動的シーンの可視化のための シナリオに基づく協調的カメラワークの生成

東海 彰吾　　松山 隆司
京都大学 大学院 情報学研究科

概要: 動的なシーンを映像として人間に提示する場合、シーン内で起こる事柄を撮影し提示するだけでなく、シーン内の状況が「分かりやすい映像」や「見飽きない映像」が望まれる。このため、本研究では、シナリオなどの事前知識に基づいてシーン内で起こる事象をカメラワーク込みの映像として提示するための、カメラワークの計画および実行を行うシステムについて考える。本研究で提案するシステムは、シーンと最終的な映像に関する事前知識を入力として、カメラ配置や協調的制御などを含めたカメラワークを立案し、さらに、シーンの状況と事前知識との間にズレが生じた場合にもシーンの状況に適応して映像の撮影および提示を行うものである。本文では、システム全体の構成とその処理の概要を述べ、システムの基本機能の例として構築した２種類のプロトタイプシステムとその実験例について説明する。

Scenario-based Cooperative Camera-work Planning for Dynamic Scene Visualization

Shogo Tokai and Takashi Matsuyama
Graduate School of Informatics, Kyoto University

Abstract: In this paper, we propose a framework of a system to visualize dynamic scene situations as intelligible and attractive image sequences for human viewers. To make these image sequences, it is necessary to use camera-work techniques adaptively to the situations. In our system, camera-work are planned based on prior knowledge about the dynamic scene, fabrications of final sequences and know-how about camera-work. This plan includes about camera-layout, cooperative camera-controls, switching and so on. However, the prior knowledge has several deviation practically. Then, this system has one more planning step and works adaptively. In this paper, we also explain two prototyping examples of the system functions to make clear that this framework is effective to visualize the scene.

1 Camera-Work for Intelligible and Attractive Scene Visualization

Most of active vision systems developed so far including the ones described in [1] [2] [3] capture images to control cameras and understand scene structures. This paper, on the other hand, addresses active camera control methods for dynamic scene visualization. There exists a large difference between these two tasks; while the former throws away observed image data after processing, the latter puts its focus upon how we can fabricate image sequences intelligible and attractive for human viewers.

Here, "intelligible" implies that viewers should be able to understand global / dynamic scene contexts from limited sequences of captured images. "Attractive" means fabricated image sequences should keep attracting viewers' interest without being felt tired or boring. If possible, moreover, they should be artistic.

As discussed in [1] [2] [3], our research group, Cooperative Distributed Vision (CDV, in short) project, offers a fundamental framework for scene visualization as well as scene understanding. To realize versatile scene visualization, we have to solve the following problems:

Camera Layout : How many and where should we put a group of cameras?

Dynamic Camera Control : How should we control camera parameters dynamically?

Image Sequence Fabrication : How should we fabricate intelligible and attractive image sequence(s) from raw image data observed by the cameras?

By *camera-work planning* we mean methods to solve these problems.

