# Development of Video Switching System in Sport Fields

Yoshimasa Umehara[1]*, Wenyuan Jiang[2], Yoshito Nishita[3], Yuhei Yamamoto[4],
Takeshi Naruo[5], Shigenori Tanaka[6], Akira Yokomichi[7], Norio Fujimoto[7],
Toshihiro Akagi[7] and Shingo Hakamata[7]

[1]Faculty of Business Administration, Setsunan University, 17-8 Ikeda-nakamachi, Neyagawa-shi, Osaka 572-8508, Japan
[2]Faculty of Engineering, Osaka Sangyo University, 3-1-1 Nakagaito, Daito-shi, Osaka 574-8530, Japan
[3]Academic Foundations Programs, Kanazawa Institute of Technology, 7-1 Ohgigaoka, Nonoichi-shi, Ishikawa 921-8501, Japan
[4]Faculty of Environmental and Urban Engineering, Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan
[5]Organization for Research and Development of Innovative Science and Technology, Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan
[6]Faculty of Informatics, Kansai University, 2-1-1 Ryozenji-cho, Takatsuki-shi, Osaka 569-1095, Japan
[7]People Software Corporation, 1-15-3, Achi, Kurashiki-shi, Okayama 710-0055, Japan

## Abstract

In Japan, the Sport Basic Plan was formulated in 2012, which mentions not only the development of highly qualified sports instructors but also new perspectives such as the provision of programs in which everyone can enjoy the value of sports together. Against this background, trials of new technologies utilizing the latest ICT equipment such as sensor devices are being made. However, in college sports, where financial resources are often limited, shooting with hand-held video cameras is the most common method, which may cause the manager to overlook important scenes of play depending on his or her skill level. This problem affects not only competitors but also spectators. To solve this problem, we develop a system for capturing video images of the entire field with multiple fixed video cameras so that the system can automatically switch from one to another video of the most appropriate camera for the respective scenes of play. The results of the demonstration experiments in basketball and futsal showed that the switching video of the proposed technology can be utilized for tactical analysis in sports.

*Keywords:* Sports; Multiple cameras; Player tracking; Video switching; Time-series difference.

## 1. Introduction

In Japan, the Sport Basic Plan (Ministry of Education, Culture, Sports, Science and Technology, 2023a) was formulated in 2012, which outlines the key principles of sport promotion for the 10-year period and the policies to be taken over five years. This plan aims at sound development of youth, creation of economic vitality, and improvement in the international status through the promotion of sports. In addition, the third phase of the plan was formulated in 2022, which mentions not only the development of highly qualified sports instructors but also new perspectives such as the provision of programs that allow everyone to enjoy the value of sports together (Ministry of Education, Culture, Sports, Science and Technology, 2023b).

Against this background, trials of new technologies have been made to aim at improving the level of competition as well as entertaining property of sports by utilizing ICT equipment such as sensor devices and video cameras (Tanaka et al., 2020; Haraguchi et al., 2011). However, in college sports, of which financial resources are often limited, capturing videos using a video camera is the most common method. Specifically, a team manager holds a video camera in his or her hand, follows the players and shoot them while adjusting the angle of view at the same time. After

practices or games, the head coach and coaches check the video and analyze players' performance. However, when a game develops rapidly, the manager may be unable to adjust the angle of view in time and miss the ball or players. The manager may even be overlooking important scenes of play depending on the skill level. This problem is not limited only to the teams competing but also affects the spectators similarly. In order to capture important scenes of play in a game, it is necessary to keep holding the camera in position throughout the game and follow the play. This imposes a burden on the cameraman, who can't enjoy the game.

In this study, we develop a system that captures videos of the entire field with multiple fixed video cameras and automatically switches from one to another video of the appropriate camera according to the respective playing scene (hereinafter referred to as "switching"). By developing this system, the cameraman does not have to hold a video camera in his or her hand and follow the playing scene but can always capture the video provided by the system that always switches to the most appropriate camera. This will allow the team managers and spectators to always capture the video of the game that captures the playing scene without fail, contributing to the advancement of play analysis for sports teams and the improvement in the entertainment value of the captured video for spectators.

## 2. Devising Video Switching Algorithms

In this study, we devise a technology that allows automatic switching to the video of the appropriate camera without the need for manual shooting. In specific, fixed multiple video cameras are installed to capture video of a game that covers the entire field. Then, a video is generated by automatically switching the camera to be used based on the position of the ball in the angle of view, the number of people on the court, and the position of the players (hereinafter referred to as "switching video").

