Building Smart and Accessible Transportation Hubs with Internet of Things, Big Data Analytics, and Affective Computing

Jie Gong[1]; Cecilia Feeley[2]; Hao Tang[3]; Greg Olmschenk[4]; Vishnu Nair[5]; Zixiang Zhou[6]; Yi Yu[7]; Ken Yamamoto[8]; and Zhigang Zhu[9]

[1]Dept. of Civil and Environmental Engineering, Rutgers Univ., New Brunswick, NJ 08901.
[2]Dept. of Civil and Environmental Engineering, Rutgers Univ., New Brunswick, NJ 08901.
[3]Dept. of Computer Science, City Univ. of New York, New York City, NY.
[4]Dept. of Computer Science, City Univ. of New York, New York City, NY.
[5]Dept. of Computer Science, City Univ. of New York, New York City, NY.
[6]Dept. of Civil and Environmental Engineering, Rutgers Univ., New Brunswick, NJ 08901.
[7]Dept. of Civil and Environmental Engineering, Rutgers Univ., New Brunswick, NJ 08901.
[8]Dept. of Computer Science, City Univ. of New York, New York City, NY.
[9]Dept. of Computer Science, City Univ. of New York, New York City, NY.

Abstract

Large transportation hubs are difficult to navigate, especially for people with special needs such as those with visual impairment, Autism spectrum disorder (ASD), or simply those with navigation challenges. The primary objective of this research is to design and develop a novel cyber-physical infrastructure that can effectively and efficiently transform existing transportation hubs into smart facilities capable of providing better location-aware services. We investigated the integration of a number of internet of the things (IoT) elements, including video analytics, Bluetooth beacons, mobile computing, and facility semantic models, to provide reliable indoor navigation services to people with special needs, yet requiring minimum infrastructure changes. Our pilot tests with people with special needs at a multi-floor building in New York City has demonstrated the effectiveness of our proposed framework.

INTRODUCTION

Transitional spaces such as bus terminals, train stations, airports, and multi-modal transportation hubs have become an increasingly important part of city's infrastructure as we are spending more and more of our lives in these spaces in today's ever more connected world. Transportation facility owners are facing growing challenges to accommodate the rising public travel demands while improving quality of service. Future transportation facilities need to be smart, providing efficient, high-quality, and equitable services to the increasingly diverse population. This is especially true for those gigantic transportation hubs because wayfinding in these facilities has always been challenges for people with special needs such as individuals with visual impairment and Autism Spectrum Disorder

(ASD) and people with difficulties in finding places, particularly persons unfamiliar with metropolitan areas.

In the United States alone, the visually impaired population has reached 6.6 million people and expected to double by 2030 (from 2010 figures) (Varma et al., 2016). According to Centers for Disease Control and Prevention (CDC), ASD is the fastest-growing developmental disorder affecting 1 in every 68 people in the US. One common and recurring obstacle that people from both groups face every day is navigation, particularly as related to mobility. Using public transportation services is the best way for them to travel. However, there are also significant hurdles in using them due to their challenges. In 2015, a study conducted at Rutgers University found that according to adult respondents on the spectrum and their family members, 35.1% of these adults with ASD have difficulty in determining directions/route (Feeley et al., 2015).

Table 1. Difficulty with Different Aspects of Walking

Difficult Aspects of Walking	Responses	Percent of Responses	Percent of Respondents
Difficulty determining directions/route	247	14.2	35.1
Crossing a street	290	16.7	41.3
Judging the distance and/or speed of	318	18.3	45.2
Walking in areas without sidewalks (on	193	11.1	27.5
Dealing with distractions while walking	282	16.2	40.1
Too many people on the sidewalk	64	3.7	9.1
Too many cars or too much traffic	257	14.8	36.6
Other, please specify:	86	5.0	12.2
Total	1737	100.0	NA

While new technologies can be included and integrated into the design and construction of new transportation hubs to make them smarter, retrofitting existing facilities to make them smarter will be a more cost-effective choice in highly developed urban settings. Current emerging mobile computing and IoT technologies, together with advances in computer vision techniques used in 3D localization and crowd analysis, will provide great opportunities in significantly improving navigation services as well as creating innovative approaches to accommodate passengers and customers. While the support for these kinds of projects is evident, few studies have systematically investigated the synergy of these technologies as a cyber-physical infrastructure to enable these services. Most studies have focused on individual technological solutions which tend to fail to deliver reliable services in large and complex transportation hubs. The purpose of this study is to explore deep integration of Internet-of-Things (beacons, surveillance cameras, facility models, and mobile devices), Big Data analytics (deep learning, localization, and computing infrastructure), and affective computing (cognitive computing) as a novel cyber-physical system to build smart and accessible transportation hubs (SAT-Hub) capable of providing better location-awareness services (e.g. navigation support) to all, especially to people with disabilities (visual impairment and ASD) and people with navigation challenges. This paper describes and presents preliminary results on a novel cyber-physical infrastructure framework that can effectively and efficiently transform existing transportation hubs into smart facilities that are capable of providing better location-awareness and personalized navigation services (e.g. finding terminals, improving travel experience, obtaining security alerts) to the traveling public, especially for the underserved populations including those with visual impairment, ASD, or simply navigation challenges.

