Detecting Soccer Goal Scenes from Broadcast Video using Telop Region

Naoki Ueda[*], Masao Izumi[†]

Abstract

We propose a simple method to detect goal scenes from broadcast soccer video by using scoring telop regions. Scoring telop regions show us scores of both teams simultaneously with games. In the first step of our proposed method, we extract the telop regions which include scores of both teams. As next step, we try to recognize the score changes on the telop regions by using inter-frame difference between consecutive two frames. Then finally we can detect goal scenes as the timing of score changings. Experimental results show the ability of our proposed method.

Keywords: Soccer video analysis, telop, goal scene detection, inter-frame difference.

1 Introduction

Nowadays, it has become ordinary way to record and save a large amount of TV programs on HDD or disc recorders at any home. And you are able to see TV programs anytime you want, and you are able to collect some kinds of your favorite sports TV programs on your own recorders. But on the other hand, amount of these recorded videos is going to be larger and larger, so you need time to find videos which you want to see from your recorded video library. Especially sports TV program, you need much time to find your favorite scenes such as homerun scenes in baseball video for example. Therefore automatic scene search methods from sports video have been studied recently, and many researches have been proposed. In soccer videos, there exists some kinds of events that you want to see. For example, they are kick-off scene, corner kick scene, free kick scene, shoot scene, passing scene, dribble scene and goal scene. Those of scenes are focused on the area of automatic sports video analysis systems. There have been many studies about these kinds of researches. Most of them are using multimodal information in order to detect certain scenes from videos. Combination of visual information such as players' trajectories, the linguistic information such as live speech or the acoustic information such as cheers, and also the break term information which shows the breaks of matches, have

[*] Graduate School of Engineering, Osaka Prefecture University, Osaka, Japan

[†] Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Osaka, Japan

been applied to detect several kinds of scenes from soccer videos [1]. And also the cooperative processing of text, sound and image have been used to detect highlight scene and to index scenes [2]. The summary information such as news articles have been applied [3]. Commentator's commentaries have been applied to detect goal scenes [4]. And the state of camera which has been used to take scenes has been estimated to extract specific events [5]. Image analysis have also been applied to detect automatic game analysis [6]. But these approaches were limited to detect specific scenes from sports videos. Focusing on the goal scenes, there exits availabilities to use detected 'scores' displayed on the video in order to detect goal scenes. The information of score telop area has been used to generate digest sports video [7][8]. In this paper, we have focused on the method by using 'scores' displayed on video as 'telop' regions [9]. Every goal scenes are the keys to change 'scores', so if we are able to detect the changing 'scores' from telop regions, goal scenes can be detected.

In section 2, the overviews of our proposed method are explained. In section 3, we describe the way of extracting telop regions from video. And the method of extracting 'scores' changes in telop regions are described in section 4. And then in section 5, the method of detecting goal scenes is explained. Section 6 shows the experimental results, and the last section, section 7 is concluded.

2   Overviews

We explain the flow of goals. That has two patterns. In *pattern 1*, firstly, when the goal occurs, the telop region disappears. Secondly when the telop region appears again, the score changes. In *pattern 2*, firstly, after the telop region disappears, the goal occurs. Secondly when the telop region appears again, the score changes. Figure 1 shows these two patterns of the flow of goals.
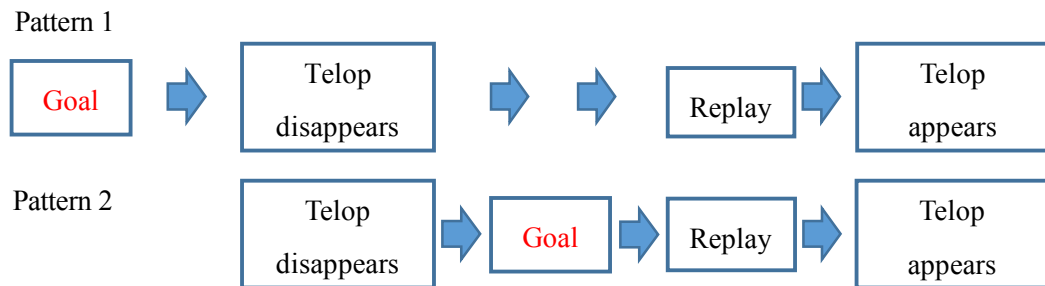

Figure 1: Flow of goals.

