Construction of Extended Geographical Database Based on Photo Shooting History

Kiyoko Iwasaki
Nara Institute of Science and Technology
8916–5 Takayama, Ikoma
Nara, Japan
kiyoko-i@is.naist.jp

Masayuki Kanbara
Nara Institute of Science and Technology
8916–5 Takayama, Ikoma
Nara, Japan
kanbara@is.naist.jp

Kazumasa Yamazawa
Nara Institute of Science and Technology
8916–5 Takayama, Ikoma
Nara, Japan
yamazawa@is.naist.jp

Naokazu Yokoya
Nara Institute of Science and Technology
8916–5 Takayama, Ikoma
Nara, Japan
yokoya@is.naist.jp

ABSTRACT

This study proposes an extended geographical database based on photo shooting history to enable the suggestion of candidate captions to newly shot photos. The extended geographical database consists of not only subject positions but also the likely shooting positions and directions estimated using the histories of the shooting positions and directions of the subjects. A user can add a caption to a photo by selecting an appropriate one from the candidate captions. The candidate captions are acquired using the shooting position and direction as a key to the extended geographical database. In this paper, we present the results of experiments for constructing the extended geographical database using a prototype system.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*indexing methods*; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Design

Keywords

extended geographical database, photo captioning, indexing, shooting position and direction

1. INTRODUCTION

The use of media recording devices such as digital cameras and camera phones is becoming more and more widespread, and users are therefore required to manage increasingly larger collections of digital photos. Adding metadata is one approach to managing these photos efficiently. For example, exchangeable image file format (EXIF) [8] specifies metadata such as the date, time, camera parameters, position, and low-level image features. Cameras that can receive a photo file in JPEG format embedded with the shooting position by connecting to a GPS receiver are also offered commercially.

A photo retrieval method using metadata [9, 24], methods and user interfaces for the generation of a caption (or a tag) as for a photo metadata [10, 13, 15, 17, 21, 23], and a method for indexing photos using event segmentation [20] have been proposed. Metadata is also useful for blog services and photo-sharing websites, which have recently become popular. These services enable users to share their stored photos. For instance, Panoramio [2] and ZoneTag [3] have shown photos placed on a 2D map. They have managed to store large collections of photos efficiently using location-based captions, such as place or facility names, as metadata. Other methods for managing geo-referenced photos have also been investigated. A photo-browsing interface that makes use of a map [18, 19, 22], methods to automatically organize photos based on geographical information [5, 11, 12, 14], and methods that suggest geographical tags for photos [4, 7, 18] have been proposed. Therefore, it is useful to add location-based metadata, such as shooting position and place or facility name of the shooting position, to a photo.

Naaman et al. [11] have proposed a method to acquire the place name of a newly shot photo by referring to a photo from a database that contains several pre-prepared photos along with their shooting positions (latitude and longitude) and names (place, facility, etc.). The shooting position of the chosen database photo is similar to the position of the new photo. Fujita et al. [5] have proposed a method to acquire the observed subject name of a photo in relation to the shooting position and direction. Although such conventional methods provide a location-based name for a photo, photos, their shot positions, and location-based names must be prepared in advance. In addition, these methods may add a caption that is incorrect or not wanted by a user, since captions are determined automatically. On the other hand, a semi-automatic photo-captioning system using a geographical database and web retrieval to add an appropriate caption to a photo has been also proposed [7].

CIVR'08, July 7–9, 2008, Niagara Falls, Ontario, Canada.

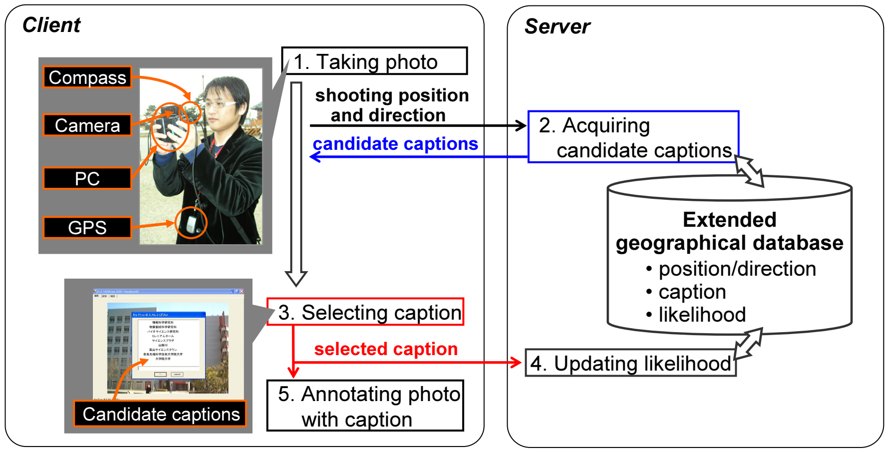


Figure 1: Flow diagram of extended geographical database construction for photo captioning.

However, for determining candidate captions for a photo, each time it is needed, these systems require reference to registered data on the shooting position of the photo and the extraction of candidate captions that are appropriate to the photo. The systems cannot acquire appropriate candidate captions if the distance to the subject is unknown or if there are occlusions with circumferential objects, since the systems determine candidate captions based on a relationship between a subject's position and the camera's position and direction.

In this study, we propose an extended geographical database containing not only subject positions but also the likely shooting positions and directions estimated using the histories of the shooting positions and directions of the subjects. Candidate captions for a photo are acquired by referring only to the data corresponding to the shooting position and direction of the photo by constructing the proposed extended geographical database. This produces interactive photo captioning while reducing the amount of reference data and processing time needed to acquire the candidate captions. By considering the history of the shooting position of a subject, cases where appropriate candidate captions cannot be acquired because of occlusion from other buildings can be prevented.

Supposing that ordinary users capture and provide captions to photos at tourist sites, which include several subjects registered in the extended geographical database, more appropriate candidate captions can be acquired by considering favorable positions for shooting a subject (shooting likelihood of a subject), inferred from shot histories. The histories of the shooting position and direction, together with a caption, are acquired by a semi-automatic photo-captioning system using the extended geographical database and a web retrieval method. The system shows candidate captions that are selected based on the shooting position and direction with respect to a user, and the user then selects an appropriate caption for the photo. In this study, we show the results of the extended geographical database construction through photos acquired by the captioning system.

