Improvement of Semi-automatic Location-based Photo Captioning Using Geographical Database

IWASAKI Kiyoko
Vision and Media Computing Lab.

Location-based Photo Captioning

Easy taking of photos using camera phone

⬇

A huge amount of photo data unorganized

Need efficient method/system

Photo captioning based of location information
Acquiring shooting position (latitude/longitude) using GPS
Shooting position based captioning
Acquiring shooting direction and distance
between shooting position and subject position
Subject position based captioning
Generating place/facility name added photos
Using retrieval, browse, and album photos

Kondo
Saito

2

Conventional Research and Problems

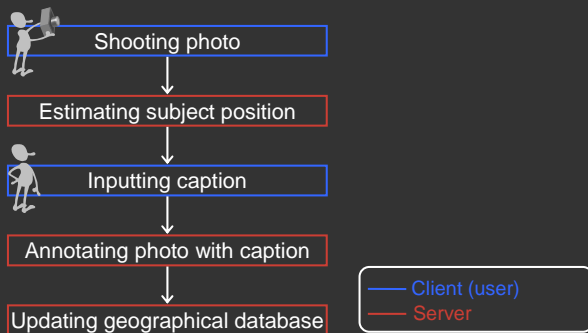- Organization of photos based on shooting position
[Naaman et al., 2004]
  Grouping photos and adding name to the groups
    The name is NOT intended by a user because of referencing database prepared in advance.
    Semi-automatic captioning to add proper caption

- Showing of caption candidates for photos to a user
[Sarvas et al., 2004]
  – Preparing a database including photos and their captions
  – Showing caption candidates based on shooting position and image similarity
  – Deciding a caption for a photo when taking the photo
    A photo does NOT processed in practical time (takes several minutes).
    Captioning on site in practical time

3

Proposed Photo Captioning

Semi-automatic location-based photo captioning

[Kondo]
[Toto]
[Saito]

[Kondo]

Taking a photo   Captioning
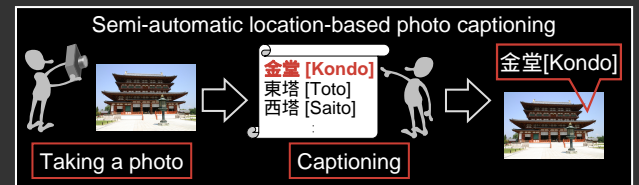
- Captioning
  – Semi-automatic: selecting a caption from candidates to add appropriate caption
  – On site   to lighten user's labor
- Geographical database
  – is stored in a server.   to share by multiple users
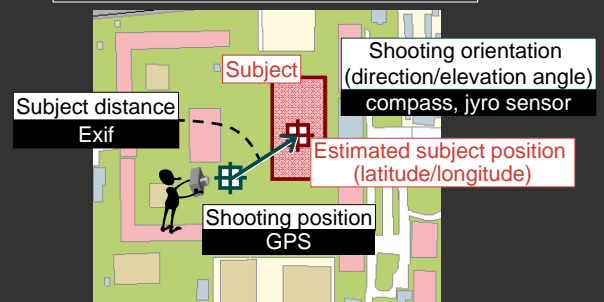 – is enhanced by dynamic update.

4

Flow Diagram of Photo Captioning

Shooting photo

⬇

Estimating subject position

⬇

Inputting caption

⬇

Annotating photo with caption

⬇

Updating geographical database

Client (user)
Server

5

Subject Position Estimation

Calculated from
– shooting position/orientation
– subject distance

Subject

Shooting orientation
(direction/elevation angle)
compass, jyro sensor

Subject distance
Exif

Estimated subject position
(latitude/longitude)

Shooting position
GPS

6

Flow Diagram of Photo Captioning

Shooting photo

↓

Estimating subject position

↓

Inputting caption

↓

Annotating photo with caption

↓

Updating geographical database

— Client (user)
— Server

Caption Inputting

Estimated subject position

↓

[DB] Acquiring caption candidates by reference to geographical DB

↓

Selecting caption — Reacquisition of caption candidates

Selection of caption

[web] Acquiring caption candidates by relevant word extraction using web retrieval

