Collaborative Guidance System Using Multiple Gaze History and Shared Photograph Collection

Rieko Kadobayashi and Azman Osman Lim
National Institute of Information and Communications Technology
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0289, JAPAN
Email: {rieko,aolim}@nict.go.jp

Abstract

In this paper, we propose a collaborative guidance system which provides information extracted shared photo collections based on a user's context. The user's gaze history is used to determine the user's situation in the real world. In our proposed system, we use a photograph

viewpoint logging (PVL) system, which was previously developed to record the real world as photographs with viewpoint information (based on the photographer's position and direction of gaze). In the prototype of the PVL system, user terminals are equipped with mobile phones with attached motion sensors, a GPS sensor, and a notebook PC for logging the position and orientation of the mobile phone. Based on the PVL system, we design the collaborative guidance system by using multi-gaze history and a shared photograph collection. After a user takes two or three photographs consecutively, the collaborative guidance system provides useful and well-processed content based on the user situation.

1. Introduction

Nowadays, sharing photographs over the Internet has become a popular activity. There are many online photo sharing services such as Flickr [1], Picasa [2], Kodak Gallery [3]. Flickr is widely used as a photograph repository, whereby photographs are tagged and browsed using folksonomic (social indexing) means. In the last few years, the rapid increase in photograph sharing has created a new trend within Internet communities.

Photologs and moblogs are also becoming popular. A photolog is a kind of blogs which mainly contains photographs rather than texts while a moblog is a mobile-enabled blog which allows users to update their blogs from a mobile phone or other mobile device. Today people can send their photos to online photo sharing sites or blogs from their camera phone using services such as Pictavision [4] and LocoBlog [5].

These photologs, moblogs, and shared online photographs could be "Consumer-Generated Media" (CGM) [6]. CGM, as the name represents, is content created by consumers available online. CGM has attracted a lot of attention recently as an effective marketing method. Word of mouth is an example of CGM and there exist a lot of word of mouth marketing sites. For example, mapcomi [7] is a site that allows users to post and share their word-of-mouth information about sightseeing, restaurant, hotels, events, and so on.

Accordingly, we can assume that people are willing to post and share their photographs and comments. In other words, it is reasonable to design a guidance system using shared photo collections collaborately generated by multiple users.

Advances in sensor technologies have realized mobile phones equipped with a variety of sensors such as a Global Positioning System (GPS) sensor, a motion sensor, and an electronic compass. The GPS sensor makes it possible to track the location of a mobile phone user. It also allows users to add geographical metadata to pictures and they can post and share "geotagged" pictures online [1, 5]. The geotagged pictures can be mapped on to a map so that users can easily know where these pictures were taken or users can search pictures using the geographical information.

An electronic compass or a motion sensor enables an application running on a mobile phone to detect the orientation of the mobile phone. This means that the application can detect which direction the mobile phone user is looking in. Using these sensors, therefore, users can easily add their gaze direction metadata to pictures.

This observation led us to develop a 3D photo-logging system which allows mobile phone users to annotate a part of the real world, e.g., scenery, buildings, monuments and publish the annotation as Blogs [8]. From the CGM point of view, the 3D photo-logging system allows users to create shared collections of photographs with annotations about objects and events in the real world. With a GPS sensor and a motion sensor, the 3D photo-logging system can obtain a user's position and gaze direction which are collectively referred to "viewpoint-information" and hence it can exploits viewpoint-based image retrieval [9, 10] to easily access to pictures of interest.

Expanding the idea of 3D photo-logging system, we have developed a gaze-based guidance system for mobile phone users [11]. A user of the gaze-based guidance system can obtain pictures and its explanations based on the user's position and gaze direction which are automatically sensed by the sensors installed in or attached to a mobile phone. The user does not need to type in keywords that express such as location and points of interest while the user can acquire information about what he or she is looking at.

In this paper, we propose a collaborative guidance system which provides information extracted shared photo collections based on a user's context. The user's gaze history is used to determine the user's situation in the real world. We indistinguishably use the term photograph viewpoint logging (PVL) system in referring to the 3D photo-logging system in this research. To elaborate on our proposed idea, we first define the concept of viewpoint information in Section 2. We describe the principle of a PVL system in Section 3. Then we introduce our idea of a collaborative guidance system in Section 4. Section 5 reviews relevant previous research work. Lastly, Section 6 concludes this paper.