In Section 2, we describe the proposed extended geographical database construction based on the shooting history for photo captioning. Section 3 describes the prototype system along with experiments on the extended geographical database construction. Finally, in Section 4, we present the summary of this study.

2. CONSTRUCTION OF EXTENDED GEOGRAPHICAL DATABASE FOR PHOTO CAPTIONING

This section describes a framework to construct the extended geographical database for photo captioning. The extended geographical database is designed for photo captioning by considering not only subject positions but also the likely shooting positions and directions estimated using the histories of the shooting positions and directions of the subjects. The extended geographical database is shared by multiple users and is updated using the users' feedback.

2.1 Overview of Extended Geographical Database Construction

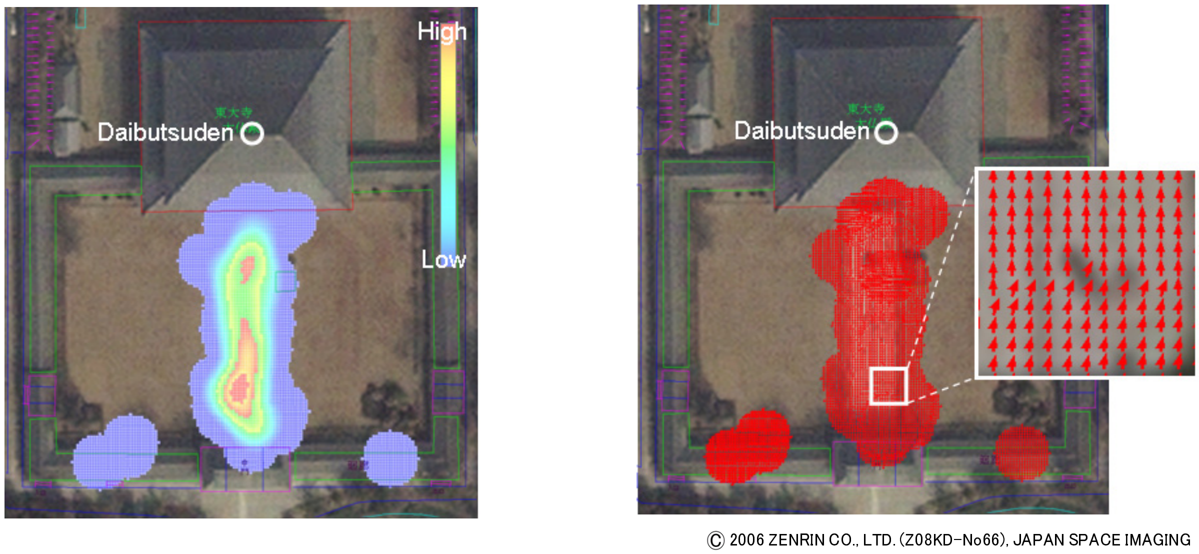
Figure 1 shows a flow diagram of the proposed framework. The proposed system acquires geo-referenced and captioned photos using a semi-automatic photo captioning [7] based on shooting position and direction to construct the extended geographical database. The client system provides functions for capturing photos that contain information on the shooting position and direction, and for adding captions to the photos. Photo captioning is performed by selecting a caption from a number of candidate captions, which are suggested by the server. The server system manages the extended geographical database. The extended geographical database is referred to obtain candidate captions for the photos and is updated using user-selected captions and information on their shooting positions and directions in order to obtain more appropriate suggestions for the candidate captions. The following are the steps involved in this process:

Step 1. A user captures a photo using a camera, and the shooting position and direction are recorded using a GPS receiver and a compass, respectively.

Step 2. The system refers to the extended geographical database using the shooting position and direction and obtains place or facility names as the candidate captions for the photo.

Step 3. The user selects the appropriate caption for the photo from the available candidate captions.

Step 4. The extended geographical database is updated using the selected caption, shooting position, and direction as feedback.



(a) Likelihood for position.

(b) Mainly-shot direction.

O subject position ← mainly-shot direction of each position

Figure 2: Likelihood of a subject at a position of interest.

Step 5. The caption is added to the photo as metadata.

In Section 2.2, we provide the details of the proposed extended geographical database. Section 2.3 explains semi-automatic photo-captioning using the extended geographical database. Finally, Section 2.4 describes the method of updating the extended geographical database using captioned photos that include the shooting position and direction.

2.2 Extended Geographical Database

The extended geographical database consists of the following items:

- geographical coordinates (latitude and longitude),
- direction,
- subject (place or facility) name,
- likelihood value of a subject shot at a specified coordinate in a particular direction.

A subject's shooting likelihood value is obtained by specifying a coordinate, direction, and subject name. The value is determined using the number of photos that have been shot at a particular position in a particular direction. Calculation of the likelihood value is detailed in Section 2.4. An example of the visualized likelihood value data of a subject is shown in Figure 2. The colors in Figure 2(a) indicate the likelihood value, and the arrows in Figure 2(b) indicate the main direction in which the shots were taken from each position. In this study, the likelihood values are set to discrete shooting directions at each position.

2.3 Semi-automatic Photo Captioning Using the Extended Geographical Database

A user acquires a photo that contains information on the shooting position and direction and adds a caption to the photo using the client system shown in Figure 1. Figure 3 shows a flow diagram for the addition of a caption to a photo. The user inputs a caption that is appropriate for a photo by the following three inputting methods:

1. [DB] Selecting a caption from candidate captions acquired by referring to the extended geographical database,
2. [web] Selecting a caption from candidate captions acquired by relevant word extraction using web retrieval,
3. [key] Entering a caption using a keyboard.

These inputting methods achieve efficient photo captioning for users. The following sections describe each inputting method.

2.3.1 [DB] *Selecting a Caption from Candidate Captions Acquired by Referring to the Extended Geographical Database*

The server acquires candidate captions by referring to the extended geographical database based on the shooting position and direction of a photo. The following are the steps in acquiring the candidate captions.

