Paper

Active Grid-based Pass Region Estimation from Multiple Frames of Broadcast Soccer Videos

Sho Takahashi (student member)[†], Miki Haseyama (member)[†]

Abstract   An Active grid-based method for estimating pass regions from broadcast soccer videos is presented in this paper. It is assumed that the pass region has a high probability of the pass succeeding. In soccer matches, players discover pass regions based on previous and current player positions. In conventional methods, pass regions are estimated by applying Active Net to only a single frame of a soccer video. In the proposed method, Active grid is applied to three-dimensional data by which frames of the soccer video are connected with the temporal dimension. The proposed method then realizes robust estimation of pass regions based on multiple frames of player positions. The proposed method was applied to actual TV programs to verify its effectiveness.

1. Introduction

Systems for analyzing video contents must be developed to realize efficient video retrieval. Many researchers in the field of multimedia processing have proposed systems for indexing or extracting specific events based on audio-visual features in news and sports programs[1)~6)]. Since soccer is one of the most popular sports, various systems for analyzing soccer videos have been proposed. Some systems perform indexing of several events in soccer videos to generate highlight scenes for a sports news program[4)~6)]. However, many viewers are interested in not only specific events but also tactics and players' skills. Hence, some alternative techniques for analyzing tactics and players' skills are necessary.

There already exist traditional methods that estimate player positions and pitch areas for soccer content analysis[7)~10)]. Furthermore, by using the estimated player positions, we have proposed some methods for analyzing tactics in soccer videos[11)12)]. Specifically, we proposed a method for estimating pass regions* from player positions within a single frame in a broadcast soccer video[12)]. However, in a soccer match, players discover pass regions based on previous and current player positions. Therefore, the accuracy of the previous method is limited. Thus, by utilizing player positions of multi-

ple frames, more successful pass region estimation can be expected.

In this paper, we propose a new method for estimating pass regions from a broadcast soccer video based on Active grid[13)]. Active grid is a method for extracting a region from a three-dimensional space. In the proposed method, Active grid is applied to three-dimensional data by which generated data from each frame of the soccer video are connected with the temporal direction. The three-dimensional data include data for previous and current player positions. Therefore, since the proposed method utilizes player positions of multiple frames by introducing Active grid, more successful pass region estimation is realized.

This paper is organized as follows. In Section **2**, we explain Active grid used in the proposed method. In Section **3**, an overview of the proposed method is explained. In Section **4**, generation of the energies for Active grid applied to a soccer video is explained. In Section **5**, the method for estimating the pass region based on Active grid is proposed. In Section **6**, the effectiveness of our extraction method is verified from experimental results. Finally, concluding remarks are presented in Section **7**.
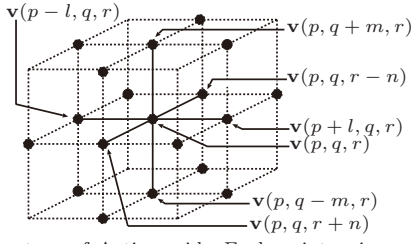
2. Active Grid

In this section, we explain Active grid[13)]. Active grid is a rectilinear grid structure model based on minimizing energy to extract a target object from three-dimensional data. Active grid is represented as parameters of a three-dimensional point ar-

Fig. 1 Structure of Active grid. Each point $\mathbf{v}$ is connected to six neighboring points.

ray. It is composed of the lattice points $\mathbf{v}(p, q, r) = (x(p,q,r), y(p,q,r), z(p,q,r))$, $0 \leq p \leq 1$, $0 \leq q \leq 1$ and $0 \leq r \leq 1$, as shown in **Fig. 1**. Each point $\mathbf{v}(p, q, r)$ is connected to 6 neighboring points, $\mathbf{v}(p - l, q, r)$, $\mathbf{v}(p + l, q, r)$, $\mathbf{v}(p, q - m, r)$, $\mathbf{v}(p, q + m, r)$, $\mathbf{v}(p, q, r - n)$ and $\mathbf{v}(p, q, r + n)$, to form a three-dimensional grid. The values p, q and r are parameters that represent the grid shape and satisfy $0 \leq p \leq 1$, $0 \leq q \leq 1$ and $0 \leq r \leq 1$. Furthermore, $l = 1/(X - 1)$, $m = 1/(Y - 1)$ and $n = 1/(Z - 1)$, where X, Y and Z are the number of lattice points in the X, Y and Z axes, respectively.