Therefore, we detect the goal scene by recognizing such a score change of the telop region from entire soccer video.

In soccer video, the camera moves freely. Furthermore, regions other than the telop will change along with the game progress. Thus, it is difficult to recognize only a score change of the telop region from original soccer video. Therefore, in this study, firstly, we extract only the telop region from the input frames by carrying out some process, such as lawn region delete, the in-

ter-frame difference, expansion processing, and labeling. Secondly, we recognize a score change of telop region. When the goal occurs, the telop region disappears, or after the telop region disappears, the goal occurs. Then, when the telop region appears again, the score changes. Thus, we recognize disappearance of the telop region, and then examine the change of the pixel in the telop region when the telop region appears again. On the basis of this tendency, we try to recognize a score change. Finally, we detect the goal scene by using the score change obtained. Figure 2 shows the outline of our proposed method.
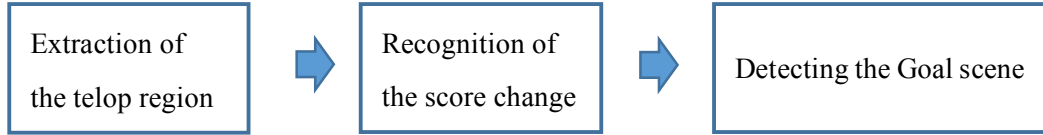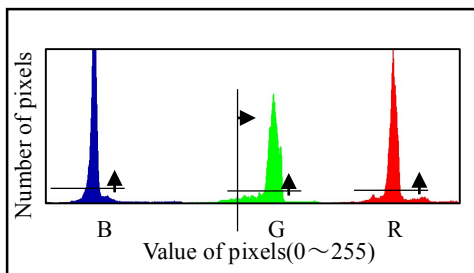

Figure 2: Flow of proposed method.

3 Extraction of telop region

In this section, we explain the method of extracting the telop region by removing the non-telop regions.

3.1 Lawn region delete

In soccer video, during the game progress, lawn region occupies most of the screen. Therefore, we use this feature. Firstly, we create an RGB histogram of the input frame (Figure 3). Secondly, if in the input frame, pixels have the value of G that are over 80, and pixels have more than 10% of the maximum value of the number of pixels in each histogram, we remove the region as lawn region. That's because they are largely responsible for the green grass. Finally, we perform a binary process for image. Figure 4 shows an example of this process.

$$\begin{cases} (0,0,0) & if \left((G > 80) \cap \begin{pmatrix} B > 0.1 \times h_B \\ G > 0.1 \times h_G \\ R > 0.1 \times h_R \end{pmatrix} \right) \\ (255,255,255) & otherwise \end{cases} \quad (1)$$

Figure 3: RGB histogram of the input frame.

(a)Input frame.   (b)Result frame.

Figure 4: Lawn region delete.

3.2 Inter-frame difference

Telop region is always located where the preformatted screen. Using this feature, we extract telop region by inter-frame difference. We tried to take out the pixel of which the value does not change over a plurality of frames in comparison to a frame having a telop region that holds the

score, "0-0". However, the value of pixels varies by background, because the telop region is translucent. Therefore, when we left only region where the result of inter-frame is zero, we cannot extract telop region well. Thus, we decided to leave the region where the result of inter-frame difference is equal to or less than a threshold value, as a candidate of telop region. With respect to the input frames consecutive, we apply inter-frame difference until the number of frames reaches the 100 sheets. Figure 5 shows an example of the results of this process. In our experiments, 100 frames were selected manually that shows the telop.

$$result(0) = 255$$

$$result(i) = \begin{cases} 255 & if & \begin{pmatrix} (absdiff(dst, src) < 40) \\ \cap \\ (result(i-1) = 255) \end{pmatrix} \\ 0 & & otherwise \end{cases} \quad (i = 1\sim100) \quad (2)$$

Where, *dst* is value of pixels which a frame that holds the score, "0-0" has, *src* is value of pixels which an input frame has, *result* is value of pixels after processing, *absdiff* is value of pixels after inter-frame difference processing.

(a) Input frame.   (b) Result frame.

Figure 5: Inter-frame difference.