↓

Selecting caption — Acquisition of caption

Selection of caption
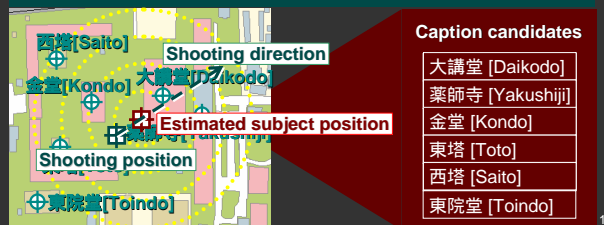
[key] Inputting caption by keyboard

↓

Caption

Geographical Database

- is stored in a server to share by multiple users.
- is initialized by data included in map software.
- is updated by: user selected frequency.
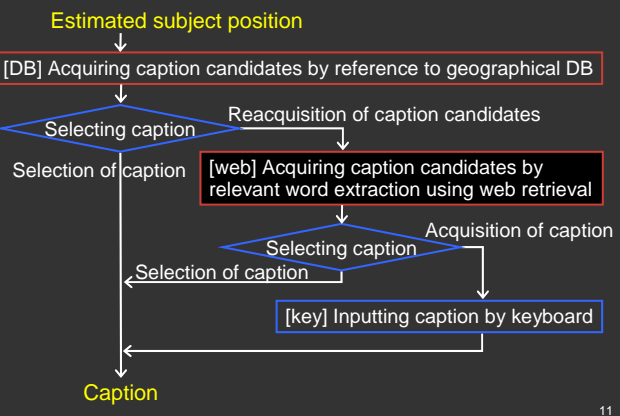 data acquired by other inputting way.

[Saito]
[Kondo]
[Daikodo]
[Yakushiji]
[Toto]
[Toindo]

Name	Latitude	Longitude	Freq.
[Yakushiji]	34.668878	135.784313	0
[Kondo]	34.668686	135.784668	1
[Toto]	34.668041	135.784865	1
[Saito]	34.668073	135.784341	1
[Daikodo]	34.668690	135.784668	1
[Toindo]	34.667815	135.784865	1
⋮	⋮	⋮	⋮

Caption Candidates Acquisition by Reference to Geographical Database
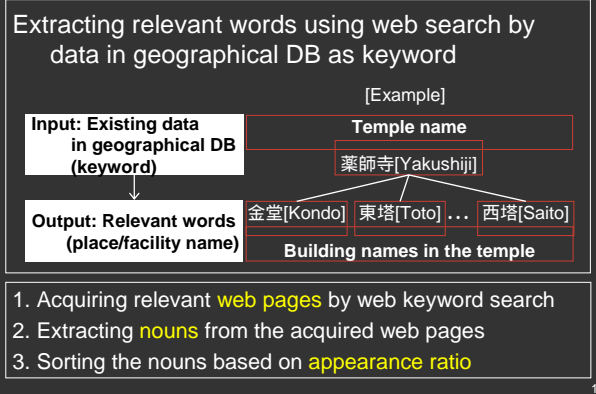
1. Acquiring data whose distance from the estimated subject position within a threshold
2. Calculating likelihood of each data based on similarity of distance and direction
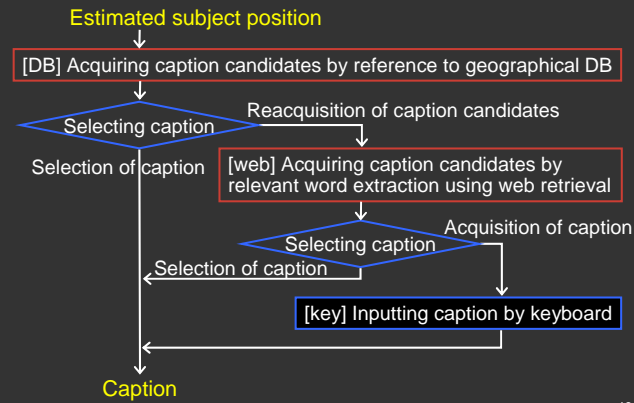3. Sorting data based on likelihood and user selected frequency

