A Web-Based Approach for Automatic Composition of an Insightful Slideshow for Personal Photographs

Kotaro Yatsugi[1], Naomi Fujimura[2], and Taketoshi Ushiama[2]

[1] School of Design, Kyushu University 4-9-1 Shiobaru, Minami-ku, Fukuoka, 815-8504 Japan
yatsugi@gospel.aid.design.kyushu-u.ac.jp
[2] Faculty of Design, Kyushu University 4-9-1 Shiobaru, Minami-ku, Fukuoka, 815-8504 Japan
{fujimura,ushiama}@design.kyushu-u.ac.jp

Abstract. Recently, the number of digital content objects is increasing rapidly with the progress of information technology. It has become important how we manage enormous digital content objects effectively and utilize them efficiently. Up to now, a lot of researches on digital content management have been reported. One of the important objectives of conventional management techniques is to search digital content objects that satisfy an information request of a user. This is based on the assumption that a user has one or more information requests. However, a user may have no information request explicitly when the user uses some kinds of devices for presenting digital content such as a digital photoframe. Such devices are expected to provide a presentation of digital content that rouses user's interest. In this paper, we introduce an approach for composing photo slideshow that attracts user's interest automatically.

1 Introduction

Digital cameras and cellular phones with camera have widely spread in human societies. In Japan, the number of households that have one or more digital cameras increased from 25% in March 2002 to 65% in March 2008[1]. The running cost of a digital camera is relatively lower than that of a film camera, and operations of a digital camera is easier than those of a film camera. Today, various types of people take photographs in everyday life, and enormous new photographs arise day by day. So a user has to manage a large amount of photographs. It is one of the most important issues how to manage a lot of personal photographs easily and effectively. On the other hand digital photographs are used for variety kinds of purposes. Effective utilization of a lot of personal photographs is another issue.

Up to now, many techniques for efficient management and effective utilization of personal digital photographs have been reported. Most of the conventional techniques for personal photograph management could be categorized into three types: search, browse and recommendation. The objective of search and browse is

J.D. Velásquez et al. (Eds.): KES 2009, Part II, LNAI 5712, pp. 623–630, 2009.

to find a photograph that satisfies a requirement of user. When a user searches photographs, an information request of the user is represented explicitly as a query. When a user browses photographs, the user can access photographs navigationally with simple operations. On the other hand, in recommendation a user does not have to give any query and operation. The system proposes some photographs to the user based on access logs and profile of the user. The above three approaches for providing photographs to a user suppose that the user has had one or more information requirements previously. The objective of those approaches is to find one or more photographs, which satisfies an information requirement of the user. We named them as *requirement-based approach*.

Many software and web services based on the requirement-based approaches have been designed based on the assumption that they are used on a personal computer and their objective is to find and provide one or more photographs which could satisfy an information requirement of a user. Figure 1 illustrates an overview of the requirement-based approach.

Today, many kinds of application software and web services for managing personal photographs based on the requirement-based approach have been provided. For example, Flickr, which is one of representative web services for personal photograph management, displays thumbnails of photographs on a map based on their location information and a user can browse them intuitively with easy and simple operations. It is possible to fulfill the requirement such as "I want to see the photograph taken in a location", "I want to know the place where a photograph was taken" and so on. Some digital cameras and cellular phones have GPS function and they can assign location information (latitude and longitude) to a photograph automatically.

However, recent few years, novel types of device such as digital photoframe for managing and utilizing personal photographs have attracted many attentions. The main objectives of a digital photoframe are to store a lot of digital photographs with easy operations and to display them continually. Most of digital photoframe displays stored photographs as a form of slideshow. A slideshow is expected to give a user comfortable feeling or to attract interest of a user. In order to fulfill such expectation, the conventional requirement-based approach is insufficient. This is because the requirement-based approach depends on information requirements of a user. However, a user may have no information requirement when the user watches a slideshow on a digital photoframe.

In this paper, we propose a novel approach for managing and utilizing photographs, which are named *interest-based approach*. This approach is used for composing a slideshow of a target set of personal photographs of a user. Up to now, some techniques for composing a slideshow have been reported. However, in the conventional techniques, photographs are ordered in terms of their file names, shooting date and time or at random. These conventional techniques are inefficient for attracting interest of a user. In our approach, the system supposes that a user has had no information requirement and composes a slideshow of personal photographs that arises interests in the user. We named such a slideshow as *insightful slideshow*. Figure 2 shows an overview of interest-based approach.