Step 1. Subject (place or facility) names and their likelihood values, which are set for a particular shooting position and direction of a photo in the extended geographical database, are acquired.

Step 2. The subject names are sorted according to their likelihood values.

Step 3. The sorted subject names are sent to a client as a list of candidate captions.

The user selects an appropriate caption for the photo from the list of candidate captions.

2.3.2 [web] *Selecting a Caption from Candidate Captions Acquired by Relevant Word Extraction Using Web Retrieval*

If the candidate captions acquired from the extended geographical database are not appropriate, new candidate captions are obtained by relevant word extraction using web retrieval. For example, a server extracts relevant names of buildings by web retrieval

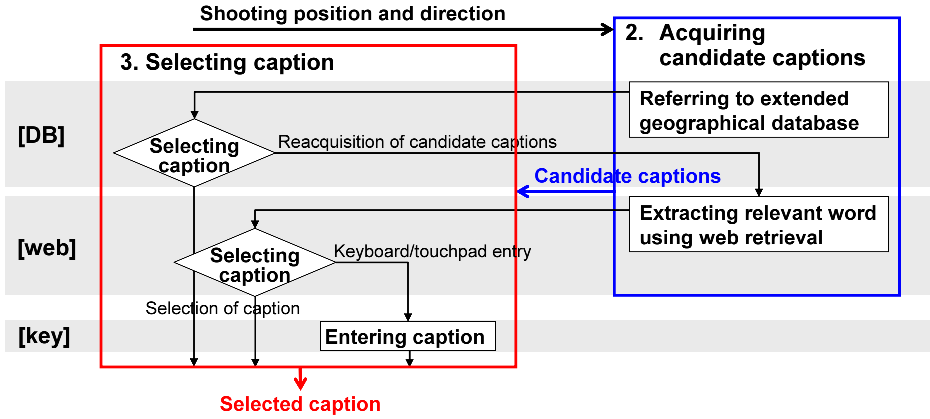


Figure 3: Flow diagram of captioning a photo.

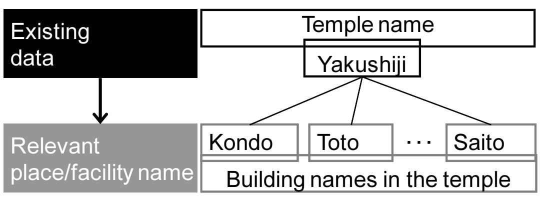


Figure 4: Relevant place or facility name.

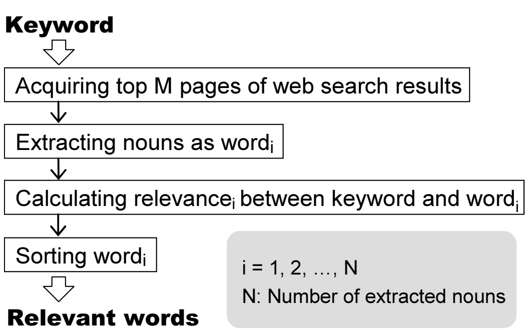


Figure 5: Relevant word extraction diagram.

using a neighboring facility's name as a keyword in the extended geographical database (see Figure 4). Sato et al. [16] proposed a relevant word extraction method using web retrieval. Since this method aims to apply the extracted words to a dictionary, the main purpose is to extract words accurately; therefore, the processing time is not considered. On the other hand, our system considers the processing time needed to achieve interactive captioning.

Figure 5 shows the flow diagram of relevant word extraction. The relevant words are acquired as follows:

Step 1. The user selects a word (or words) related to the desired caption from the candidate captions by referring to the extended geographical database. The selected word is sent to the server and is used to extract relevant words. Hereafter, the selected word is referred to as a " keyword. "

Step 2. The server searches the Web using the keyword and determines the top M web pages from the search results.

Step 3. The server extracts words without HTML tags from the web pages and obtains the part-of-speech and semantic classes of the words using a lexicon. Nouns whose semantic classes are relevant to the geographical subject names are extracted, since they are appropriate for the geographical captions of the photos.

Step 4. The indicator of relevance between the keyword and an extracted noun is obtained from Eq. (1).

$$\begin{aligned} relevance_i &= hit_{key \cap word_i} / hit_{key \cup word_i} \\ &= hit_{key \cap word_i} / (hit_{key} + hit_{word_i} - hit_{key \cap word_i}) \\ &\quad (i = 1, 2, ..., N), \end{aligned} \qquad (1)$$

where

$relevance_i$:Relevance of word$_i$,

hit_{key} :Number of hit pages that include the given keyword,

hit_{word_i} :Number of hit pages that include word$_i$,

$hit_{key \cap word_i}$:Number of hit pages that include both the keyword and word$_i$,

$hit_{key \cup word_i}$:Number of hit pages that include either the keyword or word$_i$,

N :Number of extracted nouns.

The value of $hit_{key \cap word_i}$ is estimated approximately from Eq. (2) to compute the $relevance_i$ according to the practical processing time.

$$hit_{key \cap word_i} \simeq page_{word_i|key} hit_{key} / page_{key}, \qquad (2)$$

where

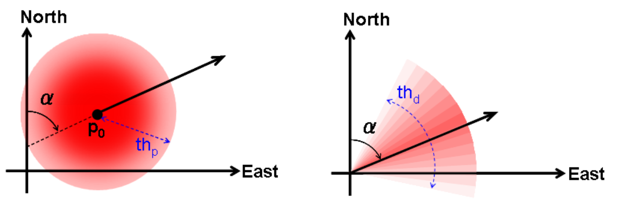
$page_{key}$: Number of web pages acquired by web retrieval using a keyword,

$page_{word_i|key}$: Number of web pages in which the word$_i$ appears.

The extracted nouns are sorted in descending order according to relevance and are presented to the user as new candidate captions.

Table 1: Hardware configuration of client system.

PC	VGN-UX90PS (SONY)	CPU: Intel Core Solo U1400 1.2 [GHz]
Camera	built-in PC	1.34 [Mpixel]
GPS receiver	BT338(GlobalSat)	accuracy of position: 10 [m]
Digital compass	InertiaCube^3(INTERSENSE)	accuracy of direction: 1 [deg]



(a) Update region for shooting position. (b) Update region for shooting direction.

