Location-based Image Viewing System Synchronized with Video Clips

Yuanyuan Wang*, Maho Nishizawa **, Yukiko Kawai ***, Kazutoshi Sumiya**

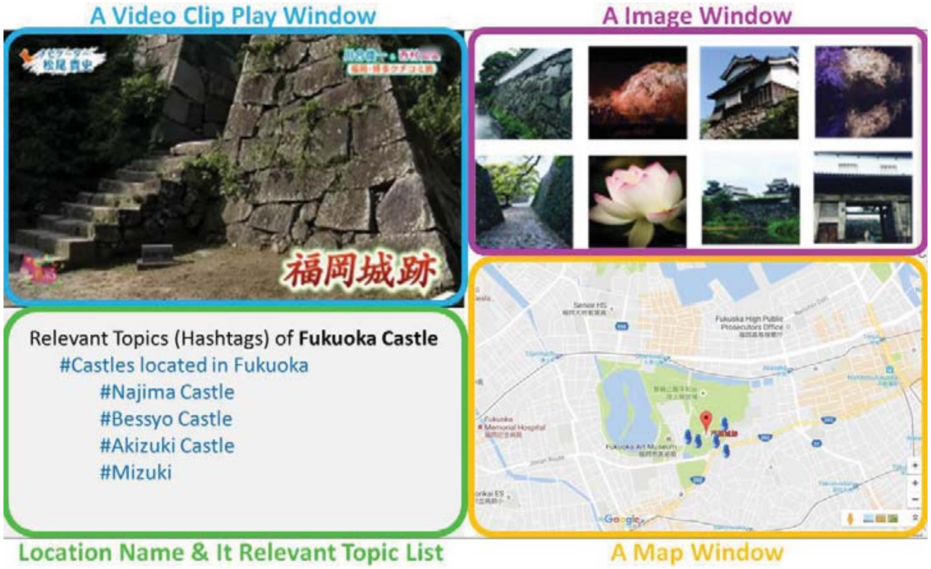
* Yamaguchi University, Japan
** Kwansei Gakuin University, Japan
*** Kyoto Sangyo University, Japan

1. Introduction

In recent years, various video clips, such as travel and educational videos, mainly provide a high-affinity geographic data, and the video clips in TV programs are often associated with closed captions. While watching TV programs, users are probably interested in some contents in the video clips and search related information through the Web. For example, users may search locations or check the current weather of tourist spots appeared on a travel channel using smartphones or tablets. However, current services cannot present location-based contents, such as location-based images, geographic maps, synchronized with video clips, and users difficult to grasp the surroundings of the geographic data, how the locations are related, and distances between them during the video clips. In particular, it is difficult to search appropriate location-based images from the huge number of users' posts on photo sharing sites, such as Instagram (2012) and Flickr (2005). For example, a user wants to grasp the surrounding scenery of Sanda that is a city located in Hyogo Prefecture, Japan by using a search query, Sanda. The search results also contain images about Sanda of Tokyo, because the search query is not appropriate. Therefore, it is necessary to analyze the semantics of the geographic data in a video clip, and supplement the video clip automatically with related information (e.g., location-based images, geographic maps).

In this work, the goal is to develop a novel automatic location-based image viewing system synchronized with video clips based on the concept of second screen service (Nandakumar & Murray 2014), (Geerts, Leenheer, Grooff, Negenman & Heijstraten 2014) by analyzing semantic structures of video clips (see Figure 1). To achieve our goal, we first extract location

Published in "Proceedings of the 13th International Conference on Location-Based Services", edited by Georg Gartner and Haosheng Huang, LBS 2016, 14-16 November 2016, Vienna, Austria.



Relevant Topics (Hashtags) of **Fukuoka Castle**

- #Castles located in Fukuoka
  - #Najima Castle
  - #Bessyo Castle
  - #Akizuki Castle
  - #Mizuki

Figure 1. An interface of the location-based image viewing system with video clips.

names which appear in closed captions of a video clip and detect their scenes of the video clip. Then, the system extracts a semantic structure that is a tree structure of extracted location names by utilizing Wikipedia categories, and detects relevant topics of location names in the semantic structure. Once a user watches a video clip, the system presents each location name appears in each scene with its relevant topic list, and images of Instagram related to each location name, are synchronized with the video clip. Also, the user can select any topic in the topic list; the system then presents images of Instagram related to the selected topic. Images can help the user easily grasp the surrounding scenery and appearances of locations in the video clip, and they can also rouse the user's interests. Location names and their relevant topics can also help the user easily obtain details and relevant knowledge of locations in the video clip.

