Soccer Video and Player Position Dataset

Svein Arne Pettersen[1], Dag Johansen[1], Håvard Johansen[1], Vegard Berg-Johansen[4]
Vamsidhar Reddy Gaddam[2,3], Asgeir Mortensen[2,3], Ragnar Langseth[2,3], Carsten Griwodz[2,3],
Håkon Kvale Stensland[2,3], Pål Halvorsen[2,3]

[1]University of Tromsø, Norway [2]Simula Research Laboratory, Norway
[3]University of Oslo, Norway [4]Tromsø Idrettslag, Norway

{svein.arne.pettersen, dag, haavardj}@uit.no vegard.berg-johansen@til.no
{vamsidhg, asgeirom, ragnarla, griff, haakonks, paalh}@ifi.uio.no

ABSTRACT

This paper presents a dataset of body-sensor traces and corresponding videos from several professional soccer games captured in late 2013 at the Alfheim Stadium in Tromsø, Norway. Player data, including field position, heading, and speed are sampled at 20 Hz using the highly accurate ZXY Sport Tracking system. Additional per-player statistics, like total distance covered and distance covered in different speed classes, are also included with a 1 Hz sampling rate. The provided videos are in high-definition and captured using two stationary camera arrays positioned at an elevated position above the tribune area close to the center of the field. The camera array is configured to cover the entire soccer field, and each camera can be used individually or as a stitched panorama video. This combination of body-sensor data and videos enables computer-vision algorithms for feature extraction, object tracking, background subtraction, and similar, to be tested against the ground truth contained in the sensor traces.

Categories and Subject Descriptors

D.2.4 [**Software Engineering**]: Software/Program Verification—*Validation*; H.3.7 [**Information Storage and Retrieval**]: Digital Libraries—*Collection*; I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding—*Video analysis*

General Terms

Experimentation, Measurement, Performance

Keywords

Soccer, body-sensors, position tracking, panorama video

1. INTRODUCTION

Video capturing and analytic systems are becoming essential tools for professional sports-clubs all over the world, reshaping how soccer is being played [3] and how athletes are being developed. Significant effort has been put into building algorithms and computer systems for tracking objects in videos, including in sports. Still, object tracking remains a hard challenge, and the accuracy of different approaches vary with the number of objects and the complexity of the scene. Vision algorithms employ combinations of various approaches like feature extraction, feature tracking, and background subtraction for detecting and tracking objects. The robustness of the algorithms are often very sensitive to the lightning and environmental conditions in the source material. In outdoor scenarios like on a soccer field, these parameters are difficult to control.

To have comparable results and repeatable experiments on the same data, datasets are often made available to be used to test proposed algorithms or mechanisms. A multitude of datasets for various purposes in multimedia research do already exists, like the TrecVid collection where both videos and ground truths are provided [9]. However, few dataset exist for multimedia and computer vision research in the soccer domain. The workshops on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS) have for quite a few years provided quality datasets with outdoor videos and ground truth for object placement, including one for soccer scenarios [2]. The provided dataset is quite primitive where videos are captured using three cameras placed at different corners of the stadium covering only parts of the field. Also included in the VS-PETS dataset is training data for use in learning background models, and data to test the tracking algorithms. Unfortunately the data lack many features we find necessary. First, the annotations of players are provided only for one camera. Second, the placement of the cameras does not relate to the usual broadcasting videos, where the videos are usually recorded from the center of one side of the field. Moreover, D'Orazio et al. [4] provide a dataset for soccer with video recorded using three HD cameras placed on both sides of the field, but the annotations for players are performed manually and only available for two minutes of the game.

Since we have not found any available dataset for a large outdoor soccer stadium, we here provide a set of videos and correlated body sensor data from the home-team players captured at Alfheim Stadium in Tromsø, Norway, during three different soccer games [1] in 2013. The body-sensor data is provided using the radio-based ZXY Sport Tracking (ZXY) system [7], which report the players' positions at 20 Hz and with an approximate error of one meter. The

MMSys '14, March 19–21 2014, Singapore, Singapore.

corresponding videos are recorded from a set of cameras positioned in the middle of one of the long-sides of the soccer field, and a high-quality wide field-of-view video of the entire stadium is also provided.