$\mathbf{p_0}$: shooting position $\mathbf{th_p}$: threshold for position

$\boldsymbol{\alpha}$: shooting direction $\mathbf{th_d}$: threshold for direction

Figure 6: Likelihood update in the extended geographical database.

The user then selects an appropriate caption for the photo from the available candidate captions.

2.3.3 [*key*] *Entering a Caption Using a Keyboard*

If the candidate captions acquired by relevant word extraction are not appropriate, the user can enter a caption using a touchpad or keyboard.

2.4 Update of Extended Geographical Database Using Captioned Photos

In Step 4 of Figure 1, the extended geographical database is updated using captions that contain information on shooting position and direction. A likelihood value of a subject at a shooting position in a particular direction is determined based on photo shooting histories.

Figure 6 shows the updated regions. The extended geographical database is updated for a region whose distance from a shooting position within th_p is given by

$$\sqrt{(\vec{p_0} - \vec{p_i})^2} \leq th_p \tag{3}$$

and a region whose angle from the shooting direction within $\frac{th_d}{2}$ is given by

$$|\frac{\pi}{2} - \alpha - \theta| \leq \frac{th_d}{2}. \tag{4}$$

The likelihood set for position p_i is calculated from Eqs. (5)-(7) on the n-th update.

$$l_n(\vec{p_i}) = l_{n-1}(\vec{p_i}) + w_p w_d, \tag{5}$$

$$w_p = \frac{1}{\sqrt{2\pi}\sigma_p} e^{-\frac{(\vec{p_0}-\vec{p_i})^2}{2\sigma_p{}^2}}, \tag{6}$$

$$w_d = \frac{1}{\sqrt{2\pi}\sigma_d} e^{-\frac{(\frac{\pi}{2}-\alpha-\theta)^2}{2\sigma_d{}^2}}, \tag{7}$$

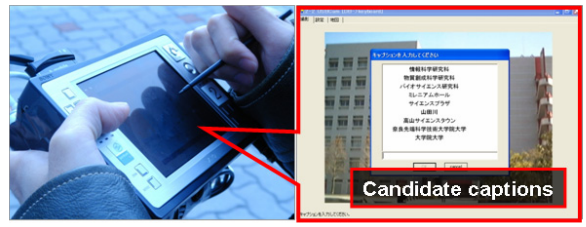


Figure 7: Selecting a caption using the client system of prototype system.

where

$\vec{p_0}$: Shooting position,

$\vec{p_i}$: Interest position,

α : Shooting direction (clockwise direction from north),

θ : Interest direction (counterclockwise angle from east),

σ_p : Standard deviation of normal distribution for distance,

σ_d : Standard deviation of normal distribution for angle.

The values of σ_p, σ_d, th_p and th_d are determined according to the accuracy of sensors when capturing and captioning photos.

3. EXPERIMENTS ON THE EXTENDED GEOGRAPHICAL DATABASE CONSTRUCTION

We have implemented a prototype system and have conducted experiments to confirm the effectiveness of the proposed extended geographical database.

3.1 Prototype System

This section explains an implemented prototype system. The prototype system consists of a client system for capturing photos that contain information on shooting position and direction and for captioning the photos, together with a server system that manages the extended geographical database.

3.1.1 *Client System*

The client system consists of a PC, a GPS receiver, a digital compass, and a camera built into the PC, as shown in Figure 1. Users capture the photos and add captions using this system. The photos contain information on the shooting position and direction as well as location-based captions as metadata. The hardware configuration of the client system is shown in Table 1. A user selects an appropriate caption for a photo using the touch screen of the client PC, as shown in Figure 7.

3.1.2 *Server System*

The server system manages the extended geographical database. When information on the shooting position and direction is received from the client system via the network, the server system

Table 2: Softwares used in the server.

Web server	Apache 1.3.27
SQL server	PostgreSQL 7.3.2
Servlet	Tomcat 5.5.3, JDK5.0
Web search engine	Google API [1]
Thesaurus	Goi-Taikei — A Japanese Lexicon [6]

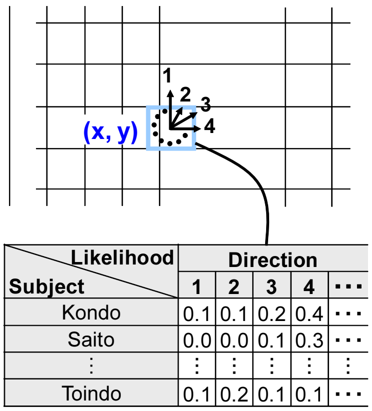


Figure 8: Data stored in the extended geographical database.

determines candidate captions for a photo by referring to the extended geographical database. Next, the server sends the candidate captions to the client, and the user selects a caption from the available candidates. In addition, the system updates the extended geographical database with the caption selected by the user, along with the information on the shooting position and direction.

Softwares used in the server are shown in Table 2. The server system consists of an HTTP web server to connect clients and the server, an SQL server to manage the databases, and a servlet to access the SQL server with information acquired from the web server. The captioning method of relevant word extraction using web retrieval, described in Section 2.3.2, is implemented using a web search engine [1] and a thesaurus [6].

Figure 8 shows the data structure of the geographical data, which is set to a particular cell. The cell, whose coordinates (x, y) are defined as the center of a region divided using a grid, is set as the likelihood of shooting a subject in a particular direction. However, all cells in a particular space do not need to be set in the extended geographical database, since there are regions comprising buildings and regions that cannot be entered by users. Tables in the extended geographical database are designed so that cell records can be added successively when the corresponding data for the cells is obtained. Details of the tables are shown in Tables 3-5.

Cell: A record is defined using coordinates specified by latitude and longitude, along with likelihood IDs.

Feature: A record is defined by the place or facility name (caption) corresponding to a subject.

Likelihood: A record is defined by the likelihood values for discrete directions corresponding to a cell ID and feature ID.

