15

SKIMMING MULTIPLE PERSPECTIVE VIDEO USING TEMPO-SPATIAL IMPORTANCE MEASURES

Toshihiko Hata[†] [†††] Tatsuo Hirose[††] Katsumi Tanaka[†††]

[†] *Industrial Electronics & Systems Laboratory, Mitsubishi Electric Corporation. 8-1-1 Tsukaguchi-Honmachi, Amagasaki, Hyogo 661-8661, Japan*

hata@img.sdl.melco.co.jp

[††] *Division of Computer and Systems Engineering, Graduate School of Science and Technology, Kobe University. 1-1, Rokkodai, Nada, Kobe, Hyogo 657-8501, Japan*

hirose@db.cs.kobe-u.ac.jp

[†††] *Division of Information and Media Sciences, Graduate School of Science and Technology, Kobe University. 1-1, Rokkodai, Nada, Kobe, Hyogo 657-8501, Japan*

tanaka@db.cs.kobe-u.ac.jp

Abstract Multiple perspective video taken by more than several tens of cameras are observed and stored for applications such as video surveillance and outside broadcasting. It is one of the most important demands to grasp what happens in the area from the multiple perspective video in a short time, but there are some problems to be solved. For example, we cannot look at a lot of video simultaneously and it is difficult to understand the entire state of a phenomenon that occurs in a broad area and is sparsely taken by a plural cameras. In this paper, we propose a new skimming method using tempo-spatial importance measures. At first, the video importance is calculated using the importance of elements captured in the video scene. The elements are objects such as buildings and cars, and events such as a temperature rising over and a batter hitting. They have their own importance based on space, time and semantics. Then some video are selected based on their importance and displayed with a map and three-dimensional graphics of the objects for skimming the multiple perspective video effectively. We discuss the tempo-spatial importance of multiple perspective video, the video importance calculation and the display methods based on the importance. We also describe our prototype and some experimental results.

Keywords: video database, skimming, multiple perspective video, tempo-spatial

H. Arisawa et al. (eds.), *Advances in Visual Information Management*

1. INTRODUCTION

Recently, much attention has been focused on archiving, retrieving and delivering of multiple perspective video data because of rapid advances in digital video technologies. The multiple perspective video means a collection of mutually-synchronized video data taken by multiple video cameras. There are many applications of the multiple perspective video such as a building security surveillance, a plant monitoring, and a TV broadcasting of a sport event etc. Video data of a camera are digitized and stored in archival devices. In some applications such as building security surveillance, not only video data but also their metadata (such as camera position data, sensory device data etc.) can be acquired and stored in a synchronized manner.(Tsukada et al., 1997)

In considering applications of those multiple perspective video, one of the most important research issues is how to provide a mechanism to present the vast volume of video data taken by several tens of cameras to users. It seems to be difficult for human beings to see and recognize the contents of multiple numbers of video taken simultaneously. Ramesh Jain et al.(Katkere et al., 1996) proposed a system that the video intervals which capture the walker extracted by image recognition techniques can be queried and retrieved from multiple perspective video data. However, the system is not available for various applications because the query is only for already-known moving objects and ad hoc. In addition, it is not impossible to grasp what happens in the area captured by a lot of cameras in a short time.

It is effective for browsing and displaying the multiple perspective video to be able to understand intuitively where a video clip or a video frame is taken. There are video browsing methods for that purpose that utilize camera field data such as camera's position and direction given by some sensors like a GPS (global positioning system) and a remote controllable camera. Arikawa (Arikawa 1998) proposed a system in which the camera icons that represent each video clip have their directions and are displayed at their positions in 3D virtual world. Joachim Sauter et al.(Sauter et al.,) proposed a way to display each video frame at its camera position and in its camera direction. But they are methods for a moving camera and not for the multiple perspective video.

On the other hand, in the area of video database systems, much works have been done about 'video skimming', which is to summarize a whole video data based on the 'importance' of scenes of the video. For example, Kanade et al.(Wactlar et al., 1996) proposed a way to do the video skimming by the combination of speech recognition techniques and information retrieval techniques. However, most works on the video skimming have been concerned with skimming a 'single' video stream data. Our attention is focused on the way to do the skimming of multiple stream video data that are taken concurrently. In this case, there are several temporal or spatial relationships among cameras, and so, we propose a way to skim multiple perspective video by considering these temporal and/or spatial relationships.

