A Method for Identification of Moving Objects by Integrative Use of a Camera and Accelerometers

Naoka Maruhashi[1], Tsutomu Terada[2], Masahiko Tsukamoto[1]
[1]Kobe University
Kobe, Japan
[2]Kobe University
PRESTO, JST
Kobe, Japan
maruhasi@stu.kobeu.ac.jp, tsutomu@eedept.kobeu.ac.jp, tuka@kobeu.ac.jp

ABSTRACT: *In sports events shown on TV or on the Internet, having the names and uniform numbers of the players displayed beside the players on the screen helps viewers to identify the players. Conventional methods of annotating and tracking objects utilize image recognition techniques with highresolution images, so players captured on camera from a long distance are too small for accurate recognition. In response to this problem, we propose a method for identifying and tracking moving objects with a combination of wearable acceleration sensors and image recognition techniques. In our method, the method recognizes moving objects by matching the context picked up by the wearable sensors with the context inferred during the image processing. Evaluation results demonstrated the effectiveness of the proposed method.*

1. Introduction

Many kinds of sports events are broadcast on TV and on the Internet. In games like soccer or basketball, where many players are displayed on the screen at the same time, it is often difficult for viewers to distinguish players and to recognize who is playing in a certain area. If the players' names, numbers, and related information are displayed beside them on the screen, it helps the viewers to identify the players and to derive greater enjoyment in watching the game. Moreover, if coaches had access to real time player tracking, they could acquire relevant information about the game, such as the total running distance of each player, and use this information to plan effective strategies.

Conventional methods of annotating and tracking moving objects utilize image recognition techniques with highresolution images. If the system uses camera (for broadcasting) or a household camera (for viewing), the images of players captured from a long distance are too small and obscure for accurate recognition because both camera images are low-resolution, and many cameras are required to prevent incorrect recognition due to players overlapping on screen. If we used a large-scale system for location tracking, in which tracking devices are deployed to each player and there are multiple detection points in the field, accurate tracking can be done. However, this is impractical for use by general people or for small-scale events due to the prohibitive cost.

In response, we have developed a method for identifying and tracking moving objects by combining wearable acceleration

sensors and image recognition technology that requires neither a large-scale location tracking system nor high-resolution cameras. The system recognizes two types of contexts: one acquired from the acceleration sensor each player wears and the other acquired from rough camera images. The system identifies the player in an image by matching these two contexts. We developed a prototype of the proposed method and performed an evaluation to determine its effectiveness.

This paper is organized as follows. In Section 2, we describe related work. The system architecture and evaluation are discussed in Section 3. We conclude the paper in Section 4.

2. Related Work

Conventional methods for object recognition that use image processing or acceleration sensors have been extensively studied recently, and some of them use context recognition techniques. Research on image processing is especially active these days, and image-based object recognition techniques are growing more accurate as a result.

For example, P. J. Figueroa et al. proposed a method for tracking soccer players that uses at least four static cameras. The tracking algorithm is based on a search of paths in a graph defined by blobs representing segmented players [1]. The different cases of occlusions or contact of these players are treated by splitting the corresponding blobs and taking into account features such as the number of components, the location of the blobs, players, trajectory, and so on. Anthony et al. propose player tracking method by using particle filter [2]. After detecting field in the image, their method detects players. To get coordinate, color and velocity information by using particle filter can make player tracking possible. Sunghoon et al. realize player tracking [3]. This method detect field by histogram, predict player position in the next image by Kalman filter and decide player position by Template matching.

Televised sporting events are one example of services utilizing such image recognition techniques, especially group sports such as soccer. Player recognition in televised sports is useful not only for helping viewers to grasp the play but also for assisting computer system with automatic scene detection which is often used for video summarization and retrieval. Yamada et al. proposed a location estimation method for detecting the position of players and the ball in soccer games by using image processing techniques. This system was highly accurate when using information specific to soccer games, such as the position of white lines[4]. However, it did not work well under poor sunlight conditions. Shimawaki et al. proposed a method to recognize a change of a scene by using color histograms and to determine whether the system should analyze the scene or not by using the location of a player, the field, and the placement of white lines [5]. They were able to achieve accurate, efficient scene detection using color distribution. Kasuya et al. proposed a method to improve player tracking by using two images that are taken synchronously by two fixed cameras placed at a high angle [6]. However, they reported several problems in their method, including false detection caused when one player is detected as two regions, and defective detection ccaused by changes of sunlight and the low-resolution images. S. Gedekli et al. proposed a computer vision system that is able to automatically extract consistent motion trajectories of soccer players from camera recordings [7, 8]. This system handles a number of difficult situations, such as fast movements, camera zooms and changes to lighting conditions. However, false detections do occur, caused by several other factors such as occlusions, player clustering, and strong shadows on the field. S. H. Khatoonabadi et al. proposed a method of tracking soccer players in goal scenes by using the information of the four white lines in the penalty area[9]. This system has missed detections when player occlusion occurred or players were mixed together. Ming's method uses eight cameras around the stadium and get images from many different angles [10]. This system classifies extraction objects into field players or goal-keepers or referees by using uniform color. In addition color information, the number of players is used to maintain the correct number of players in each team. This system realizes multiple player tracking, however it is not possible to identify individual players and it costs high because of using eight cameras. Ishii et al. propose a method that estimate 3D position of a soccer ball by using two cameras[11]. Their method detects the ball by choosing the detection algorithm based on the ball states to reduce the chance to miss the ball and the computation cost. The 3D position of the ball is estimated by the estimated 2D position of the two camera images. This system utilizes the Kalman Filter to compensate the ball position and predict the 3D ball position because loss of the ball in the image make it possible to obtain the 3D position. Kataoka et al. use monocular camera to track multiple players and ball in a soccer video [12]. They proposes a robust tracking method by combining Particle Filter and Classifer. This system try to track applying labeling and nearest neighbor algorithm. And more, it applies perspective transformation to extract player's position on the pitch. However, in the case of using image from a single camera, there are might be a lot of occluded of players. J. Liu et al. proposed a method for performing automatic multiple player detection, unsupervised labeling, and efficient tracktracking in broadcast soccer videos [13]. Although their system detects most of the players and can track them, even during long occlusions, serious video blur causes failure.