The tables are designed to register only the required cells in an area, in order to reduce the amount of data.

Table 3: Table of cells.

cell id	x	y	likelihood id
1	34.6670	135.7840	{1,3}
2	34.6675	135.7840	{2}
⋮	⋮	⋮	⋮

Table 4: Table of features.

feature id	caption
1	Kondo
2	Saito
⋮	⋮

Table 5: Table of likelihoods.

likelihood id	feature id	direction
1	1	{0.1, 0.1, 0.05, ···}
2	1	{0.0, 0.0, 0.2, ···}
3	2	{0.1, 0.2, 0.1, ···}
⋮	⋮	⋮

3.2 Extended Geographical Database Construction

3.2.1 *Overview of Extended Geographical Database Construction*

The extended geographical database was built using 2545 photos of 26 subjects in Nara Park, Japan. The Nara Park is a world heritage site with an area of 66,000 square meters. These photos contain information on shooting positions and directions as well as the captions in the form of metadata. Some of the positions of the subject facilities are shown in Figure 9. Some shot photos are also shown in Figure 10. The size of a cell needed to set the likelihood is one square meter. The likelihood updating region for position th_p is 10 m, and the region for direction th_d is 90 degrees. These parameters are decided by considering the accuracy of the GPS receiver and digital compass used in the experiments. The initial extended geographical database includes data, which consist of pairs of facility name and its geographical coordinates (latitude/longitude) in map software on the market.

3.2.2 *Results of Extended Geographical Database Construction*

Figure 11 shows the extended geographical database construction processes for Daibutsuden. With an increasing number of shot photos, the spreading cells, which are set as the likelihoods, and likelihood values are confirmed to increase. In other words, the greater the number of captioned photos added to the extended geographical database as feedback, the more appropriate the database for efficient photo captioning. When the number of captioned photos is increased sufficiently, a small cell is registered anew in the extended geographical database. This can be construed as the convergence of data for captioning. Although the number of photos for convergence depends on the size of a subject and circumferential environment, such as the observable area of a subject, in many cases, the data can be converged by constructing the extended geographical database using several hundred captioned photos.

The results of construction of the extended geographical database construction are shown in Figure 12. The colors in Figure 12 indicate the likelihood value, and the arrows in Figure 12 indicate the main direction in which the shot is taken from each position. An appropriate extended geographical database is constructed based on

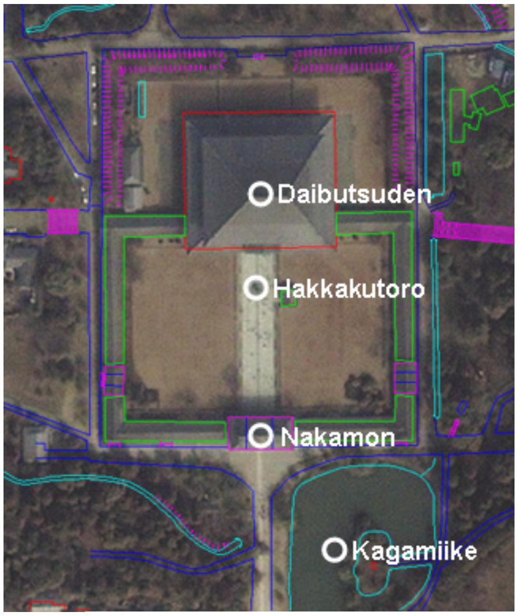


Figure 9: Position of subject facility.

Figure 10: Photos to be captioned.

the captioned photos of each facility, because most arrows indicate the tendency of the direction of the shooting position to the actual subject position. The arrows whose direction is different from the direction of the shooting position to the subject position are the result of photos in which the captioned subject is not captured at the center of the photo. For example, the shooting direction of photos shown in Figure 13 is similar to the direction of the shooting position to the subject position, whereas the shooting direction of the photos shown in Figure 14 is different from the direction of the shooting position to the subject position. This problem can be solved by excluding the corresponding photos as statistical exceptions when the system obtains more photos as feedback.

There are regions whose likelihood values are set low or are not included in these figures because users cannot enter the regions or because the subjective facilities are occluded by other buildings; in these cases, the facilities are less likely to be shot by users. For example, although a photo shown in Figure 15(b) shot from (B) in Figure 15(a) whose likelihood is high captures Nakamon, a photo shown in Figure 15(c) shot from (C) whose likelihood is low does not capture Nakamon, which is occluded by Hakkakutoro.

More appropriate candidate captions for a photo can be presented to a user since positions that are favorable and unfavorable for users are determined through the construction of the extended geographical database based on the proposed method.

4. CONCLUSIONS

This study proposes a location-based photo-captioning framework using an extended geographical database. This framework considers not only a subject position but also the likelihood of the shooting position of a subject, estimated from the histories of the position from which the subject is shot, for location-based interactive captioning. Candidate captions for a photo are acquired by referring only to the data that corresponds to the shooting position and direction of the photo, using the proposed extended geographical database. In the experiments using the prototype system, we have constructed the extended geographical database using captioned photos and their shooting positions and directions. This extended geographical database achieved efficient photo captioning by considering the positions where a subject was occluded and the favorable positions for users.

As a future study, we will develop a method for constructing a more reliable extended geographical database by processing captioned photos with statistical techniques. We will also propose applications using a constructed extended geographical database, such as a location-based video annotation system.

5. REFERENCES

[1] Google Web API. `http://api.google.com/`.

[2] Panoramio. `http://www.panoramio.com/`.

[3] ZoneTag Photos. `http://zonetag.research.yahoo.com/`.

[4] S. Ahern, M. Davis, D. Eckles, S. King, M. Naaman, R. Nair, M. Spasojevic, and J. H.-I. Yang. Zonetag: Designing context-aware mobile media capture to increase participation. In *Proc. Pervasive Image Capture and Sharing: New Social Practices and Implications for Technology Workshop*, 2006.