Active grid has an energy

$$E^*_{grid} = E^*_{internal} + E^*_{image}, \tag{1}$$

that consists of an internal energy $E^*_{internal}$ and a matching energy E^*_{image}, and they are defined as follows:

$$E^*_{internal} = \int_0^1 \int_0^1 \int_0^1 E_{int}(\mathbf{v}(p, q, r)) dp dq dr, \tag{2}$$

$$E^*_{image} = \int_0^1 \int_0^1 \int_0^1 E_{img}(\mathbf{v}(p, q, r)) dp dq dr. \tag{3}$$

E_{int} is utilized to smooth the shape of Active grid and is written as

$$\begin{aligned} E_{int}(\mathbf{v}(p, q, r)) &= \alpha E^*_{int} + \beta E^{**}_{int}, \tag{4}\\ E^*_{int} &= (|\mathbf{v}_p|^2 + |\mathbf{v}_q|^2 + |\mathbf{v}_r|^2),\\ E^{**}_{int} &= (|\mathbf{v}_{pp}|^2 + 2|\mathbf{v}_{pq}|^2 + |\mathbf{v}_{qq}|^2 \\ &\quad + 2|\mathbf{v}_{pr}|^2 + |\mathbf{v}_{rr}|^2 + 2|\mathbf{v}_{qr}|^2), \end{aligned}$$

where subscripts of $\mathbf{v}$ represent partial derivatives, and the weights α and β are parameters to control the strength of the first term and the second term, respectively, in Eq. (4). Equation (4) is defined by following two terms: the first term converges Active grid to a target object, and the second term controls smoothness of the grid shape. E_{img} is a contraction force, which draws Active grid toward a salient region in the three-dimensional data. Therefore, by defining a proper matching energy, region extraction can be realized. In this paper, E_{img} is defined as follows:

$$E_{img}(\mathbf{v}(p, q, r)) = \omega I(p, q, r), \tag{5}$$

where $I(p, q, r)$ is a value of three-dimensional data on the coordinate of lattice point $\mathbf{v}(p, q, r)$, and ω is a parameter that controls the effect of the matching energy.

3. Overview of Pass Region Estimation Based on Active Grid

In this section, we provide an overview of the proposed method. Pass regions are decided on the basis of geometry of player positions. Note that the previous method[12] estimates pass regions based on player positions. Specifically, by using player positions in a single frame, the previous method generates a matching energy for Active Net. However, since pass regions in soccer matches are decided from previous and current player position geometries, the accuracy of the previous method is limited. Therefore, by using player positions of multiple frames, the proposed method attempts to realize more successful pass region estimation. The proposed method is realized by the following two functions.

1) Generation of Three-dimensional Data (Section 4).

Since pass regions are not visible in a soccer video, the energy E^*_{image} cannot be directly defined from visual features of the soccer video. Therefore, the proposed method generates three-dimensional data for the energy E^*_{image} from player positions of multiple frames.

2) Estimation of Pass Regions (Section 5).

The proposed method applies Active grid to three-dimensional data by which generated data from player positions in each frame are connected with the temporal direction. Since the proposed method utilizes player positions of multiple frames by introducing Active grid, more successful pass region estimation is realized.

