Image Analysis Technologies to Realize "Dream Arenas"

● Daisuke Ishii ● Shinichi Akiyama ● Masahiko Toichi ● Takuma Yamamoto

Gearing up for successive major sporting events to be held in Japan around 2020, various initiatives to revitalize sports are accelerating in Japan. Fujitsu pursues the development for practical adaptation of a smart arena solution to support the Japan Basketball Association (JBA) and the Japan Professional Basketball League (B.LEAGUE) with ICT. This solution aims to achieve two objectives: strengthening players and teams, and entertaining spectators. In terms of strengthening players and teams, we have developed a high-precision motion tracking technology in order to capture the positions of moving players in a game, particularly when they are bunched together. In terms of entertaining spectators, we have developed free viewpoint video generation technology that enables the generation of video from any point of view without physical restrictions. For example, it allows for the creation of videos from viewpoint inside the court and 360-degree multi-viewpoint replay videos. This paper presents the benefits of these two types of image analysis technology.

1. Introduction

Gearing up for successive major sporting events to be held in Japan around 2020, various initiatives to revitalize sports are accelerating in Japan. Fujitsu pursues the development for practical adaptation of a smart arena solution to support the Japan Basketball Association (JBA) and the Japan Professional Basketball League (B.LEAGUE) with ICT.[1] This solution has two objectives: strengthening players and teams, and entertaining spectators.

For strengthening players and teams, the goal is to help to train high-performing players, reinforce the team, and enhance their competitiveness. To realize this, there needs to be an effective coaching system based on a scientific approach using statistical data (stats) that summarizes players' and teams' performance information.

Of all the stats, the most important information is the player positional data to analyze the team's formation as well as each player's performance. Because players constantly move during a game, it was difficult to obtain more than approximate information previously due to the nature of capturing the positional data manually. Given this, Fujitsu Laboratories has developed a high-precision motion tracking technology capable of capturing players' positions at all times using a number of cameras installed around the court.

As for the spectator entertainment aspect, we pursue the research and development of free viewpoint video generation technology that enables the generation of realistic video. This technology uses image data taken from several cameras as input to recreate 3D images, allowing for the creation of videos seen from perspectives other than those of the cameras. For example, applying this technology to a sports arena allows for the creation of videos from viewpoints inside the court as well as 360-degree multi-viewpoint replay videos. Thus, viewers can enjoy the sport from physically unrestricted points of view.

This paper describes these high-precision motion tracking and free viewpoint video generation technologies.

2. Motion-tracking technology

This section describes the high-precision motion tracking technology to continuously capture positions of moving players in a game.

2.1 Current state and challenges of motion-tracking technology

There are two types of indoor sports tracking systems. One is the beacon type to monitor the positions of players who carry transmitters on them, and the other is the image analysis type, which calculates the players' positions through video images captured by cameras. Fujitsu's tracking system currently under development has adopted the latter, which requires no equipment to be attached to the players in order not to interfere with their play.

A tracking system based on image analysis technology requires that the players always be in view, and not obstructed by other players or objects. A camera mounted on the ceiling is ideal in this sense, as it commands a bird's-eye view over the entire court and captures all the players within it unobstructed. However, it requires major installation work, which may not be available or practical for some facilities. Thus, we adopted a more accessible method to install cameras on the walls and capture the court from high positions from the side, knowing that players obscuring other players would be a problem. At this stage, we decided to install eight cameras to surround the court, allowing us to capture the players from different angles and minimize the problem of blocked views.

In this method, basically, the image of the court without the players is used as a reference background, and the players are recognized in silhouette, extracted by the differences against the background image. This system can identify individual players by the characteristic features of their figures, and tracking is relatively reliable when there is not much overlapping.

However, in a crowded situation, such as a heated moment in a goal area, players' figures merge to a great extent even if captured by multiple cameras. It is difficult to accurately discern each individual player in this situation, resulting in recognized players being confused, duplicated, or lost. Thus, high-performance tracking technology is needed to enable accurate tracking in crowded situations.

2.2 How to handle crowded situations

As a work-around, we adopted a hybrid system with a main tracking mode based on the silhouette recognition, which monitors levels of player density and switches to an alternative mode specifically developed to process tracking in crowded situations. Details of this alternative mode are described below.