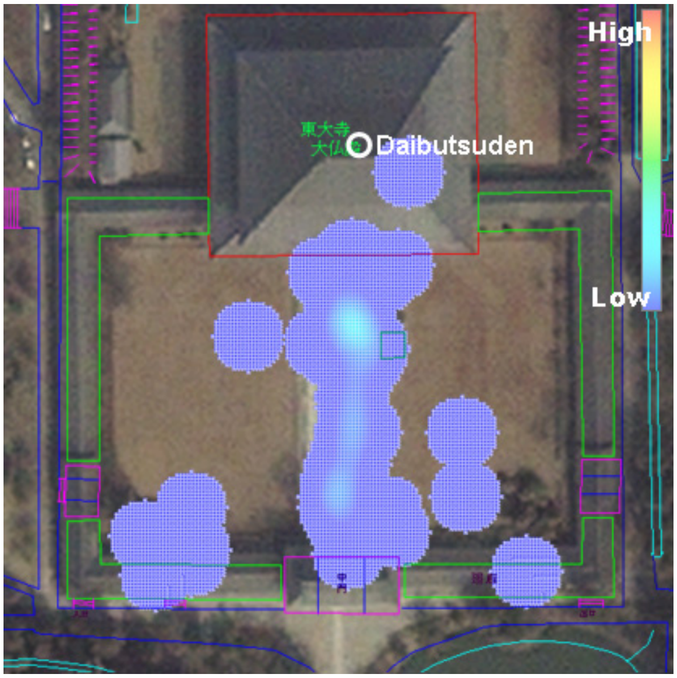
[5] H. Fujita and M. Arikawa. A photo vector field model for photo mapping considering users' lines of sight. In *Proc. Int. Joint Workshop on Ubiquitous, Pervasive and Internet Mapping*, pages 143–152, 2004.

[6] S. Ikehara, M. Miyazaki, S. Shirai, A. Yokoo, H. Nakaiwa, K. Ogura, Y. Ooyama, and Y. Hayashi. *Goi-Taikei — A Japanese Lexicon CDROM*. Iwanami Shoten, Tokyo, 1999.

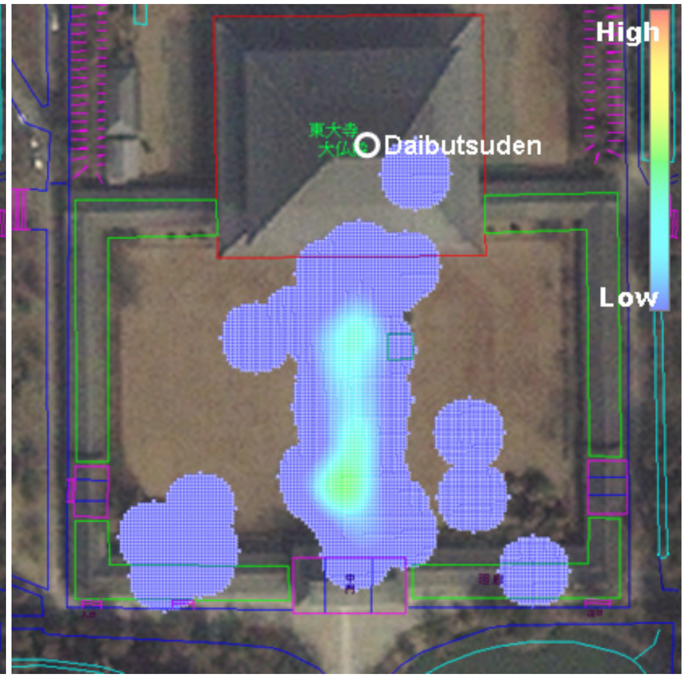
[7] K. Iwasaki, K. Yamazawa, and N. Yokoya. An indexing system for photos based on shooting position and orientation with geographic database. In *Proc. IEEE Int. Conf. on Multimedia and Expo*, 1259, 2005.

[8] Japan Electronics and Information Technology Industries Association(JEITA). Exchangeable image file format for digital still cameras: Exif version 2.2, 2002.

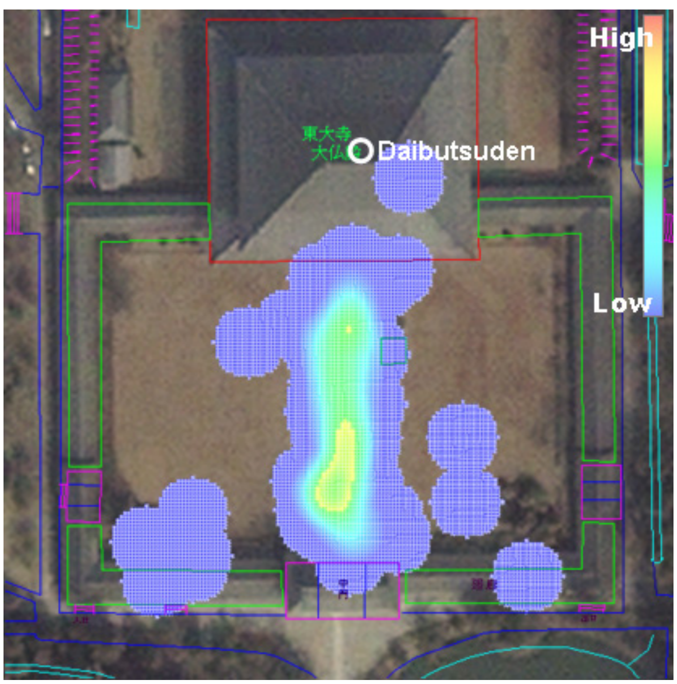
[9] A. Kuchinsky, C. Pering, M. L. Creech, D. Freeze, B. Serra, and J. Gwizdka. Fotofile: a consumer multimedia



