An Overview of Problems in Image-Based Location Awareness and Navigation

Yung-Hsiang Lu and Edward J. Delp
School of Electrical and Computer Engineering, Purdue University
West Lafayette, Indiana 47907
yunglu@purdue.edu and ace@ecn.purdue.edu

ABSTRACT

In this paper we describe some of the research issues and challenges in image-based location awareness and navigation. We will describe two systems being developed at Purdue University as testbeds for our ideas. The main system architecture combines image processing, mobility, wireless communication, and location awareness. We will describe two fundamental scenarios for using images to aid in mobile navigation problems. The first provides the ability to use a locally acquired image to determine the identity of an object, for example a building, as one roams in an area. The second problem is the use of images in a database to aid in vehicle navigation. The solution to these problems use location information, such as GPS signals, to compare and search location-annotated images in a database. We believe location information can improve the accuracy in image database search.

1. INTRODUCTION

Widely deployed wireless networks have made it possible to transmit images and video sequences to and from mobile devices. Today, users can capture, display, and transmit images or video sequences using personal digital assistants (PDAs) or mobile telephones. Even though it is possible to transmit images via wireless networks, it is unclear what applications can benefit from this technology. Sending static images does not explore the full potential of the technology. Seeing a movie on a 320×240 screen is not appealing to many users. Location-based advertisements or coupons can be sent to users; however, these promotions may be considered as spam. The real values of the technology remain to be exploited for bringing new services to mobile users.

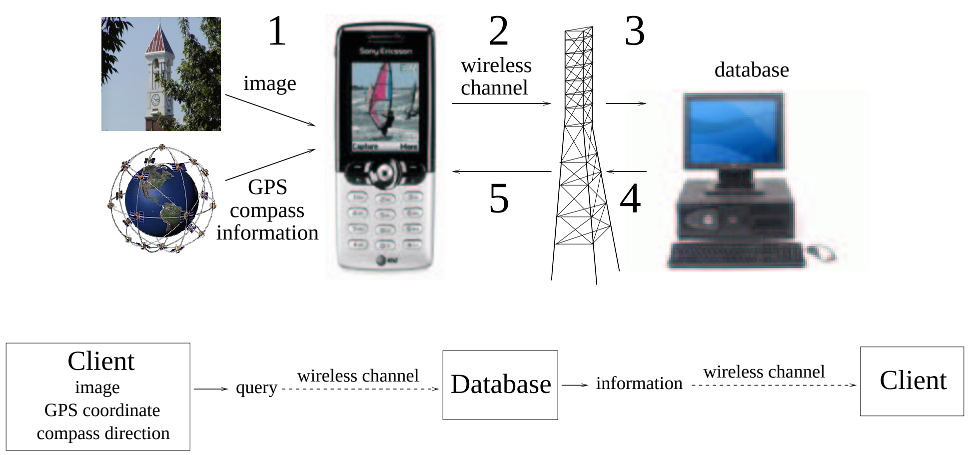

Figure 1. Location-aware image database system (LAID).

Address all correspondence to Yung-Hsiang Lu, yunglu@purdue.edu.

