ICARCV 2016

The 14th International Conference on Control, Automation, Robotics and Vision

November 13 - 15, 2016
Phuket, Thailand

IEEE Catalog Number: CFP16532-USB
ISBN: 978-1-5090-4757-4

Introduction | Session Index | Author Index | Search

Myriam Vaca, Jose Emilio Traver, Vicente Milanés, Joshué Pérez, David González, Fawzi Nashashibi:
Automated global planner for cybernetic transportation systems. 1-6

Ehsan Mihankhah, Danwei Wang:
Environment characterization using Laplace eigenvalues. 1-6

Yuanzhe Wang, Danwei Wang, Ehsan Mihankhah:
Navigation of multiple mobile robots in unknown environments using a new decentralized navigation function. 1-6

Trung-Kien Dao, Thanh-Hai Tran, Thi-Lan Le, Hai Vu, Viet-Tung Nguyen, Dang-Khoa Mac, Ngoc-Diep Do, Thanh-Thuy Pham:
Indoor navigation assistance system for visually impaired people using multimodal technologies. 1-6

Dinh-Van Nguyen, Myriam Elizabeth Vaca Recalde, Fawzi Nashashibi:
Low speed vehicle localization using WiFi fingerprinting. 1-5

Ilankaikone Senthooran, Jan Carlo Barca, Hoam Chung:
A 3D line alignment method for loop closure and mutual localisation in limited resourced MAVs. 1-6

Wei Jing, Joseph Polden, Pey Yuen Tao, Wei Lin, Kenji Shimada:
View planning for 3D shape reconstruction of buildings with unmanned aerial vehicles. 1-6

Jose R. G. Braga, Haroldo F. de Campos Velho, Gianpaolo Conte, Patrick Doherty, Elcio H. Shiguemori:
An image matching system for autonomous UAV navigation based on neural network. 1-6

Asuka Kobayashi, Hironori Suzuki, Toshiki Hirogaki, Eiichi Aoyama:
Investigation of impact task performed by humanoid robot using input shaping control method. 1-6

Dayang Nur, Salmi Dharmiza, Awang Salleh, Emmanuel Seignez:
Monocular visual odometry with road probability distribution factor for lane-level vehicle localization. 1-6

Xiaowei Fu, Xiaoguang Gao:
Network building and communication strategy designing for multi-UAVs cooperative search. 1-5

Ming-kang Shi, Hong Jiang, Song-huan Li:
An intelligent traffic-flow-based real-time vehicles scheduling algorithm at intersection. 1-5

Ya Jun Yu, Zhan Yang, Beom-Seok Oh, Yong Kiang Yeo, Qinglai Liu, Guang-Bin Huang, Zhiping Lin:
Investigation on driver stress utilizing ECG signals with on-board navigation systems in use. 1-6

Yi-Min Hsieh, Yen-Chen Liu:
Model predictive control strategy for plug-in hybrid electric vehicles. 1-6

Djamila Dekkiche, Bastien Vincke, Alain Mérigot:

2016 14th International Conference on Control, Automation, Robotics & Vision
Phuket, Thailand, 13-15th November 2016 (ICARCV 2016)

Indoor Navigation Assistance System for Visually Impaired People using Multimodal Technologies

Trung-Kien Dao[1,*], Thanh-Hai Tran[1], Thi-Lan Le[1], Hai Vu[1], Viet-Tung Nguyen[1],
Dang-Khoa Mac[1], Ngoc-Diep Do[1], Thanh-Thuy Pham[1,2]
[1]MICA International Research Institute (HUST - CNRS/UMI2954 - Grenoble INP)
Hanoi University of Science and Technology, Hanoi, Vietnam
[2]University of Technology and Logistics, Bacninh, Vietnam
[*]Email: trung-kien.dao@mica.edu.vn

Abstract—**In this paper, a complete indoor navigation assistance system for visually impaired people is introduced. Different multimedia technologies are integrated in a single system in order to provide a precise, safe and friendly navigation service. First, the environment is modeled and represented. After that, the user location is determined by combining Wi-Fi and vision information. This combination offers some benefits in comparison with single technology systems such as setup cost, computational time and accuracy. Finally, the interaction between users and the system is performed through natural Vietnamese language with the support of Vietnamese voice synthesis and recognition. The proposed the system has been successfully deployed in a school for visually impaired pupils. Evaluation with various criteria on visually impaired pupils reveals the feasibility of the solution.**

