A Gaze-based Guidance System based on a Real-world 3D Photo Logging System

Rieko Kadobayashi
National Institute of Information and Communications Technology
3-5 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0289, Japan
rieko@nict.go.jp

ABSTRACT

To achieve easy access to the information that users need in the real world, we are developing a gaze-based guidance system by extending the 3D photo-logging system we have previously developed for recording the real world. The prototype system consists of several servers running on a single machine and user terminals. 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. Collecting photographs with 3D viewpoint information (based on the photographer's position and direction of gaze) by using the 3D photo-logging system, we can use these photographs and the text explanations associated with them as guide information. Users can access the information simply by pointing and clicking their mobile phones in the real world.

Categories and Subject Descriptors

H.4.3 [**Information Systems Applications**]: Communications Applications—*Information browsers*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Interaction styles, Prototyping*

General Terms

Design, Experimentation, Human Factor

Keywords

Gaze-based services, Navigation, mobile applications, photo browsing

1. INTRODUCTION

It is worthwhile to use a person's direction of gaze as well as his or her position when he or she is navigating in the real world because navigation depends on the direction in which the person wants to go and on what kind of information the user needs most. On the basis of this idea, we are developing a gaze-based guidance system for navigating in the real world by extending the 3D photo-logging system [6] we have previously developed for recording the real world with photographs.

Our 3D photo-logging system uses viewpoint-based image retrieval [5]. The key idea is to use 3D viewpoint information [7] as metadata when taking pictures in the real world to facilitate easy access to pictures of interest. The position and orientation of the camera are regarded as a user's point of view in a 3D space and are referred to as 3D viewpoint information, which is a key element for bridging 2D images and 3D spaces (e.g., 3D computer graphics models, 3D virtual worlds, or the real world).

3D Viewpoint Information is basically a 3D vector from the observer's eye to the point where the observer is looking at. When it is not easy to obtain the 3D coordinates of the point correctly, we can use a direction vector, which starts the observer's eye to the direction the observer is facing, as the 3D Viewpoint Information. The direction vector can be calculated from the extrinsic parameter, i.e., the position and orientation, of the camera which the observer is using.

The most important feature of the 3D viewpoint information is that it enables users to search pictures intuitively without the need to specify key words [7]. 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 they can make an appropriate query. This is too restrictive for practical use.

Another retrieval method called "content-based image retrieval" [4, 8], 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.

Moreover, 3D viewpoint information is useful for organizing and analyzing data associated with spatial data. It is has many applications, such as archaeological information systems [3]. The 3D photo-logging system was designed to take full advantage of 3D viewpoint information. A prototype of our gaze-based guidance system has been developed based on the 3D photo-logging system.

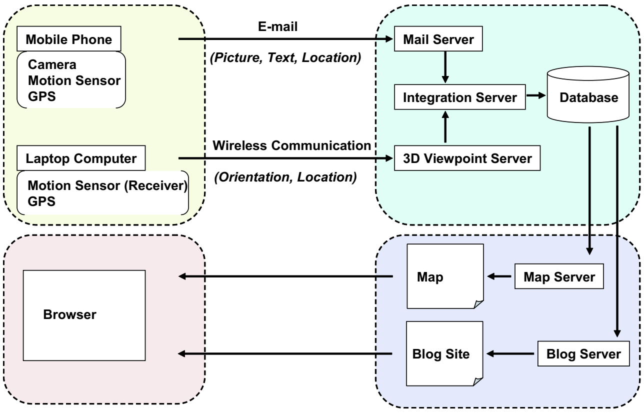

Figure 1: System configuration.

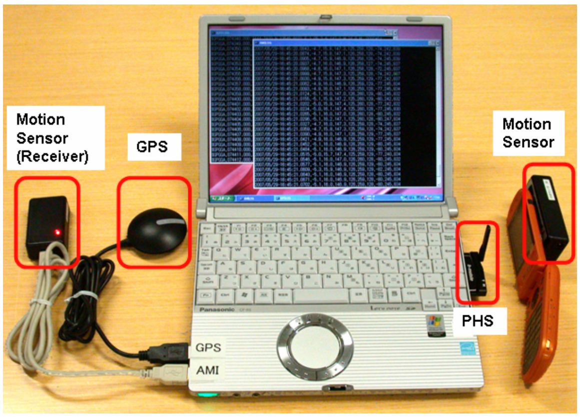

Figure 2: Hardware setup for user terminal.