In this paper, we propose a novel skimming method using tempo-spatial importance measures. At first, the video importance is calculated using the importance of elements captured in the video scene. The elements are objects such as buildings and cars, and events such as a temperature rising over and a batter hitting. They have their own importance based on space, time and semantics. Then some video are selected based on their importance and displayed with a map and three-dimensional graphics of the objects for skimming the multiple perspective video effectively.

The rest of this paper is organized as follows. Section 2 explains the multiple perspective video and its skimming. We discuss the tempo-spatial importance of multiple perspective video, the video importance calculation and the display methods based on the importance in Section 3 and Section 4. The prototype for evaluation and some experimental results are described in Section 5 and Section 6. Finally, we conclude this paper and indicate future work in Section 7.

2. MULTIPLE PERSPECTIVE VIDEO AND ITS SKIMMING

2.1. MULTIPLE PERSPECTIVE VIDEO

Multiple perspective video in this paper is live and stored video taken by more than several tens of cameras. Applications treating such multiple perspective video are video surveillance for building security, plant monitoring and super highway facilities management, broadcasting of sport and festival, and so on. Multiple perspective video has the following metadata. Those metadata describe 'when', 'where', 'why' the video are taken and 'what is captured in the video'. In addition, the metadata can be created automatically by various methods such as image sensing and speech recognition, and stored with the video data. Therefore the metadata are very useful for query and search.

- Time stamp : It specifies when the video is taken.
- Camera ID and camera field : They specify where the video is taken. The camera field consists of its position, direction, focus length and so forth. If the camera field and the position of an

object are known, it can be calculated by geometry whether the object is taken in the camera or not. Then we can know what is captured in the video by utilizing the camera field data and a geographic information system which manages objects such as buildings and facilities on a map. The camera filed can be get by some electronic equipment and devise such as a controllable camera, a GPS and a gyrocompass.

- Event : It specifies what happens in a object such as facilities and people. For example, a person intruding at night, two cars colliding at a crossing, a batter hitting a home run and so on. Event data consists of event ID, event type, time stamp and so on. There are various sensing devices on the market that create such events, for example, an optical sensor, a temperature sensor and a sound sensor. Image sensing technology is also often used practically for video monitoring and surveillance, and there are a lot of researches on this technology (Brill et al., 1998). Speech recognition technology is available to create some events from a telephone speech and a broadcast commentary.

- Annotation : It is data which a user annotates to the video for report and analysis. It is input as metadata by a keyboard and a speech recognition device.

Multiple perspective video has another feature, tempo-spatial sparseness. All of the whole area is not perfectly taken by multiple cameras but some parts of the whole area are taken sparsely in space as illustrated in figure 1. Moreover, all of the video are not recorded for a long time but some parts of the video, especially the important parts such as the latest interval and the interval around an event, are only stored.

A video server for the multiple perspective video such as the literature (Tsukada et al., 1997) captures video taken by a lot of cameras, stores the video data and the corresponding metadata, searches the appropriate video clips with the metadata, transmits and display them. We focus this search process and consider effective skimming methods for the multiple perspective video in this paper.

2.2. PROBLEMS OF ITS SKIMMING AND OUR APPROACH

There are various purposes to see the multiple perspective video such as live observation, event analysis and video editing, and we need some ways to get and see the necessary parts of video effectively for those purposes. It is one of the important demands to grasp what happens in

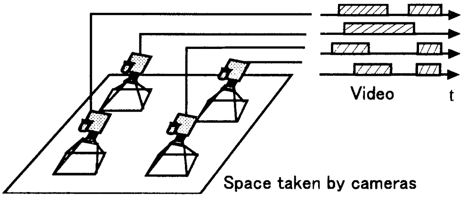


Figure 1 Tempo-Spatial Sparseness of Multiple Perspective Video

the area captured by a lot of cameras in a short time. For that purpose, we need a effective video skimming for multiple perspective video that extracts some important parts of video and display them. But there are several problems to be solved for such video skimming.

1 More than several tens of cameras

A multi-video display method is often used in which multiple video pictures are displayed simultaneously on a screen that is divided into n x n video windows such as 4 x 4. But we can see only several pictures at the same time and it is very difficult to look at precisely more than two pictures at the same time. Furthermore some unimportant video on the screen may obstruct to look at the important video. It is an essential problem how to select the important parts of video taken by more than several tens of cameras.

2 Intuitive understanding in space

We cannot intuitively understand where is taken in a video scene only by displaying the video picture such as the multi-video display method. It is more serious if there are a lot of cameras. Furthermore it is difficult to understand an event that occurs in a broad area and is taken sparsely in space and time by many cameras. We are interested in how to display a plural video in which several parts of the whole area are peeped in order to understand the entire event well.

