

# MEDIA SYNCHRONIZATION METHOD FOR VIDEO HYPERMEDIA APPLICATION BASED ON EXTENDED EVENT MODEL

Hironobu Abe<sup>1</sup>, Hiroshi Shigeno<sup>2</sup>, Ken-ichi Okada<sup>2</sup>

<sup>1</sup>Mitsubishi Electric Corporation, <sup>2</sup>Keio University

## ABSTRACT

This paper describes a proposal of an extended event model using a media synchronization method for video hypermedia applications. In this extended event model, video and metadata are synchronized by periodically inserting event information in the video multiplex. We considered the following design policies: 1) a model that is independent of the video format and delivery method, 2) the synchronization accuracy can be tuned depending on the purpose and use of the metadata. We designed the extended event model based on the above design policies, and implemented this model as an encode/decode library for Windows Media. Based on this model we developed a video hypermedia system prototype and performed evaluation experiments. The evaluation results of real time synchronization performance of the system prototype showed that in the case of sports video content a synchronization accuracy of 100 msec between video and metadata, makes our method effective for use in video hypermedia applications.

## 1. INTRODUCTION

Recently, the rapid proliferation of broadband Internet access has enabled high-speed Internet connections possible. As a consequence of this *broadbandization*, delivery of multimedia information such as voice and video is possible.

In this paper, we define metadata that combine video and related information for the realization of video hypermedia applications using streaming technologies, and design a media synchronization method of video and metadata, and describe the implementation and evaluation of our system prototype.

System requirements of video hypermedia application in this paper are, 1) a model that is independent of the video format and delivery method, 2) the synchronization accuracy can be tuned depending on the purpose and use of the metadata. Because it is a method in existing media synchronization method such as SMIL[1] to describe time information in the script as synchronization information with the metadata, it is not applicable in the delivery of

the broadcasting type without time information in the video targeted with this paper.

This system periodically inserts event information which defines time relationship between video and metadata in the video multiplex, and establishes media synchronized delivery between video and metadata.

We implemented a system prototype considering the performance of synchronization accuracy, because handling event processing has a large impact on the CPU load. And we evaluated real time synchronization performance of the system prototype.

## 2. EXTENDED EVENT MODEL

### 2.1. Design concept

In this paper, we propose extending the method of event inserting in video as a design concept of media synchronization between video and metadata.

Events are independent of the video format and delivery method, because they can be inserted at any point of the video stream. Media synchronization is accomplished by inserting information required for synchronized delivery of video and metadata. Media synchronization is also possible for the case of broadcasting video over multicast technology, because synchronization information is inserted in the video stream.

Also, depending on the intended use, event information can be inserted periodically to achieve start point and real time synchronization of video and metadata.

### 2.2. Metadata model

Spatio-Temporal Locator[2] defined in MPEG-7 Visual is a metadata model used for video hypermedia applications. But since the description cost of this model is high, we choose as data model for our research the model described in [3]. Figure 1 shows an overview of the metadata. In our model, metadata in the video stream does not necessarily contain size and position data in every frame, it only contains data in the frames we call start point, end point, and intermediate points. Frames which

have no data make an estimate based on anteroposterior frames using interpolation, as shown in Figure.1.

“Related information” is defined as an action provided to the user after clicking the metadata on the video. This model defines “Web” and “Text” as related information. “Web” sends the URL information to an external web browser as retrieved related information, and as consequence the specified web content is shown in the web browser. “Text” displays text information on top of the video as related information.

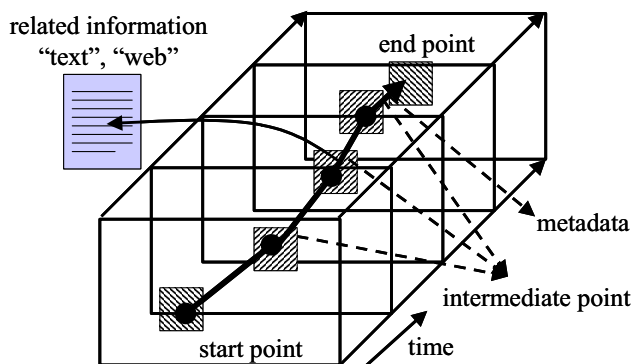


Figure 1 Overview of metadata

### 2.3. Media synchronization method

Based on this design concept, we propose a delivery method based on the extended event model for synchronized delivery of video and metadata.

