# Close-up Imaging of a Moving Object by Cooperative Multiple Pan-Tilt Cameras

Hicham BOUCHNAIF†, Yoshinari KAMEDA†, Yasuhiro MUKAIGAWA†, and Yuichi OHTA†

† University of Tsukuba
3M302, 1-1-1 Tenoudai, Tsukuba, Ibaraki, 305-8573.

**Abstract** We propose a new method for filming a moving object by fixed framing of multiple pan-tilt cameras. This filming approach assures the acquisition of high-resolution images of moving objects without blurring, which may occur when a conventional tracking system rotates a pan-tilt camera to film the object. In our approach, we decouple the framing task into two parts, a monitoring part and a stand-by part. We assign the monitoring role to one camera and the stand-by role to other cameras. While the monitoring camera is filming a moving target object, other cameras, which are in stand-by role, are directed to cover the outer areas of the field of view of the monitoring camera as much as possible by predicting the motion of the target object. We propose to estimate prediction of the moving target object based on marginal distribution obtained by projection of the probability density function of the motion model onto the circumference of the image plane. The validity of our method was demonstrated through experiments involving filming a player running around a real soccer field with switching of two to nine pan-tilt cameras.

**Key words** pan-tilt cameras, fixed framing, tracking, object frame-out probability, high resolution textures, large-scale space.

## 1. Introduction

Improvements in computer processing power, storage capacity, and access to high-speed networks through broadband technologies have increased the importance of 3D visualization systems [1] in image media. As improving the quality of produced 3D videos is the most important factor for the widespread use of these systems, the requirements for high resolution and sharp textures of moving and non-moving objects are gaining in importance. Moving objects are common targets for automated filming, especially in sport broadcasting where sharp and high-resolution images of players can help in watching player actions.

Many systems have been developed for indoor and outdoor filming of moving objects.

In some advanced automatic videography methods [2] [3], a pan-tilt camera is rotated through the tracking process. As a result, the motion of the camera is changed frequently. This frequent motion adjustment causes serious problems in video quality for filming human activities and may also result in blurring of textures. These problems become more serious in the case of tracking a soccer player who may move at high speed and perform sudden changes in both speed and direction. This results in high speed and erratic rotation of the camera, which makes acquiring sharp textures of the targets extremely difficult.

Kameda et al. [4] proposed a control method to realize a planned video composition with less adjustments of camera motion. Their approach defines some constraints to reduce camera motion for producing comfortable video for viewers.

Ozeki et al. [5] proposed a virtual frame control method to reduce unpleasant camera motion while tracking an object manipulated by an operator. In their approach, the direction of the filming pan-tilt camera is locked while the object is located inside a virtual frame that is set on a video image, and object tracking is activated just after the object goes beyond the virtual frame. When the object settles down, the camera is locked again. This method works well in the case of tracking a manipulated object; however, it does not support high-speed motion, such as in the case of a soccer player running after a ball on a soccer field.

In this paper, we introduce a new method to control multiple pan-tilt cameras to realize filming and acquisition of sharp and high-resolution textures of an object moving in a wide field. In addition, to describe the basic idea of the proposed method [6] [7], we discuss the procedure of filming.

To realize the filming method, we apply a probabilistic approach to direct pan-tilt cameras based on the position of the object and its motion velocity. Once one of the cameras starts filming the moving object, it will take over the role of filming the object and the video will be transmitted from the camera.

This paper is organized as follows: Section 2. presents an overview of the whole system and explains the method of switching pan-tilt cameras to realize fixed framing of a moving object. Section 3. explains the generic probabilistic motion model used to determine the optimal direction of the cameras to avoid missing the target. Section 4. describes the actual motion models that are applied in this paper. Section 5. presents experimental results using the proposed system. Finally, section 6. concludes with a summary of the system, and mentions future directions.

## 2. Multiple Pan-Tilt Camera System

Our system consist of a set of $m$ pan-tilt cameras zooming in. We assume all these pan-tilt cameras are set at a single location, which is sufficiently far from the field where an object moves. The reason for this assumption will be explained in the next section. We use one of these cameras for filming the object moving in the target field. We call this camera a monitoring camera, and the other $m-1$ cameras are stand-by cameras.

