A Mobile-Cloud Collaborative Traffic Lights Detector for Blind Navigation

Pelin Angin, Bharat Bhargava
Department of Computer Science
Purdue University
West Lafayette, IN, USA
{pangin, bb} @cs.purdue.edu

Sumi Helal
Dep. Computer and Information Science & Engineering
University of Florida
Gainsville, FL, USA
helal@cise.ufl.edu

***Abstract*—Context-awareness is a critical aspect of safe navigation, especially for the blind and visually-impaired in unfamiliar environments. Existing mobile devices for context-aware navigation fall short in many cases due to their dependence on specific infrastructure requirements as well as having limited access to resources that could provide a wealth of contextual clues. In this work, we propose a mobile-cloud collaborative approach for context-aware navigation, where we aim to exploit the computational power of resources made available by Cloud Computing providers as well as the wealth of location-specific resources available on the Internet to provide maximal context-awareness. The system architecture we propose also has the advantages of being extensible and having minimal infrastructural reliance, thus allowing for wide usability. A traffic lights detector was developed as an initial application component of the proposed system and experiments performed to test appropriateness for the real-time nature of the problem.**

Keywords-mobile; cloud; navigation; context-awareness

I. INTRODUCTION

Mobility is important for the quality of life. The ability to see, hear, and experience the context of the environment is critical for safety. Visually impaired or blind persons rely on their previous knowledge of an environment to navigate, usually getting help from guide dogs or the white cane, which leaves them handicapped in achieving desired level of mobility and context-awareness especially in unknown environments. Existing navigation systems for the blind and visually impaired people provide some level of help, but fail to address the important aspects of context-awareness, safety and usability. They also are not open and not designed for extensibility, which makes them unable to integrate or take advantage of the newer, more advanced technology and the wealth of relevant Internet resources. Most of these systems depend heavily on the underlying infrastructure, limiting their use in places where the infrastructure requirements are not met. Context information provided to the user by the available devices is usually very limited and the devices aiming to provide more detailed information (such as recognizing particular classes of objects) sacrifice portability, which is undesirable especially for long trips. Much can be done to enhance the experience and increase the safety and capabilities of individuals in navigating freely in buildings, college campuses, and cities. By providing maximal awareness of the environment and its contexts, without requiring any modification to the existing infrastructures, the quality and experience of navigation will be significantly enhanced for the blind user as well as other users in unfamiliar environments.

The urban world is becoming more complex every day with advances in technology, products of which such as quite cars, make it more difficult especially for the blind and visually-impaired to fully sense their environment. Existing route planning devices provide guidance in terms of directions to follow, but fail to address important safety issues such as when to cross at an intersection, which requires awareness of the status of traffic lights and dynamic objects such as cars. Accurate and fast object recognition and obstacle detection, which require the use of computationally intensive image and video processing algorithms, are becoming increasingly important for systems aiming to help the blind navigate independently and safely. The limited computational capacity and battery life of currently available mobile devices make fast and accurate image processing infeasible when the devices are to be used in isolation (i.e. without communicating with any external resources for the computations).

In this work, we propose a context-rich, open, accessible and extensible navigation system, bringing the quality of the navigation experience to higher standards. We use currently available infrastructure to develop an easy to use, portable, affordable device that provides extensibility to accommodate new services to help in high quality navigation as they become available. This paper particularly focuses on a traffic lights detector developed as an initial component of the proposed context-aware navigation system, which we aim to build on in future work. The rest of the paper is organized as follows: Section II discusses previous work in the area of mobile navigation devices; Section III describes the proposed mobile-cloud collaborative blind navigation system architecture; Section IV provides a description of the mobile-traffic light recognizer we developed as an initial component of the proposed system and experiment results are provided in Section V. Section VI concludes the paper with future work directions.

