IEEE ICCE 2014

2014 IEEE Fifth International Conference on Communications and Electronics (ICCE)

Pullman Beach Resort Da Nang
July 30 - August 1, 2014

IEEE

WELCOME

Welcome to the 2014 IEEE Fifth International Conference on Communications and Electronics (IEEE ICCE 2014) integrated in a USB version.

ICCE is becoming a reputable biennial international conference series in the scientific community on the areas of Electronics and Communications recently. Following the past successful conferences, the fifth IEEE ICCE 2014 looks for significant contributions to various topics in communication engineering, networking, microwave engineering, signal processing and electronics engineering. The conference will also include tutorials, workshops, and technology panels given by world-class speakers.

At the conference, two hundred and twenty eight (228) papers from more than 30 countries have been submitted. Among these submissions, ninety six (96) regular full papers, which will be submitted for inclusion into IEEE Xplore and twenty nine (29) poster papers have been accepted for presentation and will be organized in 18 regular and 2 special sessions. The Opening Session will host 4 keynotes. Four tutorials offered by the conference widely cover the most interesting topics on electronics and communications engineering, whose issues related to Quality of Mobile Multimedia Experience, Immersive Visual Communication with Depth, Crowdsourcing, and Free Space Materials Characterization.

The technical program focuses on hot topics on the fields of Communications Networks and Systems, Signal Processing and Applications, Microwave Engineering, and Electronic Systems. Besides, the special sessions on Crowdsourcing and Crowdsourcing Applications, and on Information Hiding and Security in Communications: Recent Developments have been added to the technical program of this event.

Beside the printed proceeding, this edition in USB is designed for readers to locate the papers by session or authors. Papers are originated as electronic files and were converted to Adobe Acrobat PDF file format for a crossplatform access. Even though the viewing quality on your monitor may vary, all papers have been printed clearly.

Enjoy your reading through!

IEEE Catalog Number: CFP1416B-USB
ISBN: 978-1-4799-5050-8

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For reprint or republication permission, email to IEEE Copyrights Manager at pubs-permissions@ieee.org. All rights reserved. Copyright ©2014 by IEEE.

Series Editors

Prof. Nguyen Xuan Quynh	Member of the State Council for Professor Title, Vietnam; Chair of the Interdisciplinary Council on Electricity - Electronics - Automation, Vietnam; Senior expert of the National Council for Science and Technology Policy, Vietnam
Prof. Phuoc Tran-Gia	University of Würzburg, Germany
Prof. Dong Ho Cho	Korea Advanced Institute of Science and Technology, Korea
Prof. Peter A. Freeman	Georgia Tech. and formally NSF, USA

Mapping Services in Indoor Environments based on Image Sequences

Quoc-Hung Nguyen[1], HaiVu[1],Thi-Thanh-Hai Tran[1]
International Research Institute MICA, HUST
CNRS/UMI 2954 - Grenoble INP Hanoi University of Science and Technology, Viet Nam
[1]{quoc-hung.nguyen, hai.vu, thanh-hai.tran }@mica.edu.vn

Quang-Hoan Nguyen[2]
Faculty of Information Technology
Hung Yen University of Technology and Education
Khoai Chau, Hung Yen, Viet Nam
quanghoanptit@yahoo.com.vn

***Abstract*— This paper describes a visual-based system that autonomously operators for both map building and localization tasks. The proposed system is to assist mapping services in small or mid-scale environments such as inside a building or campus of school where conventional positioning data such as GPS, WIFI signals are often not available. Toward this end, the proposed approaches utilize only visual data. We design an image acquisition system for data collections. On one hand, a robust visual odometry method is utilized to create routes in the environment. On the other hand, the proposed approaches utilize FAB-MAP (Fast Appearance Based Mapping) algorithm that is maybe the most successful for recognizing places in large scenarios. The building route and learning visited places are off-line process in order to represent a map of environment. Through a matching image-to-map procedure, the captured images at current view are continuously positioned on the constructed map. This is an online process. The proposed system is evaluated in a corridor environment of a large building. The experimental results show that the constructed route coincides with ground truth, and matching image-to-map is high confidence. The proposed approaches are feasible to support visually impaired people navigating in the indoor environments.**

Keywords*— *Visual Odometry, Place Recognition, FAB-MAP algorithms.