The processing flow of the proposed system is shown in Fig. 1. This system consists of a function for constructing ball and person detection models and a function for generating switching video.
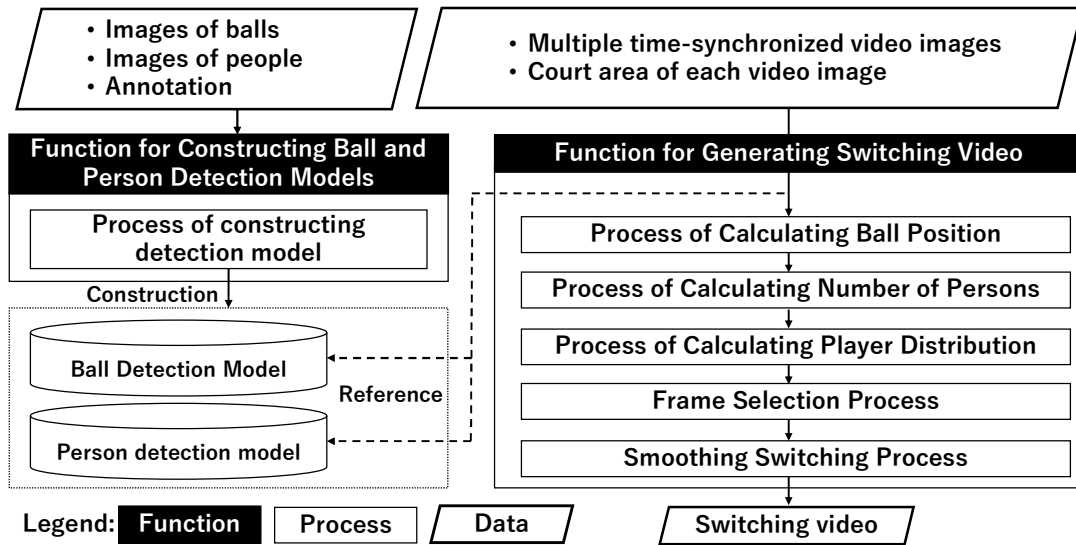


Fig. 1. Process flow

### 2.1. Function for Constructing Ball and Person Detection Models

This function is used to create a set of images in which a ball or a person is captured, as well as the coordinate values of a rectangular region on the image (hereinafter referred to as "annotation"). These are then learned by YOLOv4 (Bochkovskiy et al., 2021), an object detection and recognition algorithm, to construct a ball detection model and a person detection model.

### 2.2. Function for Generating Switching Video

Firstly, this function is used to detect the ball and players from the same frame of the video captured by multiple time-synchronized video cameras. Second, it is used to calculate the positions of the ball and persons and the number of people, then compare them to select the frame in which the play is taking place at the nearest point to the center of the angle of view. Then, the appropriate camera number is chosen from the selected frame. Finally, each frame is combined to generate a switching video.

### 2.3. Process of Calculating Ball Position

This process calculates the distance between the ball and the center of the X coordinate of the angle of view in each frame. First, the ball is detected from the frame using the ball detection model constructed by function for constructing the ball and person detection models. Second, as shown in Fig. 2, the distance between the center of the X-coordinate of the detection frame of the detected ball and the center of the X-coordinate of the angle of view is calculated from the number of pixels. Finally, by repeating this for the number of cameras, the distance between the ball and the center of the angle of view in each frame is extracted.
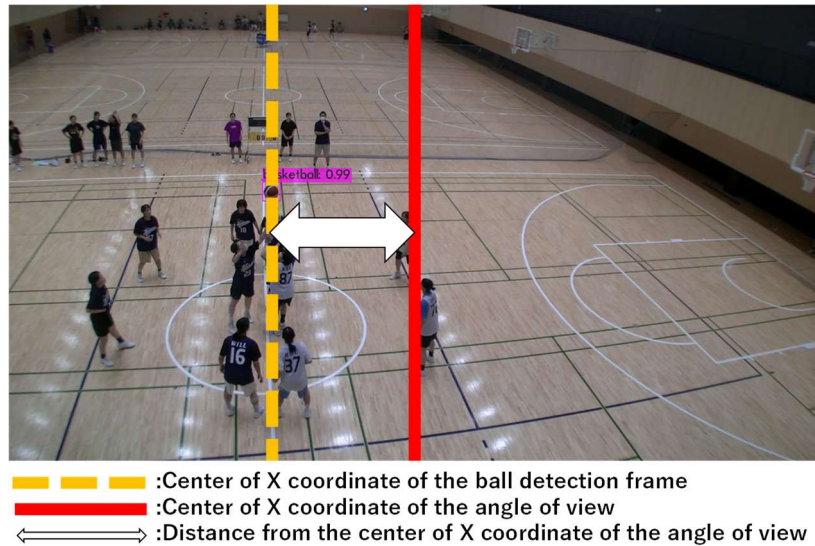


Fig. 2. Process of calculating ball position for basketball

### 2.4. Process of Calculating Number of Persons

This process calculates the number of persons in the court in each frame. First, the area of the court in each video image is manually input. Second, all persons in the image frames are detected using the person detection model. Then, concerning all the persons detected, the bottom of the center of the detection frame of each detected person is considered positional information of the person. Finally, only those persons who are located within the input area of the court are calculated. This is repeated for the number of cameras to obtain the number of persons in each frame.

