As 2020 approaches, the need for video analysis technology for analyzing sports events is increasing. Analysis results are used for various purposes in sports, but, in particular, new visual presentations that deepen viewer interest and understanding of the competition in sports programming are desirable. In sports programs, the interest of viewers is focused on the players and the ball. Using technology for detecting and tracking these objects from camera images makes it possible to visualize their movements so that viewers can understand them easily. Such technology has been studied for quite a long time, and it has been used in the refereeing of formal competitions in recent years. An original object tracking system that we have developed has been used in sports programming. Here, we describe the progress that has been made in object tracking technology based on video analysis and present examples of its application to sports programming.

1. Introduction

With 2020 in view, the expectations for technology that facilitates viewer understanding of sports competitions presented in video are rising. There have been attempts to process and analyze sports video, and techniques have been developed for judging player skill, judging the team play situation, and using computer graphics technology for visualization effects in broadcast video\(^1\)\(^2\). Video analysis is also used in professional sports for various purposes, including strengthening the technical capabilities of players, analyzing team data, planning countermeasures and tactics for use against opposing players and teams, and preventing bad judgment calls. In tennis and soccer, video analysis technology is already being used in official matches\(^3\)\(^4\).
The hierarchy of information that can be acquired by sports video analysis is illustrated in Fig. 1, where we can see how the quality of information increases in stages from physical quantities to semantic content by applying advanced analysis techniques, using video sequences as the input. First, the video sequence is analyzed to recognize players, balls, and other objects and their positions and names are acquired. These items of information can be used to measure the amount and speed of movement of each player and to visualize the movement path. Then, the detailed movements of each object and the positional relationships of the objects can be analyzed to automatically identify various events, such as shots and goals. Also, a comprehensive analysis of the data for each object and event can reveal semantic information about the game content, thus enabling viewers to better understand the state of play. Using advanced video analysis technology in this way makes it possible to acquire a wide range of information from physical data such as player position to semantic information such as a description of the competition.

Player and ball positions are particularly important items of information, because they are the basis for understanding events such as shots and goals and acquiring a higher-level understanding of the competition. Player positions can also be measured with Global Positioning System (GPS) sensors or radio-frequency identifier (RFID) tags, but devices worn by players can affect their performance, so, with some exceptions, this approach has limited use. To avoid this problem, a technique of measuring the position of an object by analyzing images has been studied as a noncontact sensor approach. Here, we focus on using image analysis for object tracking.

In television broadcasting, on the other hand, an expected application of video analysis technology is the creation of new video representations that deepen viewer understanding and interest in sports competitions. The viewers’ interest is mainly focused on the players and the ball, whose movements can be visualized in a way that is easy to understand by image-based object detection and tracking. The detection and tracking of particular photographic subjects in video sequences has been studied for a long time. This technique began with image feature matching and proceeded to highly accurate three-dimensional (3-D) position measurement by multiple-view video analysis and machine learning. Robustness was then further improved by online machine learning (Fig. 2). Machine learning includes techniques and methods that try to implement the types of learning that are natural to humans on computers. In the pretrained model approach illustrated in Fig. 2, a model is trained to recognize the patterns of image features of objects in advance, making it possible to automatically track the target object on a screen. Online learning, on the other hand, involves relearning as new data is acquired. The application of online learning to object tracking makes it possible to flexibly handle changes in the appearance of an object that result from changes in orientation and so on.

We have also been moving forward with research on video analysis for object tracking and have applied the results to sports programming. In live sports coverage, the objects to be tracked and the shooting environment vary with the competition, so the video analysis technology must be adapted for use in the competition and shooting conditions. Also, the program follows the competition, so the analysis processing should be completed during or immediately after the play. Because high-speed processing and high accuracy are both required, accuracy must be maximized while keeping the computational load low. When developing a system for sports broadcasting, it is therefore necessary to choose analysis processing that is suitable for the competition and shooting environment and find an optimal trade-off between accuracy and speed.

In the next section, we describe the progress in object tracking technology and its application to sports programming.

2. Progress in Object Tracking Technology

2.1 Matching-based object tracking technology

Matching-based object tracking technology tracks objects by matching image features between reference area and target area. The general processing flow is illustrated in Fig. 3. After preprocessing the input camera video for background differencing, etc., the tracking target area

---

*1 A wireless, noncontact means of reading and writing data to an IC RFID tag, which is a data recording medium.