I. INTRODUCTION

Understanding and representing environments have been research topics over long period of time in field of the autonomous mobile robots. These works aim to answer two questions. Given a representation of the environment, the first question is that "What does the world look like?" (or building a map of environment). In contrast to this, the localization service is to estimate the pose of object of interest relative to a position on the created map. It is to answer the second question "Where am I?" (Positioning the object of interest on the created map). To solve these questions, the positioning data come from various types of sensors such as GPS, WIFI, and LIDAR. However, these conventional source data are not always available or convenient for acquisitions, particularly, in small or moderately environments. For example, GPS systems provide the mapping services in strict conditions such as good weather, outdoor environments, no presence of buildings. It is highly cost to setup LIDAR systems in the environments, where are mid-scale areas like campus of school, hospital. WIFI systems are also not easily installing to cover such environments. To overcome these issues, this paper presents a vision-based system that utilizes only the visual data. Advantages of using such data are that it is safe, flexible, provides a very rich and valuable perception information of the environments. The proposed system aims to automatize the map building and localization services, particularly, to serve the mapping services in indoor environments.

In this work, the map building consists of creating trajectories and learning scene elements from the evaluated environment. We simultaneously collect visual data of the routes and scenes using an own-designed imaging acquisition system. A robust visual odometry technique that tends to use only one consumer-grade camera is adapted in order to build the trajectories. In order to learn places in the environment, we utilize so-called loop closure detections method [2, 3]. For the localizing task, an agent (such as vehicle, robot, and human) needs only consumer-grade camera (e.g., mobile camera of a tablet) for capturing images. A current view is matched to the place database through a probabilistic model of the FAB-MAP (Fast Appearance Based Mapping) algorithms. It is notice that our proposed system is not able to update new positions against the created map. We simply past new places using a simple motion that is based on positions of the closest neighbor places. The proposed system is evaluated in a corridor environment of a large building. Experimental results show that we successfully create a map of the evaluated environment. The results of matching image-to-map are 20% precision and 100% recall. It is feasible results for developing navigation services in the evaluated environment.

The next sections of paper are organized as follows: In Section II, we briefly survey related works. In Section III, we present our vision-based system for autonomous map building and localization tasks. We report the experimental results in Section IV. Finally, we conclude and give some directions for future works.

II. RELATED WORKS

Vision-based mapping and localizing services are fundamental topics in field of mobile robotics and computer visions. There are uncountable publications of these topics. Readers can refer a good survey in [2]. In this section, we focus on recent advanced techniques in the field of computer vision that offer substantial solutions with respect to localization and navigation services in known or unknown environments. Alcantarilla [4] utilizes well-known techniques such as Simultaneous Localization and Mapping (SLAM) and Structure from Motion (SfM) to create 3-D Map of an indoor environment. He then utilizes means of visual descriptors (such as Gauge

978-1-4799-5051-5/14/$31.00 ©2014 IEEE

Speeded Up Robust Features, G-SURF) to mark local coordinate on the constructed 3-D map. Instead of creating a prior map, Lui et al. [5] utilize a pre-captured reference sequence of the environment. Given a new query, their system desires to find the corresponding set of indices in the reference video. By means of visual SLAM techniques, some wearable applications are proposed. Pradeep et al. [8] present a head-mounted, stereo-vision for detecting obstacles in the path and warn subjects about their presence. They incorporate visual odometry techniques and metric-topological SLAM. Murali et al. in [9] estimate relative location relative of users to the crosswalks in the current traffic intersection. They develop a vision-based smart-phone system for providing guidance to blind and visually impaired. The system of Murali et al. in [9] requires supplemental images from Google Map services, therefore it is suitable with travels at traffic intersections only.

Complexity of the map building varies in function of the environment size. For example, indoor environments are more complex than outdoor environments because of the office supplies as chairs, tables, etc. Furthermore, matching a current view to a position on the created map seems to be the hardest problem in many works [2],[10]. In this work, we solve these issues through an incremental map which is able to increase accuracy of the matching procedures. Therefore, different from these systems, our approaches learn places in an environment through many trials. When new observations appear, they must be locally and globally consistent with the previous construction. These problems are able to solve through the loop closure algorithms [3],[11]. However, major differences from those work are that our proposed system utilize only visual data for both map and localization services, whereas the works in [3],[11] use GPS data for localizing visited places on the map. Furthermore, localization services suffer from limitations of consume-grade camera so that matching image-to-map's performance is acceptable.