Figure 2 shows the concept of our synchronized delivery method used in the proposed extended event model.

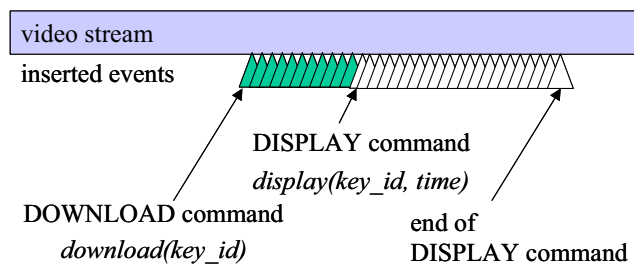


Figure 2 Concept of synchronized delivery method used in the extended event model

Our proposed synchronized delivery method is a method which periodically inserts synchronized information in the video stream to achieve media synchronization of video and metadata. Almost all video delivery services in the Internet use connectionless protocols based on UDP/IP. Our method deals with packet loss by inserting synchronized information periodically in the video.

This method inserts the minimum information required for synchronization as synchronized information in the video. Metadata is not inserted in the video itself, but delivered using other delivery methods such as HTTP.

Synchronized information in this method is defined below.

(1) DOWNLOAD command: **download (key\_id)**

When the client receives this command, it looks for metadata in internal memory corresponding to the argument *key\_id* and in case the metadata does not exist, it downloads it via the network. By calling the download command before calling the display command, start point synchronization of video and metadata is achieved.

(2) DISPLAY command: **display (key\_id, time)**

When the client receives this command, it reads the metadata corresponding to the argument *key\_id* from internal memory and displays it on the designated position of video in playback. If the designated metadata does not exist in internal memory, then the client downloads it via the network and displays it on the video. The argument *time* sets the elapsed time from starting point of the display command. Before displaying the metadata, the client calculates position and size of the metadata based on the elapsed time. By calling the display command periodically this method accomplishes real time synchronization of video and metadata.

Features of this method are summarized below.

- (1) it is independent of the video format and delivery method
- (2) it establishes start point synchronization of video and metadata by calling download command before calling display command
- (3) it establishes real time synchronization of video and metadata by calling display command periodically
- (4) by inserting synchronized information periodically, it can support connectionless type video delivery service in spite of experiencing packet loss

Figure 3 shows a flow of synchronized delivery of video and metadata.

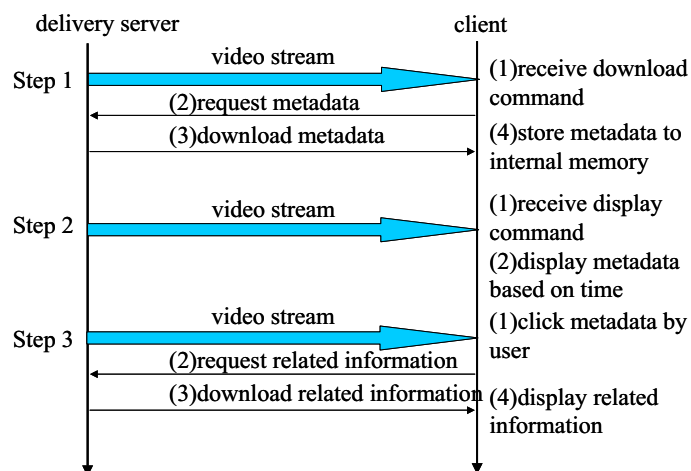


Figure 3 Flow of synchronized delivery of video and metadata

(1) Step 1: Downloading metadata

When the client receives the download command inserted in video, it downloads metadata via network and stores it in internal memory. However if the client already has the appropriate metadata in internal memory, then it does nothing.

