Captioning Photos Based on Shooting Position and Orientation with Geographic Database

Vision and Media Computing Lab.

IWASAKI Kiyoko
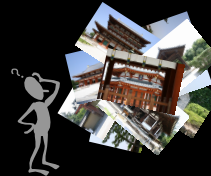

Background

Spread of digital cameras

⬇ Few methods or systems to manage photos simply

A huge amount of
photo data unorganized

2

Photo Management and Conventional Image Retrieval Methods

Conventional image retrieval methods are roughly classified into two categories:

1. a method using low-level image features
2. a method using textual information (metadata)

Mills et al. (2000) *

Retrieval by textual information out-performed visual-based retrieval for the management of personal photograph collection.

Photo management methods that add textual information to photos have been proposed.

* "Shoebox: A digital photo management system", Technical Report 2000.10, AT&T.   3

Photos with Location Information

- Exif
  - Specifying the formats of metadata for photos
  - Including caption, camera parameters, shooting position information and so on.
- GPS and camera attached cellular phone

Photos with location information will become popular in the near future.

Location-based captioning

4

Location-based Captioning

Spread of digital cameras

⬇

A huge amount of
photo data unorganized

西塔[Saito]
(West pagoda)

東塔[Toto]
(East Pagoda)

金堂[Kondo]
(Main hall)

大講堂[Daikodo]
(Lecture hall)

東院堂[Toindo]
(East hall)

鐘楼[Shoro]
(Belfry)

＊They are buildings in "薬師寺" [Yakushiji], a temple in Nara.

Efficient browse/retrieval by captioning

5

Objective

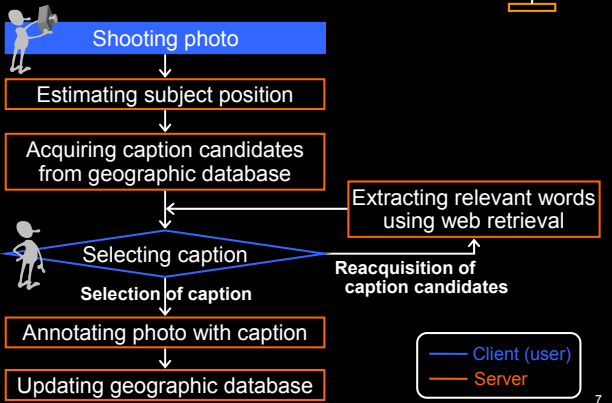**A semi-automatic photo captioning system**

Enabling users to make location-based captions for digital photographs easily

Approach

- A user selects a caption of a photo from a list of caption candidates.
- Caption candidates are acquired based on shooting position and orientation with:
  - (1) reference of geographic database.
  - (2) relevant word extraction using web retrieval.
- Geographic database is updated by adding captions selected by users as feedback.
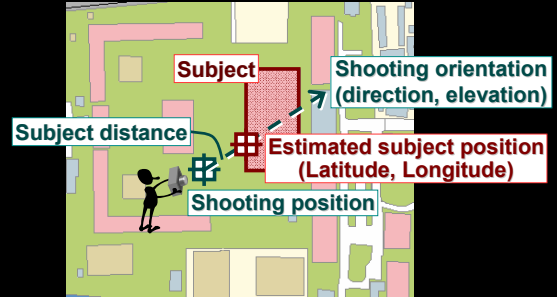
6

Flow Diagram of Photo Captioning

Shooting photo

↓

Estimating subject position

↓

Acquiring caption candidates from geographic database

↓

Selecting caption — **Selection of caption**

Extracting relevant words using web retrieval

Reacquisition of caption candidates

↓

Annotating photo with caption

↓

Updating geographic database

— Client (user)
— Server

Estimating Subject Position

To acquire caption candidates based on the subject position of a photo

Subject

Shooting orientation (direction, elevation)

Subject distance

Estimated subject position (Latitude, Longitude)

Shooting position

Acquiring Caption Candidates from Geographic Database (1/2)

Geographic database

- is stored in a server.
- initially consists of place/facility data included in map software on the market.
- is updated by using the captions selected by users as feedback.