These studied indicate that if only image processing is used for location detection or tracking, the recognition accuracy is seriously changed by conditions such as sunshine. In addition, a high-resolution camera is required for accurate recognition. Although the location of an object in an image is important for the context recognition of a player and ball [14], accurate tracking is difficult because of changes to brightness or occlusion. J. Ren et al. proposed a system for estimating the trajectory of a soccer ball [15]. That uses a local matching process in place of the Kalman filter in the case of a merged ball. Using motion information and modeling the expected appearance of a moving ball are both used to improve the detection accuracy. Nicolai et al. achieved player tracking in real time by using multiple cameras [16]. M. Beets et al. analyzed soccer matches and the contexts of objects in real time and obtained location information by using receiver and sender devices [17, 18, 19]. However, receiver devices are expensive, and the system requires many of them surrounding the field.

Next, the aim of annotation is assumed to display information about person or objects. Annotation technique utilizing image recognition is active, so there are many related works these days.

For example, Assfalg et al.propose a system that performs automatic annotation of the highlights in soccer video [20]. Hilight detection exploits visual cues that are estimated from the ball motion, playfield zone, players' position and colors of players' uniforms. Yamamoto et al. present a Web-based video annotation system (Figure. **??**) that allows users to associate Internet video content with annotations [21]. The system analyzes video content to get image information and color histgrams. It sutomatically generates aWeb document that allows the users to edit the annotation. Nitta et al. propose method to automatically generate semantic annotations to significant scenes of broadcasted sports video [22]. It considers the structure of the game and sports program. This system extracts text segments where the actions and events occur from closed-caption by searching key phrases and make annotations by extracting events and related players from segments. In addition to this, the system extract image segments from image stream by template matching to determime segments displayed annotation. Finally, it associate the textual anotations with the image segments.

To sum up: conventional methods mostly utilize only image processing. The research of Kawai et al. is an example of using both image processing and sensors [23]– their method accurately identifies human locations by matching the location obtained by sensor with remote image. However, this system requires the off-line analysis of images and sensor data, which is time-consuming.

3. System Architecture

In our proposed method, the system recognizes moving objects by using a camera and an acceleration sensor. If a sensor is not used, recognition is difficult because of the overlap of moving objects in real time and low-resolution images. Therefore, our method identifies moving objects by matching the context obtained by a wearable acceleration sensor with the related context from the image processing. In this method, "*moving objects* " means the sports players.
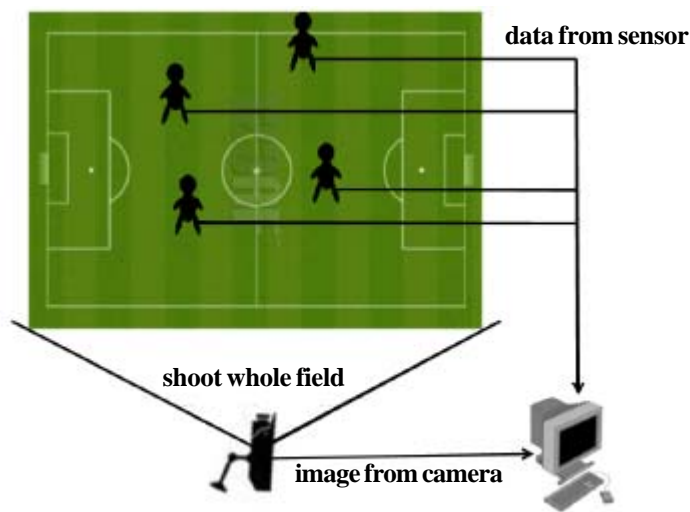

Figure 1. System architecture

Figure 1 shows the system architecture. Our aim was for this method to fulfill the following requirements.

• Can recognize multiple moving objects.
If the proposed method is used in sports (such as soccer) that feature multiple players, multiple moving objects need to be recognized.

• Can track moving objects.
If the trajectories of moving objects are identified, annotations can be displayed at the accurate location.

• Can perform object recognition in real time.
If the system is used in sports, real time processing is required for a variety of reasons. For example, coaches can acquire the running distance of each player and the played area so they can plan effective strategies during game.

3.1 Algorithm
Figure 2 shows the procedure for recognizing moving objects in our method. Objects are identified following the procedure below. Our method is focused on sports broadcasting, so we use soccer games as an example for explaining algorithms. In all our evaluations, a camera was placed at a distance from the field to grasp the whole entire playing area.

1. An image is taken by a camera. It is low-resolution and difficult to use to identify players.

2. Moving objects are extracted by using a background subtraction image technique, and the proposed method the object in the preceding frame with a nearby object in the current frame as the same object.

3. The proposed method allocates the image-based context for each tracked object. Table 1 shows examples of intended contexts. In this paper, our method uses contexts un-related to the soccer ball.

4. The sensor-based context of a player is also determined by the data obtained from the acceleration sensor he or she wear.

5. Our method identifies each player by comparing the contexts obtained from the image processing and from thesensor data (Figure 3).

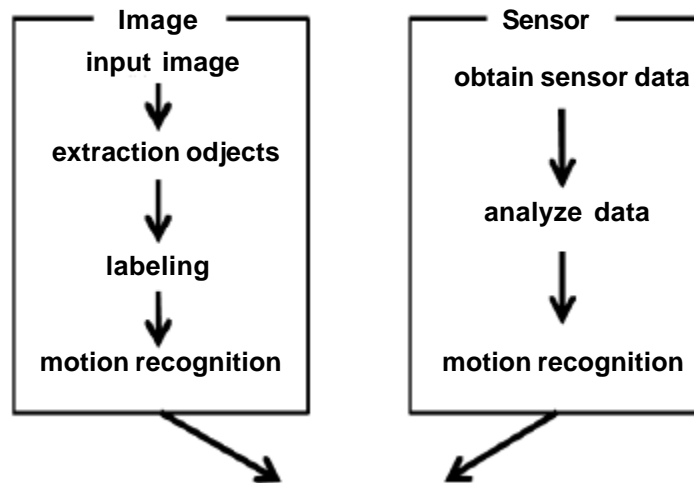This process is explained in detail in the following subsections.