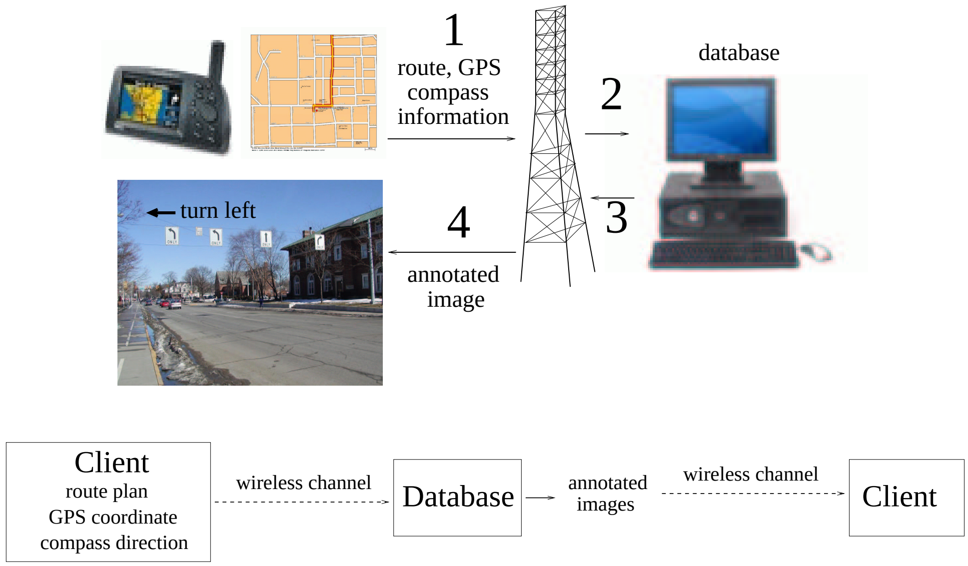

Figure 2. Differences between map-based and image-based navigation (IEN).

We believe that one area where a mobile user might benefit by having a network-connected image sensor and display is navigation. In this paper we will describe some of the research issues and challenges in image-based location awareness and navigation. We will describe two fundamental scenarios for using images to aid in mobile navigation problems.

The first is the ability to use locally acquired images to determine the identity or function of an object that one observes as one roams in an area. For example, a user is walking through a university campus and comes to a building that is unknown to the user. The user takes a picture of the building with a handheld device. This image will be sent via the wireless network to a "location aware" image database that will index the image against its content and send back information to the user about the building, such as the name, use, and type of architecture. We shall refer to this scenario as using a Location-Aware Image Database or LAID. Throughout the rest of this paper this scenario will be known as the *LAID* system and is shown in Figure 1.

The basic building blocks of the *LAID* system is a handheld client that has a camera, a GPS (global positioning system) receiver, a compass, and network connection. The user connects via the network to a server that has location-aware image database and sends the image of the object. The server extracts content-dependent features from the image and uses this along with the location and orientation information to find the best match object in the database. This information along with the matched image (possibly a thumbnail image) is returned via the network to the user.

The second scenario is using images in an database to aid in vehicle navigation. The ultimate goal here is to provide the driver with an image of a recognizable object (for example, a building or a landmark) that can help the driver determine where the driver is located. The main principle here is that people are better at recognizing objects to determine their locations than reading maps particularly when they believe they are lost and may be disoriented. We refer to this as Image Enhanced Navigation or IEN. Throughout the rest of this paper this scenario will be known as the *IEN* system and is shown in Figure 2. Images used in the *IEN* system are annotated with navigation instructions. The driver only needs to take a glance of the image to confirm the current direction. Voice-assisted GPS and the visual direction ensure that the driver follow the route correctly.

The *IEN* system consists of a GPS receiver, a map database, a route planner, a display, and wireless network connection that is located in the vehicle. The components of the *IEN* system that are not contained in currently available on-board navigation systems are the image database and the wireless connectivity. Before a road trip, a user plans the route and downloads the associated images along the route. These images, in addition to maps, help the user navigate to the destination. While the user is moving, the system automatically updates the images periodically. Each image has information about its location, orientation, time of the day, and the weather. These features provide direct and intuitive visual information. The system has a "panic" feature that allows the user who may be lost to quickly request an image from the system to orient themselves.

In the rest of this paper we will present a review of previously related work in this area and describe some of the issues that need to be addressed in designing and deploying these types of systems.

2. PREVIOUS WORK

The major operation in *LAID* is using the image (and location information) acquired by the user to index an image database. Below we provide an non-exhaustive review of image databases. In the *IEN* system, we need to use images to enhance navigation and below we review recent work in navigation.

2.1. Image Databases

Image databases use content, instead of keywords, to retrieve information. While keywords can be assigned to each image, these keywords are often subjective and cannot capture all the information presented in an image. Consequently "query-by-example"[1] is widely used for image databases. One could argue that main component of the *LAID* system is a "query-by-example" image database search. In the next section, we will describe how our approach differs.