We have earlier used these data in our previous work on the Muithu [6] and Bagadus [5, 10] systems. For example, Bagadus uses the player positions to mark and follow one or more players live during a game, and also as an input to improve both accuracy and efficiency of the the background subtraction algorithm. The background subtraction is then again used as input for selecting a dynamic seam when generating the high quality panorama image of the soccer pitch. We hope that this data, which we have generated as part of the research can be useful to others and advance research in the area of multimedia and computer vision.

2. MEASUREMENTS AND LOGS

The dataset consists of videos and correlated body-sensor data for the home team from three different games played and recorded at Alfheim stadium located in Tromsø, Norway, during November 2013. Although the sensor system in use can capture data from both teams, the away teams were not equipped with the required sensor belts during these games. We have also obtained consent from the involved players in the home team for including their body-sensor data. The data has been made available online [1] containing the games where Tromsø IL met Strømsgodset IF (Norway), Anzhi Makhachkala (Russia) and Tottenham Hotspurs. All matches were played outdoors in the evening using floodlighting of 1400 lx in all directions. We have also included manually tagged ball position for the Tottenham game.

The players' identity in the body-sensor data is randomized to protect the privacy of the players. Although the dataset can be used for general object and people tracking, attempts to re-identify individual players, create player/club performance profiles, and similar, are not allowed. The data may only be used for *non-commercial research purposes.*

2.1 The ZXY Sports Tracking Logs

The body-sensor data is recorded using the ZXY system which consists of sensor belts, worn by the athletes around his lower torso, and 11 stationary radio receivers mounted on poles or on the tribune roof around the stadium. Each radio receiver has an approximately 90 degrees field-of-view giving overlapping zones of the soccer field to provide high immunity to occlusions and signal blocking, which is necessary for reliable operation.

The current generation of the ZXY system is based on the 2.45 GHz ISM band for radio communication and signal transmissions. Each stationary radio receiver computes the position data for each player using vector-based processing of the received radio signals from the active belts. This enables direct projection of each player's position on the field without the use of conventional triangulation methods.

The 10 g body sensor includes an accelerometer that registers body movements in all 3-directional axes, a gyro, a heart-rate sensor and a compass. The accelerometer provides valuable data in addition to the more common data of distance covered in different speed categories [8]. The compass in combination with the gyro allows us to track the actual heading of the player. Due to privacy concerns and instability in the heart-rate sensor, pulse data is not included in the dataset.

The default positional sampling rate per belt is currently set to 20 Hz, transmitting in real-time to a central relational database for storage. By including all body sensor information in the same radio signal used for computing the positions, the system enables time synchronization of all raw data when stored in the database. The producer of ZXY claims accuracy in the order of 1 m for our version of the system, which conform to our own experiences and is significantly more accurate than more traditional GPS-based tracking solutions [7].

2.1.1 *Coordinate System and Layout*

The ZXY system provides a two-dimensional positional coordinate system calibrated to the station. At Alfheim, the positive x-axis points southwards parallel along the long side of the field, while the positive y-axis points eastwards parallel with the short edge of the field, as shown in Figure 1. The position $(0, 0)$ is located in the north-western corner of the soccer field, which from the position of the cameras is the lower-left corner. The soccer pitch is 105×68 m wide and hence valid in-field values for x and y are in the range of $0 \leq x \leq 105$ and $0 \leq y \leq 68$.