Figure 2. Procedure for recognizing moving objects

3.1.1 Extraction of moving objects from image
In this section we describe the extraction of objects by the background differencing technique, the processing of extracted objects, the identification of objects between frames, object tracking, and the determination of context.
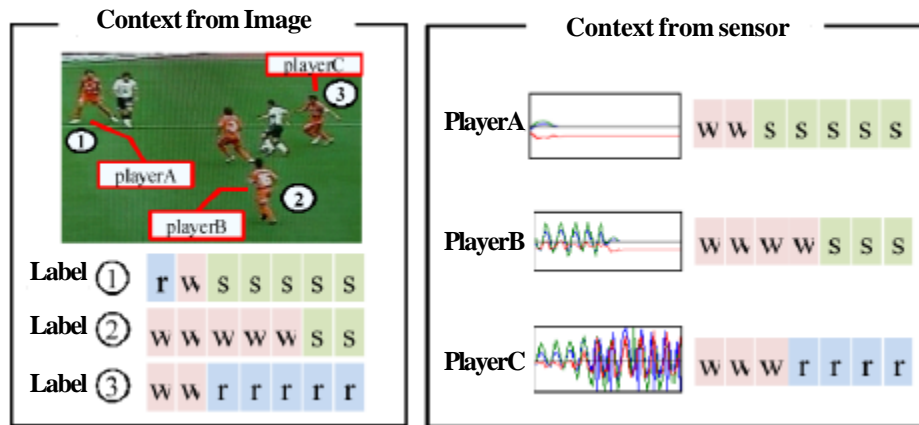
Figure 3. Context matching

First, objects in a image are recognized by image processing. Our method simply utilizes the background differencing technique to separate objects from the background. Techniques to separate objects from an unchanging background and the current frame are often utilized in computer systems. Although some conventional methods use brightness to separate objects from the background image, our method uses the difference from the initial background image. An image in which objects do not exist is used as the initial background image, and the difference between the current image and the initial background image is transformed to grayscale. Moving objects are extracted after the threshold processing and image denoising are complete.

Objects of various sizes are extracted due to noise and subtle vibrations of the camera, so regions that are too small and objects with an unnatural horizontal-to-vertical ratio are eliminated. Because most extracted players have longer length than width, this elimination is effective and most players are extracted accurately.

Next, our method determines if the extracted objects in the current frame are identical to the ones in the previous frame. A random number is assigned to each extracted object as a temporary ID. If the extracted objects are determined to be the same as those in the previous frame, the temporary number is changed to that of the object in the previous frame. The distance is calculated for every combination of the center of gravity coordinate of extracted objects in the previous and current frames. If the distance is the minimum and less than the threshold, the object in teh current frame is determined to be the same as the previous object. Figure 4 shows an example of allocation. Objects with solid lines are extracted objects in the current frame, objects with broken lines are extracted objects from the previous frame, the numbers in circles are labels of nearby objects, and the numbers in squares are the labels of nearby objects from the previous frame.

Context	Contexts recognition in image processing
stop	a player does not move
walk	a player moves slowly
run	a player moves quickly
pass	a ball leaves a player to another player
trap	a ball and a player overlap
fall down	a player's ratio of length and width changes
dribble	a ball moves with a player
shoot	a ball leaves a player to goal

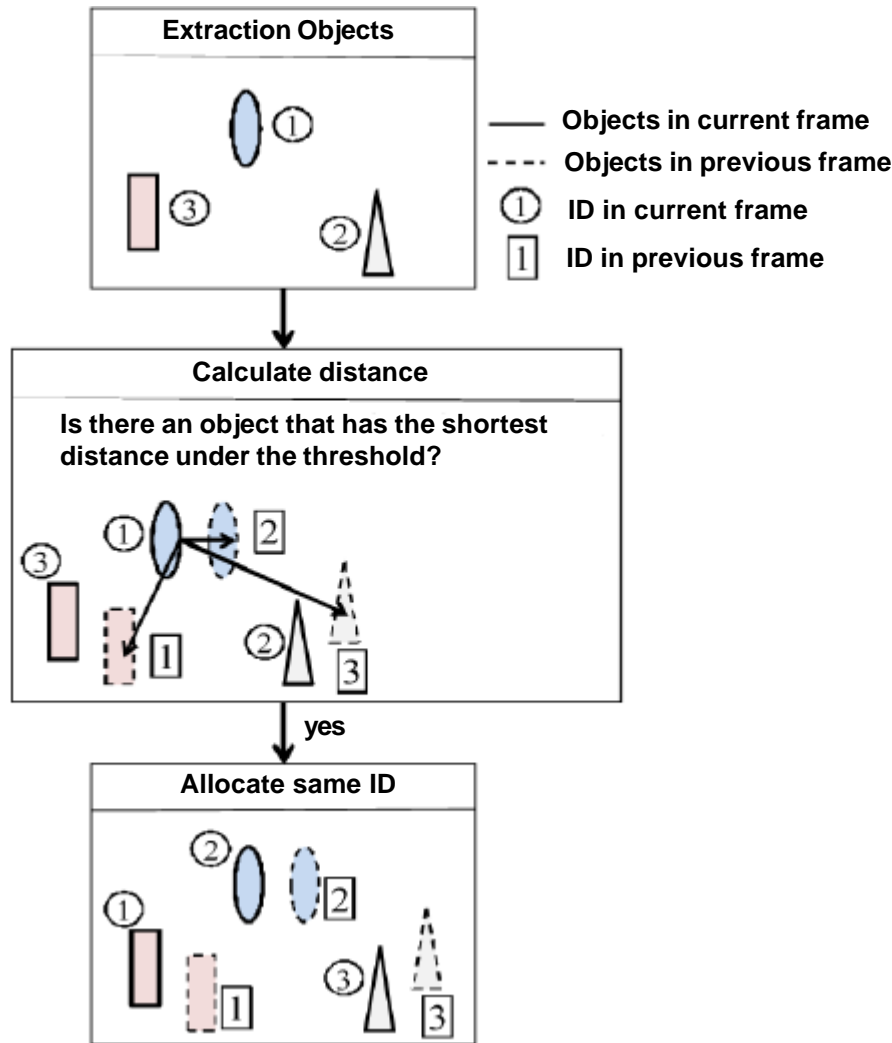Table 1. Examples of intended contexts

Figure 4. Procedure for ID allocation

If there is an object in the current frame that has no corresponding object in the previous frame, it is determined to have appeared in the image suddenly, as shown in Figure 5. In the same way, if there is an object in the previous frame that has no corresponding object in the current frame, our method determines that the object has disappeared, as shown in Figure. 6. The appearing or disappearing situation is mainly seen in merge and split. The object merge and split shows in Figure 7 and Figure 8. There are some cases where objects seem to appear and disappear in very quick succession because of various image processing conditions such as camera flash and noises. To avoid such misdetection, the center of gravity coordinate of disappearing objects is retained for a while and used to identify these types of disappearance. If an object disappear for more than a certain amount of time, it is determined that this object is no longer in the image. In addition, if an object appears close to a disappearing object within a given duration, this new object is assumed to be the same as the disappearing object. Our method assumes that the object closest to a disappearing object is the disappearing object itself, and the coordinate of the nearest object is assigned to the disappearing object.