2. The concept of viewpoint information

Viewpoint information is basically a 3D vector from the observer's eye to the point the observer is looking at. Figure 1 illustrates a model of viewpoint. When it is not easy to obtain the exact 3D coordinates of the point, we can use a direction vector, which starts at the observer's eye and continues in the direction the observer is facing, as the viewpoint information. The direction vector can be calculated from extrinsic parameters, i.e., the position and orientation of the camera that the observer is using.

The most important feature of viewpoint information is that it enables users to search pictures intuitively without the need to specify keywords [10]. Keyword-based searches require users to formulate appropriate queries; however, combining several key words is often insufficient. Users may experience difficulty combining several keywords to formulate a query that can distinguish between images that are very similar. Moreover, keyword-based searches require users to have some knowledge about the image content, so that they can make an appropriate query. This is too restrictive for practical use.

Another retrieval method called "content-based image retrieval" [12, 13], uses features such as color, shape, and texture that are automatically extracted from images. However, photographs taken in the real world, especially those taken outdoors, are usually affected by natural conditions such as sunlight and weather, and artificial conditions such as traffic and crowds. This causes diversity in the image features of the same object under various conditions over time and hence reduces the efficiency of image retrieval. Viewpoint-based searches, on the other hand, are robust against diversity in image features.

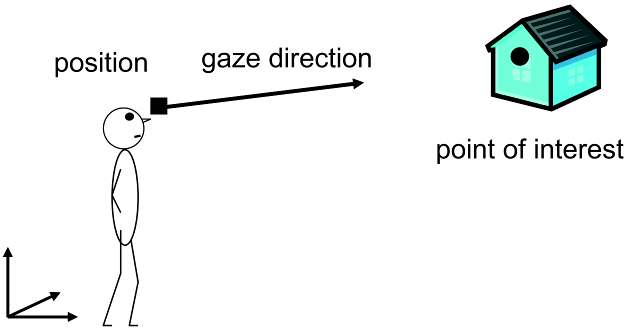

Figure 1. A model of viewpoint.

Moreover, viewpoint information is useful for organizing and analyzing data associated with spatial data. It has many applications, such as in archaeological information systems [14]. In our previous research, a prototype of the gaze-based guidance system was developed based on the PVL system, which is designed to take full advantage of viewpoint information.

3. Photograph viewpoint logging

3.1. Overview

We use the term "photograph logging" to express the method of recording objects and events in the real world as photographs. Photograph viewpoint logging (PVL) refers to photograph logging using viewpoint information. The viewpoint information is comprised of the user's position and direction of view when he or she takes a photograph. The user's position is obtained from the GPS, while the direction of view is obtained from the motion sensor.

The goal of developing the PVL system is to create a framework for recording, organizing, analyzing, searching, and presenting every object and event in the real world with "portable devices," such as mobile phones, in an intuitive and casual manner. Objects include, for example, buildings, monuments, landscapes, and archaeological sites, and events include, for example, festivals. Anything that exists or happens at a particular location can be a target of the PVL system.

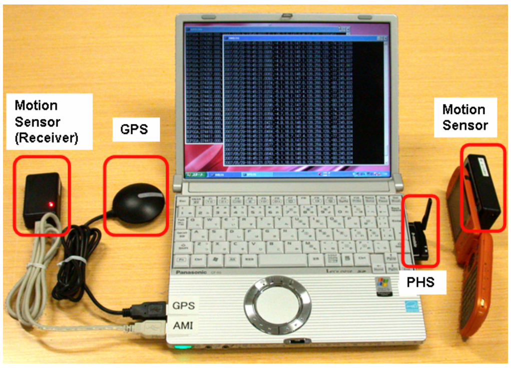

Figure 2. User terminal of the PVL system.

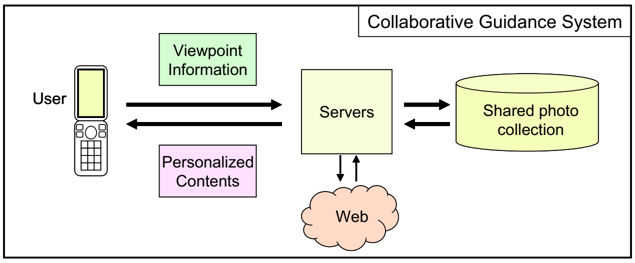

Figure 3. Collaborative guidance system configuration.