Image databases have been used in various applications, such as oil exploration, remote sensing, and medicine[2].[3] One major challenge is to find images with similar objects even though the background lighting or sensor angles are slightly different. Many techniques have been adopted to improve the accuracy of search, including content similarity,[4] user feedback[5],[6] users' sketches,[7] semantics,[8] color patterns,[9] and perception.[10] Chen et al.[11] built image pyramids in which similar images were stored in the same or adjacent levels of the pyramids. Cha et al.[12] used high dimensional grid-cell trees to encode the similarities among images. Some recent research is devoted to using portable cameras for building image databases[13].[14] Derthick[15] built a user interface on a palmtop computer for image searching. Wolf et al.[16] used smart cameras that could recognize gestures. Bennani[17] proposed building a personal image database using a home network and computers. Jeon et al.[18] presented a probabilistic approach to automatically annotate images.

2.2. Navigation System

Several studies demonstrated that maps are not the best way for route instruction and navigation. Ekman et al.[19] described the benefit of contextual information in small displays of mobile systems. Satalich[20] surveyed many user studies in navigation and wayfinding and reached the following conclusions: (a) Reading maps alone was insufficient. Additional information could help users recognize location sequences, landmarks, and routes. (b) Even with customized maps, users could still make many errors in wayfinding. (c) "You Are Here" maps in buildings did not significantly reduce the time to find rooms. Kray et al.[21] compared four different ways to present route instructions: textual, abstract direction (arrows), 2-D maps, and computer-rendered 3-D maps. They found the most commonly used strategy was matching buildings in the 3-D maps. User feedback suggested improving 3-D maps and making them more realistic. Maps are popular for navigation because they are better than text[21] and do not require substantially more storage or rendering capabilities. Vainio et al.[22] reported that users prefer photo-realistic 3-D models so that landmarks could be recognized more easily even when a user was moving. Gronbak et al.[23] built spatial hypermedia to represent buildings for architecture designs. Baus et al.[24] discussed designing navigation systems for users that combined several means of transportations to reach their destinations. Rogers et al.[25] used GPS information to provide safety advices such as lane keeping. Chan et al.[26] built a visualization system for 3-D geographical data. Jordan[27] predicted that future "intelligent maps" would include realistic images.

Our application in the *IEN* system is completely different from the "image-based navigation" for virtual reality[28] or robot control.[29] First, the images are obtained from real objects, not computer generated. Second, tours are planned based on road connections. Images are used to assist drivers; and drivers are responsible for avoiding collision. Third, the images are presented to drivers and not used for recognition by computers.

3. TECHNOLOGICAL CHALLENGES

3.1. Resource Requirements

We are currently deploying two testbeds at Purdue University to test and implement the *LAID* and *IEN* systems.

We use the following initial system requirements for *LAID*:

1. **Handheld device:** The device has 64MB of memory, a 300 MHZ processor, a 802.11a network interface, and a 320 × 240 pixel color display. The device will also have enough slots to accept up 3 peripheral devices if they are not built-in.

2. **Peripheral devices:** We have a camera capable of taking 640 × 480 color images in JPEG format. We also have a GPS receiver that either has an accurate compass or we have a compass as a separate device.

3. **Image database:** The image database is populated with images of buildings or other objects that might be observed by the user and were taken at resolutions of 1024x1024 pixels and are stored in JPEG format. Each image has location information associated with it (latitude, longitude, compass direction, and perhaps elevation), text descriptions of the structure, content dependent image features, and a 320x240 thumbnail JPEG thumbnail image. In out initial system this database in built around MySQL.

4. **Storage and processing:** Each image taken on the handheld will require approximatively 80 KB of storage, the location information is require about 1 KB. This will need to be set over the network. The result of this query will be the thumbnail image and the description. This is will approximately 150 KB of data. Our goal is to have the query return to the user in no more than 10 seconds.