When we utilize multiple successive results that were obtained by using the previous method[12], pass region estimation from multiple frames can also be realized. However, in the pass region estimation from single frames, errors caused by inaccurate detection of player positions occur independently of the previous and following frames. Therefore, the performance of the pass region estimation is degraded by errors in individual frame. On the other hand, by using multiple frames simultaneously, Active grid converges to correct regions based on the previous and following player positions, since each lattice point of Active grid depends on its neighboring lattice points. Therefore, by using Active grid, more accurate pass region estimation is realized.
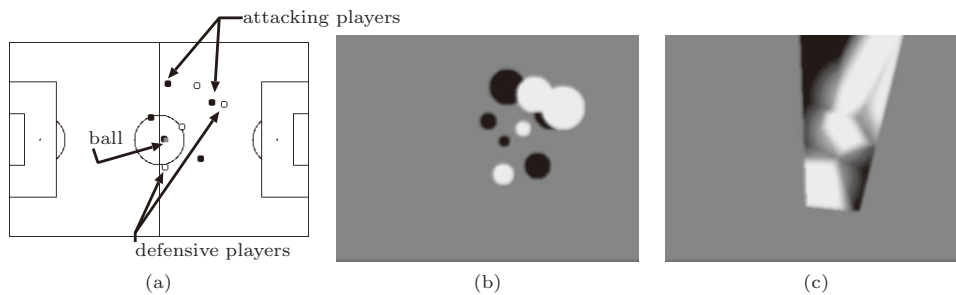
Fig. 2   An example of player positions and examples of generated data: (a) player positions, (b) A_t, (c) S_t. Black color and white color in (b) indicate attacking player regions and defensive player regions, respectively. Black color and white color in (c) indicate spaces and non-spaces, respectively. The trapezoidal region in (c) is the pitch area captured by the camera. The intensity of (b) and (c) expresses the degree of the target region.
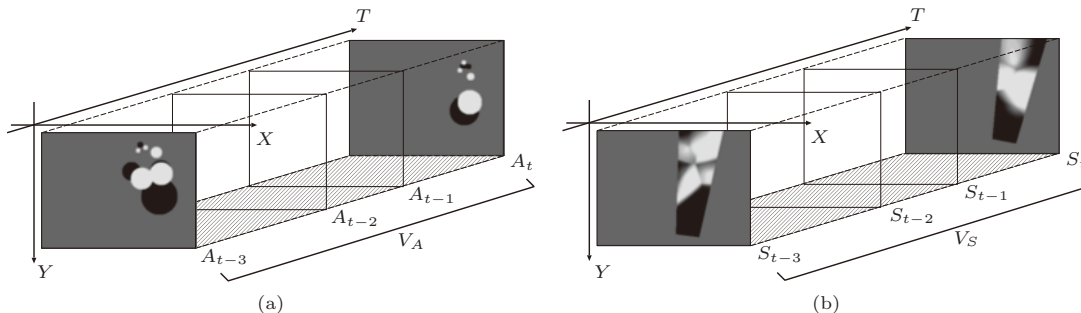
Fig. 3   Examples of three-dimensional data: (a) three-dimensional data V_A, (b) three-dimensional data V_S. The three-dimensional data V_A are generated by connecting A_t with temporal dimension, and the three-dimensional data V_S are generated by connecting S_t with temporal dimension.

4.   Generation of Three-dimensional Data

The generation of three-dimensional data based on player positions of multiple frames is explained in this section. Generally, in Active grid-based region extractions, a matching energy that is lower in a target region than in non-target regions is defined. Then lattice points are converged to target regions and region extraction becomes feasible by noting the density of the lattice points. However, since pass regions are not visible in a soccer video, the energy E^*_{image} cannot be directly defined from visual features of the soccer video. In this paper, we note that a player passes the ball to another player or a space. Thus, by using player positions of successive frames, the proposed method generates three-dimensional data that have low values in attacking player regions and spaces. Furthermore, the proposed method defines the energy E^*_{image} based on the values $I(p, q, r)$ of three-dimensional data. Specifically, the proposed method obtains the energy E^*_{image} by utilizing Eq. (5). Thus, by using Eq. (5) and the three-dimensional data, the proposed method converges Active grid's lattice points to pass regions. In the proposed method, the three-dimensional data are expressed by connecting generated data from player positions in each frame with the temporal direction. The generated data are the energy distribution of each frame