RELATED WORK

People who are have normal vision rely almost exclusively on their sight to orient themselves in a new indoor environment. As for people with visual impairment, eyesight is not a useable or reliable perception means, and they need to use alternative sensory tools to collect information to explore the environment. In spite of this need, the majority of the tools available to this population of people are not able to tell them their locations accurately, not even for navigation. For example, a white cane can help them to determine whether an area is walkable or not, but it cannot provide users their location information. Guide dogs may help to lead users to walk along known paths, but users still need other information to reason their locations when they want to change their routes, let alone to say owning a guide dog is expensive. GPS is sometimes used for localization in outdoor environments, but GPS signals can rarely be detected indoors or in dense urban areas because GPS signals are weakened and scattered by walls, roofs, and other obstructions (Agarwal et al., 2002).

Similarly, ASD individuals welcome technological solutions in order to overcome many of their daily obstacles. Among those obstacles, one common and recurring obstacle is navigation, in particular in indoor settings. Outdoor areas have signs, maps, and GPS-based navigation systems that can help a person navigate to their destination, whereas indoor navigation is often proved to be a much more difficult task. Because of this, lack of adequate navigation capabilities has limited their opportunities to use public transportation services. In many circumstances, ASD individuals may get lost or are unable to find their destinations in a complex building. In situations like these, not all ASD individuals are comfortable enough to seek help from strangers due to several reasons like communication difficulties, language problems, or social issues.

In recent years, researchers and several startups have been working on indoor GPS systems such as WiFi- or Bluetooth-based navigation approaches, and a few public facilities even have tested such approaches (i.e. SFO airport) (Indoor.rs, 2016). Some have proposed localization using the magnetic field (Li et al., 2012) while others have suggested using accelerometers and compasses on mobile devices in order to detect the speed and direction of the user (Collin et al., 2003). However, these methods are very much prone to error and may not be supported by all devices. Recently, localization using Bluetooth Low Energy (BLE) beacons has emerged as a viable method of positioning considering its wide availability and low cost (Gruman, 2014). However, these approaches with fixed beacons often require large-scale infrastructure changes and tedious sensor (beacon) installations and calibrations. These kinds of requirements are very costly and difficult to meet in large public transportation centers, all of which having frequent 3D environment changes and large volume of passengers. Needed are approaches that would require minimum infrastructure changes and sensor installations. Semantic facility model-based navigation could be a potential solution. However, relying on semantic models alone would be problematic because these kinds of facilities are simply changing so fast and models become obsolete quickly.

Another means to provide indoor localization and navigation services is computer vision based approaches. Previous work (Hu et al., 2014) explored methods to process images by image matching and estimate the location information. However, image matches are error-prone in the indoor and urban environments with large textureless areas. Some other studies have explored using Structure from Motion (SfM) to create street 3D models in the outdoor environment and recognizing the places utilizing images from Internet (Sattler et al., 2015; Torii et al., 2015; Zeisl et al., 2015). Some researchers use Bag of Words (BoW) (Cao et al., 2016) or ConvNet features (Sünderhauf et al., 2015) to represent outdoor environments for localization. Among these studies, very few of them focus on indoor scenarios, especially for an assistive localization purpose. In addition, a practical SfM model heavily relies on the richness and distinguishes of environmental features extracted from the images, which is hard to use in environments where few features are available and detected features often tend to be repetitive in space.

The rise of mobile and wearable devices as ubiquitous sensors has greatly accelerated the advancement of both general computer vision research and assistive applications. Farinella et al. (Farinella et al., 2015) uses Android phones to implement an image classification system with DCT-GIST based scene context classifier. Some others apply Google Glass and develop an outdoor university campus tour guide application system by training and recognizing the images captured by Glass camera (Altwaijry et al., 2014). Paisios, a blind researcher, creates a smart phone app for the Wi-Fi based blind navigation system (Paisios, 2012). Manduchi proposes a sign-based way-finding system and tests the blind volunteers with smart phones to find and decode the information embedded in the color marks pasted on the indoor walls (Manduchi, 2012). However, in spite of the technology promise demonstrated in these studies, few research work exist on designing user-friendly smart phone apps for helping visually impaired people to localize themselves and navigate through an indoor environment.