The system requirements for the *IEN* system include:

1. **Route and image storage:** Since a route is composed of the locations indicating turning points, a one-mile trip usually needs fewer than 10 KB storage. Suppose we want to display one image per mile. If the image resolution is 640 × 480, the size is approximately 80KB using JPEG compression. Thus, 100 KB is sufficient for one mile of route. A 500-mile road trip needs only 50MB storage and it can fit into a typical Compact Flash disk. If the trip is planned from the home, the data can be downloaded within one minute through a high speed home network (cable modem or DSL). If the trip is planned inside a vehicle, due to the limitation of wireless bandwidth, one image is downloaded every five miles.

2. **Update requirements:** If the images are updated every 30 seconds during the trip, 100 KB is transmitted every 30 seconds. The bandwidth requirement is 3.3 KB/s. If there are 100 cars in the same region (a network cell), 330 KB/s is needed for the total bandwidth. We use relay networks to reduce the bandwidth requirements among multiple vehicles. If the user is in a region where image updates are unavailable, for example, due to network congestion or weak signals, the images downloaded during route planning can still assist the user while the user is traveling.

3.2. Database Construction

The first challenge of the system is to populate the database with location-annotated images. A database is currently being constructed with images taken on the West Lafayette campus of Purdue University. When an image is taken, its location and orientation are recorded based on the GPS and compass readings. As the system scales up, an automatic method must be developed to construct the database efficiently. The city of West Lafayette has more than 100 miles of streets. In order to provide visual directions for navigation, four images are needed in each intersection. Suppose there are ten intersections for each mile of road, over 4000 images must be taken with their location information.

We are designing an automatic method to populate the database. The method contains four webcams mounted on the roof of a vehicle. The webcams face four different directions and they are connected to a laptop computer in the vehicle. The webcams take images periodically. If a camera takes an image every five seconds, 24 images will be taken for each mile when the vehicle travels at 30 miles per hour. The laptop computer is connected to a GPS receiver and a digital compass. When the cameras take images, the laptop computer also record the current location and direction. We need only the direction of the vehicle and can compute the cameras' directions easily.

This approach can quickly populate the image database with another benefit: By comparing the images in the sequence, we can determine the occlusion relationships among objects. We can calculate which objects should appear first when viewed at a particular location and direction. This information will improve the matching accuracy of LAID.

3.3. Image Comparison

Figure 3. Different viewing angles and background lighting.

A major challenge in *LAID* is that the images sent by clients may be taken with angles or background lighting different from the images stored in the database. Figure 3 shows three images taken at different times with different angles. While human beings usually have no problem recognizing these three images, it can be difficult for the database. Since each image contains the information of its location and direction, the search accuracy can be significantly improved. Moreover, the occlusion relationship obtained during image capturing can further assist enhancing the matching. This problem will have to be addresses in the way we extract features from the received image. Even though, in a general sense, we are doing "query-by-example" the location information will help us solve this problem.

3.4. Background Adjustment

The angle and lighting problem for *IEN* is the opposite. It is preferred to present the images with lighting and the angle matching what the user should be seeing. For example, if an image is taken at daytime and the user is driving at night, the images should be adjusting to match the night view. There are two solutions to this problem. First, we can use the sequence of images to separate background (sky) and objects because objects move faster in the sequence due to perspective projection on the camera. After separating the background from the objects, we can change the colors of the background to reflect the current time of a day.

The second solution is to capture multiple images at different time of a day or in different seasons. If *IEN* is widely deployed, image capture may be achieved through the collaboration with car rental companies. Because *IEN* helps drivers navigate more comfortably, it has the potential to reduce accidents because of nervousness due to feeling lost. Consequently, *IEN* may help rental companies reduce their cost due to accidents. Some rental companies consider using GPS to track their vehicles for better roadside assistance. These cars may also be equipped with digital cameras and capture images along the routes the customers travel. The privacy issue here are important but beyond the scope of this paper.

3.5. Counter Arguments

One may argue that our approach is too complicated and simpler approaches may exists. This section lists several counter arguments against our approach and the responses to these arguments.

