MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
3-23

Probabilistic Framework for Intelligent Filming by Switching Temporarily Locked Pan-Tilt Cameras

Hicham Bouchnaif   Yoshinari Kameda   Yasuhiro Mukaigawa   Yuichi Ohta

Department of Intelligent Interaction Technologies, Graduate School of Systems and Information Engineering, University of Tsukuba

{hicham,kameda,mukaigaw,ohta}@image.esys.tsukuba.ac.jp

Abstract

We propose a new filming method of a moving object by fixed framing of multiple pan-tilt cameras. This filming approach assures the acquisition of high resolution images of the moving object without blurring images, 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 a monitoring role to one camera and a stand-by role to other cameras. While the monitoring camera is filming a target object, other cameras, that are in stand-by role, are directed so as 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. The prediction is calculated based on a marginal distribution obtained by the projection of a probability density function of the motion model onto the circumference of the image plane.

The validity of our method has been proven through experiments of switching videos of two to nine pan-tilt cameras.

1. Introduction

Automated filming of moving objects have been a key technology in various fields including visualization of sports, scene analysis, video surveillance, etc. Thanks to the improvement of computer processing power, storage capacities, and high speed network access of broadband technologies, 3D visualization is becoming a focal point in image media. This has increased the needs for high resolution and sharp textures of moving and non-moving objects. Moving objects are in particular good targets for filming, especially in sport broadcasting where sharp and high resolution images of players can help for watching player actions.

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

In some advanced automatic videography methods [1][2], a pan-tilt camera is being rotated all through the tracking process. As a result, the motion of the camera is changed frequently. This frequent motion adjustment causes a serious problem of video quality for filming human activities. It may also results in blurring of their textures. This problem becomes more serious in the case of tracking a soccer player who may move at a high speed and perform a sudden change in the speed and the direction. This imposes high speed and erratic rotation of the camera, which makes acquiring sharp textures of the targets extremely hard.

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 a comfortable video to the viewers.

Ozeki et al.[5] proposed a virtual frame control 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 the object tracking is activated just after the object goes beyond the virtual frame. When the object comes to settle down, the camera will be locked again. This method works well in the case of tracking a manipulated object, however it does not support high speed motion such as the case of a soccer player running after a ball in a soccer field.

In this paper, we propose a new method to control multiple pan-tilt cameras in order to realize filming and acquiring sharp and high resolution textures of an object moving in a wide field. 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 is going to be transmitted from the camera.

2. Camera Switching Concept

In this paper we assume all pan-tilt cameras are set at one place which is sufficiently far from the field where the target moves around.

We consider a set of m pan-tilt cameras zooming in. We use one camera for filming one object. We call this camera a monitoring camera, and other $m-1$ cameras 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 wait for the object to come in their field of view. Since there are m

cameras, stand-by cameras can check only $m-1$ regions. In order to estimate the frame-out probability "FOP" of each region, the moving direction of the object is estimated based on the apparent motion on the monitoring camera. FOP is estimated for every possible regions around the FOVM, and $m-1$ best regions are covered by the stand-by cameras. When the object moves out of the FOVM and enters into the field of view of one of the stand-by cameras, the two roles are switched between the two cameras.

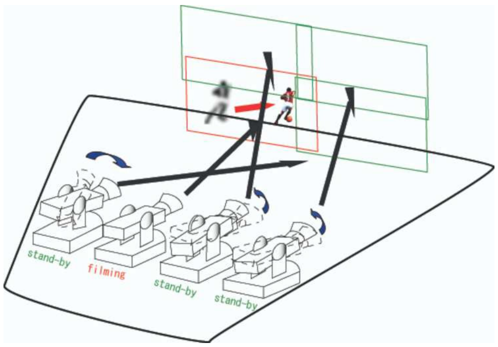

Figure 1: Directing and switching cameras.

3. Optimal Placement of Stand-by cameras

In order to estimate FOP, we utilize a probabilistic motion model of the target object. FOP is calculated by projecting the model onto the circumference of the FOVM. In this section, we propose generic form of FOP. Then we describe the implementation of FOP.

3.1 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 3), 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 which is at C_p, C_v and which is going to get out from the FOVM at the 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:

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

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

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

3.2 Frame-out Probability Function

The probabilistic model $f()$ can be formulated in various ways. In this paper, we formulate the motion model based on the current moving direction of the object. The motion model is projected on the circumference of the FOVM to estimate FOP for each zone.

We use the counterclockwise azimuthal coordinates as illustrated in (Figure 2). Let θ be the azimuthal angle and motion direction of the target be $\theta = 0$.

Suppose there is a high possibility that the object keeps moving in the same current direction and the further the object deviates from the current direction the smaller the probability becomes. The model distribution will look like the distribution illustrated in Figure 2. 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 3).