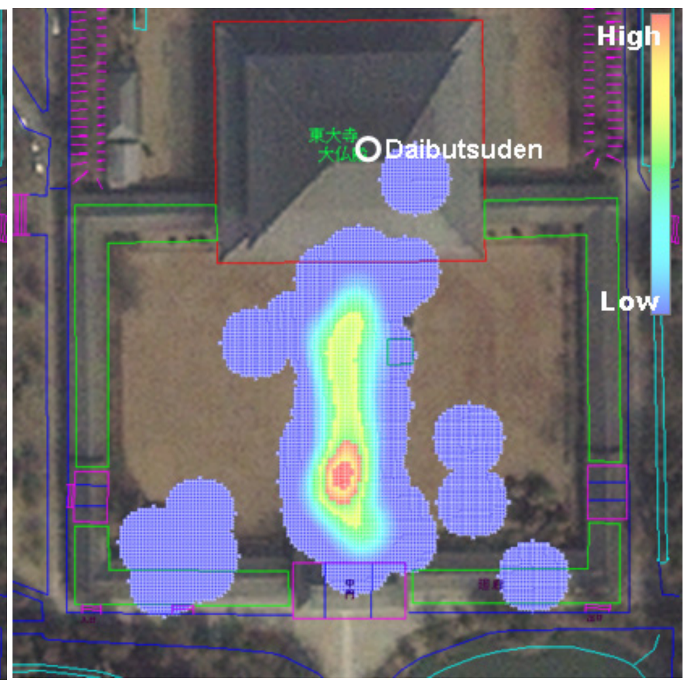
(a) 100th photo.



(b) 200th photo.

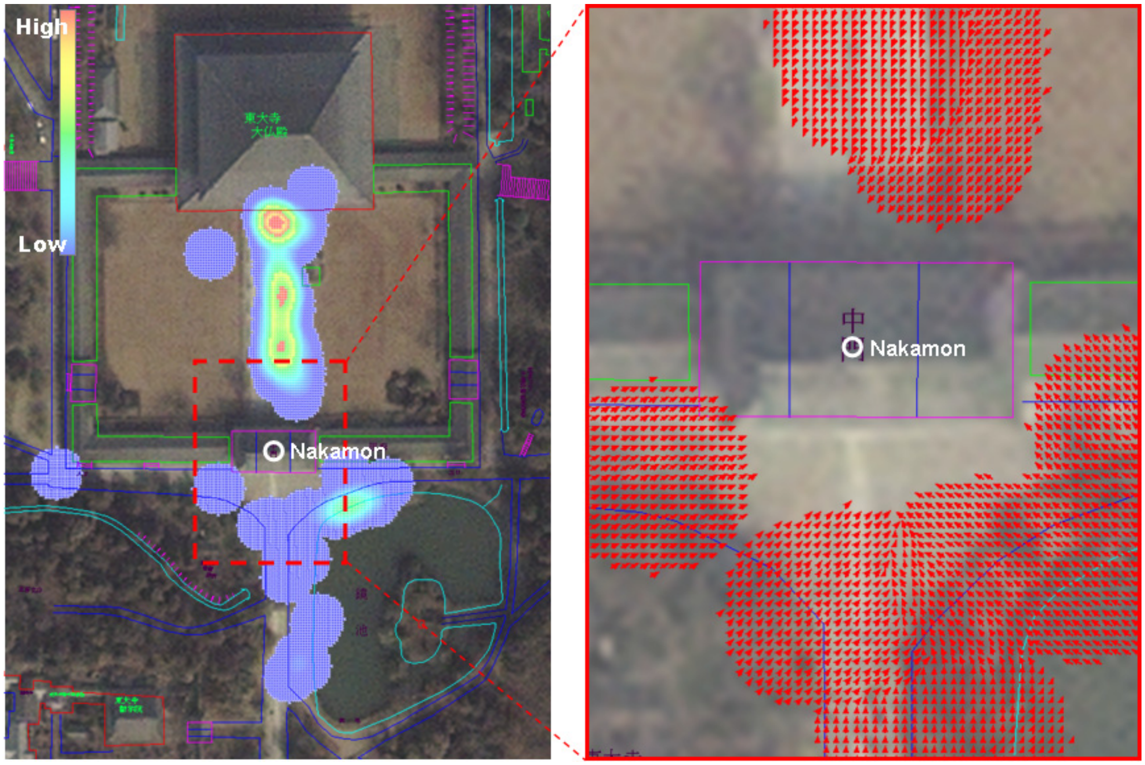


(c) 300th photo.

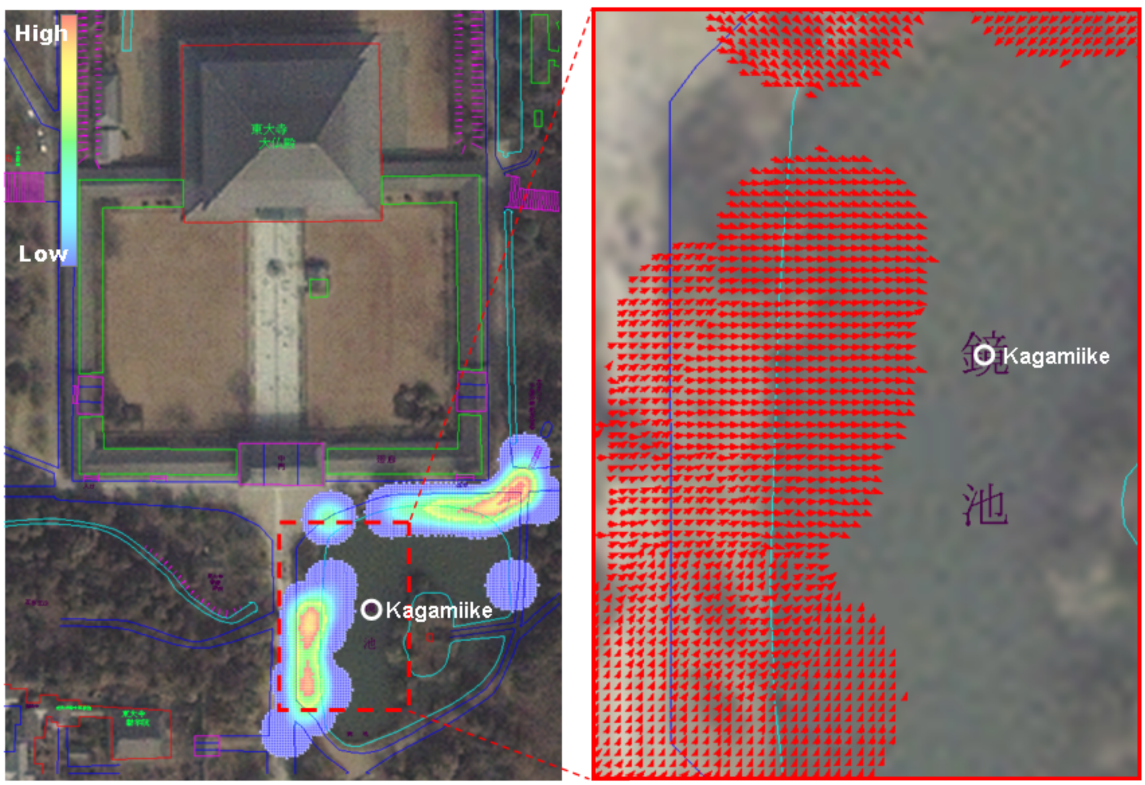


(d) 400th photo.

