Landmarks as Beacons: Pedestrian Navigation Based on Landmark Detection and Mobile Augmented Reality

Jinmeng Rao[a]*, Song Gao[a], Yuhao Kang[a] and Qingyun Du[b]

[a] GeoDS Lab, Department of Geography, University of Wisconsin-Madison, Madison, USA
[b] School of Resource and Environmental Science, Wuhan University, Wuhan, China
* jinmeng.rao@wisc.edu

Keywords: Navigation, Augmented Reality, Location-Based Services, Mobile Computing, Object Detection

Introduction

The prevalence of cloud computing, mobile devices as well as ubiquitous sensors enables citizens to produce, acquire, and utilize geo-referenced data for various Location-Based Services (LBS). Navigation as one of these services has been widely used by people. Currently, most navigation methods are map-based approaches that depend on turn-by-turn instructions displayed in a map (Giannopoulos et al., 2015). Traditional practices are based on interactive charts (Rümelin et al., 2011) and turn direction concepts (Klippel and Montello, 2007). These map-based methods may lead to ambiguity (Amirian and Basiri, 2016) as people may have different spatial cognition and interpretations due to the isolation between the instructions and the environment (Rao et al., 2017). Landmarks are salient geographic objects when compared with objects in the surrounding area (Richter and Winter, 2014; Zhu and Karimi, 2015). Some studies in the field of spatial cognition showed that landmark, as one of the three forms of spatial knowledge, is of great importance for navigating humans in complex environments (Siegel and White, 1975; Werner et al., 1997; Zhu and Karimi, 2015). Users often tend to rely on landmarks during wayfinding since they are easy to recognize and usually situate along the route (Millonig and Schechtner, 2007). However, the conversion from salient landmarks into navigational cues remains a challenge. For instance, the actual visibility of some landmarks may be ignored when navigating, which may cause navigation interruptions (Millonig and Schechtner, 2007).

In this paper, we aim to present a new pedestrian navigation method that combines landmarks and mobile Augmented Reality (AR). As a promising technology, AR provides natural visualization and interaction modes and has been used in navigation (Amirian and Basiri, 2016), driving assistant system (Bolton et al., 2015), city browser (Madden, 2011), etc. By leveraging AR, we convert the landmarks into beacons by placing virtual arrows and information panels around them on the screen, and then these beacons can be used to guide users to their destinations step by step. Different from existing AR map navigation such as Google Map AR Mode, our method takes advantage of visually salient landmarks to help users better understand where they are and where they are heading to. In this paper, we focus on building landmarks (e.g., tall buildings or sculptures) since they are common and conspicuous in human activity areas, and often near roads. Compared with the traditional methods, our method preserves the human's actual feeling and perception of the environment, leading to a more natural and practical navigation process. Our mobile prototype and current experiment also indicate that the method can precisely detect and track specific landmarks, register virtual objects, and successfully navigate users using beacons in an intuitive manner. The proposed method can be applied in many domains such as smart transportation, location-recommendation, and tourism by providing navigation services with good user experience and rich location-based information.

Method

The method proposed in this paper mainly consists of three parts: landmark detection, registration, and AR-based pedestrian navigation. Figure 1 describes the framework of the method.

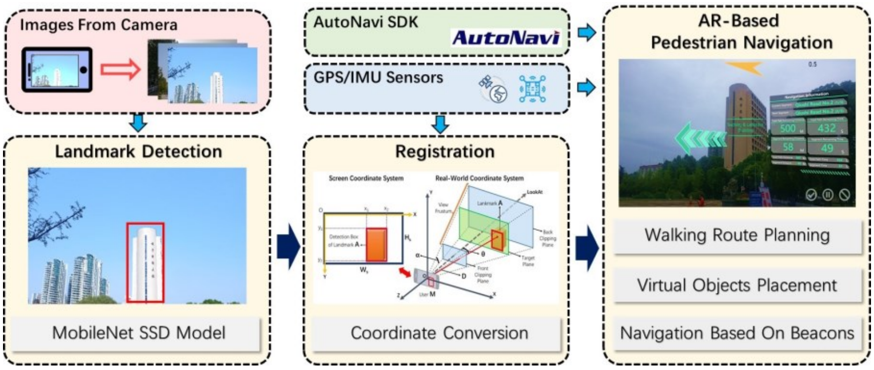

Figure 1: The framework of the method.