Keywords—**indoor navigation, navigation guidance, visually impaired people, indoor localization**

I. Introduction

Mobility, the most involved activity in everyday life, relies mostly on the vision. Mobility in an unfamiliar environment is not always an easy task even for sighted people. Visually impaired people with a very limited (even not at all) sense of sight encounter a lot of difficulties in mobility and even fall in dangerous situations. Therefore, mobility aid has an important role in assistive technologies for visually impaired people. According to [1], there are two main components in mobility: obstacle avoidance, and orientation/navigation. The work introduced in this paper focuses on navigation aid for visually impaired people.

Nowadays, a number of navigation-assistance tools for indoor and outdoor environments have been proposed [2], [3]. Localization can be considered as a principal basis for these tools. In outdoor navigation, even when GPS has become a widely used solution, many issues still need be addressed when developing this solution for visually impaired people. Furthermore, since GPS is not always available for indoor environment, different techniques have been proposed for indoor navigation such as Wi-Fi, camera, RFID (Radio Frequency Identification), etc. [4], [5]. Taken into account that each technique has its own advantages and disadvantages, this work aims to improve the person localization by combining Wi-Fi and camera based localization technologies.

This study has three main contributions. First, a complete indoor navigation system for visually impaired people that includes environment representation, person localization, path finding and user interaction is introduced. Second, a combination of vision and Wi-Fi information is proposed for person localization. This offers some benefits in comparison with single-method systems such as setup cost, computational time, and accuracy. Finally, the proposed system has been deployed and tested with real visually impaired pupils in Nguyen Dinh Chieu School for Blind Pupils in Hanoi, Vietnam. The experimental results show the feasibility of the solution.

II. System Overview

The proposed system for navigation assistance is designed according to a layered architecture as illustrated in Fig. 1. It consists of four layers.

- *Physical layer* contains equipments and materials constituted the system: some environmental cameras for vision-based human detection and tracking, a hand-held smartphone with headset for user-system interfacing and Wi-Fi-based localization, a central server, a number of Wi-Fi access points ensuring wireless communication and at the same time acting as beacons for Wi-Fi-based localization.
- *Logical layer* provides the Application Program Interface (API) and communication.
- *Functional layer* consists of all software modules, which could be run online or offline, i.e., data acquisition (RSSI and video sequence), environment representation, multi-modal user localization, path finding, speech recognition and speech synthesis for user-system communication.
- *Application layer* contains applications based on provided functionalities of the system. In our work, the application provides navigational aids for blind users and is implemented on the smartphone.

To start, the user activates the navigation-assistance application on the smartphone. Then he makes his request for navigation by speaking to the microphone. This speech will be transmitted and analyzed on the server to determine the user's target location on the pre-built map. Then the multi-modal localization module is activated to continuously estimates the user position. With help of the environment map, the shortest path from user location to the target one can be determined. Navigational indication will be generated and sent to the

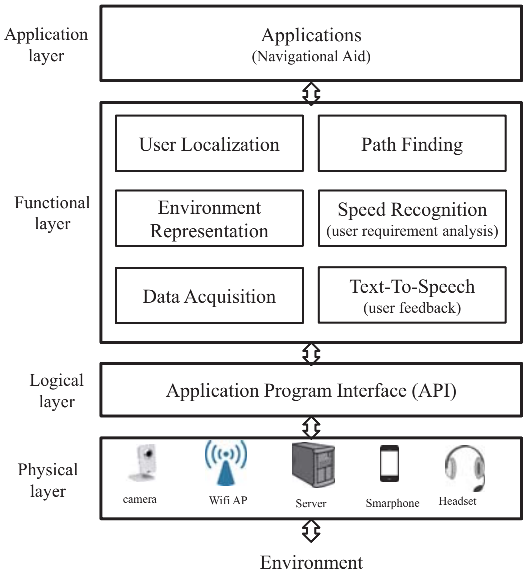


Fig. 1. Layered architecture of the proposed system.

smartphone and played as synthesized voice to the user. The localization - path-finding - indication process is repeated until the user and recompute the path until the user reaches the target location. In the following sections, the main components in the functional layer of the system are described in more detail.