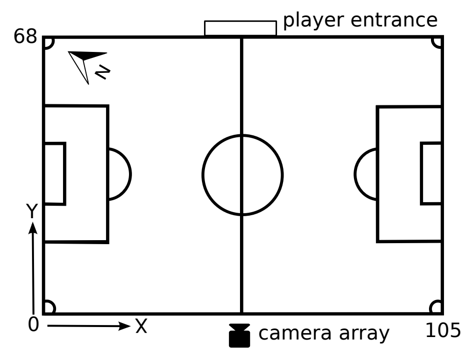


Figure 1: Layout at the Alfheim Stadium

Out-of-field values occur in the dataset whenever a player ventures outside the field, for instance during throw-ins or warm-up sessions. Headings and directional values are defined so that 0° is in the direction of the positive y-axis. The magnetic north then lies at approximately 300°.

2.1.2 *Data Format*

The logs extracted from the ZXY database are stored in plain Comma-Separated Values (CSV) ASCII text format. We provide both a raw and an interpolated dataset. The raw dataset contains the data as received from the sensors, while the processed dataset has been interpolated to have matching timestamps, simplifying queries as seen in Figure 8. Each line contains one time-stamped record from exactly one belt. Belts are uniquely identified by a tag-id, and each player wears only one belt. The tag-ids have been randomized for each game to provide a higher-level of anonymity to the players. A sample of the logs can be seen in Figure 2, where the data records has the following format:

timestamp (string) – Local Central European Time (CET) time encoded as ISO-8601 format string.

tag_id (int) – The sensor identifier.

x_pos (float) – Relative position in meters of the player in the field's x-direction.

```
'timestamp','tag_id','x_pos','y_pos','heading','direction','energy','speed','total_distance'
    ...
'2013-11-03 18:30:00.000612',31278,34.2361,49.366,2.2578,1.94857,3672.22,1.60798,3719.61
'2013-11-03 18:30:00.004524',31890,45.386,49.8209,0.980335,1.26641,5614.29,2.80983,4190.53
'2013-11-03 18:30:00.013407',0918,74.5904,71.048,-0.961152,0,2.37406,0,0.285215
'2013-11-03 18:30:00.015759',109,60.2843,57.3384,2.19912,1.22228,4584.61,8.14452,4565.93
'2013-11-03 18:30:00.023466',909,45.0113,54.7307,2.23514,2.27993,4170.35,1.76589,4070.6
    ...
```

Figure 2: Samples from the 20 Hz ZXY sensor traces

y_pos (float) – Relative positions in meters of the player in the field's y-direction.

heading (float) – Direction the player is facing in radians where 0 is the direction of the y-axis.

direction (float) – Direction the player is traveling in radians where 0 is the direction of the y-axis.

energy (float) – Estimated energy consumption since last sample. The value is based on step frequency as measured by the on-board accelerometer. The unit for this value is undefined and might be relative to each individual.

speed (float) – Player speed in meters per second.

total_distance (float) – The number of meters traveled so far during the game.

For each game, we also provide a set of the following 1 Hz values derived from the raw data:

timestamp (string) – Date formatted as ISO-8601 followed by time of day.

tag_id (int) – Sensor identifier, identifies a player on the home team.

total_distance (float) – Cumulative distance traveled so far during the match in meters.

hir_distance (float) – Total distance covered in speed category high-intensity run (i.e., speeds $\geq 19.8\,\mathrm{km\,h^{-1}}$) [8].

sprint_distance (float) – Total distance covered in speed category sprint (i.e., speed $\geq 25.2\,\mathrm{km\,h^{-1}}$) [8].

total_effort (float) – Estimated cumulative energy consumption based on the energy field in the main sample file. The unit is also here undefined.

2.2 Video Data

We provide two different sets of video data from two different cameras types (see Table 1). The first set is a collection of shutter synchronized videos that are captured using three wide-angle cameras (acA1300-30gc). The second set is a stitched high-quality panoramic video using five cameras (acA2000-50gc). Both video types cover the entire soccer field.

2.2.1 Wide-Angle Cameras

To capture different parts of the field (Figure 3), we have used three Basler acA1300-30gc cameras recording simultaneously from the center of one of the sides using 3.5 mm Kowa-LM4NCL lenses. The provided videos are processed to correct for the distortions created by the wide-angle lens. There is also sufficient overlap between adjacent cameras for applications like panorama stitching.

