A Synchronization Scheme for Multimedia Annotation

Chin-Hwa Kuo, Timothy K. Shih, Tzu-Chuan Chou

Multimedia Information NEtworking Laboratory (MINE)
Department of Computer Science and Information Engineering
Tamkang University, Tamsui, Taiwan R.O.C
e-mail: chkuo@mail.tku.edu.tw

Abstract

Multimedia annotation is of benefit to many multimedia applications in the enhancement of presentation, instruction, and communication. In the present paper, we make use of voice and electronic pen as annotation media in the designing of a multimedia annotation and playback system. Our contributions are as follows: (1) A time-based synchronization mechanism is developed which takes care of not only voice and electronic pen movement temporal synchronization issue but also electronic pen movement spatial synchronization problem. (2) File storage and replay mechanisms are devised to handle multimedia storage process and to facilitate interactive presentation process during replay phase. We implemented the proposed system on the Microsoft Windows 95 platforms in the Multimedia Information NEtworking Laboratory (MINE) at the Tamkang University.

Keywords: multimedia synchronization, multimedia annotation, computer aided instruction.

1. Introduction

An annotation mechanism is of benefit to many multimedia applications [2]. For instance, in multimedia presentation and in remote learning where annotation is utilized to enhance presentation, instruction, and communication. Up to date, several commercial multimedia presentation packages, e.g., Microsoft Power Point [6] and IBM/Lotus Freelance Graphics, are available. However, all of them are lack of multimedia annotation and replay functions. In other words, once the presentation document is formatted, its presentation form is also fixed. A presenter or instructor can only utilize the computer mouse pointer or electronic pen to annotate his/her documents during presentation process. Furthermore, these packages do not equip with the record function to replay the annotating process. Under such a circumstance, many valuable information do not capture and retain. Obviously it is useful to design an annotation and playback system to overcome the above drawbacks.

Briefly speaking, a multimedia annotation and playback system serves to enhance a presentation and to facilitate a review process. To achieve these goals, in our design which integrates the following mechanisms: (1) system input mechanism, (2) synchronization mechanism, (3) file storage mechanism, and (4) replay mechanism. The system input mechanism takes care of the input medium signals. Two time-dependent media, i.e., *voice* and *electronic pen*, are chosen as annotation media to comment prepared multimedia documents which are made by other multimedia presentation application packages. The synchronization mechanism needs to handle voice and pointer synchronization issue. We consider the temporal as well as spatial synchronization of voice and pointer. The file storage mechanism deals with the annotated file storage process. And the replay mechanism is used to facilitate the interactive process during playback. With the above mechanisms, the whole presentation are recorded and thus the audience and the students can retrieve the annotated file in an interactive manner to imitate the original presentation or only to replay the significant parts.

The present paper is outlined as follows: In section 2, we discuss the system overview of the multimedia application annotation and playback system. In section 3, we present the temporal and spatial synchronization scheme of voice and pointer used in our system. In section 4, we present the file storage mechanism and interactive replay mechanism. In section 5, a prototype of the proposed system is implemented. Conclusion and future works are given in section 6.

2. System overview

As shown in Fig. 1, the whole system process can be divided into three phases: the *annotating phase*, the *storage phase*, and the *replay phase*. Before a presentation, an instructor has the presentation documents prepared in advance. During the annotating phase, the instructor browses the presentation document, comments the document through voice, and annotates the document by using electronic pen or mouse movement. Note that electronic pen and mouse movement perform the same annotation function in our

0-7803-4053-1/97/$10.00 © 1997 IEEE

system, thus hereafter we use these two terms interchangeable. As it is well known, a lot of information produced in this phase may be valuable to retain for future review process. In the storage phase, the synchronization mechanism integrates all the above input media and the original file as an *annotated* file. This annotated file is then stored in the lecture database by following design format. Finally, in the replay phase, the original file or the annotated file can be retrieved by the users.