[Saito]
[Kondo]
[Daikodo]
Shooting direction
Estimated subject position
Shooting position
[Toindo]

Caption candidates

Caption candidates
[Daikodo]
[Yakushiji]
[Kondo]
[Toto]
[Saito]
[Toindo]

Caption Inputting

Estimated subject position

↓

[DB] Acquiring caption candidates by reference to geographical DB

↓

Selecting caption — Reacquisition of caption candidates

Selection of caption

[web] Acquiring caption candidates by relevant word extraction using web retrieval

↓

Selecting caption — Acquisition of caption

Selection of caption

[key] Inputting caption by keyboard

↓

Caption

Relevant Word Extraction Using Web Retrieval

Extracting relevant words using web search by data in geographical DB as keyword

[Example]

Input: Existing data in geographical DB (keyword)

Temple name

[Yakushiji]

Output: Relevant words (place/facility name)

[Kondo] [Toto] [Saito]

Building names in the temple

1. Acquiring relevant web pages by web keyword search
2. Extracting nouns from the acquired web pages
3. Sorting the nouns based on appearance ratio

Caption Inputting

Estimated subject position
↓
[DB] Acquiring caption candidates by reference to geographical DB
↓
Selecting caption — Reacquisition of caption candidates
Selection of caption → [web] Acquiring caption candidates by relevant word extraction using web retrieval
↓
Selecting caption — Acquisition of caption
Selection of caption →
↓
[key] Inputting caption by keyboard
↓
Caption

13

Flow Diagram of Photo Captioning

Shooting photo
↓
Estimating subject position
↓
Inputting caption
↓
Annotating photo with caption
↓
Updating geographical database

— Client (user)
— Server

14

Update of Geographical Database

Updating depends on how a caption acquired by

DB
↓
select
↓ web
↓
select
↓ key
↓
caption

Geographical DB
initial data
Increasing selected frequency
added data
Registering caption and estimated subject position
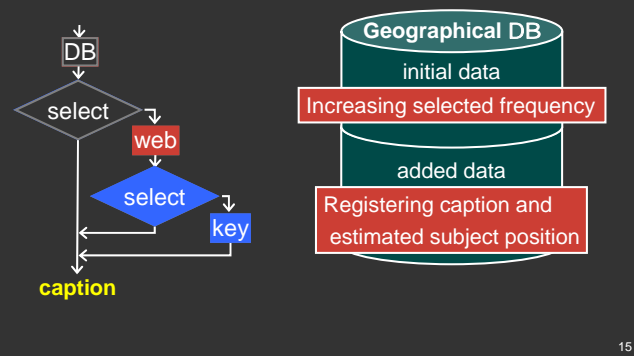
15

Problem with Current Framework

- Relevant word extraction
  Caption candidates include words which are not relevant to location information.
  -> Referencing lexicon in extraction process

- Update of geographical database
  New data is NOT registered accurate position.
  -> Storing geographical data as region

16

Elimination of Non-relevant Words

- Referencing "A Japanese Lexicon" ( ) in relevant word extraction process
- Eliminating words which are non-relevant to place or facility information

(East pagoda)   -> Caption candidate
(abbreviated name)   -> Eliminated

About 50% reduction of the number of caption candidates
-> Easier user inputting (=captioning)

17

Problem with Current Framework

- Relevant word extraction
  Caption candidates include words which are not relevant to location information.
  -> Referencing lexicon in extraction process

- Update of geographical database
  New data is NOT registered accurate position.
  -> Storing geographical data as region

18

Subject Position Estimation

Registering estimated subject position as position of data in geographic database
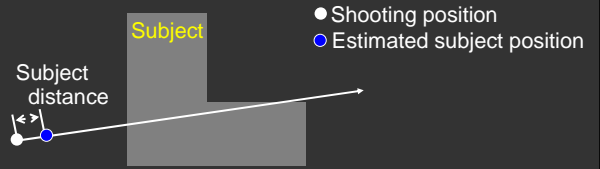
Case1: "subject distance" is acquired accurately.

