Automated Description Generation for Indoor Floor Maps

Devi Paladugu, Hima Bindu Maguluri*, Qiongjie Tian*, Baoxin Li
Computer Science and Engineering, Arizona State University
{apaladug,hmagulur,qtian5,bli24}@asu.edu

ABSTRACT

People with visual impairment can face numerous challenges navigating a new environment. A practical need is to navigate through unfamiliar indoor environments such as school buildings, hotels, etc., for which commonly-used existing tools like canes, seeing-eye dogs and GPS devices cannot provide adequate support. We demonstrate a prototype system that aims at addressing this practical need. The input to the system is the name of the building/establishment supplied by a user, which is used by a web crawler to determine the availability of a floor map on the corresponding website. If available, the map is downloaded and used by the proposed system to generate a verbal description giving an overview of the locations of key landmarks inside the map with respect to one another. Our preliminary survey and experiments indicate that this is a promising direction to pursue in supporting indoor navigation for the visually impaired.

Categories and Subject Descriptors

H.5.2 **[Information Interfaces and Presentations]**: User-centered design. K.4.2 [**Social issues**]: Assistive technologies for persons with disabilities.

General Terms

Design, Experimentation, Human Factors

Author Keywords

Visually Impaired, Indoor Floor maps, Navigation, Verbal Description

1. INTRODUCTION

Assistive technologies for the visually impaired have seen tremendous developments. Nevertheless, there still remain many practical barriers for a blind individual who strives to lead an independent and active life [1]. One such problem faced repeatedly in their everyday life is navigation, especially when visiting new places. Indoor navigation in unfamiliar places, in spite of being a big hindrance factor, has not received enough attention from researchers and developers. Sonar, camera, laser based devices and infrared based signage are some devices developed in the last decade to help with indoor navigation via obstacle detection [1]. However, these devices require sensory devices to be strategically placed inside the building, which is a constraint. Further, these devices do not provide an overall sense of the location or help with planning a visit to a new location. People still largely depend on seeking human help around the location to get directions.

A simple and effective solution is to describe a floor map, when available, in terms of a scheme of relative positions. Using a relative scheme eliminates the need for having multiple devices tracking user position in real-time. A system that generates a verbal description combined with a user-friendly interface can be very useful in helping the user to get a sense of relative positions of locations and thus might reduce the need for totally relying on "asking for directions". Say a student who is visually impaired wants to go to a local library that he has never visited before (or with which he has little familiarity). He could gain knowledge of the key locations before he gets there, if a verbal description of the major landmarks and their geographical relation are readily available. An automatic indoor map description generator that can be easily used by a visually-impaired user may thus greatly benefit the blind population in promoting and supporting a more active life style.

2. SYSTEM AND DESIGN

The first problem towards building a system to provide verbal description of a floor map is to determine the availability of a usable digital map of good resolution. A majority of public buildings have their floor plans published on their website. However, in general, it is very difficult if not impossible for a visually impaired user to go through all the links on a web site to locate and assert that a file/link is indeed a floor plan. Secondly, given an image of the floor plan, the components of the plan need to be determined automatically. Thirdly, assuming all the components of the map can be segregated, finding the right way to describe these components is another challenging task. To the best of our knowledge, there is no fully automatic solution that completely solves the problems in an end-to-end system. This demo will present a prototype system that we have developed for working around the problems presented above. The system provides a simple interface for a user to type in the search keywords for the place he is interested in visiting. The system then uses these keywords to extract the URL/domain of the relevant website. A keyword-based search is employed to parse all the URLs obtained from the website to find the appropriate floor plan, if it indeed exists. Automated image process algorithms are then deployed to extract useful landmarks in the map. Finally, the system generates a verbal description that is presented on the screen as well as read out to the user. The current version of the system has been focused on public libraries, to properly constrain the domain to some degree. The overview of the system is illustrated in Figure 1.

2.1 WebCrawler Interface

We designed an interface that takes in a set of keywords or an URL and downloads the relevant image or pdf files. We used Google's "I am feeling lucky" search URL format to obtain the first link, given the keywords. Given this URL, we use the CURL and DOM features in a simple PHP script combined with keyword search to extract the required URLs containing the floor plan(s).

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

* Indicates equal contribution.