II. RELATED WORK

Systems based on different technologies have been proposed for the task of helping the blind and visually-impaired find their way at indoor and outdoor locations. After the introduction of the Global Positioning System

(GPS) in the late 1980s, many systems based on the GPS to help the visually impaired navigate outdoors were proposed and some were commercially released. Among those systems are the LoadStone GPS (http://www.loadstone-gps.com/), Wayfinder Access (http://www.wayfinderaccess.com/), BrailleNote GPS and Trekker by Humanware (http://www.humanware.com), and StreetTalk by Freedom Scientific (http://www.freedomscientific.com). The disadvantage of these GPS based devices is that their use is limited to outdoor environments and they provide limited contextual information during navigation. Drishti [1] developed at the University of Florida provides both indoor and outdoor navigation help, taking into account the dynamic changes in the environment. InfoGrid [2] uses an RFID (radio frequency identification) tag grid to allow for a localized information system, describing the surroundings at finer granularity, but its use is limited to places meeting the infrastructure requirements. Among indoor navigation systems are Jerusalem College of Technology's system [4] based on local infrared beams informing the user of the names of specific places in a building and Talking Signs (http://www.talkingsigns.com/) based on audio signals sent by invisible infrared light beams decoding names of indoor locations. SWAN by Georgia Institute of Technology [5] uses an audio interface to guide the listener along a path, while indicating the locations of other important features in the environment. There are also more specialized efforts such as facilitating grocery shopping as aimed with ShopTalk [6]. Although ShopTalk enables users to find the specific items they are looking for at a grocery store, it requires carrying hardware including a barcode scanner, its base station, a computational unit in a backpack and a numeric keypad, which is not desirable for most users. Despite the many efforts resulting in the development of the mentioned navigation systems, the blind population today still does not have access to an easy to use, sufficiently accessible and portable device providing detailed context information to ensure safe and independent navigation both indoors and outdoors.

Systems specifically for detecting the status of traffic lights to aid blind and visually-impaired as well as color-blind drivers were also proposed. Among those are [7], which consists of a digital camera and a portable PC analyzing the video frames captured by the camera and [8], which uses a 2.9 GHz desktop computer to process video frames in real time. Although these systems provide fairly accurate recognition of traffic lights, they are cumbersome due to the hardware they depend on for image and video processing. To be complete, a context-aware navigation system needs to achieve other vision based tasks such as detecting generic moving obstacles, as in [9], which requires around 400 ms video processing time on a dual core 2.66 GHz computer, which again hinders portability.

III. PROPOSED SYSTEM ARCHITECTURE

The context-aware navigation system architecture we propose is a two-tier architecture as seen in Fig. 1. The two main components are the "Mobile Navigation and Awareness Server" (mNAS), which could be any smart phone device in the market and the "Cloud Navigation and Awareness Server" (cNAS), which is basically the Web Services Platform we will employ to support a variety of context-awareness functionalities. The mNAS, with integrated location sensing module (GPS receiver) will be responsible for local navigation, local obstacle detection and avoidance, as well as interacting with the user as well as the cloud side. It will be responsible for providing location data to cNAS, which will perform the desired location specific functionality and communicate the desired information as well as relevant context information (contextlets) and warnings of potential hazards in context back to mNAS. Total navigation will be composed of cues and prompts based on the local navigation capability provided by mNAS and the additional location and other contextual data provided by the cNAS. For instance, total navigation will be tentatively achieved with a combination of GPS signals and Wi-Fi based location tracking to achieve better accuracy and support for tracking in outdoor and indoor environments as well as when the GPS signal is lost in outdoor environments. A compass integrated into the mobile device (which some Android platforms support) will be used to determine the direction the user is facing to provide for additional accuracy in path guidance.

The mNAS acts as a server to the user performing multi-tasks and prioritizing prompts, guidance, warnings, and other on textual information release. The mNAS also acts as a thin client to the CNAS server. It provides GPS coordinates as well as other user commands (including feedback), and receive succinct coded-text which is quickly expanded by mNAS and output as audio using text to speech capabilities (TTS). The mNAS takes on the delicate

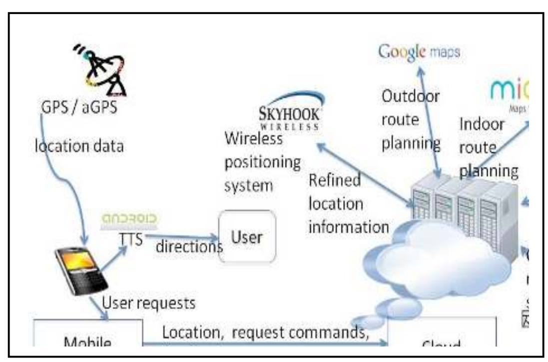

Figure 1. Proposed System Architecture

task of spatio-temporal modeling of all audio outputs based on priorities and an information release model that is cognitively acceptable.