*2 A method of extracting regions of an image that contain moving objects using the difference between the image and the previous image.
(reference area) is specified and image features are extracted from that area. In the subsequent frames, the target object is tracked by comparing the image features of the reference area with those of the search area.

The most basic algorithm is template matching, in which a full raster-scan search is carried out at the position within the search area where the sum of values, such as absolute values, or the sum of squares of the pixel-by-pixel difference, within the reference area is minimum. When rotation and scaling changes are taken into account, the computational load is very high, so the mean shift method has been proposed for fast and efficient searching.

As an intermediate method between the local search and the full search, the use of particle filters for the probabilistic estimation of object position has been proposed.

An application of matching-based object tracking is displaying the trajectory of a pitched baseball (Fig. 4). Before the ball is pitched, the location and size of the search area are specified manually and a search for the ball is made within that area. After extracting the moving objects by frame differencing, the object with image features (color, shape, and movement) closest to those of the ball is detected. The movement of the object is then predicted using a Kalman filter and the object is automatically tracked by moving the search area to the predicted position. The green dotted line in Fig. 4 approximates the curve calculated in the prediction process. Limiting the movement of the search area to the vicinity of the curve achieves highly precise tracking with few detection errors.

The processing flow for the system, including the tracking process described above, is illustrated in Fig. 5. The pitch trajectory is drawn from the position data of the detected ball and the result is composited with the original broadcast camera video. The processing from the video input to the composite video output is performed at the broadcast camera frame rate and can be used for live broadcasting. The system was used in baseball broadcasting for nine years from 2004 to 2012. The trajectory composite image was presented in slow playback after the pitch so that the ball motion could be easily understood by

---

*3 A method of scanning the horizontal scanning lines of a display screen beginning from the upper left and proceeding to the lower right.

*4 A method of detecting very large values of a density distribution function by repeatedly shifting the data center of gravity (mean value).

*5 A method of estimating the state of a dynamic system by the numerical calculation of many particles.

*6 A calculation method for estimating the state of a dynamic system using observation data that includes errors.
visualization (Fig. 6).

2.2 Three-dimensional position measurement by multiple-viewpoint video analysis

The 3-D position of an object can be calculated from two-dimensional coordinates on object images obtained by object tracking from multiple cameras that have different viewpoints. With such a multiple-viewpoint camera system, a 3-D CG trajectory can be drawn and composited on images from cameras other than a sensor camera\(^7\). Because position and speed in real space can be calculated, it is possible to provide a more detailed analysis and explanation of play.

The processing flow for 3-D trajectory CG drawing by multiview video analysis is illustrated in Fig. 7. First, the cameras are calibrated by measuring their positions and orientations with respect to the court. The calibration is carried out by photographing a grid pattern of known size and scale. The measured camera positions and orientations in the image coordinates of the objects detected by each camera are combined and used to calculate the 3-D positions of the objects. Specifically, a virtual line-of-sight vector that connects the focal point of the camera and the object coordinates on the imaging surface is extended in 3-D space and the intersections of the vector with vectors from other cameras are calculated to determine the 3-D position of the object. Using a virtual camera head that can acquire camera orientation information (amounts of pan, tilt, and zoom) in real time makes it possible to measure the 3-D position from moving camera images. We can also expect improved object tracking by processing to predict the position in 3-D space.

This multiview video analysis was applied in the analysis of a beach volleyball game as shown in Fig. 8. We recorded the final game in the 2017 Japan Beach Volleyball Women’s Championship using four cameras to follow the ball and calculated the 3-D position for each frame. The state of tracking processing for images from two of the cameras is shown in the figure.

\(^7\) A camera used to measure object position.
For the final game described above, complete 3-D ball trajectory data was collected and used for explanation in the “Sports Innovation” program (Fig. 9). The 3-D spatial analysis enabled, for example, a more detailed game analysis of changes in ball trajectory caused by the speed of the serve and the wind.

