PhotoLoop: Implicit Approach for Creating Video Narration for Slideshow

Keita Watanabe[1], Koji Tsukada[2,3], and Michiaki Yasumura[4]

[1] Meiji University, Japan
[2] Future University Hakodate, Japan
[3] Japan Science and Technology Agency, Japan
[4] Keio University, Japan
watanabe@gmail.com, tsuka@acm.org, yasumura@sfc.keio.ac.jp

Abstract. People often have difficulty in browsing a massive number of pictures. To solve this problem, we focused on the activities of people who share slideshows with their friends: that is, they often talk about the each picture shown on the display. We think these activities are useful as narrations for the slideshows. Therefore, we propose a novel slideshow system, PhotoLoop, which can automatically capture people's activities while watching slideshows using video/audio recordings and integrates them (slideshows and video narrations) to create attractive contents. In this paper, first, we describe people's behavior while watching slideshows. Next, we present the PhotoLoop prototype based on our observations. Finally, we confirm the effectiveness of the system through evaluation and discussion.

Keywords: Photograph, Slideshow, Narration, Implicit creation.

1 Introduction

As digital cameras and camera equipped mobile phones became popular in recent years, people began to take more pictures than ever before. They now face the problem of managing the large number of pictures. Many researchers focused on metadata related to each picture for effective picture management. These approaches can be divided into three categories: (1) using features extracted from pictures [1], (2) using sensor information collected during the capturing process [5, 6], and (3) using metadata manually created by users [3, 4, 5, 8]. In particular, subjective metadata created by users have advantages in reflecting their judgments or intentions which are difficult to be generated by computers. However, users often experience difficulty in adding metadata continuously.

To solve this problem, we focused on the activities of people who share slideshows with their friends: that is, they often talk about the current picture shown on the display. Chalfen also reported that people usually make conversations during slideshows to avoid silences [2]. We think these activities are useful as narrations for the slideshows. Therefore, we propose a novel slideshow system, PhotoLoop, which automatically captures people's activities while watching slideshows using video/audio recordings and

N. Streitz and C. Stephanidis (Eds.): DAPI/HCII 2013, LNCS 8028, pp. 87–96, 2013.

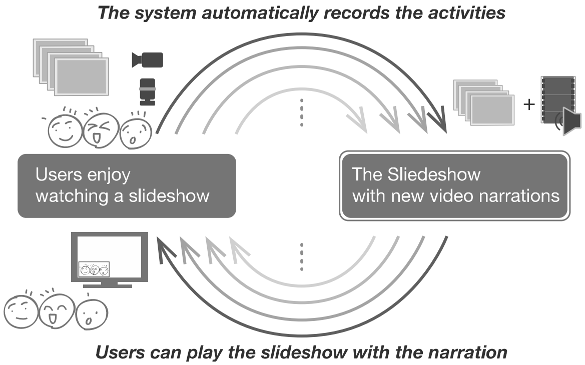
The system automatically records the activities
Users enjoy watching a slideshow
The Sliedeshow with new video narrations
+
Users can play the slideshow with the narration

Fig. 1. The concept of PhotoLoop

integrates them (slideshows and video narrations) to create attractive content. Using PhotoLoop, users can easily add video narrations to slideshows without special efforts (Figure 1).

2 Preliminary Study

We performed a preliminary study to record people's activities while watching slideshows using video/audio recordings and analyzed their conversations and behavior. The aim of this study is to explore the number, frequency, and content of people's conversations while watching a slideshow. Moreover, we also observed their behavior during the slideshow. We explain the procedure of the study. First, we selected 16 subjects (11male and 5 female, aged between 19 and 54) who all worked at the same laboratory. We prepared 30 pictures taken at a laboratory camp. 12 of the subjects had attended the camp. We classified them into four groups. Each group includes a subject who had not attended the camp. The slideshow was manually controlled by a subject who was randomly selected in each group.