- For *LAID*, is GPS and the compass information enough to determine what we are looking at? Why do we need images?
  - A user may be looking at an object that is far away and not exactly at the object in front of them, particularly when the object is sufficiently large (such as a building).
  - One example is seeing the Golden Gate Bridge from a tall building in the San Francisco Financial District. The current location (the building) is far away from the object being seen (the bridge). Due to the advancement of imaging technology, zooming can bring the object to the user easily. If we use only location and direction, we will be unable to determine the object of interest.
- For *LAID*, is the location of the wireless access point enough? Why do we need GPS or a compass?
  - Access points have no directions. Users' locations and orientations cannot be determined by the locations of the access points.
  - Even if we can use multiple access points to calculate the users' locations and direction of motion, we still have no information what object is interesting to the user, as described earlier.
- Can the use of RFID (radio frequency identification) tags solve problems of *LAID*? Why do we need images or GPS?
  - RFIDs are short range so users cannot receive the RFID signals from remote objects. It is also possible that a user receives several RFID signals from different objects. Images help determine what objects the users are really interested in.
  - It is more difficult putting an RFID in *every* object that may possibly be interesting to users. In contrast, creating an image database can be easier due to its centralized management.
- Some users prefer reading maps. Why do we need *IEN*?
  - Various studies have shown that maps are not intuitive to some people. *IEN* does not replace maps; *IEN* enhances maps.
  - Some people may feel comfortable reading maps for navigation. However, they use maps simply because there is no alternative now. If *IEN* becomes widely deployed, these users may determine that *IEN* is better than maps.
- For *IEN*, can wireless networks support the large amount of images transmitted? What happens during rush hour?

– The frequency of image update can be dynamically adjusted based on the network conditions, as long as users receive the turning instructions correctly.
– Because drivers on the same road and in proximity see the same scenes, the database can broadcast the images to adjacent vehicles. Therefore, the bandwidth requirements do not linearly increase as the number of vehicles increase.

4. CONCLUSION

This paper presents two systems that provide location-aware services using image databases. We describe the rationale of our approach and the testbeds being built at Purdue University. We list counter arguments the responses of our approach. Our system has the potential to greatly improve the services for mobile users.

REFERENCES

1. D. Zhong and S.-F. Chang, "Spatio-temporal Video Search Using The Object Based Video Representation," in *International Conference on Image Processing*, pp. 21–24, 1997.
2. V. Castelli and L. D. Bergman, eds., *Image Databases: Search and Retrieval of Digital Imagery*, Wiley, 2002.
3. R. C. Veltkamp, H. Burkhardt, and H.-P. Kriegel, *State of the Art in Content-Based Image and Video Retrieval*, Kluwer, July 2001.
4. Z. Yang and C. C. J. Kuo, "Learning Image Similarities and Categories From Content Analysis and Relevance Feedback," in *International Multimedia Conference*, pp. 1–1, 2001.
5. J. Martinez and E. Loisant, "Browsing Image Databases With Galois' Lattices," in *ACM Symposium on Applied Computing*, pp. 791–795, 2002.
6. S. Tong and E. Chang, "Support Vector Machine Active Learning for Image Retrieval," in *ACM International Conference on Multimedia*, pp. 107–118, 2001.
7. K. Matkovic, L, J. Siglaer, M. Kompast, and W. Purgathofer, "Visual Image Query," in *International Symposium on Smart Graphics*, pp. 116–123, 2002.
8. Y. Nah and P. C.-Y. Sheu, "Image Content Modeling for Neuroscience Databases," in *International Conference on Software Engineering and Knowledge Engineering*, pp. 91–98, 2002.
9. D. Stan and I. K. Sethi, "Color Patterns for Pictorial Content Description," in *ACM Symposium on Applied Computing*, pp. 693–698, 2002.
10. E. Chang, K.-T. Cheng, W.-C. Lai, C.-T. Wu, C. Chang, and Y.-L. Wu, "PBIR: Perception-based Image Retrieval-a System That Can Quickly Capture Subjective Image Query Concepts," in *ACM International Conference on Multimedia*, pp. 611–614, 2001.
11. J.-Y. Chen, C. Taskiran, E. J. Delp, and C. A. Bouman, "ViBE: A New Paradigm for Video Database Browsing and Search," in *IEEE Workshop on Content-Based Access of Image and Video Libraries*, pp. 96–100, 1998.
12. G.-H. Cha and C.-W. Chung, "The GC-tree: A High-dimensional Index Structure for Similarity Search in Image Databases," in *IEEE Transactions on Multimedia*, pp. 235–247, June 2002.
13. S. Mann, *Intelligent Image Processing*, Wiley, 2002.
14. M. Gelgon, "Using Face Detection for Browsing Personal Slow Video in A Small Terminal and Worn Camera Context," in *International Conference on Image Processing*, pp. 1062–1065, 2001.
15. M. Derthick, "Interfaces for Palmtop Image Search," in *ACM/IEEE-CS Joint Conference on Digital Libraries*, pp. 340–341, 2002.
16. W. Wolf, B. Ozer, and T. Lv, "Smart Cameras As Embedded Systems," *IEEE Computer* **35**, pp. 48–53, September 2002.
17. N. Bennani, "Integrating A Digital Camera in The Home Environment: Architecture and Prototype," in *International Symposium on Multimedia Software Engineering*, pp. 67–70, 2000.
18. J. Jeon, V. Lavrenko, and R. Manmatha, "Automatic Image Annotation and Retrieval Using Cross-media Relevance Models," in *ACM Conference on Research and Development in Information Retrieval*, pp. 119–126, 2003.