3.3 Expansion processing

We perform expansion processing by the morphology operation. That's because we extract telop region without exception from the candidate of telop region obtained by inter-frame difference. Figure 6 shows an example of an expansion processing in the binary image. If there is even one white pixel to 8 in the vicinity of the 3×3 pixels, centering on the target pixel, we replaced the target pixel to white. Figure 7 shows the results of the process.
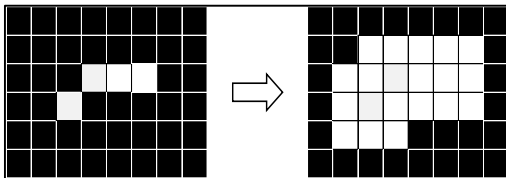

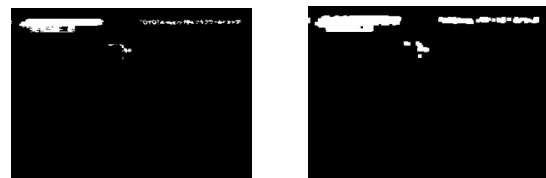Figure 6: Morphology operation.

Figure 7: Result of Morphological expansion process.

3.4 Labeling

Finally, we perform the labeling process. This is because we determine the telop region from the candidate of telop regions. Labeling is the process of allocating the same number for each con-

nected region. Figure 8 shows an example of labeling process.

We take out the minimum inclusion rectangular by obtaining the upper left coordinates and lower right coordinates of regions of every same number. Then, we decide that the region in which the number of the pixel is the largest is telop region. Figure 9 shows gzvcevkqp"qh"vgrqr region by using labeling.

0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	1	1	1	1	1	0
0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0
0	2	2	2	0	0	1	0	0	3	0
0	2	2	2	0	0	1	0	0	3	0
0	2	2	2	0	0	1	0	0	3	0
0	2	2	2	0	0	0	0	0	3	0
0	0	0	0	0	0	0	0	0	3	0
0	0	0	0	0	0	0	0	0	0	0

Figure 8: Example of labeling.
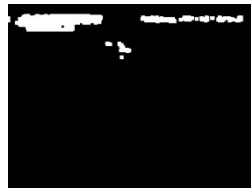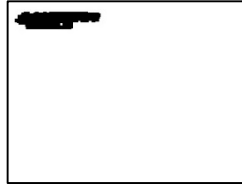
(a) Input frame.　　　　(b) Result frame.　　　(c) Extraction of the telop region.

Figure 9: Extraction of telop region by using labeling.

4　Recognition of the score change

In this section, we explain a method of extracting the score change from telop region by using the extracted telop region in Section3.

When the goal occurs, the telop region disappears, or after the telop region disappears, the goal occurs. Then when the telop region appears again, the score changes. Thus, firstly, we try to extract the timings of the disappearance and appearance again of the telop region. Secondly, we recognize the score change by examining the change of the pixel in the telop region when the telop region appears again.

4.1　Inter-frame difference

With respect to the input frames consecutive, we do inter-frame difference for each frame. Then, we leave the region where the result of inter-frame difference is equal to or less than a threshold value. Furthermore, we examine the change of the pixel of telop region. Those are because we find a frame which score of telop region changes, for the entire soccer video, as well as in section 3.

4.1.1　Recognition of disappearance of telop region

Firstly, we recognize the disappearance of the telop region. When the disappearance of the telop region starts, semi-transparency of the telop region is increased, eventually telop region disappears. By utilizing its features, a frame in which telop region begins to fade, will have a slightly different pixel values in comparison to the telop region in the frames before disappearance. Therefore, when we apply inter-frame difference continuously, if inter-frame difference of the

telop region becomes more than a threshold value, we decide it as a candidate of disappearance of telop region. Figure 10 shows the difference between telop regions before disappearance and starting disapperance.
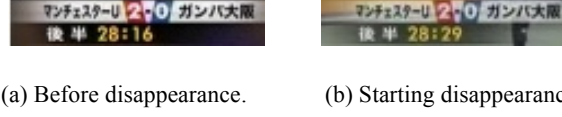

(a) Before disappearance.   (b) Starting disappearance.

Figure 10: Starting disappearance of telop region.

$$\begin{cases} terop\ region\ disapears & if\ (before - after) > 100 \\ \qquad not\ disapear & otherwise \end{cases} \quad (3)$$

Where, *before* is the number of pixels of inter-frame difference result of the previous frame, *after* is that of the current frame.

