Computers, Environment and Urban Systems 67 (2018) 110–123

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

A dialogue based mobile virtual assistant for tourists: The SpaceBook Project

Phil Bartie[a,*], William Mackaness[b], Oliver Lemon[c], Tiphaine Dalmas[b], Srini Janarthanam[c], Robin L. Hill[b], Anna Dickinson[b], Xingkun Liu[c]

[a] University of Stirling, United Kingdom
[b] University of Edinburgh, United Kingdom
[c] Heriot-Watt University, United Kingdom

ARTICLE INFO

Keywords:
Location based service
Spoken Dialogue System
Viewshed
Virtual city guide

ABSTRACT

Ubiquitous mobile computing offers innovative approaches in the delivery of information that can facilitate free roaming of the city, informing and guiding the tourist as the city unfolds before them. However making frequent visual reference to mobile devices can be distracting, the user having to interact via a small screen thus disrupting the explorative experience. This research reports on an EU funded project, SpaceBook, that explored the utility of a hands-free, eyes-free virtual tour guide, that could answer questions through a spoken dialogue user interface and notify the user of interesting features in view while guiding the tourist to various destinations. Visibility modelling was carried out in real-time based on a LiDAR sourced digital surface model, fused with a variety of map and crowd sourced datasets (e.g. Ordnance Survey, OpenStreetMap, Flickr, Foursquare) to establish the most interesting landmarks visible from the user's location at any given moment. A number of variations of the SpaceBook system were trialled in Edinburgh (Scotland). The research highlighted the pleasure derived from this novel form of interaction and revealed the complexity of prioritising route guidance instruction alongside identification, description and embellishment of landmark information – there being a delicate balance between the level of information 'pushed' to the user, and the user's requests for further information. Among a number of challenges, were issues regarding the fidelity of spatial data and positioning information required for pedestrian based systems – the pedestrian having much greater freedom of movement than vehicles.

1. Introduction

Technology has long supported tourists in experiencing the city from trip planning, to finding public transport information, to providing navigation assistance, to post-trip reminiscing (online photo sharing and blogs). The smartphone has revolutionised how travellers personalise their travel experiences but an ongoing concern of the smartphone platform is that finding relevant information on screen distracts the user from their environment - a conspicuous and somewhat hazardous activity. What is required is a more concealed technology that supports intuitive interaction but does not come between the tourist and their enjoyment of the city. This paper reports on the SpaceBook Project (EU ref.: 270,019), which focussed on designing a wearable technology that delivers relevant information (information push), and responds to user questions (information pull) while the tourist explores the city on foot. In order that it leaves the tourist both hands-free and eyes-free, the system was entirely dialogue based using only speech input and output. The system was able to provide navigation guidance as well as identify landmarks in view (e.g. statues, buildings, parks) by modelling visibility in real-time based on a Digital Surface Model (DSM) built from LiDAR data. While there are many approaches to informing a mobile user, (Google glass, haptic interfaces), to our knowledge SpaceBook was the first system to rely solely on natural language speech and text-to-speech responses to support in situ navigation and exploration of the urban environment. While there have been industry led speech-based virtual agents (e.g. Apple's Siri, Microsoft's Cortana), SpaceBook is unique in modelling context by calculating the pedestrian's field of view in order to provide situated speech-based dialogue to support both navigation and exploration in the truest sense of the word.

The system was evaluated by 42 people on the streets of central Edinburgh (a busy area crowded with visitors and traffic and with a wide variety of geographic features and topography). The evaluation had four main aims: (1) to establish the performance of continuous Automatic Speech Recognition (ASR) in a noisy outdoor environment; (2) to model object visibility in real-time and in conjunction with social media data (e.g. Flickr, Foursquare) in order to determine useful landmarks to assist in navigation tasks or information push; (3) to

* Corresponding author.
E-mail address: phil.bartie@stir.ac.uk (P. Bartie).

http://dx.doi.org/10.1016/j.compenvurbsys.2017.09.010
Received 20 June 2017; Received in revised form 27 September 2017; Accepted 27 September 2017
0198-9715/ © 2017 Published by Elsevier Ltd.

evaluate pedestrian level positioning and tracking in the urban environment; (4) to determine the optimal balance in the delivery of 'pushed' information and user requests alongside navigation instructions.

2. Background

The increase in processing power of mobile devices has enabled a new generation of mobile spatial interaction (MSI) (Carswell, Gardiner, & Yin, 2010) allowing users to interact more easily with relevant digital information in their surrounding environment. For example IT solutions exists that enable users to navigate and find out the name of a landmark using an Augmented Reality application (Chung, Pagnini, & Langer, 2016; Gu & Gu, 2016; Liarokapis, Mountain, Papakonstantinou, Brujic-Okretic, & Raper, 2006; Narzt et al., 2006). Smartphones have become the most suitable candidate for MSI because of their combination of 1) small form, 2) positioning capabilities and other sensors (acceleration, barometer, gyroscope), 3) data transfer via mobile networks, and 4) sufficient battery power for a day of use.