t. There are two kinds of passes in soccer: 1) a pass in which the ball moves to an attacking player and 2) a pass in which the ball moves to an empty space in the pitch. Therefore, in the proposed method, the energy distribution is expressed as A_t and S_t. A_t shows the energy distribution of each frame t for a pass in which the ball moves to an attacking player, and S_t shows the energy distribution of each frame t for a pass in which the ball moves to an empty space in the pitch. A_t and S_t has low values in the attacking player position or the empty space in the pitch. A_t and S_t are generated from player positions. An example of player positions is shown in **Fig. 2**(a). In Fig. 2(a), black points, white points and the gray point are the attacking player positions, defensive player positions and ball position, respectively. Examples of generated A_t and S_t are shown in Figs. 2(b) and (c). The proposed method generates the three-dimensional data V_A and V_S by connecting A_t or S_t with the temporal dimension. Specifically, the three-dimensional data V_A and V_S are generated as shown in **Fig. 3**. These three-dimensional data include the previous and current player positions.

Next, specific procedures for generating A_t and S_t at each frame t are explained.

[**Generation of A_t**]

A pass in which the ball moves to an attacking player is realized between a ball carrier and the attacking player.

Therefore, the proposed method generates A_t that has low values in attacking player regions and high values in defensive player regions. In the generated A_t, these player regions are expressed as round shapes. This is because a receiver can move around to receive the pass. The round shape of each player region is represented by using a radius r that controls the distance that a player can move to receive the pass. The radius r is computed on the basis of distance between the ball carrier and each player. Specifically, the radius r for player P_i in frame t is computed as follows:

$$r(P_i)_t = \frac{d_{t_1} v_h}{v_b}, \qquad (6)$$

where d_{t_1} is the distance between the ball carrier and player P_i, v_h is the velocity of the receiver and v_b is the velocity of the ball. v_h and v_b are defined in the proposed method as fixed values.

[**Generation of S_t**]

A pass in which the ball moves to an empty space in the pitch succeeds when the attacking players move to the space faster than the defensive players. Thus, the proposed method generates S_t in frame t by the following three steps.

(i) **Definition of degree $s(x, y)_t$**

In the proposed method, degree $s(x, y)_t$ of the space at each point (x, y) of the pitch is defined as follows:

$$s(x, y)_t = \lambda d_{t_2}, \qquad (7)$$

where d_{t_2} is the distance between the nearest defensive player and each point on the pitch, and λ is a parameter that controls the degree of the space.

(ii) **Estimation of probability $p(x, y)_t$**

The proposed method estimates the probability $p(x, y)_t$ that an attacking player reaches each point (x, y) of the pitch. Specifically, probability $p(x, y)_t$ is defined on the basis of the time that attacking players can arrive at the target space, where time is defined by the distance between the attacking players and the target space. Thus, in the proposed method, probability $p(x, y)_t$ is estimated as follows:

$$p'(x, y)_t = \gamma(d_{t_d} - d_{t_a}) + 0.5, \qquad (8)$$

$$p(x, y)_t = \begin{cases} p'(x, y)_t & (0 \leq p' \leq 1) \\ 1 & (p' > 1) \\ 0 & (p' < 0) \end{cases} \qquad (9)$$
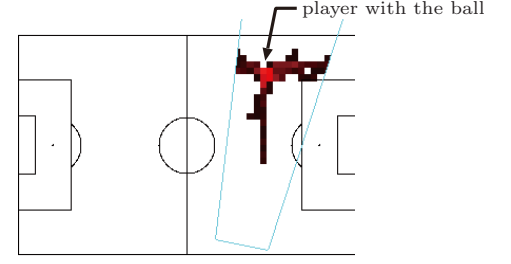
Fig. 4 An example of pass regions, which are indicated by red color.

where d_{t_a} is the distance between the target point and the nearest attacking player, d_{t_d} is the distance between the target space and nearest defensive player, and γ is a parameter that controls the probability.

