An Integrated Framework for Interactive Multimedia Presentations in Distributed Multimedia Systems *

Ramazan Savaş Aygün
Department of Computer Science and Engineering
State University of New York at Buffalo
aygun@cse.buffalo.edu
Advisor: Aidong Zhang

ABSTRACT

An interactive multimedia presentation in a distributed multimedia system requires synchronization of media streams, preprocessing media for content-based retrieval and low-bandwidth transmission over network, and user interface for interacting multimedia presentations. The power of synchronization models is limited to the synchronization specifications and user interactions. We propose an event-based synchronization model that can handle time-based actions while enabling user interactions like backward and skip. For effective transmission of multimedia data, the multimedia data needs to be preprocessed. The sprite generation and moving objects segmentation can reduce the required bandwidth significantly. We propose a method for multiresolution sprite which will allow reproduction of the video at different resolutions. The object segmentation will be extracted by generating a closed boundary for the object. Since the video data may also exist in a compressed format, we also propose to extract new features from the compressed video. We will consider compressed data that is generated by Discrete Cosine Transform (DCT) which has been used in MPEG-1, MPEG-2, MPEG-4 [8] and H263.1. The user will be provided a high-level user interface to access the contents of the presentation. We will test this integrated framework on distance education project over the Internet.

1. INTRODUCTION

In this doctoral thesis, we aim to achieve the following goals:

- to provide a synchronization model which will flexibly support interactive multimedia presentations in distributed multimedia systems,
- to achieve Quality of Service(QoS) requirements by preprocessing multimedia data which includes the generation of the background and the extraction of moving objects,
- to provide a high-level user interface for content-based access of multimedia data.

*This research is supported by NSF grant IIS-9733730.

2. MULTIMEDIA SYNCHRONIZATION

There has been many models proposed for multimedia presentations [3, 7, 4]. The major drawback of these models is the lack of support for VCR user interactions like backward and skip which are necessary for applications like distance education over network and sports applications. This originates from the information which is specified for the forward presentation by the user by stating the relationships among streams as in [1]. We propose to handle synchronization requirements by *synchronization rules* which are composed of events, conditions and actions like ECA rules. We will also consider how synchronization rules are generated from SMIL expressions. Each synchronization rule is simple and it is easier to deduce new rules from these rules for the backward presentation. A timeline for events and actions have to be generated to handle user interactions like skip and change direction. Our model [6] is an event-based model which can also handle time-based operations. It has 1) receivers to detect events, 2) controllers to check conditions and 3) actors to execute actions. A general layout of the components of the synchronization model is depicted in Figure 1.

3. VIDEO PROCESSING

If the data are transmitted over network, the background generation and the extraction of moving objects can reduce the amount of data to be transmitted significantly. If the multimedia data are compressed, features which will decrease the amount of processing need to be determined. Most of the previous video streams are compressed using Discrete Cosine Transform (DCT). The value of DC coefficient, the number of non-zero AC coefficients and the index of the absolute value of the highest AC coefficient are good features to compare two blocks. The sum of the absolute values of the highest AC coefficients indicates the existence of a boundary in the block. The highest coefficients also indicate the type of the boundary like (vertical, horizontal) which are still true after camera operations like panning and tilting.

The background generation can be performed by clustering of similar blocks at different intervals of the video stream

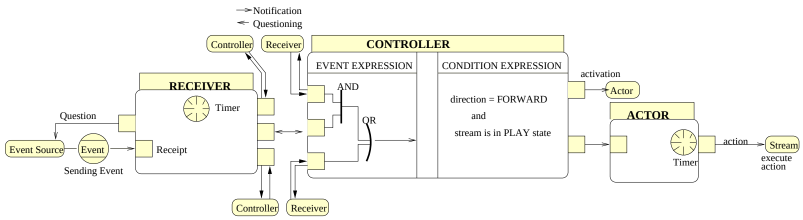

Figure 1: The receiver, controller and actor relationships.