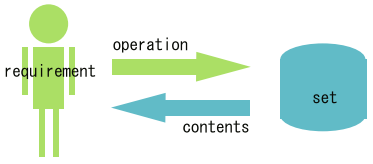

Fig. 1. Overview of Requirement-based approach

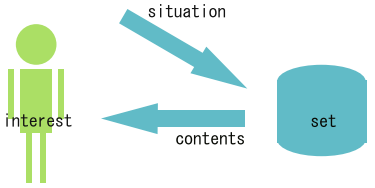

Fig. 2. Overview of Interest-based approach

We suppose that each personal photograph in a target set has a tag manually or automatically. In order to compose an insightful slideshow, the weight of a semantic relation between two photographs is calculated using tags assigned to the photographs and the Web. We suppose that co-occurrence of the tags in the Web reflects relation between the photographs. The photographs are organized into a network and an optimized route in the network is found. The route represents an effective order of photographs as an insightful slideshow.

The composition of this paper is as follows. Section 2 describes some related works. Section 3 show on overview of our approach. Section 4 describes our algorithm for compose an insightful slideshow of stored personal photographs. The summary and future plans are described in Section 5.

2 Related Works

Ba-log[2] manages photographs with the location information where a photograph has been taken. Ba-log provides an interface in which a thumbnail of each photograph is arranged on a map based on its location information. With this interface a user can browse photographs according to their geographical location.

Iwazaki et al.[3] have been proposed a technique for indexing photographs. The geographical location and shooting direction of a photographer when the photographer took a photograph are used for indexing the photograph. Pairs of keyword and location information corresponding to the keyword are stored in a database. The system finds new keywords in the database to a photograph based on its geographic location and shooting direction and recommends some of them relative to the photograph. A user selects suitable keywords from the proposed keywords, and assigned to the photograph as tags. A user can search photographs using assigned tags from various viewpoints.

Fujita et al.[4] has proposed a search interface for photographs based on gaze points of a photographer. In the interface, not only the shooting point but also the gaze point, which is the photographic subject, is considered for browsing photographs. The shooting vector of a photograph is defined as a vector from its shooting point to its gaze point. When a user clicks an object on a map, the clicked point is assumed to be a gaze point. Shooting vectors whose photographs contain the gaze point are searched from the database. The shooting vectors are represented as search results. By selecting a shooting vector in the result, the user can browse photographs about the clicked object.

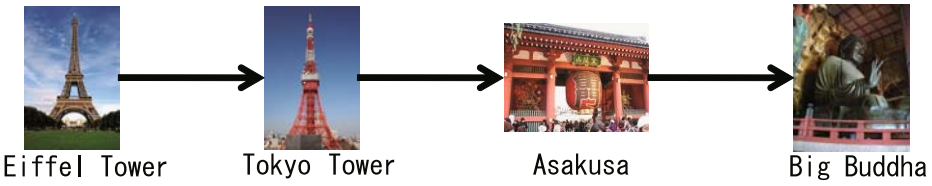

Fig. 3. An Example of Insightful Slideshow

SpaceTag [5] is a system for public information service on the geographic space. A user can make and browse virtual objects on the geographic space with a mobile device. Virtual objects in this system can be accessed in a specific location, date and time.

These systems provide easy and intuitive access of photographs using maps or geographic space. However, they are based on the requirement-based approach, so could not be applicable to attract user's interest.

Jin et al.[6] and Ishida et al.[7] use the Web as knowledge to extract an interpersonal relation network. A relation between persons in a network is estimated based on the co-occurrence of their names on the Web. When the co-occurrence of the names of two persons is high, it is considered that the two persons are closely related. The co-occurrence is derived from the number of web pages in a result of a web search. The weight of a relation is calculated with the Simpson Coefficient. In this paper, we propose a method that derives relations between photographs using the web. This is similar to the method for deriving interpersonal relation network. However, our objective is to compose a slideshow of photographs based on a derived relation network.

3 Insightful Slideshow of Photographs

The objective of our research is to develop a novel method for composing an insightful slideshow of personal digital photographs. An insightful slideshow could attract user's interests. In an insightful slideshow generated by our system, photographs are arranged in a meaningful order. Figure 3 shows an example of insightful slideshow. In this example, a photograph of "Tokyo Tower" is presented continuously after a photograph of "Eiffel Tower" because they contain the same type of buildings. Similarly, a photograph of "Asakusa" is displayed after the photograph of "Tokyo Tower" because they are popular sightseeing spots in Tokyo, and a photograph of "Big Buddha" is displayed after the photograph of "Asakusa" because they are strong related to Buddhism in Japan.

Various type of insightful slideshow of the same set of photographs can be considered. This is because various type of relation can be considered between the same pair of photographs. When new photographs are added into a target set of photographs, another type of slideshow might be organized from a different viewpoint.