E-mail: {tokai|tm}@i.kyoto-u.ac.jp

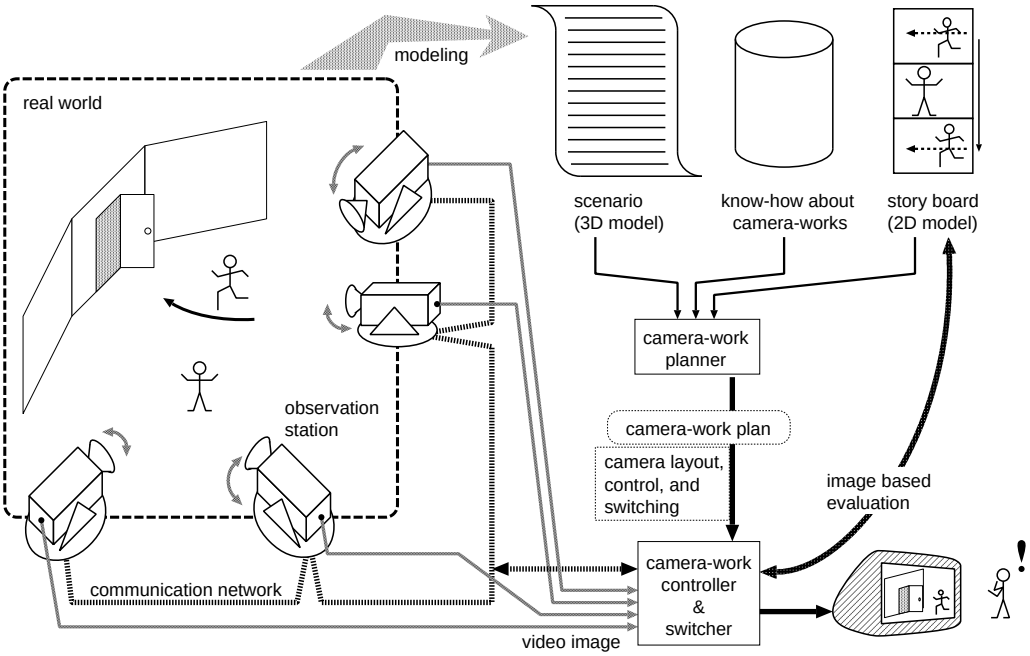

Figure 1: Framework of the scenario-based cooperative camera-work planning for dynamic scene visualization.

Since the real world includes a wide spectrum of dynamic scenes and moreover, the intelligibility and attractivity are too abstract to define computationally, it is almost impossible to attain the meaningful camera-work planning without knowledge. The following three types of knowledge can be used for the camera-work planning (see Figure 1):

Scenario Description : This specifies semantics and physical structures of the scene as well as dynamic events involved in the scene.

Story-Board Description : This specifies a group of characteristic snapshots in the image sequence(s) to be fabricated. That is, it defines the intelligibility and attractivity to be realized by the camera-work.

Know-How about Camera-Work : Many effective camera-work techiniques have been developed in cinematography[4]. They include a variety of camera layout, image framing, and camera switching techniques. We can use such know-how for the camera-work planning.

Note that the first one is described in terms of abstract semantic and/or 3D physical scene features, the second 2D image appearances taking into account psychological effects onto human viewers, and the third includes transformation rules between them.

Camera-work planning systems incorporate these three types of knowledge to solve the above mentioned three problems. In general, the planning should be done in the following two stages:

Off-Line Planning : Given a scenario description to be visualized, the system first makes a camera-work *plan* based on the knowledge.

On-Line Camera Control : Since the scenario is just a rough model of the real world scene, real world situations usually deviate from the scenario. Thus, on-line adaptive camera controls should be conducted during the scene visualization process. CDV systems such as those described in [1] [2] [3] support such on-line adaptive camera controls.

Figure 1 summarizes the framework for the dynamic scene visualization discussed above.

In this paper, we describe a scenario-based dynamic scene visualization system being developed in the CDV project, where major emphasis is put upon dynamic cooperation between distributed cameras (i.e. observation stations). That is, we believe that to fabricate intelligible and attractive image sequence(s) from those observed by the cameras, flexible inter-camera coordinations are required as well as individual dynamic camera controls.

2 System Organization

Here we describe specifications of each component of the scenario-based scene visualization system in Figure 1.

2.1 Knowledge Sources

As discussed before, three types of knowledge is give to the system:

1. Scenario Description There have been proposed several scenario/camera-work description methods and camera-work planning systems [6], [5], [7]. In [5], Christianson et al proposed the Declarative Camera Control Language, with which various types of camera-work patterns can be described. While the camera-work patterns can be used for the off-line planning, no mechanism is supported for the on-line camera control. In [6] and [7], on the other hand, on-line dynamic camera-work/interaction control methods are proposed. [6] used a state transition network to specify dynamic camera control and switching but they does not consider any cooperations between cameras which is needed to realize complicated camera-work in real world. [7] proposed a scripting method for interactive systems based on Allen's temporal interval algebra [8].

In our system, a scenario is described by an *event graph* (Figure 2), where each node denotes an event representing the dynamic 3D model of a real world scene and an arc a temporal/geometric/semantic relation between events. The simplest but most popular event graph is a series of event nodes connected by a chain of directed arcs denoting the temporal order (i.e. B-A arcs in Figure 2). Various types of semantic arcs, such as retrospection, hearsay, and illusion, may be used to enrich scenario contents.