西塔[Saito]
金堂[Kondo]
大講堂[Daikodo]
薬師寺[Yakushiji]
東塔[Toto]
東院堂[Toindo]

Name	Latitude	Longitude	Freq.
薬師寺 [Yakushiji]	34.668878	135.784313	0
金堂 [Kondo]	34.668686	135.784668	1
東塔 [Toto]	34.668041	135.784865	1
西塔 [Saito]	34.668073	135.784341	1
大講堂 [Daikodo]	34.668690	135.784668	1
東院堂 [Toindo]	34.667815	135.784865	1
⋮	⋮	⋮	⋮

Acquiring Caption Candidates from Geographic Database (2/2)

Caption candidate acquisition

1. Calculating distance between the estimated subject position and the position of each data in the geographic database
2. Extracting the data whose distance is within a threshold
3. Calculating likelihood that evaluates the relevance between each data and the content of a photo
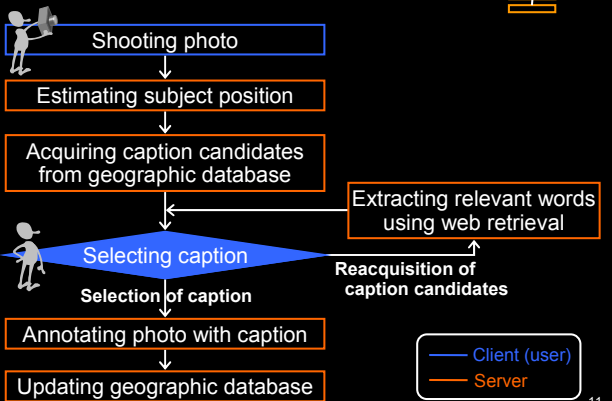4. Sorting the data in descending order of the likelihood

西塔[Saito]
金堂[Kondo]
大講堂[Daikodo]

Shooting direction

Estimated subject position

Shooting position

東院堂[Toindo]

Caption candidates

大講堂 [Daikodo]
薬師寺 [Yakushiji]
金堂 [Kondo]
東塔 [Toto]
西塔 [Saito]
東院堂 [Toindo]

Flow Diagram of Photo Captioning

Shooting photo

↓

Estimating subject position

↓

Acquiring caption candidates from geographic database

↓

Selecting caption — **Selection of caption**

Extracting relevant words using web retrieval

Reacquisition of caption candidates

↓

Annotating photo with caption

↓

Updating geographic database

— Client (user)
— Server

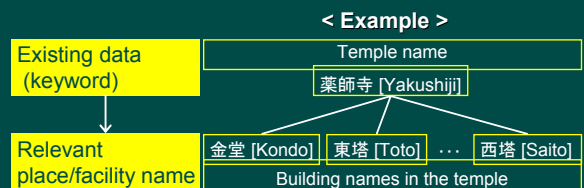## Extracting Relevant Words Using Web Retrieval (1/2)

The system extracts relevant words such as more detailed names than those stored in the geographic database.

< Example >

Existing data (keyword)

Temple name
薬師寺 [Yakushiji]

Relevant place/facility name

金堂 [Kondo]　東塔 [Toto]　…　西塔 [Saito]
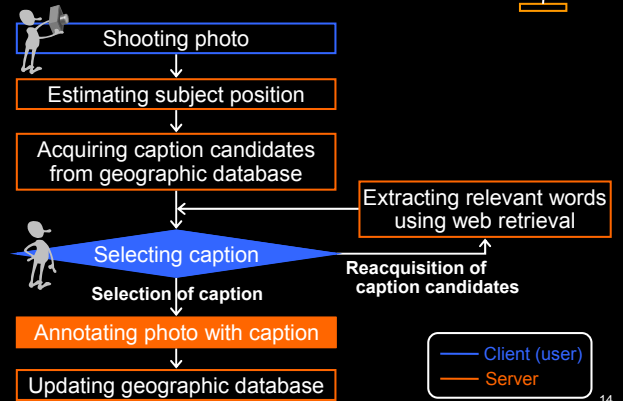
Building names in the temple

Extracting Relevant Words Using Web Retrieval (2/2)

A keyword which is a word related to the desired caption is selected from presented caption candidates by a user.

1. Acquiring the URL list of relevant web pages by web retrieval with the keyword
2. Acquiring top N pages of web search results
3. Extracting nouns from the web pages
4. Calculating relevance between the keyword and each extracted noun
5. Sorting the nouns in descending order of the relevance