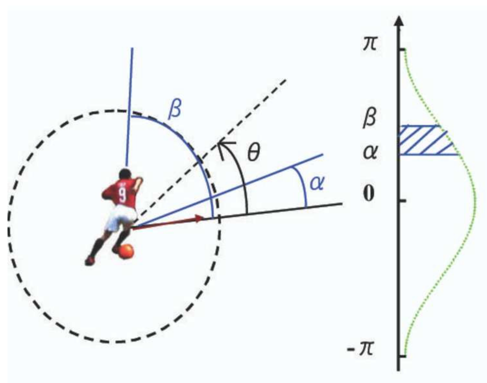

Figure 2: Azmuthal distribution.

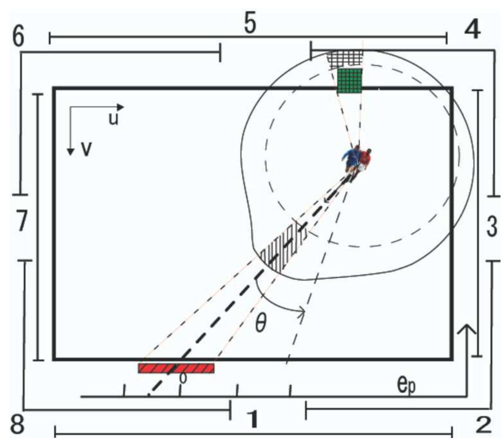

Figure 3: Azimuthal distribution, its projection on the image circumference, and the eight zones to place stand-by cameras.

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

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

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

$$f_c(\theta) = \frac{cos(\theta)}{a} + \frac{1}{2\pi} \quad (a > 2\pi) \qquad (3)$$

The second term in equation (2) is added to normalize the proposed probabilistic density function $f(\theta)$. Since $\theta \in [-\pi, \pi]$ and not $[-\infty, \infty]$, the second term is needed to fulfill the condition $\int_{-\pi}^{\pi} f(\theta) d\theta = 1$.

In order to discuss the shapes of $f_g(\theta)$ and $f_c(\theta)$ in detail, we set the image size as (640×480), the target position at $C_p(u = 500, v = 150)$, and the velocity at $C_v(-1, 1)$. We divide the circumference of the image plane into eight zones (Figure 3) and we investigate frame-out probability at each zone.

In the azimuthal coordinates, each of the probability density functions (Figure 4) has a peak at the direction $\theta = 0$. When these density functions are projected onto the circumference of the image, the distributions change their form into new distributions which we call marginal distributions.

We adopt the marginal distribution as the frame-out probability model of the object.

We examined three types of distributions: $f_g(\theta)$ with $\sigma = 0.5$, $\sigma = 3.0$,and $f_c(\theta)$ with $a = 7.0$

- $f_g(\theta)$, $\sigma = 0.5$. The variance of this distribution is small, therefore, in the neighborhood of the object moving direction, the frame-out probability is high. As a result, the marginal distribution has its maximum value around $e_p = 0$. This is illustrated in the upper graph(sigma=0.5) of Figure 4. In the marginal distribution illustrated in the graph in the middle of Figure 4 and the upper graph of Figure 5, the zone 1 has the highest probability. If we have only one stand-by camera, the best direction will be zone 1. In the case we can use more stand-by cameras, they will be directed in the order: zone 1, zone 8, zone 7 and so on.

- $f_g(\theta)$, $\sigma = 3.0$. In the azimuthal coordinates the probability distribution has its peak at $\theta = 0$ as illustrated in the upper graph(sigma=3.0) of Figure 4. On the other hand, marginal distribution has its high peaks at zone 4 and zone 3 (graph corresponding to sigma=3.0 in middle of Figure 4). Moreover, from frame-out probability calculation results (Figure 5), we can conclude that zone 4 represents the maximum probability and therefore it is the best zone to place a stand-by camera. If we have more than one stand-by camera, the priority of stand-by directions will be in the order of zone 4, 5, 3, and so on.

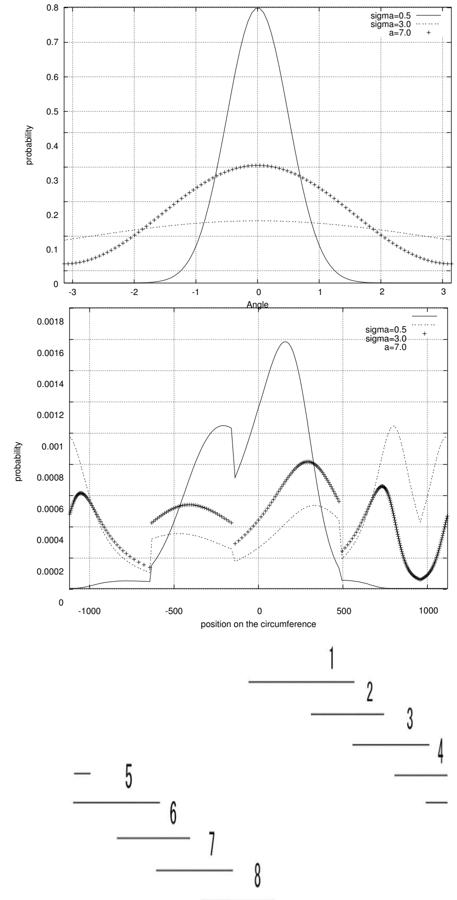

Figure 4: Probability density functions: the first figure illustrates the azimuthal distribution, the figure in the middle represents marginal distribution and the third figure represents the range of the eight zones.

