IDeixis
- Searching the Web with Mobile Images for Finding Location-Based Information

Konrad Tollmar (*,**), Tom Yeh (*), Trevor Darrell (*)
(*) MIT Computer Science and Artificial Intelligence Laboratory
(**) Lund University Department for Informatics

...what is this?

...and this?

Today - MMS

iDeixis
Image-based Point and Click
Outline

- Testing the concept: Prototype I system
- Concept grounding: Images-based searching for location-based information
- User study: Prototype II system
- Vision Technology
- Future work

Image Matching Made Simpler

- Landmark can occupy most of the image
  → can avoid segmentation
- People tend to take images of a famous landmark from the same location
  → can reduce viewpoint variation

Searching the Web with Mobile Images for Location Recognition (Yeh, Tolmar & Darrell CVPR 04)

Image Matching Test

- Whether CBIR can match location images from mobile devices to the pages on the world wide web
- 12,000 web images collected from mit.edu
- Student volunteers to take a total of 50 images from each of three selected locations: Great Dome, Green Building and Simmons Hall
- Around 60% accuracy (Note difference between similar and relevant image)

First Prototype

A perfect search result?

Small Problem?
Problem 2 - "irrelevant" web pages

Interview Study
- 20 subjects in a couple of common tourist locations around Boston
- Use of maps and tour books?
- What do you want to know about a specific location?
- How do you want to obtain that information?
- How would you take pictures that express that particular interest?

"Take pictures of things and places that you are interested in and would like to know a bit more about"

1 / What is this building called?
2 / Who is the architect?
3 / What is the bus schedule?
4 / What kind of tree is this?
5 / Where could I buy this book?

Findings
- Our subjects asked two main questions:
  - Where can I find X?
  - What is this?
- Need for both general and more specific information

Three types of search mechanisms
1 – Searching for web pages
2 – Keyword Boosting
Three types of search mechanisms

1. Searching for keywords
2. Take a picture
3. Select a keyword

Prototype Evaluation

- 16 subjects aged between 13 and 63
- Two locations:
  - The MIT Dome
  - Stockholm City Hall
- Semi-functional
  - Pre-computed image search
  - "Live links"
  - Video-recorded outdoor experiment

Finding Web Pages

- Almost no problem with use: "Internet in the phone is cool"
- Webthumbnails are too small: "I prefer text, but I see a terrible website I don't need to go there"
- Find matching web pages is not good enough. User needs additional support in finding the "right" web page that matches their need for information. "I'm looking for more specific info, and not the general that the top ranked page provide"

Using maps and keywords for finding information

- Problems with keywords, most don't understand keywords directly.
  - “Keyword lead me out in tangents”
  - “Use keyword for fast search and similar web pages for general information”
- Maps are problematic.
  - "I have no idea what this map is covering"
- Effective if a location-information ontology is available to support the search for the right information.

Quantitative comparison of search strategies.

<table>
<thead>
<tr>
<th>16 subjects / 32 trails</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Strategy</td>
</tr>
<tr>
<td>Web Pages</td>
</tr>
<tr>
<td>Extracted Keywords</td>
</tr>
<tr>
<td>MapQuest</td>
</tr>
</tbody>
</table>
Finding location images
- There are many location images on the web.

Finding location images
- There are also many other kinds of images on the web.

Finding location images
- Removing images of humans

Finding location images
- Removing man-made graphics

CBIR (Content-based Image Retrieval)
- Finding similar images by
  - Texture
  - Color
  - Shape
  - Structure
  - Region
- Preliminary experiment has shown the feasibility this application, and current approach

Experiment
- 29 MIT Locations
- 222 Images
Result

- **Accuracy**: % of top 16 images that are correct
- **Robustness**: % of correct images that fall within top 16
- **Performance**: % of test cases when there is at least one correct image within rank 16

<table>
<thead>
<tr>
<th></th>
<th>IRM</th>
<th>STAIR</th>
<th>ANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>47%</td>
<td>29%</td>
<td>40%</td>
</tr>
<tr>
<td>Robustness</td>
<td>82%</td>
<td>51%</td>
<td>76%</td>
</tr>
<tr>
<td>Performance</td>
<td>98%</td>
<td>86%</td>
<td>96%</td>
</tr>
</tbody>
</table>

IMAGE and TEXT hybrid image search

**Content-based filtering**

- Textual similarity ≠ visual similarity

**Future work**

- Evaluation of functional prototype with live image database
- New designs for browsing large search results
- Multi-modal interface
- Image-based Object Search

**Related Work**

- Camera-equipped Mobile Devices
  - AR-PDA (Gausemeier & Bruederlin 2003)
- Content-based Image Retrieval
  - IBM’s QBIC (Niblack et al 1993) / Webseek (Smith & Chang 1995)
- Location Recognition
  - SLAM (Montemerlo 2002)
- Location-based Information Retrieval
  - Mapquest.com / GUIDE (Cheverst 2000)
Conclusion

- A new image-based paradigm for location-aware computing
- Our method can refer to distant locations
- Directly matching similar web pages provide an intuitive way of searching for information that involves much less interaction than conventional interfaces, but can also yield in search results that sometimes are too general.
- Keyword-based interface more cognitive demanding but can find more specific information faster.