Camera	Basler acA1300-30gc	Basler acA2000-50gc
Resolution	1280×960	1920×1080
Frame rate	30 fps	25 fps
Lens model	3.5 mm *Kowa-LM4NCL*	8 mm *Azure-0814M5M*
Use	single wide-angle videos (1280×960)	stitched panoramic video (4450×2000)

Table 1: Properties of the cameras

Figure 3: The views from the wide-angle cameras

2.2.2 High-Quality Panorama

The provided set of high-quality cylindrical panorama video is stitched together from the images of five Basler acA2000-50gc cameras with 8 mm *Azure-0814M5M* lenses. The cameras are placed inside the control room at Alfheim so that the video includes audience, which can, for instance, be used for analyzing crowd behavior. An example frame from the panorama video for the two first games is shown in Figure 4. For the Tottenham game, the camera array is moved closer to the field (see [1]).

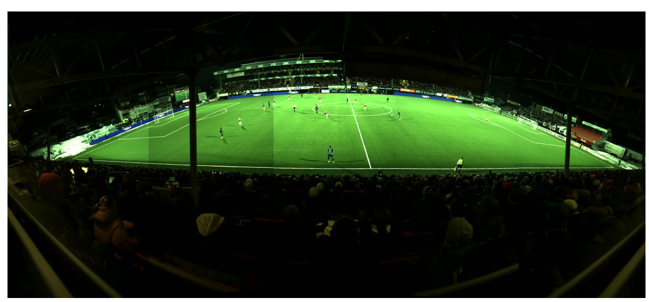


Figure 4: High quality panorama video

2.2.3 Video Format

Even though the videos are captured using different camera models, they are stored using the same codec with same properties, except for resolution and Frames Per Second (FPS). All the videos are H.264 encoded using the libx264 software. The encoding parameters are listed in Table 2. All other parameters are selected by the profile. Each camera stream is split into files of three-second segments, containing exactly $3 \times FPS$ frames. The file names encodes a four-digit sequence number followed by a timestamp of the first frame in the file, represented with nanosecond accuracy. This al-

profile	high
preset	ultrafast
tune	zerolatency
Colorspace	YUV 420 planar
GOP	3 x fps
Files	xxxx_YYYY-MM-DD hh:mm:ss.000000000.h264

Table 2: Parameters used for H264 encoding

lows for easy synchronization with the ZXY data as the time stamp of the sensor data can be matched to the video segment file names.

2.3 Sensor Coordinates to Video Pixels

The players' positions on the field obtained from the ZXY system are represented using the Cartesian coordinate system with the origin in the lower left corner. The cameras are mounted at the middle of the field under the roof and inside the control room. In order to locate and track a person in the video using the sensor positions, we need a transformation from the sensor coordinates to the image pixels for all valid pixel coordinates in a video frame ($ZXY(x,y) \rightarrow pixel(u,v)$). In this respect, we have calculated a 3×3 transformation matrix using fixed, known points on the field as shown in Figure 5. Sample matrices are given in [1].

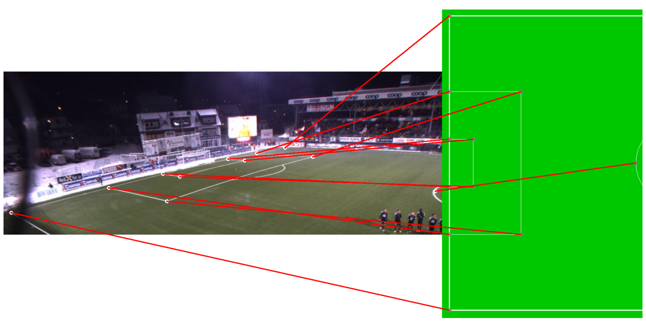


Figure 5: Pixel mapping between the video images and the ZXY tracking system.