3 Summary of events that occur for a long time

It is an important issue to summarize video data even for single stream video. But there are peculiar problems for multiple stream video because those video streams are related each other in time, space and semantics. We are also interested in how to summarize multiple perspective video and how to structure the summarized data that include the video parts and their metadata.

In this paper, we especially consider the former two problems for the applications that have a kind of locality on the importance of each camera in space and time. The locality means that most of the cameras are not important all the time, but the importance of each camera changes in turn and the number of cameras to be seen is limited at a time. At first in our approach, the importance of video is calculated using the importance of elements captured in the video scene. One type of the elements is an object such as facilities, buildings, people and cars, and another type is an event that is an action of object or a change of its state such as someone intruding, a temperature rising over and a batter hitting. The elements have their own importance based on space, time and semantics as described in section 3. The video importance is decided by summing up the importance of elements that is captured in the video scene. Then some video are selected based on their importance and displayed with a map and three-dimensional graphics of the objects for skimming multiple perspective video. Three display methods are described in section 4.

3. TEMPO-SPATIAL IMPORTANCE

As described before, the importance of video is decided by the importance of elements which are captured in the video scene. The importance of an element changes dependently on its position in space, the time when it is captured, and the relationship among elements. In this section, we consider the tempo-spatial importance of the elements and we describe the calculation of video importance.

3.1. TIME-DEPENDENT IMPORTANCE

We can divide the time-dependent importance broadly into two types.

1 Importance in a micro-view of time

An event is not only important at the moment when it occurs but also important for some time before and after it occurs. Then it has an importance distribution in time which origin is the time when it occurs. We can think of various kinds of distribution forms dependently on the kind of event and applications. To speak generally, an event is more important around the origin and is less far from the origin such as a normal distribution in figure 2 (a). Figure 2 (b) is available if the state of affairs is important before an event occurs. For example, we want to find a cause of the event, but the sensor is low sensitive and the cause was taken in the video a long time before the event occurred. Figure 2 (c) can be used if the state after an event occurs is important. For example, the cause of the event is known in advance and we want to see the state caused by the event well. We call this kind of importance as importance in a micro-view of time because we pay attention only to one event in a micro time interval.

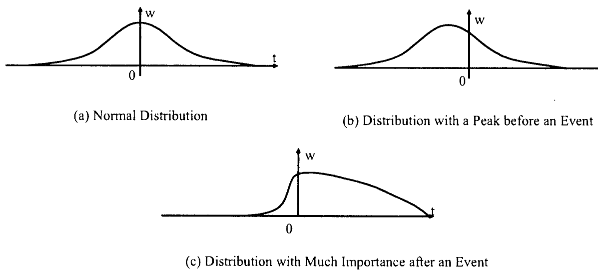


Figure 2 Examples of Event Importance Distribution

2 Importance in a macro-view of time

The event importance may change dependently on time attributes and temporal relationship among events as the following. We call this kind of importance as importance in a macro-view of time.

- Importance depend on time attributes

  The event importance may change with attributes of time such as a time period and a season. For example, an event of people walking in an office building is important at midnight but is not necessary in the daytime for security. Suppose that the important distribution $w_C(t)$ for an event class C and an important coefficient value $c_A(t)$ caused by the importance depend on time attribute A can be given, the importance $w_i(t)$ of an event instance i when it occurs at time t_i is described in the following equation.

$$w_i(t) = c_A(t) \times w_C(t - t_i) \tag{1}$$

- Importance depend on temporal relationships among events

  The event importance also may change by relationships between the event and the other events. For example, when many events of a kind occur in a short period, the importance of the first and last few events may be higher than those of the other events as illustrated in figure 3. The relationships are various for applications and it is burdensome to define such relationships by hand for each application. We want to find the method that extracts some features in any group of events and makes use of those features to decide the importance depend on temporal relationships among events. This is one of our future works.

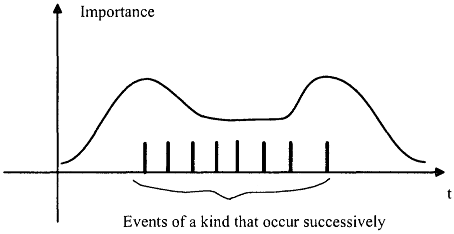


Figure 3 Importance Depend on Temporal Relationships among Events

3.2. SPACE-DEPENDENT IMPORTANCE

The space-dependent importance can be also classified into two types.