Landmark Detection

By utilizing the object detection method, the names and bounding boxes of landmarks on the screen can be obtained, and these landmark detection results can be utilized to register virtual objects (i.e., navigation arrows and panels) into the real world (i.e., the places around landmarks). We use MobileNet SSD, a deep learning model that combines MobileNet (Howard et al., 2017) and Single Shot Multibox Detector (Liu et al., 2016) to achieve fast and precise landmark detection. We manually selected 20 visually salient landmarks (tall buildings, statues, etc.) from the Wuhan University campus. In addition, we collected and labelled about 4000 images of these landmarks to establish a landmark detection dataset in Pascal VOC format for the model training.

Registration

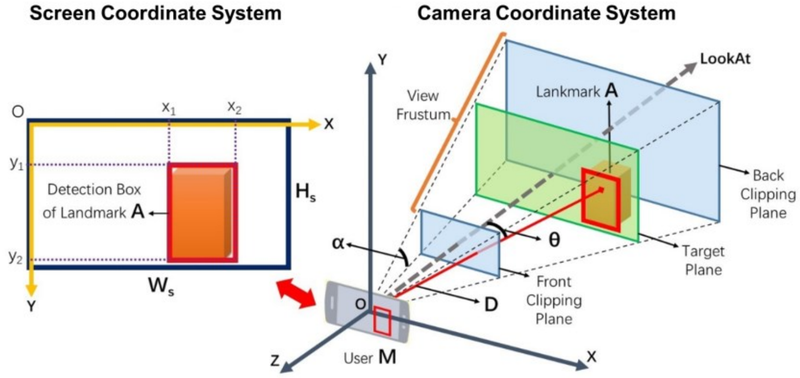

Figure 2: The conversion between coordinate systems (modified from Rao et al., 2017).

Registration means the correct alignment between the virtual object and the real world. To achieve precise registration, the coordinate conversion between Screen Coordinate System (SCS) and Camera Coordinate System (CCS) need to be determined. As is shown in Figure 2, the detection box of a landmark can be described by four pairs of coordinates in the SCS. For each pair of coordinates (X, Y) in the SCS, its corresponding coordinates (X_c, Y_c) in the CCS can be determined by Equation (1):

$$\begin{bmatrix} X_c \\ Y_c \\ 1 \end{bmatrix} = \begin{bmatrix} r & 0 & 0 \\ 0 & r & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -W_s/2 \\ 0 & -1 & H_s/2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{1}$$

where W_s and H_s are the width and height of the screen; r is the scaling factor defined by Equation (2):

$$r = \frac{2 \times \tan(\alpha/2) \times Z_c}{H_s} \tag{2}$$

where α is the vertical view angle of the camera. The Z value Z_c can be determined by Equation (3):

$$Z_c = D \times \cos(\theta) \tag{3}$$

where D is the distance between the user and the landmark; θ is the included angle between the camera LookAt direction and the landmark direction;

Thus, every screen coordinates (X, Y) of the landmark detection boxes can be converted to the corresponding camera coordinates (X_c, Y_c, Z_c), and the registration of virtual objects can be achieved based on these coordinates. In addition, we also use the Kernelization Correlation Filters (KCF) (Henriques et al., 2014) to achieve stable tracking in real-time after the registration.

AR-Based Pedestrian Navigation

In this paper, the navigation process is equivalent to the following task: Given a walking route R that consists of several road segments, the user will be guided to the ending point of the first road segment s_1 from the origin, and then guided to the ending point of the next road segment s_2. This process repeats until the user reaches the destination. In our method, when the user chooses a destination, a walking route (e.g., the shortest distance route) will be generated based on the user's current location and the destination. The user then can use the camera to detect the nearby landmarks, and the detected landmarks are retrieved from an online POI database (AutoNavi) according to their IDs. Once detected, the landmarks will be surrounded by virtual objects indicating where to go as well as other navigation information.