PROPOSED APPROACH

Our proposed solution to provide reliable indoor navigation services in major transportation hubs relies on integration of Internet-of-Things (beacons, surveillance cameras, facility models, and mobile devices), data analytics (deep learning and localization), and affective computing (cognitive human factors). The proposed cyber-physical system is designed to require minimum infrastructure changes as it leverages existing cyberinfrastructures such as surveillance cameras, facility models, and mobile phones, and incorporates a minimum number of beacons to achieve reliable navigation services. Figure 1 shows four essential elements in our proposed framework. In the following, we detail the technical innovations in each of these elements.

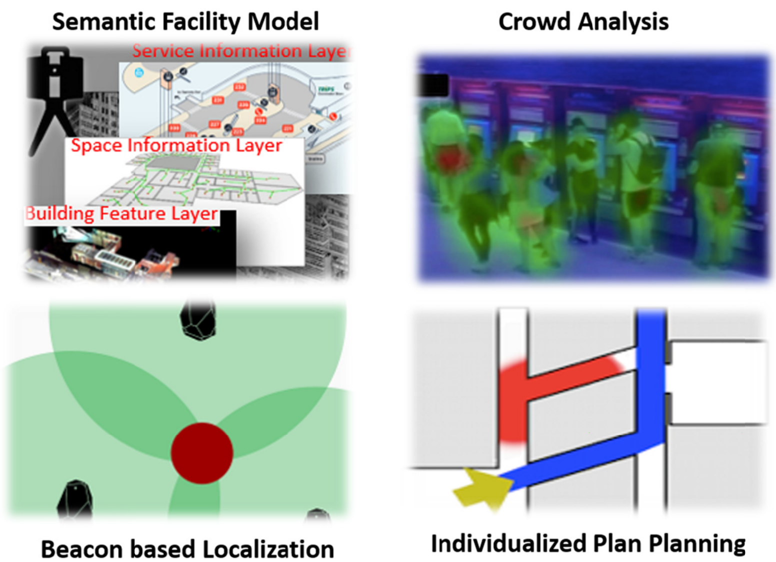


Figure 1. Proposed Framework to enable personalized indoor navigation assistance to people with special need

3D semantic facility model based localization

We use a 2D-3D registration approach to register smart phone images from multiple users with a pre-built semantic 3D facility model to infer absolute 3D locations of users. In our proposed framework, we develop a base framework in which facility users and the 3D semantic model of a facility can collaboratively work to realize robust and real-time localizations. To facilitate the process, we create a database of feature distributions for key positions in the facility based on the point cloud data and semantic facility model. We discard or discount those features that are from facility elements that will likely change over the time, and boost those with more permanent installations. The key positions will be determined based on how distinctly features distribute at selected vantage views from candidate key positions. We further assert that passive image capture and registration approach may not be effective whereas providing some directions to the users will greatly accelerate the converging process during localization because of two reasons: (i) Poor coverage—data collected from people without directions will likely have poor coverage of scenes with informative features; (ii) Data quality—without directions, data gathered is uncoordinated, resulting in low quality with more noise, making it difficult to process it, e.g., capturing the scene under different angles/positions, abruptly shaking device during capture, etc. This leads to the need of beacon based localization.

Beacon-based indoor localization

When vision based localization fails, beacon-based indoor localization is the backstop to ensure the availability of adequate navigation services in our proposed framework. However, apart from simply deploying a dense network of fixed Bluetooth beacons with known locations, a unique feature of the proposed work is the utilization of the 3D semantic model for the beacon installation. Both installing and calibrating beacons are tedious and challenging. Therefore, the use of 3D model will make the installation and maintenance of both fixed and mobile beacons more effective. 3D locations of beacons can be planned either interactively or automatically in the 3D digital model for the best coverage, and visualized in a virtual reality display for each installation. When a user comes into the facility, his/her App will be able to detect at least three of the beacons with known locations to obtain a relatively accurate location (from a meter to several meters). Then the 2D-3D registration approach will be used to further refine and track the location of the user.