III. PROPOSED APPROACHS

A. *Imaging acquisitions system*

We design a compact imaging acquisition system to capture simultaneously scenes and routes in the environments.

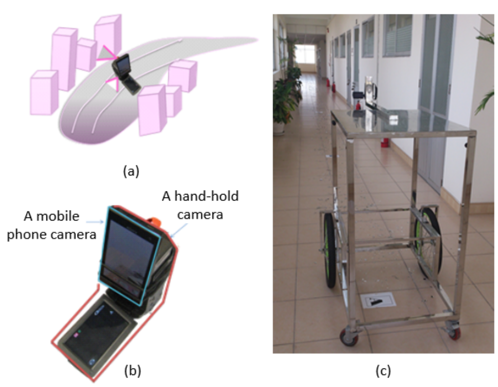

Fig. 1. (a) A schematic view of the visual data collection scheme. (b) The proposed imaging system. Mobile phone camera is attached on rear of a hand-hold camera. (c) The cameras are mounted on wheel-vehicle.

A schematic view the system is shown in Fig. 1(a). It has two cameras. One captures scenes around the environment. The second one captures road on the travel. Setting of two cameras is shown in Fig. 1(b). These cameras are mounted on a wheel-vehicle, as shown in Fig. 1(c). The detail of the collected data in an evaluated environment is described in Section IV.

B. *The proposed framework*

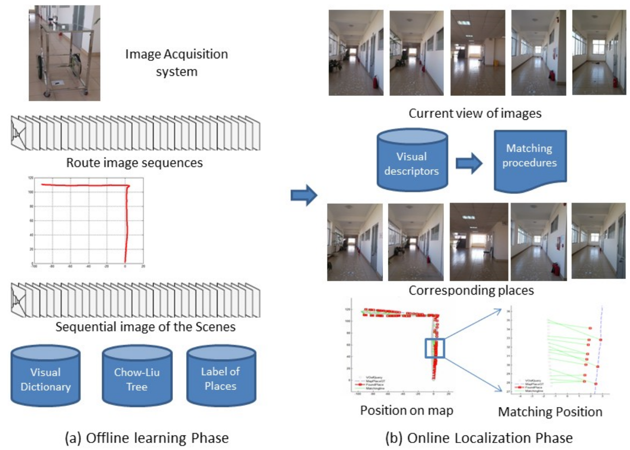

Fig. 2. The framework of the proposed system

- **Offline learning**: Using the collected visual data, this phase creates the trajectories and learnt the places along the travels. The techniques to construct the map and learning the places are described in Sec.III.C, Sec.III.D, respectively. Because scenes and route images are captured concurrently, the constructed map contains visited places and their corresponding positions.
- **Online localization**: A current observation is described using visual words. These data are associated matching procedure to the places which are indexed in a database. The current pose thus is localized on the constructed map.

C. *Route building based on visual odometry techniques*

To build route of the travel , we utilize a visual odometry method proposed by Van Hamme et al [1]. The method is based on the tracking of ground plane features. Please refer [1] for details of the algorithms. In this section, we describe our adaptations to the algorithms in [1].

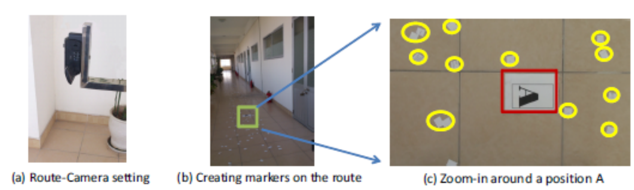

Fig. 3. The scheme for collecting road images. (a). Setting of the camera. (b) Scattering makers on the evaluated route. (c). A zoom-in version around a position in (b).

The algorithms in [1] are designed to take into account the uncertainty on the vehicle motion as well as uncertainty on the extracted features. However, well-known issues for visual odometry techniques are that they need to estimate precisely correspondences between the features of consecutive frames. Once the feature correspondences have been established, we can reconstruct the trajectory of the vehicle. Our image acquisition system setups a camera so that it is perpendicular to the ground plane, as shown in Fig. 3(a). We also scatter markers in the whole journey as shown in Fig. 3(b-c). Because the detected features are projected on a ground plane, results of the detecting and matching features are more precisely. We compare our results with/without using man-makers in Fig. 7 The learning places aims to visually present scenes along the travel. These visual presentations need to be easy implementation and efficient distinguishing scenes. To adapt with these issues, we involve the FAB-MAP algorithms [3] which are recently successful for matching places in routes over long period time. The FAB-MAP is a probabilistic appearance-based approach to place recognition.