(iii) **Generation of S_t**

The proposed method generates S_t based on degree $s(x, y)_t$ and probability $p(x, y)_t$. A value $I(x, y)_t$ of S_t in each point of the pitch is defined as follows:

$$I(x, y)_t = I_{max} - p(x, y)_t s(x, y)_t, \qquad (10)$$

where I_{max} is the maximum value of S_t.

5. Estimation of Pass Regions

In this section, the method for estimating pass regions based on player positions of multiple frames is explained. In soccer matches, since the player with the ball passes on a radial course, pass regions become radial shapes whose center is the position of the player with the ball. Therefore, the proposed method defines the shape of Active grid with a cylindrical shape. The center of Active grid in each frame is the position of the player with the ball. In order to estimate pass regions, three-dimensional data are generated as described in the previous section. The three-dimensional data have low values in target player regions and target space regions. Thus, by using a matching energy based on the values of the three-dimensional data, the proposed method converges Active grid lattice points to pass regions. Then, by noting the density of Active grid lattice points, the possibility that a pass succeeds to each region of the pitch that is divided into a lattice shape is calculated. Specifically, the possibility that a pass succeeds is calculated by the following steps.

(i) The proposed method applies Active grids to generated three-dimensional data A_t and S_t.

(ii) In order to calculate the density of Active grids in
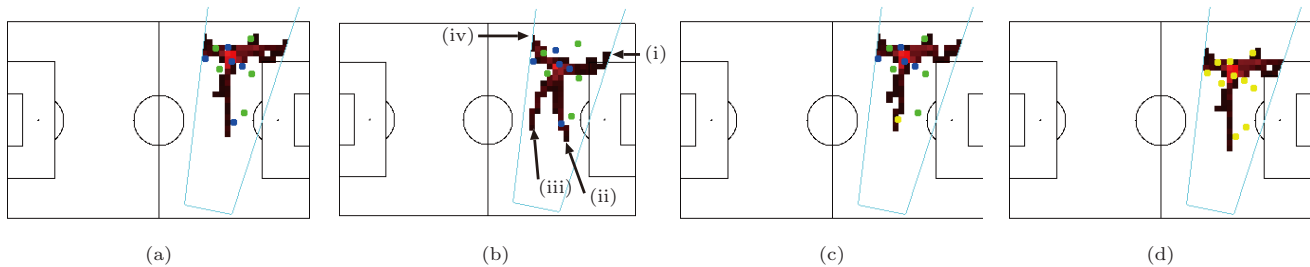
Fig. 5 Pass region estimation results: (a) the proposed method, (b) the previous method[12], (c) case 1 in which inaccurate player positions are included, (d) case 2 in which inaccurate player positions are included. In this figure, red regions are the estimated pass regions, blue points are attacking players, green points are defensive players, and yellow points are players for which positions are inaccurate.

some regions of the pitch $D_{ij}(i = 1, 2, \ldots, n, j = 1, 2, \ldots, m)$, the proposed method counts Active grid lattice points of each region D_{ij} in the target frame.

Consequently, the proposed method obtains the pass regions that have dense Active grids' lattice points. An example of the estimated pass regions is shown in **Fig. 4**. In Fig. 4, red regions express the pass regions and the intensity of the red color is the density of Active grids lattice points.

6.　Experimental Results

In this section, the effectiveness of the proposed method is confirmed by showing results of experiments. For the experiments, we used pass scenes from actual soccer programs in which there were no panning or zooming shots. The player positions can be obtained by using conventional methods[7]~[10]. However, in order to exclude the influence of inaccurate player positions, we manually provided player positions for this experiment. The number of Active grid lattice points is $52 \times 36 \times 3$, with parameters $\lambda = 15$, $\gamma = 0.2$, $\omega = 5.0$, $v_h = 24$[km/h], $v_b = 30$[km/h], and $I_{max} = 255$. In the iteration of Active grid, lattice points that exist on the outer surfaces of the lattice are fixed at all steps. Thus, Active grid covers all of the three-dimensional data at all steps. A part of the results is shown in **Fig. 5**(a), and (b). Results obtained by using the previous method[12] are also shown for comparison. The converged results of Active grid and the estimated pass regions in a target frame are shown in Fig. 5. This figure shows pass regions in red, and the intensity of the red color is the density of Active grid lattice points or the density of Active Net lattice points.