Video based crowd analysis

A unique component of our proposed approach is integrating crowd analysis into indoor navigation services. Traditional indoor navigation services rarely consider contextual information when providing navigation guidance. However, this could be an important issue for people with special needs. For example, ASD individuals may prefer to choose paths that have less dense crowds due to psychological factors; people with visual impairment try to avoid large open space due to difficulty to find references for localization; and people in wheel chairs can navigate along paths with less crowds far more conveniently than along those with large crowds. In our proposed framework, we analyze the video feeds in real-time from surveillance cameras in the facility to evaluate the density of crowds in different parts of the facility. The analysis results will be incorporated into path choices.

Context-aware navigation guidance

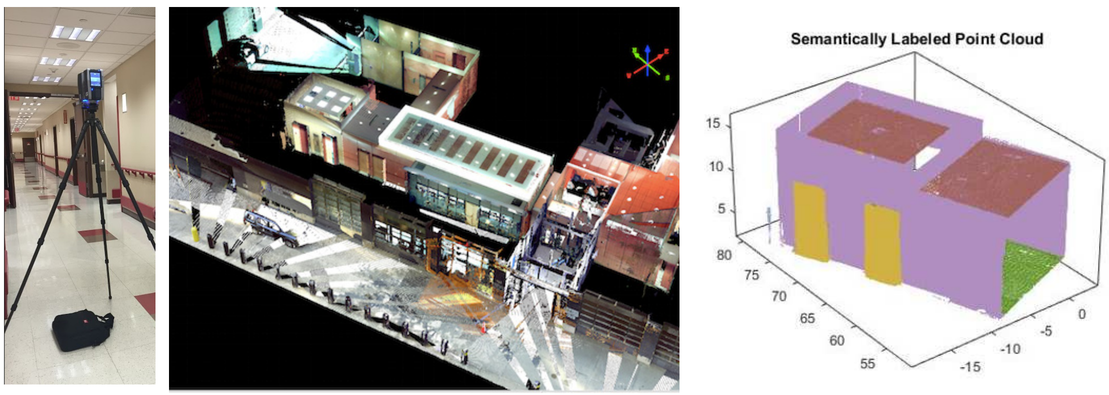
The proposed framework also includes a user-centric, activity-aware and feedback-enabled services with the support of the surveillance camera system to provide human crowd analysis results. In our framework, path planning for a user is made based on the following five factors: 1) Both the user's current location and his/her destination; 2) The user's planned schedule (for example the time to take a bus); 3) The special needs of the user; 4) The semantic 3D models with all the important facility labels; and 5) The crowd analysis results from the surveillance cameras. This is a graph planning problem with multiple cost attributes, and probably the graph and the path need to be updated if the path is not very short. As examples, a visually impaired or wheelchair user should avoid stairs. We will also need to adapt the path based on user's feedback. If an ASD user gets stuck and panic at certain location, the App will need to re-route the path, probably will also need to put them to wait if certain areas that they have to pass are too crowded and their time still allow them to wait.

PRELIMINARY RESULTS

In order to test the proposed framework, we have conducted pilot tests at a multi-floor facility in New York City. The facility has high definition surveillance cameras in place and it provides services to people with visual impairment. Estimote beacons (Estimote, 2016) are installed in the facility during our study. Although the facility was not a true transportation hub, it provides a great opportunity to test and validate our proposed approach with the easy reach to one of the groups we would like to provide services. The pilot provides foundational knowledge to expand our approach to transit stations and transportation hubs. In the following, we describe the development and testing of the framework at this facility.

3D semantic model and image model registration

As the first step, we utilized a terrestrial laser scanner (Figure 2a) to create a high-fidelity 3D model of the facility. The facility is represented with colorized 3D point cloud (Figure 2b) with dense annotation of building elements (Figure 2c). The creation of dense annotation is realized with a semi-automated segmentation and labeling tool developed as part of this project. Basically, the tool segments the point clouds through a region growing method (Rusu, 2010) and the segmented point clouds are manually annotated.



(a) (b) (c)

Figure 2. (a) Facility modeling with terrestrial laser scanning; (b) Colorized point cloud data of the facility; (c) Point clouds with dense annotation of building elements (in this case, elevator doors)

In this component, we also investigated registration of mobile phone image of the user with the 3D semantic model to provide user more accurate location and orientation information to get to his/her desired location. The registration between mobile images and facility point cloud data is solved by determining the projection between corresponding pixels/points. Denote a point as $C = [X, Y, Z, 1]^T$, and a pixel as $c = [u, v, 1]^T$. The projection from a 3D point on to a 2D pixel could be expressed as:

$$c = \boldsymbol{A}[\boldsymbol{R}|\boldsymbol{t}]C \qquad (1)$$

Where A includes intrinsic camera parameters, R and T are extrinsic camera parameters, including rotation and translation of the camera, according to the reference coordinate. Figure 3 shows alignment of a user view of the elevation from his mobile phone with the point cloud data.