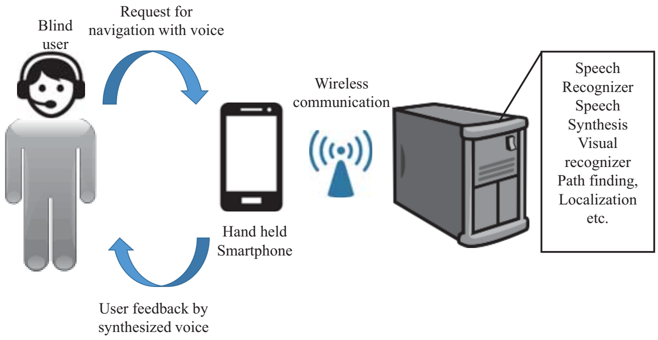


Fig. 2. Illustration of user and system

III. Map Representation and Path Finding

In a pervasive system, the physical world needs to be digitalized to a form of data structure which is called the environment model. There are a number of requirements for this model to be fulfilled depending on the nature of each specific application, including accurately mimicking of the physical environment characteristics, combining the environment with sensor networks or other artifacts of pervasive systems, and providing the data semantic to the system. Without informative and interactive models, the system capacity will be limited. In this study, the environment module is designed to (1) providing environment data for localization and navigation services (including semantic data of environment), (2) visualizing in 3D space, and (3) acting as a comprehensible interface for environment description. A structured XML model is targeted to achieve generic characteristics while preserving semantics and correctness of data.

Since the main target of an environment model is to describe human interaction with the environment, each entity in environment is managed and has its properties well defined. By modeling the environment in object-oriented databases, we can manage to describe interactions between objects, human and other factors such as time, conditions, etc. Starting from a CAD drawing of the environment, an automatic script will help to translate the drawing into desired structured XML under a prescribed schema as shown in Fig. 3. Resultant XML contains most of environment description but lack of semantics data as well as details of objects such as: walls height, area definition, area name or sensors details, etc. The developer will then need to manually add more detailed information to the description. The final XML document should follow the aforementioned schema and is processed automatically to transform into other database formats for other purposes.

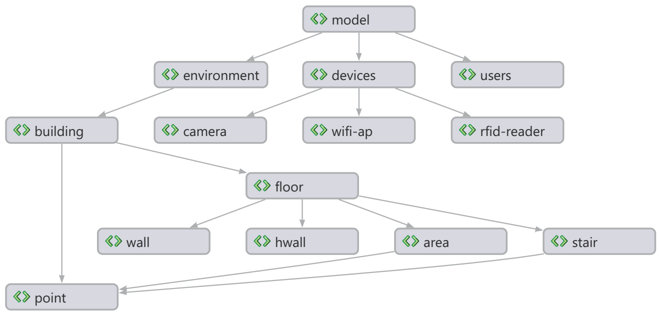


Fig. 3. XSchemas for XML description

In this work, the targeted environment is a dynamic one with multiple moving objects. One of the main services is path planning for user navigation, i.e., calculating the optimal path to navigate efficiently in the environment. Path planning is a popular research topic which could be solved using some variant of IDA [6], based on the well-known A* algorithm [7]. Most of the solutions approach the environment in uniformly shaped grids and making use of Voronoi subtraction/addition as well as geometric computation such as visibility graph in order to prevent collisions. The optimal path planning in dynamic environments require the system to be aware of moving obstacles in order to prevent collisions on their pre-calculated path. A technique to obtain information from the vector-based environment model and deriving optimal path-planning scheme in real time is proposed with geometric computation of a so-called visibility graph. The environment in this case is considered in 2D space with projections of objects (e.g., walls, doors, robots, obstacles, etc.) into floor planes. As a result, the environment at any given time is a non-simple polygon with holes representing obstacles. The optimal path in this case would be a path for the user to move inside the polygon without crossing holes.

IV. User Localization

Many localization systems with different architectures, configurations, accuracies and reliabilities have been proposed. In this study, a mechanism for multi-modal combination of multiple localization technologies proposed in [8] is applied, in which, Wi-Fi and camera are the main technologies being used. The location of users resulted from the localization service can be tracked through the platform.

