Automatic Control of a Robot Camera for Broadcasting Based on Cameramen’s Techniques and Subjective Evaluation and Analysis of Reproduced Images

Daiichiro Kato¹), Tetsuo Katsuura²) and Hideo Koyama²)

¹) NHK Science and Technical Research Laboratories
²) Faculty of Engineering, Chiba University

Abstract With the goal of achieving an intelligent robot camera system that can take dynamic images automatically through humanlike, natural camera work, we analyzed how images were shot, subjectively evaluated reproduced images, and examined effects of camerawork, using camera control technique as a parameter. It was found that (1) A high evaluation is obtained when human-based data are used for the position adjusting velocity curve of the target; (2) Evaluation scores are relatively high for images taken with feedback-feedforward camera control method for target movement in one direction; (3) Keeping the target within the image area using the control method that imitates human camera handling becomes increasingly difficult when the target changes both direction and velocity and becomes bigger and faster, and (4) The mechanical feedback method can cope with rapid changes in the target’s direction and velocity, constantly keeping the target within the image area, though the viewer finds the image rather mechanical as opposed to humanlike.


Keywords: robot camera, automatic shooting, camera work, human camera handling, control algorithm, subjective evaluation experiments

Introduction

In the field of broadcasting, recent years have seen the development of new and diversified image and information services in step with the remarkable progress made in digital technology. At the same time, the ongoing convergence of communications and computer technologies is making it possible to provide multimedia services, which combine images, data, and other media in various ways, through Integrated Services Digital Broadcasting (ISDB). Broadcasting stations are thus being required to produce an enormous number of programs in response to new broadcast services and digital-based multi-channel systems. Against this background, expectations are growing for a program production support system that can efficiently create contents (featuring various types of information and data in addition to conventional video and audio) in a high-quality multimedia format attractive to viewers (Mitsumine et al., 1995; Enami, 1996; Kim, 1997; Yamanouchi et al., 1997; Hayashi et al., 1999). With this in mind, the authors have been researching an intelligent robot camera system as one new way of supporting program production. This system aims to shoot objects automatically and produce natural images that appear to have been taken by human camerawork. Furthermore, by making effective use of auxiliary information like three-dimensional coordinate data to display, for example, the path and velocity of a moving object as needed, we aim to create a flexible system that allows viewers to enjoy video and information in a selective manner.

To date, there have been many instances of robot cameras and remote-controlled cameras in the broadcasting field (Ikeda, 1987; Sugawa et al., 1989; Michael et al., 1989; Ray, 1990; Robert et al., 1991). For example, observation cameras and weather cameras for rooftop installation on broadcasting stations and servo cameras for use in sports events have been developed. Recent research, moreover, has been focusing on integrated robot-camera control equipment plus action cameras that achieve three-dimensional motion in combination with a crane (Aoyama, 1993; Ohtsubo et al., 1997).

Research into automatic tracking cameras, moreover, includes studies on detecting and following the position of an object using color information and ultrasonic waves. Case studies have been reported on the automatic camera shooting of commentators inside news studios and that of animals in wild settings (Yamanaka, 1993; Mitsuzuka et al., 1994; Yamanaka et al., 1995; Yokota et al., 1999; Ishii, 1993). The work at program-production sites, however, has traditionally been recognized as “work done by people,”
and the role of robot cameras has been assumed to be a supplementary one at best. This way of thinking originates in the fact that no attempts have been made to reproduce images by human-like camerawork despite the fact that robot cameras can display an object on a screen by panning at uniform velocity, can capture an object at screen center and keep it there, etc. As a result, images taken with a robot camera still appear to shift in a mechanical manner, and their application has been severely limited in broadcasting. In response to this state of affairs, we have been researching the incorporation of human-engineering characteristics so as to achieve intelligent robot cameras that can be used effectively in a wide range of broadcasting applications.

To realize intelligent robot cameras, analysis of the techniques used by broadcast cameramen in actual studio work is indispensable. Some camerawork analysis systems have already been constructed (Kato et al., 1994) and analysis of cameramen’s techniques have been performed through basic experiments employing actual programs and studios. These experiments revealed the following (Kato et al., 1995; Kato et al., 1996; Kato et al., 1997).

(1) When shooting a moving object, the cameraman surveys the overall situation with his naked eye, estimates the speed of the moving object, and moves the camera appropriately.