### 2.5. Process of Calculating Player Distribution

In this process, distribution of players is obtained by calculating the average of the distances from the position of the players within the angle of view to the center of the angle of view. First, just as the process of calculating the ball position, the distance between the person within each frame in the angle of view and the center of X coordinate of the angle of view is calculated. Then, this calculation is repeated for the number of players, and the average is calculated to obtain the average distance from all of the players in the angle of view to the center of angle of view. This makes it possible to select the frame in which players are concentrated at the center of the angle of view in case that there are multiple frames with the same number of players.

### 2.6. Frame Selection Process

This is a process of comparing the outputs from the functions of calculating ball position, the number of persons, and player distribution, then selecting and combining the frame images captured by the appropriate camera to create the switching video. Selection is made in the following order of priority:

1. The camera with the ball being closest to the center,
2. The camera with the largest number of players in the angle of view,
3. The camera with player distribution being closest to the center.

### 2.7. Smoothing Switching Process

In the frame selection process, the switching video is created by comparing the ball, player positions, and number of players for each frame. Therefore, it is considered that the wrong selection of frames due to wrong detection or failure of detection results in inappropriate camera numbers being scattered throughout the frame, causing the video to flicker

as one video switches to another at high speed. In order to suppress this flickering phenomenon, switching is smoothed by decision by majority at a regular frame interval (hereafter referred to as "smoothing interval") in this process.

### 2.8. Trials in Basketball
In this experiment, we verify the switching accuracy of the proposed method by comparing the switching video generated by the proposed method and the existing method. Additionally, we interview the instructors of Kansai University Women's Basketball Club to evaluate the usefulness of the system from the instructor's point of view.

### 2.8.1 Experiment Outline
In this experiment, first, a switching video is manually created to serve as correct-answer data using video taken with a video camera from the position shown in Fig. 3 for approximately 2 minutes from the beginning of the game between the two halves of a team. For a video camera, A SONY / FDR-AX40 capable of shooting at 4K quality at 60 fps was used. Second, the proposed method and the existing method (Iguchi et al., 2002) are applied to the video of the same game above to generate switching video. Finally, the correct-answer data is compared with the switching videos created with the proposed and the existing methods respectively to evaluate the precision, recall, and F-measure.
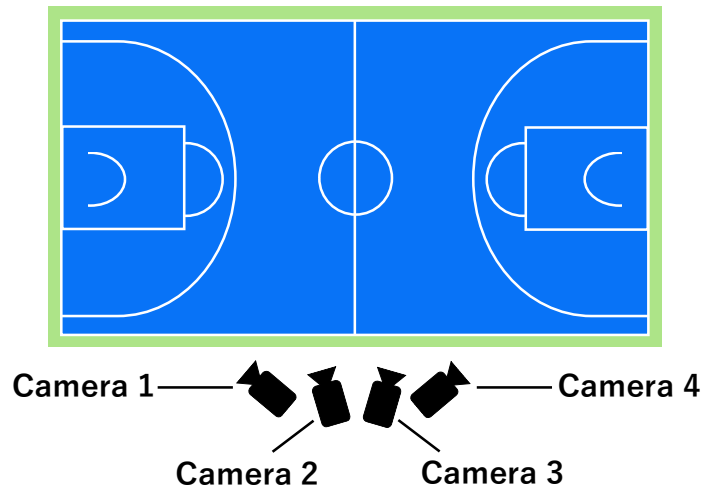


Fig. 3. Position of installing video cameras for basketball

### 2.8.2 Experimental Results and Discussion
The evaluation results of the proposed and existing methods shown in Table 1, and examples of switching video selection shown in Fig. 4. Table 1 shows that from the existing method, precision is 0.41, recall is 0.45, and F-measure is 0.43, indicating that it is difficult to select appropriate frames. On the other hand, from the proposed method, precision is 0.72, recall is 0.71, and F-measure is 0.71, indicating that appropriate frames can be adopted compared with the existing method. Although the precision was 0.72 and there were some cases that switching failed in some frames, it was found that the smoothing approach allowed the flickering phenomenon to be lowered to just one time. The visualization results shown in Fig. 4 also indicate that this method allowed appropriate selection of a camera.