3.2. System configuration and description

In this section, we discuss the basic configuration and description of the PVL system. Our prototype system consists of a server machine and user terminals. A database server, mail server, and integration server are run on the server machine. The PVL database consists of a shared photograph collection, in which the photographs with viewpoint information are stored. A map server and blog server are used to retrieve information from the PVL database.

The user terminals are equipped with mobile phones to which motion sensors are attached, a GPS sensor, and a notebook PC for logging the position and orientation of the mobile phone, as shown in Figure 2. The motion sensor obtains the orientation of the mobile phone when the user takes a photograph. The motion sensor data is transmitted by wireless connection to a receiver, which is connected to the laptop PC via a universal serial bus (USB) port. The GPS sensor obtains the location of the mobile phone.

To obtain the posture of the user from his or her mobile phone, we used AMI601-CG [15] developed by Aichi Micro Intelligent Corporation. The AMI601-CG is an evaluation kit of AMI 601, a six-axis G2 motion sensor, which has a three-axis magnetic sensor and a three-axis accelerometer [16]. We can calculate the direction of the user's gaze from data about the user's posture obtained with the mobile phone.

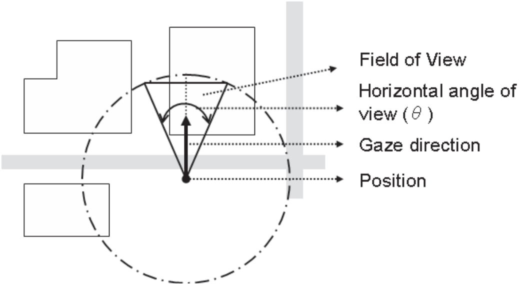

Figure 4. Field of view for POI estimation.

When a user takes a photograph with a mobile phone, he or she sends the photograph to the mail server via e-mail accompanied with a text explanation. Usually the user writes the name or identifier in the “subject” field of the e-mail and writes some comments in the body of the e-mail so that these text explanations can be transformed into the title and description of a blog entry. This increases the readability of the blog.

The motion sensor attached to the mobile phone detects the orientation of the mobile phone every second and sends the information with a timestamp to the laptop PC by a wireless link. The location of the phone is detected by the GPS sensor connected to the PC through a USB port. Note that the position of the mobile phone can be obtained by the GPS sensor installed in the mobile phone itself; however, the reading will not be as accurate. The data on the position and orientation of the mobile phone is automatically uploaded to the 3D viewpoint server. The integration server determines the correct position and orientation data for the picture by comparing the timestamp in the e-mail containing the picture and text with that contained in the position and orientation data. It then stores this data in the database.

Every time the mail server receives e-mail with a picture attached, the picture is processed and stored in the database along with metadata. The metadata include the viewpoint information as well as a picture ID, e-mail date, title (subject of the e-mail), and description (content in the body of the e-mail). The blog server then updates the blog site so that the most recent picture is shown as the latest entry.

4. Proposed collaborative guidance system

In this section, we introduce a collaborative guidance system using multi-gaze history and a shared photograph collection compiled by the photograph viewpoint logging system (PVL system). The collaborative guidance system