2.4 Sources of Error

Although the ZXY sensor system in use has a significantly higher precision compared to traditional GPS based ones [7], we still observe situations where the pixels are not perfectly mapped to the sensor data. Figure 6 gives an example of such a situation where the player is not perfectly in the center of the tracking-box. However, for the purpose here, we believe that the dataset is accurate enough to provide a ground truth to be used in various video object tracking scenarios.

Inaccuracies in our current mapping between the video and sensor data is also a source for such errors (see Figure 6). New and better mapping functions might thus be investigated to improve the accuracy of the data.

Sometimes the ZXY sensor belts also fail or are removed during a game, and sometimes belts not in use are accidentally remotely activated by the control system and transmit data. In these cases, intermittently or permanent abnormalities do occur in the data. We have not removed these abnormalities from the dataset as they should be expected and handled by any practical algorithm using such data.

Figure 6: Small mismatch of sensor positions and video pixels.

3. EXAMPLES OF USE

The provided dataset has many use cases. This section provides some use cases and examples that illustrated the type of information that can be extracted from the data.

3.1 Computer Vision Algorithms

A good amount of research in the computer vision field is focused on tracking and detection of objects in complicated scenarios. Some of the sub-problems involved with scene understanding are background subtraction, feature extraction, feature tracking, etc. An outdoor stadium like Alfheim provides enough challenges with varying weather and uncontrolled light conditions. Performance evaluations of the algorithms can be tricky in such cases without proper ground truth data. Moreover, the existence of tracking data can also be used for further improvements of the algorithms.

3.1.1 Background Subtractions

One use case we explored where to perform real-time background subtraction to extract the pixels belonging to the players. We used the player position data to speed up the algorithm and also improve the precision and robustness to noise. Some of the results can be seen in Figure 7 where we have used the Zivkovic background subtraction algorithm [11].

(a) Without sensor positions (b) Using sensor positions

Figure 7: Result from a background subtraction algorithm with improved performance in both speed and quality with the help of tracking data.

3.1.2 Tracking Players

Tracking humans, especially when several persons interact and have overlapping movements, is challenging. In this respect, we have used the ZXY data directly in combination with video to follow individual players (see for example Figure 6) and parts of the team directly [5, 10, 7]. If such tracking is performed by a video analysis tool, the sensor data can be used as a ground truth to verify the accuracy of the proposed algorithm.

3.1.3 Automatically Generation of Video Summaries

The combination of video with sensor data can also be used to make automatic video summaries or extract complex events. For example, when we used the data in Bagadus [5], we described a query like "all the events where defender X is in the other team's 18-yard box in the second half", and the video summary started playing in a few milliseconds. In the context of analyzing the videos for making video summaries, this dataset can again be used to compare the accuracy of the algorithms by using the time intervals returned from queries to the dataset. For example, summaries over "all sprints faster than $Y\ m/s$", "all players running backwards in the center circle" or "all runs where a player has a negative acceleration larger than $Z\ m/s^2$". Figure 8 shows an example query (assuming the dataset is inserted into a PostgreSQL database) where we find "all situations where player with tag 15 is closer than 4.5 meters to any of the other players", and the returned time intervals used to generate video summaries are displayed per player in Figure 9.

```
WITH match_data AS (
  SELECT "timestamp", "tag_id", point("x_pos", "y_pos") AS pos
    FROM raw_data)
SELECT md."timestamp", md."tag_id", array_agg(fnd."tag_id")
  FROM "match_data" AS md
  INNER JOIN "match_data" AS fnd
    ON (md."timestamp" = fnd."timestamp"
        AND md."tag_id" != fnd."tag_id"
        AND (circle(md."pos", 4.5) @> fnd."pos"))
    WHERE md."tag_id" = 15
  GROUP BY md."timestamp", md."tag_id"
  ORDER BY md."timestamp", md."tag_id";
```

Figure 8: Simple PostgreSQL query demonstrating how to find the intervals and tags when any other tag is closer than 4.5 meters.