2.1 Results

We recorded the activities of the users while watching the slideshow with video/audio and wrote down all conversations. Next, we summarized the number of conversations and the length of the slideshow for each group (Figure 2). On average, the subjects spent 11:32 min watching the 30 pictures and had 367 conversations. In other words, the subjects spent 23 seconds and had 12 conversations per picture. Next, we analyzed the contents of the conversations during the slideshows to explore their features. First, we found "Questions and answers" that is, a subject asked a question such as "What is it?!", "When did it happen?", or "Who is he/she?" and other members answered to the question. For example, a subject asked "Where is this road?" while watching Figure 3 (A) and another subject answered "The road is located under the ropeway".

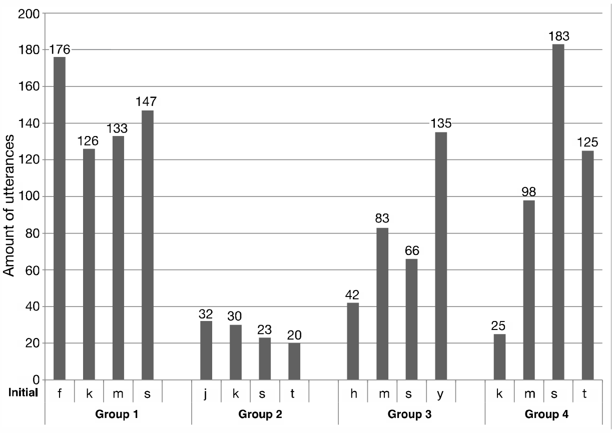

Fig. 2. Number of conversations for each group member

We also found "Recall by question": that is, a subject asked a question such as "It was a warm day, wasn't it?" to conform his/her memory. For example, a subject said "We were in the same group in the workshop, weren't we?" while watching Figure 3 (B), and another subject answered "Yes, we worked together!"

Some subjects mentioned their impression of the picture. For example, a subject mentioned "I felt fine because of the weather." while watching Figure 3 (C).

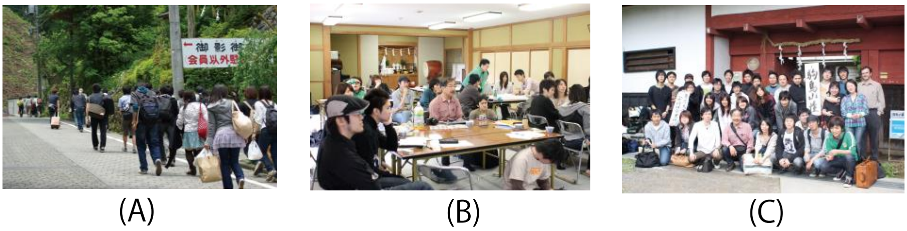
(A) (B) (C)

Fig. 3. Examples of pictures in preliminary study

As shown above, we observed many types of conversations related to the pictures. These conversations often contained information that was not found in the pictures themselves. Moreover, the conversations were sometimes different between groups. For this reason, we thought that the system can obtain various narrations for the slideshow by recoding conversations from multiple groups.

Furthermore, many subjects performed gestures to express their interests or emotions during conversations. For example, they often pointed at an object or a person in the picture with their fingers when talking on them.

3 PhotoLoop

We propose a novel slideshow system, called PhotoLoop, based on the results of the preliminary study. PhotoLoop can automatically capture people's activities while watching slideshows using video/audio recordings and integrates them (slideshows and video narrations) to create attractive contents.

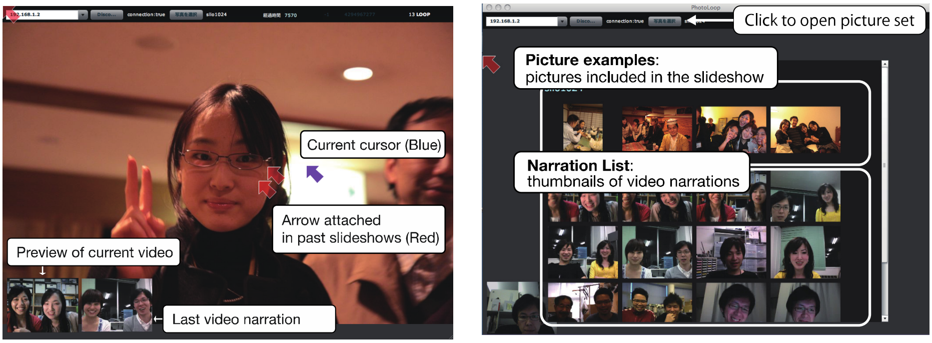

Fig. 4. Screenshot of the main window and slideshow browser

In this section, we explain the basic features of PhotoLoop: "video narration", "overlap recording", and "pointer logging".