Although much of players' figures may overlap in a crowd, their heads are relatively better exposed to view. Thus, we chose to capture and track the movements of players' heads during crowded situations. However, heads are much smaller than entire figures, and they contain fewer features that are easily identified. Simply changing the focus targets does not offer reliable tracking. When players gather during crowded situations, their heads are also close to one another. We needed a means to ensure reliable tracking while removing the interference of other players in close proximity.

2.3 High-precision tracking technology with camera selection capability

We have developed high-precision tracking technology with camera selection capability for processing crowded situations. This technology maximizes the advantages of capturing images from a number of angles, as it selects those images that contain less overlapping and thus more useful for the image analysis.

More specifically, we have developed two processing technologies. One is tracking order scheduling technology, which determines which players can be tracked most reliably. The other is multi-camera priority evaluation, which selects the images to process according to the scores assigned in terms of the clarity of heads in the images captured by multiple cameras.

1) Tracking order scheduling technology

Because it is easier to reliably track players who are constantly captured in the video image, the system determines in advance which players are more easily identifiable by degrees of hiding and puts them in order of processing. The head images that have been processed are removed from the subsequent data processes for identifying other players. This procedure allows the technology to narrow down the focus area for identifying the head of a player whose image is mostly hidden by other players and therefore difficult to detect and track reliably.

The order of priority is determined based on two parameters: the level at which the player is hidden from the camera view, and the level of resemblance between the target player and the surrounding images. By predicting each player's movements based on past tracking data, the system estimates their positions against other players and the background, according to which the overall priority order is determined.

2) Multi-camera priority evaluation technology

Individual players also look different when seen from different angles. The system selects the cameras that have better image disposition and removes others to execute 3D position detection for enhanced reliability.

The multi-camera priority evaluation uses the same parameters as those described in 1) above, excludes those cameras that did not capture the target player clearly, and assigns weight on each camera according to the parameters. This dynamic camera selection and weighting system is illustrated in **Figure 1**. During detection, by evaluating the unique characteristics of a head in each image of the selected cameras and comprehensively evaluating based on the spatial constraints constituted by the geometric relationship between the cameras, the player's position in this space is determined.

2.4 Verification of tracking performance

We conducted a performance test of the system with the high-precision tracking technology with camera selection capability, which we described in the previous subsection. **Figure 2** shows an example of the results from processing a crowded scene. On the left, the 2D rendering shows the relative positions of the players in the court and the cameras that captured them. The video image captured at the indicated time is shown on the right. We can see that different cameras are selected at different times according to the situations of the players.

In this test, we used eight cameras and recorded a basketball game. We extracted 18 segments, each consisting of an offensive unit time of 24-second interval (shot clock) and totaling 420 seconds, for the test. The ratio of crowded situations in this sample was 60%.

For the experiment, each sequence was divided into 10-second fragments and verified for successful tracking of particular players. **Figure 3** shows the test results. The developed technology has made significant improvement on the tracking performance in crowded situations, achieving 94% overall tracking accuracy, with an average improvement of 31%. From these results, we have verified that reliable tracking can be achieved with the high-precision tracking technology with camera selection capability even when players crowd together in a goal area.