In order to confirm the usefulness of the switching video generated by the proposed method, we interviewed the head coach and manager of the women's team of Kansai University Basketball Club. As a result, they answered that the proposed method is useful for tactical analysis because it enables them to grasp the positions of players involved in pass plays.

Table 1. Evaluation results of proposed and existing methods

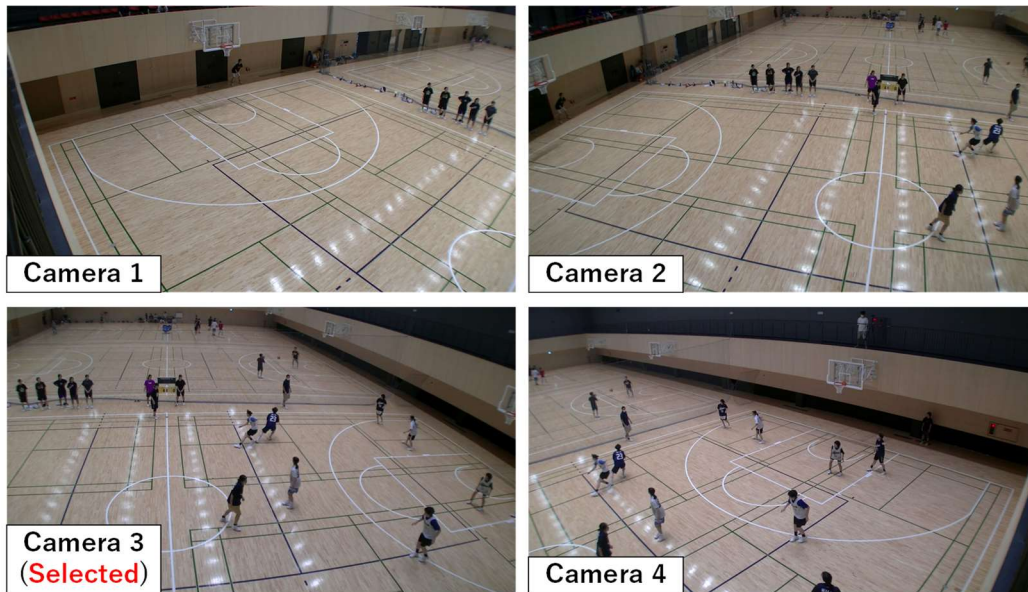| methods | Switching | Flickering phenomenon | Evaluation (Average of 4 cameras) | | |
|---|---|---|---|---|---|
| | | | Precision | Recall | F-measure |
| Manually | 20 | 0 | — | — | — |
| Existing | 387 | 203 | 0.41 | 0.45 | 0.43 |
| Proposed | 25 | 1 | 0.72 | 0.71 | 0.71 |

Fig. 4. Example of switching video selection for basketball

Based on the above results, the switching video of the proposed method is considered useful in the field of sports owing to its applicability to tactical analysis as well as highly entertaining property.

## 3. Proposal of Video Switching System

In this study, we develop a video switching system with the aim of social implementation of the video switching algorithm we devised and contribution to the promotion of college sports. The system inputs video from fixed video cameras shooting at multiple locations, and outputs switching video that switches to the appropriate camera according to the playing scene through the switching video playback function, the function for narrowing down generation range, the function for specifying analysis area, and the function of automatically generating switching video. The number and positions of the video cameras can be freely set.
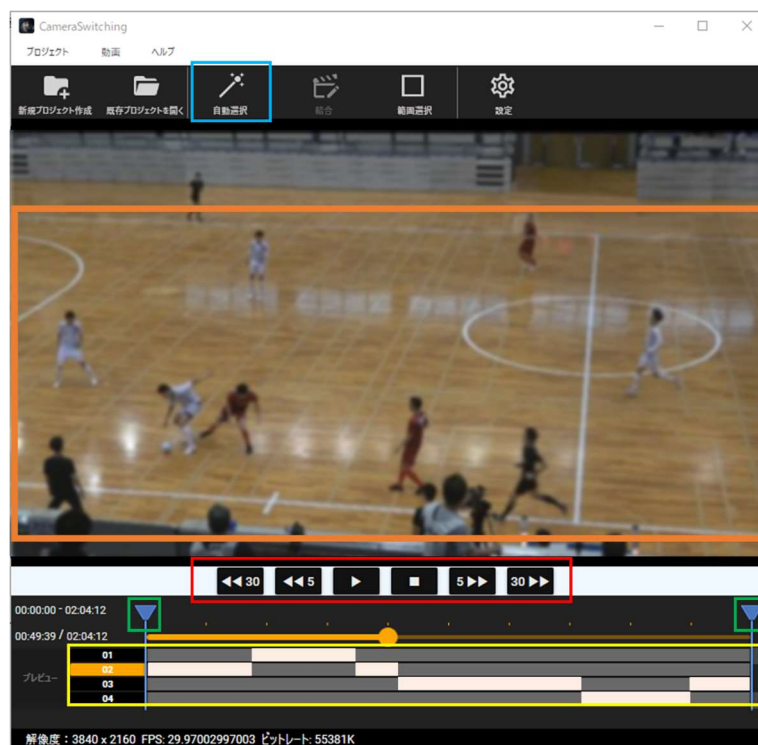