(2) Depending on the direction and size of the moving object, the cameraman determines its optimal position within the screen (viewfinder screen) and the range of tolerance with respect to this position.

(3) In the event that camera speed and predicted object speed deviate during shooting and the range of tolerance within the screen is exceeded, the cameraman corrects the camera’s speed. This is performed, however, without sudden compensations. In addition, the following results were obtained by subjective evaluation of reproduced images.

(1) The four main factors in evaluating camerawork are continuity, vividness, sensitivity, and human warmth (Kato et al., 1997; Kato et al., 1998).

(2) Images taken in a manner similar to that of a cameraman in actual shooting, in which the camera is kept just ahead of a moving object, received relatively high evaluations in terms of humanlike characteristics and sensitivity (Kato et al., 1998).

This time, using camera control technique as a parameter, we analyze the way in which images are taken and perform subjective evaluation tests to examine the effects of camerawork.

**Method**

An experiment was performed in which an object was shot while varying camera control conditions, namely the position adjusting velocity curve for the object within the screen, tolerance range of target position, and time (duration) of position adjustment, and subjective evaluation tests were performed on the reproduced images.

**Position adjusting velocity curve of the target**

The position adjusting velocity curve of the object within the screen is the velocity curve that occurs when returning the object to its optimal position after it has exceeded its tolerance range within the screen. For each adjustment, the curve shows the time required to make the adjustment normalized in terms of maximum object speed within the screen at this time (the maximum value of the ratio of the difference in object position on the screen between video frames to the video frame period (1/30 sec)) (Fig. 1). The position adjusting velocity curve features slight fluctuations in speed, a characteristic of human operation. Past analyses have shown that the curve is independent of object velocity and size, and exhibits a quasi curve shape. In the experiment performed to analyze how the position adjusting velocity curve can affect camerawork, we first performed data analysis on position adjusting velocity curves and then selected curves for use in the comparison experiment.

Specifically, a circular target 10 cm in diameter was placed 5.5 m from the camera and driven in the horizontal direction on a path 3.6 m long, and three cameramen with experience of 10, 12, and 15 years were asked to track and shoot the target. Three velocities were employed for the target equivalent to 4.7 deg/sec, 6.7 deg/sec, and 10.1 deg/sec in panning speed, and three view angles were used in which the diameter of the target corresponds to 1/2, 1/4, and 1/6 of the vertical view angle. Five consecutive shots were made for each of these velocities and view angles, which were selected on the basis of results from previous camerawork-analysis experiments using real programs. The camerawork data so obtained were analyzed, and after selecting position adjusting velocity curves, subjective evaluation tests were performed.

**Tolerance range of target position and position-adjustment time**

Another experiment was performed with parameters being the tolerance range of target position with respect to optimal position within the screen, and the duration of position adjustment when target position exceeds the tolerance range. The target was shot under three sets of control conditions, named Condition A, Condition B, and Condition C. In Condition A, tolerance range and the time of position adjustment were set to the same shooting conditions of a cameraman (average values obtained from past experiments involving nine cameramen). In Condition B, they were set 30% smaller than the shooting conditions of a cameraman. And in Condition C, correction to the optimal position is forcibly done every video frame (every 1/30 sec) in a manner typical of mechanical control.
Conditions A and B constitute feedback-feedforward control, which imitates the way the cameraman manipulates the camera, using "human (typical) data" for position adjusting velocity curves, and Condition C is mechanical feedback control. The experiment examined two cases. In the first, movement of the object is limited to one direction, and in the second, movement and speed of the object are varied. For each of the above sets of control conditions, subjective analysis was performed on obtained images after analyzing the way in which the object was shot.

With regard to the above two cases of subjective evaluation, all images were shot with experimental camera-control equipment that accurately reproduces the camerawork of cameramen characterized by slight fluctuation in speed (Kato et al., 1996; Kato et al., 1997). Similar to the above shooting experiment involving cameramen, a circular target 10 cm in diameter was placed 5.5 m from the camera and driven in the horizontal direction on a path 3.6 m long. The target was followed and shot with the video camera (Fig. 2). The optimal position of the target within the screen corresponds to the average value of data obtained from the work of nine cameramen in past experiments.

For the subjective evaluation tests, sample video was recorded on a VTR for each set of experiments and shown to test viewers. Specifically, eight broadcasting engineers and seven ordinary people were asked to evaluate the reproduced images at seven levels on a bipolar scale using the four main factors of continuity, vividness, sensitivity, and human warmth (Kato et al., 1998) as evaluation axes. In these experiments, no special attention was paid to differences between viewers; that is, the values obtained from all viewers with regard to the above factors were used to calculate averages for use as evaluation values. To compare the various experimental parameters (control conditions), we first checked significance by analysis of variance (ANOVA) that analyzed the effects of nominal independent variables on dependent (continuous) variables. We then performed a post hoc test (Fisher PLSD) and evaluated the parameters by a significance standard of 5%.

**Results**

*Position adjusting velocity curve of the target*

The frequency of occurrence with respect to the time (normalized time) at which maximum velocity occurs in position adjusting velocity curves was plotted for camerawork data obtained from repeated shooting by three
Cameramen's Techniques and Subjective Evaluation

Two graphs are shown, one for the case of panning to the right (clockwise) and the other for panning to the left (counterclockwise). From the figure, it can be seen that the maximum velocity occurred most frequently before the normalized time of 0.5 for camera panning to the right and after 0.5 for camera panning to the left. The normalized areas of position adjusting velocity curves were then calculated and average values determined for each of the experimental conditions described above, as shown in Table 1. As can be seen, normalized areas take on values from 0.45 to 0.6. Based on these results, the position adjusting velocity curve taken to be most representative of this data was selected and defined as “human (typical) data”. Its curve shape is shown in Fig. 4. Then, for the sake of comparison, we selected three other position adjusting velocity curves. The first was selected randomly from the above data and defined as “human (random sampling) data”. The second was obtained by taking moving averages of “human (typical) data” and defined as “human (moving-average) data”, which is easy to apply to mechanical control. The third corresponds to a well-known mechanical control technique and is defined as “asymmetrical-triangle control data”. The target was therefore shot on the basis of these four position adjusting velocity curves and the images obtained were presented for subjective evaluation.

Figure 5 shows the results of subjective evaluation for the experiment having the target’s position adjusting velocity curve as parameter. In this experiment, target velocity was fixed at 6.7 deg/sec in terms of panning velocity and target size was varied. In the case of small target size, no significant difference can be seen among the four position
adjusting velocity curves. As target size increases, however, evaluation with respect to all factors drops in the case of “asymmetrical-triangle control data.” Evaluation also drops slightly for continuity and sensitivity in the case of “human (random sampling) data” and “human (moving-averages) data”. “human (typical) data”, on the other hand, maintains high evaluation regardless of target size. In contrast, Fig. 6 shows the results of subjective evaluation when target size is fixed to 1/4 and velocity is varied. Although evaluation tends to drop for all factors with respect to velocity in the case of “asymmetrical-triangle control data,” no significant

Fig. 3 Frequency of occurrence with respect to the time (normalized time) at which maximum velocity occurs in position adjusting velocity curves. Total value of three cameramen, three velocities, three view angles. (a) Case of panning to right direction (CW). (b) Case of panning to left direction (CCW).

Fig. 4 Normalized position adjusting velocity curves for subjective evaluation tests
differences can be observed for the other three position adjusting velocity curves.

**Tolerance range of target position and position-adjustment time**

Table 2 shows the results of measuring maximum normalized fluctuation in target position within the screen for each of the conditions (Condition A, Condition B, and Condition C) when following and shooting the target moving from left to right. Normalized fluctuation is determined by measuring the amount of change in target position within the screen per video frame (1/30 sec) and normalizing that against screen width. For Conditions A and B, fluctuation becomes large as target size or velocity increases, and the resulting image appears to shake significantly within the screen. In Condition C as well, increase in target size or velocity increases fluctuation, but the amount of fluctuation is not as large. Adjustment time is about 300 msec for Condition A, but only about 33.3 msec for Condition C. As a result, shaking frequency of the target in the screen is higher in the latter case.

Figure 7 shows the results of subjective evaluation for the experiment with the target moving in only one direction (left to right). As target size increases, evaluation does not drop all that much for Conditions A and B, but drops significantly for Condition C. Moreover, as target velocity increases, the factors of continuity and vividness are evaluated highly for Condition C. Here, however, human warmth receives a low evaluation as the resulting image exhibits a mechanical feeling. Comparing results for Conditions A and B, it can be seen that while differences are slight, Condition B produces significantly higher evaluation for continuity and vividness for target sizes of 1/6 and 1/4.