2.5 Summary

Aiming to realize automated stats generation

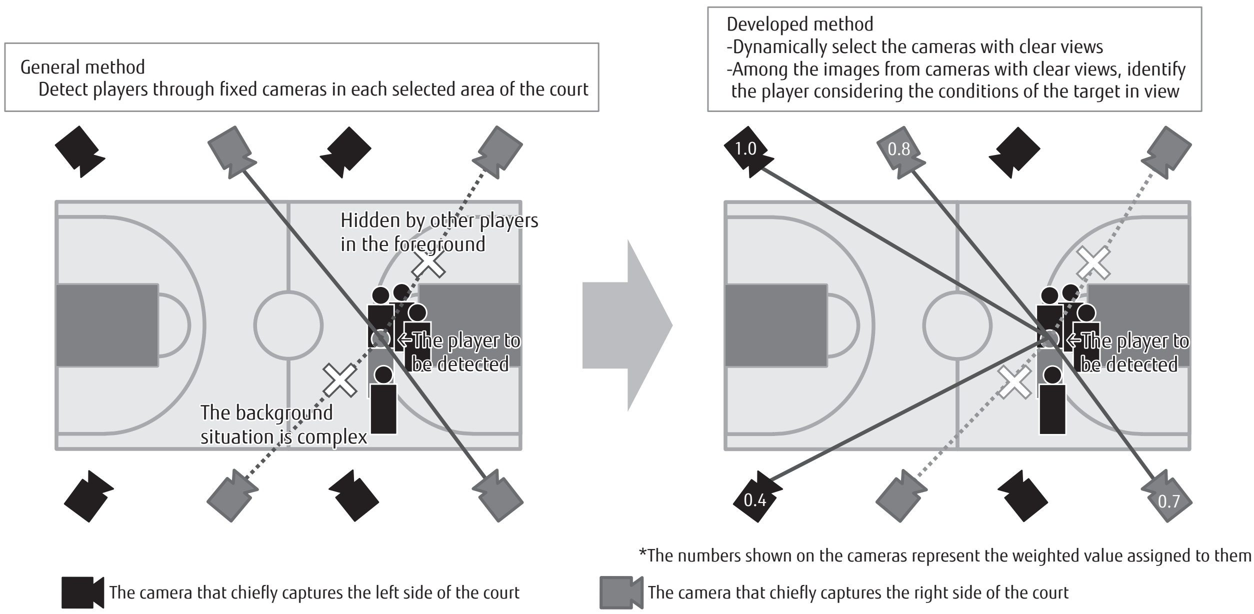

Figure 1
Dynamic camera selection and weighting for each player.

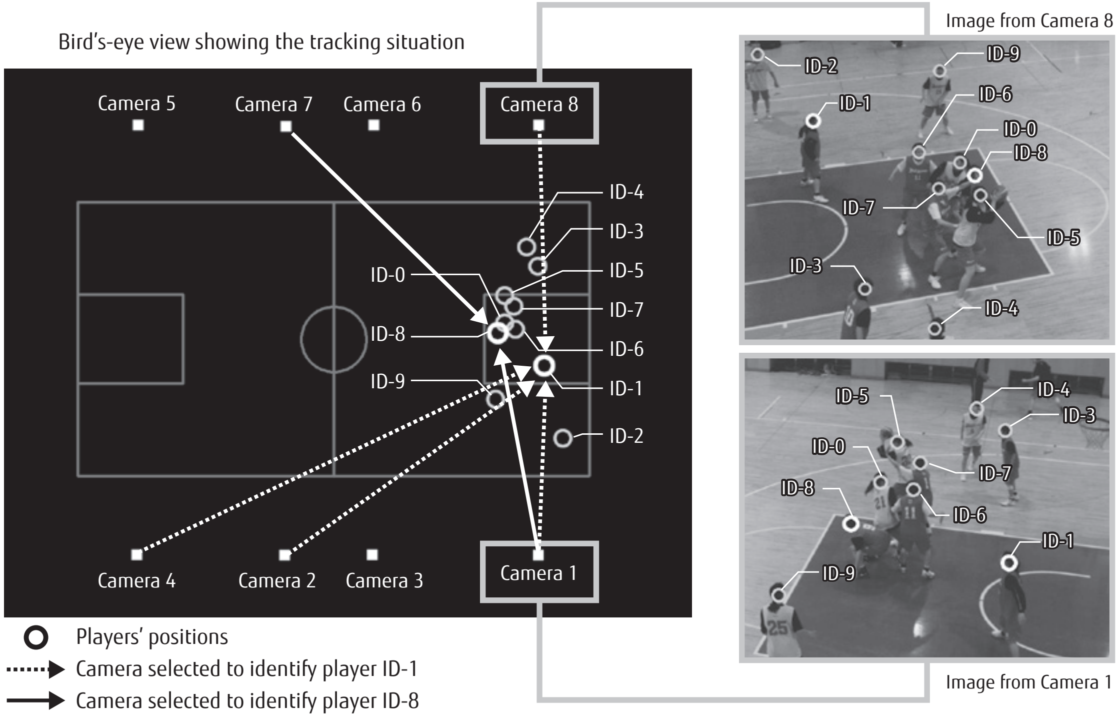

Figure 2
Example of camera-selection and tracking in crowded situation.