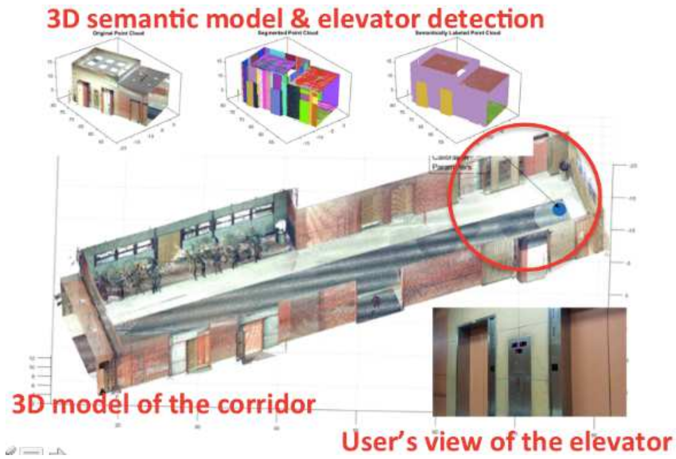


Figure 3. Registration of mobile phone image with the 3D semantic model

Deep learning for crowd analysis

We have been studying deep learning methods to improve the accuracy of crowd density estimation for the low- or mid- density crowd (Zhang et al., 2015), and tackle the high-density crowd using a regression-based method (Lempitsky and Zisserman, 2010; Chen et al., 2012). So far we have obtained very promising results on both crowd counting and crowd density estimation (Figure 4) based on convolutional neural networks (CNNs) (Jia et al., 2014). Though other convolutional neural networks have been used for crowd detection (Zhang, Li, Wang and Yang, 2015), our proposed pixel-wise calculation structure of the neural network is novel for the application of crowd density detection. From a high level perspective, the program would take as input a single color frame from the surveillance footage and output a form of "heat map" showing where people are at in the image and how many people there are. The "heat map" visualizes the count of the number of people per pixel of the image. Since a person takes up more than one pixel – and the sum of the total values within the body of a person is 1, the value per pixel is low. Where multiple people are occluding one another, we expect a higher value in that area. That is, even though that specific pixel only shows part of one person, the program should use surrounding pixels to determine that one person is occluding another. From this, any portion of the image can be considered, and within that portion the count of people can be determined. Additionally, these values can be averaged over time to compare the density of people per period of time.

The detection process is performed by a convolutional neural network. A brief explanation of how this works is as follows: an artificial neuron (which exists in the form of code) "looks" at the values in a tiny patch of pixels in the image. Each neuron has a certain pattern of values it "likes" to see in this patch. The closer the patch matches what the neuron likes to see, the higher value the neuron itself outputs. Additional layers of neurons then look at the output of the previous layers, themselves each liking their own pattern from that pervious layer. In this way, early neurons might like to see something like lines while the later layer neurons like to see combinations of lines in certain shapes. Finally, the entire network of neurons is made to like the appearance of people or groups of them. The neurons are trained to like the patterns they do, by training them on manually annotated data with density of people as described above. That is, known input is given to the network, the output is compared with the expected true output, and all the neurons are adjusted to more closely make the network's output match the expected output.

Our results showed the network had a prediction error of ~10% the count of people per image frame in the camera footage tested. This accuracy is acceptable for general statistics, the crowd avoidance navigation, and crowd simulation verification. This accuracy comes from a small training set of data (due to the large amounts of time required to annotate data). We believe accuracy would improve simply with more ground truth data without any improvements to the network itself. Figure 4 shows example detection results using an early version of the network being applied on publicly available data. The later networks were trained and designed confidential video footage at the facilities we were testing at. The later footage includes more challenging data, particularly in regards to occlusions.

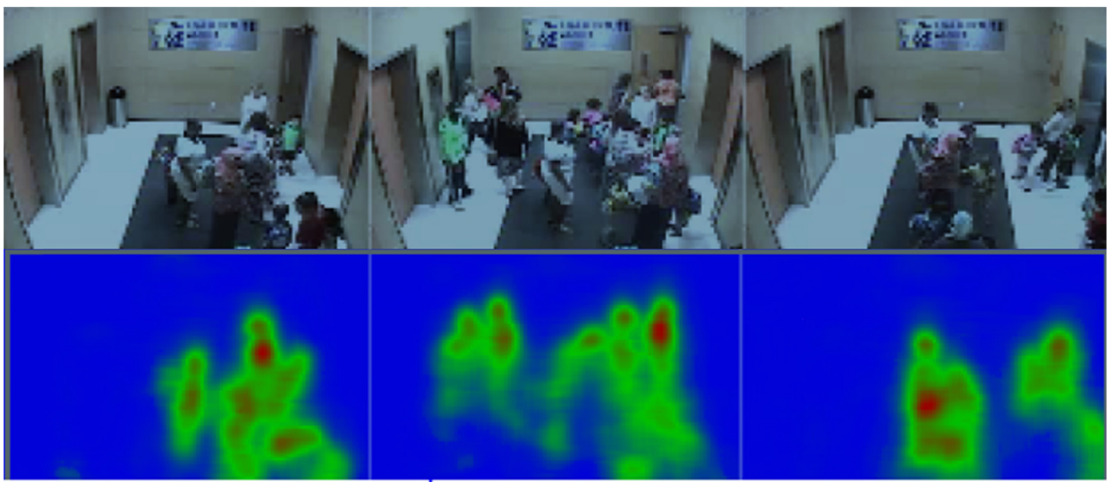


Figure 4. Crowd detection using deep learning

Beacon-based indoor localization

We installed Estimote Beacons in the facility to test the performance of beacon-based indoor localization. With Estimote beacons, we explored two methods of positioning using Bluetooth "beacons": trilateration and fingerprinting. Our goal was to determine which method would yield a position that was closest to the real position of the device. Trilateration works under two assumptions: (1) We know the ground truth positions of all of the beacons installed, and (2) the distances calculated using the received signal strengths are accurate. This second assumption is problematic because of the interference that may be caused by obstructions and other devices. Fingerprinting, on the other hand, is the process by which a "snapshot" of the area's radio landscape is taken before localization is actually done (Subhan et al., 2011). Fingerprint-based localization involves comparing the current radio conditions around the device with this snapshot, which consists of multiple "fingerprints." Whereas trilateration required a very high accuracy (for the RSSIs - received signal strength indicators) in order to precisely determine a position, fingerprinting naturally assumes that the RSSIs are error-prone. This is reflected in the algorithm, which defines a margin of error for the measured data RSSI in relation to the fingerprint RSSI. Furthermore, the algorithm also assumes that the client may naturally miss one of the beacons in the fingerprint (potentially from walking around or due to congestion). Thus, we are able to almost guarantee that a position will be computed and that this calculated position is very near to the real position. The resulting fingerprint map using three Estimote Beacons (Estimote, 2016) is shown in Figure 5. By comparison between the two approaches, the fingerprinting is a very viable and very robust method of localization and is a preferred approach to provide location-based services inside large, complex transportation hubs.

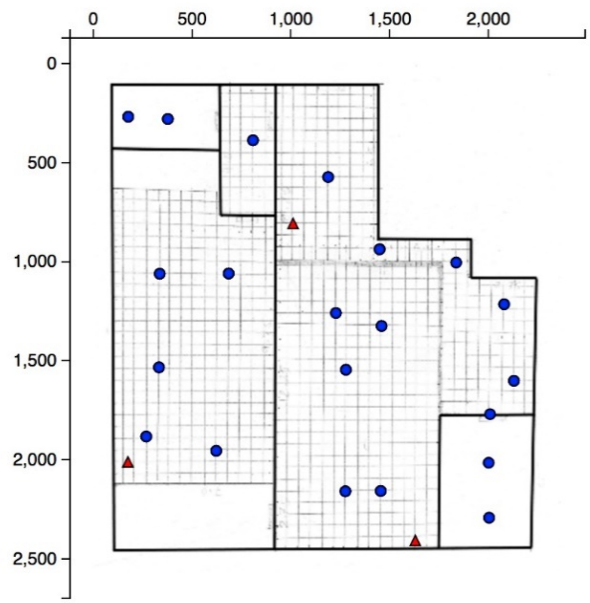


Figure 5. Server-generated map of fingerprints (blue circles) and beacons (red triangles) in the test area. Grid lines on hand-drawn floor plan represent tiles on floor. Axes represent pixel coordinates. 76 fingerprints were taken at 21 locations (average: ~3.6 fingerprints per location). In this visualization, the unit of both axes is in pixels.