Each time the image taken, its visual descriptors are detected and extracted. In our system, we utilize SURF extractors and descriptors for creating on a visual vocabulary dictionary. A Chow Liu tree is used to approximate the probability distribution over these visual words and the correlations between them. Fig. 4(a)-(b) shows the extracted features and visual words to build visual dictionary.

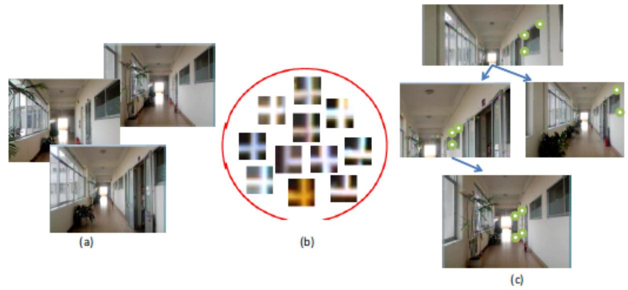

Fig. 4. FAB-MAP algorithm to learn places. (a) SURF features are extracted from image sequences. (b) Visual words defined from SURF extractors. (c). Co-occur of visual words by same object

Beyond the conventional place recognition approaches that simply compares image similarity between two visual descriptors. The FAB-MAP involves co-occur visual word of same subject in the worlds. For example, Fig. 4(c) shows window subject in various contexts, but several visual words are co-appearances.

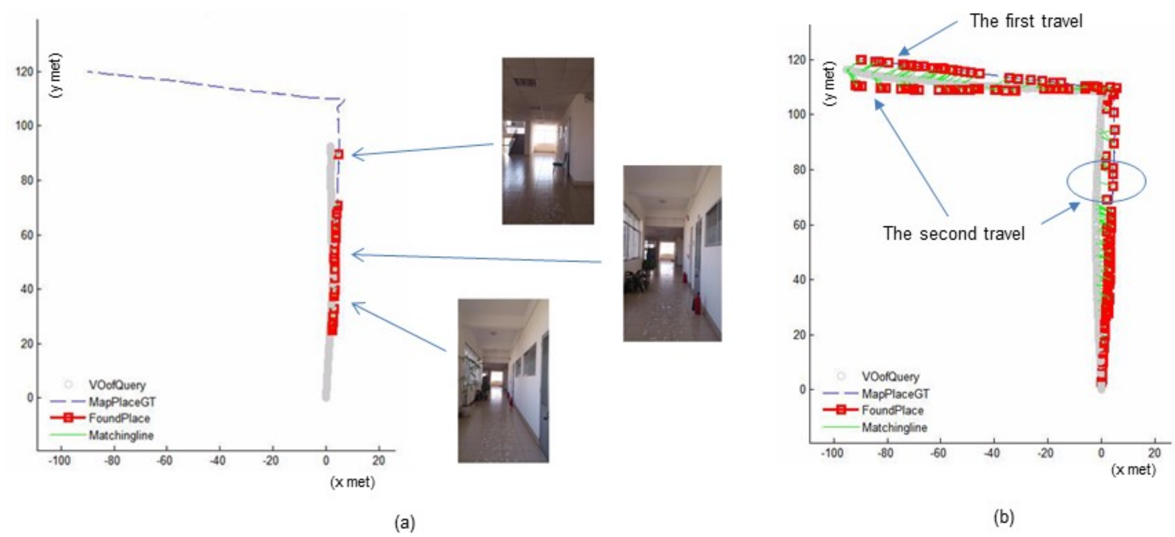

Fig. 5. (a) The places are learnt and their corresponding positions are shown in the constructed map data. (b) Many new places are updated after second trial

Consequently, the distinct scenes are learnt from visual training data. For updating new places, we implement captured images through several trials.

For each new trial, we compare the images with the previous visited places which are already indexed in a place database. This procedure calls a loop closure detection. These detections are essential for building an incremental map. Fig. 5 (a) shows only few places are marked by the first travel, whereas various places that are updated after the second travel as shown in Fig. 5(b).