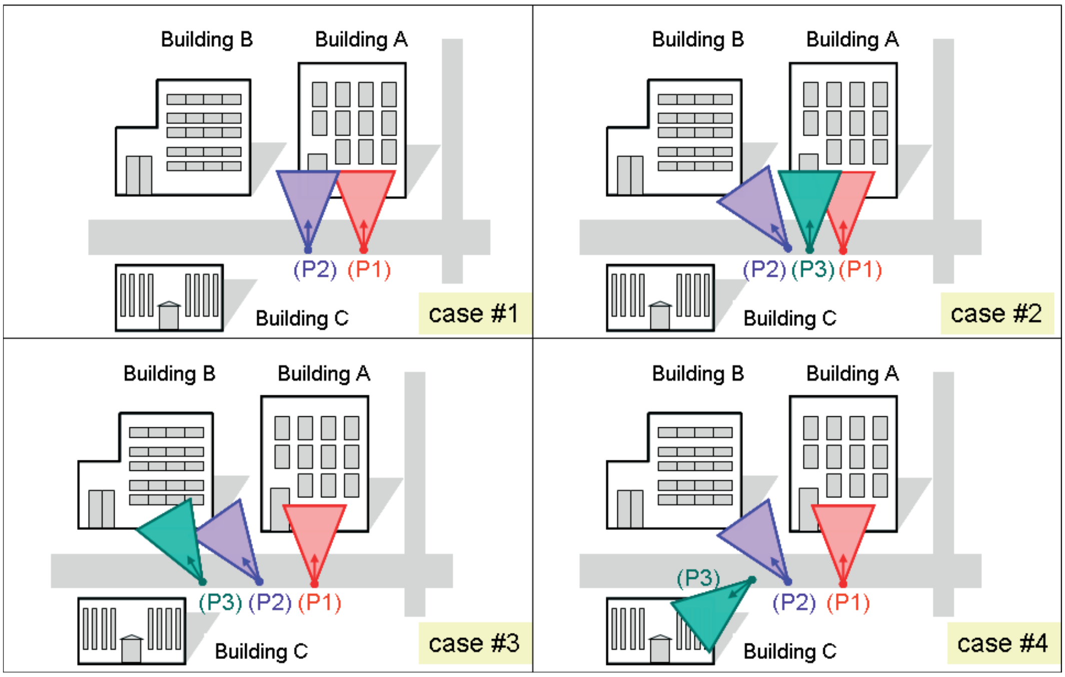

Figure 5. POI estimation based on the multiple gaze history of a user.

is used to support users' activities based on the user's situation (e.g., location, conditions, etc.) in the real world. The system configuration is depicted in Figure 3.

We made several assumptions when designing the collaborative guidance system. First, we assume that users are willing to publish and share their photographs online. These shared photographs are stored in a database using the PVL system. Second, users annotate photographs to show, for example, subjects of the photos and their opinions or feelings about the subjects. Third, photographs of objects such as monuments, buildings and landscape are available beforehand with guidance. More the users post their photographs to the system, the more they are given detailed content about what they want to know.

The collaborative guidance system is comprised of the following procedures:

(a) A person uses a mobile phone equipped with the PVL system;
(b) A user takes pictures using the PVL system while walking along the street;
(c) A shared PVL database is used to collect the photographs taken by users;
(d) A server is used to estimate the target object or the point of interest (POI);
(e) A multiple gaze history is used to estimate the user situation;
(f) Based on the POI estimation and user situation, the user guidance content is provided; and,
(g) The content is displayed on the mobile phone.

The prototype of the collaborative guidance system was configured based on the PVL system. However, there are two additional features in the configuration of the collaborative guidance system. First, a PVL server that is a combination of web, map, blog, and database servers is used to estimate the POI of the user and the user situation. Second, the PVL server sends the content to the user's mobile phone.

4.1. POI estimation

In this section, we describe how to estimate the POI of the photograph using the following two steps:

Step 1 Object identification
Step 2 POI determination

In object identification, first, a field of view (FOV) is determined by using the viewpoint information of the photograph. The position of viewpoint information gives the position of the user. Meanwhile, the direction of viewpoint information is the gaze direction, which is the user's view direction. The position of viewpoint information forms the center point of a circle, as shown in Figure 4. A triangle is obtained by using the horizontal angle of view (θ), in which the horizontal angle of view can be obtained by simply taking the width of the picture or the scope of the camera lens.

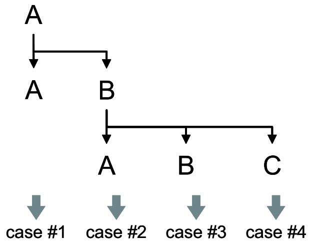

Figure 6. Flow chart for determining the four possible cases.

By interposing the gaze direction in the middle of the horizontal angle of view as the viewpoint position, and using this as a reference point, the triangular FOV can be determined. Second, the intersection of the FOV and a digital vector map is determined in order to extract the polygons, which can be an object such as a building, monument, etc. After the mapping process, the polygons are identified as objects.

