

A Mobile Location Search System with Active Query Sensing

Felix X. Yu, Rongrong Ji, Tongtao Zhang, and Shih-Fu Chang
Department of Electrical Engineering, Columbia University
{yuxinnan,rrji,ttzhang,sfchang}@ee.columbia.edu

ABSTRACT

How should the second query be taken once the first query fails in mobile location search based on visual recognition? In this demo, we describe a mobile search system with a unique Active Query Sensing (AQS) function to intelligently guide the mobile user to take a successful second query. This suggestion is built upon a scalable visual matching system covering over 0.3 million street view reference images in New York City, where each location is associated with multiple surrounding views and panorama. In online search, once the initial search result fails, the system will perform online analysis and suggest the mobile user to turn to the most discriminative viewing angle to take the second visual query, from which the search performance is expected to greatly improve. The AQS suggestion is based on both offline salient view discovery and online viewing angle prediction and intelligent turning decision. Our experiments show our AQS can improve the mobile location search with a performance gain as high as 100%, reducing the failure rate to only 12% after taking the second visual query.

Categories and Subject Descriptors

H.4.3 [Information Systems Applications]

General Terms

Algorithms, System, Measurement

Keywords

Mobile Visual Search, Mobile Location Recognition

1. APPLICATION SCENARIOS

With the popularity of mobile handheld devices, we are seeing increasing demand and promising progresses in mobile location search using visual recognition techniques. The mobile user takes a photograph of his/her surrounding view and sends the captured image over the wireless link to the server, which matches the reference images of the candidate locations stored in the server to identify the true location of the mobile user. Such visual matching capabilities can be found in several commercial systems such as snaptell [1], Nokia point and find [1], kooaba [1], as well as research prototypes [2–4]. Complementary GPS or network based lo-

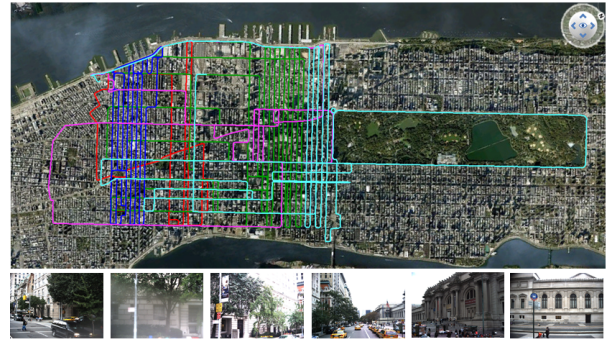


Figure 1: Top: Location coverage of 0.3 million reference images in our system; Bottom: six views of an example location in New York City.

calization techniques, mobile location search provides more precise localization plus indication of orientation. For instance, GPS information is typically in the range of 20-100 feet without orientation; but using visual search, improvement to about 10 feet plus orientation information is feasible. Furthermore, visual based localization can still be used if GPS signals are not available, as the case of urban jungles.

However, current visual-based localization systems still lack robust performance deemed acceptable in practice. In a preliminary trial we conducted, only less than 50% of the visual location queries were successful. This is consistent with the moderate accuracy (0.4-0.7 average precision) reported in the literature so far. Often the failure cases can be attributed to false alarms caused by the confusing locations with high visual similarities.

Distinct from prior works, we focus on a novel system [5] to improve the mobile visual search experience in subsequent user interaction, namely how to guide users to find the right queries that are more likely to work. We develop the system based on the observations that there exist certain unique salient views for each target location. For instance, certain views of a location are more discriminative since they contain unique “signatures” such as distinctive landmarks or skylines. In contrast, other views are less useful in distinguishing the given location from others because they contain common and confusing objects (trees, walls, etc.).

We demonstrate a mobile location search system with Active Query Sensing (AQS) capabilities to guide a user in turning the camera to capture the best views for subsequent queries. Our current prototype includes 0.3 million geographical reference images¹ of Manhattan, New York City,

¹We thank NAVTEQ for providing the NYC image data set, Dr. Xin Chen and Dr. Jeff Bach for their generous help.

