Smart phone application for indoor scene localization

Nabeel khan
Computer Science Department
University of Otago
nabeel@cs.otago.ac.nz

Brendan McCane
Computer Science Department
University of Otago
mccane@cs.otago.ac.nz

ABSTRACT

Blind people are unable to navigate easily in unfamiliar indoor environments without assistance. Knowing the current location is a particularly important aspect of indoor navigation. Scene identification in indoor buildings without any Global Positioning System (GPS) is a challenging problem. We present a smart phone based assistive technology which uses computer vision techniques to localize the indoor location from a scene image. The aim of our work is to guide blind people during navigation inside buildings where GPS is not effective. Our current system uses a client-server model where the user takes a photo from their current location, the image is sent to the server, the location is sent back to the mobile device, and a voice message is used to convey the location information back to the user.

Categories and Subject Descriptors

K.4.2 [**Computers and Society**]: Social Issues—*Assistive technologies for persons with disabilities*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Shape*

General Terms

Verification, Design, Measurement.

Keywords

Blind people, Indoor Navigation, Features, Voice.

1. INTRODUCTION

The number of blind people in the world is about 39 million with another 246 million with visual impairments. A wide range of products based on GPS are available to assist blind people during outdoor navigation, but GPS is not accurate indoors and it cannot provide details such as which floor the user is on. The emergence of low cost touch-based smart phones with cameras in the last few years have made it possible to develop cheaper navigation solutions based on computer vision techniques. Smart devices are already in use by blind people for reading emails, messages etc. So it is relatively easy for a blind person to use a smart phone application. Our proposed smart phone application can be used by a blind person to take a photo of the current scene and determine their location whenever they want. The proposed vision based solution can be used in any building for which a suitable database of images has been collected.

ASSETS'12, October 22–24, 2012, Boulder, Colorado, USA.
ACM 978-1-4503-1321-6/12/10.

2. PROPOSED SYSTEM

The proposed system uses a client server paradigm. The client side refers to an Android application currently running on an *HTC Wildfire S* (2.3 Gingerbread) smart phone. The server uses computer vision techniques to match the query photo sent by the phone [3]. Our system consists of two independent modules:-

1. **Mapping:-** The building intended for navigation needs to be mapped first. A large number of images covering all locations within a building are captured and stored along with the corresponding location information on the server. Images are represented by features which represent the unique parts of an image and are used for image matching. Our server uses a Visual Bag of Word (BoW) approach based on SIFT type of features for image matching [2, 3]. The generated features from the stored mapped images are clustered by approximate k-means to obtain the visual word's frequency for every image. The server stores this information in a data structure referred as an "inverted index" and then waits for the client request. The index is generated off-line and plays a key role in robust image matching.

2. **Scene Localization:-** Upon user request, the application takes a photo of an indoor place from the camera, sends the photo to the server using a wireless Internet connection and waits for a reply. When the server gets the photo, it uses the index to quickly retrieve 100 mapped images most similar to the query photo. These similar images are then ranked in ascending order giving higher ranks to the images closer to the query photo. If the top three ranked images vote for the same indoor location, the server simply returns that corresponding location. Otherwise geometry information is checked between the query and the top ranked images one by one via a fundamental matrix computation. If a reasonable number of features of the top ranked database image (at least 20%) indicate a relationship against the query photo, then the location of that ranked image is returned followed by a voice message on the phone.

3. APPLICATION DESIGN

The interface of our application is shown in Figure 1. Blind people can use a voice powered intelligent virtual assistant to launch the application on the phone. The application starts with a welcome message and waits for user input. The blind person clicks the scene localization button and gets a voice message regarding the current location. The phone can be held either in landscape or portrait position. The simple interface of our application provides high accessibility, although a voice activated button would further improve the interface.