In the object identification process, it is possible that more than one object will be identified from the photograph. Thus, it is necessary to determine which object is the target object, or the so-called point of interest (POI). If the photograph contains only one object, then the POI is that object. However, if the photograph contains more than one object, then the POI is determined based on the size of the objects. If the size of one object is greater than the size of the other object, then the POI is the larger object and vice versa. If the size of one object is equal to the size of the other object, the object nearest to the user position is selected as the POI.

4.2. Multi-gaze history

In the PVL system, the collection of stored photographs is updated continually in the PVL database. The PVL database is assumed to be shareable among users. One of the functions of the PVL server is to determine the user's situation based on multi-gaze history. First, the PVL server analyzes consecutively the photographs received by the integration server. Based on the sequence of POIs of the photographs, the PVL server decides the user's situation. Lastly, the PVL server sends the most appropriate content to the user.

In order to describe how the PVL server utilizes the multi-gaze history, we use the scenario shown in Figure 5. In the figure, the sequence of POIs corresponds to the user position and direction the user is facing when taking photographs. Suppose a user initially stands at location P1 and takes the first photograph in the viewpoint towards building A. Then, the user moves further away, and stands at location P2, and takes the second photograph. Finally, the user moves again and stands at location P3, and takes the third photograph. According to the three consecutive photographs, we analyze the multi-gaze history and categorize them into four possible cases as follows:

Case #1 When the second photograph is taken and the result of gaze history is building $A \rightarrow$ building A, the POI is building A.

Case #2 When the third photograph is taken and the result of gaze history is building $A \rightarrow$ building $B \rightarrow$ building A, the user cannot decide which building to go into.

Case #3 When the third photograph is taken and the result of gaze history is building $A \rightarrow$ building $B \rightarrow$ building B, the POI is building B.

Case #4 When the third photograph is taken and the result of gaze history is building $A \rightarrow$ building $B \rightarrow$ building C, the user is taking a stroll or has not settled on his or her destination.

The above cases can be simplified into a flow chart as depicted in Figure 6. Based on the cases mentioned above, the PVL server determines the user's situation and proposes the most suitable content to the user.

4.3. Content for collaborative guidance system

After a user takes two or three photographs consecutively, the PVL server determines the user situation and sends the most appropriate content to the user's mobile phone. Since the mobile phone is a portable device constraint, the content is limited to a certain size. However, before the most suitable content can be sent to the user, the PVL server needs to decide which content will be sent. This is based on the user situation by looking at all the possible cases, as described in the previous section. In these cases, we propose an example of content of the collaborative guidance system by assuming a user is taking photographs using a mobile phone while walking along the street. The proposed content of the collaborative guidance system is dynamically changed according to the user situation. For example, every time the user takes another photograph the displayed content changes interactively according to the criteria of the cases. Figure 7 shows an example of the content of the collaborative guidance system.

Next, we elaborate on the content that corresponds to the derived cases. The left image of Figure 7 represents Case 1 or Case 3, where the POI is building A or B, respectively. Therefore, a generic description is provided, and icons are displayed in soft key pop-ups at the bottom of the mobile phone screen. The generic description contains an image and text about the buildings. For example, building A is a restaurant. Information about the type of restaurant, menu,