19. I. Ekman and P. Lankoski, “What Should It Do? Key Isssues in Navigation Interface Design for Small Screen Devices,” in *Conference on Human Factors and Computing Systems*, pp. 622–623, 2002.
20. G. A. Satalich, “Navigation and Wayfinding in Virtual Reality: Finding Proper Tools and Cues To Enhance Navigation Awareness,” Master’s thesis, University of Washington, 1995.
21. C. Kray, C. Elting, K. Laakso, and V. Coors, “Presenting Route Instructions on Mobile Devices,” in *International Conference on Intelligent User Interfaces*, pp. 117–124, 2003.
22. T. Vainio and O. Kotala, “Developing 3D Information Systems for Mobile Users: Some Usability Issues,” in *Nordic Conference on Human-computer Interaction*, pp. 231–234, 2002.
23. K. Gronbak, P. P. Vestergaard, and P. Orbak, “Towards Geo-spatial Hypermedia: Concepts and Prototype Implementation,” in *Conference on Hypertext and Hypermedia*, pp. 117–126, 2002.
24. J. Baus, A. Kruger, and W. Wahlster, “A Resource-adaptive Mobile Navigation System,” in *International Conference on Intelligent User Interfaces*, pp. 15–22, 2002.
25. S. Rogers, P. Langley, and C. Wilson, “Mining GPS Data To Augment Road Models,” in *International Conference on Knowledge Discovery and Data Mining*, pp. 104–113, 1999.
26. C. S. L. Chan, T. K. Y. Chan, and E. C. Prakash, “A System for Geographical and Spatial Data Exploration on The Internet,” in *Pan-Sydney Workshop on Visualisation*, pp. 23–29, 2001.
27. L. E. Jordan, “Dawn of A New Era: The Image Information Age,” in *Geoscience and Remote Sensing Symposium*, pp. 2228–2230, 1996.
28. H. W. Kang and S. Y. Shin, “Tour into The Video: Image-based Navigation Scheme for Video Sequences of Dynamic Scenes,” in *ACM Symposium on Virtual Reality Software and Technology*, pp. 73–80, 2002.
29. J. Hong, X. Tan, B. Pinette, R. Weiss, and E. M. Riseman, “Image-based Homing,” *IEEE Control Systems Magazine* **12**, pp. 38–45, February 1992.