The cNAS server will act as an integrator of select information sources that can provide more refined and critical context to the user. For instance Micello (http://micello.net) is a service that provides Navigational maps and an API to navigate inside buildings. It is an emerging service that must not be ignored and its benefits should be made available and accessible to the blind and visually impaired user. CNAS and our architecture make the inclusion of a service like Micello as well as other future services possible.

IV. TRAFFIC LIGHTS DETECTOR

The ability to detect the status of traffic lights accurately is an important aspect of providing safe guidance during navigation. The inherent difficulty of the problem is the fast image processing required for locating and detecting the status of traffic lights in the immediate environment. As real-time image processing is demanding in terms of computational resources, mobile devices with limited resources fall short in achieving accurate and timely detection. An accurate traffic lights status detection service would benefit not only the blind and the visually-impaired, but also the color-blind as well as systems like autonomous ground vehicles and even careless drivers. The shortcomings of mobile devices in terms of computational power and short battery life in providing this type of service can easily be compensated with the wealth of resources made available by Cloud Computing providers.

A. Detector Architecture

The system architecture we developed for the traffic lights detector consists of two main components: The mobile component can be any smart phone device with an integrated camera (readily available in the market today) and the cloud component is a set of servers made available by Cloud Computing providers dedicated to perform specific tasks as needed to provide context-awareness. The mobile component is responsible for communicating the location-specific information gathered through sensors (such as the integrated camera) along with the desired function to the cloud component, which processes the information received and sends back a response as appropriate.

Fig. 2 shows a schema of the traffic lights detector system developed as an initial component of the context-aware navigation system proposed. The mobile component of the system is an Android (http://www.android.com) based mobile phone and the Elastic Compute Cloud service of Amazon Web Services (http://aws.amazon.com/ec2/) is used to host the cloud component, where the server is responsible for receiving video frames from the Android mobile device, processing to detect the presence and status of traffic lights in the frame and sending a response as appropriate back to the mobile device over a TCP connection. The status of the traffic lights as detected by the remote server is communicated to the user via the text-to-speech interface of

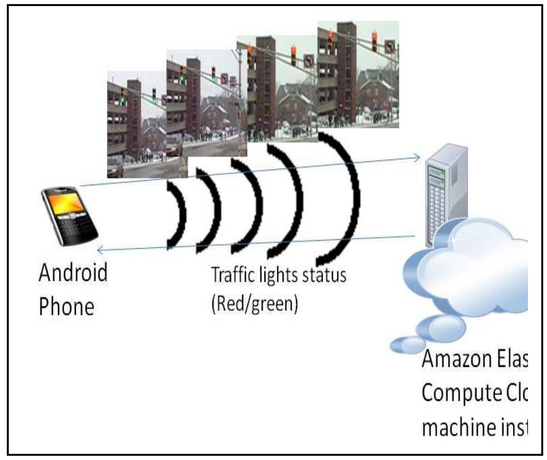

Figure 2. Traffic lights detector system

the Android platform. The traffic lights detector application running on the cloud component uses the OpenCV (http://opencv.willowgarage.com/) implementation of the AdaBoost algorithm for fast object detection, which is explained in the next part.

