HỌC VIỆN KỸ THUẬT QUÂN SỰ

ISSN - 1859 - 0209

Tạp chí KHOA HỌC & KỸ THUẬT

Journal of Science and Technology

CHUYÊN SAN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG - SỐ 5

HÀ NỘI

Số 164

10-2014

TẠP CHÍ
KHOA HỌC VÀ KỸ THUẬT

CHUYÊN SAN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Le Quy Don Technical University

Journal of
Science and Technology

No. 5 (10-2014)

SECTION ON INFORMATION AND COMMUNICATION TECHNOLOGY

Office:	236 Hoang Quoc Viet Str., Nghia Tan, Cau Giay Dist., Ha Noi
P.O. Box:	7EN-280, Nghia Tan, Cau Giay, Ha Noi
Tel:	069.515329
E-mail:	chuyensan.cntt@mta.edu.vn

MỤC LỤC

NGHIÊN CỨU KHOA HỌC **Trang**

TABLE OF CONTENTS

PHÒNG THÔNG TIN KHQS/HỌC VIỆN KTQS

CÁC ẤN PHẨM:

- ❖ Tạp chí Khoa học và Kỹ thuật xuất bản 6 số/năm
- ❖ Bản tin điện tử
- ❖ Các chuyên đề và thông tin phục vụ lãnh đạo không định kỳ

CÁC DỊCH VỤ:

- ❖ Cung cấp thông tin theo địa chỉ và hỏi đáp
- ❖ Thông tin thư mục chuyên đề
- ❖ Cung cấp bản sao tài liệu gốc, bản dịch, băng ghi âm, ghi hình về CT-XH và KHCN
- ❖ Viết các chuyên đề, tổng quan, phân tích theo yêu cầu
- ❖ Quảng cáo cho các đơn vị nghiên cứu và kinh doanh

ĐỊA CHỈ LIÊN HỆ

Ban Tạp chí KH và KT, Phòng Thông tin KHQS, Học viện KTQS.

Hòm thư: 7EN-280, 236 Hoàng Quốc Việt, Nghĩa Tân, Cầu Giấy, Hà Nội.

Tel.: 069.515329

E-mail: chuyensan.cntt@mta.edu.vn

HOAN NGHÊNH MỌI Ý KIẾN ĐÓNG GÓP CỦA ĐỘC GIẢ

TẠP CHÍ KHOA HỌC VÀ KỸ THUẬT

Giấy phép xuất bản số 231/GP-BTTTT ngày 25/02/2011

In tại xưởng in Học viện Kỹ thuật Quân sự.

THỂ LỆ VIẾT VÀ GỬI BÀI
CHO CHUYÊN SAN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

1. Bài gửi đăng phải là công trình nghiên cứu của tác giả, chưa đăng và chưa gửi đăng ở bất kì tạp chí nào khác.

2. Bài gửi đăng bằng tiếng việt hoặc tiếng anh theo khuôn dạng kích thước A4 (một cột), font chữ Times New Roman, cỡ chữ 12, lề trên và lề dưới 4cm, lề phải 3cm và lề trái 4cm.

3. Các hình vẽ phải rõ ràng, chuẩn xác. Nếu bài có ảnh thì phải rõ nét. Hình vẽ và ảnh phải được chú thích đầy đủ.

4. Bài gửi phải được soạn thảo sử dụng LaTex (tải mẫu tại http://fit.mta.edu.vn/files/chuyensan/template.rar).

5. Tài liệu tham khảo, trích dẫn phải có đủ các thông tin theo trình tự sau: Họ và tên tác giả (hoặc chủ biên), tên sách (tên bài báo/tạp chí, tên báo cáo khoa học), nơi xuất bản, nhà xuất bản, năm xuất bản, trang trích dẫn.

6. Ghi rõ họ, tên, học hàm, học vị, nơi làm việc, số điện thoại của tác giả kèm theo các tệp chứa nội dung bài báo (cả tệp pdf và nguồn) và gửi qua email: chuyensan.cntt@mta.edu.vn.

7. Bài viết phải có tên bằng tiếng Việt và tiếng Anh. Mỗi bài cần có kèm theo phần tóm tắt bằng tiếng Việt (cỡ chữ 10, tối đa 150 từ) cung cấp những thông tin cần thiết về kết quả chính của bài viết.

8. Cấu trúc bài báo bao gồm các phần: dẫn nhập, nội dung khoa học của bài và kết luận (viết thành mục riêng). Bài báo phải đưa ra được các kết quả mới, các ứng dụng mới hoặc phải nêu được hiện trạng, những hướng phát triển cơ bản của vấn đề được đề cập, khả năng nghiên cứu, phát triển và ứng dụng tại Việt Nam. Bài viết không quá 15 trang.

9. Với bài thông tin khoa học, tin ngắn khoa học kỹ thuật: là các bài dịch, tổng thuật, tổng quan về các vấn đề khoa học kỹ thuật có tính thời sự.