An event node includes:

- Semantic Scene Features: type of the event and/or atmosphere of the scene, e.g. fighting, thrilling adventure, happy dining, solemn ceremony and so on.
- Background Scene Characteristics: overall geometric and illumination structures of the scene and their dynamic variations: e.g. soccer field, crowded downtown, conference room, and so on.
- Foreground Object Characteristics: attributes and dynamics of objects requiring focused imaging. Sometimes mental features and moods of objects may be associated with physical characteristics. For example, a tall man in a red shirt rushes out through the door crying loudly.

2. Story-Board Description This is described by a series of 2D sketches specifying how each *shot* in the finally fabricated image sequence looks like (Figure 1). That is, it is the goal specification for the scene visualization. It contains the information about viewing angle, image framing, camera position, motion, and switching.

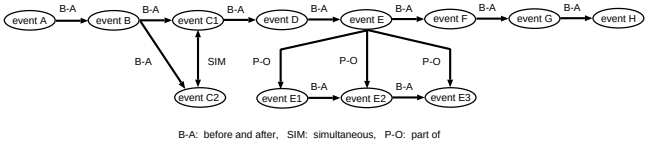

Figure 2: Event graph.

In addition, each sketch is associated with an event ID(s) in the event graph. Note that in general, associations between events and sketches are M : N. That is, it is very popular that an event is visualized by a series of shots taken from different cameras. We call a continuous video sequence taken by a camera *physical shot*. On the other hand, a single image frame sometimes includes multiple physical shots representing different events: e.g. a group of scientists in a conference room are discussing about the earth looking a video image transmitted from a satellite. We call such composite shot *logical shot*. That is, each sketch specifies characteristics of a logical shot, while an event corresponds to a physical scene.

3. Know-How about Camera-Work Since the story-board is just a rough and abstract goal specification, we need additional knowledge to attain the goal under the given scenario description. Know-how about camera-work specify heuristic rules to take intelligible and attractive image sequences under various scenario situations. They include rules for camera layout, dynamic camera control, and switching. Note that camera control and switching rules specify not only actions of each individual camera but also coordination methods among distributed cameras.

2.2 Camera-Work Planning

Off-Line Planning Given three knowledge sources described above, the camera-work planner (Figure 1) reasons about effective camera-work for the scene visualization.

1. First, for each event in a given scenario, the planner determines geometric camera layout, dynamic camera action, and temporal camera switching and coordination. Since there exist many different possible camera-work to visualize a given event, the planner uses sketches in a given story-board to select the most effective camera-work rule. Note that the camera-work plan generated at this stage specifies physical shots obtained by the cameras placed in the scene.

2. Then, the planner determines an image composition plan to fabricate logical shots specified in the story-board. Note that while most of logical

shot compositions are realized by 2D image processing, virtual images may be synthesized based on the 3D scene information restored from multi-viewpoints image sequences. Note also that the planner should make camera coordination plans across multiple events to generate well synchronized/organized logical shots.

On-Line Camera Control After placing a group of cameras according to the designed camera layout, the camera-work and logical shot composition plans are loaded onto a group of observation stations and the camera-work controller & switcher respectively (Figure 1). Then,

1. The cameras stand by and objects in the real world start the actions specified in the scenario.

2. Each observation station captures an image sequence by controlling camera parameters according to the camera-work plan. The acquired image sequence is delivered to the controller & planner. As noted before, each observation station should adaptively control its camera since the scene usually deviates spatially and temporally from the plan. Moreover, multiple observation stations should cooperate with each other through communications to control their cameras. These adaptive and cooperative camera controls are realized using sketches in the story-board as goal specifications.

3. In the camera-work controller & switcher , a series of logical/physical shots are fabricated from a group of raw image sequences captured by observation stations. Note that the camera-work controller & switcher itself may generate virtual image sequences based on the 3D scene information restored from multi-viewpoints image sequences. Thus, it should dynamically communicate with observation stations to realize smoothly connected and/or well synchronized logical shots. The smoothness and synchronization are evaluated at the 2D image level referring the story-board (Figure 1). In this sense, we may call it a director and/or a composer.