Fig. 5. UI of video switching system

### 3.1. Switching Video Playback Function

As shown in the red frame in Fig. 5, this function allows playback while switching between the videos input from multiple video cameras. It allows controlling playback, stopping, fast forwarding, and reversing of the video. The playback position can be selected arbitrarily using the seek bar. It also allows switching to any video by selecting it from the yellow frame in Fig. 5.

### 3.2. Function for Narrowing Down Generation Range

Video captured by a video camera may include pre- and post-game scenes or video taken while the video camera is being prepared for installation. In addition, it is not necessary to analyze all frames of a video, while it is also possible to generate switching video for only some of the playing scenes. Accordingly, as shown in the green frame in Fig. 5, this function can be used to narrow down the range of automatic generation of switching video using the seek bar.

### 3.3. Function for Specifying Analysis Area

Since there are referees, spectators, etc. near the field, there is a risk of their appearance in the captured video, which may cause failure in appropriate camera selection due to their movements. Therefore, as shown in the orange frame in Fig. 5, this function is used to specify the analysis area on the video to achieve analysis targeting only the players on the field.

### 3.4. Function of Automatically Generating Switching Video

As shown in the blue frame in Fig. 5, only by clicking the button, this function can be used to automatically select the appropriate camera and generate the switching video based on the knowledge of the video switching algorithm we have devised. Emphasizing comfortable operation feel to use as a client system, the following two improvements are added to the video switching algorithm of this system.

● Improved frame selection process

In the frame selection process, the frame to be switched is selected based on the person or ball detected by deep learning. However, the processing cost of deep learning is so high that if it is incorporated into client software, it may cause an enormous amount of waiting time for analysis results, which has a risk of significantly impairing the comfort of the system. Consequently, emphasizing light operation, the camera with the highest number of pixels in the time-series difference is adopted with priority in the frame selection process.

● Analysis Frequency Setup

With video switching algorithms, analysis is performed for all frames. However, there is a problem that the selected frame changes continuously for each frame, causing flickering phenomenon. In fact, flickering phenomenon is observed in a trial experiment in basketball. Therefore, setting the analysis frequency as a parameter, the frame selection process is performed for each analysis frequency rather than for all frames. This suppresses the flickering phenomenon. It also contributes to reduction of the analysis time as the number of frames to be analyzed can be narrowed down.

## 4. Demonstration Experiment

In this experiment, the usefulness of the system is evaluated by applying the video switching system to the video of a futsal game and checking the analysis time and switching accuracy.

### 4.1. Experimental Plan

Two experiments are conducted as follows: "Experiment 1: experiment on the analysis Frequency" and "Experiment 2: experiment on the smoothing interval". Experiment 1 is performed to check the analysis time that varies with the analysis frequency and clarify the optimal analysis frequency. Experiment 2 is performed to check the switching accuracy that varies with the smoothing interval and clarify the optimal smoothing interval.

### 4.2. Experimental Data

In this experiment, an actual futsal game is shot from the camera position shown in Fig. 6. Video cameras to be used are SONY / FDR-AX40 (4K quality, 30 fps) and JVC / Sports Coaching Cam (Full HD quality, 60 fps). The video of 15 minutes from the kickoff is extracted as experimental data.

Fig. 6. Position of installing video cameras for futsal

### 4.3. Experiment 1: Experiment on Analysis Frequency
In Experiment 1, this system is applied to the videos shot by two cameras, one with 4K picture quality and the other with Full HD picture quality, to check the analysis time depending on the analysis frequency.

### 4.3.1 Experimental conditions
The higher the analysis frequency is set, the lighter the operation feeling can be. On the contrary, there is a possibility that the video switching cannot respond to the play transition and the system may continue to select an inappropriate camera. Thus, in this experiment, the analysis frequency is set to 100 ms, 200 ms, 500 ms, 1,000 ms, 1,500 ms, 2,000 ms, and 3,000 ms. The target analysis time is set to 8 hours, assuming that the user can confirm the results of one game shot during the daytime in one night.