Then, if the result of inter-frame difference in a frame after five frames is equal to or less than a threshold value, exists, we recognize that telop region disappears like Figure 11. That's because we confirm the disappearance of the telop region. In other cases, we decided that it is mere reduction of pixels.

$$\begin{cases} Disappearance & if\ \ 5frame\_after < 200 \\ Not\ disappearance & otherwise \end{cases} \quad (4)$$

Where, *5frame_after* represents the number of pixels of the difference result after five frames.
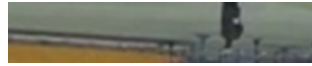

Figure 11: Disappearance of telop region.

4.1.2 Recognition of appearance of telop region

Then, we recognize the appearance of telop region. After telolp region disappears, we continue to do inter-frame difference. Then, the result of inter-frame difference sometimes is 0. Thus, we complemented the result of inter-frame difference to 1 in order to continue processing in order to continue processing. When telop region begin to appear, the result of the difference is large. That's because the region having a pixel value similar to the telop region increases. Thus, after the telop region disappears, we continue to perform the inter-frame difference. Because we try to find appearance of telop region again. When the result of inter-frame difference becomes more than a threshold value, becomes, we recognize that telop region appears. To extract the fully appeared telop, we decide that a region after the 10 frame is telop region. Figure 12 shows the difference of these two telop regions. In other cases, we decided that the score didn't changed.

$$\begin{cases} Appearance & if\ \ after > disappear \times \dfrac{2}{3} \\ Not\ appearance & otherwise \end{cases} \quad (5)$$

Where, *disappear* is the number of pixel of the difference result in frames before telop region disappears. *after* represents the number of pixels of the difference result in the current frame.

(a) Starting appearance.　　　(b) After 10 frames.

Figure 12: Appearance of telop region.

4.1.3 Recognition of the score change of telop region

If, when the telop region appears again, we examine the change of the number of pixels between the result of inter-frame difference in a frame before telop region disappears and that after it appears again. In addition if it is equal to or more than a threshold value, we recognize that the score change occurs in the frame. Figure 13 shows aspects of the score change.
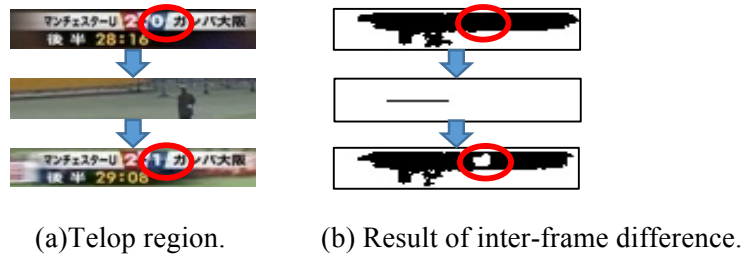

(a)Telop region.　　　(b) Result of inter-frame difference.

Figure 13: Recognition of the score change.

$$
\begin{cases}
Score\ change & if\ (disappear - appear) \\
& > disappear \times 0.025 \quad (6) \\
Not\ score\ change & otherwise
\end{cases}
$$

Where, *disappear* is the number of pixels before telop region disappears, *appear* represents the number of pixels after telop region appears.

5　Detection of the goal scene

In this section, we explain the method of estimating the goal scene by using a score change of
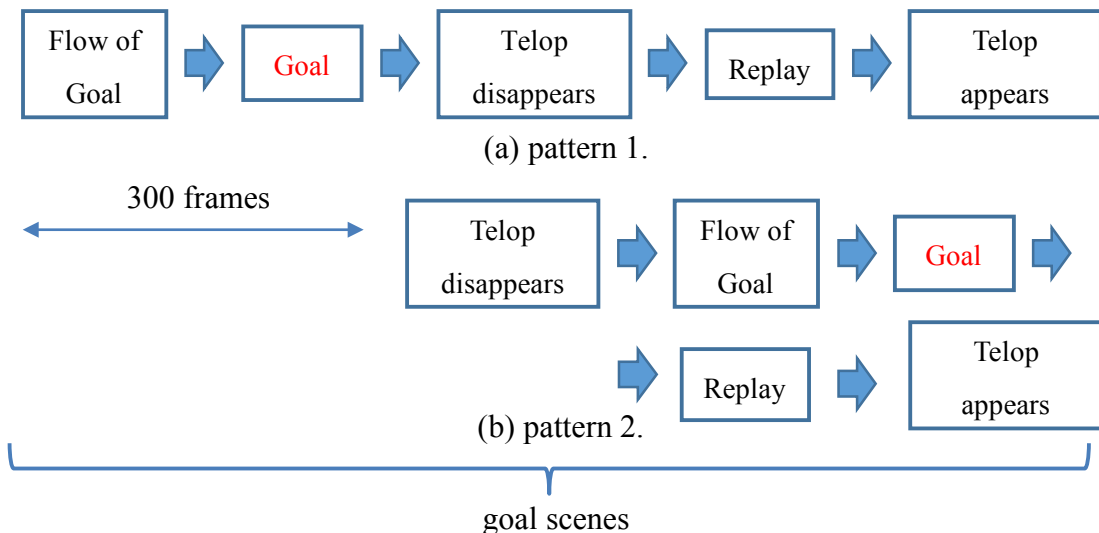