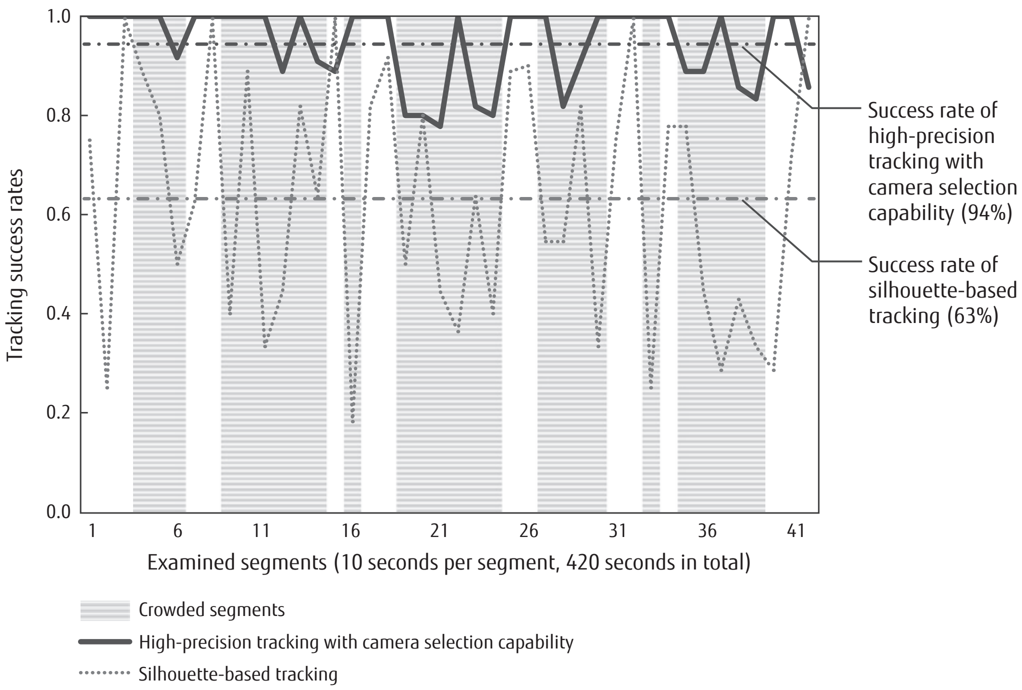

Figure 3
Results of performance evaluation of the high-precision tracking technology with camera selection capability.

from sports plays, we have developed high-precision tracking technology with camera selection capability to address the challenge of reliably tracking of players in crowded situations. As this technology allows Fujitsu to capture players' locations with high accuracy when they are bunched together, we now have insight into realizing an automated motion tracking system that does not confuse target figures. Future tasks include the development of an automatic stats generation system and feature enhancement for practical operation.

3. Free viewpoint video generation technology

This section describes the technology that allows the rendering of images from any chosen viewpoint on the court.

3.1 Trends in free viewpoint video generation technology

Research into free viewpoint video generation started in the 1990s, and it was first put into practical use in effects for films[2] and advertisements. Today, it is used in replay videos during live sports broadcasting.[3), 4)]

The major methods known today include: creating a multi-viewpoint sequence by combining images from frames captured with different cameras,[5] and generating 3D images of a target player from captured images to digitally render plays from any desired perspective.[6), 7)] The former technique is relatively easy to realize and renders high-definition pictures, but it requires many cameras in order to obtain a smooth viewpoint-shifting effect, and the limited perspective flexibility is a problem. The latter has the advantage of greater perspective flexibility and is highly compatible with CG synthesizing, as it digitally reconstructs the target images from 3D data. However, the clear separation of figures from the background is important for attaining high image quality. It is essential to establish a method that reliably achieves the separation in any environmental conditions and with any background images if it is to be introduced to various sporting venues.

At Fujitsu Laboratories, we are developing a system based on the latter technique, as it has great potential for various services to which 3D information analysis may be applied.