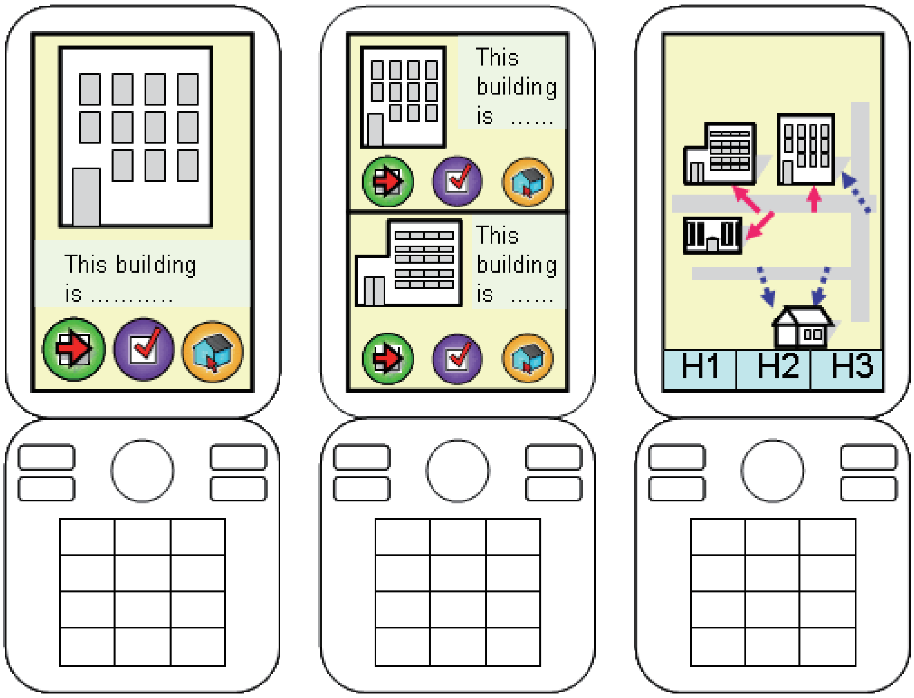

Figure 7. An example of content of the collaborative guidance system: Left: The POI is building A and building B for Cases #1 and #3, respectively. A generic description is provided, and icons are also displayed at the bottom of the mobile phone screen. Center: In Case #2, two alternative buildings are shown so that the user can easily decide which one to choose. Right: In Case #4, a map of the surrounding area is shown in the main window. Positions and directions of the three most recent positions where the user took photographs are indicated by the solid arrows. Other users' photographs with their positions and directions are also shown, whereby the dotted arrows represent the viewpoint information of other users' photographs.

etc., is displayed on the mobile phone. The center image of Figure 7 represents Case 2, in which the user cannot decide which building to enter; therefore, two alternative buildings are shown so that the user can easily decide which one to choose. On the right side of Figure 7, the user is walking around without a specific destination (Case 4); therefore, a map of the surroundings is displayed in the main window of the mobile phone. The positions and directions of the three most recent places where the user shot photographs are marked by arrows on the map. In addition, the three most recent photographs that were taken, along with their position and direction, are also marked on the map so that the user can simply decide which one to choose.

Beside the main content above, the collaborative guidance system also provides an additional content tailored to the user based on his or her profile stored in the mobile phone. Icons are used to link to the additional content. In order to elaborate more about the additional content, we present two examples for Case #1 and Case #4 in the next following paragraph.

Figure 8 illustrates an example of icons to be displayed in a soft key pop-up content for Case #1. There are three icons. Assume that building A is a restaurant. The first icon contains the most relevant information of interest about the target object. For example, the information is about the best table for corresponding user to sit at in order to be able to enjoy a garden view or listen to a jazz band in the restaurant. In the right side of the figure, the first icon displays a pop-up overlap window, which contains content about the garden view of the restaurant. The second icon contains a summary of useful information about other events happening at the same restaurant, which user may be interested in after having dinner there. This information is processed and obtained from the PVL database based on the user profile. The last icon could be used for further details about the restaurant, such as user comments, related topics, etc. For Case #2, the content of the icons is similar to Case #1 or Case #3.

On the other hand, the icon-based content for Case #4 is different from the others. This is because the main content

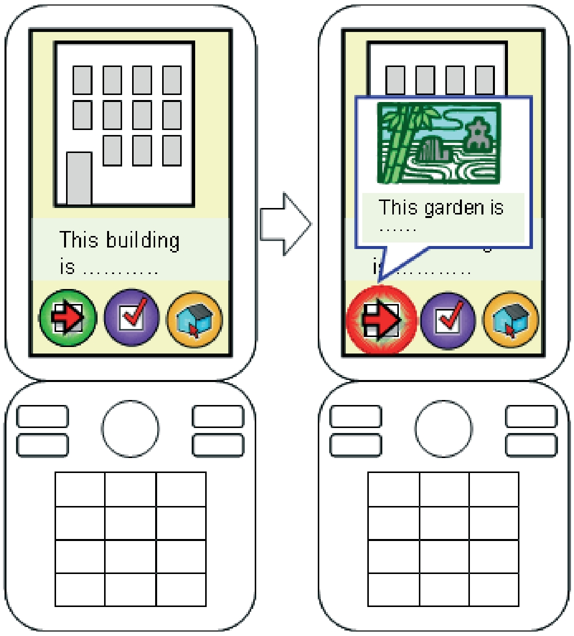

Figure 8. Example of additional content for Case #1.