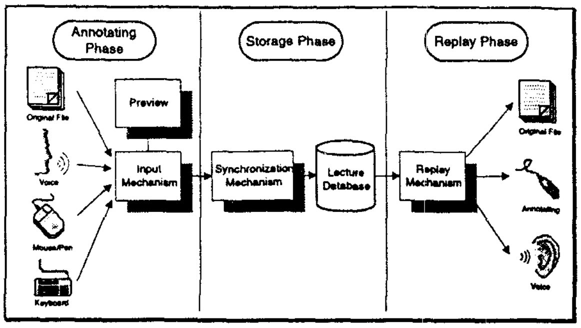

Figure 1. The multimedia annotation and playback system process

Several mechanisms are developed to accomplish the above described process which include: (1) system input mechanism, (2) synchronization mechanism, (3) file storage mechanism, and (4) replay mechanism. We describe the functions of these mechanisms as follows:

System input mechanism

The system input mechanism takes care of the signals such as commentary media and key press events. Two time-dependent media, i.e., *voice* and *electronic pen*, are chosen as annotation media to comment prepared multimedia documents. These documents are made by other multimedia presentation application packages. Note that the presentation documents are *only* one of our system inputs.

Synchronization mechanism

The synchronization mechanism handles media presentation synchronization issue. Since voice and pointer movement are taken as system inputs, the synchronization mechanism needs to deal with the *inter-media synchronization* problem [4]. Moreover, the window size may be changed during the replay phase. Thus the pointer movement spatial synchronization issue are also necessarily to be considered.

File storage mechanism

The file storage mechanism deals with the annotated file storage process. As shown in figure 1, its inputs are the outputs of synchronization mechanism. To facilitate the replay process, we devise a data structure and store the data with two specialized indexes. One is the time index for the indication of play time and the other is the slide index for the indication of slide sequence.

Replay mechanism

The replay mechanism performs the replay functions. A graphical user interface is designed to help users to select the desired document. It also includes pause, forward, backward functions. Accompany the time index and the slide index, users can arbitrarily select replay starting position in the time-based manner or in the slide-based manner.

With the above mechanisms, the whole presentation process are recorded and thus the audience and the students can retrieve the annotated file in an interactive manner to imitate the annotated presentation, to replay the significant parts, or to retrieve the original documents. In this manner, the overall application effects can be significantly enhanced.

3. Synchronization mechanism

Due to multimedia synchronization issue significantly affects the presentation quality, the synchronization mechanism plays a vital role in the designing of multimedia applications. A number of different techniques for multimedia synchronization have been proposed to satisfy diverse requirements. These techniques can be classified as either *distributed* or *local*. The distributed approaches implement network protocol-based synchronization and the local approaches are used in single sites for multimedia synchronization. In [1] and [3], a general problem description and proposed methodologies are presented. Readers consult the above references for further detail.

As shown in Fig. 1, the original file, mouse movement event, key press event, and instructor voice are inputs to the system. The original file is usually presented in a slide by slide manner. And the instructor makes comments on each slide by using voice and annotates the slides by using electronic pen. These events are need to playback synchronously. Thus, the synchronization mechanism need deal with the inter-media synchronization of these system inputs.

Here a time-based temporal synchronization mechanism is utilized. Once the instructor starts the annotating process, system is also starting to format the annotating data. At each time interval which is one second in our design, all system input signals are recorded. For instance, as shown in fig. 2a, in the first second system has $voice_1$ data, mouse down event at time t_1, mouse movement event at time t_2, and mouse movement event at time t_3. In the second second, system has $voice_2$ data, mouse movement event at time t_4, and mouse movement event at time t_5. In the third second, system has a silence interval and mouse up event at time t_6. In the fourth second, system has voice$_4$ data and key press event which may indicate another slide is selected. Then the synchronization mechanism constructs the data format in a sequence as shown in fig. 2b. In this example, mouse downs at the time t_1 and ups at time t_6. During this time interval, several mouse movement events may occur. Notice that the system input mechanism detects the voice input signal at every time interval to justify that particular time interval belongs to talk spurt or silence gap. If it is a talk spurt, no matter how long of that talk spurt, the whole time interval voice data is stored. Notice also that the storage data of the mouse event includes the time instant such that the replay time can be duplicated in the replay phase.

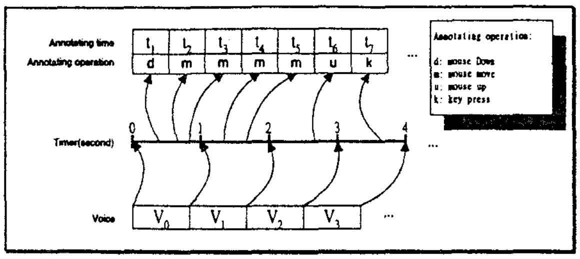

Figure 2a. A sample of time-based temporal synchronization

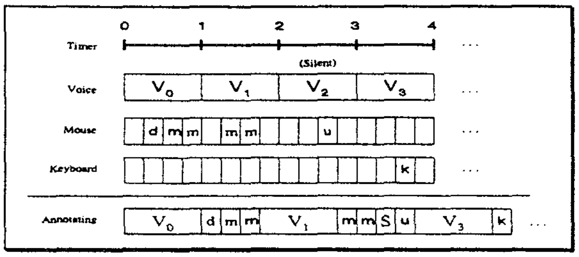

Figure 2b. A sample of time-based temporal synchronization

During the replay phase, the window may be resized. Then the spatial relationship is changed. In order to handle such a situation, a simple algebraic transformation spatial synchronization scheme is included in the synchronization mechanism.