(a) pattern 1.

(b) pattern 2.

goal scenes

Figure 14: Method of estimating the goal scenes.

telop region recognized in section 4. The goal scene has two patterns. It occurs before telop region disappears or after that. Therefore, we extract both of them. We can confirm the flow of the goal scene in the experiment of the five games when we go back to 300 frames from disappearance of the telop region. Then finally we decided that 300 is the best number of frames experimentally. Thus, we estimate frames to be the goal scene, from 300 frames before the telop region disappears to appearance of it. Figure 14 shows two patterns of goal scenes which we use the periods of goal scenes in our experiments.

6 Experimental results

We have tested the effectiveness of the proposal method through experiments using real soccer game broadcast videos.

6.1 Environment

In this study, the subjects of the experiments are five matches of broadcast video of the soccer games that are Gamba Osaka vs. Manchester United, semi-final of Club World Cup performed on December 18, 2008, Japan vs. Lebanon, performed on March 3, 2004, Japan vs. Singapore, performed on March 31, 2004, England vs Italy, performed on July 24, 2012, Czech vs Japan, performed on April 28, 2004. The analysis was carried out by using the OpenCV [10] that is a library of programming functions mainly aimed at real-time computer vision, originally developed by Intel's research center in Nizhny Novgorod. The library is cross-platform and free for use under the open-source BSD license. This time, we determined manually the start frame of the treatment in which telop region exists. We evaluated this experiment by using the precision ratio and recall ratio. Precision ratio represents whether it is able to correctly extract the goal scenes. Recall ratio represents whether it is able to extract without omission the goal scenes.

$$precision = \frac{\sum TP}{\sum TP + \sum FP} \ (7) \qquad recall = \frac{\sum TP}{\sum TP + \sum FN} \ (8)$$

Where, ΣTP is correctly extracted goal number, ΣFP is erroneous extraction number, ΣFN is the extraction leakage number.

The following shows the condition of this experiment.

- resolution: 320 x 240 pixels

- frame rate: 30 frames/second

- number of images: Match A = 33000, Match B =172000, Match C = 170000

    Match D = 170796, Match E = 161789
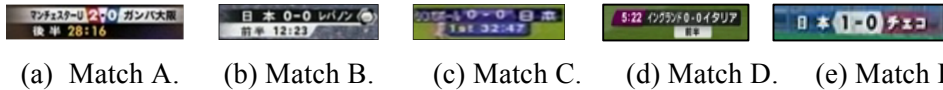

(a) Match A.   (b) Match B.   (c) Match C.   (d) Match D.   (e) Match E.

Figure 15: Telop region in three matches.

And Figure 15 shows five different telop regions extracted from five different matches.

6.2 Result

Table 1 shows the detected results of the goal scenes in this study.

Table 1: Detected results of the goal scenes in this study.

	Match A	Match B	Match C	Match D	Match E	Total	Recall ratio[%]	Precision ratio[%]
Number of extraction	5/5	3/3	3/3	0/0	0/1	11/12	91.7	100

If the goal scenes are contained from 300 frames before disappearance of the telop region to a frame of appearance of the telop region, we decided that we were able to extract the goal scenes correctly.

6.3 Discussion

We compared the accuracy of this study and previous studies [11] [4] [1]. Table 2 shows the detected result of the goal scenes in previous studies.

Table 2: Detected results of the goal scenes in previous studies.