We use the POI database and route planning results provided by AutoNavi, a web mapping and navigation provider. Figure 3 depicts the design and placement of virtual objects. These arrows and information panels are used as instructions to guide directions, mark destinations, and display key navigation information such as road names, remaining time, remaining distance, etc. When a landmark is detected, its bounding box on the screen can be mapped into the camera view frustum (i.e., the red dotted box in Figure 3). Thus, virtual objects can be superimposed around the landmarks.

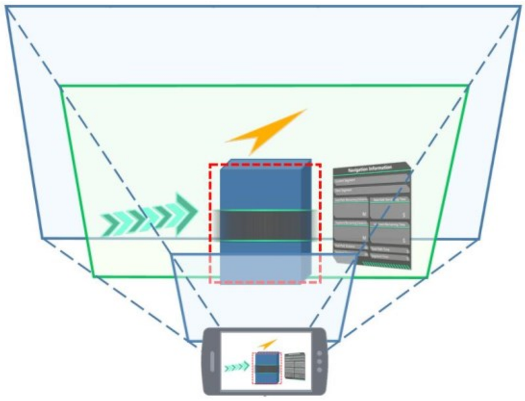

Figure 3: The design and placement of virtual objects.

Experiment and Result

We trained our landmark detection model on the collected landmark images. We used the MobileNet SSD model implemented in the MXNet GluonCV, a deep learning toolkit in computer vision. The model was trained with appropriate training parameters for 50 epochs on Google Colaboratory. As is shown in Figure 4, the training loss decreased significantly while the mean-average-precision (mAP) score on the validation set increased during the training process. The average precision scores in Figure 5 on the test set indicate that the model is able to precisely detect these landmarks on campus. Examples of landmark detection results can be found in Figure 6.

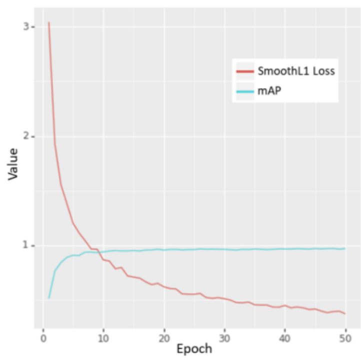

Figure 4: Curves of the loss and the mean average precision.

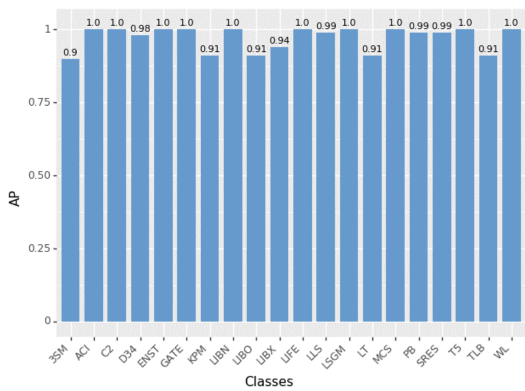

Figure 5: AP scores for all landmarks in our campus dataset.

Figure 6: Examples of landmark detection results.

We also developed a mobile AR prototype and conducted a pedestrian navigation experiment at the Wuhan University campus to validate the proposed method and to test its performance. The experiment area is shown in Figure 7. The user is first required to input a destination in the prototype system, then a walking route between user's current location and the destination will be calculated. The detected landmarks (in red) and the generated walking route (in blue) are also shown on the map. As can be seen in the screenshots in Figure 8, the user was guided step by step through the planned walking route and reached the destination in the end. The frame rate of the application is around 10 Frames Per Second (FPS), which indicates an acceptable time cost.

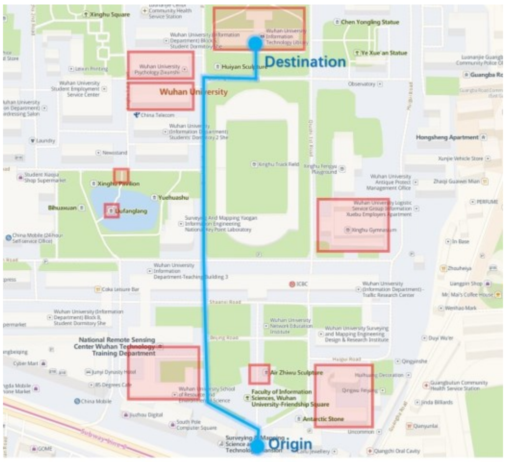

Figure 7: Experiment with the landmarks and the walking route between origin and destination.