Object tracking is performed by using this procedure for each frame. However, the tracking accuracy is expected to be low when several objects are recognized as one due to their close proximity.

3.1.2 Contexts from image processing
The context of each object is identified by examining the trajectory of the extracted object. For example, a context such as stop, walk, or run is identified by calculating the distance of movement between the positions of an object in the current flame

and the previous flame. In order to avoid the influence of noise, our method calculates the distance using not two continuous images but rather the distances of the two images (the current image and the image of *n* frames before). After this, our method stores the context for use in the matching phase.
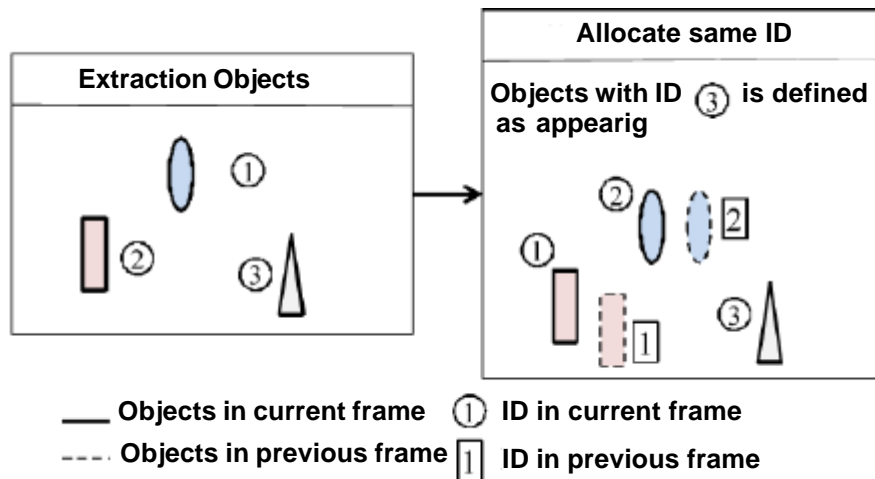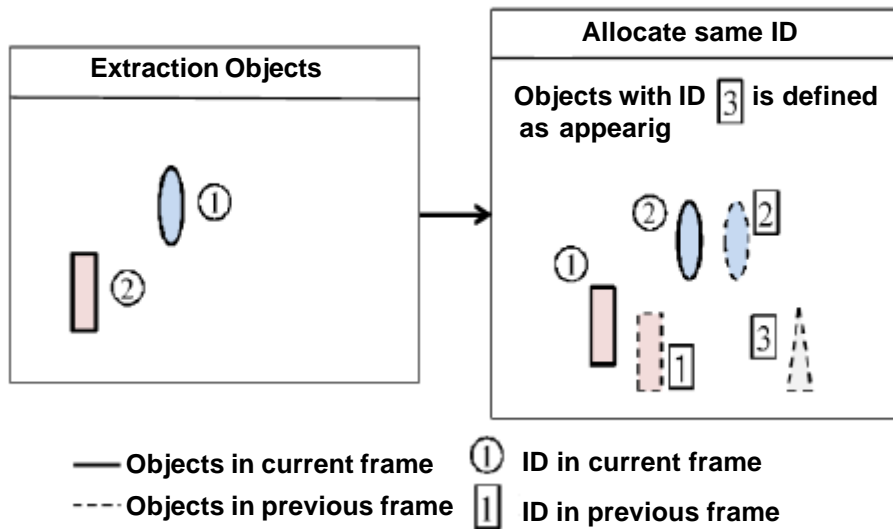

Figure 5. Object appearing

Figure 6. Object disappearing

3.1.3 Contexts from acceleration sensor

The context obtained from the acceleration sensor is identified by analyzing the numeric data obtained from a sensor. Where to place the sensor and how many of them to use is determined by referring to the contexts that are relevant to a soccer game. In our method, each player wears a sensor on his or her foot to record the contexts related to him or herself. The dispersion value in each axis for a constant period is calculated on the basis of sensor data. The window in which the peak of repeated contexts can exist is set to 0.8 seconds for recognizing certain repeated behaviors such as walking and running. Next, the context is identified by using the variance of the sensor data in the window as the feature value. Since the variance gets bigger as the context changes from stop to run, our method can determine the current context by setting thresholds for all contexts. A threshold is determined for each player. Our method stores the context derived from the acceleration sensor as well as that from image processing.

