Active Query Sensing for Mobile Location Search

Felix. X. Yu, Rongrong Ji, Shih-Fu Chang

Digital Video and Multimedia Lab
Columbia University

November 29, 2011
Outline

• Motivation
• Problem Justification
• Active Query Sensing
  – Offline Salient View Learning
  – Online Active Query Sensing
• Evaluation
• User Interface
• Conclusion
Outline

• Motivation
• Problem Justification
• Active Query Sensing
  – Offline Salient View Learning
  – Online Active Query Sensing
• Evaluation
• User Interface
• Conclusion
Motivation

• Mobile Visual Search
  – Search information about products, locations, or landmarks
  – By taking a photo of the target of interest using the mobile device

• Commercial Systems
  – Snaptell
  – Nokia point and find
  – Google goggles
  – Kooaba

• Research Prototypes
  – Nokia Stanford MVS system
  – PKU mobile landmark search
  – MIT sixth senses
  – MSRA Photo2Search prototype
Performance of Mobile Visual Search Varies

• Search success depends on
  – How distinctive object classes are
  – View variation and object deformation
  – Quality of captured images

• Not every view of an object or location is distinctive enough as a successful query
  – Only 47% of the first query succeed in our location search endeavor on Manhattan 0.3m street view dataset
  – 0.4-0.7 average precision in the state-of-the-art location search systems ([Knopp 2010] [Girod 2011] etc.)
Motivation

How to guide the user to take the mobile query?

- Which view will be the best query?
  - For example, in mobile location search:
  - Or in mobile product search:

- A General Yet Untouched Problem for searching locations, objects and landmarks.
Motivation

• Possible Solutions
  – Solution I
    • User takes a video query by scanning the surrounding
    • Then the mobile end selects one or several distinct frames
      (views) to be sent to the server

• Solution II
  • User takes a query at his/her will
  • Once this query fails, the server suggests a best viewing
    angle to take the second query
  • Intelligently guide the mobile user to take a second query

We focus on solution II, enabling users and machines to collaborate

Relieve user from trial and error
Motivation

• Research Objective
  – Use mobile location search as a case study
    • We will develop a general framework that can be extended to mobile object retrieval and product search
  – Why not GPS?
    – GPS may not be reliable in the dense urban jungle areas
    – Visual based solutions can also be used to refine GPS solutions
    – Our research goal is general: a special case of multi-view mobile search
Outline

• Motivation
• Problem Justification
• Active Query Sensing
  – Offline Salient View Learning
  – Online Active Query Sensing
• Evaluation
• User Interface
• Conclusion
NAVTEQ 0.3M NYC Data Set

- 300,000 images of 50,000 locations in Manhattan
- Collected by the NAVTEQ street view imaging system
NAVTEQ 0.3M NYC Data Set

• Location Sampling
  – Locations are imaged at a four-meter interval on average
  – Six camera views for each location separated by 45°

• Visual Data Organization
  – Six views (images)
  – Also provide panorama (used for visualization in this work)
Problem Justification

• Our goal
  – Once the first query fail, how to take the best second query to correctly find his/her current correct location.
  
• How much can AQS help
  – The percentage of locations that are partially searchable.

Targets of AQS

Subsample 200 locations to “# of searchable views” using cropped Google street view.
Outline

• Motivation
• Problem Justification
• Active Query Sensing
  – Offline Salient View Learning
  – Online Active Query Sensing
• Evaluation
• User Interface
• Conclusion
Active Query Sensing System

• Active Query Sensing Demo
  (refer the demo video)
Active Query Sensing System

Basic Idea and Workflow

- **Offline**
  - Salient view learning for each reference location
- **Online**
  - Viewing angle prediction of the first query
  - Suggest new views by majority voting
Outline

• Motivation
• Problem Justification
• **Active Query Sensing**
  – Offline Salient View Learning
  – Online Active Query Sensing
• Evaluation
• User Interface
• Conclusion
Million Scale Image Matching

Extend and modify from [Nister 06][Schindler 07]

- **Offline**
  - Codebook construction and BoW quantization
  - Inverted location/image indexing

- **Online**
  - Inverted indexing based search + spatial verification
Location Score Calculation

Location \( i \) 1 2 3 4 5 6

Location \( j \) 1 2 3 4 5 6

Location \( k \) 1 2 3 4 5 6

View 1 View 2 View 3 View 4 View 5 View 6

Small scores Large scores
Offline: Salient View Learning

• Question
  – Among 6 views of a location, which is most informative?

• Definition
  – **Salient view** is the most discriminative view in a location

[Diagram of six views with a geographical scatter graph and a query over the salient view leading to successful results]
Offline Salient View Learning

• Solution I: Content-based prediction
  – Occurrence of attributes/content
    • Such as detecting discriminative objects in this image
  – Visual word discriminability distribution
    • Step I
      • Measure the frequency of the discriminative words
      • Use term frequency and inverted document frequency
      • Build a TF-IDF frequency histogram
      • Example of discriminative words
    • Step II
      • Train SVM on to predict the saliency
Offline Salient View Learning

- Solution II: Self retrieval testing
  - Goal: Determine whether a view is salient
  - Approach: Using the reference image of each view to retrieve the corresponding location

- Step I
  - For each reference image, search the database to obtain the score distributions of both positive and negative results

- Step II
  - Evaluate the score distribution's robustness to measure the online query easiness

An ideal score distribution

\[
Saliency = \frac{\sum_{i=1}^{N} \sum_{j=1}^{i} \text{score}(j) \text{rel}(j) / i}{\sum_{i=1}^{N} \sum_{j=1}^{i} \text{score}(j) \text{rel}(j) / i},
\]

(1) \( N \) is the number of returned locations.