Figure 11: Extended geographical database construction process: Daibutsuden.



(a) Likelihood for position (left) and mainly-shot direction (right) of Nakamon.



(b) Likelihood for position (left) and mainly-shot direction (right) of Kagamiike.

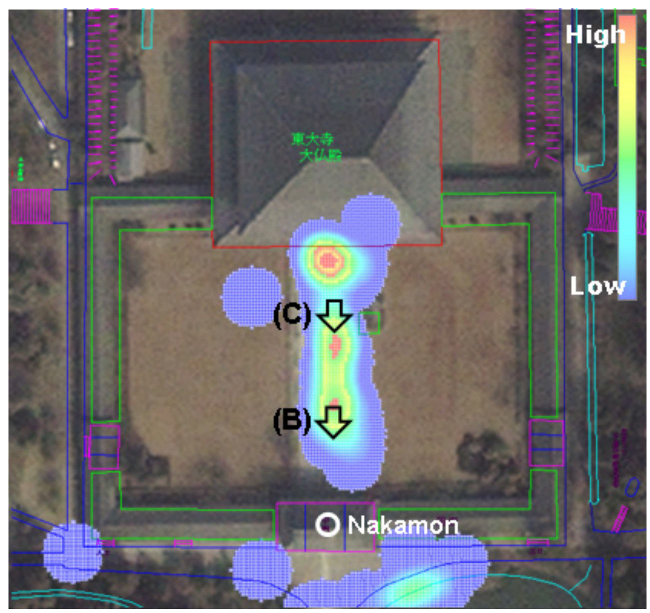
O subject position ← mainly-shot direction of each position

Figure 12: Results of extended geographical database construction.

organization and retrieval system. In *Proc. SIGCHI Conf. on Human factors in computing systems*, pages 496–503, 1999.
[10] H. Lieberman, E. Rosenzweig, and P. Singh. Aria: An agent for annotating and retrieving images. *IEEE Computer*, 34(7):57–61, 2001.
[11] M. Naaman, A. Paepcke, and H. Garcia-Molina. From where to what: Metadata sharing for digital photographs with geographic coordinates. In *Proc. 10th Int. Conf. on Cooperative Information Systems*, pages 196–217, 2003.
[12] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. In *Proc. 4th ACM/IEEE-CS Joint Conf. on Digital Libraries*, pages 53–62, 2004.
[13] M. Naaman, R. B. Yeh, H. Garcia-Molina, and A. Paepcke. Leveraging context to resolve identity in photo albums. In *Proc. 5th ACM/IEEE-CS Joint Conf. on Digital Libraries*, pages 178–187, 2005.
[14] A. Pigeau and M. Gelgon. Building and tracking hierarchical geographical & temporal partitions for image collection management on mobile devices. In *Proc. 13th ACM Int. Conf. on Multimedia*, pages 141–150, 2005.
[15] R. Sarvas, M. Viikari, J. Pesonen, and H. Nevanlinna. Mobshare: controlled and immediate sharing of mobile images. In *Proc. 12th ACM Int. Conf. on Multimedia*, pages 724–731, 2004.
[16] S. Sato and Y. Sasaki. Automatic collection of related terms from the web. In *Proc. 41st Annual Meeting of the Association for Computational Linguistics*, pages 121–124, 2003.
[17] B. Shneiderman and H. Kang. Direct Annotation: A drag-and-drop strategy for labeling photos. In *Proc. Int. Conf. on Information Visualization*, pages 88–95, 2000.
[18] B. K. Smith, E. Blankinship, A. Ashford III, M. Baker, and T. Hirzel. Image maps: Exploring urban history through digital photography. In *Proc. Digital Cities*, pages 326–337, 2000.
[19] D. D. Spinellis. Position-annotated photographs: A geotemporal web. *IEEE Pervasive Computing*, 2(2):72–79, 2003.
[20] A. Stent and A. Loui. Using event segmentation to improve indexing of consumer photographs. In *Proc. 24th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 59–65, 2001.
[21] T. Tan, J. Chen, P. Mulhem, and M. Kankanhalli. Smartalbum: a multi-modal photo annotation system. In *Proc. 10th ACM Int. Conf. on Multimedia*, pages 87–88, 2002.
[22] K. Toyama, R. Logan, A. Roseway, and P. Anandan. Geographic location tags on digital images. In *Proc. 11th ACM Int. Conf. on Multimedia*, pages 156–166, 2003.
[23] L. Wenyin, S. Dumais, Y. Sun, H. Zhang, M. Czerwinski, and B. Field. Semi-automatic image annotation. In *Proc. Human-Computer Interaction - Interact'01*, pages 326–333, 2001.
[24] Y. Wu, E. Y. Chang, and B. L. Tseng. Multimodal metadata fusion using causal strength. In *Proc. 13th ACM Int. Conf. on Multimedia*, pages 872–881, 2005.

Figure 13: Photos whose shooting directions were similar to the directions of shooting positions to subject positions.

Figure 14: Photos whose shooting directions differed from the directions of shooting positions to subject positions.



(a) Likelihood of Nakamon and shot vector of photos

(b) Nakamon (observed from (B)).

(c) Occlusion of Nakamon by Hakkakutoro (observed from (C)).

Figure 15: Photos shot at different likelihood positions.