Mobile Navigation and Augmentation utilizing Real-World Text

Mike Eissele and Thomas Ertl

Institute for Visualization and Interactive Systems (VIS) University of Stuttgart

Abstract

Mobile, smart devices are already available in everyday life. Most of them offer a rich feature set, including camera, GPS antenna, or even WIFI modules. Navigation software utilizing GPS to acquire location data became very popular on these mobile clients. However, due to hardware restrictions GPS positioning cannot be used for indoor scenarios. In contrast, this paper is focused on navigation for indoor environments targeting on a low-cost solution, without the requirement of additional hardware installations or markers inside buildings. A server-based optical character recognition service is used to map images containing unique text passages – acquired with a mobile client – to additional meta data, e.g. locations or orientation. We also very briefly discuss that augmented-reality methods can be used to provide an intuitive presentation and interaction system by overlying real-world images with additional information like navigation symbols or internet hyperlinks.

1 Introduction

Many applications are designed for mobile devices like PDAs, Smartphones, or Ultra Mobile PCs, however, navigation is still one of the most popular software used on these appliances. Commercial applications to support users in path finding for outdoor scenarios are already widely available. In contrast, systems that help users to find locations inside buildings – i.e. meeting points, rooms, persons, etc. – are not yet common and most of the available solutions require additional hardware setups inside the target environment, as the GPS signal cannot be received inside buildings. These installations range from low-cost visual markers (1D/2D barcodes) up to expensive ultrasonic or electro-magnetic beacons that have to be deployed inside the building at suitable positions, i.e. good visible locations for optical markers. Both cases require an involvement of the facility management to evaluate either if a high number of visual markers can be placed inside the building or a costly installation of beacons is possible.

In contrast, the system we propose makes use of commonly installed good-recognizable features at visible locations: Text phrases of, e.g., door plates, posters, or labelled installations are used to indirectly identify the user's current location. A database of mappings for text paragraphs to additional meta data, most importantly the location inside the building, is used in combination with a server-based OCR system. This service is used by the developed client software running on a PDA and allows users to capture images of text inside the building to access additional information. The presented prototype makes use of position and orientation of captured text for a navigation application and can also display additional textual meta data like language translations or word explanations. The main advantage of our system is that no further installations in buildings are required in order to use the system.

2 Location System Overview

The localization system we propose is separated into two components: A server connected to a database to perform optical character recognition and text matching and a hardware specific client application to perform navigation tasks and access additional information to real world texts. An overview of the system is depicted in Figure 1.

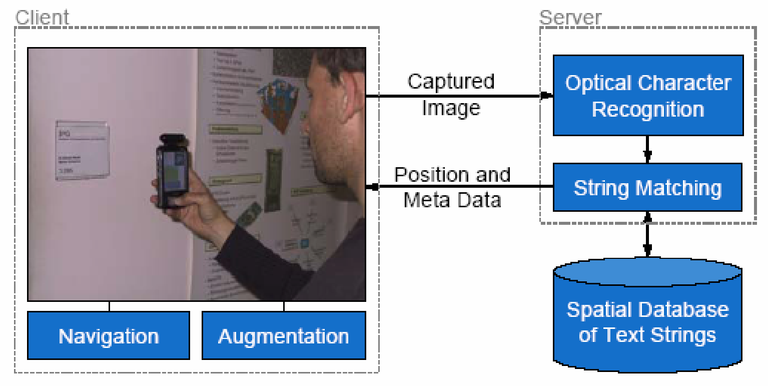

Figure 1: System Overview

In order to be able to deploy the proposed system in an environment, unique text phrases have to be identified. Good examples for such text strings inside buildings are: Door plates or room numbers, unique placards or posters, labelled access ports or switches, etc.

2.1 Location Based Services using Augmentation

With the information returned to the client two primary services are provided to the end users: Indoor navigation and text augmentation with additional information. The proposed location technique can provide highly accurate position and direction information without requiring any additional hardware installations. For that purpose we equipped the server text-string database with additional information about the orientation of each text entity inside the building. In addition, the server calculates the position of the camera relative to the text entity during the perspective correction of a received image. Note that our prototypical implementation does not perform a perspective correction we therefore assume that the camera is facing the text orthogonally, at a distance of about 30cm. The server combines both peaces of information and transmits the global position and orientation to the client. The user's current view direction can be shown in map drawings, to help them finding their orientation relative to the map (Figure 2, yellow-blue pointer, bottom part). Utilizing orientation for the visualization of navigation information also offers the possibility to introduce augmented-reality techniques, thereby eliminating the requirement of users having to figure out directions themselves. Augmenting the image captured with the client requires a consistent visualization of the navigation symbols, relative to the orientation of the text entity seen in the image. We therefore project the navigation direction, calculated by shortest path finding, to the plane of the text entity. This new navigation direction is used to select and render navigation hints directly onto the real-world captured image (Figure 2, upper right).

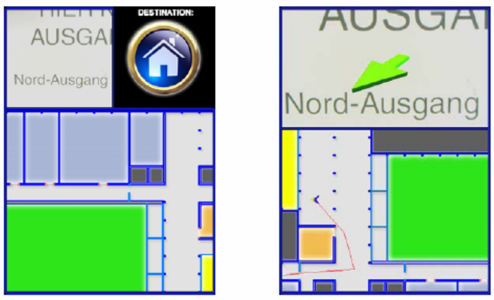

Figure 2: Navigation client interface. Left: real-time camera view with destination and floor map. Right: augmented view after the image capturing showing the direction relative to the sign, map displays location with direction

2.2 Information Augmentation of Real-World Texts

In a similar way as the systems proposed by (Kameyama et al. 2005) and (Jagannathan & Jawahar 2005, I) for translation of texts to other languages, the server component in our system can attach additional meta data to the recognized text. Even further, dynamic information can be generated, based on multiple context aspects – e.g. client device type, user settings, time, date, etc. – and transmitted to the client. Using small information icons, also displayed using AR methods, result in an intuitive user interface where users simply have to click on the icons embedded in the captured image to access the additional data. The required information and the image-space positions for the overlay icons are generated by the server. Therefore, the text-string database allows defining meta data to text entities, but also for arbitrary substrings of each stored text line.

The demonstrated interplay of camera images and direct interaction on augmented texts using a pen device characterizes an intuitive user interface to access relevant information of the user's current environment. The interaction is thereby limited to single pen tips or button pushes, multi-modal buttons or textual entries using techniques like virtual keyboards or handwriting recognition are replaced by a simple image capturing of the relevant text phrases.

3 Conclusion and Future Work

We have briefly shown an intuitive mobile system that supports users in accessing information and navigation inside buildings. The system does not require any additional hardware installations like location systems or markers. Existing wireless infrastructure – WLAN or GPRS – is used to transmit captured images containing text to a server, which tries to find a mapping of text phrases to additional meta data like position, orientation, or language translation utilizing OCR. The returned result is presented on the client device as an overlay to the realworld images using augmented reality techniques. This allows an intuitive presentation, especially of navigation symbols, which users easily understand and benefit from. The prototypical implementation of the recognition system also shows that the performance is still adequate for the presented application even for big configurations of 5000 text entities.

In future we will further optimize the proposed system including the perspective correction of the captured images and an image compression on the client device for transmission to the server. We also want to investigate a hybrid system where parts of the OCR method – i.e. the document layout analysis – are performed on the client device to be able to transfer onlyimportant parts of the image to the server.

References

Kameyama, T.; Takahashi, T.; Yamazaki, M.; Yamaguchi, T.; Koga, M.; Mine, R. (2005): Camera-based Kanji OCR for mobile-phones: practical issues. In Proceedings of the Eighth International-Conference on Document Analysis and Recognition, Vol. 2, pages: 635- 639.

Jagannathan, L.; Jawahar, C.V. (2005, I): Crosslingual Access of Textual Information using Camera Phones. In Proceedings of the International Conference on Cognition and Recognition.

Jagannathan, L.; Jawahar, C.V. (2005, II): Perspective correction methods for camera based document analysis. In Proceedings of the First International Workshop on Camera-based Document Analysis and Recognition (CBDAR), pages 148-154.

Contact

Mike Eissele

Institute for Visualization and
Interactive Systems (VIS)

University of Stuttgart
Universitätsstr. 38
70569 Stuttgart

Phone: +49-(0)711-7816-246
Fax: +49-(0)711-7816-340
E-Mail: eissele@vis.uni-stuttgart.de

Thomas Ertl

Institute for Visualization and
Interactive Systems (VIS)

University of Stuttgart
Universitätsstr. 38
70569 Stuttgart

Phone: +49-(0)711-7816-246
Fax: +49-(0)711-7816-331
E-Mail: ertl@vis.uni-stuttgart.de