In the annotating phase, system record the window size as ($Sx \times Sy$). In the replay phase, system also record the window size as ($Tx \times Ty$). When the window size changes from $Sx \times Sy$ to $Tx \times Ty$, the mouse movement position needs to transform by following the formula:

$$X = X_0 \times S_X \div T_X$$

$$Y = Y_0 \times S_Y \div T_Y$$

where X: the translated horizontal coordinate;
Y: the translated vertical coordinate;
X_0: the original horizontal coordinate;
Y_0: the original vertical coordinate.

4. File storage and replay

In this section, the file storage mechanism is illustrated. As mentioned before, the annotated file consists of the original file and the annotating part which includes voice, key press events, and mouse movement events. These two parts are stored separately. In the annotating part, we list its data structure as follows:

```
struct annotating_data
{
        int         waveSize;
        BYTE       *waveData;
        Clist <Annotation> AnnoList; // Clist is a list template
};
and
struct Annotation
{
        char        eventFlag;
        CPoint      mousePos;
        WORD        keyInfo;
        DWORD       timeInterval;
};
```

The annotation data are recorded in a list of `annotating_data` data structures. Each node of this list is a record of one second. Every `annotating_data` node contains an Annotating list data structure which comprises the Event Flag, Mouse Positions, Keyboard Press Information and Time Interval to record the annotation events happened in that second. In the above `Annotation` data structure, the variable `eventFlag` consists of four kind of annotating operations, i.e., mouse down (d), mouse move (m), mouse up (u), and key press (k). The variable keyInfo can be any keys or special function keys such as page down, page up, and clear annotation (delete).While the one-second voice data obtained, the annotation list appends to the annotating_data data structure.

To facilitate the replay process, as shown in fig. 3, the annotated file contains two special indexes, i.e., the time index and the slide index. The time index is to indicate the time sequence of an annotated file. The slide index is to indicate the slide sequence. Users select this index number to change their replay starting slide or replay sequence. In a similar manner, users skip the undesired slides during the replay phase.

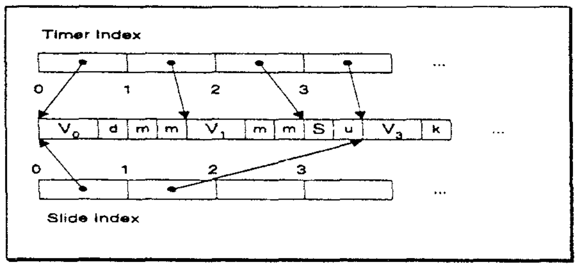

Figure 3. The Index Mechanism

5. Implementation

We implemented the proposed multimedia annotation and playback system on Microsoft Windows 95 platform by using Visual C++ [5]. Voice input signal is treated by a Sound Blast card which uses the PCM encoding scheme. To facilitate the interactive process, we design a graphical user interface (GUI) as shown in fig. 4. By using this interface, users can select to retrieve the desired annotated file or the original file. After an annotated file is loaded, users just click the start bottom to replay the annotated document. Under the above circumstances, the file is replayed in the speed as the annotating process. Fig. 5(a) depicts a Power Point document. The annotated file is shown in fig. 5(b). Due to the restriction of this paper presentation, the voice portion does not appear. However, from the file storage data structure one can see that voice and pointer movement are presented simultaneously and synchronously.