- $f_c(\theta)$, $a = 7.0$. This distribution behaves like $f_g(\theta)$ at $\sigma = 0.5$ as far as the top priority zone is concerned (zone 1), but it prioritizes zone 2 which is closer to the target over zone 8 as shown in the lower graph of Figure 5 and the middle graph of Figure 4. When a is increased $g(\theta)$ tends to behave like $f(\theta)$ at $\sigma = 3.0$.

The best candidate zones may have similar probability values. If there are fewer cameras than the number of candidates we may exchange the placing order depending on other factors such as the camera response time etc.

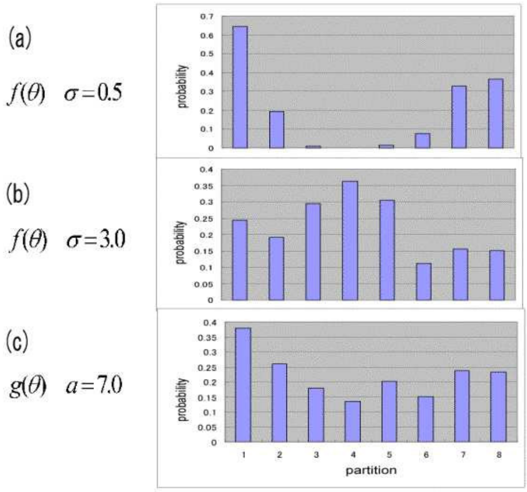

Figure 5: Frame-out probability at each zone.

4. Experiments

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

Our system exploits nine Sony EVI-D100 pan-tilt cameras at max. 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. The cameras are controlled using VISCA protocol over RS-232C cables and all the cables are connected to a USB hub through a serial-USB converter.

Object detection is performed using temporal differencing, and stand-by cameras are directed based on our frame-out probability model described in the above sections. Figure 6 and Figure 7 illustrates the results of our filming using four pan-tilt cameras. The monitoring camera which frames the player inside its FOV is indicated by a red rectangle, and other cameras are standing-by for future coming of the player. Our fixed framing result was successfull, and current implementation runs at 29 fps.

5. Conclusion

We have presented a novel approach for filming a moving object by switching temporarily locked multiple pan-tilt cameras to realize fixed filming. Switching is performed by assigning a monitoring role to one camera, and a stand-by role to other cameras. Stand-by camera placement is determined based on the probabilistic motion model of the moving object. We have implemented a system that has more than four cameras based on the proposed method. The experimental results have proven its efficiency.

Several issues have yet to be addressed such as target object's texture extraction, and expanding the frame-out probability model based on other criteria.

Figure 6: Tracking result 1, each of stand-by-cameras take a position and wait for the target

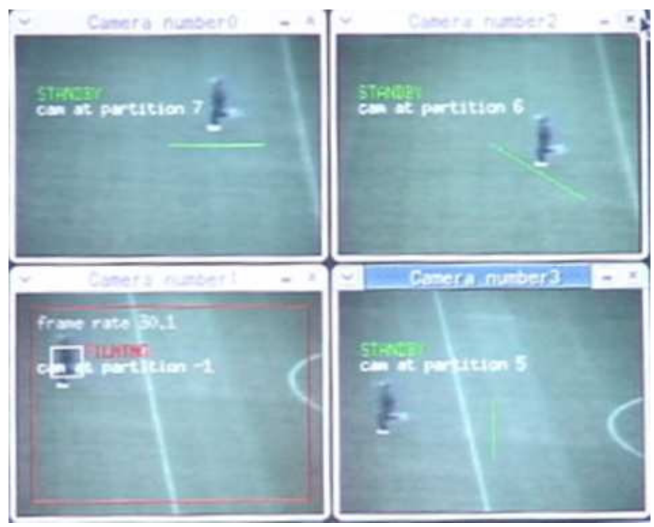

Figure 7: Tracking result 2, each of stand-by-cameras take a position and wait for the target

References

[1] 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.

[2] M.Greiffenhagen et al, "Statistical modeling and performance characterization of a real-time dual camera surveillance system," CVPR, pp.2335-2342, 2000.

[3] T. Koyama et al.: "Live Mixed-Reality 3D Video in Soccer Stadium," ISMAR, pp.178-187, 2003.

[4] Y. Kameda et al: "Reduction of Camera Motion Adjustments Under a Planned Video Composition with Pan-Tilt Camera," ACCV, Vol.1, pp.216-221, 2004.

[5] M.Ozeki et al: "Camerawork for intelligent video production -capturing desktop manipulations," ICME, pp.41-44, 2001.