Figure 8: Several screenshots from the navigation experiment using our developed mobile AR prototype.

Discussion and Conclusion

In this paper, we presented a novel AR pedestrian navigation method that uses landmarks as beacons to guide users' wayfinding. Specifically, landmarks are detected using a vision-based model and then augmented by virtual arrows and panels with navigation information. We developed a mobile AR prototype and conducted a pedestrian navigation experiment at the Wuhan University campus. The result shows that our method can well detect landmarks and guide users to their destinations using AR beacons in an intuitive manner. Also, the result shows the potentials for combining AR-based visual clues and map-based route planning to facilitate pedestrian navigation. Our method also has some limitations. For example, the landmark detection model is deployed on a cloud computing environment, which leads to network dependence. Also, the method requires collecting a large number of images of the selected landmarks to train the detection model, and the density of landmarks in the area may also affect the navigation performance.

Overall, the proposed method takes the user's visual concern and environmental perception into account and takes advantage of landmarks to achieve pedestrian navigation, which provides natural user experience. Future applications based on the method will focus on smart transportation, location-recommendation, and tourism, which could facilitate people and help them have a better understanding of the place which they visit or live in.

Acknowledgements:

Support for this research was provided by the University of Wisconsin - Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. This research was also supported by the National Natural Science Foundation of China (Project No. 41371427).

References

Amirian, P., & Basiri, A. (2016). Landmark-based pedestrian navigation using augmented reality and machine learning. In Progress in Cartography (pp. 451-465). Springer, Cham.

Bolton, A., Burnett, G., & Large, D. R. (2015, September). An investigation of augmented reality presentations of landmark-based navigation using a head-up display. In Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 56-63). ACM.

Giannopoulos, I., Kiefer, P., & Raubal, M. (2015, August). GazeNav: gaze-based pedestrian navigation. In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (pp. 337-346). ACM.

Klippel, A., & Montello, D. R. (2007, September). Linguistic and nonlinguistic turn direction concepts. In International Conference on Spatial Information Theory (pp. 354-372). Springer, Berlin, Heidelberg.

Henriques, J. F., Caseiro, R., Martins, P., & Batista, J. (2014). High-speed tracking with kernelized correlation filters. IEEE transactions on pattern analysis and machine intelligence, 37(3), 583-596.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Springer, Cham.

Madden, L. (2011). Professional augmented reality browsers for smartphones: programming for junaio, layar and wikitude. John Wiley & Sons.

Millonig, A., & Schechtner, K. (2007). Developing landmark-based pedestrian-navigation systems. IEEE Transactions on Intelligent Transportation Systems, 8(1), 43-49.

Rao, J., Qiao, Y., Ren, F., Wang, J., & Du, Q. (2017). A mobile outdoor augmented reality method combining deep learning object detection and spatial relationships for geovisualization. Sensors, 17(9), 1951.

Rümelin, S., Rukzio, E., & Hardy, R. (2011, October). NaviRadar: a novel tactile information display for pedestrian navigation. In Proceedings of the 24th annual ACM symposium on User interface software and technology (pp. 293-302). ACM.

Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. In Advances in child development and behavior (Vol. 10, pp. 9-55). JAI.

Werner, S., Krieg-Brückner, B., Mallot, H. A., Schweizer, K., & Freksa, C. (1997). Spatial cognition: The role of landmark, route, and survey knowledge in human and robot navigation. In Informatik'97 Informatik als Innovationsmotor (pp. 41-50). Springer, Berlin, Heidelberg.

Zhu, R., & Karimi, H. A. (2015). Automatic selection of landmarks for navigation guidance. Transactions in GIS, 19(2), 247-261.

Richter, K. F., & Winter, S. (2014). Landmarks: GIScience for intelligent services. Cham, Switzerland: Springer International Publishing.