The use of multiple cameras for 3-D object position measurement has already been applied in professional sports refereeing. For example, professional tennis tournaments have adopted a “challenge system”, in which CG images can be used to ascertain whether a ball was in bounds or out of bounds when a player disputes a referee’s judgment. The serve of professional tennis players may exceed 200 km/h, making it difficult for a referee to follow the ball by eye. A system referred to as “Hawkeye” makes it possible to track such serves. This system uses more than 10 high-speed cameras that are installed around the court, the ball is tracked from each camera image, and the 3-D position is calculated for each frame. Decisions such as whether a serve was in or out are made with high accuracy in 3-D space and the result is visualized with a CG image for easy understanding. This technology is also used in soccer, and “Goal Line Technology” was adopted for automatically making goal or no-goal decisions in the 2014 Brazil World Cup.

The 3-D positions of players are also being automatically measured in soccer. In Japan, the company Data Stadium uses TRACAB, a product of the US company ChyronHego, to provide position data for players and the ball for J-League matches and other events. This data serves as a basis for providing statistical information such as the calculated speed of players and their distance traveled, and heat maps, which are attracting attention as new data that enhances the appeal of watching sports events.

2.3 Video analysis by machine learning

There have been many applications of machine learning to object tracking in recent years, increasing its accuracy and tracking performance. The processing flow for tracking that uses a pretrained model is shown in Fig. 10. In this approach, a model is trained with images of the target object as positive examples and images that are not the target

---

*8 Visualization graphs that express the strength or weakness of individual values in a data matrix.
object as negative examples by extracting image features such as the brightness gradient and color histogram. The classifier, such as a support vector machine, is trained using the supervised learning paradigm. In operation, the features are extracted from the candidate object area and the object is detected according to the decision of the classifier.

The pretrained model approach is suitable for live broadcasting, which requires immediate, high-speed processing, because the object is tracked without updating the pretrained classifier. On the other hand, the tracking performance strongly depends on the training data, so stable operation is difficult to maintain when the object appearance changes or when the lighting changes such as during outdoor events.

As a way to cope with such changes in image features, an online learning tracking method in which the classifier is updated for each image acquired during operation has been proposed. In this approach, tracking robustness is increased by repeatedly retraining the classifier on positive-example image features that are extracted from the peripheral area of the object detected in the current frame and negative-example features that are extracted from the outside area. This approach is suitable for tracking under changing operating conditions, but considerable time is required to update the classifier for each frame. It is thus difficult to deal with fast-moving objects, although online learning applied to tracking can be used in sports broadcasting in the absence of fast-moving objects.

Here, we present an example of visualizing trajectories that takes advantage of the features of online learning for curling competitions. A slow-moving object has a smaller amount of movement within the camera image, so it is not always necessary to perform tracking processing at the frame rate of the camera. For example, a curling stone moves at a relatively low speed, so online learning can be applied to track the stone. Furthermore, curling players sweep the ice with the brush and the appearance of the stone changes from moment to moment (Fig. 11). In this case, online learning enables robust processing in which tracking is not interrupted, even with severe object occlusion.

In the live broadcast of the Japan Curling Championships in 2017 and 2019, the trajectories of stones slid in the competition were displayed cumulatively (Fig. 12). The passing of a stone changes the conditions of the ice on the path and thus modifies the ease of sliding and path bending. Displaying the cumulative stone trajectories made it possible to visualize the ice conditions, which are not easy to see in normal camera images, and thus made it possible to explain the state of play in an easy-to-understand way.

3. Conclusion
The analysis of sports video to detect and track players and the ball provides object position information to generate new video representations. Video-based object-tracking techniques have been researched for a long time, and

---

*9 A graph that shows the statistical distribution of colors in an image.
*10 A model training method in which the data is labeled as correct or incorrect.
we have applied the technology in sports programming. Here, we have described the progress in object tracking technology and presented examples of its application to baseball, beach volleyball, and curling competitions. The development of this technology has enabled object tracking that is robust against changes in visibility and speed, and provided new visual representations of information in sports broadcasting. On the other hand, the applications of the various tracking methods are often limited to particular types of competition and shooting environments, and an object tracking technique that can be applied to a wide range of situations is a challenge for the future.

Machine learning has become mainstream in recent years, and video analysis technology that uses machine learning has a range of applications, including automatic camera work, automatic switching, and free-viewpoint video generation in addition to object tracking. Its use is also expanding greatly. Techniques for understanding human poses from camera video are also becoming common, and motion recognition and event recognition are also advancing. Future technology that automatically acquires semantic information capturing the state of play in a game is expected. We will continue to monitor the most recent technological trends and promote the research and development of video analysis technology to create sports programming that is more appealing and easier to understand.

References