3.1.4 Context matching

Our method determines the player ID for each moving object by matching the context from the image with that from the
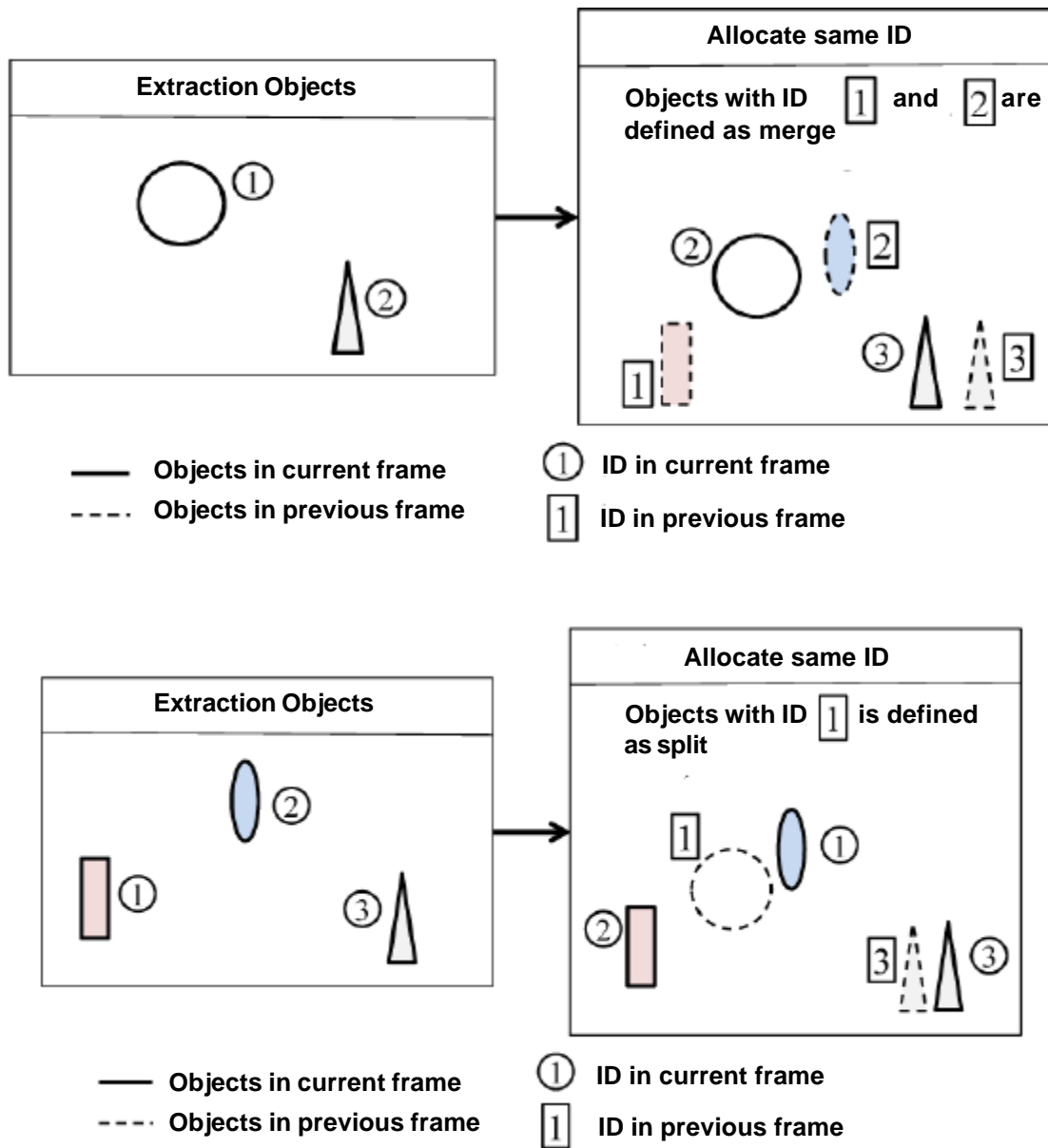
Figure 7. Object merge

acceleration sensor. Figure 9 shows the procedure for this matching. All the objects with broken lines in the previous frame have an ID (the player ID determined in the previous frame). First, in the current image our method finds the object that is closest to the IDed object from the previous frame. If the object is closest and the distance is less than the threshold, the object in the current frame takes the same ID as the previous object. After this, we compare each context that is a combination of objects that are closest together. If the number of same contexts exceeds the threshold, the combination of closest objects takes the same ID. If the number of same contexts does not exceed the threshold, our method uses context rather than distance to find the matching target. We count the number of same contexts in each combination of objects between the current and previous frames. When the combination of objects with the maximum number of same contexts is found, the matching target is decided. Figure 10 shows how we compare each context of playerB determined by the sensor with each context of label1 or label2 or label3 determined by the image processing.
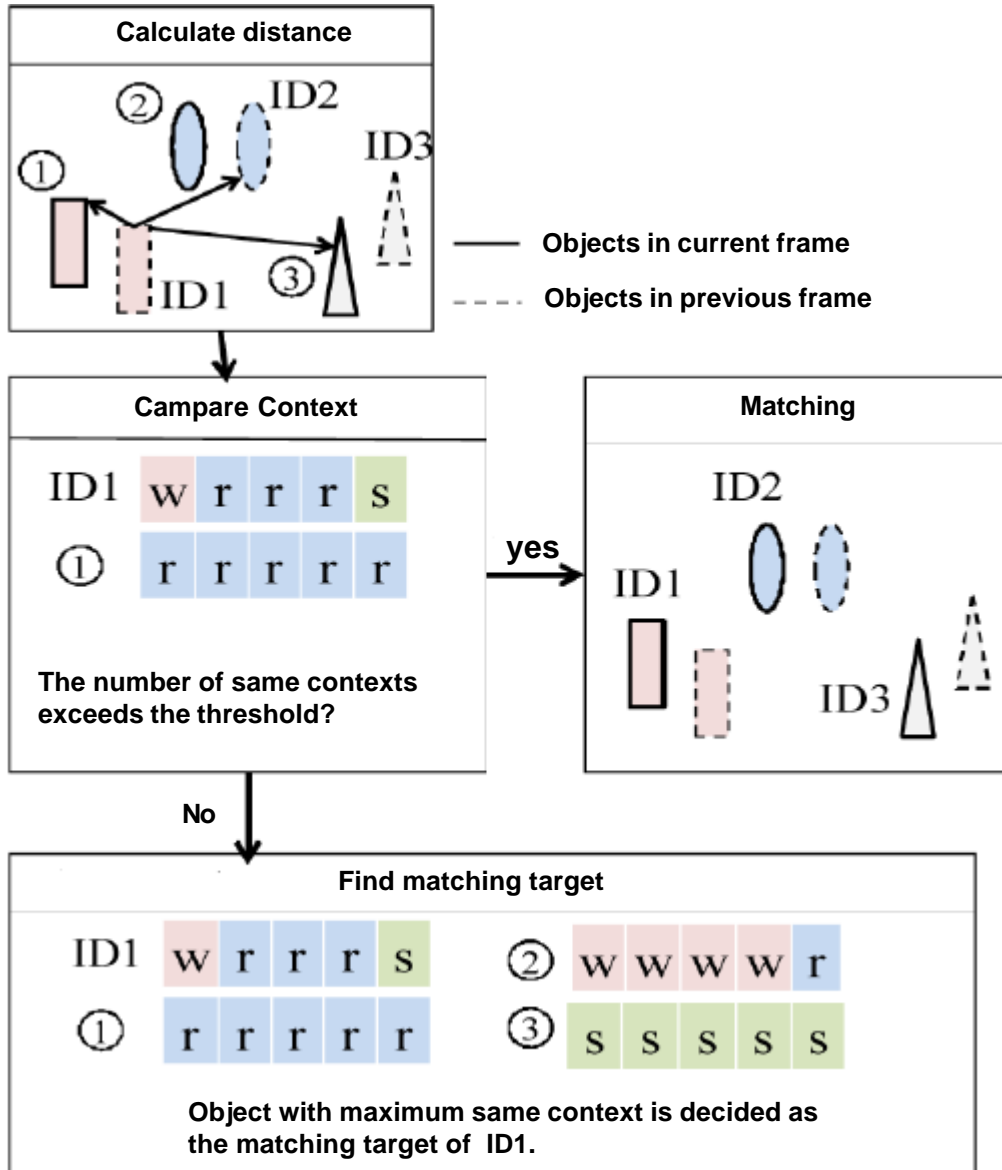
Calculate distance