D. Matching image-to-map procedure

Given a current view, its position on the map is identified through a place recognition procedure. We evaluate the current observation at location L_i on the map by its probability when given all observations up to a location k:

$$\rho(L_i|Z^k) = \frac{\rho(Z_k|L_i)\rho(L_i|Z^{k-1})}{\rho(Z_k|Z^{k-1})} \quad (1)$$

Where Z_k contains visual words appearing in all observations up to *k-1*; and Z^k presents visual words at current location *k*. These visual words are defined in the learning places phase. A probability $p(Z_k|L_i)$ infers observation likelihood that learnt in the training data. In our system, a L_i is matched at a place k∗ when *argmax(p(Z*$_k$*|L*$_i$*))* is large enough (through a pre-determined threshold T = 0.9). Fig. 6 shows an example of the matching procedure.

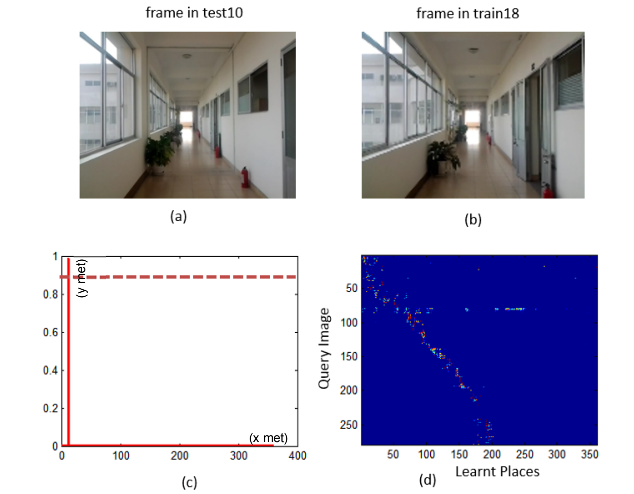

Fig. 6. (a) Given a current observation, (b) the most matching place. (c) The probability $p(L_i|Z^k)$ calculated with each location k among K = 350 learnt places. (d) Confusion matrix of the matching places with a sequential collected images (290 frames).

Given an observation as shown in Fig. 6(a), the most matching place is found at *placeID* = 12. The probability $p(L_i|Z^k)$ is shown in Fig. 6(c) with a threshold *value = 0.9* whose the maximal probability is *placeID = 12*. A confusion matrix of the matching places for an image sequence is shown in Fig. 6(d). This example shows that we can resolve almost places in a testing phase.

IV. Experimental Results

A. Evaluation Environments

- **Setting up environments:** We examine the proposed method in a corridor environment of a building, where is 10th floor of International Research Institute MICA-Hanoi University of Science and Technology (HUST). A 3-D model of the evaluation environment is shown in Fig. 7(c).
- **Database collection:** Two camera devices are mount into a vehicle as shown in Fig. 1(c). A person moves at a speed of 1.25 foot/second along the corridor. The total length of the corridor is about 60 m. We collect data in four times (trials), as described in TABLE I.

TABLE I. Three rounds data results

Trials	Total Scene images	Total road images	Duration
L1	8930	2978	5:14
L2	10376	2978	5:30
L3	6349	2176	3:25
L4	10734	2430	4:29

B. Experimental results

Results of the constructed map using original work of Van Hamme et al [1] is shown in Fig. 7(a), whereas the reconstructed travels using proposed method are shown in Fig. 7(b). As shown, the results of route building from three travels are quite stable. All of them are matched to ground truth that are plotted in green dash-line in a model 3-D of the evaluation environments, as shown in Fig. 7(c). The proposed method gives substantial results comparing with the original one [1]. We believe that creating highly textures on the ground plane is efficient for detecting and matching the features. Even the original algorithms [1] are designed to be robust with uncertainty of the detected features; but more precisely the features matching more higher quality creating the map. We continue evaluating the proposed system with aspects of the place recognition rate on the created map. To define visual word dictionary as described in Sec.III.D, we use collected images from L1 trial. About *W=1300* words are defined in our evaluation environments. We then use dataset from L4 travel to learn visited places along the travel. Totally, *K = 140* places are learnt. The visual dictionary and descriptors of these places are stored in XML files.

