Semantic Synchronization: Producing Effective Reminiscence Videos for Dementia Sufferers

Noriaki Kuwahara, **Kazuhiro Kuwabara**, **Shinji Abe**, **Kiyoshi Yasuda**[*] and **Nobuji Tetsutani**[†]
ATR Intelligent Robotics and Communication Laboratories
2-2-2 Hikari-dai, Seika-cho, Soraku-gun, Kyoto 619-0288 Japan
{kuwahara, kuwabara, sabe}@atr.jp

Abstract

This paper proposes a framework of "semantic synchronization" to create a reminiscence video for people with dementia from their old photo albums by making use of Semantic Web technologies. To create an effective reminiscence video, visual effects such as zoom and pan, and audio effects shch as narrations should be added with the consideration for the semantics of these elements. We show how the proposed framework coordinates above two different types of media in a reminiscence video.

1 Introduction

Annotations attached to photos are used to facilitate sharing of those photos on the Web among users, as evidenced by the multitude of popular photo-sharing Web sites such as Flickr [Flicker]. These photo-sharing services provide various styles of presentation, popular one being a slide show. There is the type of a slide show video known "reminiscence video," which has been used as an effective means to present familiar images to people with dementia so that the caregivers can have some rest while a dementia patient concentrates on watching the video. Making a reminiscence video, however, is not an easy task for the family caregivers while they try to look after the family member suffering dementia. To make it easier to create a reminiscence video, we have proposed a way to make use of annotations attached to the photos [Kuwahara *et al.*, 2004]. In designing our authoring tool, we employed de facto standard vocabularies as much as possible such as Dublin Core [DublinCore], FOAF [FOAF], and Image Regions [ImgReg] to describe annotations, so that materials from the Internet or their own materials published can easily be incorporated in the future.

In a slide show video, not only a visual but also an audio effect, such as BGM (background music) or narration, plays an important role. More specifically, our previous experiments of showing the reminiscence video to people with dementia [Kuwahara *et al.*, 2005] revealed the following findings:

1. Adding narrations is certainly effective in producing an attractive reminiscence video because the dementia sufferers restored their attention to the video due to the dictated narrations even after they seemed to become bored with the video.
2. Zooming into familiar people's faces in a photo is also indispensable, because it relieves their difficulities in recognizing the people in the photos due to their loss of perception.
3. Synchronizing the narration and zooming is important. For instance, if the narration is about a specific person in the photo, this narration should be dictated as this person's face is zoomed in.

Synchronized Multimedia Integration Language (SMIL) [SMIL] has been proposed and adopted to present multimedia contents on the Web. This specification is mainly targeted towards the presentation of multimedia contents that have already been designed and created.

In contrast, we need to consider a more abstract semantic level of relationships between visual and audio elements when we create an attractive slide show video from a set of annotated photos. We call this level of synchronization of content elements "semantic synchronization," emphasizing the importance of coordination of visual and audio elements at the semantic level. In this paper, we demonstrate how "semantic synchronization" coordinate these different media elements in a reminiscence video.

2 Reminiscence Video Authoring Tool

Figure 1 shows the architecture of our annotation tool based on the photo and narration database, utilizing the standard vocabularies. Our tool consists of two parts. The first part provides a GUI for annotating photos in the database, and the second part generates a reminiscence video. The creator specifies keywords related to the theme of the video, and our tool retrieves the photos and narrations from the database by utilizing the annotation data. Regarding the narration database, typical narration texts are generated by using simple rules based on Japanese grammar and sets of keywords, and a professional narrator dictated these texts. The keywords are used as annotation data for the narration. After the creator selects the set of the photos, our tool automatically renders the reminiscence video.

[*]also with Chiba Rosai Hospital, Japan
[†]also with Tokyo Denki University, Japan