ID2

ID3

① ②

ID1 ③

— Objects in current frame

- - - Objects in previous frame

Campare Context

ID1 | w | r | r | r | s

① | r | r | r | r | r

The number of same contexts exceeds the threshold?

yes →

Matching

ID2

ID1

ID3

No ↓

Find matching target

ID1 | w | r | r | r | s   ② | w | w | w | w | r

① | r | r | r | r | r   ③ | s | s | s | s | s

Object with maximum same context is decided as the matching target of ID1.

Figure 9. The matching procedure

Player B | w | w | w | w | w | w | r | r | r | s

label1 | s | s | s | s | s | r | r | r | r | r | **Same context : 7**

label2 | w | w | w | w | r | r | r | r | r | r | **Same context : 7**

label3 | r | r | r | r | r | r | w | r | r | r | **Same context : 2**

Figure 10. Comparing contexts

If the number of same contexts does not exceed the threshold for a long period of time, we use the combination of objects with the maximum number of same contexts. Alternatively, if the time period is short, we use the shortest distance to find the matching target.

3.2 Implementation

We implemented a prototype of our proposed method. Microsoft Visual C++ 2008 Express Edition was used for the implementation and Open Source Computer Vision Library OpenCV) was used for the image processing. OpenCV, which is a free library of programming functions developed and released by Intel Corporation for image processing and recognition, can perform high-speed, advanced processing such as edge detection, color conversion, and optical flow detection. The laptop was used for the implementation was a LaVie G made by NEC (OS: Windows 7, CPU: Intel(R) Core(TM) i3 2.27 GHz, RAM: 4.00 GB). The proposed method used not only real time data from aWeb camera and sensor but also data that had been stored in advance. The acceleration sensor we used was made by Wireless Technologies, Inc. (size: $36.5 \times 39 \times 10$ mm, weight: 17 g, transfer band: maximum 700 Kbps, communication distance: maximum 10 m, sampling frequency: 25 Hz) Figure 11 shows a multi-screen screenshot of the prototype. Figure 11(a) shows the graphed sensor data in x-axis, y-axis and z-axis. Figure 11(b) shows a view of object tracking. Figure 11(c) shows a screen on which annotations are displayed.
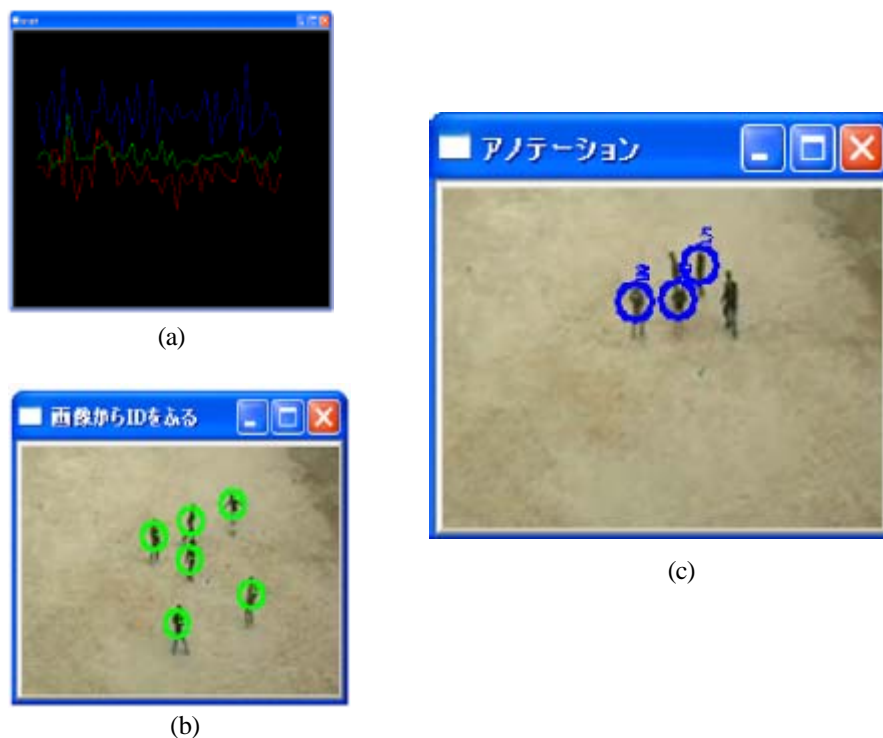

(a)

(c)

(b)

Figure 11. Screenshot of system

Object	Sensor	Image	Identication
3players	85%	71%	94%
3players	75%	56%	69%
3players	80%	67%	84%
4players	69%	50%	68%
4players	83%	66%	70%
5players	77%	60%	63%
6players	77%	59%	55%
6players	73%	58%	63%