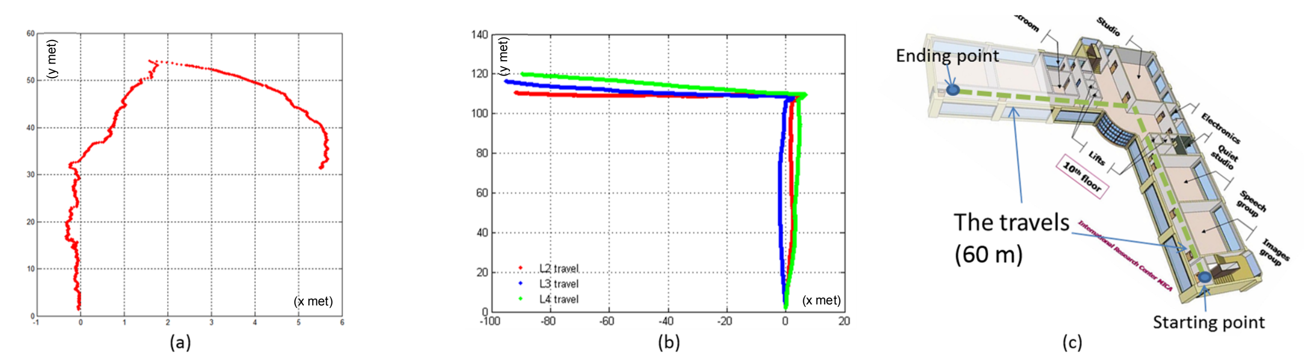

Fig. 7. (a) The travel reconstructed using original works [1]. (b) Results of three time travels (L2, L3, and L4) using proposed method. (c) A 3-D map of the evaluation environment. The actual travels also plotted in green dashed line for comparing results between (a) and (b).

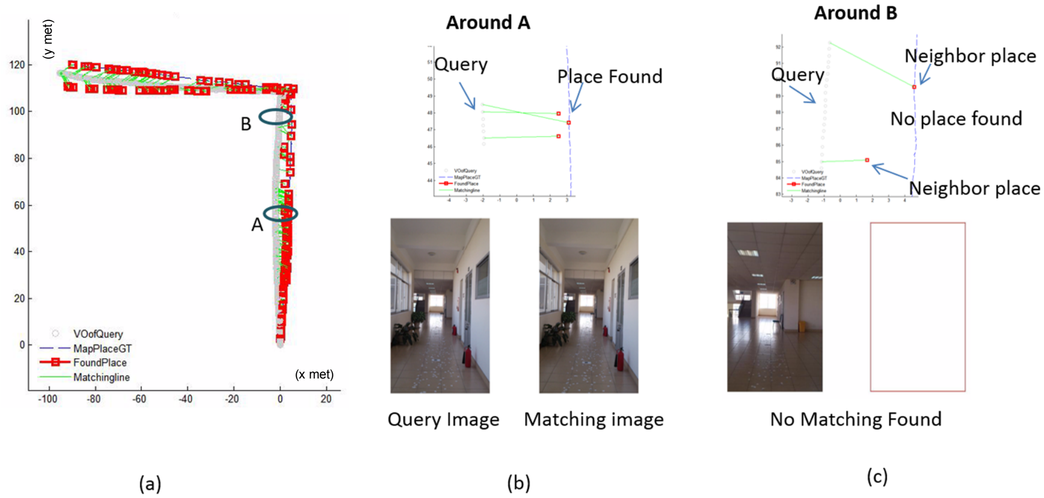

Fig. 8. (a) Results of the matching image-to-map with L3 trial. Two positions around A and B are given. (b)-(c): current view is on the left panel (query image); matching is on the right panel. Upper panel is a zoom-in around corresponding positions.