B. Object Detector

AdaBoost [10] is an adaptive Machine Learning algorithm used commonly in real-time object recognition due to the short detection time it allows for. It is based on rounds of calls to weak classifiers (classifiers whose detection rates are slightly better than random guessing) to focus more on incorrectly classified data samples at each stage to increase classification accuracy. The traffic lights detector of the developed system uses a cascade of boosted classifiers based on the AdaBoost algorithm and haar-like features [11] to detect the presence and status of traffic lights in a video frame captured by the camera of the Android mobile phone. Detectors (separately for red and green traffic lights) were trained on 219 images of traffic lights obtained from Google images (http://images.google.com) as well as pictures taken at the Purdue University campus locations. The training dataset includes pictures taken under different conditions (such as clear/snowy weather) as well as from different angles to ensure completeness. The classifiers were trained with 8 stages of the cascaded boosting algorithm and the minimum recall of each stage (the number of traffic lights detected out of all in the dataset) was set to 0.95 during training as it is important not to miss the presence of any traffic lights in the scene.

C. Traffic Lights Detection Challenges

The problem of providing real-time feedback about the status of traffic lights in the immediate environment faces challenges even when a mobile-cloud collaborative approach as explained is taken. One of the main concerns about this approach is the time it takes to send the video frames to the remote server for processing and to receive a response. The real-time nature of the problem requires response times ideally less than 1 second to provide accurate and safe guidance to the blind or visually impaired user. While the server having sufficient computation resources takes negligible time to process the received frames, network latency could create a bottleneck on the timeliness of the response to be received by the user. Continuous Internet connectivity is another problem faced by the proposed approach. Signals from wireless networks would be weak or mostly unavailable at outdoor locations, which is the main setting the application is supposed to work at. However, availability of data plans by major cell phone carrier companies today alleviates this problem. Many people are already subscribed for these data plans for a low monthly cost for continuous connectivity. Another major challenge faced is the short battery life of the mobile device. The continuous video recording approach taken in the current system exhausts the battery of the mobile device too soon, causing service interruption. A power-optimized approach as explained in the next part will need to be employed to ensure continuous guidance to the user.

D. Proposed System Enhancements

The traffic lights detector system developed is an initial attempt to demonstrate the effectiveness of a mobile-cloud collaborative approach for context-aware blind navigation. The system can be enhanced in many ways to ensure high quality service to the users. To overcome the problem of service interruption due to short battery life, video capture by the mobile device should be performed sparingly based on previous knowledge of the location of traffic lights. This information is readily available in maps extracted for use in GPS devices, where locations of traffic lights are marked as points of interest. The only extra requirement to make use of this information is having a GPS receiver on the mobile device, which many devices already do. The GPS receiver is already an indispensable component of a blind navigation system due to its use in location tracking for route planning. With a simple modification to the current system, the mobile device would only need to capture frames at locations close to the GPS coordinates of a traffic light and send them to the server for processing, which would save battery life. Yet another modification that could prove useful for the system is performing a simple preprocessing of the frames captured to get lower resolution versions which could be transferred to the cloud in a shorter time for processing. A schema of the proposed system architecture is seen in Fig. 3.

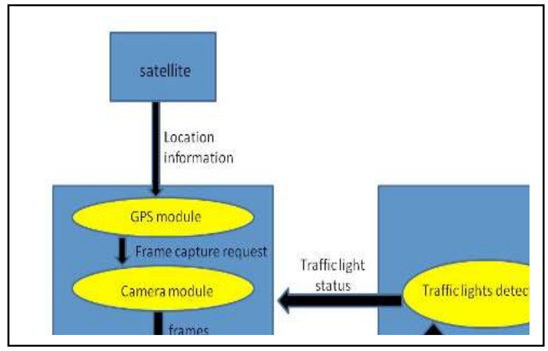

Figure 3. Enhanced traffic lights detector schema

V. EXPERIMENTS

The two most important aspects of the traffic lights detection problem are timeliness of response and accuracy. The real-time nature of the problem necessitates response times of less than 1 second as stated before, while high accuracy of detection should be achieved to ensure safety of the user.

Experiments were performed to test the accuracy and response time of the traffic lights detector application developed. Test data used in the experiments consists of video recordings at outdoor locations of the Purdue University campus (disjoint with the training data), which include scenes of different traffic lights, a sample of which can be seen in Fig. 4 and Fig. 5 shows the detector output for the sample data. The Android application developed was installed on an HTC mobile phone, connected to the Internet through a wireless network on campus. The sample task in the experiments involved processing five different resolution level versions of 934 video frames.