Table 2. Individual accuracy

Figure 12. Situation of evaluation
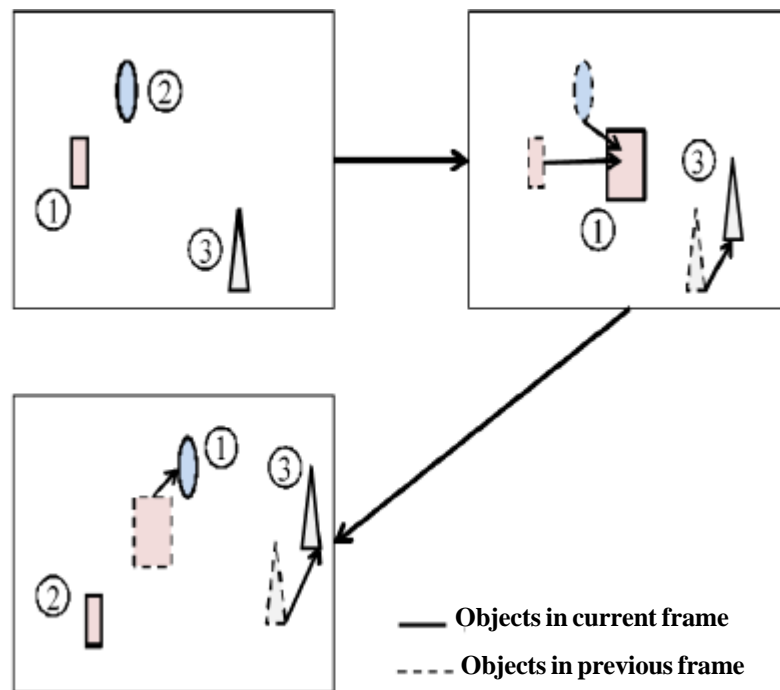

Figure 13. Image used evaluation

Figure 14. The cause of incorrect identication1

3.3 Evaluation

We evaluated our proposed method to determine its recognition accuracy by using sensor data and images from a lowresolution camera (Everio GZ-MG575 (720×480 pixel, 8.5 Mbps VBR), made by Victor Corp). The video camera was set approximately 10 away from the players and approximately 15 meters off the ground, which is based on the assumption that using only a video camera is inadequate for identifying objects. The 12shows situation of evaluation indicate where we set the video camera and the ??shows what kind of images we use. Each player wore an acceleration sensor on one foot. We collected data for two or three minutes. The thresholds for recognizing context from images and sensors were decided by giving the accuracy of context the most weight. The correct location of all players were manually prepared as answer and the correct contexts of each player were manually prepared as answer before evaluation. If the output from our method compared with the answer data is an approximate value, the identification is determined to be correct. The results are shown in Table 2. The first line shows the number of objects, and the second and third lines show how accurate the context obtained from the sensor and image technology are. The Fourth line shows the identification accuracy. In an experiment using just three people (the top line), the order of the contexts was decided in advance. In the other experiment, all participants played soccer freely.
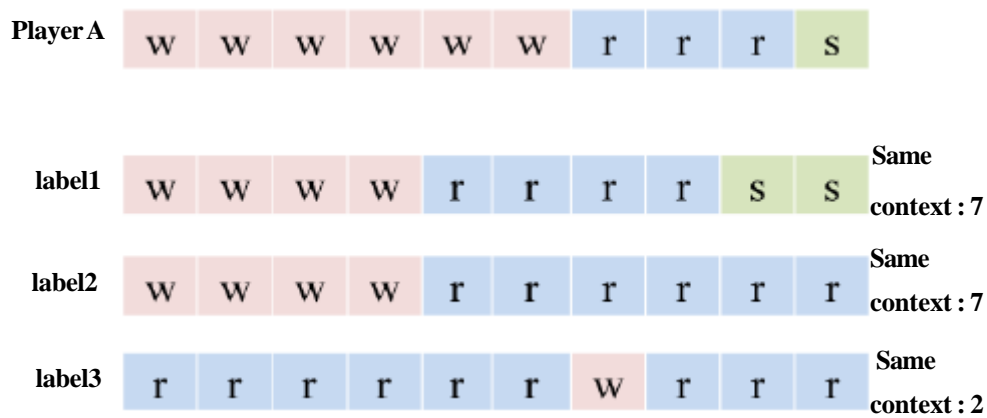
Player A	w	w	w	w	w	w	r	r	r	s

label1	w	w	w	w	r	r	r	r	s	s	Same context : 7

label2	w	w	w	w	r	r	r	r	r	r	Same context : 7

label3	r	r	r	r	r	r	w	r	r	r	Same context : 2

Figure 15. The cause of mistaking identication2

Player B	w	w	w	w	w	w	r	r	r	s

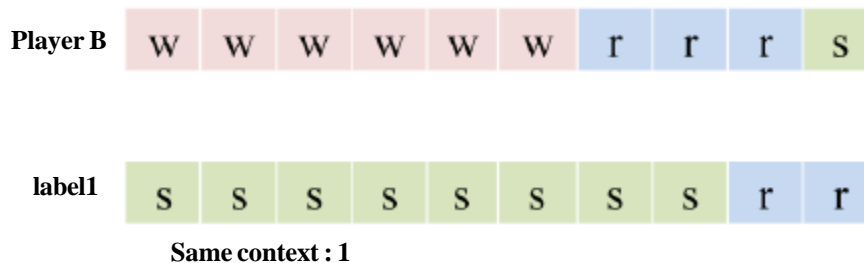| label1	s	s	s	s	s	s	s	s	r	r

Same context : 1

Figure 16. The cause of mistaking identication3

The results showed that the accuracy suffered as the number of objects increased. This is because labels rapidly interchanged due to splitting and merging. This situation is shown in Figure 14. In our experimental image, a player was connected with many matching targets when the matching targets were decided by using the number of same contexts because the contexts of all the player is similar (see Figure 15). Moreover, the context of the players did not correspond with the correct matching target for many frames (Figure 16).

4. Conclusions

We developed a method for identifying and tracking moving objects by using a combination of wearable acceleration sensors and image recognition technology. The proposed method recognizes two types of contexts, one of which is

acquired from acceleration sensors worn by soccer players and the other from rough camera images. Our method identifies the player in an image by matching these two contexts. We created a prototype of the proposed method and demonstrated its effectiveness through evaluation.