2. 3D PHOTO LOGGING

2.1 Overview

We use the term "photo logging" to express the method of recording objects and events in the real world as photographs. 3D photo logging means photo logging with 3D viewpoint information. The 3D viewpoint information is composed of a person's position and direction of view when he or she takes a photograph. His or her position is obtained from a global positioning satellite (GPS) while his or her direction of view is obtained from the motion sensor.

The goal of building the 3D photo logging 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 3D photo-logging system.

2.2 System Configuration

Our prototype system consists of a server machine and user terminals. A database server, a mail server, and an integration server are run on the server machine. Figure 1 shows the configuration of the 3D Photo Logging system.

Figure 3: Experimental use of 3D photo-logging system.

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. Figure 3 shows our trial use of the 3D photo-logging system at a tourist site.

To obtain the posture of the user from his or her mobile phone, we used AMI601-CG [1] developed by the Aichi Micro Intelligent Corporation. 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 [2]. We can calculate the direction of the person's gaze from data about his or her posture obtained with the mobile phone.

2.3 Processing Flow

When a person takes a photograph with a mobile phone, he or she sends the photo to the mail server via e-mail accompanied with a text explanation. Usually he or she 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

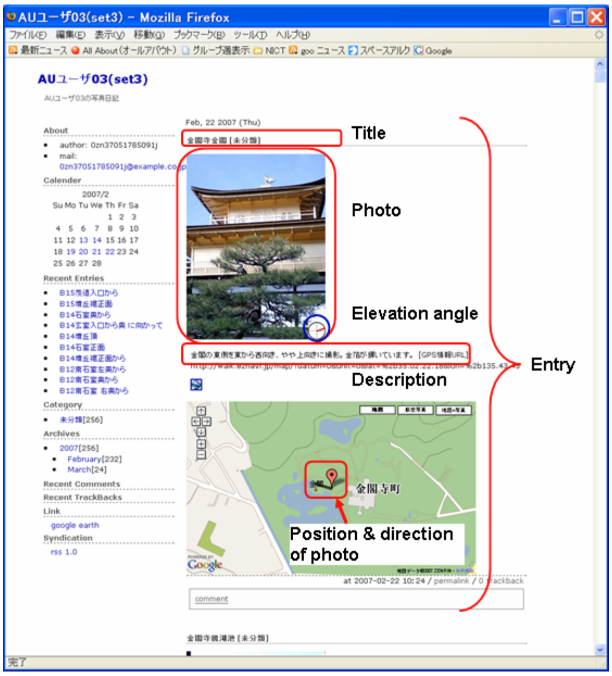

Figure 4: Sample of person's Blog site generated using photo-logging data.

Figure 5: Map-based user interface. The markers denote the positions of the person when he or she took a picture, and the lines denote the direction in which the he or she was facing.

the correct position and orientation data for the picture by comparing the timestamp in the e-mail containing the picture and text with that in data on the position and orientation. 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 is stored in the database along with metadata. The metadata includes the 3D viewpoint information as well as 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.

2.4 User Interface

Visualizing photographs in several different ways is easy because they and their associated 3D viewpoint information are stored in a general-purpose database. The system enables users to browse photographs "blog style," by treating each photograph and corresponding explanation as a blog entry, displaying them in reverse-chronological order. Figure 4 shows an example of a blog generated using the photo-logging data.

Alternatively, the collected photographs can be displayed on a 2D or 3D map. When people browse photographs on a 2D map, markers with line segments are displayed on the map to show the position of the camera, i.e., the location from which the pictures were taken and the direction in which the camera was facing. The orientation of the camera is decomposed into an azimuth and an elevation angle, and the line segment attached to the markers represents the azimuth angle. Clicking on a marker, a user can see a photograph with the corresponding elevation data represented graphically. Figure 5 shows a map-based interface.

2.5 Searching Images

The 3D photo-logging system uses multiple criteria for searching photographs, namely, the distance between two cameras, the degree of similarity in the azimuth angles, the degree of similarity in the elevation angles, the degree of similarity in the directions of gaze represented as 3D vectors, and the shooting date. After the user chooses a picture from those shown on the map, sets the values for the criteria, and clicks the search button, the system displays the results on the map graphically, i.e., with markers and line segments.Users can search photographs intuitively by interactively changing the value for the criteria. Figure 6 shows an example of an image search. One picture retrieved from the search results is displayed on the map.

3. GAZE-BASED GUIDANCE WITH MOBILE PHONES

We have implemented a proof-of-concept system by using the 3D photo-logging system. The system configuration is depicted in Figure 7. Suppose that we have a large collection of photos of tourist sites stored in the 3D photo-logging system. That is, each photo has 3D viewpoint information. Mobile phone users could then search for photographs of the tourist sites they are currently visiting by pointing and clicking with their mobile phones.