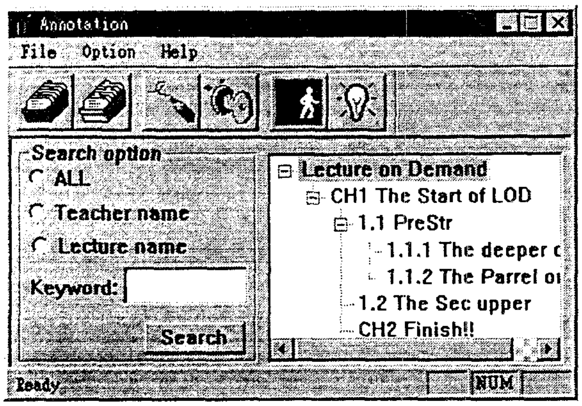

Figure 4. Lecture selection dialog box

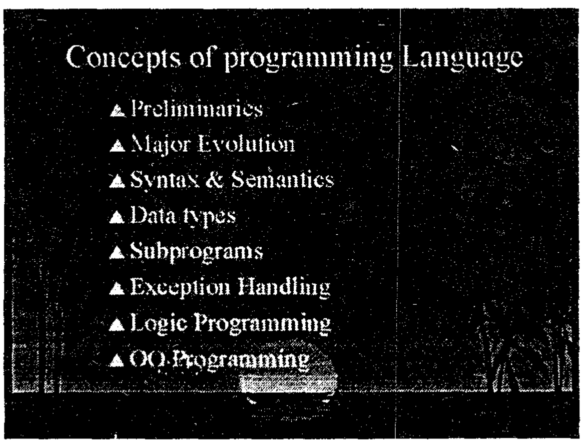

Figure 5a. A sample of before annotating

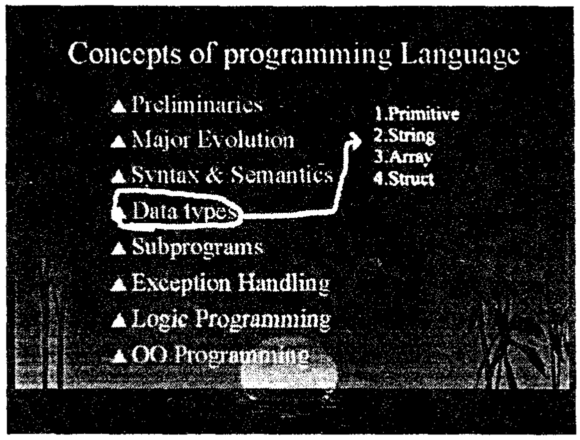

Figure 5b. A sample of after annotating

In the case of users want to interact with the system during replay phase. They just make a click on the screen then the designed GUI, i.e., fig. 6, will appear. By pressing the corresponding bottom, users can pause, forward, backward, and skip the slides. Furthermore, users can select a particular slide to play by using the index dialog box.

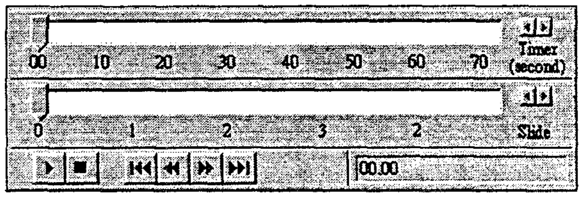

Figure 6. the index dialog box

6. Conclusions and future works

In the present paper, we design and implement a multimedia annotation and playback system. The proposed system can be used to enhance the presentation, instruction, and communication processes. Two time-dependent media, i.e., voice and pointer are chosen as annotation media to comment a prepared document. In our work, both temporal and spatial synchronization issues are taken into account. Furthermore, a file storage scheme and a replay mechanism are developed to facilitate user's interactive review process. In this manner, a lot of valuable information produced in the presentation process are retained.

At the present time, the annotation and playback system only considers the case of local multimedia systems. The extension to distributed networking environments is also beneficial to many other applications such as remote learning, computer supported cooperative work (CSCW). The part of work is currently underway.

7. References

[1] Borko Furht, *Multimedia Systems and Techniques*, Kluwer Academic Publisher, 1996.

[2] Francois Fluckinger, *Understanding Networked Multimedia Applications and Technology*, Prentice Hall, 1995.

[3] Lynane Ehley, Mohammad Iiyas, and Borko Furht, "A Survey of Multimedia Synchronization Techniques", IEEE Conference on Multimedia Computing and Systems, 1994.

[4] Ralf Steinmetz and Klara Nahrstedt, *Multimedia: computing, communications & applications*, Prentice Hall, 1995.

[5] Jeffrey Richter, *Advanced Windows (updated for Windows 95 & Windows NT 4.0)*, Third edition, Microsoft Press, 1996

[6] Badgett, Tom, *Teach yourself... PowerPoint for Windows*, MIS Press, New York , 1994