3 Prototype System

Currently we are developing a prototype system based on the framework proposed above. Here, we show two simulation results of the camera-work planning: (1) camera layout for 2D static scenes including multiple objects and obstacles and (2) scenario-based camera control and switching in 3D dynamic scenes. Simulation results demonstrate that our approach is very promising.

3.1 Planning Layout of Multiple Cameras in 2D Static Scenes

Where to place a group of cameras is one of major problems to be solved in the off-line camera-work planning. We developed an optimization method for the camera layout.

First we assume the followings:

- The scene is two dimensional and static.
- The background scene is defined as a rectangular area, in which foreground objects, obstacles, and cameras are placed (Figure 3(a)).
- An object is represented by a circle with a specific "front face" direction (Figure 3(a)). Each point on the circular object surface is associated with a weight representing the importance for the visualization. In the current simulation, we used the following function to model the weight distribution over the surface:

$$Importance(\alpha) = \frac{1+\cos(\alpha)}{2}, \qquad (1)$$

  where α denotes the angular distance from the front face direction.
- An camera is modeled by a projection center with a fixed viewing angle (i.e. fixed zoom, Figure 3(b)(c)). Note that this angle specifies the size of the image frame (i.e. area covered by an image). In addition, each camera is associated with a list of foreground objects to be imaged.

First, the size and shape of the background scene, locations and characteristics of objects and obstacles, and the number and viewing angles of cameras are given to the camera layout planner. Then, the position and viewing direction of each camera, (x, y, θ), is determined by optimizing the following evaluation function.

$$E_{total}(x,y,\theta) = \sum_{i \in Object-List} \{E^i_{visibility}(x,y,\theta) \times E^i_{position}(x,y,\theta) \times E^i_{size}(x,y,\theta)\}, \qquad (2)$$

where $Object - List$ denotes a list of the objects to be imaged by the camera.

Each component evaluation function is defined for an object-camera pair as follows:

$E_{visibility}$**: Object Surface Visibility** This evaluates how well the object looks in the captured image. We used the following function (see Figure 3(a)):

$$E_{visivility}(x,y,\theta) = \int_{visible} Importance(\alpha)\ \cos\beta\ \cos\alpha\ d\alpha, \qquad (3)$$

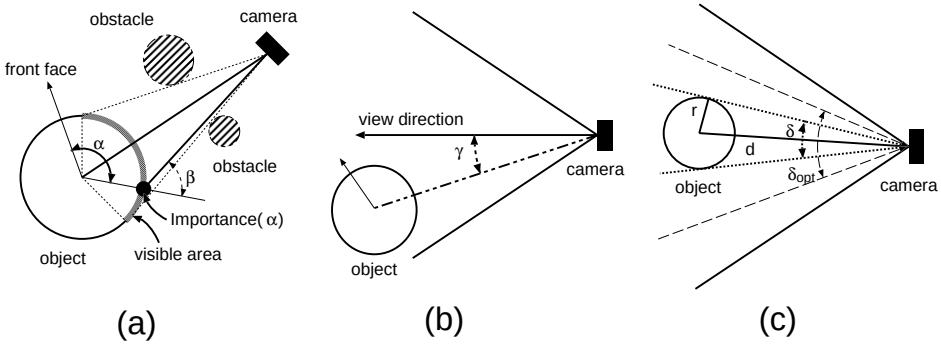

Figure 3: Evaluation functions for the camera layout.

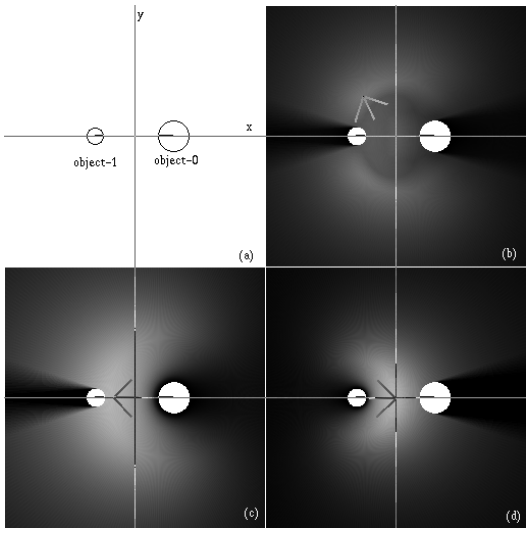

Figure 4: Optimal camera layouts: (a) a pair of objects to be visualized, (b) optimal camera layout for simultaneous imaging of the objects, (c),(d) optimal camera layouts for object-0 and object-1, respectively.

