Technical Article

Development of Technology for Generating Panorama and Visualization Using Two-viewpoint Images

Takeshi Naruo [1]*, Yoshito Nishita [2], Yoshimasa Umehara [3], Yuhei Yamamoto [4], Wenyuan Jiang [5], Kenji Nakamura [6], Chihiro Tanaka [7], Kazuma Sakamoto [8] and Shigenori Tanaka [9]

[1]Organization for Research and Development of Innovative Science and Technology, Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan
[2]Academic Foundations Programs, Kanazawa Institute of Technology, 7-1 Ohgigaoka, Nonoichi-shi, Ishikawa 921-8501, Japan
[3]Faculty of Business Administration, Setsunan University, 17-8 Ikeda-nakamachi, Neyagawa-shi, Osaka 572-8508, Japan
[4]Faculty of Environmental and Urban Engineering Department of Civil, Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan
[5]Faculty of Engineering, Osaka Sangyo University, 3-1-1 Nakagaito, Daito-shi, Osaka 574-8530, Japan
[6]Faculty of Information Technology and Social Sciences, Osaka University of Economics, 2-2-8, Osumi, Higashiyodogawa-ku, Osaka 533-8533, Japan
[7]Former Organization for Research and Development of Innovative Science and Technology, Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan
[8]Faculty of Production Systems Engineering and Sciences, Komatsu University, Nu 1-3 Shicho-machi, Komatsu-shi, Ishikawa 923-8511, Japan
[9]Faculty of Informatics, Kansai University, 2-1-1 Ryozenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

Abstract

Research on tracking and performance analysis of athletes using video images has been actively conducted with the aim of improving athletes' competitive performance. However, when filming plays in field sports, it is difficult to capture the entire field with a single camera without filming from a specific point, such as a spectator's seat on the corner side, because the field is long sideways. Even if the entire field is captured, the players at the back of the field appear small, making analysis difficult. Other issues include the fact that since many coaches and analysts film where the play is progressing, it is hard for them to track the ball seamlessly when the position of play changes significantly depending on the position of ball. To solve this problem, we develop in this study a technology to automatically generate panoramic video images that cover the entire field by using two video cameras. Using this technology, we aim to generate panoramic images of the entire field that makes it possible to surely measure and analyze all players and the ball.

Keywords: Image Processing; Panorama; Sports informatics.

1. Introduction

In Japan, the use of ICT in sports has been actively promoted mainly by Ministry of Internal Affairs and Communications, Ministry of Education, Culture, Sports, Science and Technology, and Sports Agency. For example, the "Task Force for Utilization of Sports Data" established in 2017 is studying the utilization of data for health and video distribution, with the aim of making sports more familiar through the utilization of ICT and fostering greater awareness of health among the public as a whole.

Against this social backdrop, there has been an increasing number of cases in the sports field where data is utilized to strengthen players and teams, such as in visualizing player performance and acquiring basic data for tactical planning. These cases of data acquisition are mainly conducted by wearable sensor devices and video analysis that make full use of IoT, and related technological developments (Jiang et al., 2018; Tanaka et al., 2020) are being made on a daily

Publisher's Note: JOURNAL OF DIGITAL LIFE. stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

basis. Jiang et al. (2018) focused on deep learning that can recognize people and animals with high accuracy, and devised a robust player identification and positioning method for use in American football, where occlusions of player images commonly occur. Additionally, Tanaka et al. (2020) proposed a method for tracking players, which is robust to occlusion by images complementing each other using trajectories obtained from videos captured from multiple points. In particular, focusing on tactical analysis etc., when using sensor devices, it is indispensable for players to wear the devices, and although it is possible to acquire data on one's own team, it is difficult to acquire data on the opponent's team. Thus, research and development on data acquisition by video analysis (Jiang et al., 2022; Kataoka et al., 2010; Beetz et al., 2005; D'Orazio et al., 2010; Baca et al., 2009) is being conducted.

On the other hand, although this type of data analysis is used by many professional teams, it has not yet been widespread among local sports teams and club activities in school education due to the high cost involved in its introduction and the need for experts and analysts who operate the equipment. Interviews with athletic teams in the school educational field revealed it is limited to such a level that managers use consumer-use video cameras to film play, and that the data is compiled manually for use in review. For these reasons, the proposers have been conducting research and development based on the belief that technology that enables more advanced analysis using simple conventional filming methods is needed to widely embody application of ICT to sports in society.