As illustrated in Figure 1, while the monitoring camera is filming the object, stand-by cameras are directed to surround the field of view of the monitoring camera , "FOVM," and to wait for the object to come into their field of view. As there are $m$ cameras, stand-by cameras can cover only $m-1$ regions. To reduce the risk of losing track of the object and realize continuous filming of this object while it is moving inside the target field, the frame-out probability, "FOP," for every possible region (Figure 2) around the FOVM is estimated. These FOPs are then sorted in descending order, and $m-1$ best regions are covered by the stand-by cameras. The estimation of the FOP of each region is based on the direction of movement of the object, which is derived from the apparent motion on the monitoring camera.

When the object moves out of the FOVM and enters the field of view of one of the stand-by cameras, the two roles are switched between the two cameras.

## 3. Optimal Direction of Stand-by Cameras

To achieve robust tracking of the object, a quantitative model to evaluate its direction of movement and its position on the filming camera is needed. The problem of optimal direction of stand-by cameras consists of finding the optimal direction of stand-by cameras when the direction of the monitoring camera and the apparent motion velocity and position of the moving object in this camera are given. Here, optimal direction means assuring the highest probability of capturing the moving object by a stand-by camera, when this object goes out of the field of view of the monitoring camera.

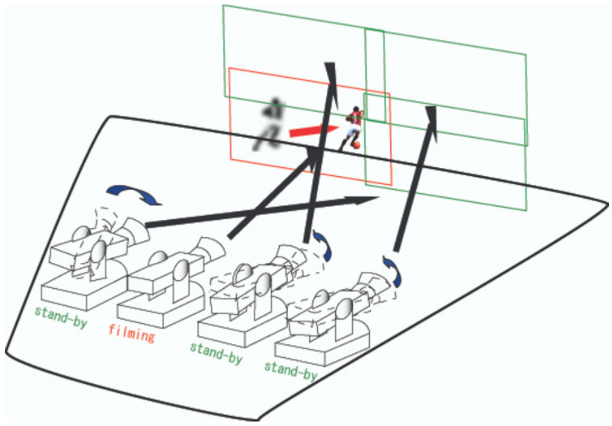


Figure 1 Switching multiple cameras

We introduce a probabilistic motion model of the object to solve this problem. Moreover, to simplify the problem, we reduce the number of directions that can be taken by the cameras by imposing the constraint that stand-by-cameras can take only 8 directions with respect to the monitoring-camera. As illustrated in Figure 2, the stand-by camera framing position has 8 possibilities with respect to the monitoring camera (up, down, left, right, up-right, up-left, down-right, down-left).

In the next section, we will explain how the stand-by camera parameters (pan and tilt values) and how the 8 positions explained above are determined.

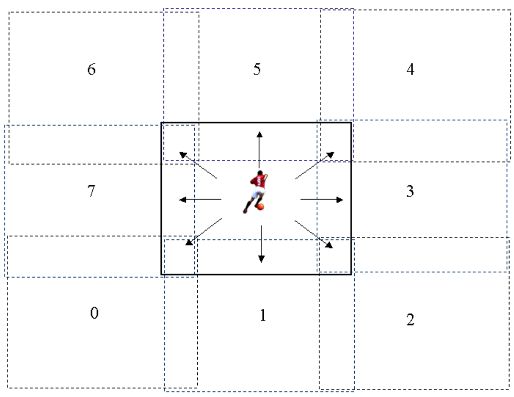


Figure 2 Possible framing positions of stand-by camera with respect to monitoring-camera

### 3.1 Determination of Stand-by Camera Parameters

Let $\theta_h$ and $\theta_v$ be the horizontal FOVM and the vertical FOVM of the camera, respectively. We assume that in each camera, the focal point coincides with its axis of rotation. As the object moves in a large-scale space and the distance between the cameras can be ignored compared to the distance between the moving object and the cameras. The two cameras can be considered in the same place and this approximation simplifies our task. Pan values that can be taken by stand-by cameras can be expressed by the equation (1), where $P_f$ is the actual pan value of the monitoring camera, $P_s$ represents the possible values of the stand-by camera, $\lambda$ is a parameter that can have the values -1, 0, or 1 depending on the rotation direction of the stand-by camera (left, middle, or right, respectively), and $\gamma_h$ corresponds to the overlap angle between stand-by camera and monitoring camera. This is illustrated in Figure 3. Similarly, the tilt values of the stand-by cameras can be expressed by the equation (2), where $T_f$ represents the actual tilt value of the monitoring camera, $T_s$ is the tilt value that will be taken by the stand-by camera, $\beta \in \{-1, 0, 1\}$ specifies the rotation direction of the stand-by camera (down, middle, up), and $\gamma_v$ specifies the vertical overlap angle between the two cameras.