13

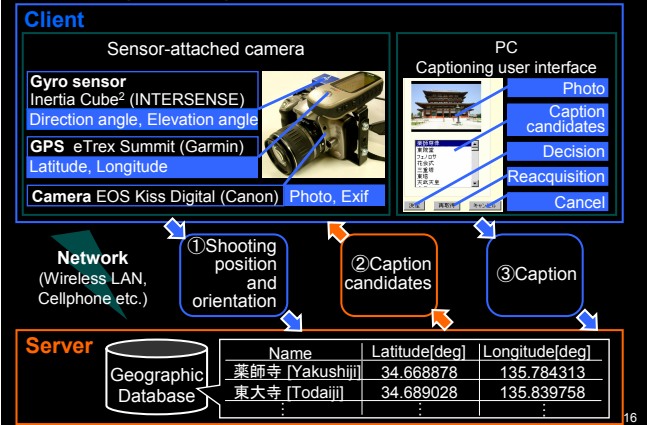## Flow Diagram of Photo Captioning

- Shooting photo
- Estimating subject position
- Acquiring caption candidates from geographic database
- Extracting relevant words using web retrieval
- Selecting caption
- **Selection of caption**
- **Reacquisition of caption candidates**
- Annotating photo with caption
- Updating geographic database

— Client (user)
— Server

14

Feedback on Geographic DB Based on User Selection

Source of the selected caption:

- Existing geographic data
  - →Updating the frequency of the word
- Relevant word extraction
  - New caption
    - →Registering the word on the geographic database with the estimated subject position
  - Existing caption
    - →Updating the position and the frequency of the word

15

Prototype System

Client

Sensor-attached camera

Gyro sensor
Inertia Cube² (INTERSENSE)
Direction angle, Elevation angle

GPS eTrex Summit (Garmin)
Latitude, Longitude

Camera EOS Kiss Digital (Canon) Photo, Exif

PC
Captioning user interface
- Photo
- Caption candidates
- Decision
- Reacquisition
- Cancel

Network
(Wireless LAN,
Cellphone etc.)

①Shooting position and orientation
②Caption candidates
③Caption

Server
Geographic Database

Name	Latitude[deg]	Longitude[deg]
薬師寺 [Yakushiji]	34.668878	135.784313
東大寺 [Todaiji]	34.689028	135.839758
:		

16

Preliminary Experiment – Captioning Photos

Photos to be captioned

- The photos are shot at "薬師寺" [Yakushiji], a temple in Nara.
- A subject of the photos is the building named "金堂" [Kondo].
- User aims to append caption "金堂" [Kondo] to the photos.
- An initial geographic database has
  - data entry of "薬師寺" [Yakushiji].
  - no data entry of "金堂" [Kondo].

Geographic Database

Name	Latitude[deg]	Longitude[deg]
薬師寺 [Yakushiji]	34.668878	135.784313

17

Processes of Captioning Photos (1/2)

Photo

Caption candidates

薬師寺（奈良県）
薬師寺駐車場（奈良県）
南都銀行（奈良県）
奈良西ノ京郵便局（奈良県）
モービル（奈良県）
西の京病院（奈良県）
観音寺（奈良県）

西塔
白鳳時代
統天皇
天武天皇
修二会
奈良駅
金堂

Geographic DB reference
"薬師寺" [Yakushiji] is ranked 1st.

Relevant word extraction
"金堂" [Kondo] is ranked 19th.

18

Processes of Captioning Photos (2/2)

Geographic DB
Adding "金堂" [Kondo]

Name	Latitude[deg]	Longitude[deg]
薬師寺 [Yakushiji]	34.668878	135.784313
金堂 [Kondo]	34.667972	135.784306

金堂
薬師寺駐車場（奈良県）
薬師寺（奈良県）
西ノ京（奈良県）
墨運堂（奈良県）
奈良西ノ京／京都郵便局（　県）
南都銀行（奈良県）

Geographic DB reference
"金堂" [Kondo] is ranked 1st.

Experiment – Evaluation of Captioning

Photos to be captioned

"薬師寺" [Yakushiji]: 72 photos (9 facilities)

"法隆寺" [Horyuji]: 49 photos (12 facilities)

Assuming that the user selects a facility name as the caption for a photo whose content is the facility.

Presented Rank Order of Caption

Averaged rank of user-selected captions in presented candidates

	Reference of geographic database		Relevant word extraction		Number of non-captioned photos
	Rank	Number of photos	Rank	Number of photos	
"薬師寺" [Yakushiji]	1.4	56	23.8	6	10
"法隆寺" [Horyuji]	2.8	37	117.6	8	4

Candidates acquired by :
- relevant word extraction ➜ captioning takes much time
- reference of geographic database ➜ captioning easily

Summary

- A semi-automatic photo captioning system based on shooting position and orientation
- Experiments using a prototype system
  - Feasibility of adding appropriate captions to photos based on location information
  - Presenting more appropriate caption candidates from a database after its incremental update during captioning process

Future Work
  - Improvement of matching between the position data in geographic DB and the estimated subject position
  - Experiments in wide area by various users