A. Wi-Fi based

The first step is to model the signal power according to the distance, the following equation arises:

$$P(r) = P_0 - 10n\log\frac{r}{r_0} - k_d\sum_{i=1}^{n_w}\frac{d_i}{\cos\beta_i}, \quad (1)$$

where $P(r)$ is the Wi-Fi signal power at distance r from the Wi-Fi access point, P_0 is a reference power at distance r_0, n is the path-loss exponent, n_w is the number of obstacles (walls or floors) between the access point and the smartphone, k_d the attenuation factor per wall/floor thickness unit, d_i is the thickness of the i^{th} obstacles and β_i is the angle of incident on the i^{th} obstacles. These variables are shown on Fig. 4.

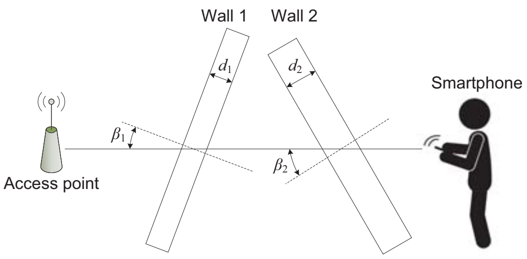


Fig. 4. Wi-Fi signal propagation through obstacles

Thus, from the measurement of $P(r)$, an estimation of r can be achieved, let's call it $\hat{r}$. In fact, the measure of r can be modeled as a normal probability distribution following:

$$\Pr(r|P) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(r-\hat{r})^2}{2\sigma^2}\right). \quad (2)$$

Moreover, σ and $\hat{r}$ are supposed to be proportional:

$$\sigma = k_{sigma}\hat{r}, \quad (3)$$

with k_{sigma} is constant.

B. Camera based

A number of methods have been proposed for human detection and localization using camera, among which the most cited human detector (HOG-SVM) is introduced in [9]. However, this detector works originally on still images and is significantly not computationally efficient, which is unsuitable for real-time applications. In an indoor-environment context, certain cameras are implemented at fixed locations with a supposition that all background objects are immobile and only people move in the scene. Relying on these assumptions, a background subtraction technique is applied to efficiently detect moving objects at the first stage. Then, regions including moving objects are checked to determine whether they contain humans using HOG-SVM detector. Specifically, the steps in Algorithm 1 are carried out for each camera. Fig. 5 illustrates intermediate results after each step of the process.

Algorithm 1: Algorithm of camera-based localization

Data: Current frame I, background model BG
Result: User location (x, y) in real world coordinates

Step 1: Background subtraction using BG [10];
Step 2: Noise and shadow removal using density-based score fusion scheme following learning approach [11];
Step 3: Human verification using HOG-SVM detector applied on extended moving regions [9];
Step 4: Human localization in world coordinates - the human foot point is considered and re-projected in the floor plane using calibrated parameters of the camera

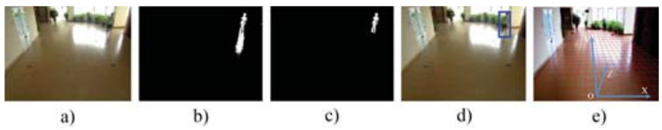


Fig. 5. Illustration of different steps of human localization: a) input frame; b) background subtraction result; c) shadow and noise removing result; d) HOG-SVM verification; e) human location in world coordinates (red dot)

C. Multi-modal Combination

To determine the user location, space is divided into grid points. At each point, the user appearance probability and precision are calculated with information provided from all available localization technologies. Since each localization technology provides results at different moments, the time past since the last result received from a technology to the current system time can affects the final result. Besides, each application may require a different acceptable precision, so on the basis of the given precision from each application, the location with highest probability is chosen as user localization result:

$$\mathbf{x}_{res} = \mathbf{x} \quad \text{if} \quad \Pr(\mathbf{x}) = \max_{i=1}^{k}\{\Pr(\mathbf{x}_i)\} \wedge \xi(\mathbf{x}) \leq \xi^*, \quad (4)$$

where $\Pr(\mathbf{x})$ and $\xi(\mathbf{x})$ are probability and precision at location $\mathbf{x}$, ξ^* is the acceptable precision given by the application, k is the number of grid points, and $\mathbf{x}_{res}$ is the localization result.