3.2 Process flow of free viewpoint video generation

The free viewpoint video generation technology being developed at Fujitsu Laboratories employs a number of cameras installed on the walls or behind the gallery benches so that the cameras surround the court (**Figure 4**). This configuration enables the capturing of all images and objects within the court, including the players, referees, the ball, and baskets, with almost all of the cameras installed. On the other hand, the background images, such as gallery benches, are captured only by the cameras that are set opposite of them.

As the number of cameras and their directions differ, we have adopted a method to apply a 3D reconstruction process individually to each target after the separation of images. The objects in the foreground (players, the ball, etc.) are processed using a volume intersection method, which can achieve fairly detailed 3D reconstruction. The background image is three-dimensionally reconstructed using an applied version of a stereo vision method.

3.3 Issues with conventional foreground/background separation technology

For free viewpoint video generation, the accuracy of the separation of foreground from background has

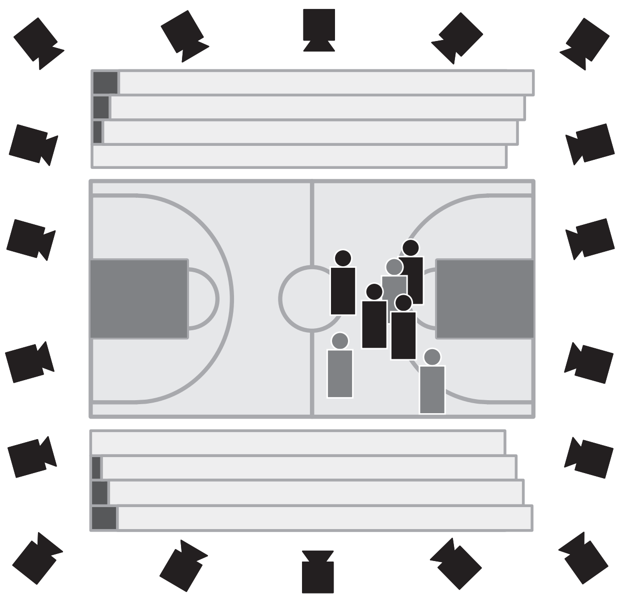

Figure 4
Camera positions for the free viewpoint video generation system.

significant impact on the final output because the foreground mask (the area of the foreground object) is used in later stages of image processing. For this reason, the enhancement of object extraction is indispensable for achieving high-quality free viewpoint video generation.

Foreground/background separation is an important basic technology in video image processing, and there are several established methods. For example, the chroma keying method[8] involves filming objects in front of a single-hued sheet, usually in green or blue, and removing that color in the background to extract the foreground. Background subtraction[8] is a method to compare two images, one with only the background and the other with target objects, extracting the foreground by the difference between the two.

Chroma keying is employed, for example, at TV studios when they want to superimpose images of people or objects that do not exist there to create a synthetic image. Background subtraction is used when detecting moving objects in visual surveillance. These techniques for foreground/background separation pose the following two issues in terms of their application to the free viewpoint video generation in sporting events.

First of all, chroma keying cannot be used in soccer (football) or basketball games because the games take place in specified playing fields where unicolor background sheets cannot be installed. Additionally, in the case of basketball, team uniforms often have colors similar to the background floor or walls, which makes it difficult to adopt background subtraction.

3.4 Development of foreground extraction method with applied deep neural network

Given these issues, we have decided to employ deep neural network (DNN),[9] which made remarkable advancement in AI technology in recent years, to develop a highly-accurate foreground extraction method. More specifically, our system utilizes convolutional neural networks,[9] a version of DNN designed for image processing, to discriminate the foreground elements by pixels. As a result, the area depicting a human figure can be extracted with high accuracy. Furthermore, with requirements to make the foreground mask smooth in the rendered space, graph cuts[10] are applied for global optimization.

As CNN can employ several filters, it can extract various characteristics of the target object, such as the edges and overall shapes as well as colors. This feature enables CNN to clearly discern the object in the foreground even if it has colors similar to the background. Applying the graph cuts to optimize the entire image helps to remove noise and extract a high-quality mask.