1 Importance in a micro-view of space

An element is not only important at the point where it exists but also important around it in space. Then it has an importance distribution in space which origin is the place where it exists. Various kinds of distribution forms are available dependently on the kind of element and applications. As the same as the time-dependent importance, an element is more important around the origin and is less far from the origin in general. An example of the importance distribution in space is illustrated in figure 4 (a). We also call this kind of importance as importance in a micro-view of space.

2 Importance in a macro-view of space

The importance of an element can change dependently on space attributes and spatial relationships among elements as the following. We call this importance as importance in a macro-view of space.

- Importance depend on space attributes

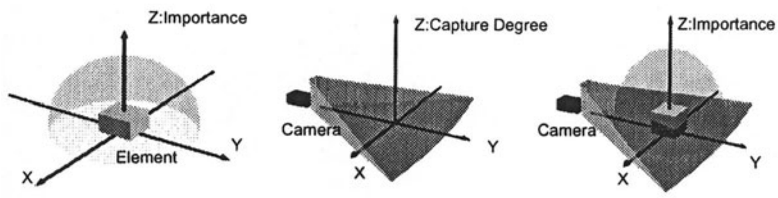
  The element importance may change with attributes of space such as geographical mesh data. For example, a heavy rain and flood warning is more important in a region with a soft ground. Suppose that the importance distribution $w_C(x, y)$

for an element class C and an importance coefficient value c_A caused by the importance depend on space attribute A can be given, the importance $w_i(x, y)$ of an element instance i at point (x_i, y_i) is described in the following equation.

$$w_i(x,y) = c_A(x,y) \times w_C(x - x_i, y - y_i) \tag{2}$$

- Importance depend on spatial relationships among elements

  The element importance also may change by relationships between the element and the other elements. For example, in a soccer game, a foul event near a goal object is more important and an area where some famous players are playing closely may be important. It is also our future work to find the method that extracts some features in any group of elements in space and furthermore in time and space, and makes use of those features to decide the importance depend on spatial or tempo-spatial relationships among elements.



(a) Importance Distribution of Elemaent (b) Capture Degree Distribution of Camera (c) Video Importance

Figure 4 Importance Calculation in Space

3.3. CALCULATION OF VIDEO IMPORTANCE

We explain the calculation method of video importance based on tempo-spatial importance, especially based on importance in a micro-view of time and space.

1 Calculation in space

   The video importance is calculated based on the elements in the video and how well those elements are taken in the video. For example, the importance distribution of an element is a form in figure 4 (a). A camera also has a spatial distribution as illustrated in figure 4 (b) to calculate how well the elements are captured. This distribution is decided by camera field such as a direction and a focus length of the camera. We call it as a capture degree. We define the video importance for an element as a product of the importance of the element and the capture degree of the camera as illustrated 4 (c). And the video importance is the sum of all the video importance for each element. Suppose that the importance of element j is w_j, the capture degree of camera i is e_i, the importance W_i of video captured by camera i is described in the following equation.

$$W_i = e_i \times \sum_j w_j \tag{3}$$

2 Calculation in time

The video importance calculated in space has to be renewed periodically because the event importance changes in time as illustrated in figure 2. Then display of multiple perspective video changes automatically as the replay of those video are in progress. Suppose that the importance of event k at time t is $w_k(t - t_k)$ where t_k is the time when the event occurs and the capture degree of camera i at time t is $e_i(t)$, the importance $W_i(t)$ of video captured by camera i is described in the following equation.

$$W_i(t) = e_i(t) \times \sum_k w_k(t - t_k) \tag{4}$$

A controllable camera and moving camera which camera field changes are available with the camera degree $e_i(t)$.

4. DISPLAY FOR THE VIDEO SKIMMING

After the video importance is calculated, some video are selected based on their importance and displayed with a map and three-dimensional graphics of the objects for skimming multiple perspective video effectively. We explain three display methods in the following.

4.1. SYNCHRONIZED DISPLAY IN SPACE

N of the most important video are selected and displayed on a map where each video window is placed near its camera field. (See figure 11) Those video are replayed synchronously based on the time where they were taken. Looking at what happens at several places by the synchronized replay on a map, it is possible to understand the event more accurately in less time. The number N should be altered interactively or automatically by user's intention, screen layout not to hide the map with many video windows, display ability of the machine and so forth. Furthermore, the following optional methods are useful.

- Enhancement: The video window is displayed with enhancement such as its size and color based on the video importance so that more important video can be seen more conspicuously.

- Quality: The quality of video such as image resolution and frame rate is altered based on the video importance so that more important video is displayed more clearly and precisely. It is also useful for making use of the system resources such as cpu power and network bandwidth effectively.