The probability at point $\mathbf{x}$ is defined as

$$\Pr(\mathbf{x}) = \prod_{j=1}^{k}\Pr_j(\mathbf{x})R_j\exp(-\lambda_j\Delta t_j), \quad (5)$$

where $\Pr_j(\mathbf{x})$ is the probability component corresponding to the j^{th} technology at point $\mathbf{x}$, R_j is reliability constant (i.e., camera-based results is more reliable compared to those of Wi-Fi based ones), λ_j is time decay constant of the j^{th} technology, Δt_j is the time difference from the last result current processing time, and k is the number of localization technologies in use. The precision at point $\mathbf{x}$ is defined as

$$\xi(\mathbf{x}) = \min_{j=1}^{k}\{\xi_j(\mathbf{x})\}, \quad (6)$$

where $\xi_j(\mathbf{x})$ is the precision of the i^{th} technology at point $\mathbf{x}$. For localization technologies providing results consisting of location and precision, $\Pr_j(\mathbf{x})$ can be extracted from normal distribution using the empirical rule:

$$\Pr_j(\mathbf{x}) = \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left(-\frac{d_j^2(\mathbf{x})}{2\sigma_j^2}\right), \quad (7)$$

where $\sigma_j = 3\xi_j$, $d_j(\mathbf{x})$ is the distance from point $\mathbf{x}$ to user location provided by the i^{th} technology.

The system, as illustrated in Fig. 5, uses results from multiple technologies as inputs and gives out user location to applications as output. Event-driving approach is used, i.e., each input will provide new localization results to ensure real-time availability. API module contributes as an open platform to extract useful information from inputs and store in the database. Information extraction module extracts necessary information from database as well as receives the precision limitation from applications then provides them to next module. Calculating module does the main processing function and produces user localization result. Depending on applications and their required precisions, results would be different. Afterwards, the outputs including user location, precision and probability are sent to applications and fed back to be stored in database for next calculation tasks.

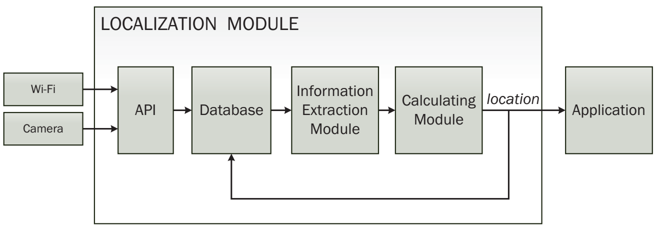


Fig. 6. General architecture of localization platform

V. USER-SYSTEM INTERACTION

The system is designed following a multi-modal multi-user model, in which, visually impaired users interact with the system by haptic buttons and speech. The interaction architecture is based on the PAC (Presentation-Abstraction-Control) pattern which is selected by his highly modularity, parallelism and distribution attributes [12]. This pattern models the whole system as a hierarchy of interactive objects, where *Presentation* entities define inputs and outputs perceived by users, *Abstraction* entities implement the functions that the system is able to perform, and *Control* entities maintain consistency between *Abstraction* and *Presentation* ones.

This system is composed of two main parts: Smartphone (which represents either smartphones or tablets) and Server. The Control entity of the Smartphone part is linked to the Control entity of Server part by a distributed messaging mechanism (using ZeroMQ [13]) through a wireless network.

A. Smartphone Part

The Smartphone part is modeled as an interaction object, as illustrated in Fig. 7. When the user presses a button and speaks a request (Presentation entity), the voice will sent to an Automatic Speech Recognition module (Abstraction entity) to analyze, validate and extract the desired destination. If user's request is valid, **MainActivity** (Control entity) will send a request with destination as parameter to Server for begin new navigation guide. Feedback information such as *where to go*, *how to go*, and *does user get his destination* will be achieved by speech synthesis and vibration.

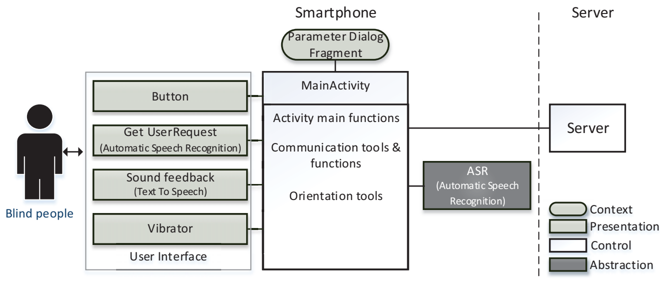


Fig. 7. Main interactive object on Smartphone part

B. Server Part

When receiving a navigation assistance request with destination as parameter, the Server will repeatedly gets the user location, finds path to destination, creates and sends feedbacks to the user through his smartphone. Localization service holds the Abstraction role which determines user location and path finding (see Fig. 8).

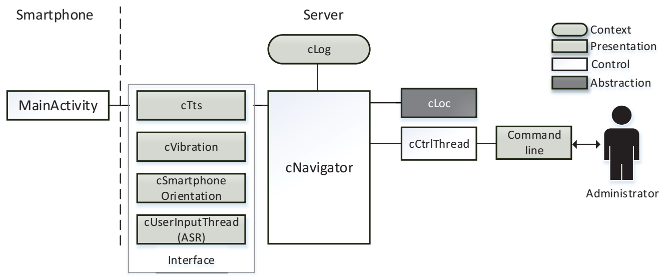


Fig. 8. Main interactive object on Server part

In Fig. 8, **cTts**, **cVibration**, **cSmartphoneOrientation** and **cUserInputThread** (ASR) packages interface between **MainActivity** (Control entity on Smartphone), with **cNavigator** (Control entity on Server). The **cLog** package records the information needed to debug, and evaluate the system, history part of interaction context. The **cCtrlThread** package manages the interaction so that the administrator can see the system activities and get the control if necessary. To ensure distributed attribute interface classes are inherited from **cZmq** (see Fig. 9), interactive packages are modeled and implemented as active concurrent entities and inherit from **cThread** class which encapsulates all the functions of an active concurrent entity.

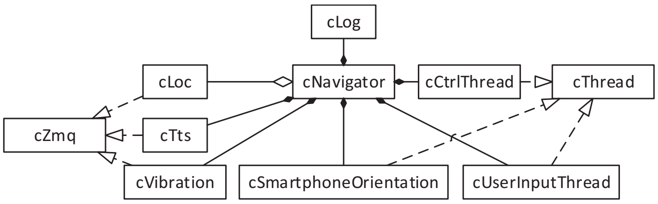
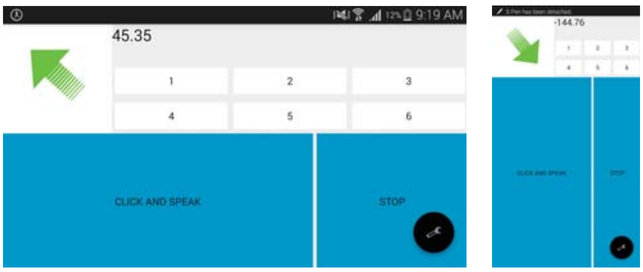


Fig. 9. Interaction class diagram of Server part

C. Interaction

On smartphone's main interface, the screen is divided into several zones, as shown in see Fig. 10. The lower zone is designed to receive the user navigation requests by using two large buttons: left button to start, right one to stop; while the top zone is reserved to testers by showing device orientation and configurable buttons corresponding to main tasks. The moving direction is shown in the top-left corner in both vertical and horizontal modes. To make a request for a destination, the user needs to press lower left button, say the destination, then release. Once the system detects a navigation assistance request, the smartphone vibrates to confirm and forwards this request to server. On receiving navigation assistance information, the smartphone will feedback to the user by Vietnamese speech synthesis such as *turn left*, *turn right*, *go straight n meters*, *stop*, etc., with the help of a Vietnamese TTS (text to speech) module [14]. When the user arrives to the destination, the smartphone will notify by voice and vibration.

On the server, the administrator can see the user information, identifier, location, orientation, as well as his desired destination. This is useful to evaluate the system or test the usability by the Wizard of Oz method [15].