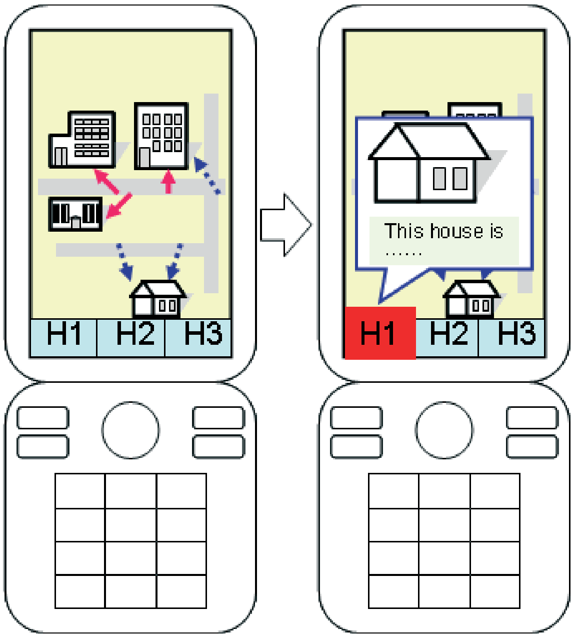

Figure 9. Example of additional content for Case #4.

of Case #4 is part of a map of the surroundings, which is different from a building, as in the other cases. Here, we show the difference by naming the icons with letters and numbers, e.g., H1, H2, and H3. Figure 9 shows an example of text-based icons to be displayed in a soft key pop-up content for Case #4. Since the user does not know where to go or visit, the first, second, and third icons contain information of interest about other recent user events in order to help the user decide which one to select. In the right side of Figure 9, the first text-based icon displays a pop-up overlap window, which contains content of interest about the house. These photographs and text are retrieved from photographs taken by other users in the PVL database.

5 Related Work

Among many photo sharing services, Flickr [1] is one of the most popular service. It uses tagging mechanism which allows people to add labels to photographs taken by others as well as themselves. This mechanism enables users to easily organize and find photographs. They can add geographical information as a tag to photographs in order to view these geotagged photos over maps.

Flickr and other online photo sharing services aim to provide users with sharing experiences and communication among users. Vronay and Davis [17] argued that sharing photos through online is not as compelling as sharing photos face-to-face. This is because the text annotation for the sharing photos online does not convey the emotion and feelings as sharing photos face-to-face. However, our motivation is not to give a function for conveying the emotion and feelings related to photographs but to give a proper guidance in a real world using collaboratively collected photo collections.

Attempts at photo blogging from mobile phones were also pursued by Pictavision [4]. Photo blogging system is also called Moblogging. Moblogging is mobile-enabled blog that allows users to post photographs from anywhere. Examples of the Moblogging are LifeBlog and LocoBlog. The LifeBlog is a digital photo album tool designed with mobile phone photographers and bloggers in order to allow users to view, search, edit, and share the photographs and messages. LocoBlog is a mobile phone application which supports location-based mobile photo blogging. Users of LocoBlog can post pictures with geographical information and view photographs on the map.

“MobShare” is a mobile phone based picture sharing system that enables immediate, controlled, and organized sharing of mobile pictures, and browsing, combining, and discussion of the shared photographs [18]. The system is based on a client-server architecture. It focuses on new ways of promoting discussion in sharing photographs and enabling the combination and comparison of personal and shared pictures.

In [19] the current digital co-present solutions that includes tabletop interfances is presented. It is great for sharing digital photos, but lack the portability. Leonard and Marsden [20] propose a mobile application, which allows users to share photos with other co-present users by synchronizing the display on multiple mobile devices. This work different from our work because the contents that are displayed on the mobile phone is asynchronized manner. One obvious weakness of this work is that the group size is limited to four participants at one time.