When people explore an unfamiliar environment, particularly a cityscape, they spend a lot of time looking around. Most tourist systems require the user to interact with a graphical interface, which distracts them from appreciating the city or to paying attention to obstacles in their path (Heuten, Henze, Boll, & Pielot, 2008). Speech, on the other hand is one of the most natural forms of interaction, and is particularly suitable when a user is carrying out tasks that occupy their view (e.g. driving, walking) or makes physical demands (e.g. opening doors, carrying, physically aiding others). The ambition of SpaceBook was to focus on this speech interaction in order to build a hands-free, eyes-free application that enabled users to explore and be guided around a city. The system used speech as its only user interface, without any use of the phone's display, such that the phone remained concealed in a pocket or bag. Interaction was via headphones and microphone. Speech interfaces in industry such as Siri, Cortana, Alexa, and Google Assistant, normally respond to only a single utterance from the user, such as a web search or command, with minimal follow-up conversation. In contrast, the SpaceBook system deployed a Spoken Dialogue System where multi-turn sequences of interaction are employed in a long-running conversation with the user, lasting many minutes. Such dialogue based interaction can better reflect the collaborative nature of exploratory learning, with well understood interaction benefits (Cai, Wang, MacEachren, & Fuhrmann, 2005). However this conversational style of interaction (Lemon, 2012) is challenging because it has to perform tasks and meet the user's goals across long sequences of turns, maintaining an accurate representation of the context at all times. Standard research on dialogue systems focusses on single tasks, such as restaurant search (Young, Gasic, Thomson, & Williams, 2013) or flight booking, where the user's location is static. The dialogue system for SpaceBook had to manage a much more challenging situation, with multiple tasks (e.g. navigation, Points-of-Interest, Question-Answering) within a dynamic location-based system, and so it constitutes one of the most complex spoken dialogue interfaces yet created. The main novel contribution of the SpaceBook dialogue system is its location-based Interaction Manager (see Section 4.7) (Janarthanam, Lemon, Bartie, et al., 2013; Janarthanam, Lemon, Liu, et al., 2013) which handled multiple conversational threads (Lemon & Gruenstein, 2004). The system also used a continuous speech recogniser, which was always listening to the user, rather than a push-to-talk system or one triggered by `hot-words' such as "OK Google" or "Alexa", which are used in current commercial systems.

Location Based Services (LBS) enable more effective system interactions by automatically including the user's location in the search. The pioneer of LBS was Cyberguide (Long, Aust, Abowd, & Atkeson, 1996) which could calculate its location indoors using infrared beacons and outdoors using the Global Positioning System (GPS), providing location customised information to tourists. The system demonstrated that mobile computing was able to usefully adapt information delivery based on location and place histories, offering an alternative to a human tour guide. A similar system for blind pedestrians was designed by R. G. Golledge, Klatzky, Loomis, Speigle, and Tietz (1998), and evaluated by Loomis, Golledge, and Klatzky (1998), that proposed speech based input coupled with spatialized sound to convey information about the immediate environment. A wide variety of location aware applications followed, including GUIDE a virtual guidebook (Davies, Cheverst, Mitchell, & Friday, 1999), GEONOTES for attributing space using virtual tags (Espinoza et al., 2001), way-finding applications (A. J. May, Ross, & Bayer, 2005; Andrew J. May, Ross, Bayer, & Tarkiainen, 2003), friend finding (Strassman & Collier, 2004), urban gaming (Benford et al., 2006), and EASYGO for public transport information (Gartner et al., 2007).

The initial uptake of LBS by the population was fairly slow, which can be partially attributed to poor user experiences, service unreliability, and a lack of perceived ownership benefits (Chincholle, Goldstein, Nyberg, & Eriksson, 2002). Furthermore many potential users were concerned about issues of privacy (Duckham & Kulik, 2006) and security (Cahill et al., 2003). Their established ubiquity has arrived with mobile computing platforms that are continuously geolocated using solutions such as Global Navigation Satellite Systems (GNSS) and WiFi positioning, with freely available location aware applications (e.g. Google Maps, TripAdvisor, AirBnB). Most applications use a simple measure of distance to determine geographical relevance when filtering information. For example, the closest park determined in Euclidean space, or the nearest supermarket using network space. People, however, often refer to items in vista space (Montello, 1993), defined as the region visible from a location (Fig. 1).

Augmented Reality applications (e.g. Layar, Wikitude) overlay name labels on an image captured from a smartphone's camera to notify the user of surrounding features, which is a development of the point-to-query interaction pioneered by Geowand (Egenhofer, 1999), where the user's position and the orientation are used to control the data filter. A more recent example of this which includes a visibility filter, is Zapp (Meek, Priestnall, Sharples, & Goulding, 2013) which runs on a smartphone allowing the user to discover information about things in their line of sight, such as the geology of a distant hill. Prior to this visibility modelling was included in the Edinburgh Augmented Reality System (EARS) (Bartie & Mackaness, 2006) which through a speech interface supported pedestrian urban exploration. EARS announced landmarks as they came into view, enabling the appropriate audio keyword so that the user could ask for more details about any previously announced landmark by name via the speech interface. The system supported free exploration but not navigation, and the visibility information was pre-calculated for 86 selected landmarks within the pilot study region (Edinburgh city). There are other examples of systems which avoid visually distracting the user, such as through haptic feedback (Heuten et al., 2008) or abstract sound (S Brewster, 1997; S. Brewster, 1998) but these tend to be limited in what they can communicate and are not intuitive as they require the user to learn how the interface presents information.

3. Design aspects

Various forms of information are required to support the process of exploration which is a complex dynamic process. Urban exploration is about having the capacity to roam freely, retaining a sense of where you are, while acquiring spatial knowledge and a sense of place (Reginald G. Golledge, 1992). Wayfinding is just one component of exploration; much has been written on the role that landmarks play as confirmatory cues, or at key decision points along a route (Duckham, Winter, & Robinson, 2010; Richter & Winter, 2014; Sorrows & Hirtle, 1999). Augmented information supporting this process often comes in map form (with its associated cognitive effort). However in dialogue based systems the cognitive effort comes in providing useful

Download English Version:

https://daneshyari.com/en/article/4965110

Download Persian Version:

https://daneshyari.com/article/4965110

Daneshyari.com