(a) (b)

Fig. 10. Interface on smartphone: (a) Horizontal mode, and (b) Vertical mode

To reduce interaction by hand between blind people and smartphone, interaction by speech in noisy environment was designed in the application (Vietnamese speech is used in this project). According to the pre-defined scenario of the project, user (blind pupils) talks to smartphone to make a request to find the way from his/her place to another place in the school. The speech request is recorded by smartphone and be transmitted to the server. A speech recognition module is designed on the server to recognize the users request and send the command to the next module. Using client-server architecture, the speech recognition module can receive requests from many clients and process the requests correctly.

First, usually used requests for the defined scenarios were investigated through a survey with 50 blind pupils to know how they give a voice demand to the system. With different starting points and different destinations, 30 sentences were selected for recording to cover all possible sentence templates and keywords. Blind speakers cannot use smartphone directly; therefore, a technician helps them with a special recoding tool developed to speak out the sentences and record the utterances. To resolve the problem of robust automatic speech recognition in noisy environment, multi-style training method was used (depends on number of levels of noise, number of recorded sentences, number of speakers, etc.) For each speaker, a set of 30 sentences were recorded 16 times in different places: study room, corridor and club hall, in different moments: playing time (noise sources: people speaking, laughing, shout, sport activities) or learning time (noise sources: murmurs, fan). It was supposed to cover all possible noise levels of the real environment of the application. The completed corpus contains around 10 hours of audio signal for 20 voices.

VI. Experimental Results

For evaluation, the whole system is implemented in Nguyen Dinh Chieu School for Blind Pupils in Hanoi, Vietnam. Ten Wi-Fi access points (EnGenius ECB300) are installed in a single-floor area with nine rooms and a corridor, together with four cameras (Axis M1054) along the corridor, as depicted in Fig. 11. For the localization test, a person holding a smartphone (see Fig. 12) moves along a path with predefined turning points. Each time reaching one point, the user is asked to press a button to signal the system saving a timestamp which will later be used to interpolate the ground-truth for accuracy evaluation.

Table I summarizes the localization results in 3 different configurations: using only Wi-Fi, using only camera, and using both of them; while Fig. 13 exposes the details in the last case. Four error indices are used, including maximal value, average value, root mean squared value, and error value with reliability of 90%. It can be observed that the combination method has significantly improved the localization accuracy compared to the Wi-Fi-only system, with the average error is 42.5% lower. However, the error of 1.71m at reliability of 90% is obviously higher than acceptable safety requirements for the deployment in real applications, and further improvement in localization accuracy is essential.

TABLE I
LOCALIZATION RESULTS WITH DIFFERENT CONFIGURATIONS

Error (m)	**Wi-Fi only**	**Camera only**	**Multimodal**
Maximal	4.48	4.5	4.79
Average	1.55	0.88	0.89
Root mean squared	1.99	1.09	1.10
With reliability of 90%	3.17	1.69	1.71

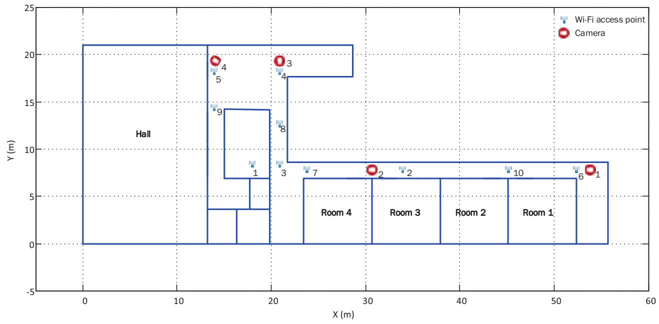


Fig. 11. Testing environment

Fig. 12. Experiment with blind people

For the interaction by voice, from the dataset, the speech recognition module is implemented using Sphinx toolkit of CMU. For dependent speaker evaluation, the 16th utterance of each speaker is extracted to test set, while other utterances (from 1st to 15th) are used to train. The sentence error rate is 13.4%, word error rate is 6.4% and word accuracy is 97%. While testing with the visually impaired pupils, after getting familiar with the system, each one starting at a certain location asks the system to give instruction to a desired destination. All the tests for this scenario is successful without additional help.

VII. Conclusion

The system introduced in this paper has been implemented and tested with visually impaired pupils. Experimental results show that the proposed technologies are suitable for the navigation assistance for the visually impaired people. However, so far, the accuracy of the localization solution (1.71m with reliability of 90%) is insufficient in real applications. Therefore, the system is currently only recommended for using in safe environments, or using together with some traditional tools, before further improvement is achieved.