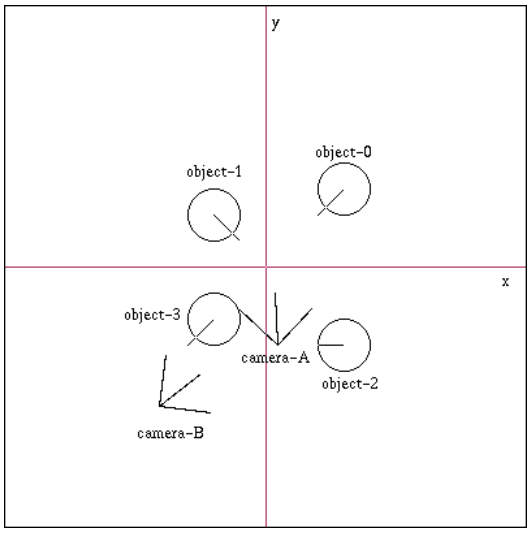

Figure 5: Optimal layout of a pair of cameras for four objects.

where α denotes the angular distance of a surface point from the front face direction and β the angle between the surface normal at that point and the view direction from the camera. The integral covers only those surface points that can be seen from the camera without being interfered by obstacles.

$E_{position}$**: Object Position in the Image** We assume the object is to be captured at the center of the image. Then, the following function evaluates the goodness of the object position (see Figure 3(b)):

$$E_{position}(x, y, \theta) = \frac{1}{2}(1 + \cos\gamma), \tag{4}$$

where γ denotes the angle between the object center direction from the camera and the view direction of the camera.

E_{size}**: Object Size in the Image** The object size in the captured image is also an important factor in visualization. We assume that a certain optimal object size is specified in the story-board. Then the following function evaluates the goodness of the object size (see Figure 3(c)):

$$E_{size}(x,y,\theta) = \begin{cases} \frac{1}{2}(1+\cos\frac{\delta-\delta_{opt}}{\delta_{opt}}\pi) & \text{if } \delta \le \delta_{opt} \\ \frac{1}{2}(1+\cos\frac{\delta-\delta_{opt}}{\pi-\delta_{opt}}\pi) & \text{if } \delta_{opt} < \delta \le \pi \end{cases} \tag{5}$$

where $\delta = 2\sin^{-1}(r/d)$, r denotes the radius of the object, d the distance between the camera and the object, and δ_{opt} the pre-specified optimal size parameter.

We conducted several simulations to examine the

effectiveness of the above mentioned camera layout method. Figure 4(a) illustrates the geometric configuration of a pair of objects to be visualized. Figure 4(b) shows (1) the spatial distribution of $E_{total}(x, y, \theta)$ and (2) the optimal camera position and its viewing direction when both objects are required to be imaged simultaneously by a single camera. To depict (1), the optimal view direction, θ^*, is first computed at each position and $E_{total}(x, y, \theta^*)$ is encoded by the gray level: the brighter the gray level is, the higher value the evaluation function takes. (2) is depicted by a group of three line segments: their intersection point denotes the camera position, the central segment the view direction, and the pair of marginal ones the viewing angle. Figures 4(c) and (d) show the optimal camera layouts when each object is required to be imaged by a single camera, respectively.

Figure 5 illustrates the optimal layout of a pair of cameras when camera-A and camera-B are used for imaging {object-0, object-1} and {object-2, object-3}, respectively, where $\{\cdot\}$ denotes the list of objects to be imaged by a camera.

While these simulation results are simple and include many points to be improved, we believe they showed practical utilities of our framework. Currently we are developing a novel camera layout method which utilizes the story-board as the evaluation function.

3.2 Dynamic Camera Coordination for Smooth Camera Switching

Here we demonstrate the importance of the on-line coordinated camera control and switching in visualizing dynamic scenes.

Suppose a scenario description specifies that "A man is running along the long straight path at the constant speed." and the story-board requires that his zoomed-up face should be captured continuouly since changes of his facial expressions are the crucial factor for visualizing the scene.