In shooting with a conventional consumer video camera, there is a limit to the range of an image that can be captured using a single camera. To capture an object completely with a single camera, it is necessary to maintain a sufficient distance from the object. For example, the standard angle of view of a consumer camera is approximately 50 degrees, and the width of a soccer court (or field) is 120 m. Therefore, capturing the entire court requires a position approximately 30 m away from the soccer court. However, players become smaller and more difficult to analyze as the distance between the field and camera increases. It is also possible to shoot different parts of the field with a plurality of cameras; however, the analysis of the images then needs to be performed with consideration of the positional relationship between the cameras, and the efficiency of the analysis deteriorates.

To solve these problems, a method of capturing a panoramic video image (Naruo et al., 2022) while overlooking the entire field was developed. However, currently, capturing a panoramic video image of the entire field requires the use of dedicated equipment or a camera capable of capturing images in many directions simultaneously, such as an all-sky spherical camera. The former is very costly due to the use of specialized equipment, while the latter is not suitable for analyzing images because its imaging range is the whole and consequently the captured image of players and balls are very small. With the recent development of smartphones, it is possible to capture a panoramic picture, but it is not easy to do so for video images. Based on the above, we develop a technology that allows anyone to easily create panoramic video images by using multiple 4K cameras, which have become available inexpensively in recent years, to capture images in different directions and combine them later.

2. Overview of the Proposed System
This system consists of a panoramic picture generation process and a panoramic video images generation process (see Fig.1). The processing flow of panoramic picture generation consists of 1) reading frame pictures, 2) detecting feature points, 3) estimating and refining camera parameters, 4) applying projection method, 5) exposure compensation, 6) determining stitching lines, and 7) combining left and right images.

The panoramic video images generation process combines the panoramic pictures generated from each frame picture to generate a panoramic video.

Video images captured by two video cameras serve as input data for this camera system. There are no strict settings or values for parameters such as position, height, and camera angle in capturing the input images. The presumption is that one video camera scans the left half of the field and the other scans the right half of the field, and what is common to both video cameras is shown. The output data is a panoramic video image generated by combining two video images.

The process of generating a panoramic image is as follows. After reading the video frames, a frame that generates a panoramic image is read from two images, which comprise its input data. Feature points are detected, and the feature count is calculated from the read left camera and right camera frame images. The feature points of the left and right frame images are matched (i.e., corresponding points detection). Camera parameters are estimated and refined based on the results of the corresponding points detection of the left and right frame images. A cylindrical projection method is applied to the left and right frame images, and the brightness of the images are adjusted for exposure compensation. In superimposing the left and right images, the stitching line is determined by the center of the common portion to be

stitched. In the composition of the left and right images, the left and right images are combined based on the processing results so far to generate a panoramic image.
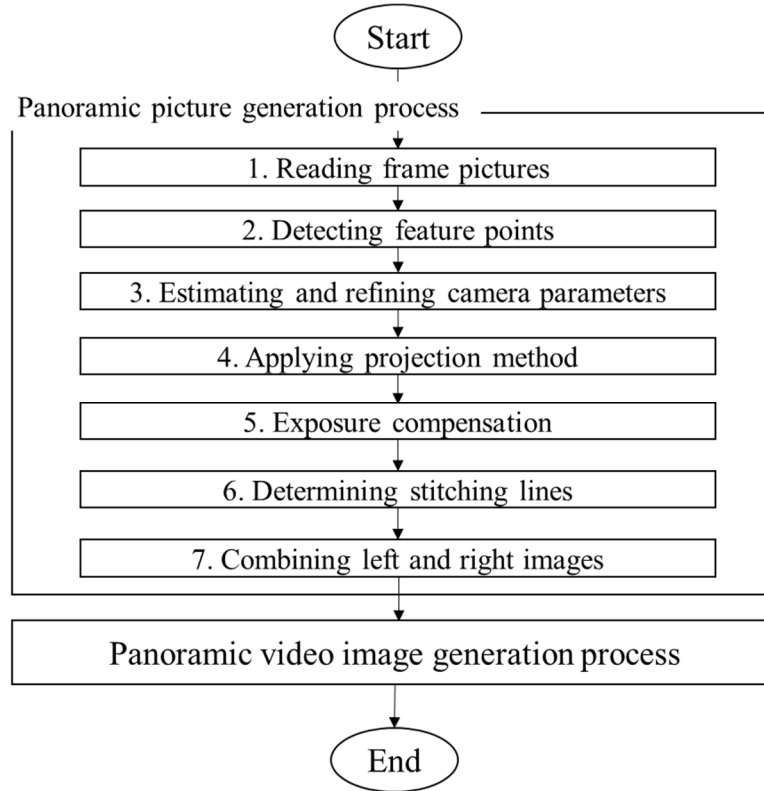