- Map operation: A map operation such as scrolling and scaling is useful to see a map from various viewpoints, so that the display method should correspond to the map operation. Namely, it alters the representative video to be displayed according to the display area of map. The video importance is independent on the display area of map, but there are two methods that the N representatives to be displayed are selected from all the video or selected from the video in which the display area are taken.

4.2. TRACEABLE DISPLAY IN SPACE

In tracking a moving object or observing a propagating phenomenon such as smoke in fire, it is possible to see the trace with the synchronized display in space if the event importance for such phenomenon are set high level. In addition, it is effective to display the trace itself on a map. The important video change in turn as the replay is in progress. We can see the trace of moving objects by displaying a still image of the video that was important in the past and has become unimportant and disappeared. (see figure 5)

4.3. SERIALIZED DISPLAY WITH SPACE NAVIGATION

This display method is useful to look at each important video more precisely and understand a three-dimensional structure around the video and between them. In this method, each video is displayed by the order based on the video importance. The video is also displayed in three-dimensional (3D) world that consists of 3D graphic objects around the video. After one video is over, we walk through the 3D world to the next video as illustrated in figure 6. We have not considered the detail

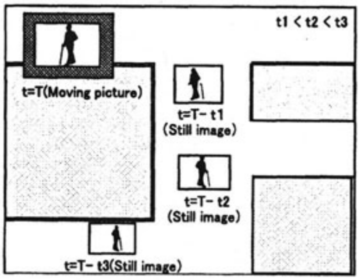


Figure 5 Traceable Display in Space

of the method yet. The order of video, the display time for video, the viewpoint in 3D world and so on can be decided by several things such as the place and time of video, the requested time to navigate besides the video importance. This is also one of our future works.

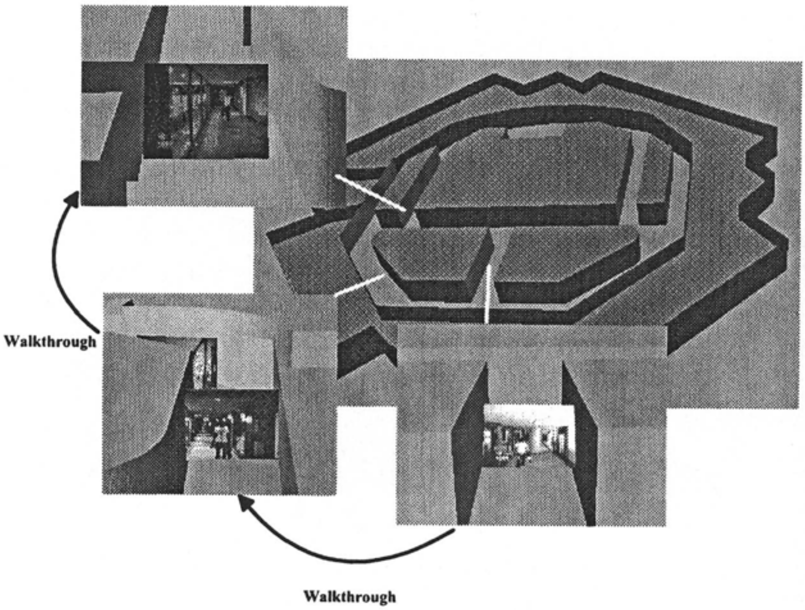


Figure 6 Serialized Display with Space Navigation

5. PROTOTYPE

We developed a prototype to evaluate some of the methods proposed in this paper. The system construction, the implemented functions and the software construction are described in the following.

1 System construction

The system construction of the prototype is illustrated in figure 7. Multiple perspective video are taken by multiple cameras and recorded on video tapes. Those video data are digitized by a capture board attached to a personal computer, encoded and stored on to a hard disk drive in an AVI file format. A video editing tool on the market is used to edit the multiple video streams such as adjusting the start frame of each stream. The metadata are created by hand with a text editor and stored in a text file. When the skimming program that we implemented is started, the system loads the map data, the metadata and the video data, and the system displays them on screen. The skimming program is written by C++ language and runs on a personal computer with dual Pentium® II and Windows NT 4.0®. The program uses Windows Media^{TM} Player Control to replay video.