10. Trước khi in chính thức HĐBT sẽ gửi bản bông cho tác giả để soát lỗi lần cuối trong thời hạn từ 1-2 ngày. Do vậy, tác giả là người chịu trách nhiệm cuối cùng không chỉ về nội dung khoa học mà cả về các lỗi in ấn, lỗi chính tả (nếu có).

DANH SÁCH PHẢN BIỆN KHOA HỌC

1.	TS Nguyễn Lê **Anh**	*Đại học Xây dựng Hà Nội*
2.	TS Võ Nguyễn Quốc **Bảo**	*Học viện Công nghệ Bưu chính Viễn thông TPHCM*
3.	PGS TS Lê Anh **Cường**	*Đại học Công nghệ, ĐHQGHN*
4.	TS Phạm Quang **Dũng**	*Đại học Bách khoa Hà Nội*
5.	PGS TS Đinh **Điền**	*Đại học Khoa học tự nhiên, ĐHQGTPHCM*
6.	TS Lê Vũ **Hà**	*Đại học Công nghệ, ĐHQGHN*
7.	TS Trần Thị Thanh **Hải**	*Viện Nghiên cứu MICA, Đại học Bách khoa Hà Nội*
8.	TS Vũ **Hải**	*Viện Nghiên cứu MICA, Đại học Bách khoa Hà Nội*
9.	TS Tạ Chí **Hiếu**	*Học viện KTQS*
10.	TS Huỳnh Trung **Hiếu**	*Đại học Công nghiệp TPHCM*
11.	TS Trần Văn **Hoài**	*Đại học Bách khoa TP Hồ Chí Minh*
12.	PGS TS Nguyễn Xuân **Hoài**	*Đại học Hà Nội*
13.	TS Phạm Ngọc **Hùng**	*Đại học Công nghệ, ĐHQGHN*
14.	PGS TS Lê Thanh **Hương**	*Đại học Bách khoa Hà Nội*
15.	TS Hà Thu **Lan**	*Học viện Công nghệ Bưu chính Viễn thông Hà Nội*
16.	PGS TS Lương Chi **Mai**	*Viện CNTT, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*
17.	TS Lý Quốc **Ngọc**	*Đại học Khoa học tự nhiên, ĐHQGTPHCM*
18.	TS Nguyễn Hồng **Quang**	*Đại học Bách khoa Hà Nội*
19.	TS Trần Thái **Sơn**	*Viện CNTT, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*

20.	TS Nguyễn Việt **Sơn**	*Viện Nghiên cứu MICA, Đại học Bách khoa Hà Nội*
21.	PGS TS Trần Đức **Tân**	*Đại học Công nghệ, ĐHQGHN*
22.	PGS TS Ngô Quốc **Tạo**	*Viện CNTT, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*
23.	TS Nguyễn Phương **Thái**	*Đại học Công nghệ, ĐHQGHN*
24.	PGS TS Huỳnh Quyết **Thắng**	*Đại học Bách khoa Hà Nội*
25.	TS Nguyễn Thị **Thủy**	*Đại học Nông nghiệp Hà Nội*
26.	PGS TS Trịnh Nhật **Tiến**	*Đại học Công nghệ, ĐHQGHN*
27.	PGS TS Đặng Thành **Tín**	*Đại học Bách khoa TP Hồ Chí Minh*
28.	TS Trần Minh **Triết**	*Đại học Khoa học tự nhiên, ĐHQGTPHCM*
29.	PGS TS Nguyễn Linh **Trung**	*Đại học Công nghệ, ĐHQGHN*
30.	TS Lê Minh **Tuấn**	*Sở Khoa học Công nghệ Hà Nội*
31.	TS Trần Xuân **Tú**	*Đại học Công nghệ, ĐHQGHN*
32.	TS Phạm Trần **Vũ**	*Đại học Bách khoa TP Hồ Chí Minh*

AUTOMATIC LOCALIZATION AND NAVIGATION ASSISTANCE FOR VISUALLY IMPAIRED PEOPLE USING ROBOT

Nguyen Quoc Hung[1], *Tran Thanh Hai*[1], *Vu Hai*[1], *Nguyen Quang Hoan*[2]

Abstract

This paper presents a new support system for the visually impaired navigation in small pervasive environments using mobile robots. The system uses computer vision techniques to build environment map, locate and detect obstacle. The proposed method enables the robot to quickly know its location on the map. The object detection and recognition algorithms are performed in real time in order to avoid static and dynamic obstacles in the environment. To help the visually impaired person navigating in the environment, the proposed method aims to find the optimal path that is the shortest and the most secure path for the moving of the human. In order to interact between human and robot, we design and develop an interface using touch screen and vibration patterns of a smart-phone device. The system is evaluated on a number of blind pupils. The experimental results confirmed that the proposed system is feasible to deploy in practical application.

Bài báo này trình bày một hệ thống trợ giúp định vị và dẫn hướng tự động cho khiếm thị hoàn toàn mới trong môi trường cảm thụ diện hẹp sử dụng robot. Hệ thống sử dụng kