IDeixis - Image-based Deixis for Finding Location-Based Information

Konrad Tollmar, Tom Yeh and Trevor Darrell

MIT CSAIL, 200 Technology Square, Cambridge, MA 02139
{konrad, tomh, trevor}@csail.mit.edu

Abstract. In this paper we describe an image-based approach to finding location-based information from camera-equipped mobile devices. We introduce a point-by-photograph paradigm, where users can specify a location simply by taking pictures. Our technique uses content-based image retrieval methods to search the web or other databases for matching images and their source pages to find relevant location-based information. In contrast to conventional approaches to location detection, our method can refer to distant locations and does not require any physical infrastructure beyond mobile internet service. We have developed a prototype on a camera phone and conducted user studies to demonstrate the efficacy of our approach compared to a map-based alternative.

1 Introduction

Location-based information services offer many promising applications for mobile computing. While there are many technologies for determining the precise location of a device, users are often interested in places that are not at their exact physical location. There are no common or convenient means to make a pointing (deictic) gesture at a distant location with existing mobile computing interface or location-based computing infrastructure.

In this paper, we present Image-based Deixis (IDeixis), an image-based approach to specifying queries for finding location-based information. This is inspired by the growing popularity of camera-phones and leverages the fact that taking a picture is a natural and intuitive gesture for recording a location. The key idea is that with a camera phone, users can point at things by taking images, send images wirelessly to a remote server, and retrieve useful information by matching the images to a multipurpose database such as the World Wide Web. Here we describe a scenario to illustrate an instance of the general idea:

"Mary is visiting campus for the first time ever. She is supposed to meet a friend at "Killian Court". She is uncertain if this place is the "Killian Court". She takes an image of the building in front her and sends it to the server. This image is then used to search the web for pages that also contain images of this building. The server returns the 5 most relevant web pages. By browsing these pages, she identifies the names 'Killian Court' and 'The Great Dome' and concludes that this is the right place."

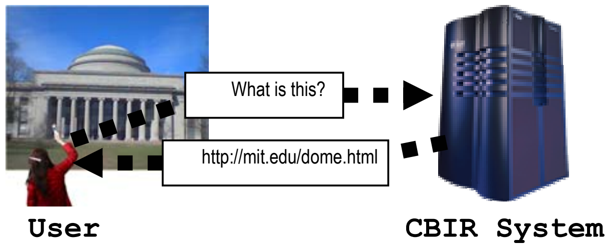

Fig. 1. Mobile Image-based Deixis.

Our system consists of two major components: a client-side application running on a mobile device, responsible for acquiring query images and displaying search results, and a server-side search engine, equipped with a content-based image retrieval (CBIR) module to match images from the mobile device to pages in a generic database (Figure 1).

We first review related work in the literature, and describe the findings of an interview study on user needs for location-based information services. We then describe a prototype constructed to demonstrate the technical feasibility of the concept of the image-based deixis. Finally, we report the results of a user study using a second prototype designed to test and compare different interface approaches.

2 Related Work

Image-based deixis touches four related areas: augmented reality with camera equipped mobile devices, content-based image retrieval, location recognition in robotics and wearable computing, and location-based information retrieval.

2.1 Camera-equipped Mobile Devices

Camera-equipped mobile devices are becoming commonplace and have been used for a variety of exploratory applications. The German AP-PDA project built an augmented reality system on camera-equipped iPAQ to assist electricians in appliance repair [6]. Images of the problematic appliance are taken, sent to a remote server for geometry-based model recognition, and finally augmented with useful information pertinent to that particular model. At HP, mobile OCR applications were developed using a pen-size camera to capture images of text for archiving storage or machine translation [16]. FaceIT ARGUS is a commercial system that provides remote face recognition to law enforcement agency on mobile devices [5]. Mobile image matching and retrieval has been used by insurance and trading firms for remote item appraisal and verification with a central database [3]. These systems are successful cases of information retrieval made possible by camera-equipped mobile devices, but they require specific models (e.g. for appliances) and are unable to perform generic matching of new images.

2.2 Content-based Image Retrieval

Content-based image retrieval systems can perform generic image matching and have been developed in the past decade for the application of multimedia data mining and archival search. One of the first such systems was IBM's Query-By-Image-Content (QBIC) system [13]. It supported search by example images, user-drawn pictures, or selected color and texture patterns, and was applied mainly to custom, special purpose image databases. In contrast, the Webseek system [19] searched generically on the World Wide Web for images and incorporated both keyword-based and content-based techniques; the keyword search returned a set of images among which users can further search by color histogram similarity with relevance feedback. The Diogenes system used a similar dual-modal approach for searching images of faces on the web [2]. A face detector operated as a screening process to filter out non-face images from the set of initial images returned by a keyword-based query. Then, a face recognition module processed the remaining images to identify particular faces. To this day, these systems have not been applied to the task of recognizing locations from mobile imagery.

2.3 Location Recognition

The notion of recognizing location from mobile imagery has a long history in the robotics community, where navigation based on pre-established visual landmarks is a well-known technique. The task of simultaneously localizing robot position and mapping the environment (SLAM) has received considerable attention [12,10]. Similar tasks have been addressed in the wearable computing community, with the goal of determining the environment a user is walking through while carrying a body-mounted camera [22].

Closely related to our work is the wearable-museum guiding system built by Schiele and colleagues that utilizes a head-mounted camera to record and analyze the visitor's visual environment [17]. Computer vision techniques based on oriented edge histograms were applied to recognize objects in the field of view. Based on the objects seen, the system then estimated the location in the museum and displays relevant information. The focus of this system was on remembering prior knowledge of locations – which item is exhibited where – rather than finding new information about locations. Murphy et al. generalized this type of technique and combined it with a probabilistic model of temporal dynamics, so that information from a video sequence could be used for the SLAM task with a wearable device [23]. In these robotics and wearable computing systems, recognition was only possible in places where the system has physically been before.

2.4 Location-based Information Retrieval

There are already location information services offering on-line maps (e.g. www.mapquest.com), traffic reports, and marketing opportunities on mobile devices. An often-discussed commercial application of location-based information is proximity-based coupon delivery. In a typical scenario, a merchant is notified when a potential customer visits a nearby retail outlet, upon which the customer can be delivered a coupon or offered a special promotional deal. The comMotion project [11] extended this idea to allow user-side subscription-based and location-specific content delivery. The GUIDE system was designed to provide city visitors location specific information customized to their personal needs [4]. The identity of the location is provided by the underlying cell-based communication infrastructure, which is also responsible for broadcasting relevant tourist information.

In [4], Cheverst concluded that the ability to filter information based on location made such an electronic tour guide a very useful tool. However, constraining the scope of available information to a specific location must be done carefully. For example, some users reported frustration when they were interested in something visible in the distance but were unable to find anything about it in the tour guide because the available information was restricted to their current physical location.

In [9], Kaasinen examined the usage of location-based information through interviews. He found that even if new technologies, such as GPS, could somewhat reliably provide users with their current location, the real need for location-based information often went beyond merely an answer to "Where am I?". For instance, a GPS-enabled mobile Yellow Page might tell users a list of nearby restaurants, but users realized they would want to know more, such as menus and reviews. This study highlighted the need for more comprehensive services in terms of geographic coverage, variety (number of services offered) and depth (amount of information available).

3 IDeixis - Image-based Deixis

IDeixis is intended to be a pointing interface paradigm and location-based computing technique which combines the ubiquity of a new generation of camera-phones with CBIR and the world wide web. There are two main components to a location-based computing system: the specification or sensing of location, and querying a database for location-relevant content. We will discuss the image-based approach to each in turn.

In our IDeixis system users specify a particular location by pointing to it with a camera and taking images. The location can be very close, or it can be distant – it just must be visible. IDeixis allows users to stay where they are and point at a remote place in sight simply by taking photographs. IDeixis does not require any dedicated hardware infrastructure, such as visual or radio-frequency barcode tags, infrared beacons, or other transponders. No separate networking infrastructure is necessary besides what is already made available by existing wireless service carriers, for example, General Packet Radio Service (GPRS) and Multimedia Messaging Service (MMS).

Having specified a location, a location-based information service needs to search for geographically relevant messages or database records. While geographic cues may well become a standard metadata tag on the web, they are not at present commonly available. However one form of location cue is already ubiquitous throughout the internet—images of actual places. If we can develop a method to match images from mobile cameras to images on the internet, we can gather web pages based on the geographic location of the camera-equipped device.

We believe the wealth of information already contained in the web can be exploited for location-based information services. Indeed, keyword-based search engines (e.g. Google) have established themselves as the standard tool for this purpose when working in known environments. However, formulating the right set of keywords can be frustrating in certain situations [8]. For instance, when the user visits a never-been-before place or is presented with a never-seen-before object, the obvious keyword, name, is unknown and cannot be used as the query. Image-based deixis could be desirable in this situation: the intent to inquire upon something is often inspired by one's very encounter of it - the place or object in question is conveniently situated right there. With a camera phone, an image-based query can be formed simply by pointing with the camera and snapping a photo.

Web resources exhibit a high correlation between semantic relevancy and spatial proximity, an observation that has been noted and widely exploited by existing search technologies. Pieces of knowledge close together in cyberspace tend to be also mutually relevant in meaning [1]. An intuitive reason is that web developers tend to include both text and images in authoring pages. In practice, current web-image search engines, such as Google, use keywords to find relevant images by analyzing neighboring textual information such as caption, URL and title [7]. In our work, this process is reversed. An image is used to find matching images of the same location. In many situations, finding these images on the web can lead us to the discovery of useful information for a particular place in textual form.

4 FIRST PROTOTYPE

A pilot prototype was designed to find out whether searching for location-based information on the World Wide Web by matching images from a mobile device is practical and useful. We tested whether current CBIR algorithms can be effectively applied to images of locations and how feasible it was to implement a real system with existing camera-equipped mobile devices and wireless network infrastructure.

4.1 System Design

We built our first prototype on a Nokia 3650 phone taking advantage of its built-in camera (640×480 resolution) and the support for Multimedia Messaging Service (MMS), using C++ on Symbian OS [21]. To initiate a query, the user points the

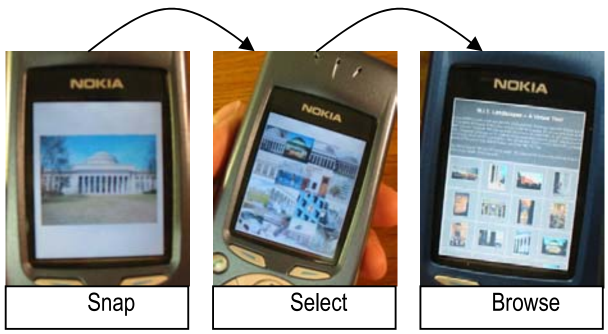

Fig. 2. A walkthrough of the first prototype. User snaps an image to send as a query (left). The result is displayed as a thumbnail mosaic (center). Selecting a thumbnail image brings up a source webpage for browsing (right).

camera at the target location and takes an image of that location, which is sent to a server via MMS.

We designed our system with an interactive browsing framework, to match users' expectations based on existing web search systems. For each query image, the search result will contain the 16 most relevant candidate images for the location indicated by the query image. Selecting a candidate image brings up the associated web page. The user can browse this page to see if there is any useful information (see Figure 2).

As reviewed above, the idea of searching for images based on their content has been a research focus in computer vision for many years. However, the performance of general CBIR systems is far from perfect—the problem often lies in the classic vision problems of image segmentation and viewpoint variation. Fortunately finding images of location does not suffer from these limitations since it can be performed with analysis over the entire image, making general image segmentation unnecessary. Users ask about a location most likely because they are physically there and there are a much smaller number of physically common viewpoints of prominent landmarks than in the entire viewsphere of a common object. If an image-based location query is ambiguous, the interactive paradigm allows a user to browse from multiple matches (as in common web searches) and/or to snap a few more pictures from a different viewpoint. These observations led us to believe that it would be possible for existing CBIR to achieve good performance when used to recognize images of location.

4.2 Experiment and Results

The first prototype was built to evaluate whether CBIR can match location images from mobile devices to the pages on the world wide web. For our initial experiments, we restricted ourselves to a known domain, a single university campus, both for web searching and when initiating mobile queries. We began by constructing an image database consisting of 12,000 web images collected from the mit.edu domain by a web crawler. Test query images were obtained by asking student volunteers to take a total of 50 images from each of three selected locations: Great Dome, Green Building and Simmons Hall. Images were collected on different days and with somewhat different weather conditions (sunny/cloudy); users were not instructed to use any particular viewpoint when capturing the images.

We tested the image matching performance of two simple CBIR algorithms: windowed color histogram [20] and windowed Fourier transform [15]. Principal component analysis was used for finding the closest image in terms of Euclidean distance in the feature space. These are among the simplest CBIR methods in the literature, and partial success with these methods would suggest even greater performance with more advanced techniques.