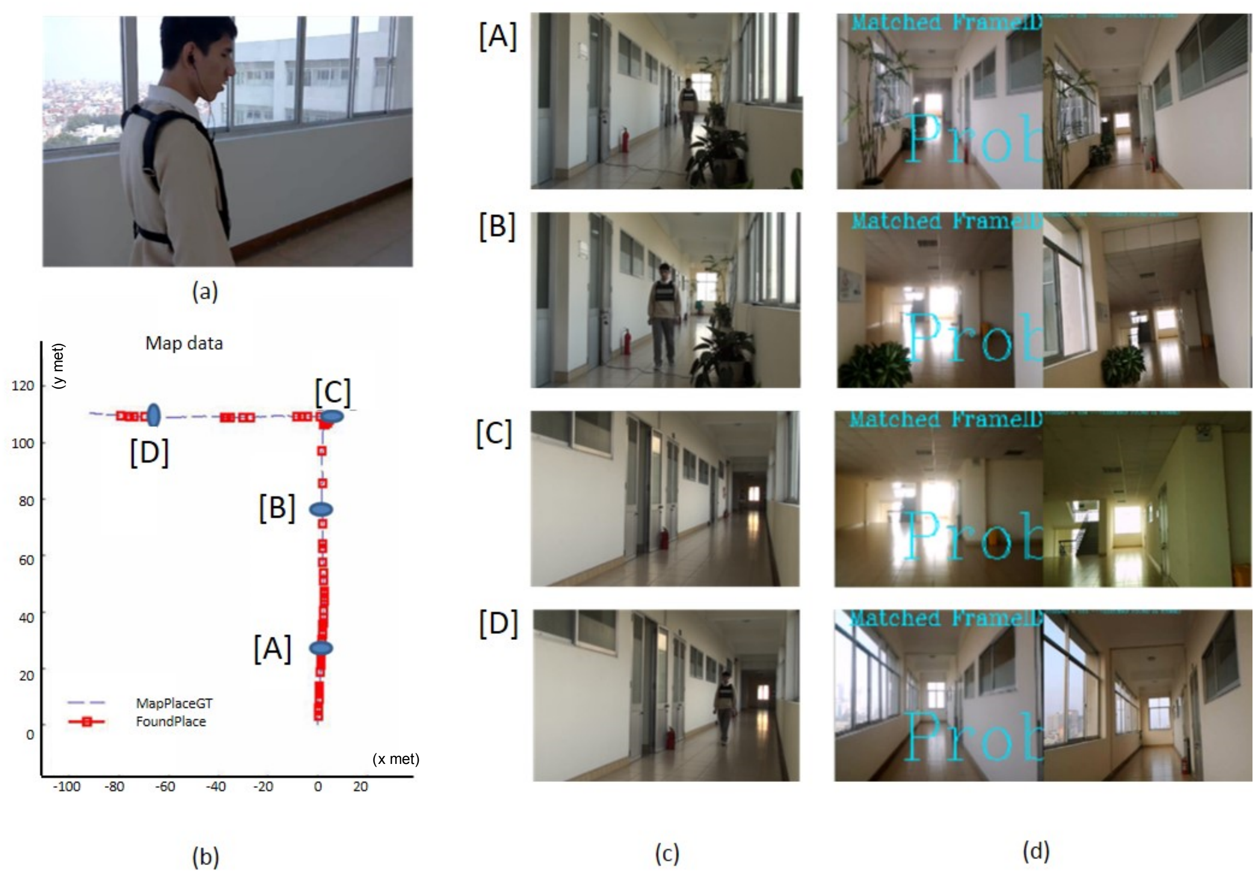

Fig. 9. (a) A blind pupil wearing Samsung tablet and ear-phone for voice control. (b) The results of his travel in the evaluated environment. The places on his travel are marked on the constructed map in red-rectangles. Four examples at [A], [B], [C], and [D] are marked on the map. (c) Corresponding images at [A], [B], [C], and [D] captured by a surveillance camera, for monitoring his travel. (d) - right panels are current views at those places captured by the camera on tablet. Left panels are corresponding matched places taken out from the place database.

The collected images in L2 and L3 travels are utilized for the evaluations. Visually, some matching places results from L3 travel are shown in Fig. 8. Two demonstrations are shown in Fig. 8 (around position A and position B). Case A shows a query image (from L3 trial) is matched to a visited place. Therefore, its corresponding position on the map is able to localize. A zoom-in version around position A is shown in the top panel. "No place found" in Case B means that the query image was not found from the place database. For the quantitative measurement, we then evaluate the proposed system using two criteria: ***Precision*** is to measures total place detected from total query images, whereas ***Recall*** is to measure correct matching places from detected places.