### 4.3.2 Experimental results
The results of this experiment shown in Table 2. Table 2 indicates that the analysis time for the 4K camera is 8 times longer than that for the Full HD camera at 100 ms. However, the processing time for the 4K camera at 100 ms was longer than the target value despite a 15-minute video, while at 1,000 ms or less, the processing time was less than one hour, which is the same level as that for the Full HD camera.

Then, based on the results in Table 2, we summarized the analysis time assumed for the whole video of one futsal game (1 hour and 20 minutes) with a 4K camera in Table 3. Table 3 shows that analysis frequencies of 500 ms or less are assumed to take longer than the target value. On the other hand, at 1,000 ms, the analysis time is approximately 4 hours and 30 minutes, which is far below the target value.

Table 2. Analysis time for
4K and Full HD camera

| Frequency | 4K | Full HD |
|---|---|---|
| 100ms | 8h38m26s | 0h56m24s |
| 200ms | 4h22m38s | 0h27m38s |
| 500ms | 1h44m50s | 0h10m38s |
| 1,000ms | 0h52m12s | 0h05m01s |
| 1,500ms | 0h33m55s | 0h03m20s |
| 2,000ms | 0h25m18s | 0h02m30s |
| 3,000ms | 0h16m36s | 0h01m34s |

Table 3. Analysis time assumed
for whole video of futsal game

| Frequency | 4K |
|---|---|
| 100ms | approximately 45 hours |
| 200ms | approximately 23 hours |
| 500ms | approximately 9 hours |
| 1,000ms | approximately 4 hours and 30 minutes |
| 1,500ms | approximately 3 hours |
| 2,000ms | approximately 2 hours and 15 minutes |
| 3,000ms | approximately 1 hours and 30 minutes |

From the above, it is clear that 1,000 ms is the optimum for the analysis interval of this system.

### 4.4. Experiment 2: Experiment on the Smoothing Interval

In this experiment, the switching accuracy is confirmed by comparing the manually created switching video as the correct data with the analysis results of the system. Regarding the switching accuracy, we check the coincidence ratio between the cases when selection of the adjacent camera is not allowed and when it is allowed. This is because it is sometimes difficult to determine the optimal camera between two cameras with partially overlapping video recording areas even with visual check, and consequently two patterns are evaluated in consideration of these cases.

### 4.4.1 Experimental conditions

In this experiment, a camera with Full HD picture quality, which is commonly used, is applied to the system. At this time, the analysis interval is set to 1,000 ms, which was made clear in Experiment 1. Fig. 7 shows the videos from each camera applied to the proposed system and the analysis area specified in the system.
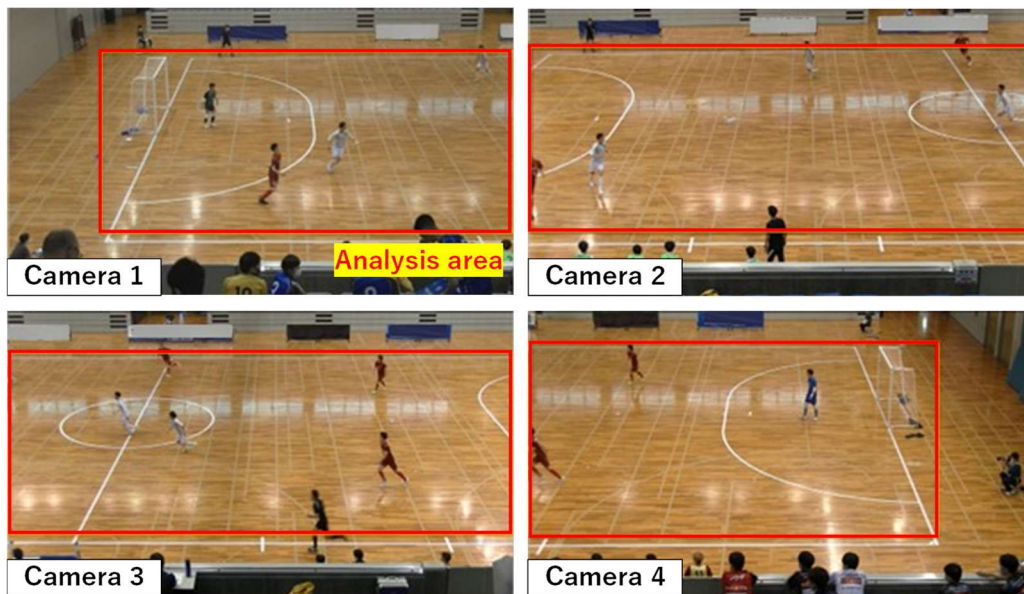