We next examined the amount of transitional displacement when following and shooting the target for each of the three control conditions when varying the direction and velocity of the moving target. Results are shown in Fig. 8. Transitional displacement is the target's displacement from its optimal position within the screen at the instant when it changes direction. It is a value normalized against screen width. A transitional displacement of “1” corresponds to the case when the target moves completely outside the screen. Figure 8 (a) shows results for a decrease in target velocity from 6.7 deg/sec to 4.7 deg/sec and Fig. 8 (b) those for an increase in target velocity from 4.7 deg/sec to 10.1 deg/sec when changing direction (turning). As shown, transitional displacement is smallest under Condition C in all cases. Under this condition, the target can be kept within the
screen even if the moving target suddenly changes direction. Under Conditions A and B, however, the target moves outside the screen in the case of maximum target size (1/2) and acceleration on changing direction. This result shows that the control method that imitates human camera operation cannot perform rapid position adjustment. In other words, there are limits to what human camera operation can perform.

The results of subjective evaluation when varying the direction and velocity of the moving target are shown in Fig. 9. Comparing these results with those for the target moving in only one direction, we see that evaluation improves overall for Condition C. Evaluation of human warmth, however, is low, because of the same feeling of mechanical movement. As target size increases, Conditions A and B show low evaluation for continuity and vividness although evaluation of sensitivity and human warmth does not drop significantly. These results also included a case in which the target moved outside the screen, but even this was considered to be humanlike by the test targets and was not felt to be uncomfortable.

**Discussion**

On examining the results of these subjective evaluation tests, no significant differences in evaluation scores could be observed between parameters on the whole. This is probably due to the fact that, with the aim of achieving a robot camera that can shoot video with natural camerawork, a subjective evaluation would be performed after slightly changing a parameter based on cameramen's shooting techniques. As shown in section 3, however, significant difference could be observed by dispersion analysis. With the above in mind, the following
discusses each set of experimental results in order.

The experiment with position adjusting velocity curve within the screen as parameter revealed no distinct differences in results by subjective evaluation in the case of small target size. This is probably because they are difficult to visually distinguish. Depending on shooting conditions, it is not necessarily the case that “asymmetrical-triangle control data”, which has been considered for configuring circuits as an extension of existing control technology, cannot be used in robot camera drive systems. As target size increases, however, evaluation of all image factors drops when using “asymmetrical-triangle control data”. On the other hand, the use of human data maintains high evaluation regardless of target size. According to a post hoc test, though, there are also conditions for which significance cannot be judged by target size between the three sets of human data. In contrast, significance can clearly be established for “asymmetrical-triangle control data,” that is, for mechanical control. With regard to target velocity, clear significance cannot be observed compared to the above case when varying target size. The reason for this is thought to be that as target velocity (or in other words, camera panning velocity) increases, background movement also speeds up, and target oscillation becomes difficult to follow.
Nevertheless, if we evaluate shooting conditions in a comprehensive manner, human position adjusting velocity curve data can be judged effective for use in the control of robot cameras that are known to be capable of taking dynamic video images.

Next, for the experiment with target-position tolerance range and position adjustment time as parameters, Conditions A and B that imitate operation by cameramen exhibit high evaluation overall compared to Condition C, a mechanical control technique. As target size increases, Condition C shows a large drop in evaluation, which is thought to be due to clearly recognizable movement resulting from frequent compensation. We surmise that test targets were not comfortable with this movement. As
target velocity increases, however, Condition C shows improved evaluation for the factors of continuity and vividness. Human warmth, on the other hand, receives a low evaluation in this case, as the reproduced image feels too mechanical. Comparing Conditions A and B, it can be seen that, while differences are slight, Condition B shows higher evaluation for continuity and vividness for target sizes of 1/6 and 1/4. We surmise the reason for this to be large target oscillation in the case of Condition A, as reflected by the normalized fluctuation values in Table 2. In a post hoc test for target sizes of 1/6 and 1/4 and target velocities of 4.7 deg/sec and 6.7 deg/sec, Conditions A and B achieve significance with respect to Condition C. For a target velocity of 10.1 deg/sec, though, it was confirmed that Condition C achieves significance. This is attributed to the difficulty of shooting a quickly moving target by human camera techniques. The possibility of producing favorably evaluated images by mechanical control was therefore demonstrated.