3.5 Foreground/background separation performance assessment

Supposing the application of the system to a basketball game, we installed eighteen 4K camera in a gymnasium to make a sample video for a performance assessment. The cameras were equipped with a global shutter feature so that all cameras simultaneously captured the same moment. The cameras and the server are connected with fiber optic cables, and the recorded images are collected on the server for processing.

In real B.LEAGUE games, the restricted areas around the goal posts are sometimes painted in the team color. In order to recreate this situation, we had our mock teams wear uniforms of similar colors to that of the paint area, and we recorded the game.

Based on this recorded sample video, we compared the developed foreground extraction method with a conventional method using background subtraction[8] based on color distribution differences. The reference image and foreground masks are shown in **Figure 5 (a)** and **Figure 5 (b)** respectively. The conventional method partially failed to capture the player's uniform, which is the same color (red) as the background. In contrast, the developed method fully captured the player.

We then processed these foreground masks using the volume intersection technique to recreate 3D images and rendered them from a perspective different from the camera viewpoints, which is shown in **Figure 5 (c)**. The image using the conventional method contains artifacts[note)] in the player in the foreground caused by a defect in the foreground mask. The player in the back is also missing the left leg of his pants because they were a similar color to the background. The developed method, by comparison, has rendered a high-quality image with very few artifacts.

note) Errors and distortions resulted from the image processing.