Based on these knowledge sources, the camera-work plan illustrated in Figure 6 is generated at the off-line planning stage. The plan specifies (1) a pair of cameras are placed at the same side along the path, (2) each camera tracks the face by dynamically rotating the view direction*, and (3) the image sequence taken by camera-1 should be switched to that taken by camera-2 when both image sequences can be smoothly connected. Here we assume the smoothness is evaluated by the apparent face motion against the backgound scene in captured image sequences.

This camera-work plan is loaded onto a pair of observation stations and the camera-work controller & switcher in Figure 1. When the action in the scene

*For simplicity, we assume only the 2D panning is allowed for each camera.

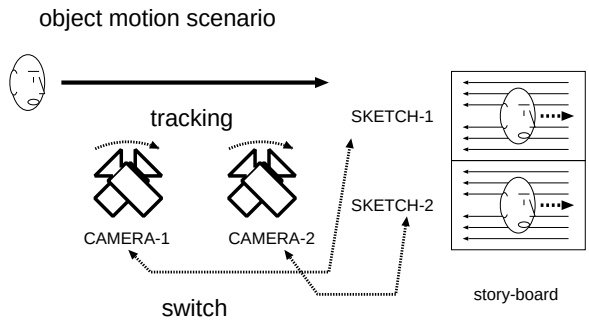

Figure 6: Camera-work plan for dynamic scene visualization.

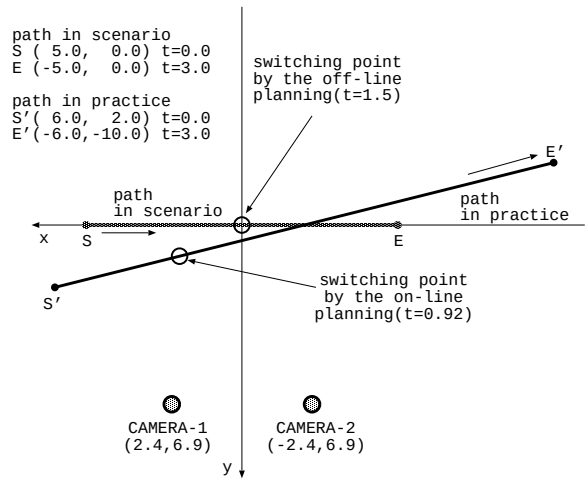

Figure 7: Deviation of object motion.

is started, the object detection and tracking process such as described in [2] [3] is executed at each observation station. Then, the camera-work controller & switcher monitors a pair of image sequences captured by the observation stations and determines the optimal camera switch timing.

As noted before, the actual scene usually deviates spatially and temporally from the scenario. Figure 7 illustrates the geometric configuration of the scene, the camera layout, and the object motion path described in the scenario. Here we assume that the actual object motion path deviates from the plan as shown in the figure. In what follows, we will demonstrate the importance of the on-line camera control in determining the optimal camera switch timing.

In the current simulation, the switch timing is evaluated by the difference in the camera rotation speed. The reason for this is as follows. Firstly, since both cameras are tracking the object, the object image stays fixed at the center of the image frame. Thus, human viewers perceive the object motion speed based on the optical flow of the background

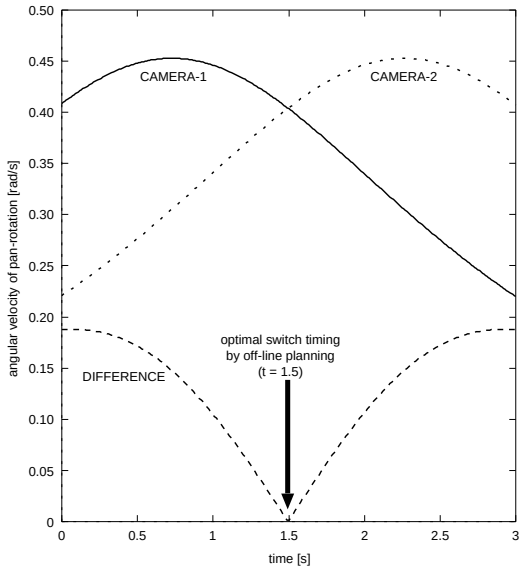

Figure 8: Optimal camera switch timing determined by the off-line camera-work planning.

