Computers, Environment and Urban Systems 57 (2016) 118–129

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Practical simulation of virtual crowds using points of interest

Matti Pouke *, Jorge Goncalves, Denzil Ferreira, Vassilis Kostakos

Center for Ubiquitous Computing, University of Oulu, Finland

ARTICLE INFO

Article history:
Received 23 May 2015
Received in revised form 4 February 2016
Accepted 11 February 2016
Available online xxxx

Keywords:
Smart cities
Virtual environments
Pedestrian modelling

ABSTRACT

We present a computational method that exploits points of interest (POIs) to generate realistic virtual pedestrians for a city model, i.e., a simulated crowd. Our method is validated using mobility traces collected longitudinally from a city-wide free and open Wi-Fi network in downtown Oulu, Finland. Analysing this data, we first construct a time-varying Origin–Destination matrix that describes how individual pedestrians in our city move at different times and places. We compare this ground-truth against a random pedestrian model to investigate how the latter underestimates or overestimates movement at various locations or times of day. By identifying these deviations, we can calibrate a weighted model that uses POIs from OpenStreetMap to adjust the simulated crowd. Our results show a significant accuracy improvement over the random model, while at the same time our work is readily applicable to simulating crowds in other cities (real and virtual) as long as POI can be defined spatially.

1. Introduction

Understanding how humans navigate in urban spaces has been the interest of scientific disciplines ranging from psychology through civil engineering to computer science. The specifics of the navigation depend on how an individual understands the physical setting in which the navigation takes place. How the physical setting is specifically understood always depends on each individual. However, environments such as cities contain properties which are more or less common to all individuals navigating them. Kevin Lynch called "the public image" of a city to its collective understanding (Lynch, 1960). This public image is the overlap of the individual images, and according to the study data of Lynch et al., this can be described consisting of the following properties: *paths*, *edges*, *districts*, *nodes and landmarks.* Paths are the channels through which individuals move. Edges divide areas physically or metaphorically. Districts are geographical medium-to-large areas of the city. Nodes are strategic points that can be street junctions or other types of crossings or convergence of paths. Landmarks are distinctive physical points, which can be observed visually such as churches, parks, skyscrapers or mountaintops (Lynch, 1960).

The acquisition of spatial knowledge, both metric and qualitative, begins as soon as an individual starts navigating in a new environment (Montello, 1998). Through increased exposure and familiarity, the quality of the spatial knowledge increases, and with enough exposure, an individual can connect separately learned places in a larger spatial understanding (Montello, 1998). It is suggested that individuals organise spatial knowledge according to *anchor points*, salient locations that form the cognitive map that the individual uses to navigate. Besides geographical points, such as landmarks, anchor points can be path segments, nodes or even distinctive areas, similar to city properties categorized by Lynch (Lynch, 1960), (Golledge & Gärling, 2002). Anchor points of an individual usually consist of home, work place (or similar) as well as other locations that are somehow meaningful for the navigation of said individual. Especially striking locations, such as famous landmarks might be common to almost every visitor of the city while some anchor points are shared between various demographics (Couclelis, Golledge, Gale, & Tobler, 1987; Golledge & Gärling, 2002; Golledge & Spector, 1978).

As individuals navigate between places, they usually associate locations according to their closest anchor points (Couclelis et al., 1987). These associations form regions connected to anchor points and individuals tend to displace these regions within the direction of the anchor points, causing metric distortions in their mental models (Couclelis et al., 1987). This is related to the findings of human tendency to store spatial information according to hierarchical categorization (Couclelis et al., 1987). A largely cited example is the experiment of Stevens and Coupe (Stevens & Coupe, 1978), where US citizens systematically considered the city of Reno, Nevada to be east of San Diego, California because the state of Nevada is associated to being east from California (Stevens & Coupe, 1978).

The empirical macro-scale analysis performed by Manley et al. (Manley, Addison, & Cheng, 2015) supports the theory that anchor points play a dominant role in urban navigation. Their findings from observing 700,000 minicab routes within London suggest that urban anchor points are more suitable for basis of urban travelling models

* Corresponding author at: Center for Ubiquitous Computing, Erkki Koiso-Kanttilan katu 3, door E, P.O. Box 4500, FI-90014, University of Oulu, Finland.
E-mail addresses: matti.pouke@ee.oulu.fi (M. Pouke), jgoncalv@ee.oulu.fi (J. Goncalves), denzil.ferreira@ee.oulu.fi (D. Ferreira), vassilis@ee.oulu.fi (V. Kostakos).

http://dx.doi.org/10.1016/j.compenvurbsys.2016.02.004

instead of cost-minimisation by shortest routes. Accordingly, the minicab drivers rarely followed route choices given by shortest path routing methodologies but instead repeatedly established common anchors in route choices (Manley, Addison and Cheng, 2015).

As stated in (Golledge & Spector, 1978), while anchor points are always subjective, groups of people might share some anchor points. In a city scale this means that, college students might share their campus as a common anchor point as well as people living in the same neighbourhood might share anchor points related to that neighbourhood. However, the work of (Bromley, Tallon, & Thomas, 2003) points out that different demographics can also be segregated according to time, at least when examining inner city use. Daytime visitors of inner city visit different locations than night time visitors and often belong to different demographics.