Fig. 7. Video images from each camera and its analysis area

In this experiment, the smoothing intervals are set to 0 (no smoothing), 2, 4, 6, 8, and 10. The larger the smoothing interval is, the more local wrong determination can be decreased; however, it may not be able to deal with rapid transitions in the players' play.
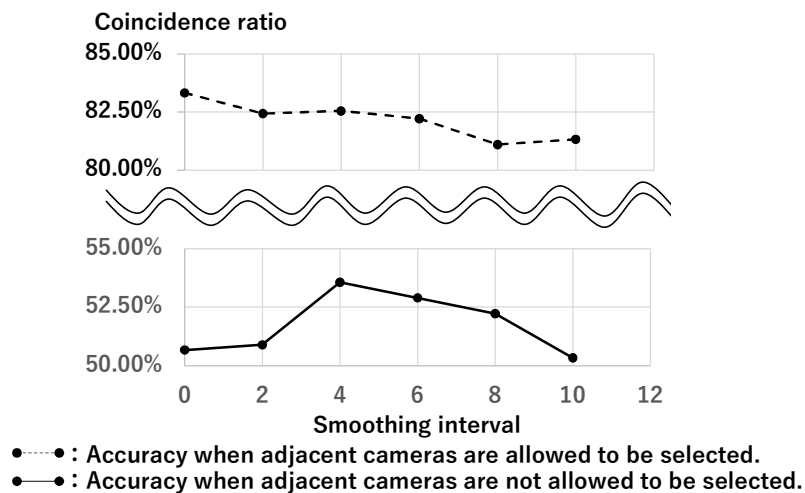


・・・●・・・ : Accuracy when adjacent cameras are allowed to be selected.
——●—— : Accuracy when adjacent cameras are not allowed to be selected.

Fig. 8. Accuracies with and without allowing selection of adjacent cameras

### 4.4.2 Experimental results
The accuracies with and without allowing the selection of adjacent cameras are shown in Fig. 8. Fig. 8 clearly indicates the following items.

● Improved accuracy by way of a smoothing process
In Fig. 8, the coincidence ratio without smoothing is 50.67%. When a smoothing interval is set up, the coincidence ratio converged at the smoothing interval 4 and increased up to 53.56%. This clearly shows the usefulness of the switching smoothing process.

● Successful generation of appropriate switching video
In Fig. 8, the accuracy has improved most at the smoothing interval. However, the coincidence ratio is 53.56%, which means that almost half have selected the wrong camera. However, In Fig. 8, the coincidence ratio is 82.56 at the smoothing interval 4, suggesting that the majority have selected the adjacent camera. The video images also indicated that players' play was captured in both cameras of the correct answer and the adjacent one, which we confirmed were switching video without an uncomfortable feeling.
From the above, it was made clear that the proposed system is capable of generating appropriate switching video.

● Wrong determination due to detection of non-athlete movement
Checking the switching video that had wrong determination except about the adjacent camera, cases of wrong determination were observed here and there due to the detection of movements other than those of players.

First, a case of detecting the movement of a spectator was observed. As shown in Fig. 9, since this experiment was conducted by shooting video from behind the spectators' seats, when a spectator stood up and moved, an occlusion occurred between the court that was the analysis area and the spectator. Since the spectator appeared larger than the player in the distance, the camera in front of the spectator was selected instead of the camera that should have been selected.



Fig. 9. Example of detecting spectator movement (Camera 1)

In some cases, the motions of referees and reserve players were detected. As shown in Fig. 10, the referee and reserve players were moving around the court, generating occlusions with the court, which were wrongly detected as a players' movements.



Fig. 10. Example of detecting movement of referees and reserve players (Camera 2)

Development of Video Switching System in Sport Fields
Yoshimasa Umehara, Wenyuan Jiang, Yoshihito Nishida, Yuhei Yamamoto,
Takeshi Naruo, Shigenori Tanaka, Akira Yokomichi