$$P_s = P_f + \lambda(\theta_h - \gamma_h) \quad (1)$$

$$T_s = T_f + \beta(\theta_v - \gamma_v) \quad (2)$$

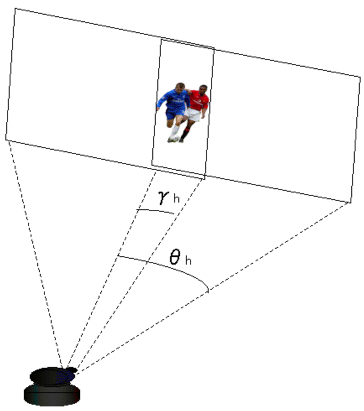


Figure 3 Approximation of pan value

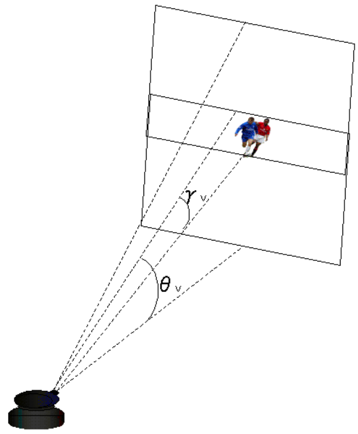


Figure 4 Approximation of tilt value

### 3.2 Generic Form of Frame-out Probability

In our framework, a target object is monitored on the image plane of the monitoring camera. Suppose $C_p = (u, v)^T$ is the position of the object in 2D space of the image plane of the monitoring camera, and $C_v = (\dot{u}, \dot{v})^T$ is the velocity of the object. $e_p$ denotes a point on the image circumference (Figure 5), and $t$ is the time that the target object takes to reach the position, $e_p$.

We describe the frame-out probability of the target object at $C_p$ with $C_v$ and the target object will go out of the FOVM at point $e_p$ at time $t$ by $f(e_p, t, C_p, C_v)$.

Then, the frame-out probability that the object goes out of the FOVM through the zone $[e_{p_i}, e_{p_{i+1}}]$ after $T_1$ seconds can be expressed as follows:

$$E^{T_1}_{e_{p_i} e_{p_{i+1}}} = \int_{e_{p_i}}^{e_{p_{i+1}}} \int_{T_1}^{\infty} f(e_p, t, C_p, C_v) dt de_p \quad (3)$$

By placing the stand-by cameras in the zones with the highest probability, we can realize good coverage with reduced risk of missing the target object.

Note that this formulation can describe any type of motion model embedded in $f(e_p, t, C_p, C_v)$.

## 4. Frame-out Probability Function

The probabilistic model, $f()$, can be formulated in various ways. In this study, we formulated the motion model based on the current direction of movement of the object. The motion model is projected on the circumference of the FOVM to estimate FOP for each of the zones described in Figure 5.

We use the counterclockwise azimuthal coordinates as illustrated in (Figure 6). Let $\theta$ be the azimuthal angle and $\theta = 0$ be the motion direction of the target.

Suppose there is a high possibility that the object keeps moving in the current direction and the further the object de-

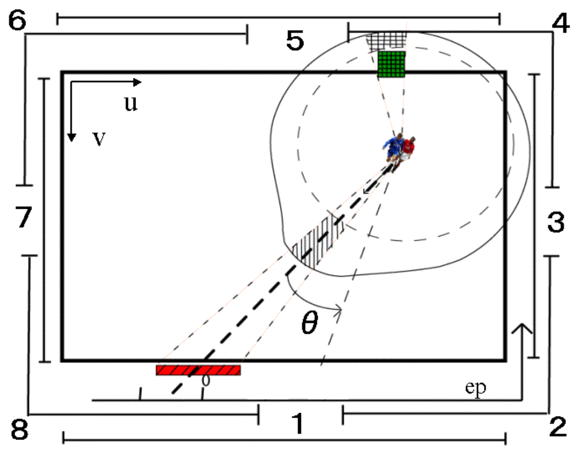


Figure 5 Illustration of the azimuthal distribution, its projection on the image circumference, and the 8 zones to place stand-by cameras.