As a result, our proposed novel system enables users to view images and location names with their relevant topics satisfy and joyfully during a video clip without additional search.

2. System Overview

The process flow of our automatic location-based image viewing system synchronized with video clips is described as follows:

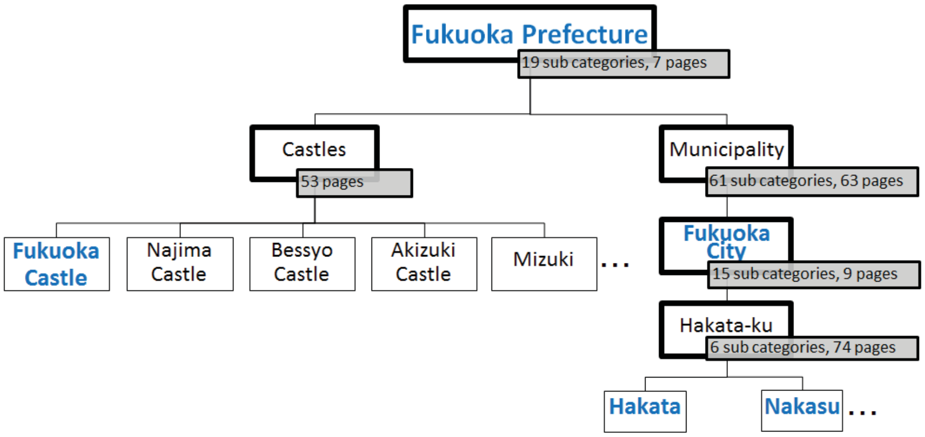


Figure 2. An example of a tree structure.

(1) Detecting location names and scenes of a video clip.

The system extracts location names using a morphological analyzer and their appearance time in closed captions from a MPEG-2 Transport Stream file of a video clip. In this work, we define one scene describes one location name. The system then divides the scenes of each location name if the next location name appears in the closed captions along the timeline of the video clip. In addition, if the same location names continuously appear in short time, they are determined in one scene. Furthermore, one scene is too short if the next location name appears within *T* seconds, this next location name will be eliminated. The time of each scene *T* is set to 3 seconds.

(2) Extracting semantic structure of a video clip.

The system extracts a tree structure of location names from Wikipedia categories using Wikipedia API (2006). Then, the system detects relevant topics if they have parallel relationships with location names from the semantic structure. A tree structure of a travel program called "New discovery! Tabipula" is shown in Figure 2, blue characters denote location names detected from closed captions of this video clip. "Hakata" and "Nakasu" have the parallel relationship, and "Fukuoka City" contains "Hakata." Therefore, geographical relationships between every two location names can be determined by using the extracted tree structure. However, a tree structure has a lot of nodes, and one location name may in many different tree structures. To solve them, we determine important categories if they contain a large number of reference pages. Then, we filter nodes of a tree structure if the reference pages of each node more than 5 pages. In addition, we select a tree structure of one location name by comparing the number of reference pages of its superordinate categories in different tree structures.

(3) Presenting images synchronized with video clips.

The system presents location names and their relevant topics in a list. Also, the system searches images from Instagram by matching hashtags and location names (relevant topics). Here, the system presents top-8 images by counting the number of "Like" from other users in Instagram.

3. User Study and Discussion

The purpose of this user study is to evaluate the effectiveness of presenting relevant topics and images with video clips. We evaluate four patterns of the same scenes from video clips as follows:

- (a) Location names synchronized with scenes (detailed information: text)
- (b) Relevant topics synchronized with scenes (detailed information + relevant information: text)
- (c) Images of location names synchronized with scenes (detailed information: images)
- (d) Images of relevant topics synchronized with scenes (detailed information + relevant information: images)

The video clips used for evaluation is from two travel programs in Japan, respectively. One is called "New discovery! Tabipula" introducing Fukuoka Prefecture in Japan with a total viewing time about 5 minutes 30 seconds; the other one is called "Eetoko" introducing Shiga Prefecture in Japan with a total viewing time about 2 minutes 30 minutes.

The study is completed by six participants, who never been to Fukuoka Prefecture and Shiga Prefecture. They completed the following 4 items (**Content Understanding**: *Q*1, **Interest-Arousing**: *Q*2, **Supplement Effects**: *Q*3, *Q*4) in a questionnaire after they watched the four patterns (a)~(d) of the same scenes from video clips.