- Subject
- Subject distance
- Shooting position
- Estimated subject position
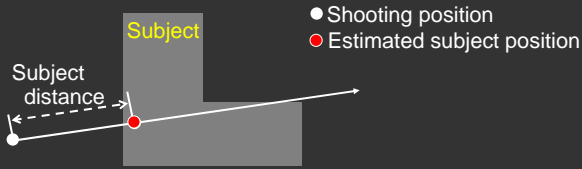
Accurate subject position is registered.

Problem of Subject Position Estimation
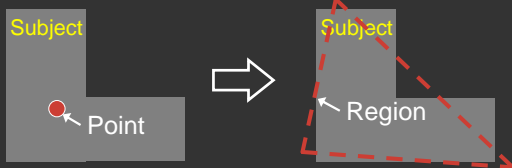
Case2: "subject distance" includes error.

- Subject
- Subject distance
- Shooting position
- Estimated subject position

Inaccurate subject position is registered.

Subject Region

- Geographical data is registered as region not as point.
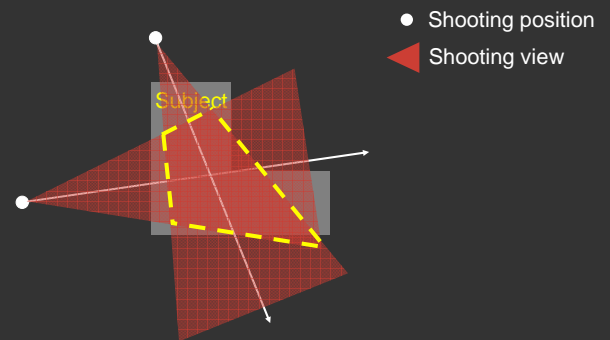- The region indicates subject existing region approximately.

- Subject
- Point
- Subject
- Region

- The region is estimated from users' shooting view.

Subject Region Estimation

Using intersection region of shooting view
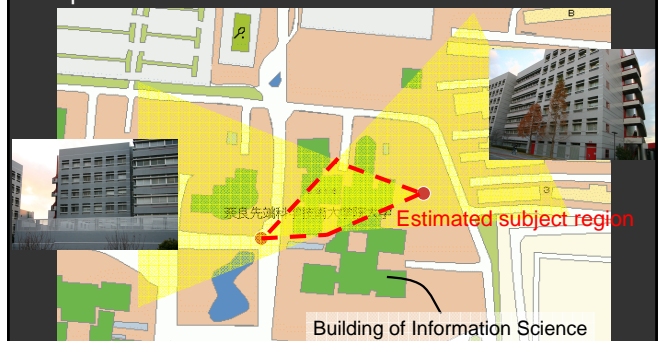

- Shooting position
- Shooting view
- Subject

Subject Region Estimation

Using intersection region of shooting view

- Shooting position
- Shooting view
- Subject
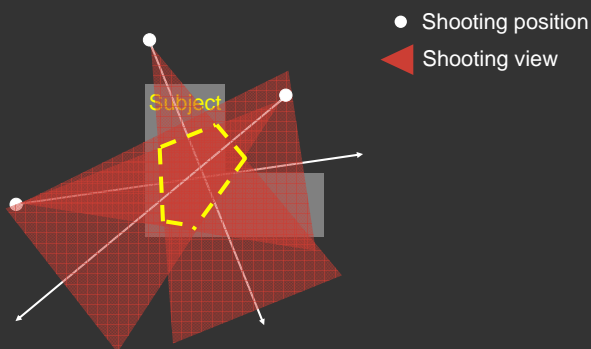
Example of Subject Region Estimation

Estimated from two photos and their shooting position and orientation information

Estimated subject region

Building of Information Science

Summary

Two improvements for current photo captioning
framework:
- Elimination of non-relevant words using lexicon
  -> Reducing the number of caption candidates which
  are presented to a user
- Subject region estimation of geographical data

Future Work

- Verification of subject region estimation
- Design of geographical database for subject region
  and hierarchical data structure

25