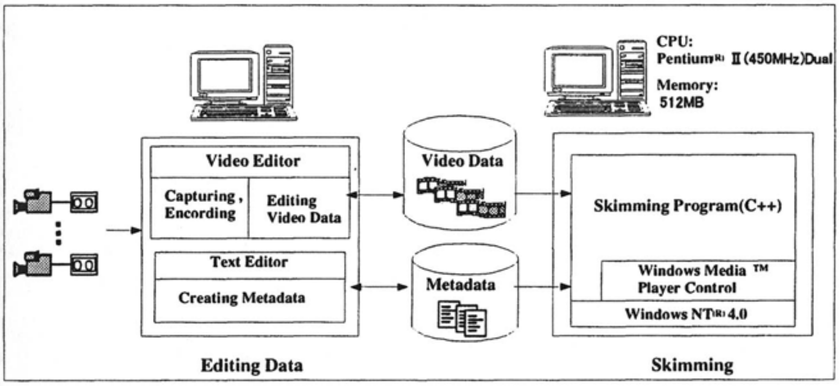


Figure 7 System Construction of Prototype

2 Implemented functions

The video importance is calculated based on the element importance defined in a text file. In addition, we implemented a function that makes the importance of video higher near around a mouse position on the map. Then the video importance is calculated based on the element importance and the mouse position. This function enables a user change the video importance dynamically based on the place where a user want to look at closely. It is useful to see the neighborhood of the important video or the important object. After checking whether each element is in a camera field or not, namely clipping, the importance in space is calculated by the following equation using only the importance of element existing in the camera field.(See figure 8)

$$W_i = \sum_j (w_j \times F_i)/(l_{ij} \times S_i) \tag{5}$$

In this equation, w_j is the importance of element j, S_i is the size of an imaging device, F_i is the focus length of camera i, l_{ij} is the distance between camera i and element j. This equation approximately represents the size of element that is projected on a focal plane. Normal distribution is used for the event importance.

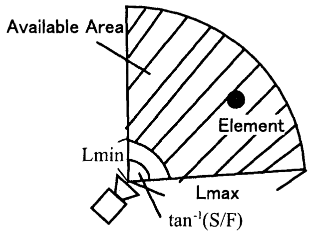


Figure 8 Clipping Element in Calculation of Video Importance

Three types of element, object, event and mouse position can be selected and each coefficient of those importance can be specified interactively for experimenting several variations of the importance calculation. The importance is recalculated periodically to cope with a change in time of the event importance and the screen is renewed. The importance is also calculated in every occurrence of the mouse event to cope with the mouse movement. N of the most important video are selected and displayed on a map with the size and frame color enhancement.(See figure 11) Those video are replayed synchronously based on the time where they were taken. Several parameters of the video importance calculation such as the number of video windows N can be altered interactively while the skimming program is executing. The map and the video windows on it can be scrolled and scaled.

3 Software construction

The skimming program consists of four threads as illustrated in figure 9. The interaction thread is a main thread of the program. It reads some data files such as the metadata and creates some internal worktables in an initial process. After the initial, it receives user's operations such as replay on the control panel and moving mouse on the map window, and sends messages to the other threads. For example, it sends a replay message to the display thread when a replay button is clicked on the control panel, and it sends a video change message to the importance thread when a mouse pointer is moved on the map window. The interaction thread also displays the map. The importance thread receives messages from the interaction thread and the event thread, calculates the video importance and updates the importance table that describes the importance and ranking for each video. The display thread receives messages from the interaction thread and the importance thread, makes a change in the display table that describes how to display each video based on the video importance, and updates the screen. The event thread monitors the captured time of the video being replayed and the occurrence time of event, and sends messages to the importance thread if some event exists within the certain amount of replay time. The display thread uses Media Player Control to display multiple videos synchronously. The controls read and replay each video data in accordance with given parameters such as a file name and a start time. When the display thread starts to replay a new video because the video importance changes, it gets the current playback time from one of the controls that is still replaying a video and gives the playback time as a start time to the new control for replaying the video. Media Player Control dose not have the exact synchronized replay ability, and neither this implementation does. But it is sufficient to evaluate the skimming methods displaying several video windows simultaneously as described in the section 6.

6. EXPERIMENTAL RESULT

With the prototype, we did the two experiments that were intended to evaluate our skimming method especially for moving objects. We describe the outline of the experiments and the results.

6.1. EXPERIMENT 1

With nine video cameras, we sparsely took the multiple perspective video of the scene where two toy cars and two toy trains were running on

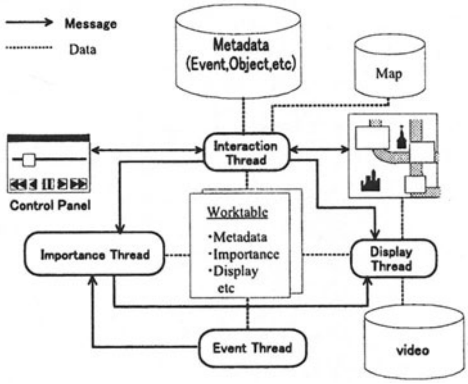


Figure 9 Software Construction of Skimming Program

rails and they sometimes collided each other.(See figure 10) We defined two types of event that were object colliding and object passing. The following were found out.