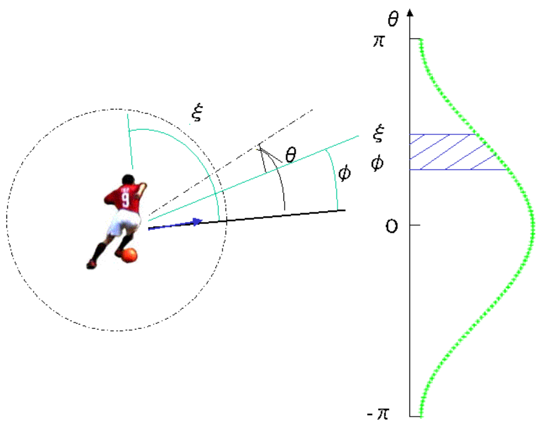


Figure 6 Illustration of the proposed azimuthal distribution.

viates from the current direction the smaller the probability becomes. The model distribution will look like the distribution illustrated in Figure 6. This distribution has only target velocity as a parameter. To achieve our objective, we project this distribution onto the circumference of the image plane (Figure 5).

The following two density functions, $f_g(\theta)$ and $f_c(\theta)$, can be considered as an approximation of the distribution described above.

$$f_g(\theta) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(\frac{-\theta^2}{2\sigma^2}\right) + \qquad (4)$$

$$\frac{1}{\pi}\int_{\pi}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} exp\left(\frac{-\theta^2}{2\sigma^2}\right) d\theta$$

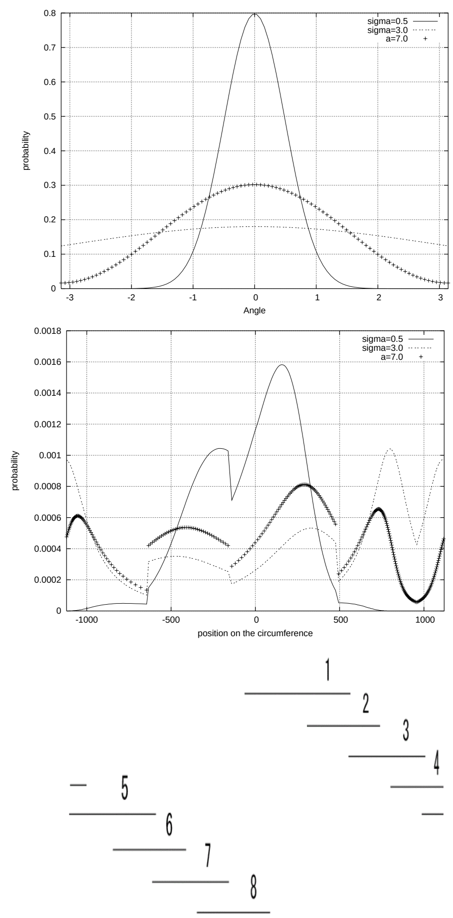


Figure 7 Probability density functions: the first figure illustrates the azimuthal distribution, the figure in the middle represents marginal distribution, and the third figure represents the range of the eight zones.

$$f_c(\theta) = \frac{cos(\theta)}{a} + \frac{1}{2\pi} \ (a > 2\pi) \qquad (5)$$

The second term in equation (4) is added to normalize the proposed probabilistic density function, $f(\theta)$. As $\theta \in [-\pi, \pi]$ and not $[-\infty, \infty]$, the second term is needed to fulfill the condition, $\int_{-\pi}^{\pi} f(\theta) d\theta = 1$.

As illustrated in Figure 7, by changing the parameters in the equations 4 and 5, we can change the behavior of the system. For example if the variance is small in $f_g(\theta)$, the marginal distribution will have its maximum value around the direction of movement of the object; therefore, the zones that are in the direction of movement will represent the highest FOP. This kind of distribution can be chosen when the object is moving at high speed, and seems very unlikely to suddenly change its direction. On the other hand, if the object can move in different directions in a complicated manner, such as in the case of soccer players, it is better to use a distribution with small variance. This can prioritize the zones that are close to the apparent position of the object in the image plane. Hence, we can prevent losing track of the object if it suddenly changes direction.