Video Narration. As mentioned above, PhotoLoop automatically records users' activities using a camera and a microphone while they watch slideshows. In this paper, we call these video/audio recordings as "video narration": the video includes facial expressions/ gestures and the audio includes explanations/impressions of the pictures. Figure 4 (right) shows a screenshot of the slideshow browser. The system presents thumbnails of pictures and video narrations included in the slideshow. Users can select multiple video narrations that are shown during the slideshow. When the users push the start button, the system presents the selected video narrations below the screen along with a preview of real-time video (Figure 4 (left)). Although all videos are played at the same time, the system plays only one narration (audio track) to avoid confusion. The users can easily hear another narration, just by clicking another video. When they finish watching the slideshow, the system automatically creates a new video narration and saves it to the database. Thus, PhotoLoop helps users create a video narration just by watching slideshows without any special operation.

Overlap Recording. As mentioned above, PhotoLoop can record the activities of users while they watch the slideshow with video narrations created at the previous slideshow. We call this feature "overlap recording". Using this function, the system can add various explanations and impressions to the pictures from multiple viewpoints. Moreover, some people may attach unexpected narrations while hearing previous narrations by another group. Thus, PhotoLoop can add further attractiveness to the slideshow by recording new video narrations while the users watch the slideshow with past narrations.

Pointer Logging. As mentioned in the results of the preliminary study, users performed various gestures to explain pictures while they watched the slideshow. Here, we focused on one of the most popular gesture, “pointing at a person/object in the picture”, to express the main target. To record this action in a practical way, we provide the “pointer logging” function to help users add arrow cursors to the picture just by clicking the mouse button, as shown in figure 4 (left). We prepared two types of pointing device: a gyro mouse and a common mouse. While the gyro mouse was suited for casual slideshows (e.g., in a living room), we also supported a common mouse for general use. The cursors recorded in past slideshows are shown as red arrows. Figure 5 shows the basic usage of the PhotoLoop.

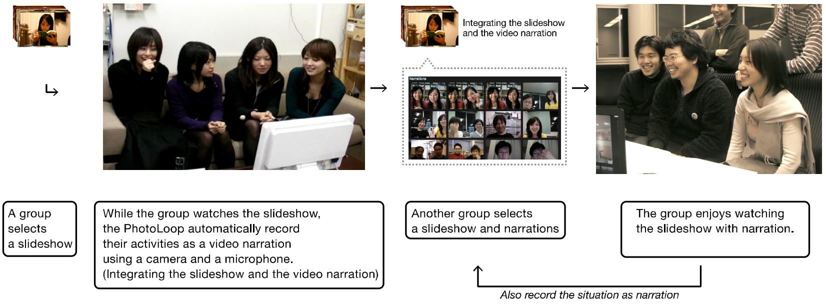
Integrating the slideshow and the video narration
A group selects a slideshow
While the group watches the slideshow, the PhotoLoop automatically record their activities as a video narration using a camera and a microphone. (Integrating the slideshow and the video narration)
Another group selects a slideshow and narrations
The group enjoys watching the slideshow with narration.
Also record the situation as narration

Fig. 5. Basic usage of the PhotoLoop

The PhotoLoop mainly consists of a PC (Windows XP), an LCD, a USB camera, a microphone, and a mouse, as shown in Figure 6. The system records video and audio using the USB camera and the microphone. We developed frontend software to control the above devices and play slideshows and video narrations using Adobe Air, as well as backend software to record video narrations using Adobe Flash Media Server . The resolution of the video is 320 x 240 pixels and the frame rate is 15 fps.

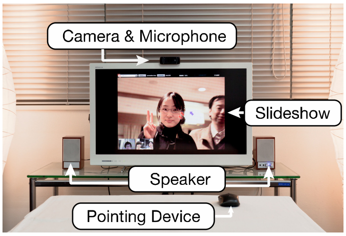

Fig. 6. System architecture of PhotoLoop

3.1 Scenarios

In this section, we describe three usage scenarios of the PhotoLoop.