- When all nine video streams were displayed at the same time without any importance as illustrated in figure 11(a), the eyes were changed one after another to the video that captured a moving toy. We sometimes missed to see a collision because we paid much attention to a toy passing. We could observe four video windows simultaneously in the neighborhood which area was about $100cm^2$ on a 21" monitor, without overlooking any event on each video.

- When an event importance of collision was set higher than that of passage and the four of the most important video were displayed with enhancement as illustrated in figure 11(b), we did not pay so much attention to a passage that we did not overlook a collision. This method has an effect that a user can know events with less importance without overlooking events with higher importance.

- It was easy to understand how a toy moves around on the map if an event importance of passage was set high. But we misunderstood the moving direction of the toy if its direction on the map and its direction on the video displayed on the map were not matched.

- It was desirable to change a video window size slowly because changing it rapidly and frequently made the eyes tired.

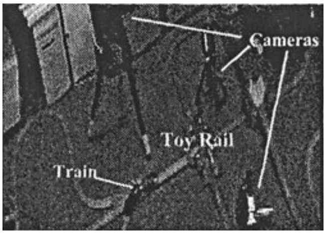
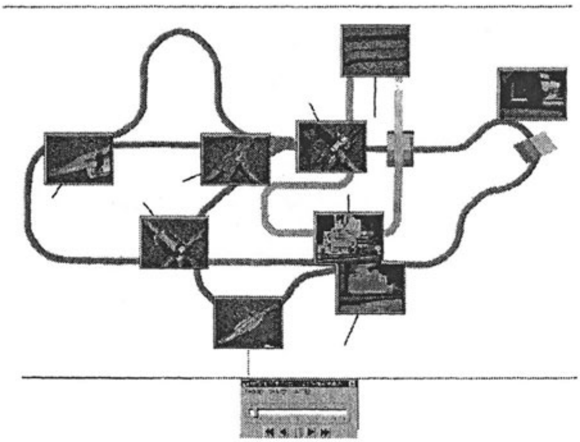
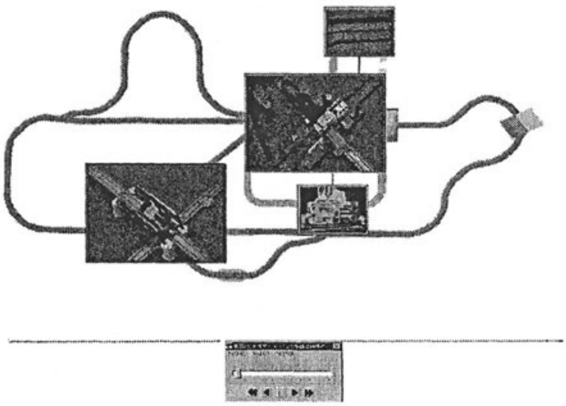


Figure 10 Experiment 1: Synchronized Display in Space with Toy Trains



(a) Synchronized Display in Space without Importance



(b) Synchronized Display in Space with Enhancement

Figure 11 Skimming Examples in Experiment 1

6.2. EXPERIMENT 2

With ten video cameras, we sparsely took the multiple perspective video of the scene where some persons were walking in a building floor. We supposed two stories. One is that a man was a suspicious person and he walked around at night with a few people walking. We defined the event that someone passed through a camera. The following were found out.

- The video that captured the suspicious man were displayed one after another as the event importance changed as illustrated in figure 12(b), so that we could easily follow him on the map. We did not mind so much the video that captured other people even if the video were sometimes displayed. The traceable display in space has not been implemented yet, but the moving path was left in the head enough by this display method.

- When he was in the blind spot of a camera for a while, the importance of the video became less so that the video window was disappeared and any other video were not displayed. Some measures are necessary to avoid this problem.

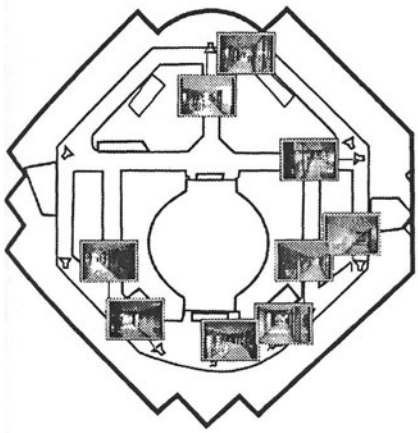
Another story is that the suspicious man walked around in the daytime with many people walking. Only the mouse position was used for the importance calculation so that we followed him interactively.