## 5. Experiments

To evaluate the performance of our system, we performed experiments involving filming a player running on a soccer field. The experiments were conducted in Kasumigaoka National Stadium, Tokyo.

Our system uses a maximum of nine Sony EVI-D100 pan-tilt cameras. As illustrated in Figure 8, the composite signal of each camera is input into a multiviewer and the 9-split output video from the multiviewer is captured with VGA resolution. Note that the 9-split video is used just for object motion estimation. The cameras are controlled using the VISCA protocol over RS-232C cables and all the cables are connected to a USB hub through a serial-USB converter. A PC-controlled video switcher will be installed in order to provide a video stream of zoomed image of the object.

Object detection is performed using temporal differencing, and stand-by cameras are directed based on our frame-out probability model described above.

The system consists of a number of modules, each of which is dedicated to one camera. A thread is created for each camera module. The modules all run concurrently using the pthread technique and they communicate with each other via a common shared memory structure into which the 9 split output video from the multiviewer is captured. Figures 9 and 10 illustrate the results of filming using three pan-tilt cameras. Figures 11 and 12 illustrate the results of filming using five pan-tilt cameras. The monitoring camera that frames the player inside its FOV is indicated by a red rectangle, and other cameras are standing-by for the player coming into their FOV. Our fixed framing results were successful, and the current implementation runs at an average of 29 fps.

## 6. Conclusion

We presented a method for close-up imaging of a moving object by switching multiple pan-tilt cameras so as to realize fixed filming. Switching is performed by assigning a monitoring role to one camera, and a stand-by role to other cameras. Placement of stand-by cameras around the viewing region of the monitoring camera is determined based on the probabilistic motion model of the moving object. We have implemented a system with more than three cameras based on the proposed method, and its effectiveness was confirmed by experiment.

Several issues remain to be addressed, such as improving the frame-out probability model to make the system more stable, realizing extraction of target object's texture, and integrating the method with a 3D visualization system.

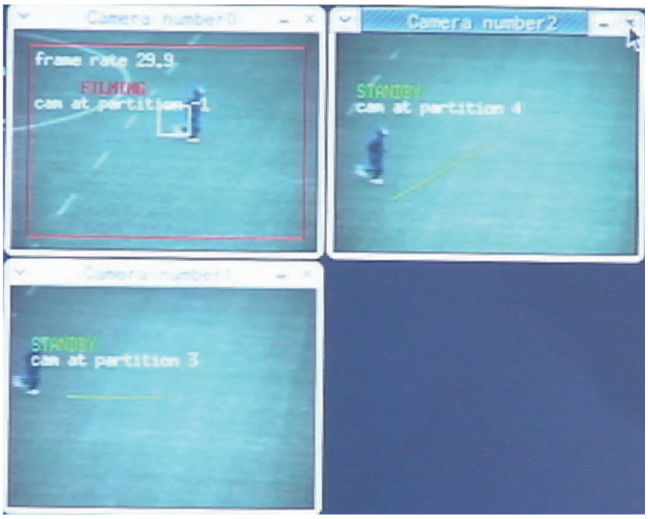


Figure 9 Tracking result 1 using 3 cameras; while the monitoring camera is filming the object, each of the stand-by cameras takes a position and waits for the target

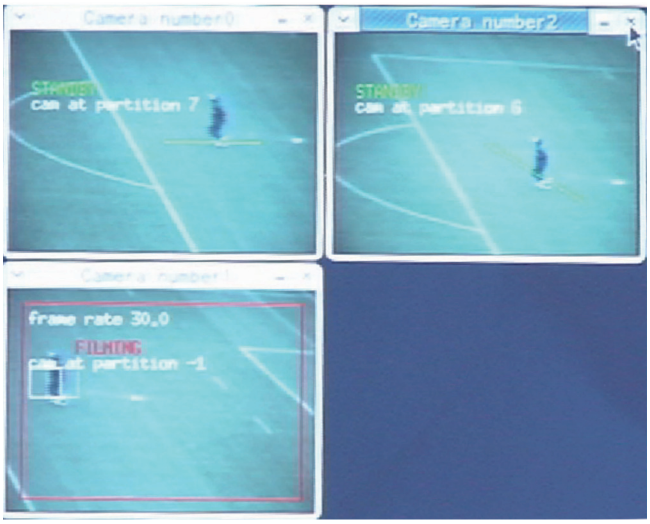


Figure 10 Tracking result 2 using 3 cameras; while the monitoring camera is filming the object, each of the stand-by cameras takes a position and waits for the target