Fig.1. Processing flow of this system

The technology of eliminating the misalignment of the joint in generating panoramic video image is important. Regarding panoramic composition of two video images, composition without the feeling of incompatibility can be achieved by determining the stitching line from each video and blending the boundary of the joint area. Algorithms for determining the stitching line include the Voronoi diagram, dynamic programming, and the graph cut. However, regarding the process of determining the stitching line using these algorithms, the image generation time and the result of the joint portion of the generated panoramic image are different. Therefore, evaluation experiments were conducted for different algorithms used in generating a panoramic video. Table 1 lists the times required to generate a video with each algorithm. Fig. 2 shows clips of the panoramic videos generated by each method. The video time was15 minutes, and a PC with the following configuration was used to generate the videos: CPU-AMD Ryzen 7 4800H 2.9 GHz, and RAM-32 GB. Generation time was converted to the time when using 1 hour 30 minutes of Full HD video for one soccer match. As shown in Table 1, there was not much difference in processing time for the method using the center as the stitching line, the Voronoi diagram, and the dynamic programming method. Nonetheless, the method using the center as the stitching line can generate a panoramic image in the shortest time. Additionally, as shown in Fig. 2, the method of using the center as the stitching line and the Voronoi diagram showed no distortions at the boundary; however, the dynamic programming method outputted an image with a wavy feel near the boundary. Therefore, in this study, we adopted the method using the center as the stitching line from the generation time and generated image.

The OpenCV library is used for panoramic picture generation and panoramic video generation.

Table 1. Panoramic video generation time for each algorithm

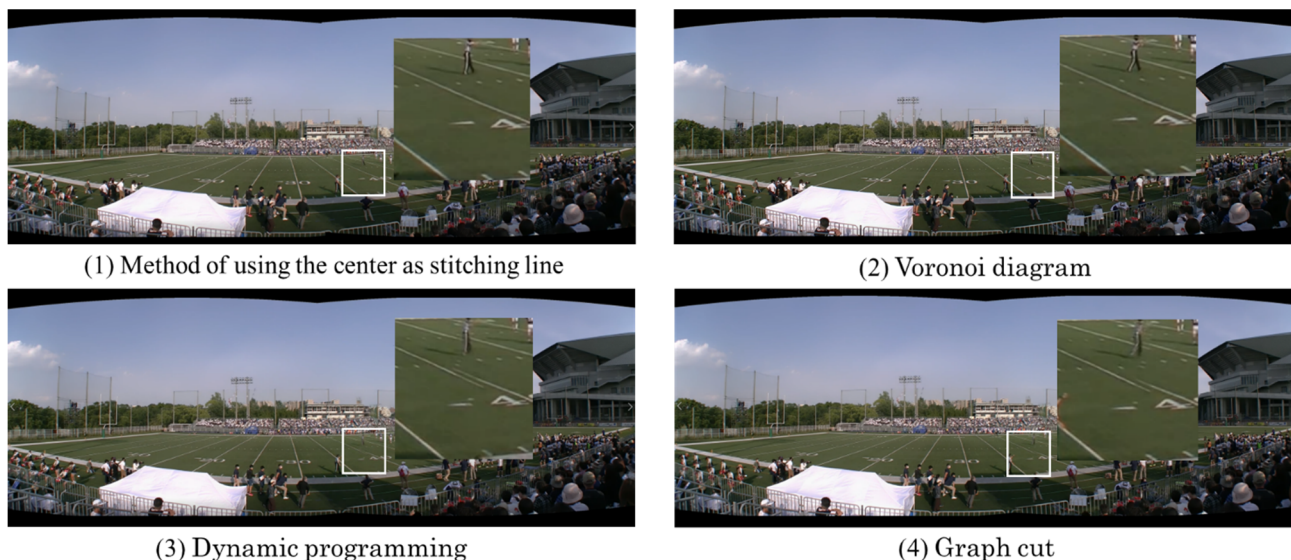| No.	Algorithms	Generation time
1	Method of using the center as stitching line	5 hours 59 minutes
2	Voronoi diagram	6 hours 13 minutes
3	Dynamic programming	7 hours 0 minutes
4	Graph cut	32 hours 5 minutes

(1) Method of using the center as stitching line

(2) Voronoi diagram

(3) Dynamic programming

(4) Graph cut

Fig.2. Panoramic video generation results for each method

3. Demonstration Experiment

In this experiment, two video cameras (SONY FDR-AX40) are used to film the field sports (soccer and American football games) in 4K resolution. A panoramic video is generated from the captured video. Usefulness of the system is examined by checking whether the center line and sidelines are accurately connected in the generated video, and whether the players and the ball moving across the center line, which is the joint area, are reproduced without any misalignment.