A threshold for matching place $T = 0.9$ is predetermined. TABLE II. shows results of the precision and recall with L2 and L3 trials. The recall of L3 travel is clearly higher than L2. The main reason is that some "new" places, where were not learnt from L4 trial, are updated after running L2 trial. Therefore, more "found" places are updated through L3 travel. Recall obtains 100% inferring that it is high confidence at the detected places despite of changing view-point of scenes (because of different travels). It is notice that although experimental results report low precisions (due to case of "no place found" of Case B in Fig. 8), it is acceptable for the localization service. Some neighbor places (as shown in the zoom-in version of case B in Fig. 8) suggest us interpolating positions of the query image. Furthermore, an agent (a robot or wheel-vehicle) moves approximately at speed of 10 cm/sec and computational time of the matching image-to-map procedure at 1 frame/sec. Therefore, the precision at 20% (averagely) means that by moving a distance of 100 cm (10 frames captured), 2 positions (within a distance of 1 m) are confidence located (100% recall).

TABLE II. RESULT OF THE MATCHING PLACE

Travels	Precision	Recall
L2	12%	100%
L3	36%	100%

We then preliminary deploy the proposed approaches to assist navigating services to a blind pupil. He is asked to travel in the evaluated environment (as shown in Fig. 7(c)), where is new environment from his knowledge. Currently, we setup a voice control system to guide the blind pupil. Furthermore, for creating ground truth of the localization services in this evaluation, we setup a surveillance camera in the environment. This system also supports us monitoring his travels.

The blind pupil wears a tablet with frontal camera and an ear-phone as shown in Fig. 9(a). According to guidance by voice control, he could go whole travel in the corridor environment. The image data collected from the tablet is send through a WIFI network. Results of the matching image-to-map is shown in Fig. 9(b) in red spots. The query and matching results of some places (e.g., [A], [B], [C], [D]) are shown in Fig. 9(c) and Fig. 9(d), respectively. Distances between these places are from minimal values of 0.3 m (around [A]) to maximal values of 5m (around [C], [D]). The blind pupil is asked to travels three times. We obtain average performances of the matching procedure that are ***18% precision*** at ***85% recall***. These results are not so far from TABLE II. In other words, we could match precisely positions the blind people using the constructed map. These results are feasible to deploy automatic navigating system.

V. CONCLUSIONS

In this paper, we presented a vision-based system for both autonomously map building and localizing services. We successfully created the map of the indoor environment using the visual odometry and learning places. The results of matching image-to-map are high confidence. Therefore, the proposed system is able to provide us deploying navigating services in the indoor environments. The proposed system directs to support visually impaired peoples in such environments. A current issue of the proposed approaches is that the number of "*place found*" is quite limited in the experiments. The motion models and particle filter algorithms suggest us directions to further research. Further evaluations with complex environments are also needed.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] Hamme, D.V., P. Veelaert, and W. Philips, Robust visual odometry using uncertainty models, in Proceedings of the 13th international conference on Advanced concepts for intelligent vision systems 2011, Springer-Verlag: Ghent, Belgium. p. 1-12.

[2] Bailey, T. and H. Durrant-Whyte, *Simultaneous localization and mapping (SLAM): part II state of the art.* Robotics & Automation Magazine, IEEE, 2006. **13**(3): p. 108-117.

[3] Cummins, M. and P. Newman, FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. Int. J. Rob. Res., 2008. **27**(6): p. 647-665.

[4] Fernández Alcantarilla, P., Vision based localization: from humanoid robots to visually impaired people, in Electronics2011, University of Alcala: Ph.D. Thesis.

[5] Liu, J.J., C. Phillips, and K. Daniilidis. Video-based localization without 3D mapping for the visually impaired. in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on.

[6] Bigham, J.P., et al. VizWiz:LocateIt - enabling blind people to locate objects in their environment. in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on.

[7] Winlock, T., E. Christiansen, and S. Belongie. Toward real-time grocery detection for the visually impaired. in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on.

[8] Pradeep, V., G. Medioni, and J. Weiland. Robot vision for the visually impaired. in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on.

[9] Murali, V.N. and J.M. Coughlan. Smartphone-based crosswalk detection and localization for visually impaired pedestrians. in Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on.

[10] Fraundorfer, F. and D. Scaramuzza, *Visual Odometry : Part II: Matching, Robustness, Optimization, and Applications.* Robotics & Automation Magazine, IEEE, 2012. **19**(2): p. 78-90.

[11] Newman, P. and H. Kin. SLAM-Loop Closing with Visually Salient Features. in Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on.