In this work, we examine the use of crowdsourced points of interest (POIs) as a means to generate anchor points into city microsimulations. The rationale for exploiting POIs as an alternative for such a purpose, is the ease of which they can be acquired. Our attempt is not to provide a complete microsimulation system that would provide alternatives to complete activity-based microsimulation models, such as the social force model (Helbing & Molnar, 1995). Instead, we study POIs in isolation as time varying inner city anchor points. This can help, for example, in identifying after-hour crime hotspots (Nelson, Bromley, & Thomas, 2001). Here, temporal data is acquired from a time-varying Origin–Destination (OD) matrix, a result from analysing municipal Wi-Fi network user data. In addition to examining crowdsourced POIs as anchor point data, this work contributes to pedestrian microsimulation research by demonstrating how a municipal Wi-Fi network can be used to simulate granular pedestrian mobility.

2. Related work

2.1. A cursory overview of travel behaviour studies

Simulation of the total flow of pedestrians across a city has been repeatedly addressed in research concerning topological studies of street networks. Space Syntax has been used to describe how and why people move through certain routes within a city (Hillier & Hanson, 1984). Space Syntax literature suggests that roughly 70% of pedestrian volume at particular street segments can be predicted from the closeness connectivity metric of the street network, while Jiang and Jia (2011) claim that a weighted version of PageRank is a more suitable metric. They also argue that the underlying street network is the most influential factor in guiding pedestrian movements, and therefore randomly moving agents and real pedestrians move essentially in a same way through the same street network. A space syntax study by Lerman and Omer (2016) combined land use, physical properties of road sections as well as demographics information with street connectivity to study the relative contribution of each of the aforementioned dimensions to pedestrian movement. Their findings substantiate the claims of previous studies (Hillier & Hanson, 1984; Jiang & Jia, 2011); street connectivity was the most significant contributor to pedestrian movement and can in itself cause changes in the physical properties of road sections and land use (Lerman & Omer, 2016).

However, city topology alone is not sufficient to fully describe all travel behaviour aspects, and therefore travel behaviour has been studied across multiple disciplines such as geography, urban planning, transportation and even computer science research.

A land-use approach has been often used in transportation research since the early 20th century. It describes the characteristics of travel behaviour between different types of land use, such as the traffic between residential zones and industrial zones. Alan M. Voorhees (2013) described how travel between different types of origins and destinations roughly follows gravitational laws, with different types of destinations generating certain types of “pull” towards the origins. The amount of the pull in all types of destinations depends on the size of the origin and destination, as well as the travel time between them. However, how the theoretical pull is calculated, depends on the type of the travel, i.e., the types of the origin and the destination (Voorhees, 2013). This gravitational law is still often revisited in research, such as in the work of Simini, González, Maritan, and Barabási (2012); their radiation model can predict mobility patterns using population density as the only input data, eliminating the need for parameter adjustments. Land use has a different effect on various aspects of travel behaviour, such as trip generation, distance travelled and choice of mode (M. G. Boarnet, Joh, Siembab, Fulton, & Nguyen, 2011). Criticism towards the effect of land use to travel behaviour has been presented by for example Boarnet and Sarmiento (1998); Stead (2001) and Ewing, Deanna, and Li (1996). In any case, transportation planning and land use continue to meet heavily in research, as seen in (Waddell, Ulfarsson, Franklin, & Lobb, 2007).

Besides land use, it is possible to also consider smaller-scale segregation of locations into urban research, while time also plays an important role. Land use research usually focuses on travel behaviour between, for example, residential zones, industrial zones and shopping districts. However, there are also significant differences in the use of inner city locations according to the time of the day and week (Bromley et al., 2003).

Activity based models have often been used to estimate Travel Behaviour since the 1990s. Such models rely on the fact that people travel because they have needs, activities to which they must tend. How these activities are scheduled, given various conditions, such as household characteristics, properties of potential destinations and the state of the transportation network, is what activity based approaches seek to answer. However, activity based approaches have received criticism for their complexity and intense data requirements (Ettema & Timmermans, 1997). Activity-based models rely on realistic modelling on schedules of individuals and households which is not a simple task; properties such as speed of spatial knowledge acquisition and its role in scheduling decisions, assigning activities to utilities and interaction between household members are difficult to model as observed by (Kay W. Axhausen & Gärling, 1992). Since this observation, studies have emerged to tackle these problems. For example, the work of (Arentze & Timmermans, 2005) works in simulating spatial knowledge acquisition and (Zhang, Timmermans, & Borgers, 2005) model household interaction. However, intense data requirements of activity-based models are still a problem. It has been speculated whether it is even possible to gather exhaustive dataset for a truly precise activity based model (K W. Axhausen, 1998). While large-scale data collection efforts have been made, it is difficult to find a representative set of participants willing to commit to a long-term data gathering effort (K. W. Axhausen, Zimmermann, Schönfelder, Rindsfüser, & Haupt, 2002).

Since the 1990s, activity based approaches have been common in travel behaviour studies. In a 2001 survey, Timmermans categorized these approaches into: constraints based models, utility-maximizing models, computational process models and microsimulation models. Microsimulation models – such as ours – attempt to simulate individual activity patterns according to probability conditions, while the other approaches infer rules and parameters from empirical data (Timmermans, Arentze, & Joh, 2002).

Microsimulations can either simulate all aspects of activity based approaches or concentrate on certain properties. RAMBLAS (Veldhuisen, Timmermans, & Kapoen, 2000) and TRANSIMS (Nagel & Rickert, 2001) are examples of microsimulation models that replicate city-wide traffic according to multiple parameters. There are, however microscopic simulations that are not activity-based approaches, but simulate pedestrian activity at the micro-level movement of individual pedestrians concentrating only on detailed aspects of pedestrian flow. These models can effectively analyse and simulate pedestrian flows and interactions through narrow spaces with varying sets of rules such as lanes and pathways, estimate concepts from vehicular traffic such as level of service and estimate effects of various types of pedestrians, such as the obese or the elderly, on pedestrian flow (Galiza & Ferreira, 2012; Guo, Wong,