Watching Slideshows after Trips. Most people take many pictures when they go on trips. They often watch these pictures together as a slideshow after they come home. Although this process often becomes an important part of their memories of the trip, most people do not record their activities while watching the slideshow. PhotoLoop helps users enrich their memories of trips by recording these activities without effort.

Sharing Video Narrations with Friends. When people watch slideshow after an event using the PhotoLoop, the system can generate a video narration that contains users' subjective impressions and explanations of the slideshow. We think these narrations are useful, especially for their friends/families who did not attend the event. For example, when users want to share the slideshow of their family trip with their grandparents who live apart, the grandparents probably prefer the slideshow with video narrations. Moreover, when their grandparents use PhotoLoop to watch the slideshow, the initial users can easily check the reactions of the grandparents.

Simple Annotations to Pictures. PhotoLoop can add a score to each picture – showing the importance of the picture within the slideshow – using the audio level of the narration and number of arrow cursors. Moreover, when the system extracts texts from the narrations using speech recognition techniques, these texts may work as annotations to the pictures. Although this method has a problem with recognition rate, the system can possibly help users add annotations to pictures just by watching slideshows as usual.

4 Evaluation

In this section, we verified the effectiveness of PhotoLoop through an evaluation. The main aim of this evaluation was to explore whether the video narrations can improve the attractiveness of the slideshow. We selected eight subjects (all males, aged between 19 and 27) who regularly use computers and digital cameras, and we randomly divided them into four pairs. First, the experimenter asked each pair to watch a slideshow using PhotoLoop twice. When they watched the slideshow at first, the system presented the slideshow alone. The second time, the system showed the slideshow along with the video narration that had been recorded with another pair. The second slideshows were performed after all first slideshows were finished. The slideshow included 15 pictures used in the preliminary study. After the second slideshows, the experimenter obtained subjective feedback from the subjects through questionnaires and oral interviews. Before the evaluation, the experimenter explained the basic function of PhotoLoop; that is, the system automatically records the subjects' activities during the slideshow and presents them to another pair.

In addition, we provided two wireless mice to allow both subjects in each pair to control the arrow cursor and the slideshow.

4.1 Results

To gain subjective feedback, we set two questions and scored the answers on a scale of 1 to 5 as follows: "Q.1: Did you enjoy watching the slideshow? (1: very boring – 5: very amused)" and "Q.2: Did you discover new information from the slideshow? (1: no information – 5: plenty of information". The average score of Q.1 was 4.75 (S.D.=0.46) at first slideshow and 4.25 (S.D.=0.70) at second slideshow. The average score of Q.2 was 4.25 (S.D.=0.46) at first and 4.38 (S.D.=0.51) at second.

Most participants responded favorably to Q.1 and Q.2 for both the first and second slideshows. Here we shows typical comments from the subjects as follows:

1. I was very amused just by hearing the comments from the other group.
2. I was amused by the person who said whatever he felt. The conversations between A and B were interesting.
3. The conversations between C and D reminded me of the details of the event. I preferred when the conversations occurred less frequently. I was impressed that the other group explained the picture from a different viewpoint.
4. I felt pleased to see that other people enjoyed the slideshow.
5. I was favorably impressed by C since he remembered various details of the event.
6. I was little bothered by the second slideshow since the pictures were the same.
7. I could enjoy the second slideshow because of the video narrations.
8. I enjoyed adding arrow cursors.
9. I sometimes focused on listening to the narration

4.2 Consideration

In this section, we consider the results of the evaluation in terms of "video recording", "pointer logging", and "video narration".

Video Recording. We had been concerned that users might hesitate to talk in front of the camera and microphone. However, most subjects seemed to act as usual while watching the slideshow as shown in the comment (B). Meanwhile, we observed several inappropriate conversations, such as "He looked like an awful monster!" and "His shirt was quite twisted around his neck. That looked so bad!". These "frank" comments may become attractive contents; however, we should also consider the risk that these comments may have negative influence on human relations.