[2]. The cluster which contains more elements than the others is the candidate cluster which contains the blocks for the background. If the video source is MPEG-compressed video, the P and B blocks which contain motion information can be eliminated before clustering because these blocks contain the moving objects. Since the change between consequent frames are small, I frames at specific intervals are used to generate the background. The generation of the sprite at different resolutions using a sprite pyramid will also be considered to reproduce the video at the required resolution.

The extraction of moving objects can be performed by using the background model. Common patterns that exist in the objects and the background as well as objects' shadows are two major problems. To reduce the effect of these problems, the edges in the background and the edges in a frame are compared. The edges of the object may be removed because of the overlapping or the edges cannot be detected because of the similarity on the edge of the object and the background. After detection of edges, morphological operations are applied to detect the video object in [5]. These morphological operations do not consider the contents of the frame and may produce false results. In our case, when the edges are eliminated or cannot be detected, we follow the information of the highest gradient in the neighbor pixels to connect the edges. The edges that overlap with the background edges are removed. Once a closed object boundary is created, the inside of the region is filled to obtain the object.

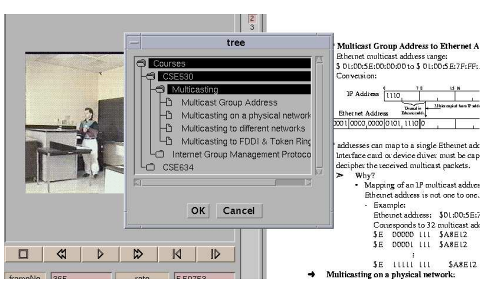

Figure 2: User Interface

4. HIGH-LEVEL USER INTERFACE

Most of the user interfaces for multimedia presentations only consider VCR-type user interactions like play, pause and resume. Applications like distance education need high-level user interactions based on the contents of the presentation. One of our applications is distance education over network. Finding topics in a multimedia presentation is very hard. The multimedia presentation is presented like a book which has a table of contents. The student can view any section of the presentation like a book. The user can also set bookmarks in the presentation to review those sections of the presentation. These are some of the reasons why the synchronization model has to support skip and backward functionalities. A segment of the user interface is shown in Figure 2.

5. EXPECTED RESULTS AND FUTURE WORK

The proposed synchronization module provides flexibility in synchronization specification and enables interactive multimedia presentations in distributed multimedia systems. The generation of the background and the extraction of the moving objects will help the applications that require low-bandwidth provide more reliable presentations. We want to integrate the synchronization module, the generation of the sprite and the extraction of moving objects with a high level user interface for distance education application where students can access multimedia presentations like accessing books. We will test the integrated system from different locations in Europe and within United States.

6. REFERENCES

[1] The cwi smil page.http://www.cwi.nl/smil.

[2] R. S. Aygun and A. Zhang. Stationary Background Generation in MPEG compressed Video Sequences. In *International Conference on Multimedia Expo (to appear)*, Tokyo, Japan, August 2001.

[3] B. Bailey, J. Konstan, R. Cooley, and M. Dejong. Nsync - a toolkit for building interactive multimedia presentations. In *Proceedings of ACM Multimedia*, pages 257–266, September 1998.

[4] I. Herman, N. Correira, D. A. Duce, D. J. Duke, G. J. Reynolds, and J. V. Loo. A standard model for multimedia synchronization: Premo synchronization objects. *Multimedia systems*, 6(2):88–101, 1998.

[5] C. Kim and J.-N. Hwang. An integrated scheme for object-based video abstraction. In *ACM*, pages 303–311, October 2000.

[6] R. S. Aygun and A. Zhang. Interactive multimedia presentation management in distributed multimedia systems. In *Proc. of Int.Conf. on Information Technology: Coding and Computing*, Las Vegas, Nevada, April 2001.

[7] J. Schnepf, J. Konstan, and D. Du. FLIPS: Flexible Interactice Presentation Synchronization. *IEEE Selected Areas of Communication*, 14(1):114–125, 1996.

[8] T. Sikora. The mpeg-4 video standard verification model. *IEEE Trans. Circuits Syst. Video Technology*, 7:19–31, February 1997.