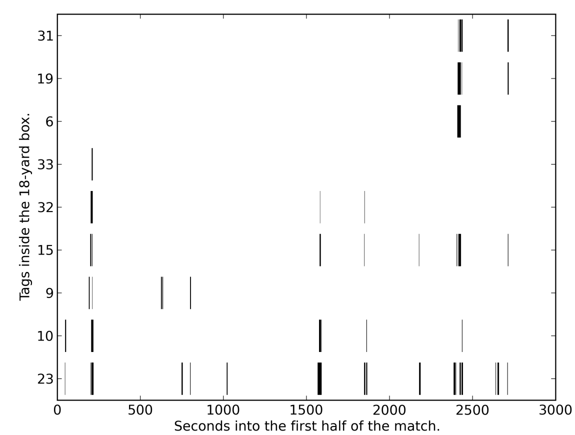

Figure 3: All tags from the sample dataset highlighted when inside the opponent 18-yard box.

records from above query are ordered according to time, and as shown in table 1, the system starts the video playback on average in 784 milliseconds.

4.3 3-meters distance from another player

Another example of a video summary could be to find the situations where a striker (X) is closer than 3 meters to the defender (Z) he should not be close to in order to be available for passes. In pseudo-code, this can be expressed as

```
WITH striker AS (
    SELECT timestamp, point(x_pos, y_pos) AS pos
    FROM zxy_database
    WHERE tag_id = X
)
WITH defender AS (
    SELECT timestamp, point(x_pos, y_pos) AS pos
    FROM zxy_database
    WHERE tag_id = Y
)
SELECT striker."timestamp"
FROM striker, defender
WHERE striker.timestamp = defender.timestamp
   AND striker.pos IN circle(defender.pos, 3)
ORDER BY striker.timestamp
```

Here, we check if the striker at a given time is inside a circle with a radius of 3 meters around the defender. On our prototype system, the video summary starts playing after 1051 milliseconds on average.

4.4 Limits and opportunities

As can be seen from the examples above and the resulting startup latencies, our system provides unique opportunities to automatically generate video summaries in arena sport scenarios. It is obvious that there are limitations in types of summaries that can be generated, but the system also enable summaries that are impossible to automatically detect using video analysis, e.g., heart-rate or energy spent. The queries may also be more complex by for example combining the queries above (e.g., all sprints above 10 mps in the second half in the opponents 18-yard box where the player has a heart-rate above 150), or looking at how players are positioned relative to each other, e.g., if the defenders are on a straight line.

5. DEMO

In this demo, we present the automatic video extraction capabilities of Bagadus. We show how we have integrated the camera array and the sensor data. We demonstrate the prototype system using an example data set recorded at Alfheim Stadium (Tromsø, Norway), and the conference participants will be able to experiment with the system using both predefined database queries and write their own queries. The resulting video should start within a second or two.

Acknowledgments

This work has been performed in the context of the *iAD* center for Research-based Innovation (project number 174867) funded by the Norwegian Research Council.

6. REFERENCES

[1] ZXY Sport Tracking. `http://www.zxy.no/`, 2013.
[2] M. Ajmal, M. Ashraf, M. Shakir, Y. Abbas, and F. Shah. Video summarization: Techniques and classification. In *Proc. of ICCVG*, volume 7594 of *Lecture Notes in Computer Science*, pages 1–13, 2012.
[3] C.-M. Chen and L.-H. Chen. A novel approach for semantic event extraction from sports webcast text. *Multimedia Tools and Applications*, pages 1–16, Dec. 2012.
[4] U. Damnjanovic, V. Fernandez, E. Izquierdo, and J. Martinez. Event detection and clustering for surveillance video summarization. In *Proc. of WIAMIS*, pages 63–66, 2008.
[5] A. Divakaran, K. A. Peker, and H. Sun. Video summarization using motion descriptors. In *Proc. of SPIE Conf. on Storage and Retrieval from Multimedia Databases*, 2001.
[6] P. Dizikes. Sports analytics: a real game-changer, 2013. http://web.mit.edu/newsoffice/2013/sloan-sports-analytics-conference-2013-0304.html.
[7] P. Halvorsen, S. Sægrov, A. Mortensen, D. K. Kristensen, A. Eichhorn, M. Stenhaug, S. Dahl, H. K. Stensland, V. R. Gaddam, C. Griwodz, and D. Johansen. Bagadus: An integrated system for arena sports analytics – a soccer case study. In *Proc. of ACM MMSys*, pages 48–59, Mar. 2013.
[8] D. Johansen, P. Halvorsen, H. Johansen, H. Riiser, C. Gurrin, B. Olstad, C. Griwodz, Å. Kvalnes, J. Hurley, and T. Kupka. Search-based composition, streaming and playback of video archive content. *Multimedia Tools and Applications*, 61(2):419–445, 2012.
[9] D. Johansen, H. Johansen, T. Aarflot, J. Hurley, Å. Kvalnes, C. Gurrin, S. Sav, B. Olstad, E. Aaberg, T. Endestad, H. Riiser, C. Griwodz, and P. Halvorsen. DAVVI: A prototype for the next generation multimedia entertainment platform. In *Proc. of ACM MM*, pages 989–990, Oct. 2009.
[10] B. Li and M. Sezan. Event detection and summarization in sports video. In *Proc. of CBAIVL*, pages 132–138, 2001.
[11] S. A. Pettersen, D. Johansen, H. Johansen, V. Berg-Johansen, V. R. Gaddam, A. Mortensen, R. Langseth, C. Griwodz, H. K. Stensland, and P. Halvorsen. Soccer video and player position dataset. In *Proc. of ACM MMSYS*, Mar. 2014.
[12] S. Sægrov, A. Eichhorn, J. Emerslund, H. K. Stensland, C. Griwodz, D. Johansen, and P. Halvorsen. Bagadus: An integrated system for soccer analysis (demo). In *Proc. of ICDSC*, Oct. 2012.
[13] H. K. Stensland, V. R. Gaddam, M. Tennøe, E. Helgedagsrud, M. Næss, H. K. Alstad, A. Mortensen, R. Langseth, S. Ljødal, O. Landsverk, C. Griwodz, P. Halvorsen, M. Stenhaug, and D. Johansen. Bagadus: An integrated real-time system for soccer analytics. *ACM Trans. Multimedia Comput. Commun. Appl. (TOMCCAP)*, 10(1s):14:1–14:21, Jan. 2014.