Pointer Logging. In this evaluation, we observed more conversations including reference terms (e.g., "that" or "this") than those in the preliminary study. This change may have resulted from the "arrow cursor" function: that is, users often have conversations like "What's this?" when they add arrow cursors in the pictures. We found these conversations quite interesting as we had not expected the arrow cursors to affect the conversations. Meanwhile, several subjects talked about objects just by pointing to them with the cursor (without clicking). Since the current system recorded cursor positions only when the subjects clicked the mice, others could not understand the meaning of the reference terms in such cases. To solve this problem, we plan to develop another logging/display method that automatically records/visualizes the cursor movement and emphasizes the cursor when clicked. In addition, we found a unique use of the arrow cursors: some users draw pictures using arrow cursors; others attached arrow cursors archly to everywhere in the picture. These are also interesting findings for us since we had not expected these usages of the arrow cursor.

Video Narration. Some subjects informally watched the slideshow with their own narration after the evaluation. In such cases, the subjects often reacted to their own conversations in the narration: for example, they often started laughing in response to their previous laughter. Next, we observed comments from the subjects who were the friends of persons shown in the pictures as follows: “I easily understood the situation though I had not participated in the event” and “I found the arrows were useful to understand the conversations. I really like it!”. Meanwhile, the number of conversations at the second slideshow was less than that at the first slideshow. These changes may arise from the fact that users were sometimes too focused on the narration and pictures and forgot to talk with each other as shown in the comment (I).

5 Related Work

There are three approaches to help users add subjective metadata: (1) supporting users to attach metadata manually [3, 4], (2) attaching metadata while capturing pictures [5, 6, 7], and (3) extracting metadata from the picture sharing process [8, 9]. Shneiderman [3] proposed a system that helps users attach labels (e.g., personal names) to pictures by drag&drop. Sigurbjörnsson [4] proposed a system that recommends tag candidates using WordNet. In contrast, WillCam [5] is a novel digital camera that can help photographers add their ideas regarding pictures via pointers. ContextCam [6] proposes a context-aware video camera that provides time, location, persons and event information using several sensors and machine learning techniques. Capturing the Invisible [7] designs real-time visual effects for digital cameras using simulated sensor data. Moreover, Aria [8] focuses on communication using pictures; that is, sending pictures by e-mail. The system can automatically create descriptions of pictures from messages written in the e-mail. Chi et al. [9] also proposed a system that supports users attaching annotations to picture sets through text chats. PhotoLoop is unique in helping users add video narrations to slideshows in a simple manner by automatically recording the users’ activities (e.g., conversations and behavior) during the slideshows.

Balabanovic et al. [10] proposed a method to add narrations to a picture set efficiently. Their research is similar to PhotoLoop in that it focuses on picture narrations. However, their system aimed to record only intentional narrations: the user manually pushes the record button and starts explaining the picture. While this system is suited for creating accurate narration, it requires users' motivation and attention to create narrations. Moreover, this system cannot record new narrations while playing previous ones. PhotoLoop is unique in automatically creating video narrations of slideshows by recording the users' activities during the slideshows using a video/audio system.

There are many research projects that focused on relationship between people and photographs [14,15,16,17,18,19]. For example, David et al. [15] had pointed that some photo sharing conversations incorporate comments about the meaning and value of photos to those who took them. They have been discussing that these comments and conversations would be good to save and associate with the photos for future personal reference and consumption. Almost all works investigated how people communicate with photographs in daily life. They have not developed system yet based on the result. However, these works will be meaningful to improve PhotoLoop in the future.

There have been several projects that focused on communication by sharing video/audio data. CU-Later [11] is a communication support system that uses video messages in remote locations and different time zones. This system records users' activities while eating with a video/audio system and shares them with friends/family in remote locations. LunchCommunicator [12] supports communication between family members (the lunch-creator and the lunch-consumer) using automatic capturing/playing techniques during the preparation/consumption of the lunchbox. The EyeCatcher [13] helps photographers capture a variety of natural looking facial expressions of their subjects by keeping the eyes of the subjects focused on the camera without the stress usually associated with being photographed. PhotoLoop is unique in focusing on users' unconscious reactions while watching slideshows and utilizing them as video narrations for the slideshows.