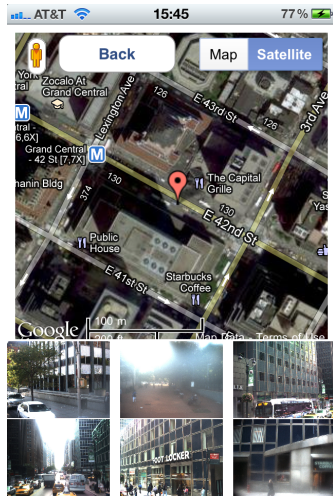


Figure 2: Our system finds and plot the top matched location (indicated by the geo tag) and associated surrounding views and points of interest in response to the given query. The user may verify the correctness of the result by comparing the surrounding views and the points of interest returned by the system.

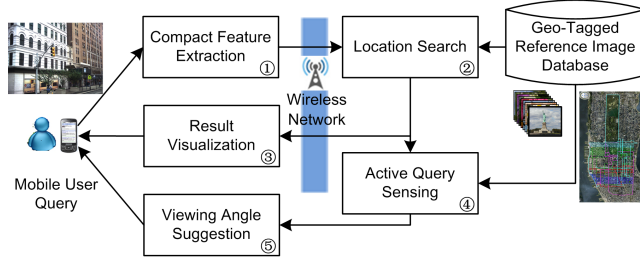


Figure 3: Workflow of the active query sensing system

covering about 50,000 locations each with six surrounding views (see Figure 1).

2. SYSTEM WORKFLOW

The procedure of our proposed AQS system is as follows (shown in Figure 3):

Step 1: First, user captures a query image of his surrounding using the camera phone. And the image (or alternatively the extracted compact descriptor [4]) is sent to the server through the wireless link.

Step 2: The location search system receives the visual query, searches matched images, and returns the top matched location (Figure 2).

Step 3: After checking and comparing the surrounding views and points of interest (as shown in Figure 2), if the user determines the predicted location is wrong, he may trigger active query sensing function to ask for the suggestion of the best query view.

Step 4: Upon user's request, our system calls the active query sensing module to come up with the best view for the next query. It sends specific instructions to the user (e.g., turn to the right about 30 degrees, turn up about 30 degrees etc.). Figure 4 shows a sample interface, indicating how the user is guided to follow the suggested view change.

Step 5: User follows the suggestion to turn the camera and take the next query image and send it for another location search, as shown in Figure 4.



Figure 4: Our AQS interface guides users to the best view that is predicted to have a better chance of successful location search. (upper left subwindow: the previous query image, upper right arrows: the view angles of previous query, current and suggested viewing angles; the main window: the current camera view). User follows the suggestion by trying to align the green arrow with the target arrow (red).

3. SYSTEM IMPLEMENTATIONS

Our current prototype includes a search server and client applications using both the actual mobile phones (either iPhone or Android) and desktop simulations. The demos on mobile phones may use pre-loaded street images or live street image captures as queries. The mobile applications communicate with the server through Wi-Fi or cellular services. The desktop simulation uses online street map services on which users can roam to any location in Manhattan and crop any view from the interface to simulate mobile search. Our search system implemented a codebook of one-million visual words using hierarchical k-means clustering [2] of SIFT descriptors, with inverted indexing to facilitate real-time search over 0.3 million reference images. Based on the implementation on a regular workstation, each query takes about 1.2 seconds.

Our AQS system implemented several unique ideas [5] - (1) an offline process for analyzing the saliency of each view associated with each location in the database (2) an online process to estimate the likely view of the current query, and determine the optimal view change in order to disambiguate the uncertainty of the top confusing locations. Details can be found in [5].

4. REFERENCES

- [1] snaptell.com, pointandfind.nokia.com, kooaba.com.
- [2] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *CVPR*, 2006.
- [3] B. Kaneva, J. Sivic, A. Torralba, S. Avidan, and W. Freeman. Matching and predicting street level images. In *Workshop on Vision for Cognitive Tasks, ECCV*, 2010.
- [4] B. Girod et al. Mobile visual search. In *IEEE Signal Processing Magazine*, July, 2011.
- [5] F. Yu, R. Ji, and S-F. Chang. Active query sensing for mobile location search. In *ACM Multimedia*, 2011.