(2) Step 2: Synchronized displaying of metadata

When the client receives a display command inserted in video, it then calculates display position and size of the metadata in internal memory based on elapsed time and displays it on the video. Every time the client receives this command, it updates the display of metadata.

(3) Step 3: Displaying related information

If the user clicks the metadata displayed on video, the client gets related information via the network and performs the display process.

Our proposed extended event model is implemented as encode /decode libraries at the target video delivery platform. Synchronized delivery of video and metadata is achieved by applying the encode library to authoring software which annotates video and metadata and the decode library to playback software.

### 3. IMPLEMENTATION

We implemented a prototype video hypermedia application for the purpose of performance evaluation of the extended event model. We selected Windows Media [4] as a base video delivery platform, which is widely used on the Internet.

Windows Media supports a script region in addition to video and audio, and user-defined events in this script region can be added.

The recommended minimum event synchronization accuracy in Windows Media platform is 100 msec. We implemented encode/ decode libraries under the condition that minimum event synchronization accuracy is 100 msec, and evaluated real time synchronization performance of the implemented libraries by experimentation.

We implemented video hypermedia authoring software which reads video data, displays metadata on the display, edits position, size and related information of metadata and encode video to Windows Media inserting synchronized information using the Windows Media encode library.

Windows Media video and metadata authored by authoring software is delivered from delivery server using Windows Media Service and Internet Information Service bundled in the Windows operating system.

We implemented video hypermedia playback software which receives Windows Media video and downloads metadata from the delivery server based on inserted synchronized information and displays the metadata on the video using Windows Media decode library.

When Windows Media Player receives Windows Media video inserted synchronized information, it plays

only video because synchronized information is treated as user-defined event.

## 4. EXPERIMENTATION

### 4.1. Evaluation policy

We prepared evaluation experiments of real time synchronization performance on our implemented prototype system.

In the case of implementing this method on Windows Media platform, the minimum event synchronization accuracy was 100 msec. In this evaluation experiment, we evaluated real time synchronization performance of video and metadata in the prototype system.

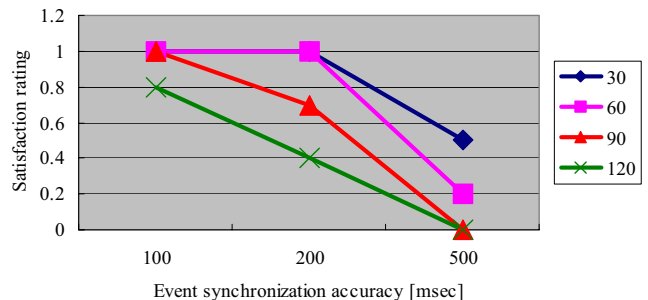
We prepared two types of data as evaluation data for real time synchronization performance. Table 1 shows the description of evaluation data.

**Table 1 Description of evaluation data**

Type	Profile	Number	Average Moving Speed
Data 1	Rectangle object	1	30 dots/sec
		1	60 dots/sec
		1	90 dots/sec
		1	120 dots/sec
Data 2	Baseball: hitting	5	11.59 dots/sec
	Baseball: fielding	5	37.72 dots/sec
	Baseball: running	5	61.09 dots/sec

### 4.2. Base performance evaluation experiments

Using authoring software, we opened videos in Data 1 and edited metadata to objects whose average moving speed are different, and encoded them to Windows Media changing the event synchronization accuracy parameter: 100, 200 and 500msec. Video encoding parameters were set as follows. Encoding bit rate: 1Mbps, frame rate: 30 frames/sec, video size: 320x240.



**Figure 4 Evaluation results of base performance**

Using playback software, we conducted evaluation experiments of display performance based on users satisfaction rating when delivering prepared evaluation data, which included different object moving speeds to 5

users. Figure 4 shows the evaluation results of base performance.