The average response times, which were determined by the time period between capturing a frame and receiving the response from the server running at Amazon Elastic Compute Cloud about the traffic lights status, were measured for each frame resolution level as determined by a Java platform-specific measure. A resolution level of 0.75 stands for the original frame as captured by the camera, whereas the lower resolution levels represent compressed versions of the same set of frames, where image quality falls with decreasing resolution level. Response times for the original frames were found to be around 660 milliseconds on average, which are acceptable levels for the real-time requirements of the problem. We also saw that response time decreases further when lower-quality, compressed versions of the frames are sent to the remote server instead of the originals.

Figure 4. Test data sample from detection experiments

Figure 5. Detector output on sample data

The recall values for each resolution level were also recorded and while resolution levels of 0.50 and 0.30 resulted in the same recall values achieved by processing the original frames, the resolution level 0.1 resulted in 15% decrease over the original recall and the 0.05 level had a 23% decrease. These results are promising in terms of being able to get fairly accurate results even with lower quality, smaller sized image files being sent to the cloud component for processing.

VI. CONCLUSION

In this paper we proposed an open and extensible architecture for context-aware navigation of the blind and visually impaired. The system proposed is based on collaboration between everyday mobile devices and the wealth of location-specific information resources on the Web as well as the computational resources made available by major Cloud Computing providers, allowing for richer context-awareness and high quality navigation guidance. We also described a traffic lights detector system developed as an initial component of the context-aware navigation system proposed and provided experimental results.

Future work on the navigation system proposed will involve efforts in many different aspects including robust obstacle detection; integration of important context information into route planning such as traffic lights status and dynamic/static obstacles information as well as infrastructure-independent indoor route planning.

REFERENCES

[1] L. Ran, A. Helal, and S. Moore, "Drishti: An Integrated Indoor/Outdoor Blind Navigation System and Service," 2nd IEEE Pervasive Computing Conference (PerCom 04).

[2] S.Willis, and A. Helal, "RFID Information Grid and Wearable Computing Solution to the Problem of Wayfinding for the Blind User in a Campus Environment," IEEE International Symposium on Wearable Computers (ISWC 05).

[3] P. Narasimhan, "Trinetra: Assistive Technologies for Grocery Shopping for the Blind," IEEE-BAIS Symp. Research on Assistive Technologies (2007).

[4] Y. Sonnenblick, "An Indoor Navigation System for Blind Individuals," Conference on Technology and Persons with Disabilities (1998).

[5] J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, and F. Dellaert, "SWAN: System for Wearable Audio Navigation," IEEE International Symposium on Wearable Computers (ISWC 07).

[6] J. Nicholson, V. Kulyukin, D. Coster, "ShopTalk: Independent Blind Shopping Through Verbal Route Directions and Barcode Scans," The Open Rehabilitation Journal, vol. 2, 2009, pp. 11-23.

[7] Y.K. Kim, K.W. Kim, and X.Yang, "Real Time Traffic Light Recognition System for Color Vision Deficiencies," IEEE International Conference on Mechatronics and Automation (ICMA 07).

[8] R. Charette, and F. Nashashibi, "Real Time Visual Traffic Lights Recognition Based on Spot Light Detection and Adaptive Traffic Lights Templates," World Congress and Exhibition on Intelligent Transport Systems and Services (ITS 09).

[9] A.Ess, B. Leibe, K. Schindler, and L. van Gool, "Moving Obstacle Detection in Highly Dynamic Scenes," IEEE International Conference on Robotics and Automation (ICRA 09).

[10] Y. Freund, and R.E. Schapire, "A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting," Journal of Computer and System Sciences, vol. 55, 1997, pp. 119-139.

[11] R. Lienhart, and J. Maydt, "An Extended Set of Haar-Like Features for Rapid Object Detection," IEEE International Conference on Image Processing (ICIP 02).