Acknowledgment

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.04-2013.32.

References

[1] M. Hersh and M. A. Johnson, *Assistive technology for visually impaired and blind people*. Springer Science & Business Media, 2010.

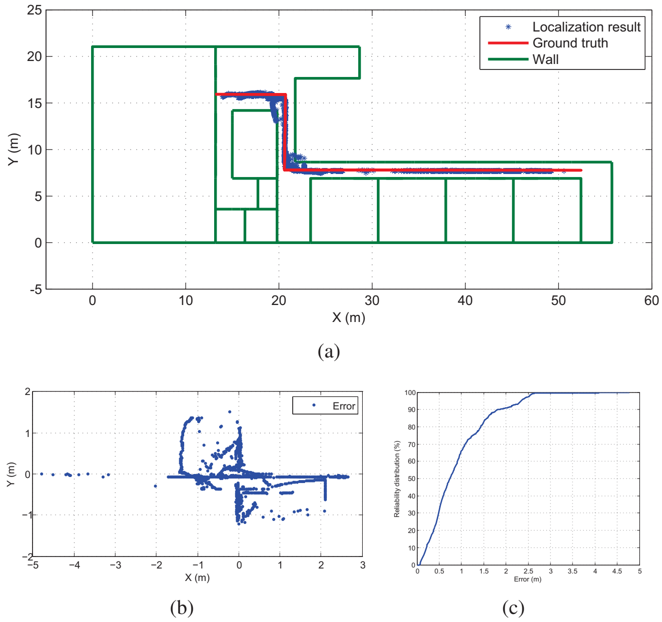


Fig. 13. Localization results: (a) Ground-truth and estimiated locations; (b) Error distribution; (c) Reliability distribution

[2] Y. Tian, X. Yang, C. Yi, and A. Arditi, "Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments," *Machine Vision and Applications*, vol. 24, no. 3, pp. 521–535, 2013.
[3] L. A. Guerrero, F. Vasquez, and S. F. Ochoa, "An indoor navigation system for the visually impaired," *Sensors*, vol. 12, no. 6, pp. 8236–8258, 2012.
[4] A. Almaaitah, K. Ali, H. S. Hassanein, and M. Ibnkahla, "3D passive tag localization schemes for indoor RFID applications," in *Communications (ICC), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1–5.
[5] R. Mautz, "Indoor positioning technologies," Ph.D. dissertation, Habilitationsschrift ETH Zürich, 2012, 2012.
[6] R. E. Korf, "Depth-first iterative-deepening: An optimal admissible tree search," *Artificial Intelligence*, vol. 27, no. 1, pp. 97–109, 1985.
[7] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *Systems Science and Cybernetics, IEEE Transactions on*, vol. 4, no. 2, pp. 100–107, 1968.
[8] H.-L. Nguyen, E. Castelli, T.-K. Dao, V.-T. Nguyen, and T.-T. Pham, "Multimodal combination of GPS, WiFi, RFID and step count for user localization," in *Ubiquitous Information Technologies and Applications*. Springer, 2014, pp. 675–681.
[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.
[10] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 2. IEEE, 2004, pp. 28–31.
[11] T. T. T. Pham, H. Vu, and A. T. Pham, "A robust shadow removal technique applying for person localization in a surveillance environment," in *Proceedings of the 6th International Symposium on Information and Communication Technology*. ACM, 2015, p. 41.
[12] J. Coutaz, "Pac: An object oriented model for implementing user interfaces," *ACM SIGCHI Bulletin*, vol. 19, no. 2, pp. 37–41, 1987.
[13] P. Hintjens, *ZeroMQ: Messaging for Many Applications*. " O'Reilly Media, Inc.", 2013.
[14] T. Van Do, D.-D. Tran, and T.-T. T. Nguyen, "Non-uniform unit selection in Vietnamese speech synthesis," in *Proceedings of the 2nd Symposium on Information and Communication Technology*. ACM, 2011, pp. 165–171.
[15] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of Oz studies: why and how," in *Proceedings of the 1st International Conference on Intelligent User Interfaces*. ACM, 1993, pp. 193–200.