6. Conclusion and future work

In this paper, we proposed a collaborative guidance system for real-world applications by extending the PVL system, which records photographs together with their corresponding viewpoint information (position and gaze direction) and explanations. The collaborative guidance system provides useful information retrieved from shared photo collections based on the user's context that is estimated from the user's gaze history. We believe that the collaborative guidance system will make it easier to access real-time information in the real world.

We intended to improve the collaborative guidance system with a new method of achieving operational usability. In order to develop a collaborative guidance system that is aware of the user's situation (e.g., location-based information, mood, preferences), in the future we will also focus on investigating the PVL server to distill the content at the proper level of abstraction depending on the user situation. Further research is required to examine the challenges of the collaborative guidance system in the field of personal privacy.

References

[1] Flickr. Available: http://www.flickr.com/.

[2] Picasa. Available: http://picasa.google.com/.

[3] Kodak Gallery.

Available: http://www.kodakgallery.com/.

[4] Pictavision. Available: http://www.pictavision.com/.

[5] LocoBlog. Available: http://www.locoblog.com/index.php

[6] Pete Blackshaw. The Pocket Guide to Consumer-Generated Media. Available: http://www.clickz.com/showPage.html?page=3515576.

[7] Mapcomi. Available: http://mapcomi.jp.

[8] R. Kadobayashi, K. Kayama, T. Umezawa, and I. E. Yairi, "Sensing human activities and environment for creating knowledge cycle", In *Proc. of the 1st Int. Symposium on Universal Commun.*, National Institute of Information and Communications Technology (NICT), 2007.

[9] R. Kadobayashi and R. Furukawa, "Combined use of 2D images and 3D models for retrieving and browsing digital archive contents", In *Proc. of SPIE-IS&T Electronic Imaging 2005, Videometrics VIII*, vol. 5665, pp. 134–143, 2005.

[10] R. Kadobayashi and K. Tanaka, "3D viewpoint-based photo search and information browsing", In *Proc. of the 28th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 621–622. 2005.

[11] R. Kadobayashi, "A gaze-based guidance system based on a real world 3D photo logging system", In *Proc. of MIRW / MGuides 2007 Workshop*, pp. 37–40, 2007.

[12] M. Flickner, H. S. Sawhney, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: The QBIC system", *IEEE Computer*, vol. 28, pp. 23–32, 1995.

[13] J. R. Smith and S. F. Chang, "VisualSEEk: A fully automated content-based image query system", In *Proc. of ACM International Conference on Multimedia*, pp. 87–93, 1996.

[14] P. Drap, A. Durand, M. Nedir, J. Seinturier, O. Papini, R. Gabrielli, D. Peloso, R. Kadobayashi, G. Gaillard, P. Chapman, W. Viant, G. Vannini, and M. Nucciotti. "Photogrammetry and archaeological knowledge: Toward a 3D information system dedicated to medieval archaeology — A case study of shawbak castle in Jordan", In *Proc. of the ISPRS Int. Workshop 3D-ARCH 2007*, 2007.

[15] Aichi Micro Intelligent Corporation, *Computer Graphics Controller (AMI601-CG) Instruction Manual AMI-MI-0132E*, 2007. Available: http://www.aichi-mi.com/3_products/601-cgmanual_e.pdf.

[16] Aichi Steel Corporation, *AMI601 Delivery Specifications*, ver.1.3_061011 edition, 2006. Available: http://www.aichi-mi.com/3_products/ami601ev1.3_061011.pdf.

[17] D. F. Vronay and S. J. Davis, "PhotoStory: Preserving emotion in digital photo sharing", Virtual Worlds Group, Microsoft Research.

[18] R. Sarvas, M. Viikari, J. Pesonen, and H. Nevanlinna, "MobShare: Controlled and immediate sharing of mobile images", in *Proc. of the 12th Annual ACM Int. Conf. on Multimedia*, pp. 724–731, 2004.

[19] M. Balabanovic, L. Chu, and G. J. Wolff, "Storytelling with Digital Photographs", in *Proc. of the CHI 2000*, pp. 564–571, 2000.

[20] M. A. K. Leonard and G. Marsden, "Co-present photo sharing on mobile devices", in *Proc. of the MobileHCI 2007*, pp. 277–284, 2007.