Information	The number of extraction	Recall ratio[%]
Acoustic [11]	24/26	92.3
Commentator's commentaries [4]	5/13	38.5
The break term [1]	1/2	50

In comparison to the previous studies in Table 2, our method obtained higher ratio. 'Acoustic' method in Table 2 could obtain 92.3% recall, but in this method, they were able to extract only 'important' scenes, such as goal scenes, corner kick scenes, free kick scenes, etc. And they couldn't distinguish which is 'goal', or which is 'free-kick', and so on. So precision of extracting 'goal' scenes in 'Acoustic' method is very low. On the other hand, our method can obtain higher ratio in both recall and precision. We used the score change of telop region that was the common information in any soccer video. Therefore, we consider that the accuracy of detection has improved. We consider that in match E detection failed because the transparency of the telop region was high and the pixel change was intense. As a solution to this problem, it is conceivable to distinguish between a pixel change around the telop region and a pixel change in the other area. And we also should think about automatic threshold values detecting in order to fit each video condition.

6   Conclusion

In this study, we proposed a method of detecting the goal scenes in soccer video, by recognizing the score change of telop region. We obtained high precision ratio and recall ratio in this exper-iment. We show that this study is very effective in detection of goal scenes, comparison to the previous studies.

We have some future issues. Firstly, that is to recognize which team's score changed. We consider that we can realize this by recognizing the position of a goal mouth reflected in the video when goal scenes occur. Secondly, that is to get automatically a start frame of the process when we extract the telop region, because we got it manually. We consider that we can do that by using the feature quantity such as a straight line appearing in the telop region. Finally, that is to confirm that the proposed classification criteria can adequately detect the telop region for a wider set of matches from different nations and broadcasters because we tested the effectiveness of the proposal method through experiments using the small sample size of evaluation videos (five soccer matches, presumably all from Japanese broadcasters). As a future goal, we will acquire the flow to the goal by combining our method of acquiring the goal scenes and the method of acquiring the position information of players and a ball. By collecting scenes related to similar goal flow, it is expected to be utilized for tactical analysis.

References

[1] H. Atobe, M, Izumi, K. Fukunaga, "Event detection from soccer video by Using break term information", Vol.106, No.606, pp. 25-30, PRMU, 2007.

[2] S. Miyauchi, N. Babaguchi and T. Kitahashi: "Highlight Detection and Indexing in Broadcast Sports Video by Collaborative Processing of Text", Audio, and Image, Systems and Computers in Japan, Vol. 34, No. 12, pp. 22-31, 2003.

[3] N. Fukino, Q. MA, K. Sumiya, K. Tanaka, Generating Football Video Summery Using News Article", DEWS, 8-P-03, 2003

[4] I. Yamada, M. Sano, H. Sumiyoshi, M. Shibata, N. Yagi, "Automatic Generation of Segment Metadata for Football Games Using Announcer's and Commentator's commentaries", IEICE Trans on Information and Systems, Vol. J89-D, No.10, pp. 2228-2237, 2006.

[5] Y. Iwai, J. Maruo, M. Yachida, T. Echigo, H. Miyamori, and S. Iisaku, "A Framework of Visual Event Extraction from Soccer Games", Asian Conf. on Computer Vision, pp. 222-227, 2000.

[6] Y. Nakagawa, "Automation of the Soccer Game Analysis", UNISYS TECHNOLOGY REVIEW, Vol.76, pp.21-38, 2003.

[7] T. Tamura, C. Xiaocqin, "Detection of Telop from Broadcasted Soccer Video for Making a Digest Video", IEICE, General Conf., D-11-34, 2012.

[8] H. Arai, H. Kuwano, S. Kurakake, T. Sugimura"Detection of Telop in Video Data", IEICE Trans on Information and Systems, Vol. J83-D2, No.6, pp.1477-1486, D-II, 2000.

[9] N. Ueda, M. Izumi, "Detecting Soccer Goal Scenes from Broadcast Video using Telop Region", Asian Conference on Information Systems, pp. 22-28, 2016.

[10] "OpenCV", http://opencv.jp/, (2017.6.20 accessed).

[11] T. Shiozaki, S. Ohira, M. Honda, K. Shirai,"Soccer Video Indexing based on Acoustic Signal Processing", IPSJ Forum on Information Technology, pp. 107-108, 2004.