6 Conclusion

In this paper, we propose a novel slideshow system, PhotoLoop, which automatically captures people's activities while watching slideshows using a video/audio recording system and integrates them (slideshows and video narrations) to create attractive content. We designed a prototype based on observations of subjects while watching slideshows and confirmed the effectiveness of the system through evaluation and discussion.

Acknowledgements. This work was partly supported by JST PRESTO program.

References

1. Veltkamp, R.C., Tanase, M.: Content-Based Image Retrieval Systems: A Survey. Technical Report UU-CS-2000-34, Utrecht University (2000)
2. Chalfen, R.: Snapshot versions of life. Bowling Green State University Press, Bowling Green OH (1987)

3. Shneiderman, B., Kang, H.: Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos. In: Proceedings of InfoVis 2000, pp. 88–95 (2000)
4. Sigurbjörnsson, B., van Zwol, R.: Flickr tag recommendation based on collective knowledge. In: Proceeding of WWW 2008, pp. 327–336 (2008)
5. Watanabe, K., Tsukada, K., Yasumura, M.: WillCam: a digital camera visualizing users' interest. In: Extended Abstracts of ACM CHI 2007, pp. 2747–2752 (2007)
6. Patel, S.N., Abowd, G.D.: The ContextCam: Automated Point of Capture Video Annotation. In: Mynatt, E.D., Siio, I. (eds.) UbiComp 2004. LNCS, vol. 3205, pp. 301–318. Springer, Heidelberg (2004)
7. Hakansson, M., Ljungblad, S., Holmquist, L.E.: Capturing the invisible: designing context-aware photography. In: DUX 2003: Proceedings of the 2003 Conference on Designing for User Experiences, pp. 1–4 (2003)
8. Lieberman, H., Rosenzweig, E., Singh, P.: Aria: An Agent For Annotating And Retrieving Images. IEEE Computer 34(7), 57–61 (2001)
9. Chi, P.-Y., Lieberman, H.: Intelligent assistance for conversational storytelling using story patterns. In: Proceedings of ACM IUI, pp. 217–226 (2011)
10. Balabanovic, M., Chu, L.L., Wolff, G.J.: Storytelling with digital photographs. In: Proceedings of ACM CHI 2000, pp. 564–571 (2000)
11. Tsujita, H., Yarosh, S., Abowd, G.: CU-Later: A Communication System Considering Time Difference. In: Adjunct Proceedings of Ubicomp 2010, pp. 435–436 (2010)
12. Kotani, N., Tsukada, K., Watanabe, K., Siio, I.: LunchCommunicator: Communication Support System using a Lunchbox. In: Adjunct Proceedings of Pervasive 2011 pp. 9–12 (2011)
13. Tsukada, K., Oki, M.: EyeCatcher: A digital camera for capturing a variety of natural looking facial expressions in daily snapshots. In: Floréen, P., Krüger, A., Spasojevic, M. (eds.) Pervasive 2010. LNCS, vol. 6030, pp. 112–129. Springer, Heidelberg (2010)
14. Crabtree, A., Rodden, T., Mariani, J.: Collaborating around collections: informing the continued development of photoware. In: Proceedings of ACM CSCW 2004 (2004)
15. Frohlich, D., Kuchinsky, A., Pering, C., Don, A., Ariss, S.: Requirements for photoware. In: Proceedings of ACM CSCW 2002 (2002)
16. Martin, H., Gaver, B.: Beyond the snapshot from speculation to prototypes in audiophotography. In: Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 55–65 (2000)
17. Swan, L., Taylor, A.S.: Photo displays in the home. In: Proceedings of the 7th ACM Conference on Designing Interactive Systems (DIS 2008), pp. 261–270 (2008)
18. Taylor, A.S., Swan, L., Durrant, A.: Designing family photo displays. In: Proc. ECSCW 2008, pp. 79–98. Springer, London (2008)
19. Voida, A., Mynatt, E.D.: Six themes of the communicative appropriation of photographic images. In: Proceedings of ACM CHI 2005, pp. 171–180 (2005)