As described above, there was a tendency for switching accuracy to decline due to the movements of those other than players, captured within the court. This problem can be solved by the following measure. First, for the function for specifying the analysis area of the system, specification with a rectangle was adopted giving priority to ease of manual specification. However, as shown in the orange frame in Fig. 5, in specifying some area with a rectangle, areas other than the court are also contained within the specified area, and the movements of spectators, referees on the sides of the court, and reserve players become factors that reduce the accuracy. Therefore, as shown in Fig. 11, it is possible to suppress movements other than those of players from entering the analysis area, by setting up the vertices of the area arbitrarily rather than using a rectangle and specifying the court area precisely.
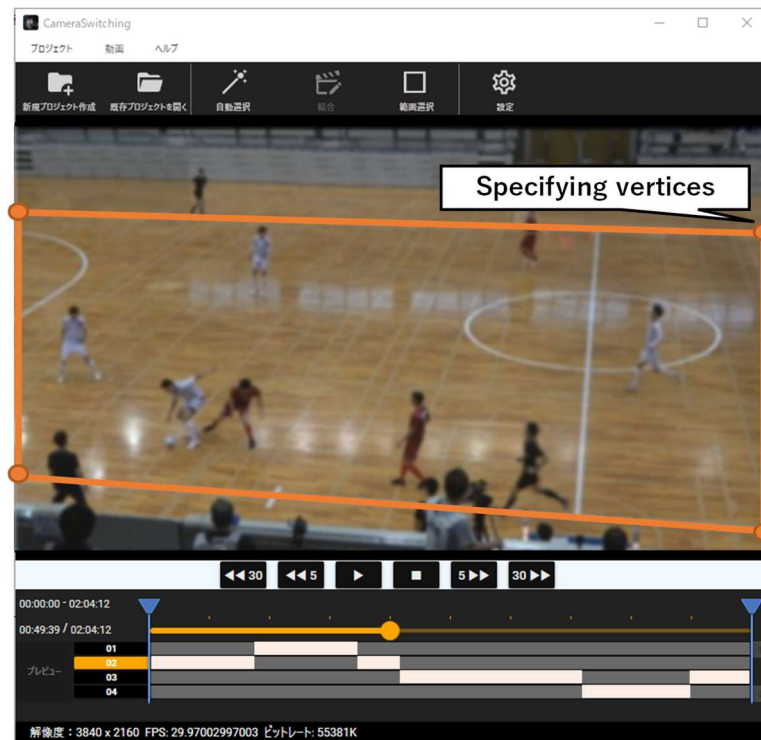


Fig. 11. Function for specifying analysis area by specifying vertices

In addition, the pixels of the time-series difference are labeled. Concerning the difference that spread out to the outside of the court, it is considered the difference caused by those other than players and excluded in the frame selection process. By doing so, we believe this problem can be addressed.

● Applicability to Other Sports
From the above results, the usefulness of this system was confirmed in basketball and futsal. Since this system determines the camera to be selected based on the time-series difference caused by the movements of the players, it is considered applicable to outdoor field sports such as soccer as well. On the other hand, its application to sports where players move simultaneously across the entire court, like volleyball, is considered difficult. However, the applicability of these must be quantitatively verified in the future.

The above results indicate that the system is capable of creating switching video by calculating the time-series difference of respective frames from the video images captured by multiple video cameras. It is clear that this enables creation of the same level of switching video as capturing by manual tracking at a low cost.

5. Conclusions
In this study, we proposed a technology of generating switching video for college sports by automatically selecting an appropriate camera from among the multiple video cameras installed based on the positions of the ball or players. Through experiments and interviews, it was found that the switching video of the proposed technology can be utilized for tactical analysis.

In addition, as a social implementation of the findings of this research and development, we developed a video switching system, establishing an environment in which anyone can easily capture switching video simply by setting

up cameras in the field. In the future, we aim at contributing to the development of the sports field by improving the functions of this system and promoting its widespread use. This paper has been translated and revised based on the technical description (Umehara et al., 2023) published in the Image Lab.

## References

Bochkovskiy, A., Wang, Y. and Liao, M. (2021). YOLOv4: Optimal Speed and Accuracy of Object Detection, arXiv preprint arXiv:2103.13343.

Haraguchi, T., Taki, T. and Hasegawa, J. (2011). "Development of Automated Tracking System with Active Cameras for Figure Skating", *IEEJ Transactions on Industry Applications*, IEEJ, *131*(4), 565-571.

Iguchi, Y., Doi, M., Manabe, Y. and Chihara, K. (2002). "Automatic Multi-Camera Control and Switching for Sports Game Broadcasting Based on Real-Time Video Analysis", *Journal of the Institute of Image Information and Television Engineers*, ITE, *56*(2), 271-279.

Ministry of Education, Culture, Sports, Science and Technology. (2023a). the Sport Basic Plan, https://www.mext.go.jp/component/a_menu/sports/detail/__icsFiles/afieldfile/2012/04/02/1319359_3_1.pdf.

Ministry of Education, Culture, Sports, Science and Technology. (2023b). 3rd Sport Basic Plan (outline), https://www.mext.go.jp/sports/content/000021299_20220316_1.pdf.

Tanaka, C., Yamamoto, Y., Jiang, W., Tanaka, S., Nakamura, K. and Nakajima, S. (2020). "Research for Pass Matchup Analysis Considering Movement Records in American Football", *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, SOFT, *32*(1), 580-589.

Umehara, Y., Jiang, W., Nishita, Y., Yamamoto, Y., Naruo, T., Tanaka, S. and Yokomichi, A. (2023). "Development of Video Switching System for Sports", *Image Laboratory*, Japan Industrial Publishing Co., Ltd, *34*(8), 23-28.