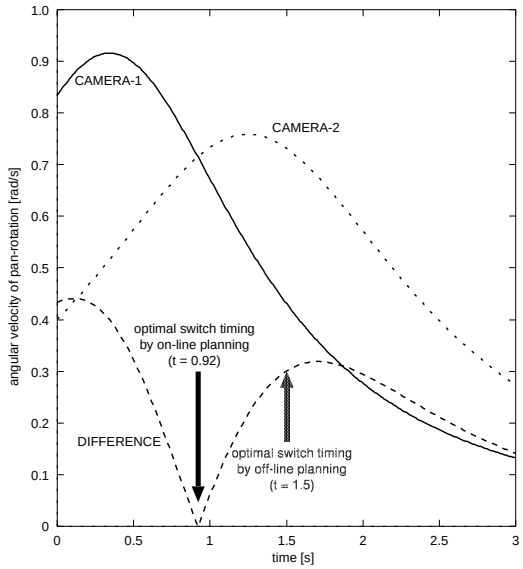

Figure 9: Optimal camera switch timing determined by the on-line camera control.

scene. Assuming the distance of the background scene from the cameras is constant, the camera rotation speed uniquely determines the strength of the optical flow. In other words, by switching the cameras when their rotation speeds coincide with each other, human viewers perceive the object as moving at the constant speed even if the camera is switched from one to the other. Note that to realize more smooth camera switching, we should control the zoom so as to make the object sizes in the pair of captured image sequences coincide.

Figure 8 illustrates temporal variations of the rotation speeds of camera-1 and camera-2 when they are tracking the object along the path specified in the scenario. The optimal camera switch timing is determined as $t = 1.5$ sec and the object location at that time is shown in Figure 7. Figure 10(a) shows the image sequence fabricated from the pair of image sequences taken by camera-1 and camera-2, assuming the object moves as specified in the scenario and the camera is switched at $t = 1.5$ sec.

If we directly applied this planned camera-work to the actual scene, we would obtain such a meaningless image sequence as shown in Figure 10(b), which demonstrates the necessity of the on-line adaptive camera control.

Figure 9 illustrates temporal variations of the rotation speeds of camera-1 and camera-2 when they are adaptively tracking the actual object motion shown in Figure 7. The camera-work controller & switcher dynamically monitors these camera motion speeds and switches the cameras at $t = 0.92$ sec (see Figure 9). Figure 10(c) illustrates the image sequence fabricated by this on-line adaptive camera control and switching method, where the soomthly connected image sequence is fabricated.

4 Discussions

In this paper we proposed a framework of senario-based cooperative camera-work planning for dynamic scene visualization. Its novel features are

- Introduction of three types of knowledge sources: scenario, know-how about camera-work, and story-borad.
- Off-line camera-work planning followed by on-line dynamic camera control and switching.
- Cooperation among distributed active cameras (i.e. observation stations) to adaptively capture intelligible and attractive image sequences.
- Logical and virtual image shots fabrication from multi-viewpoint image sequences.

While we have shown practical utilities of our approach with several simulations, the following technical developments are required to implement a scene visualization system that can work in real world scenes.

- Description languages for the knowledge sources and the camera-work plan
- Knowledge-based camera layout and dynamic camera-work planning for 3D dynamic scenes
- Plan-guided dynamic camera control for scene/object visualization