2.2 Key Landmark Localization

In this step, common landmarks of the domain (e.g., libraries) are located in a map. The processing steps also include text detection and recognition. Below are the major processing steps:

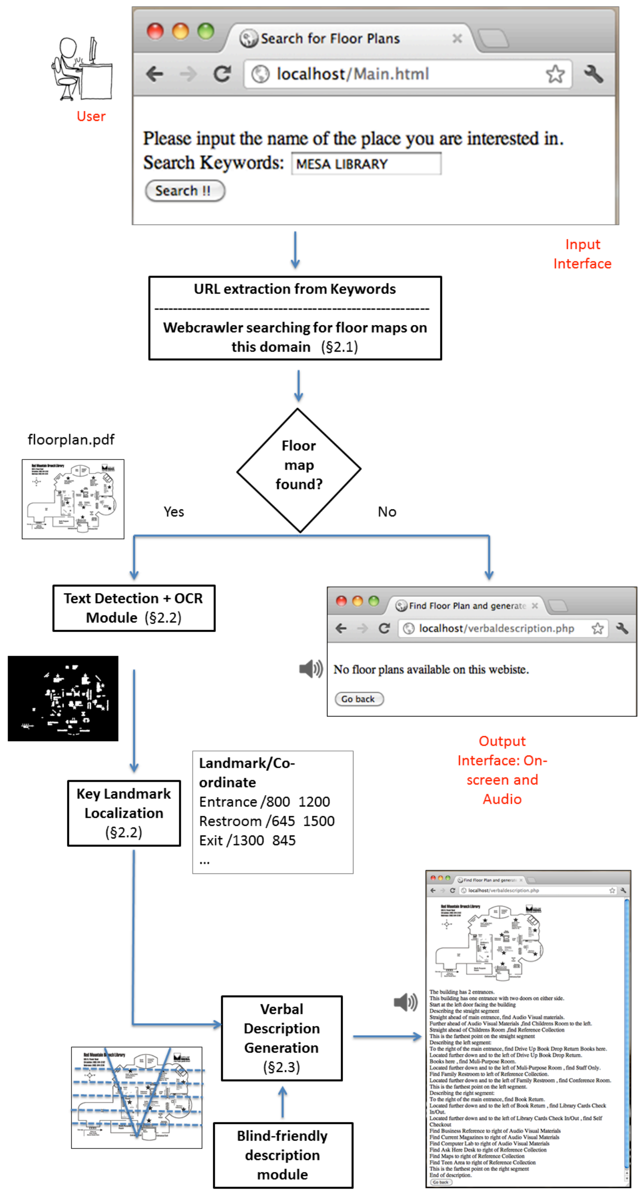

Figure 1. An overview of the proposed system.

Pre-processing: This step includes removing the color information, detection and removal of the legend, line removal.

Text Detection: We use spatial frequency variations along vertical, horizontal and diagonal direction for each pixel to estimate if a pixel contains text [2].

Text Region Post-processing: Text blocks are detected and the orientations of text are estimated and compensated for OCR.

OCR: We integrate Tesseract [3] to process the segmented text regions to obtain the text corresponding to each region.

Auto-correction and OCR: For initial refinement, we use a standard dictionary and spell check on every detected word to correct some bad detection and recognition results.

For words corresponding to key locations, we save the centroid of each detected text box and the corresponding text for the subsequent stages.

2.3 Verbal Description Generation

We performed a small-scale case study and analyzed how blind users verbally describe a building that they are familiar with and also tried to learn how they usually receive layout information from sighted individuals. We asked a small group of blind users to describe a building on campus that they are familiar with, to another blind individual who is new to the building. We analyzed the descriptions and tried to mimic the observations when designing the verbal description, resulting in the “Blind-friendly description model” shown in Figure 1.

The most natural way to describe a building as seen from our case study can be summarized as follows: Describe the rooms or landmarks encountered when walking straight into the main entrance door. Then describe locations to the left and to the right either with respect to the objects described previously or with reference to the entrance. As an attempt to simplify this task, we divide the map into three segments: “straight”, “left” and “right”. It is intuitive to divide into three regions as shown in Figure 1. Once we have each point categorized into a segment, we further divide the map into grids, to give the user a sense of distance between the relative points. Based on the grid location and segment, we use simple grammatical rules to generate verbal description.

3. SUMMARY

The contribution of this work is three-fold. First, a simple interface that eliminates all the work needed from a user to obtain a map is implemented. Second, automated algorithms were designed to factorize the map into the necessary components needed for description generation. Thirdly, a simple but efficient way to describe a floor plan learnt from the visually impaired users during a case study is proposed. The system will be demonstrated on the conference.

4. ACKNOWLEDGEMENT

This material is based upon work supported in part by the National Science Foundation under Grant No. 0845469. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

5. KEY REFERENCES

[1] Giudice, N.A., & Legge, G.E. (2008). Blind navigation and the role of technology. In A. Helal, M. Mokhtari & B. Abdulrazak (Eds.), Engineering handbook of smart technology for aging, disability, and independence (pp. 479-500): John Wiley & Sons.

[2] Z. Wang, et al., "Instant tactile-audio map: enabling access to digital maps for people with visual impairment," ACM SIGACCESS conference on Computers and Accessibility, Pittsburgh, Pennsylvania, USA, 2009.

[3] http://code.google.com/p/tesseract-ocr/