Path planning and navigation assistance

The path planning element is encapsulated in a mobile application which leverages user location information (computed from the registration of image captured by user and the 3D facility model, and beacon-based localization), semantic facility model or simply a floor plan of the facility, and crowd analysis results to make decisions on paths that consider users' personal need. The mobile application is capable of providing multi-mode sensory feedback such as vibration and voice to users to achieve assistive navigation. Figure 6 shows an example test scenario where a user used the mobile app to navigate our studied facility using both the beacon-based localization and floor-plan-based path planning algorithms we developed on Android smart phones. Our study has shown that the app is capable of providing personalized travel guidance utilizing semantic 3D model, crowd analysis results, and strategically placed beacons.

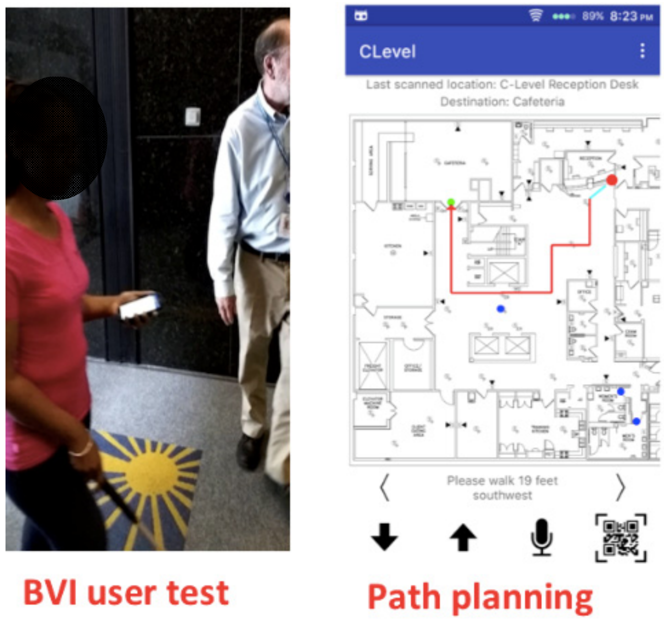


Figure 6. Beacon-based localization and floor-plan-based path planning.

CONCLUSION

This project investigated a novel cyber-physical infrastructure framework that can effectively and efficiently transform existing transportation hubs into smart facilities that are capable of providing better location-aware services (e.g. finding terminals, improving travel experience, obtaining security alerts) to the traveling public, especially for the underserved populations including those with visual impairment, ASD, or simply those with navigation challenges. We conducted our pilot test at a multi-floor building in New York City to evaluate the feasibility of our proposed framework. This initial test has demonstrated that it is feasible to integrate our proposed Internet of the Things elements (including video analytics, BLE beacons, mobile phone apps, and LiDAR-scanned 3D digital models) into a coherent framework to provide navigation services to people with special need. Future improvements would include using the 3D model to automatically determine information about the surveillance camera scene (such as camera pose and environment structure). This will not only improve the accuracy of the network, but more importantly provide a way in which the network can be generalized to all cameras in a facility without specific training the network to each individual camera. This could also make the network viable for completely different facilities and useable in any location, which will be our follow-on work. Future research will also focus on the best beacon and fingerprinting density and RSSI margin of error. Another area of interest is the best method for selecting the user’s current coordinates during fingerprinting. Existing services automatically assume that the location that the user selects on the floor plan is correct. However, there is no way for the user to actually know if they are correct or are off by inches or feet. Thus, a better method for self-localization during fingerprinting is also certainly a future area of research.

Lastly, it is in our team's agenda to test this framework in several public transit hubs in New York City and New Jersey.

ACKNOWLEDGEMENT

This research was performed under two appointments (one for Zhu/Gong and one for Tang) to the U.S. Department of Homeland Security (DHS) Science & Technology (S&T) Directorate Office of University Programs Summer Research Team Program for Minority Serving Institutions, administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and DHS. ORISE is managed by ORAU under DOE contract number DE-SC0014664. This work is supported in part by the National Science Foundation through Award EFRI -1137172, VentureWell (formerly NCIIA) through Award 10087-12, NIST/US Ignite Global City Teams Challenges(GCTC) . The work is also supported by Rutgers Center for Advanced Infrastructure and Transportation – a national University Transportation Research Center supported by USDOT. We would like to thank the following individuals for their partnerships in the GCTC effort: Barbara Campbell of the NYS Commission for the Blind, William H. Seiple of Lighthouse Guild, Ed Hoff of NJ Transit and Michael Lysicatos of Passaic County Department of Planning & Economic Development. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of DHS, DOE, NSF, USDOT, VentureWell or ORAU/ORISE

REFERENCE

Agarwal, N., Basch, J., Beckmann, P., Bharti, P., Bloebaum, S., Casadei, S., Chou, A., Enge, P., Fong, W. & Hathi, N. (2002), Algorithms for Gps Operation Indoors and Downtown, *GPS solutions*, **6**(3), 149-160.