The method to calculate users' satisfaction rating is described here. Synchronized video and metadata using playback software is shown to users. Based on the difference between the target object and metadata, users are asked to evaluate the performance as follows

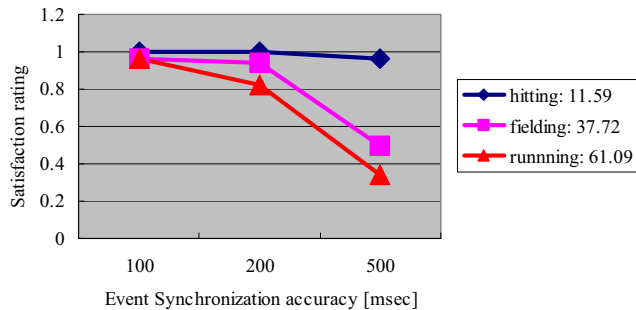
- (1) target object and metadata are out of alignment: 0
- (2) target object and metadata are in alignment: 1
- (3) between 1 and 0: 0.5

The averages of all users are then computed.

### 4.3. Evaluation experiments for sports video

Using authoring software, we edited metadata to videos in Data 2, and encoded them to Windows Media using as event synchronization accuracy the following values: 100, 200 and 500msec. Encoding parameters of the video were the same parameters used at the base performance evaluation.

Using playback software, we conducted evaluation experiments with 5 users by delivering data which included objects moving at different speeds and asking users to evaluate their satisfaction rating. Figure 5 shows the evaluation results for sports video.



**Figure 5 Evaluation results for sports video**

The method to calculate users' satisfaction rating is the same method used at the base performance evaluation.

### 4.4. Summary

According to Figure 4, if we set the target value of the users' satisfaction rating to 0.8, and the event synchronization accuracy parameter to 100 msec, the synchronized display performance is satisfactory in the case of delivering metadata with the target object moving at an average speed of 120 dots/sec which is the fastest in the evaluation data set. However, in the case of 200 or 500 msec, as the target object moves faster, the synchronized display performance gradually decreases.

Based on the results of base performance evaluation, according to Figure 5, if we set the target value of the users' satisfaction rating to 0.8, and the event

synchronization accuracy parameter to 100 msec, we also found that the synchronized display performance is satisfactory in the case of delivering metadata with the target object having an average speed equal to the profile "running", which in average is the fastest moving speed.

Therefore, if the event synchronization accuracy parameter is 100 msec, our method proved to be able to satisfactorily deliver metadata with objects moving at a maximum average speed of 120 dots/sec for video at 30 frame/sec and 320x240 display size. Also, since the maximum value of the average moving speed of objects in sport videos is usually around 100 dots/sec, our prototype system is applicable to video hypermedia applications for such sport video content.

## 5. CONCLUSION

In this paper, we considered the system requirements for a video hypermedia application with the objective of achieving media synchronization of video and metadata.

We considered the system requirements shown below.

1) a model that is independent of the video format and delivery method, 2) the synchronization accuracy can be tuned depending on the purpose and use of the metadata. Based on the system requirements, we proposed and designed an extended event model and implemented it as encode/decode libraries on the Windows Media platform.

We implemented a prototype system of a video hypermedia application based on the implemented libraries, and conducted experiments regarding its real time synchronization performance.

Considering the results of our evaluation experiments in the case of sport video content, a synchronization accuracy of 100 msec between video and metadata, makes our method effective for use in video hypermedia applications.

## 6. REFERENCES

- [1] W3C Synchronized Multimedia Home page, <http://www.w3c.org/AudioVideo/>.
- [2] ISO/IEC JTC 1/SC 29/WG 11, *MPEG-7 Overview (Version 10)*, N6828, October 2004.
- [3] H. Abe, J. Kanda, K. Wakimoto, H. Shigeno and K. Okada, "Video hypermedia authoring for video delivering service based on a combination of automatic tracking and manual editing," *5th International Workshop on Image Analysis for Multimedia Interactive Services WIAMIS2004*, in CD-ROM Proceedings, April 2004.
- [4] Microsoft Windows Media Home page, <http://www.microsoft.com/windowsmedia/>.