In the experiment for which results are shown in Fig. 5, we used a scene that shows a pass to a striker. In this scene, attacking players attack to the right side

of the result image, the striker is in the center of the right side of the result image, and the player with the ball is on the left side of the striker. As shown in Fig. 5(a), and (b), estimation of the pass regions can be achieved by the proposed method and the previous method. In Fig. 5 (b), a pass region (specifically Fig. 5 (b)-(iii)) in which the player with the ball cannot pass the ball is estimated by the previous method. In the previous method, that pass region is estimated by noting only this frame. Since players move in this scene, the proposed method does not accurately estimate that pass region by using multiple frames of player positions.

Experimental results are shown in Fig. 5(c) and (d) to explain the influence of inaccurate player positions. In these figures, yellow points are players with inaccurately detected positions. As shown in Fig. 5(c) and (d), the accuracy of the player position does not severely affect the final results. Since the proposed method calculates the possibility that a pass succeeds from the player position geometry by using a matching energy, the proposed method realizes robust estimation of pass regions despite the inaccurate detected player positions.

Finally, we calculate Precision and Recall from 147 scenes. These are defined as follows:

$$\text{Precision} = \frac{\text{Num. of correctly estimated regions}}{\text{Num. of all estimated regions}}, (11)$$

$$\text{Recall} = \frac{\text{Num. of correctly estimated regions}}{\text{Num. of passes}}, \quad (12)$$

where the number of correctly estimated regions is calculated by counting the actual passes in the estimated regions. In this experiment, we applied the proposed method and the previous method[12] to some scenes that have different numbers of frames. These experiments are the same as the experiments in the previously reported paper[12]. Precision and Recall are shown in **Tables 1** and **2**. The accuracy was calculated according to the following two conditions. First, cases in which a ball was passed to a player who is out of the camera frame

Table 1 Accuracy of pass region estimation that includes the case of the receiver being out of camera shot.

	The proposed method	The previous method[12]
Recall	0.85	0.85
Precision	0.41	0.27

Table 2 Accuracy of pass region estimation that excludes the case of the receiver being out of camera shot.

	The proposed method	The previous method[12]
Recall	0.89	0.89
Precision	0.50	0.40

were regarded as failures (see Table 1). Furthermore, the above cases were excluded for calculation of Precision and Recall (see Table 2). As shown in Tables 1 and 2, the proposed method realized an improvement of Precision. Since the previous method utilizes only player positions of single frames, pass regions in which the direction of player movement is not considered were estimated. On the other hand, by utilizing the player positions of multiple frames, the proposed method estimates pass regions in which the direction of player movement is considered. According to these results, the proposed method can provide a more accurate result.

7. Conclusion

In this paper, an Active grid-based method for estimating pass regions has been presented. The proposed method realizes robust estimation by generating three-dimensional data based on player positions of multiple frames. The effectiveness of our proposed method was shown by experimental results. The results of the proposed method show various pass courses in a soccer match. Therefore, by using the proposed method, we can obtain information on soccer tactics and players' skills. In the proposed method, the importance of the pass course in terms of soccer tactics was not calculated. Since the pass is a very important element for a goal in soccer matches, the importance must be calculated in future works.