Altwaijry, H., Moghimi, M. & Belongie, S. (2014), Recognizing Locations with Google Glass: A Case Study, *IEEE Winter Conference on Applications of Computer Vision*, IEEE, pp. 167-174.

Cao, J., Chen, T. & Fan, J. (2016), Landmark Recognition with Compact Bow Histogram and Ensemble Elm, *Multimedia Tools and Applications*, **75**(5), 2839-2857.

Chen, K., Loy, C. C., Gong, S. & Xiang, T. (2012), Feature Mining for Localised Crowd Counting, *BMVC*, pp. 3.

Collin, J., Mezentsev, O. & Lachapelle, G. (2003), Indoor Positioning System Using Accelerometry and High Accuracy Heading Sensors, *Proc. of ION GPS/GNSS 2003 Conference*, pp. 9-12.

Estimote. (2016), Esimote.

Farinella, G. M., Ravì, D., Tomaselli, V., Guarnera, M. & Battiato, S. (2015), Representing Scenes for Real-Time Context Classification on Mobile Devices, *Pattern Recognition*, **48**(4), 1086-1100.

Feeley, C., Deka, D., Lubin, A. & Mcgackin, M. (2015), Assessment of Transportation and Mobility Adults on the Autism Spectrum in Nj, NJ Department of Health.

Gruman, G. (2014), What You Need to Know About Using Bluetooth Beacons, *Smart User (blog), InfoWorld, July*, **22**.

Hu, F., Zhu, Z. & Zhang, J. (2014), Mobile Panoramic Vision for Assisting the Blind Via Indexing and Localization, *European Conference on Computer Vision*, Springer, pp. 600-614.

Indoor.rs. (2016), Indoors.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. & Darrell, T. (2014), Caffe: Convolutional Architecture for Fast Feature Embedding, *Proceedings of the 22nd ACM international conference on Multimedia*, ACM, pp. 675-678.

Lempitsky, V. & Zisserman, A. (2010), Learning to Count Objects in Images, *Advances in Neural Information Processing Systems*, pp. 1324-1332.

Li, B., Gallagher, T., Dempster, A. G. & Rizos, C. (2012), How Feasible Is the Use of Magnetic Field Alone for Indoor Positioning?, *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, IEEE, pp. 1-9.

Manduchi, R. (2012), Mobile Vision as Assistive Technology for the Blind: An Experimental Study, *International Conference on Computers for Handicapped Persons*, Springer, pp. 9-16.

Paisios, N. (2012), Mobile Accessibility Tools for the Visually Impaired, Citeseer.

Rusu, R. B. (2010), Semantic 3d Object Maps for Everyday Manipulation in Human Living Environments, *KI-Künstliche Intelligenz*, **24**(4), 345-348.

Sattler, T., Havlena, M., Radenovic, F., Schindler, K. & Pollefeys, M. (2015), Hyperpoints and Fine Vocabularies for Large-Scale Location Recognition, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2102-2110.

Subhan, F., Hasbullah, H., Rozyyev, A. & Bakhsh, S. T. (2011), Indoor Positioning in Bluetooth Networks Using Fingerprinting and Lateration Approach, *2011 International Conference on Information Science and Applications*, IEEE, pp. 1-9.

Sünderhauf, N., Shirazi, S., Dayoub, F., Upcroft, B. & Milford, M. (2015), On the Performance of Convnet Features for Place Recognition, *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, IEEE, pp. 4297-4304.

Torii, A., Arandjelovic, R., Sivic, J., Okutomi, M. & Pajdla, T. (2015), 24/7 Place Recognition by View Synthesis, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1808-1817.

Varma, R., Vajaranant, T. S., Burkemper, B., Wu, S., Torres, M., Hsu, C., Choudhury, F. & McKean-Cowdin, R. (2016), Visual Impairment and Blindness in Adults in the United States: Demographic and Geographic Variations from 2015 to 2050, *JAMA ophthalmology*.

Zeisl, B., Sattler, T. & Pollefeys, M. (2015), Camera Pose Voting for Large-Scale Image-Based Localization, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2704-2712.

Zhang, C., Li, H., Wang, X. & Yang, X. (2015), Cross-Scene Crowd Counting Via Deep Convolutional Neural Networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 833-841.