4 Composition of Insightful Slideshow

This section describes a technique for deriving semantic relations between photographs, and how to compose an insightful slideshow of the photographs based on their semantic relations.

4.1 Deriving Semantic Relations between Photographs

We suppose that every photograph has been assigned one or more tags manually or automatically. The weight of a relation is calculated based on co-occurrence of tags on the web. We think that the weight of a relation between photographs A and B is high, when the co-occurrence of the tags assigned to A and B is high. Suppose that $T = (t_1, t_2..., t_n)$ is the set of all tags to be assigned to a set of target photographs, and a tag which is assigned to a photograph p is represented as $tag(p)$. The weight of a relation between two photographs p_1, p_2 is calculated using the Simpsons coefficient as formula (1)

$$S(t_1, t_2) = \frac{page(t_1) \cap page(t_2)}{min(|page(t_1)|, |page(t_2)|)} \tag{1}$$

Here, t_i is a tag to be assigned to a photograph, and $page(t_i)$ is the set of all Web pages which contain t_i in their texts. We think that the relation of tags t_i and t_j is strong when the value of $S(t_i, t_j)$ is large. And the relation of the photographs with the tags t_i and t_j is also high when the relation of tags t_i and t_j is strong.

4.2 Construction of Slideshow

Here, we consider a method for composing a slideshow that includes photographs in a target set. Each photograph appears only one time in a slideshow. We assume that continuously presenting two photographs whose weight of relation is high could give a user comfortable impression. We think that a route in a network contains all nodes and edges whose weight of relation is high represents a comfortable presentation order of photographs, where the weight of each edge represents the weight of a relation. We can obtain the most appropriate presenting order of photographs in a slideshow, when we discover the route where the summation of weights of all edges in the route is maxim. However, the problem of finding such route in a network from all the combinations is NP-hard. For answering this problem, we use the nearest neighbor method[8] as an approximate solution method. The following shows our algorithm for deciding a route in a network for an insightful slideshow.

1. We create a network in which each node represents a photograph and each edge represents a relation between two photographs. The edge between node p_1 and p_2 is represented as (p_1, p_2).
2. Calculate of the weight of every relation in the network. The weight of a edge (p_1, p_2) is calculated by means of the Simpsons coefficient $S(tag(p_1), tag(p_2))$.

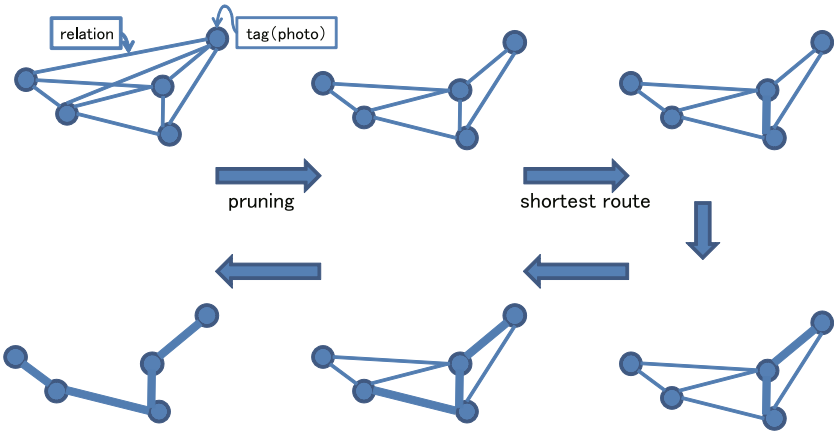

Fig. 4. Overview of Shortest Route Detection

3. For each edge (p_1, p_2), if its weight $(S(tag(p_1), tag(p_2))$ is less than a threshold, it is pruned.
4. All nodes in the network are inserted into a unvisited node list V.
5. For each edge (p_1, p_2), its distance $D(p_1, p_2)$ is calculated by formula (2).

$$D(p_1, p_2) = \frac{1}{S(tag(p_1), tag(p_2))} \tag{2}$$

6. The edge whose distance is minimum is found then the nodes of the edge is set as terminal nodes. And, the nodes are deleted from V.
7. In all the edges from the terminal nodes to a node in V, the edge with the minimum distance is found. And, the nodes of the edge are deleted from V.
8. If V is empty, this process is stopped, otherwise go to step 7.

Figure 4 shows an overview of the process. A circle represents a photograph, a thin line represents a relation, and a thick line represents a path in a selected route.