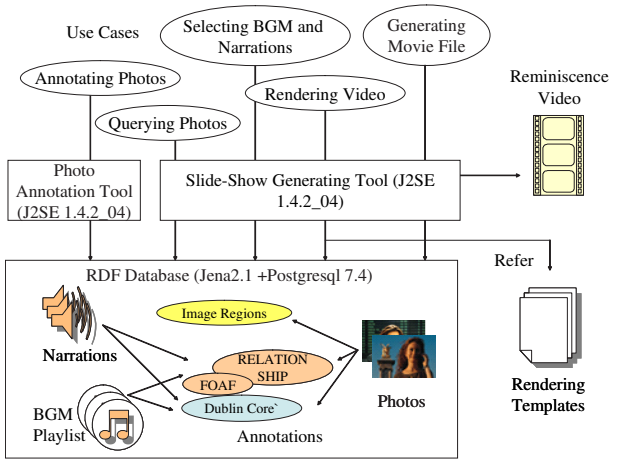

Figure 1: Annotation Tool Overview

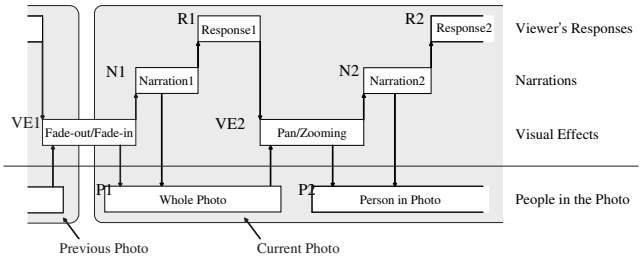

Figure 2: Semantic Synchronization Model

3 Semantic Synchronization

In the process of rendering with "semantic synchronization," the visual effects to be added are determined first. There are two types of visual effects. One is the "Fade-in/Fade-out" type that is applied when a photo displayed is being changed. The other type is the "Pan/Zooming" visual effect that focused the viewer's attention on certain regions of the photo. The latter effect should be accompanied by a appropreate narration. In order to satisfy the caregivers' needs, the region chosen for the "Pan/Zooming" effect should be synchronized correctly with the narration. In addition, we also need to consider the proper length of pause provided to viewers with sufficient time for responding to the narration.

Next, all the collected elements (the narrations, the pause, the visual effects, the persons in photos) are placed onto a video time-line, taking into account the semantic binding between the elements. Figure 2 gives an illustrative example of such an arrangement. This figure shows that N1 (Narration1) and P1 (Whole Photo) should be shown at the same time. N2 (Narration2) and P2 (Person in Photo) also exhibit the same relationship. There are also other kinds of relationships for arranging the elements on a time-line. In this example, we assume that P1 corresponds to the whole image of the photo.

Semantic synchronization is mainly concerned with creating temporal constraints between objects. The results of such constraints can be represented using the model of the interval temporal logic [Allen, 1984]. We describe below the constraints in Fig. 2, and we use this description as the template for rendering reminiscence videos.

```
{VE1 overlaps P1, VE1 meets N1, N1 during P1,
 N1 meets R1, R1 meets VE2, P1 overlaps VE2,
 VE2 overlaps P2, VE2 meets N2, N2 meets R2,
 N2 during P2, ...}
```

When we produce an interactive reminiscence video that can switch the photos, visual effects, narration and BGM according to the responses of the dementia sufferers, we need to reallocate each media element on the time-line under the limitation of the available content resources and the content playback time. The interval temporal logic can be used to dynamically adopt the media elements according to the viewer's reponse [Katsura, 2001].

4 Conclusion and Future work

This paper presented a concept of "semantic synchronization" between visual and audio elements to create an effective reminiscence video. We plan to evaluate the proposed concept by making a series of reminiscence video for dementia sufferers. Through this evaluation, we also plan to clarify the vocabulary to describe temporal constraints among various visual and audio elements, and integrate the vocabulary into the reminiscence video authoring tool.

Acknowledgments

We would like to thank Mr. Koji Saito for implementing our tool. This research was supported by the National Institute of Information and Communication Technology.

References

[Kuwahara *et al.*, 2004] Kuwahara, N. *et al.* Using Photo Annotations to Produce a Reminiscence Video for Dementia Patients. In *Demo Papers* of the 3rd International Semantic Web Conference (ISWC 2004), Japan, Hiroshima, November, 2004.

[Kuwahara *et al.*, 2005] Kuwahara, N. *et al.* Reminiscence Video - Helping At-Home Caregivers of People with Dementia -. In *Proceedings* of IFIP WG 9.3 HOIT 2005, U.K., York, April, 145–154, 2005.

[Allen, 1984] Allen, J.F. Towards a General Theory of Action and Time. Artificial Intelligence, Vol. 23, No. 2, 123–154, 1984.

[Katsura, 2001] Katsura, T. A Management Model for Synchronization Constraints of Interactive Multimedia Presentation (in Japanese). UNISYS TECHNOLOGY REVIEW, No. 71, NOV, 23–40, 2001.

[Flicker] `http://www.flickr.com`

[SMIL] `http://www.w3.org/AudioVideo/`

[DublinCore] `http://dublincore.org/`

[FOAF] `http://www.foaf-project.org/`

[ImgReg] `http://www.bnowack.de/w3photo/pages/image_vocabs`