These volunteers were used as test subjects to evaluate the performance of these algorithms. For each test query, the search result consisted of the 16 closest images in the database to the query image using the selected distance measure (color histogram or Fourier transform). We asked the test subjects to decide how many of these candidate images are actually similar to the original query images. Figure 3 summarizes the performance we obtained from the tested image matching metrics. It shows the percentage of tries our test subjects found at least one similar image on among the first 16 candidate images.

Unfortunately, the underlying MMS-based communication infrastructure had in practice an average turnaround time of 25sec, which required much patience of the user. We expect as MMS becomes more popular, with better technology implemented by wireless service carriers, the turnaround time can be drastically improved. We are also currently exploring image compression methods and better network protocols to improve interactive performance.

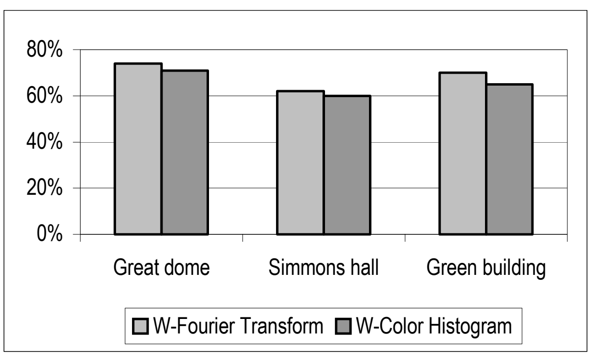

Fig. 3. Average retrieval success for each CBIR algorithm shown in percentages of 150 test attempts that returned at least some relevant images on the first page of candidate images.

4.3 Discussion

Anecdotal analysis of the example webpages in Figure 4 shows that we can in principle find web pages with images that perfectly match the image of the current

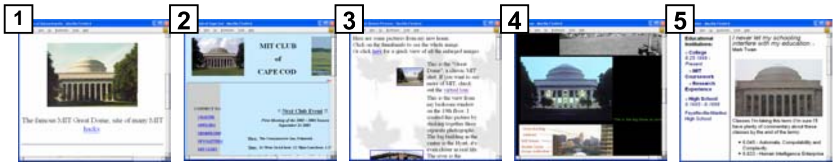

Fig. 3. Some examples of found webpages: (1) MIT Gallery of Hacks, (2) MIT Club of Cape Cod's official website, (3) Someone's picture gallery, (4) Someone's guide on Boston, and (4) Someone's personal homepage.

location, yet the web page that contains that image may be poorly related to the location – for example, a photo gallery of someone who visited a museum rather than the official page of that museum. With an interactive paradigm, users can keep searching for a more appropriate match but this may take considerable time. Additionally, we observed that people often searched for more specific questions beyond "what is that". They needed to go back and forth between the thumbnail mosaic and the web browser to examine many web pages for a specific piece of information. To do so on the small-screen device can be too cumbersome to be feasible.

5 INTERVIEW STUDY

In order to better understand the browsing after location-based information using mobile devices we conducted a interviews study before we designed the second prototype. The interview study involved 20 subjects and took place in common tourist locations. First we went through a list of questions about location-based information. Then, we accompanied subjects as they walked around and encouraged them to talk out loud about what they were seeing and describe the kind of information the would appreciate in that context. The data collected allowed us to address the following two questions:

- How do people currently use maps and tour books while visiting an unfamiliar location?

- What do people want to know about their specific location, and how do they want to obtain the information?

Regarding the first question, we were told that maps and tour books often lack detailed information. Very few subjects reported bringing them in everyday life. But virtually everyone reported carrying a map when traveling to a new place. One interesting finding was the tendency of people to overstate the usefulness of a street map. On second thought, they would retract such statement as they quickly realized they actually wanted to know more than what a map could provide, such as specific details about buildings and artifacts they were seeing around them.

Based on our walking interview, we found that there are many specific questions asked only by certain individuals, like "what kind of bike is this", "what is the name of this tree", and "when does the city empty these garbage bins." However, there were also general questions shared by most of the subjects. These questions range from historic information and events to names of buildings and makers of public artworks. We identified the two most commonly asked questions were "Where can I find X?" and "What is this?" Often times, these questions were followed by requests for time-related information such as business hours and bus schedules.