3.1. Soccer Game Video

As a soccer game video, we film a soccer team's practice game at the soccer stadium in the Takatsuki campus of Kansai University. Install the video cameras at the position and angle shown in Fig. 3. Also, the video cameras are set up on tripods on the scaffold in front of the left and right goals (see Fig.4). Each of the frame pictures extracted from the captured video is shown in Fig.5. It was confirmed that the center line was captured in both videos. From these two videos, a panoramic video image is generated according to the procedure described above.

The panoramic image generated by this experiment is shown in Fig.6. It was confirmed that a panoramic image can be generated by the processing flow described above. The center line is also joined without any problem. In this experiment, although the sideline became a U-shape because it was shot from the front of the goal, away from the center, a panoramic image closer to a rectangle can be created by using images shot in two directions from near the center line.
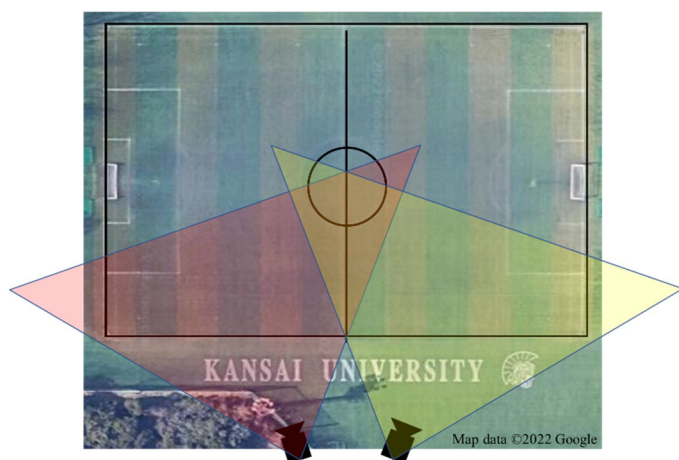

Fig.3. Cameras position and angle
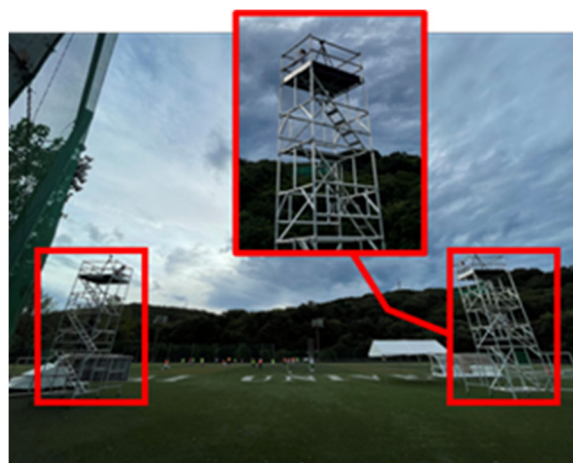(Filming from top of scaffold)