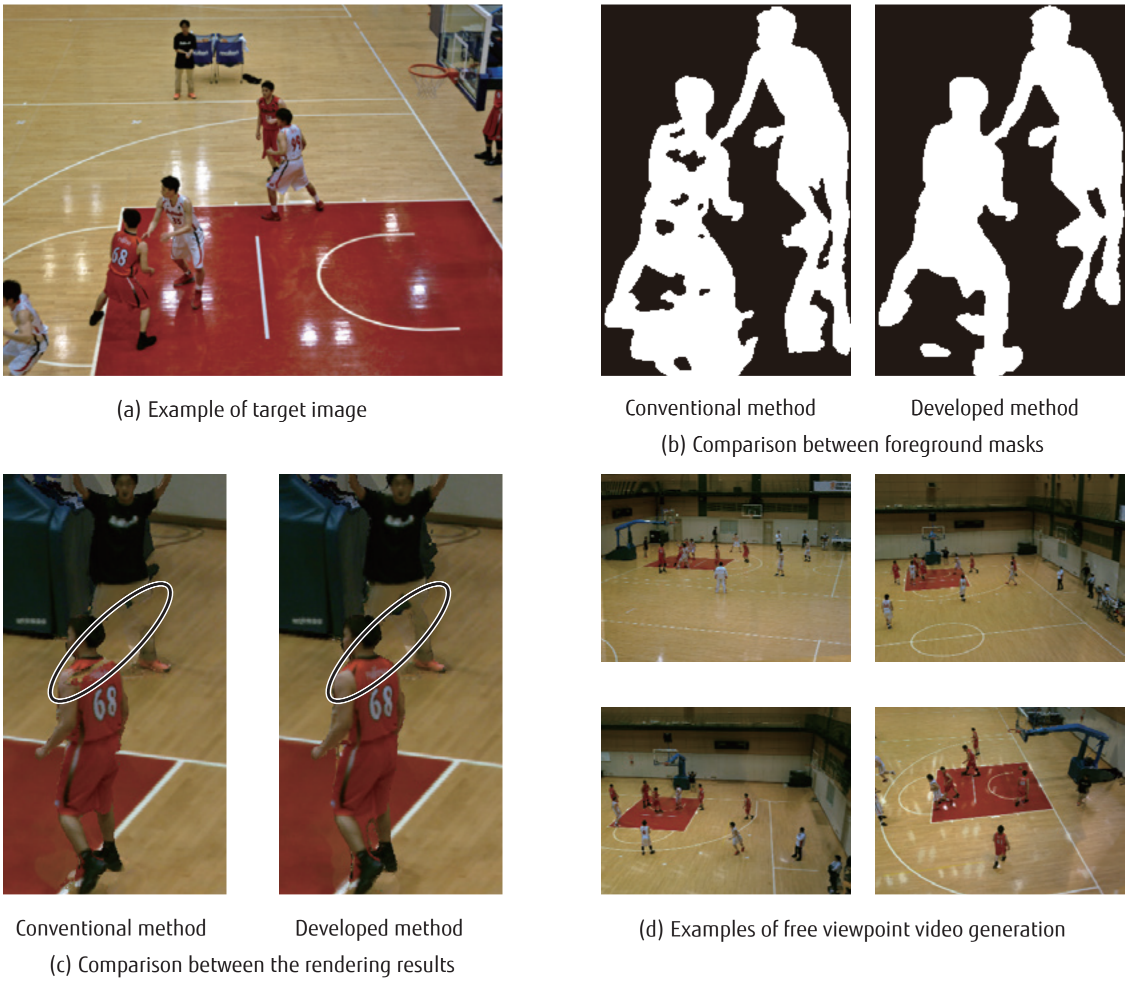

Figure 5
Effect of the developed foreground extraction method.

We then compared the foreground masks extracted through the foreground/background separation methods against a correct mask, prepared manually to perfection, by F-score (harmonic mean of precision and recall). The results were thus 0.67 for the conventional method while the developed method improved it to 0.77, which confirmed the improvement quantitatively. **Figure 5 (d)** shows the images processed by the developed method and rendered from various viewpoints different from the camera perspectives.

3.6 Summary

We have developed a technology that highly enhances the precision of foreground extraction for free viewpoint video generation by applying the DNN-based extraction method. This technology enables free viewpoint video generation reliably even in situations where several players are bunched together or the court is painted in the team color.

As the rendering process involves the reconstruction of spatial 3D data, it is highly compatible with CG and virtual reality (VR). We are also considering the incorporation of head-mounted displays, which would facilitate a new spectating experience—to enjoy a game from inside the game court.

4. Conclusion

In this paper, we described motion tracking technology and free viewpoint video generation technology, which have been developed to realize the "Dream Arena" promoted by B.LEAGUE. The smart arena solution introduced in this paper addresses the challenges facing basketball to popularize the sport, and it enables freely tagging statistical data to space, time, and video images using the image analysis technology of Fujitsu and Fujitsu Laboratories.

Our future tasks will involve the application of both the motion tracking technology and free viewpoint video generation technology to real sports games where players engage in real actions. We will also look into the possibility of offering a new spectating experience that integrates an overall viewpoint with realistic excitement as spectators enjoy viewing the game while checking out individual player's play data. In this way, we strive to realize a Dream Arena that will satisfy the intellectual appetite for absorbing the games and also offers ways in which each play can be easily appreciated.

References

1) Fujitsu: Fujitsu ICT to Support the Future of Basketball in Japan.
http://www.fujitsu.com/global/about/resources/news/press-releases/2016/0921-01.html
2) Matrix.
https://www.warnerbros.com/matrix/
3) KDDI: au BLUE CHALLENGE–au Future Stadium (in Japanese).
https://www.au.com/pr/soccer/
4) Intel: 360-Degree Replay Technology Brings Basketball Fans into the Future.
https://iq.intel.com/360-degree-replay-technology-brings-fans-into-the-future-of-sports/
5) 4DReplay.
http://4dreplay.com/
6) T. Kaneda, et al.: Virtualized Reality: Constructing Time-Varying Virtual Worlds From Real World Events. Proc. of IEEE Visualization, pp. 277–283 (October 1997).
7) W. Matsuike, et al.: Image Based Visual Hulls. ACM SIGGRAPH2000, pp. 369–374 (2000).
8) M. Okutomi, ed.: Digital Image Processing [Revised edition]. Computer Graphic Arts Society (2015).
9) I. Goodfellow, et al.: Deep learning. MIT Press (2016).
10) Y. Y. Boykov, et al.: Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. ICCV 2001 (2001).

Daisuke Ishii
Fujitsu Laboratories Ltd.
Dr. Ishii is currently engaged in research on image recognition and comprehension.

Shinichi Akiyama
Fujitsu Ltd.
Mr. Akiyama is currently engaged in the project athlete sensing solution.

Masahiko Toichi
Fujitsu Laboratories Ltd.
Mr. Toichi is currently engaged in the R&D of image processing and 3D spatial recognition.

Takuma Yamamoto
Fujitsu Laboratories Ltd.
Mr. Yamamoto is currently engaged in computer vision research.