### References

[1] Takayoshi Koyama, Itaru Kitahara, and Yuichi Ohta, "Live Mixed-Reality 3D Video in Soccer Stadium," International Symposium on Mixed and Augmented Reality, pp.178-187, 2003.

[2] R. Collins et al., "A system for video surveillance and monitoring:Vsam final report," Tech report CMU-RI-TR-00-12, Robotics Institute, CMU, 2000.

[3] Y. Kameda, S. Nishiguchi, and M. Minoh, "CARMUL: Concurrent Automatic Recording for Multimedia Lecture," Proc. IEEE International Conference on Multimedia and Expo, pp.129-132, 2003.

[4] Y. Kameda, Y. Atarashi, S. Nishiguchi, and M. Minoh, "Reduction of Camera Motion Adjustments Under a Planned Video Composition with Pan-Tilt Camera," Proc. Asian Conference on Computer Vision 2004, Vol.1, pp.216-221, 2004.

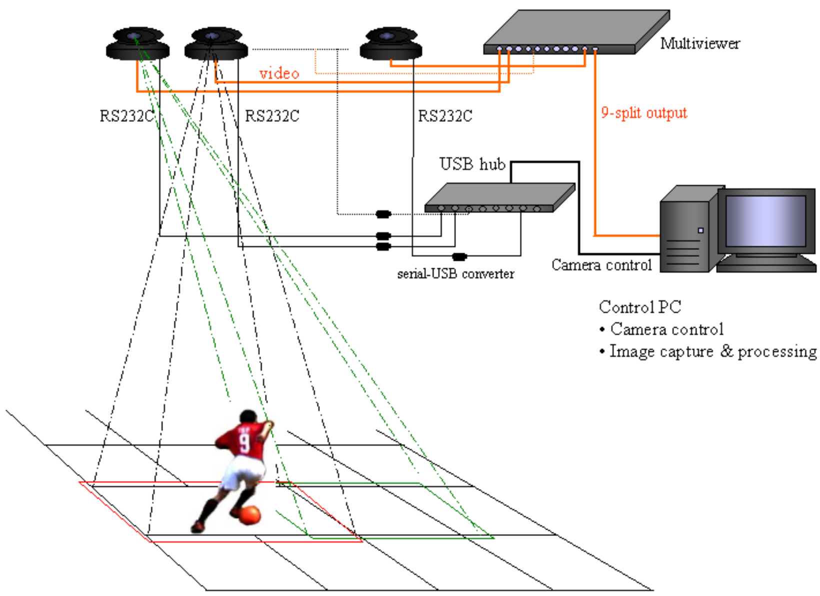


Figure 8　System overview

Figure 11　Tracking result 1 using 5 cameras; while the monitoring camera is filming the object, each of the stand-by cameras takes a position and waits for the target

Figure 12　Tracking result 2 using 5 cameras; while the monitoring camera is filming the object, each of the stand-by cameras takes a position and waits for the target


[5] M. Ozeki, Y. Nakamura, and Y. Ohta, "Camerawork for intelligent video production -capturing desktop manipulations," Proc. IEEE International Conference on Multimedia and Expo, pp.41-44, 2001.

[6] H. Bouchnaif, Y. Kameda, Y. Mukaigawa, and Y. Ohta, "Tracking a Soccer Player by Switching Multiple Pan-Tilt Cameras" Forum on Information Technology (FIT2004), I-038, pp.83-84, 2004.

[7] H. Bouchnaif, Y. Kameda, Y. mukaigawa, and Y. Ohta, "Probabilistic Framework for Intelligent Filming by Switching Temporarily Locked Pan-Tilt Cameras" Proc. IAPR Machine Vision Applications, pp.108-111, 2005.