Fig.4. How to install cameras

(1) Image captured by camera on left

(2) Image captured by camera on right

Fig.5. Captured images

Fig.6. Generated panoramic image

3.2. American Football Game Video

Next, we film an American football game held at Nagai Stadium. Install the video cameras at the position and angle shown in Fig. 7. In this case, we used two video images shot from the spectator seats near the center line extension, at a slight angle to the left and right. This picture clearly indicates that the entire field is closer to a rectangle. Fig.8 shows a panoramic image generated from the images of an American football game filmed at Nagai Stadium. In this case, we used two video images shot from the spectator seats near the center line extension, at a slight angle to the left and right. This picture clearly indicates that the entire field is closer to a rectangle.
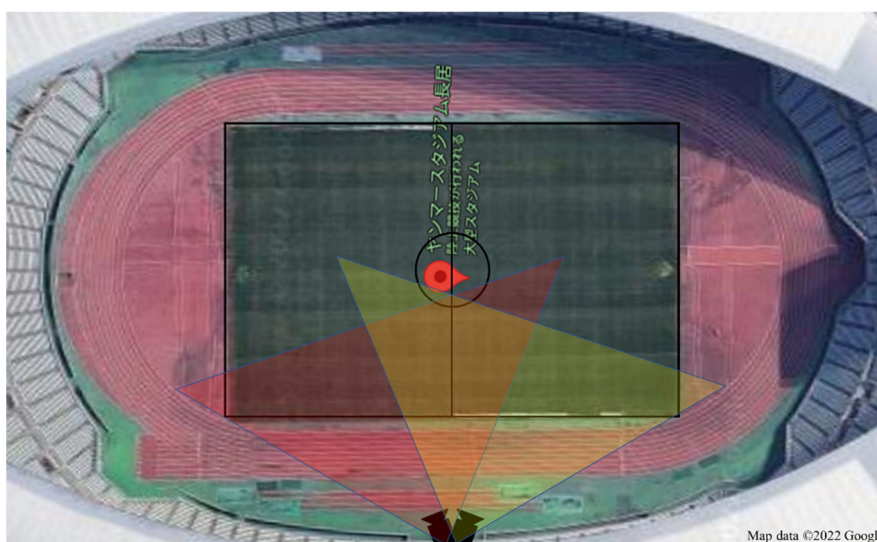

Fig.7. Cameras position and angle (Filming from back row of stadium)

Fig.8. Generated panoramic image (Filming position changed)

3.3. Consideration

In this experiment, video cameras to take pictures were set up on tripods mounted on a platform. The height of the camera was not considered significant, because panoramic images could equally be captured with a tripod not mounted in a high position. It is also assumed that the distance from the sidelines to the camera cannot be secured sufficiently. It is possible to take pictures using three or more cameras that cover the entire field, and generate a panorama from the captured images.

Like this, the quality of generated panoramic images depends on the filming method, such as filming position and angle of view. Points to be remarked include installing both cameras near the center so that they are almost parallel without increasing their angle as mentioned above as well as setting the angle of view so that the field is kept just inside, without including the outside of the filed too much within the filming range. The more the two images contain overlapping images in the middle, the easier it is to match feature points, which is advantageous for panoramic image generation.

Next, the motions of the player straddling or moving across the center line that is the joint, which was a concern in this experiment, was confirmed with the panoramic image. As a result, although deviation by a few frames was seen in some cases, the motions are generally reproduced seamlessly, which is within the range that has no problems in practical use.

4. Discussion

The panoramic image generated by this research and development is displayed using a viewer tool developed by the authors, as shown in Fig.9. This tool is equipped with functions of zooming in or out on the area of interest with a simple operation, easily moving the area to view while zooming in, and quickly playing back the scene by placing a bookmark on the scene of interest.

In addition, the generated panoramic image can cover the entire field within a single image, making it effective for grasping the entire game. When we interviewed a college basketball coach, he gave his opinion saying, "The video image showing the whole is desirable for analysis. We want to see the field as a whole that shows which players made what kind of motions." The panoramic video image generated in this study allows the user to grasp the motions of respective players throughout the game, complying with the coach's wishes.

In addition, since the entire field is filmed, this panoramic image can be used for a variety of analyses. For example, it can also be utilized for tactical analysis of the team with regard to formation. It can also be adapted to the analysis of automatic detection and tracking of players by image processing using deep learning, as conducted in the preceding study.

However, the current method of determining the stitching line with the algorithm also has a problem that the process is time-consuming. In the future, it will be necessary to improve the algorithm to reduce the processing time.
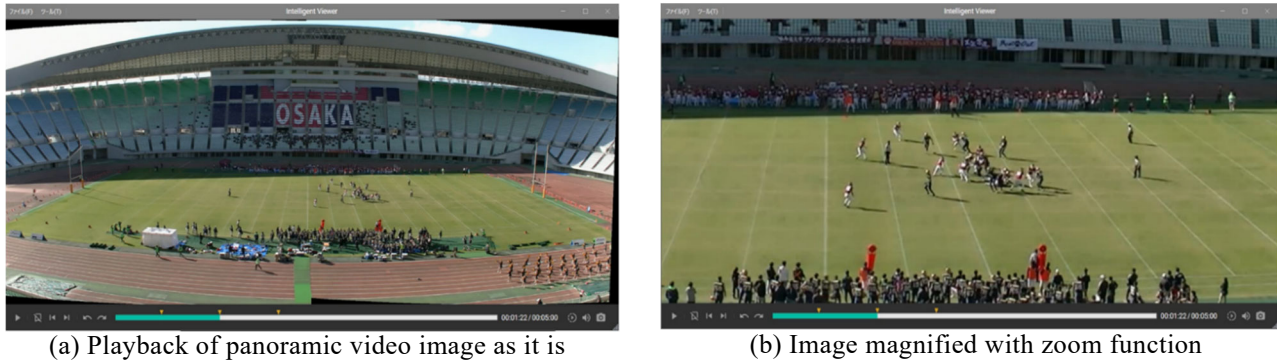
Development of Technology for Generating Panorama and Visualization Using Two-viewpoint Images
Takeshi Naruo, Yoshito Nishita, Yoshimasa Umehara, Yuhei Yamamoto, Wenyuan Jiang, Kenji Nakamura,
Chihiro Tanaka, Kazuma Sakamoto, Shigenori Tanaka