Most subjects claimed that they would rather ask someone nearby, be it a fellow visitor, a street vendor, or a policeman, than deferring the question for later, such as looking it up on-line at home. The bottom line was they need the information "right here" and "right now", or it would not worth their effort. Even when they had a mobile phone, they found it unlikely to call up someone to ask for information. The only exception was when they were supposed to meet someone and they needed to get the directions to the meeting location.

From these interviews, we concluded that location-based information services which provided access to a generic information service such as the world wide web, and which were initiated by a real-time query (e.g., "What is this place?") followed by a browsing step, would complement the users' experience in an unfamiliar setting and meet their needs for a location-based information service. But we also observed that in some case these webpages was not sufficient. To overcome this problem, we developed a new "keyword bootstrapping" approach in which image-based keywords are used for interactive searching. The first steps of the process are as before, with a user taking a picture of a location, and image search returning a set of matching images and associated web pages. Salient keywords are automatically extracted from the image-matched web pages. These keywords can then be submitted to a traditional keyword-based web search. In the museum example, the museum name (e.g. "Louvre") would presumably appear as a keyword on many of the matched pages, and a Google keyword search with "Louvre" would very likely yield the museum's official homepage. With this approach the relevant homepage can be found even when it contains no image of the location itself.

6 Second protype

Our second prototype explored the image-based keyword bootstrapping paradigm and compared it to the thumbnail browsing approach of the first prototype. We also compared performance to a baseline map-based method where users can click on a desired location on a small map to find information. We used the same hardware as the first prototype but used a web interface developed in XHTML Mobile Profile with JavaScript extension [14]. We implemented three search strategies:

- **Searching for web pages:** the search result consists of a list of matched web pages containing similar images of the query image. Each page is displayed as a thumbnail accompanied by a text abstract of its content. Selecting a thumbnail brings up the full content of the page on the screen (see Figure 5).
- **Searching for keywords:** Automatically extracted keywords are displayed side-by-side with the thumbnail image. Selecting a keyword initiates a keyword-based search on Google to find more information (see Figure 5).

- **Searching by mobile MapQuest:** We used a GPS-coordinate-based query to retrieve from MapQuest a map covering the surrounding area. This is meant to serve as the baseline for evaluating image-based approaches.

To compare the usability of these three approaches without the issue of network latency, we simulated some functions and pre-cached portions of the web interface and result pages for known landmarks. In this part of the study, we focused on the relative ease of use of the three interfaces rather than evaluating the matching success of the CBIR algorithms.

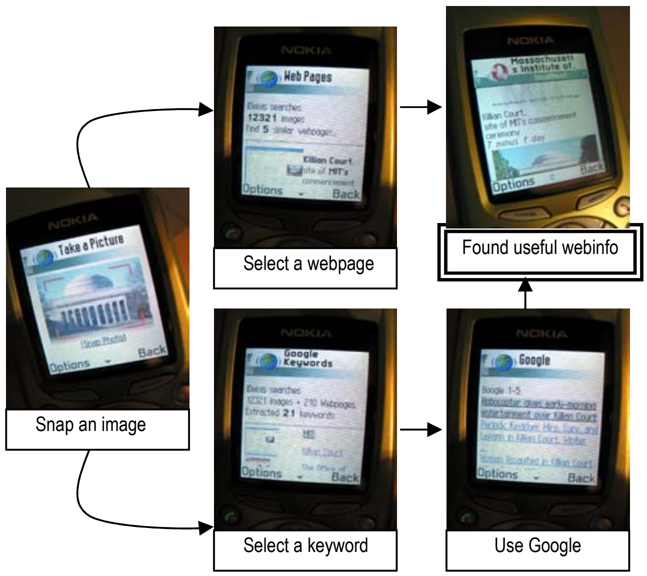

Fig. 5. Two image-based search strategies for location-based information. First strategy (Top) displays the results directly on the screen in thumbnails. Second strategy (Bottom) provides a set of keywords for a user to select and submit to a text-based search engine.

6.1 Method

We conducted the testing of our prototype at two different outdoor sites under fairly good weather conditions with 16 subjects aged between 13 and 63 evenly split between genders. A survey on their background of the technology revealed that all of them use the Internet regularly. Most of them (14 out of 16) owned a cell-phone; 11 used it as their primary phone. Some (6 out of 16) owned a digital camera; 4 used it frequently. The testing of the prototype was conducted in two steps:

1. We handed the subject a Nokia 3650 camera phone and asked her to walk around taking pictures of any particular landmark on the site she would like to know more about.