When changing the direction and velocity of the moving target, subjective evaluation improved overall for Condition C. This result is thought to be due to the manner of capturing the image when the target changes direction. In the case of maximum target size (1/2), the target may move outside the screen under Conditions A and B, reflecting the limitations of human camera operation. Under Condition C, however, the target can be accurately followed although the resulting movement tends to appear mechanical. This result also reveals a strong point of mechanical control.

The above results indicate that an optimal control method for on-site robot cameras would be a configuration that skillfully incorporates the best features of human-based camera control and mechanical feedback control. For example, control that imitates human camera operation could be applied to an athlete-tracking camera when the movement of the object is known to be one-way, as in ski-jump and short-distance track and field events. On the other hand, mechanical feedback control could be applied when it is difficult to predict change in object's motion as in soccer and rugby field matches. In this way, a totally new type of dual-use control method could be developed that produces images that in the past could not be taken by human cameramen. We therefore see the possibility of achieving a new form of image expression unknown up to new through the use of robot cameras.

Conclusions

For the purpose of achieving an intelligent robot camera system that can take dynamic images automatically through humanlike, natural camera work, we analyzed how images were shot, subjectively evaluated reproduced images, and examined effects of camerawork, using camera control technique as a parameter in shooting. The following results were revealed.

For target movement in only one direction, relatively high evaluation was obtained for images taken by camerawork based on feedback-feedforward control that imitates camera operation by a cameraman.

1. High evaluation was obtained for the position adjusting velocity curve of the target within the screen when using human data; In particular, when using “human (typical) data,” which features high frequency of occurrence in shooting-by-humans experiments, excellent evaluation was observed (See Fig. 3, Table 1, and Fig. 5).

2. For target movement in only one direction, relatively high evaluation was obtained for images taken by camerawork based on feedback-feedforward control that imitates camera operation by a cameraman (See Fig. 7).

3. When varying both the direction and velocity of the moving target, keeping the target inside the screen proved to be difficult for control methods that imitate human camera operation in the case of large target size and velocity (1/2, 10.1 deg/sec). Test targets, however, thought this to be human in character and did not feel overly uncomfortable with such occurrences (See Figs. 8 and 9).

4. Mechanical feedback control was found to be capable of dealing with sudden changes in target direction and velocity and of keeping the target within the screen. Reproduced images by such control, however, were not highly evaluated in terms of human camerawork (See Figs. 8 and 9).

In short, incorporating control methods that imitate human camera operation into robot camera systems enables humanlike, sensitive images to be obtained. At the same time, control methods that imitate human camera operation cannot always keep the target within the screen if the moving target should make sudden changes in direction and velocity. On the basis of these results, we can therefore consider a new control configuration that skillfully incorporates the advantages of both human camera operation and mechanical feedback control.

In future research, we plan to perform additional experiments and studies to find an optimum control algorithm for broadcast robot cameras and ways of customizing control to specific programs, all with the ultimate goal of developing a new and useful program production support system for broadcasting.

Acknowledgments. The authors would like to extend their deep appreciation to Director Osamu Yamada, Assistant Director Kazumasa Enami, General Manager Kazuo Fukui of Multimedia Services, and chief researchers Shigeru Shimoda and Hiroshi Fukushima of NHK Science and Technical Research Laboratories, and Assistant Professor Yuichiro Kume of Tokyo Institute of Polytechnics, for their
guidance throughout the course of this research. We would also like to thank Mr. Akio Ishikawa and Mr. Takao Tsuda of Multimedia Services, Toshiyuki Higuchi of Tokyo Institute of Polytechnics, and the cameramen and many others at NHK Broadcast Engineering Center and NHK Technical Service for their gracious help with the experiments.

References

Aoyama T (1993) Integrated robot camera control system. NHK domestic report 4-2 8-H6
Ray L (1990) NBC camera robotics system. SMPTE Journal 99 3: 203-208

Received: September 9, 1999
Accepted: November 30, 1999
Correspondence to: Daiichiro Kato, NHK Science and Technical Research Laboratories, 1-10-11, Kinuta, Setagaya-ku, Tokyo 157-8510, Japan
e-mail: katodai@strl.nhk.or.jp