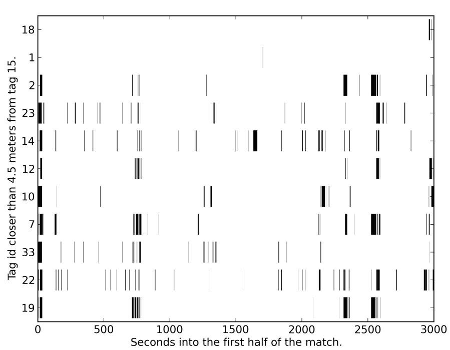


Figure 9: Time intervals returned to the video summary system for the query in Figure 8.

3.2 Sports Analytics

There are several other ways to use the sensor data, both alone and combined with video. For example, one can display how a player has moved and use these data to extract events from the video, or just use the position data for mobility simulations.

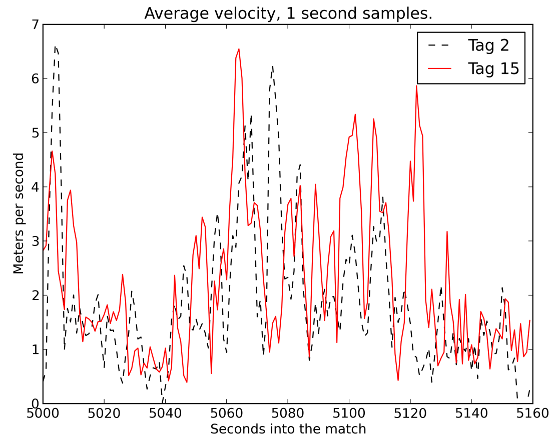


Figure 11: A player's velocity over time.

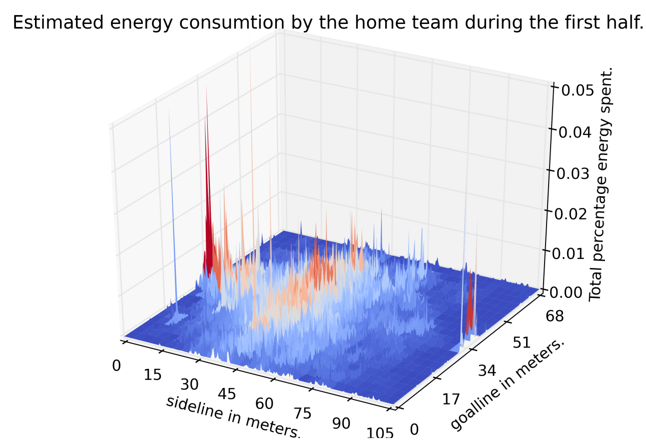


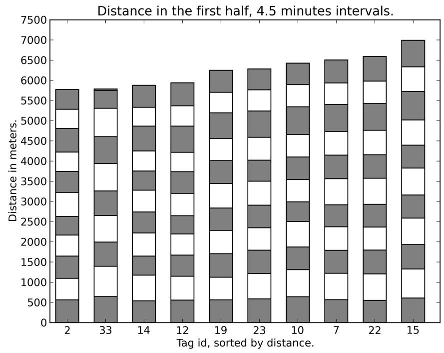
Figure 12: Estimated energy usage for the entire team in the first half.

For example, if the total distance moved is to be analyzed in the video, the ground truth may be extracted (e.g., as shown Figure 10(a)). Furthermore, the same data can be used to plot the complete movement pattern during the game (e.g., as depicted in Figure 10). Similarly, analyzing the speed is straight forward where an example of the ground truth is given in Figure 11. Finally, there are several ways to extract information over a group of people (the team). One example is given in Figure 12 displaying where on the field the the players use the energy, i.e., how much of the total energy is spent on a particular location.

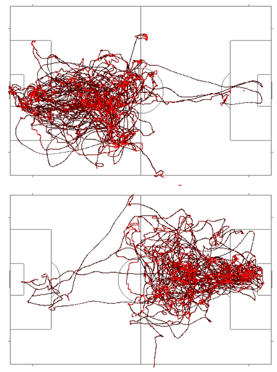
Thus, there are numerous variations of the sensor data that can be used to extract events (and ground truths) in order to extract information from the corresponding videos provided.