2. We let the subject use and evaluate the prototype system on the phone to perform the task of searching location-based information, using each of the three search strategies.

6.2 Results

All our subjects found it very intuitive to express their interests in a particular place by taking pictures in an attempt to look for location-based information. This gives us substantial, if not conclusive, evidence to claim that IDeixis is indeed a simple and intuitive way to specify the locations of interest.

All the testing sessions were recorded on video. We later analyzed the video sequence manually and extracted from it the time it took to locate a piece of information at each search attempt as well as the subjective evaluation of the quality of information found by the subjects (see Table 2). We found that the second strategy did a better job in finding specific information. The third approach failed to provide useful information more than half of the time.

Although all of our subjects were familiar with the Web, none of them had any experience in surfing the Web on a mobile device. Nonetheless, basic web browsing on the device was not a problem. However, the small screen and the low resolution created many design tradeoffs. For instance, the mobile MapQuest was rejected by most subjects as too small and basically unreadable, as one subject commented: “I have no idea what this small map is covering”.

In the second prototype, the search result is presented in both thumbnail images and text. Most subjects preferred text to images and explained that it was mainly attributed to the limited screen size. Although most subjects complained that the web page thumbnails were too small, they could not deny the fact that such visual aid was somewhat useful, as one subject commented on the text/thumbnail tradeoff: ”I prefer text, but if I see a terrible website [on the thumbnail] I don’t need to go there”.

On searching for web pages: We observed among subjects an interesting disparity in their perception of how specific search results were. One subject commented upon this by reflecting that “similar web pages are very different in content”; he thought the search result was too specific. However, another subject complained that the information the top ranked pages provided was too general.

On searching for keywords: Several subjects reported difficulties in understanding the searching for keywords strategy, as complained by a subject: “the keywords lead me out in tangents”. It might be an interface issue since both searching for similar web pages and searching for keywords strategies used a similar graphical design. Despite this, we also saw evidence of the effectiveness of the searching for keyword strategy. Ten subjects found some information that they were looking for in less than 5 steps (i.e. take a picture → search for keywords → select some keywords for Google → choose a webpage from the result returned by Google → browse and evaluate the information available on the web page). One subject suggested that he would “use keyword for fast searches and similar web pages for general information.”

On searching by mobile MapQuest: Almost all subjects expressed the opinion that at first they thought such a map-based interface could be very helpful. But after using the map interface and the image-based approach, several commented that a map-based strategy might not be as helpful as it first seemed; in our tests users with the map-based interface often failed to find specific information about a location.

Table 1. Quantitative comparison of search strategies. Each row shows, under each search strategy, the number of attempts where the user found relevant information within 30 sec, or spent more than 30 sec, or did not find anything useful at all.

Search Strategy	<30 Sec	>30 Sec	Did not find
Web Pages	6	4	22
Extracted Keywords	10	12	10
MapQuest	4	10	18

7. Future Work

While we have demonstrated the feasibility of several CBIR techniques for our IDeixis system, it remains a topic of ongoing research to find the optimal algorithm to support searching the web with images acquired by camera-equipped mobile devices. We are exploring machine learning techniques which can adaptively find the best distance metric in a particular landmark domain. Additional study is needed to evaluate the performance of mobile CBIR with a range of different camera systems, and under a wider range of weather conditions (e.g., we did not consider snow in our studies to date.)

Currently the cameras on our prototypes have a fixed focal length and there is no simple way to specify a very distant location that occupies a small field of view in the viewfinder. We plan to augment our interface with a digital zoom function to overcome this, or implement a bounding box selection tool and/or use camera phones with adjustable zoom lenses when they become available.

As discussed earlier, even with image-based search of location-based information, additional context will be needed for some specific searches. Keyboard entry of additional keywords is the most obvious option, but with equally obvious drawbacks. Allowing users to configure various search preferences can be another option. An appealing interface combination would be to consider keywords obtained via speech input at the same time an image-based deixis was being performed (e.g. "Show me a directory of this building!").

In addition, it is fairly likely that mobile devices in the near future will incorporate both camera technology as well as GPS technology, and that geographic information of some form may become a common meta-data tag on certain web homepages. Given these, we could easily deploy a hybrid system which would restrict image matching to pages that refer to a limited geographic region, dramatically reducing the search complexity of the image matching stage and presumably improving performance.