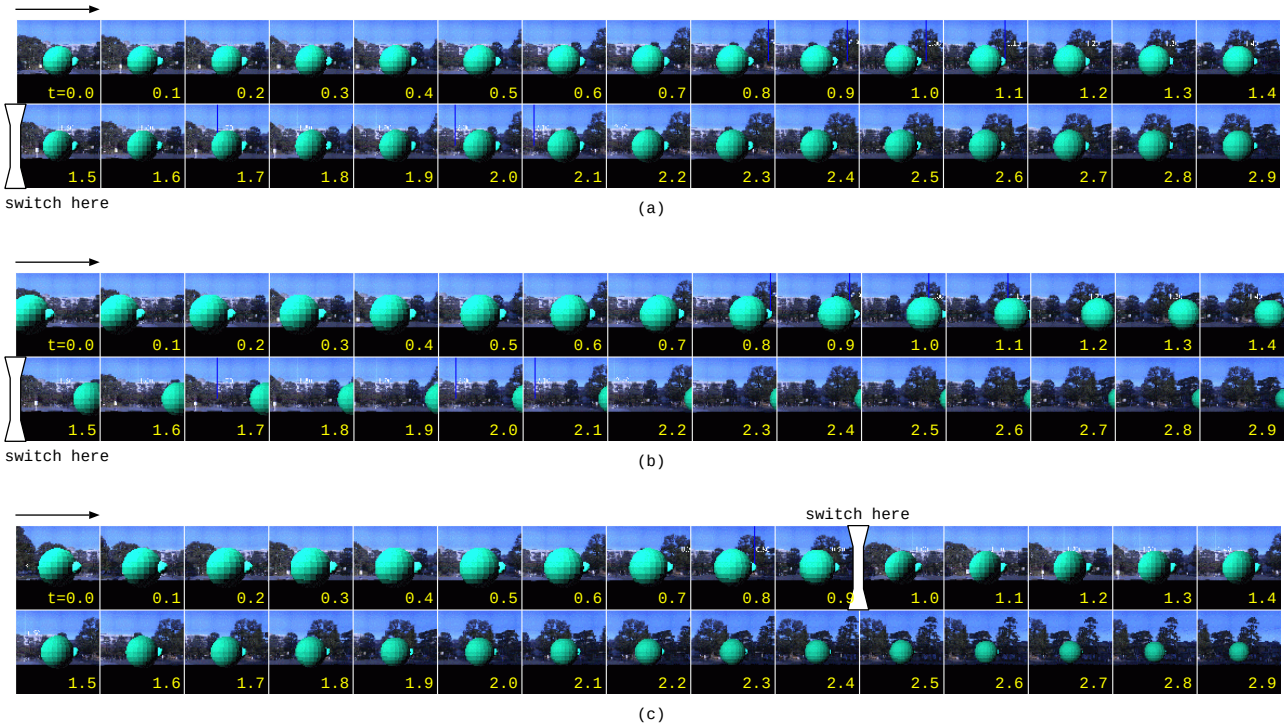

Figure 10: Fabricated image sequences. Each sequence starts at the left of the upper row followed by the lower row including a mark denoting the point of the camera switching.

- Dynamic cooperation protocols for well organized/synchronized multi-viewpoint visualization
- Image sequence switching and virtual image generation for intelligible and attractive image sequence fabrication
- Computational method of evaluating intelligibility and attractivity.

References

[1] Matsuyama, T.: Cooperative Distributed Vision – Embodying Intelligence Through Integration of Visual Perception, Action, and Communication –, Proceedings of MIRU '98, pp.I-343–352, 1998 (in Japanese).

[2] Matsuyama, T. and Wada, T.: Temporal Coordination between Perception and Action for an Active Vision System: Real-Time Object Detection and Tracking with a Fixed Viewpoint Pan-Tilt-Zoom Camera, Proceedings of MIRU '98, pp.I-359–394, 1998 (in Japanese).

[3] Matsuyama, T., Wada, T., and Maruyama, M.: Cooperative Object Tracking by Multiple Active Vision Agents, Proceedings of MIRU '98, pp.I-365–370, 1998 (in Japanese).

[4] Arijon, D.: Grammar of the Film Language, Focal Press Ltd., London, 1976

[5] Christianson, D.B., Anderson, S.E., He, L., Weld, D.S., Cohen, M.F., and Salesin, D.H.: Declarative Camera Control for Automatic Cinematography, Proceedings of AAAI '96, pp.148-155, 1996.

[6] He, L., Cohen, M.F., and Salesin, D.H.: The Virtual Cinematographer: A Paradigm for Automatic Real-Time Camera Control and Directing, SIGGRAPH'96, pp.217-224, 1996.

[7] Mase, K., Pinhanez, C.S., and Bobick, A.F.: Scripting Method Based on Temporal Intervals for Designing Interactive Systems, Trans. of IPSJ, Vol.39, No.5, pp.1403-1413, 1998 (in Japanese).

[8] Allen, J.F.: Towards a General Theory of Action and Time, Artificial Intelligence, Vol.23, pp.123-154, 1984.