(2) \( \text{score}(j) \) is the location matching score: the maximal score of its six views.

(3) \( \text{rel}(j) \) is the relevance judgement of the \( j \)th returned location, 1:correct, 0:incorrect.
Saliency Measure

Can be also used to access the location difficulty

- We define the confidence of each location as proportional to its saliency (max of the six views)
- Useful for the user to determine when AQS is reliable

Geographical location confidence distribution
Outline

• Motivation
• Problem Justification
• **Active Query Sensing**
  – Offline Salient View Learning
  – Online Active Query Sensing
• Evaluation
• User Interface
• Conclusion
Online Active Query Sensing

Examples of “junk” images:

Examples of query images selected by user:

• Assumption
  – User is unlikely to take “junk image” with no hope for finding the true target.
  – The first query, even unsuccessful, can be used as “probe” to narrow down the solution space.
Online Active Query Sensing

• Formulation
  – Maximizing uncertainty reduction

\[ v_2 = \max_{\mathbf{v}} \left( H(l|q_1) - H(l|q_1, \mathbf{v}) \right) \]

• A subsequent query should
  – maximize the uncertainty reduction (information gain)
  – across candidate locations, narrowed down by the query already been taken
  – in terms of getting the best view
Online Active Query Sensing

• In practice, we find maximizing uncertainty reduction can be well approximated by:
\[ \max_v \sum_{l \in L_N} Saliency(l, v) \]

• \( L_N \) is the candidate locations narrowed down by the first query.

• In other words, AQS suggests the user to turn to the “most salient” view by a weighted sum or majority voting in terms of \( Saliency \).
Online Active Query Sensing

For each location, we have its most salient view. The majority of the salient views decides the suggested (second) query.
What about current view unknown?

- Step I: Predict the view angle of the first query

Offline Training: Train view prediction classifiers offline

Online Prediction: View alignment based on the image matching

Our solution is to combine them both
What about current view unknown?

- Step II: Majority voting in terms of view change

  - Salient View (Offline)
  - Viewing Angle Prediction
  - View Change

  ![Diagram](image)

  Turn 90 degrees to the right
Outline

• Motivation
• Problem Justification
• Active Query Sensing
  – Offline Salient View Learning
  – Online Active Query Sensing
• Evaluation
• User Interface
• Conclusion
Query Set for Evaluation

• Manually cropped queries from Google Street View
  – 226 random locations
  – For each location: six query images are cropped from viewing angles similar to the view orientations used in the database

• A returned reference location is considered relevant
  – if it has visual overlapped view with the query image
Evaluation

- **Baseline**
  - Random view suggestion
  - Dominant view suggestion
    - Always change to front view
- **Salient view selection**
  - AP based
  - Saliency based
- **Oracle view suggestion**
  - Always suggest the ideal view

---

**Overall Performance**

![Failure rates over successive query iterations.](image)
Examples

First query

AQS suggested query

First query

AQS suggested query
Outline

• Motivation
• Problem Justification
• Active Query Sensing
  – Offline Salient View Learning
  – Online Active Query Sensing
• Evaluation
• User Interface
• Conclusion
User Interface

• We have developed an iPhone application of the proposed AQS system
  – The client communicate with the server through wireless network

• Time
  – Visual matching requires <1s
  – Query sensing suggestion is almost in real time
  – Query delivery latency depends on wireless link
User Interface

- How does the user know the first query is incorrect
  - Panorama
  - Geographical map context
- How to guide the user to take the second query
  - Compass, camera icon
- Point of interest
AQS Examples

<table>
<thead>
<tr>
<th>First Query</th>
<th>First Result</th>
<th>Ask AQS</th>
<th>Taking Second Query</th>
<th>Second Result</th>
</tr>
</thead>
</table>

1. First Query: Take an image
2. First Result: Found a location
3. Ask AQS
4. Taking Second Query: Follow the green arrow and turn
5. Second Result: Found a location
Outline

• Motivation
• Problem Justification
• Active Query Sensing
  – Offline Salient View Learning
  – Online Active Query Sensing
• Evaluation
• User Interface
• Conclusion
Conclusion

• We have designed the AQS system to
  – Intelligently guide user to take the second query without repetitive trial and error
  – Reduce the error rate from 53% to 12%

• The proposed approach is general. It can be extended to multi-view object search
  – If the database has multiple views of products, AQS can be used to suggest best view for searching

• We have demo about AQS on Thursday Morning Demo Session
Thank you

Any Questions?

We would like to also thank NAVTEQ for providing the NYC image data set, Dr. Xin Chen and Dr. Jeff Bach for their generous help.