- *Q*1: Could understand the video clips.
- *Q*2: Felt spread your interests.
- *Q*3: Write down relevant topics and images that are not related to the video clips

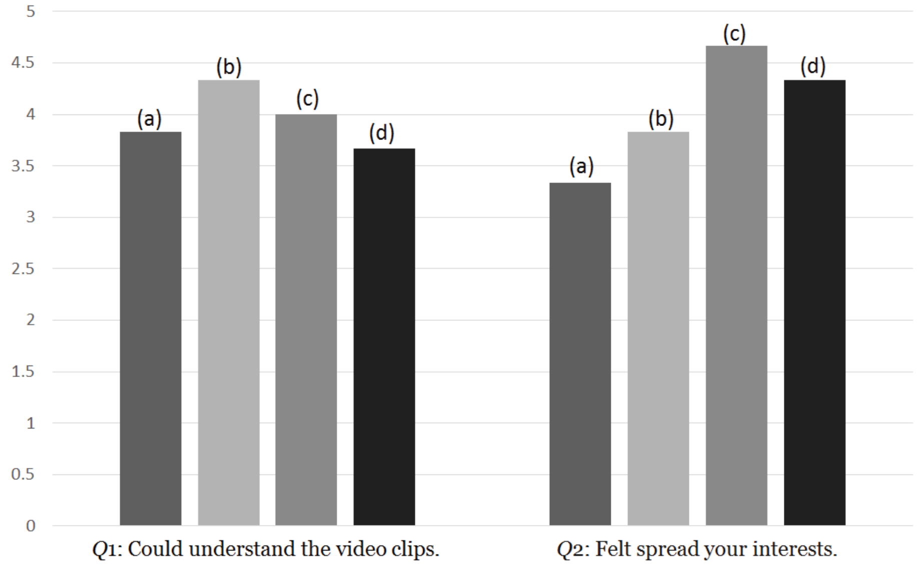


Figure 3. The results of *Q*1 and *Q*2 in the questionnaire.

- *Q*4: Write down relevant topics and images that you are interested in.

Figure 3 illustrates the average ratings of *Q*1 and *Q*2 in the questionnaire based on a five-level Likert scale. High user ratings indicate good results. The results and findings are shown as follows:

- *Q*1 for (b) gains a high rating, it can be said that it is most effective for understanding the video clip when presenting relevant information in text synchronized with scenes.
- *Q*2 for (c) and (d) gain a high rating, the ratings of images are higher than those of location names and their relevant topics in the text.
- In *Q*3 for the low relevant topics and images written by subjects, subjects felt several unknown topics are not related to the video clips. We need to present the relevances of locations in the video clips and their relevant topics. In addition, several general images about people or flowers are not related to locations in the video clips. It is necessary to extract images by considering features of locations.
- In *Q*4 for the interesting topics and images written by subjects, subjects are interested in when they watch the video clips. Then, the proposed

method is able to arouse the subjects' interests and can help the subjects enjoy the video clips.

4. Conclusion

In this paper, we developed a novel automatic location-based image viewing system synchronized with video clips based on semantic structures of the video clips. In order to extract semantic structure of a video clip, we utilize closed captions of the video clip. The system creates a tree structure of location names in the video clip by using Wikipedia categories and detects relevant topics of location names in the tree structure. Finally, we conducted a user study with the proposed system. The results of user study showed that the proposed system can help users enjoy video clips with relevant topics and images that suit users' interests.

In the future, we plan to improve the methods for presenting relevant topics and images by considering the meanings of hashtags in Instagram, and consider presenting relevant information of other types (e.g., voice, web pages, and microblogs). Another future direction is to extend the system by user interactions. For example, the system can be extended to allow viewers to decide whether and how to show the supplementary information.

Acknowledgment

This work was partially supported by JSPS KAKENHI Grant Numbers 26280042, 16H01722.

References

Geerts, D., Leenheer, R. D., Grooff, D., Negenman, J., & Heijstraten, S. (2014). In front of and behind the second screen: Viewer and producer perspectives on a companion app. In TVX 2014: Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video (pp. 95-102). ACM, New York, NY, USA.

Flickr (2005) Yahoo!. https://www.flickr.com/. Accessed 23 August 2016

Instagram (2010) Facebook. https://www.instagram.com/. Accessed 23 August 2016

Nandakumar, A., & Murray, J. (2014). Companion apps for long arc TV series: Supporting new viewers in complex storyworlds with tightly synchronized context-sensitive annotations. In TVX 2014: Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video (pp. 3-10). ACM, New York, NY, USA.

Wikipedia API (2006) SampleAPI. http://wikipedia.simpleapi.net/. Accessed 23 August 2016