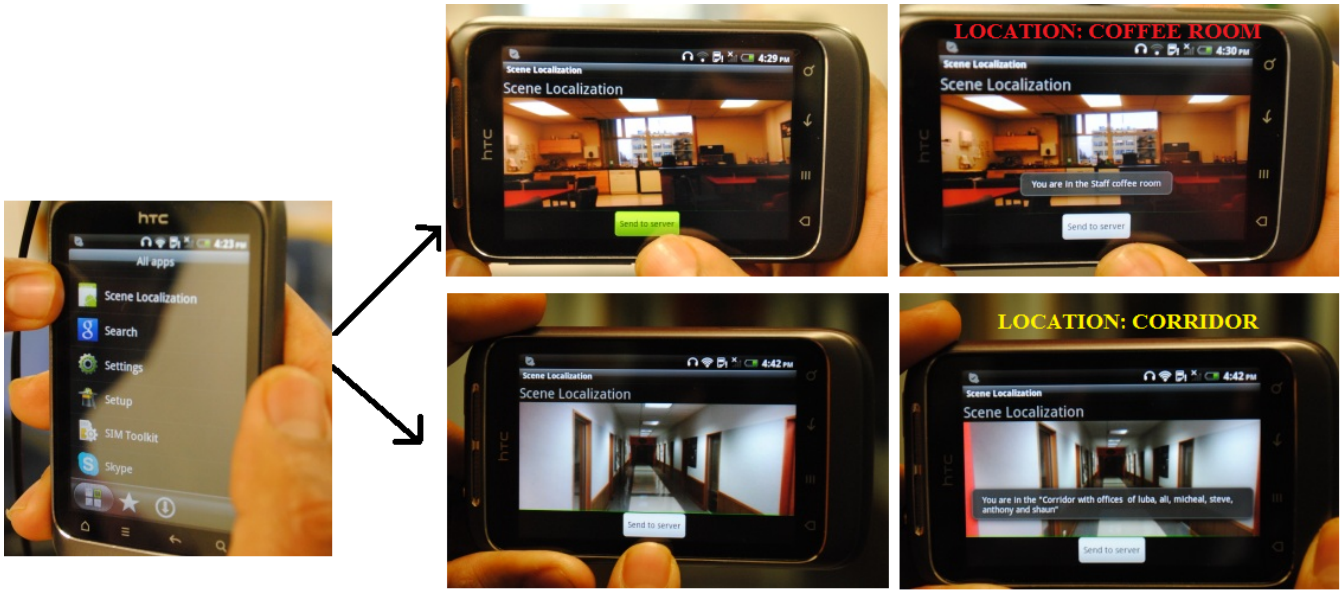

Figure 1: User interface of our application. The scene localization button captures and sends the photo to the server. The returning location information from the server is communicated to the blind person via a voice message.

4. EXPERIMENTS

The application takes 2-4 seconds on average for localization and is evaluated on indoor images taken from office buildings where scene matching is challenging due to high self-similarity between the images. Sometimes our application fails to find a match for the query photo, the user is then instructed to turn around and take another photo. In our experiments, we consider such cases a wrong match for strict analysis. Moreover mapped images are taken from a different camera for unbiased system evaluation. The used datasets [1] and results are as follows:-

1. **CS Indoor:-** It contains images of our computer science building. 1586 images are captured from three floors of the building representing 25 places such as corridors, labs, halls etc. For testing, we used a smart phone to take pictures during the night, noon and early morning. The results in Table 1 show that system did not perform well at night. This can be attributed to the glass structure of our building which results in reflections at night therefore resulting in wrong matches. Nevertheless the overall performance of the system is quite good and can further go up by excluding "no location" as wrong matches in our evaluation.

Table 1: Matching results by our system on CS Indoor

Datasets	Test Images	Match performance	No Location
Night	223	75.19%	25
Noon	293	89.95%	9
Morning	234	87.23%	6

2. **Commerce Indoor**:- It contains images of another office building covering about 2 floors. 864 images are currently captured representing 14 places like corridors, stairs, atrium etc. For evaluation, we used a smart phone to take 137 photos and the system gives 92.7% accuracy with 10 wrong matches.

5. CONCLUSION

The aim of our current work is to provide a cheap assistive technology for blind people. The proposed application can guide blind people during indoor navigation. The system works very well in buildings displaying high self-similarity between locations and should work even better in heterogeneous buildings where matching is relatively easier. The incorporation of GPS to load the appropriate mapped index on the server can make our application scalable to any number of buildings. Currently, the system works entirely from two-dimensional images and provides localisation at a room-based level. We are investigating three-dimensional matching strategies that will allow for finer scale localisation with a final target being a full indoor navigation system.

6. ACKNOWLEDGMENTS

We will like to thank Disability Information and Support Center at the Otago university for arranging interviews with blind people to capture the requirements analysis.

7. REFERENCES

[1] N. Khan. Indoor images of office buildings (nz). `http://www.cs.otago.ac.nz/pgdweb/nabeel/Downloads/`.

[2] N. Khan, B. McCane, and G. Wyvill. Sift and surf performance evaluation against various image deformations on benchmark dataset. In *Proc. of IEEE DICTA*, pages 501–506, 2011.

[3] N. Y. Khan, B. Mccane, and G. Wyvill. Homography based Visual Bag of Word Model for Scene Matching in Indoor Environments. In *Proc. of IVCNZ*, 2011.