In our future work, we will focus on an algorithm for image processing to improve the identification accuracy. We also plan to expand the possible contexts because at the moment the context of each player is too similar. This should help improve the accuracy. Moreover, we want to recognize the timing of context changes from images and sensors more accurately, which could perhaps be done if players wore colored bibs. Overall, we want to implement improvements to make our method easier to use.

5. Acknowledgements

References

[1] Figueroa, P. J., Leite, N. J., Barros, R. M. L. (2006).Tracking Soccer Players Aiming Their Kinematical Motion Analysis. *Computer Vision and Image Understanding*, 101 (2) 122–135.

[2] Dearden, A., Demiris, Y., Grau, O. (2006). Tracking Football Player Movement from a Single Moving Camera Using Particle Filters. *In*: Proceedings of the 3rd European Conferenece on Visual Media Production (CVMP), 29–37.

[3] Choi, S., Seo, Y., Kim, H., Hong, K.-S. (1997). Where Are the Ball and Players?: Soccer Game Analysis with Color-based Tracking and Image Mosaick. *In*: Proceedings of the International Conference Image Analysis and Processing (ICIAP), 196–203.

[4] Yamada, A., Shirai, Y., Miura, J. (2002). Tracking Players and a Ball in Video Image Sequence and Estimating Camera Parameters for 3D Interpretation of Soccer Games. *In*: Proceedings of the 16th International Conference on Pattern Recognition, 303–306.

[5] Shimawaki, T., Miura, J., Sakiyama, T., Shirai, Y. (2006). Ball Route Estimation in Broadcast Soccer. *In*: Proc. of ECCV{2006 Workshop on Computer Vision Based Analysis of Sports Environments (CVBASE), 26–37.

[6] Kasuya, N., Kitahara, I., Kameda, Y., Ohta, Y. (2008). Robust Trajectory Estimation of Soccer Players by Using Two Cameras. *In*: Proceedings of the 19th International Conference on Pettern Recognition (ICPR), 4 p.

[7] Gedikli, S., Bandouch, J., Hoyningen-Huene, N. V., Kirchlechner, B., Beetz, M. (2007). An Adaptive Vision System for Tracking Soccer Players from Variable Camera Settings. *In*: Proceedings of the 5th International Conference on Computer Vision Systems (ICVS).

[8] Beetz, M., Gedikli, S., Bandouch, J., Kirchlechner, B., Hoyningen-Huene, N. V., Perzyhlo, A. (2007). Visually Tracking Football Games Based on TV Broadcast. *In*: Proceedings of International Joint Conference on Articial Intelligence (IJCAI).

[9] Khatoonabadi, S. H., Rahmati, M. (2009). Automatic Soccer Players Tracking in Goal Scenes by Camera Motion Elimination. *Image and Vision Computing*, 27 (4) 469–479.

[10] Xu, M., Lowey, L., Orwell, J. (2005). Architecture and Algorithms for Tracking Football Players with Multiple Cameras. *Vision, Image, and Signal Processing*, 152 (2) 232–241.

[11] Ishii, N., Kitahara, I., Kameda, Y., Ohta, Y. (2007). 3D Tracking of a Soccer Ball Using Two Synchronized Cameras. *Lecture Notes in Computer Science, Advances in Multimedia Information Processing (PCM)*, 4810 (3) 196–205.

[12] Kataoka, H., Hashimoto, K., Aoki, Y. (2011). Player Position Estimation by Monocular Camera for Soccer Video Analysis. *In*: Proceedings of the International Conference on Instrumentation, Control, Information Technology and System Integration (SICE), 1985–1990.

[13] Liu, J., Tong, X., Li, W., Wang, T., Zhang, Y., Wang, H. (2009). Automatic Player Detection, Labeling and Tracking in Broadcast Soccer Video. *Pattern Recognition Letters*, 30 (2) 103–113.

[14] Ariki, Y., Kubota, S., Kumano, M. (2006). Automatic Production System of Soccer Sports Video by Digital Camera Work Based on Situation Recognition. *In*: Proceedings of the 8th IEEE International Symposium on Multimedia (ISM), 851–860.

[15] Ren, J., Orwell, J., Jones, G. A., Xu, M. (2009). Tracking the Soccer Ball Using Multiple Fixed Cameras. *Computer Vision and Image Understanding*, 113 (5) 633–642.

[16] von Hoyningen-Huene, N., Beetz, M.(2009). Rao-Blackwellized Resampling Particle Filter for Real-Time Player Tracking In Sports. *In*: Proceedings of the 4th International Conference on Computer Vision Theory and Applications (VISAPP), 464–470.

[17] Beetz, M., Kirchlechner, B., Lames, M. (2005). Computerized Real-Time Analysis of Football Games. *IEEE Pervasive Computing*, 4 (3) 33–39.

[18] Beetz, M., Hoyningen-Huene, N. V., Bandouch, J., Kirchlechner, B., Gedikli, S., Maldonado, A. (2006). Camera-Based Observation of Football Games for Analyzing Multiagent Activities. *In*: Proceedings of the 5th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS), 42–49.

[19] Beetz, M., Stammeier, T., Flossmann, S. (2004). Motionand Episode Models for (Simulated) Football Games: Acuisition, Representation, and Use. *In*: Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS), 1370–1371.

[20] Assflag, J., Bertini, M., Colombo, C., Bimbo, A. D., Nunziati, W. (2005). Semantic Annotation of Soccer Videos: Automatic Highlights Identification. *Computer and Vision Image Understanding*, 92 (2-3) 285–305.

[21] Yamamoto, D., Nagao, K. (2004). iVAS: Web-based Video Annotation System and its Application. *In*: Proceedings of the 3rd International Semantic Web Conference (ISWC).

[22] Nitta, N., Babaguchi, N., Kitahashi, T. (2001). Automated Annotation to Significant Scened Based on Structure of Broadcasted Sports Video. *Transactions of the Institute of Electronics, Information and Communication Engineers of Japan (IEICE)*, J84-D-2 (8) 1838–1847.

[23] Kawai, J., Shintani, K., Haga, H., Kaneda, S. (2005). Identification and Positioning Based on Motion Sensors and a Video Camera. In: Proceedings of the 69th Annual Conference The 4th IASTED International Conference on WEB-BASED EDUCATION.

[24] Open CV, http://opencv.jp/.