(a) Playback of panoramic video image as it is   (b) Image magnified with zoom function
Fig.9. Playback of generated panoramic video images with viewer tool

5. Conclusions

In this study, we proposed a technology for automatically generating a panoramic video image. The usefulness of the system was verified through a demonstration experiment. As a result, a panoramic image covering the entire field was generated from the video images filmed with two video cameras. It was confirmed that there was no tilt or distortion of the players in the generated image. Furthermore, as the overlap of the centerline at the joint as well as the seamless motion of the players and the ball straddling the centerline of the joint were confirmed without significant deviations, it is considered sufficient for practical use at the actual site. It can also be said that it is a technology that can be utilized for analyzing performance of players and team tactics throughout the field.

Furthermore, we have developed a technique that can be used for analyses of player performance and team tactics over the entire field by automatically generating panoramic images from one direction. However, if a plurality of panoramic images can be automatically generated from images taken from multiple directions, and the images can be switched via the viewer tool, the range of analyses could be further expanded. Therefore, in the future, in addition to a zoom function, we will consider implementing a function to switches video viewpoints via the viewer tool.

References

Jiang, W., Yamamoto, Y., Tanaka, S., Nakamura, K., Tanaka, C. (2018). Research for Identification and Positional Analysis of American Football Players Using Multi-Cameras from Single Viewpoint, *Japan Society of Photogrammetry and Remote Sensing*, 57(5), 198-216. https://doi.org/10.4287/jsprs.57.198

Tanaka, S., Yamamoto, Y. Jiang, W., Nakamura, K., Seo, N. Tanaka, C. (2020). Research for Tracking Athletes Using Video Images from Multiple Viewpoints, *Japan Society for Fuzzy Theory and Intelligent Informatics*, *32*(4), 821-830. https://doi.org/10.3156/jsoft.32.4_821

Jiang, W., Yamamoto, Y., Nakamura, K., Tanaka, C., Tanaka, S., Naruo, T., Matsuo, R. (2022). Construction of Training Data for Deep Learning for Player Detection Using Sports Video, *IEICE (Institute of Electronics, Information and Communication Engineers) Journal D*, *J105-3*(1), 75-88. https://doi.org/10.14923/transinfj.2021SKP0029

Kataoka, H., Aoki, Y. (2010). Robust Football Players Tracking Method for Soccer Scene Analysis, *IEEJ transactions on electronics, information and systems*, 130(11), 2058-2064. https://doi.org/10.1541/ieejeiss.130.2058

Beetz, M., Kirchlechner, B., Lames, M. (2005). Computerized Real-Time Analysis of Football Games, *IEEE Pervasive Computing*, 4(3), 33-39. https://doi.org/10.1109/MPRV.2005.53

D'Orazio, T., Leo, M. (2010). A Review of Vision-Based Systems for Soccer Video Analysis, *Pattern Recognition*, 43(8), 2911-2926. https://doi.org/10.1016/j.patcog.2010.03.009

Baca, A., Dabnichki, P., Heller, M., Kornfeind, P. (2009). Ubiquitous Computing in Sports, *Journal of Sports Sciences*, 27(12), 1335-1346. https://doi.org/10.1080/02640410903277427

Naruo, T., Yamamoto, Y., Tanaka. S., Nishita, Y., Umehara, Y., Jiang, W., Nakamura, K., Sakamoto, K., Tanaka, C. (2022). Development of Panorama Image Generation and Visualization Technology Using Two-viewpoint Video, Image Laboratory, 33(8), 28-32.