- We could easily follow him with the mouse operation. This mouse operation was very good because the video we wanted to see could be displayed only by moving the mouse near around the place.

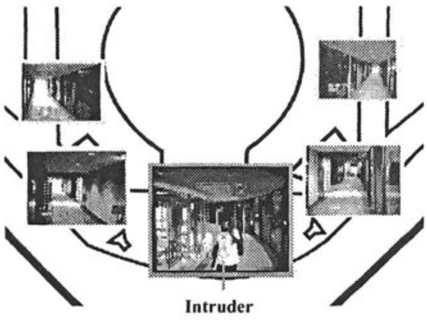
- Inappropriate video were sometimes displayed because the camera field was decided only by the position, direction and focus of camera and the building structure was not considered.

7. CONCLUSIONS

In this paper, we introduced the video importance calculated by the importance of elements that is taken in the video, and then applied the importance to the display methods with a map and 3D world for skimming the multiple perspective video effectively. We found the effectiveness for some of the proposed methods by the experiments described in section 6. But those experiments are basic and primitive. We want to evaluate them in practical experiments or real applications which are bigger scale, have more than several tens of cameras, have complicated



(a)Synchronized Display in Space without Importance



(b) Synchronized Display in Space with Enhancement

Figure 12 Skimming Examples in Experiment 2

combinations of events and so on. We are also planning to study the method that extracts some features in a group of elements and makes use of them to decide the tempo-spatial importance, and the serialized display method with space navigation.

8. ACKNOWLEDGMENTS

This research was supported by the Japanese Ministry of Education Grant-in-Aid for Scientific Research on Priority Area: "Advanced Database,"(area no.275 (08244103)) and Research for the Future Program of Japan Society for the Promotion of Science under the project "Researches on Advanced Multimedia Contents Processing,"(project no. JSPSRFTF97P00501).

References

Tsukada, A., Hata, T., Matsuda, F., Sato, K. and Ozaki, M. (1997) "Video Data Management in Media Controller: a Distributed Multimedia Surveillance System," *Lecture Notes in Computer Science, Vol.1309, pp.231-240 1997.*

Katkere, A., Schlenzig, J., Gupta, A., Jain, R.(1996) "Intearctive Video on WWW: Beyond VCR-like Intefaces," *Fifth International World Wide Web Conference, Computer Networks and ISDN Systems, Vol.28, pp.1559-1572, 1996 .*

Arikawa, M.(1998) "Spatial Browsing for Video Data Using Spatio-Temporal Descriptions of Cameras," *IPSJ SIG Notes, Vol.98, No.58, pp.225-232, 1998.*

Sauter, J., Lüsebrink, D. "The Invisible Shape of Things Past," *http://www.artcom.de/.*

Wactlar, H.D., Kanade, T., Smith, M.A. and Stevens, S.M.(1996) "Intelligent Access to Digital Video: Informedia Project," *IEEE Computer, 29(5), pp.46-52, May 1996.*

Brill, F.Z., Olson T.J. and Tserng, C.(1998) "Event Recognition and Reliability Improvements for the Autonomous Video Surveillance System," *DARPA98, pp.267-283, 1998.*

Biographies

Toshihiko Hata received his B.E. and M.E. degrees in Communication Engineering from Osaka University in 1980 and 1982, respectively. He is a group manager of Industrial Electronics & Systems Laboratory, Mitsubishi Electric Corporation. He is also a doctoral candidate in Graduate School of Science and Technology, Kobe University. His research interests include multimedia systems, video database and multimedia networking.

Tatsuo Hirose is currently a graduate student of the Department of Computer and Systems Engineering at Kobe University. He received his B.E. in Computer Science from the Kobe University in 1999. His research interests include multimedia database.

Katsumi Tanaka received his B.E., M.E., and Ph.D. degrees in information Science from Kyoto University in 1974, 1976, and 1981, respectively. In 1986, he joined the Department of Instrumentation Engineering at Kobe University as an associate professor. Since 1994, he has been a professor of the Department of Computer & Systems Engineering and the Division of Intelligence Science (Information and Media Sciences, later), Graduate School of Science & Technology at Kobe University. His interests include object-oriented databases, historical database models, hypermedia systems, and multimedia information systems. Dr. Tanaka is a member of the ACM, IEEE Computer Society, and the Information Processing Society of Japan (IPSJ).