When a person is visiting a tourist site, he or she can receive a picture of his or her approximate location by accessing our experimental Web site. The system automatically locates the person's position and the direction in which he

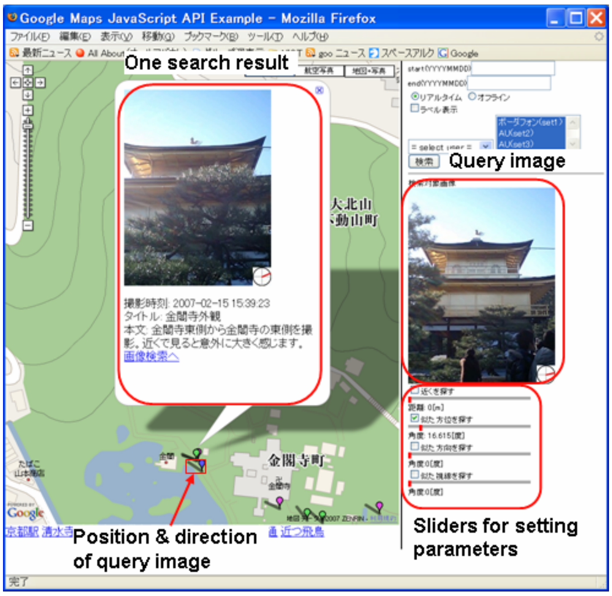

Figure 6: Image searches. A query image (on the right) and the corresponding search results (shown as markers on the map).

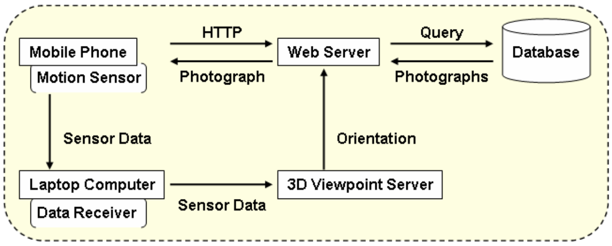

Figure 7: Configuration of searching function for mobile phones.

or she is moving, and uses this information to query the database. Note that the prototype system uses a motion sensor attached to the back of the mobile phone to detect the camera's orientation. In the current system, users receive pictures, as shown in Figure 8; however, they can either receive pictures with text or text only so that they can easily obtain information about what they are looking at.

4. CONCLUSION AND FUTURE WORK

We are developing a Gaze-based guidance system based on a real-world application of a 3D photo-logging system. The 3D photo-logging system records the picture together with its corresponding 3D viewpoint information and its explanation. Once the image database is created, we can use the data to send mobile phone users information on the basis of the direction of their gazes and their positions.

We believe that gaze-based guidance systems will increase the ease of access to real-world information in the real world. We intend to develop methods of finding the most appropriate photograph and/or information by using

Figure 8: Example of gaze-based guidance for mobile phones.

contextual data as well as the direction of a person's gaze and his or her position.

5. REFERENCES

[1] Aichi Micro Intelligent Corporation. *Computer Graphics Controller (AMI601-CG) Instruction Manual AMI-MI-0132E*, 2007. `http://www.aichi-mi.com/3_products/601-cgmanual_e.pdf` (accessed 31 May 2007).

[2] Aichi Steel Corporation. *AMI601 Delivery Specifications*, ver.1.3_061011 edition, 2006. `http://www.aichi-mi.com/3_products/ami601ev1.3_061011.pdf` (accessed 31 May 2007).

[3] P. Drap et al. Photogrammetry and Archaeological Knowledge: Toward a 3D Information System Dedicated to Medieval Archaeology: A Case Study of Shawbak Castle in Jordan. In *Proceedings of the ISPRS International Workshop 3D-ARCH 2007*, 2007.

[4] M. Flickner et al. Query by Image and Video Content: The QBIC System. In *IEEE Computer*, volume 28, pages 23–32, 1995.

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

[6] R. Kadobayashi, K. Kayama, T. Umezawa, and I. E. Yairi. Sensing Human Activities and Environment for Creating Knowledge Cycle. In *Proceedings of the First International Symposium on Universal Communication*. National Institute of Information and Communications Technology, 2007.

[7] R. Kadobayashi and K. Tanaka. 3D Viewpoint-based Photo Search and Information Browsing. In *Proceedings of the 28th Annual International ACM SIGIR Conference (SIGIR 2005)*, pages 621–622, 2005.

[8] J. R. Smith and S. F. Chang. VisualSEEk: a Fully Automated Content-Based Image Query System. In *Proceedings of ACM International Conference on Multimedia*, pages 87–93, 1996.