Acknowledgment

References

1) I. Ide, K. Yamamoto, R. Hamada and H. Tanaka: "An Automatic Video Indexing Method Based on Shot Classification," IEICE Trans. Inf. & Syst. Vol. J82-D-II, No.10, pp. 1543-1551 (Oct. 1999)
2) N. Nitta and N. Babaguchi: "Story Segmentation of Broadcasted Sports Videos for Semantic Content Acquisition," IEICE Trans. Inf. & Syst. Vol. J86-D-II, No.8, pp. 1222-1233 (Aug. 2003)
3) H. Miyamori: "Automatic Annotation of Tennis Action for Content-Based Retrieval by Collaborating Audio and Visual Information," IEICE Trans. Inf. & Syst. Vol. J86-D-II, No.4, pp. 511-524 (Apr. 2003)
4) A. Ekin, A.M. Tekalp and R. Mehrotra: "Automatic soccer video analysis and summarization," IEEE Trans. on Image Processing, Vol.12, pp.786-807, (Jul. 2003)
5) B. Li, H. Pan and I. Sezan: "A general framework for sports video summarization with its application to soccer," IEEE Int'l Conf. Acoustics, Speech and Signal Processing, Vol. 3, pp.169-172, (Apr. 2003)
6) R. Leonardi, P. Migliorati and M. Prandini: "Semantic indexing of soccer audio-visual sequences: a multimodal approach based on controlled Markov chains," IEEE Trans. Circuits and Systems For Video Technology, Vol.14, No.5, pp.634-643, (May. 2004)
7) T. Watanabe, M. Haseyama and H. Kitajima: "A soccer field tracking method with wire frame model from TV Images," IEEE Int'l Conf. Image Processing, vol.3, pp.1633-1636, (Oct. 2004)
8) D. Farin, S. Krabbe, P. H. N. de With, and W. Effelsberg: "Robust camera calibration for sport videos using court models," SPIE Storage and Retrieval Methods and Applications for Multimedia 2004, Vol.5307, pp.80-91, (Jan. 2004)
9) Y. Ohno, J. Miura, I. Ide, S. Sakai and H. Tanaka: "Tracking players and estimation of the 3D position of a ball in soccer games," 15th Int'l Conf. on Pattern Recognition, Vol.1, pp. 145-148, (Sep. 2000)
10) M. Xu, J. Orwell, and G. Jones: "Tracking football players with multiple cameras," IEEE Int'l Conf. Image Processing, vol.5, pp.2909-2912, (Oct. 2004)
11) H. Kon, M. Haseyama: "Players Clustering Based on Graph Theory for Tactics Analysis Purpose in Soccer Videos," IEICE Trans. Fund. Vol.E90-A, No.8, pp.1528-1533, (Aug. 2007)
12) S. Takahashi, H. Kon, M. Haseyama: "Active Net-Based Non-interception Region Estimation in Soccer Videos," IEICE Trans. Inf. & Syst. D, vol. J92-D, no. 4, pp. 501-510, (Apr. 2009)
13) A. Doi, S. Fujiwara, K. Matsuda and M. Kameda: "3D Volume Extraction and Mesh Generation Using Energy Minimization Techniques," Int'l Symp. on 3D Data Processing Visualization and Transmission, pp.83–86, (Jun. 2002)

Sho Takahashi received his B.S. and M.S. degrees in Electronics and Information Engineering from Hokkaido University, Japan in 2008 and 2010, respectively. He is currently a Ph.D. student of the Graduate School of Information Science and Technology, Hokkaido University. He is currently a Research Fellow of the Japan Society for the Promotion of Science. His research interest is semantic analysis in videos.
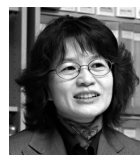
Miki Haseyama received her B.S., M.S. and Ph.D. degrees in Electronics from Hokkaido University, Japan in 1986, 1988 and 1993, respectively. She joined the Graduate School of Information Science and Technology, Hokkaido University as an associate professor in 1994. She was a visiting associate professor of Washington University, USA from 2005 to 2006. She is currently a professor in the Graduate School of Information Science and Technology, Hokkaido University. Her research interests include image and video processing and its development into semantic analysis. She is a member of the IEEE, IEICE, Institute of Image Information and Television Engineers (ITE) and Acoustical Society of Japan (ASJ).