3.3 Other Use Cases

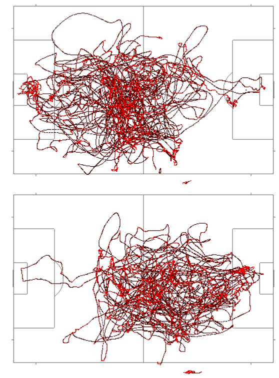
As mentioned earlier, the tracking data helps not only for evaluation of algorithms but also for improving them. But the application is not restricted to computer vision field. Other interesting application scenarios can be the football analytics itself. A wide variety of manual or semi-manual work performed to annotate videos can be fully automated. Thus, enabling the coaches and the players to use the system during the game. Researchers interested in developing



(a) Total distance moved during the first half of the match, again divided into 10 smaller intervals of 4.5 minutes.



(b) Defender.



(c) Midfielder.

Figure 10: Player movements. In figures (b) and (c), the upper image shows first half, the lower image shows second half, and the higher intensity of the red color means more time spent in that area.

analytic tools can use the dataset to develop and test them.

4. CONCLUSION

We present a multi-sensor data set encompassing the tracking data, the video data and a simple mapping between the two. The data is collected for the home team from several official soccer league matches. The body-sensor data includes timestamps, player positions, speed, heading, and several other data types. The video data includes two sets, i.e., one is a collection of wide field of view videos recorded using three cameras, and the second is a high quality panorama video stitched from five cameras. The dataset can be used in several ways. We provided two detailed use-cases and mentioned some others. We hope that the dataset can enable researchers to develop systems and algorithms for the soccer scenario without being hindered by the lack of equipment or the engineering and legal problems stemming from the collection of the data itself.

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. The authors also acknowledge the support given by Kai-Even Nilssen, who has been helpful with the practical installations; Susann Dahl Pettersen, who operates the ZXY system during the games; and Tromsø IL and their athletes for giving us their concent to participate in the dataset.

5. REFERENCES

[1] Soccer Video and Player Position Dataset. `http://home.ifi.uio.no/paalh/dataset/alfheim/`, 2013.

[2] J. Crowley, J. Ferryman, S. Maybank, and T. Tan. Outdoor people tracking: Football data (two views). VS-PETS'03 dataset, University of Reading, `http://www.cvg.rdg.ac.uk/slides/pets.html`, Oct. 2003.

[3] P. Dizikes. Sports analytics: a real game-changer. MIT News Mar. 4, Massachusetts Institute of Technology, Mar. 2013.

[4] T. D'Orazio, M. Leo, N. Mosca, P. Spagnolo, and P. Mazzeo. A semi-automatic system for ground truth generation of soccer video sequences. In *Proc of AVSS*, pages 559–564, 2009.

[5] 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.

[6] D. Johansen, M. Stenhaug, R. B. A. Hansen, A. Christensen, and P.-M. Høgmo. Muithu: Smaller footprint, potentially larger imprint. In *Proc. of IEEE ICDIM*, pages 205–214, Aug. 2012.

[7] H. D. Johansen, S. A. Pettersen, P. Halvorsen, and D. Johansen. Combining video and player telemetry for evidence-based decisions in soccer. In *Proc. of icSPORTS*, 2013.

[8] M. D. Mascio and P. Bradley. Evaluation of the most intense high-intensity running periods in English FA Premier League soccer matches. *Journal of Strength & Conditioning Research*, 27(4):909–915, April 2013.

[9] P. Over, G. Awad, M. Michel, J. Fiscus, G. Sanders, W. Kraaij, A. F. Smeaton, and G. Quéenot. Trecvid 2013 – an overview of the goals, tasks, data, evaluation mechanisms and metrics. In *Proceedings of TRECVID 2013*. NIST, USA, 2013.

[10] 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.

[11] Z. Zivkovic and F. van der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7):773–780, 2006.