Finally, further study is much needed to evaluate the usability of the overall method in various contexts beyond tourist and travel information browsing, including how to best present browsing interfaces for particular search tasks.

8 Conclusion

We have proposed an image-based paradigm for location-aware computing in which users select a desired place by taking a photograph of the location. Relevant web pages containing matching images are found using content-based image retrieval with the web or other large database. In contrast to conventional approaches to location detection, our method can refer to distant locations and does not require any physical infrastructure beyond mobile internet service.

We explored two interface paradigms, one based on thumbnail browsing of directly matched pages and one based on bootstrapped keyword search. In the latter approach keywords are extracted from matched images and used for a second-stage search, so that even if the best page for a particular location contains no image of the location, it still can be found using our system. We evaluated our prototype systems and found that finding relevant location-based information in this way required minimal interaction.

REFERENCES

1. Arasu A, Searching the Web. *ACM Transactions on Internet Technology*, 1(1), Aug. 2001, Pages 2–43.
2. Aslandogan, Y.A., and Yu, C. Multiple evidence combination in image retrieval: Diogenes searches for people on the web. In *Proc. of ACM SIGIR 2000*, 88-95.
3. Built-in camera mobile phones increasingly used for business purposes. *Nikkei Net Business*, Dec. 9 2002.
4. Cheverst, K., Davies, N., Mitchell, K., and Efstratiou, C. Developing a context-aware electronic tourist guide: some issues and experiences. *Proc. of CHI 2000*, 17-24.
5. FaceIT, www.visionics.com.
6. Gausemeier I., and Bruederlin B. Development of a real time image based object recognition method for mobile AR-devices. In *Proc. of the 2nd AFRIGRAPH*, 2003, 133-139.
7. Google, Google's PigeonRank, http://www.google.com/technology/pigeonrank.html
8. Jansen, B. J., Spink, A., Bateman, J., and Saracevic, T., Real life information retrieval: A study of user queries on the Web. *ACM SIGIR Forum 32*, 1 (1998), 5-17.

9. Kaasinen E., User needs for location-aware mobile services, *Personal and Ubiquitous Computing* (2003) 7: 70–79.

10. Kosecka, J., Zhou, L., Barber, P. and Duric, Z. Qualitative Image-Based Localization in Indoors Environments, *CVPR 2003*, (To Appear).

11. Marmasse N. and Schmandt C., Location-aware information delivery with commotion. *HUC 2000 Proceedings*, 157-171.

12. Montemerlo, M., Thrun, S., Koller, D. and Wegbreit, B. FastSLAM: A factored solution to the simultaneous localization and mapping problem, In *Proc. of the AAAI National Conference on Artificial Intelligence*, 2002.

13. Niblack W., R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos and G. Taubin. The QBIC project, *SPIE Storage and Retrieval for Image and Video Databases*, 1993, 173-187.

14. Nokia, Next Generation Mobile Browsing, http://www.nokia.com/nokia/0,,32912,00.html

15. Oliva, A., and Torralba, A. Modeling the shape of the scene: a holistic representation. *Intern. Journal of Computer Vision*, 42(3), 147-175.

16. Pilu, M. and Pollard, S. A light-weight text image processing method for handheld embedded cameras. *Tech. Report*, HP Laboratories, March 2002.

17. Schiele, B., Jebara, T., and Oliver, N. Sensory-augmented computing: wearing the museum's guide. *Micro. IEEE*, 21(3), 2001, 44-52.

18. Smith J. and Chang S., An Image and Video Search Engine for the World-Wide Web, *Symposium on Electronic Imaging*, IS&T/SPIE, Feb. 1997.

19. Smith, J.R., and Chang, S.F. Searching for images and videos on the WWW, *Multimedia Systems*, 1995, 3-14.

20. Swain, M., & Ballard D., Indexing via color histograms. *Intern. Journal of Computer Vision*, 1990, 390-393.

21. Symbian OS – The Mobile Operating System, http://www.symbian.com/.

22. Starner, T., Schiele, B. and Pentland, A. Visual Contextual Awareness in Wearable Computing. In *Proc. of ISWC*, 1998, 50-57.

23. Torralba, A., Murphy, K.P., Freeman, W.T., and Rubin, M.A. Context-based Vision System for Place and Object Recognition. In *Proc. of ICCV 2003*, 273-280.