4.3 Example

This section shows an example of slideshow composition based on our proposed method. Figure 5 shows four photographs to be composed into a slideshow. One tag has been assigned to each photograph. The tags are “Eiffel Tower”, “Tokyo Tower”, “Asakusa” and “Big Buddha”. Table 1 shows the Simpsons coefficient values and the distances for every combination of photographs. The number assigned to an arrow represents the distance between photographs. Figure 6 shows a slideshow of four photographs that is composed by our proposed method. In this figure, an arrow represents an order of displaying two photographs. The number over an arrow represents the distance between two photographs. Intuitive semantic relations can be assumed between each pair of neighboring photographs

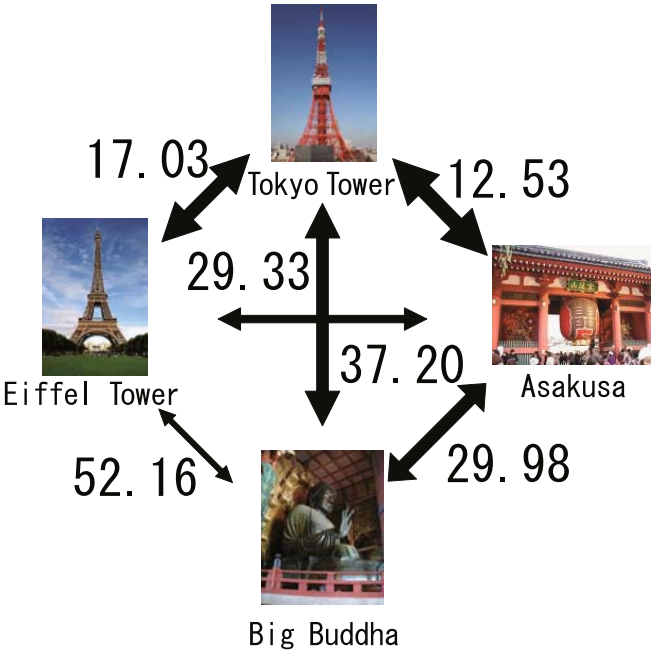

Fig. 5. Example of Photographs and Distances between Them

Table 1. Example of the Simpsons Coefficient Values and Distances between Photographs

rank	tag1	tag2	simpsons	distance
1	Tokyo Tower	Asakusa	0.0798	12.53
2	Eiffel Tower	Tokyo Tower	0.0587	17.03
3	Tokyo Tower	Big Buddha	0.0341	29.33
4	Asakusa	Big Buddha	0.0334	29.98
5	Eiffel Tower	Asakusa	0.0269	37.20
6	Eiffel Tower	Big Buddha	0.0192	52.16

Fig. 6. A Slideshow as a Result of Our Method

in the slideshow. The term under an arrow shows an example of such semantic relation. Therefore it could be expected that the slideshow would give a user a kind of comfortable impression.

5 Conclusion

In this paper, we proposed a method for composing an insightful slideshow automatically. This method provides a slideshow with natural flow by reveal an insight of user's intention.

We have a plan to develop a technique of automatic assignment of tags to a content object by using map information and Web information. Additionally, it is scheduled to evaluate our proposed method by using a large-scale set of photographs.

In this paper, we focus on personal photographs as target content, however the proposed method can be applicable to the other types of digital content. For instance, when a user is watching a news clip, user's personal photographs that have been taken in the place concerned to the news clip. A user could enjoy a new type of information content in which public content and personal content are mixed.

References

1. Cabinet Office.: Survey of consumer behavior. adoption rate (2008), `http://www.esri.cao.go.jp/jp/stat/shouhi/shouhi.htm`
2. Hiroki, U., Kosuke, N., Tetsurou, T., Ohmukai, I., Hideaki, T.: Ba-log. The Japanese Society for Artificial Intelligence (2004)
3. Kiyoko, I., Kazumasa, Y., Naokazu, Y.: Indexing Photos Based on Shooting Position and Orientation. MIRU (2005)
4. Hideyuki, F., Masatoshi, A., Koji, O.: Retrieval System for Digital Photographs using Observing point of Photographs. DEWS (2003)
5. Ken, M., Megumi, N., Hiroyuki, T., Kambayashi, Y.: Design and Implementation of the SpaceTag Prototype System: An Object System with Spatio-temporarily Limited Access(Special Issue on Knowledge and Information Sharing). Transactions of Information Processing Society of Japan 41(10), 2689–2697 (2000)
6. YingZi, J., Yutaka, M., Mitsuru, I.: Extracting Artist Network from World Wide Web. The Japanese Society for Artificial Intelligence (2006)
7. Keisuke, I., Yutaka, M., Yuki, Y.: Mr. Network Extraction: A Network Extraction System from the Web. The Japanese Society for Artificial Intelligence (2007)
8. Murano, T., Matsumoto, N.: Comparison of Approximate Methods for Traveling Salesman Problem. IEICE technical report. Nonlinear problems 102(625), pp.1–6 (2003)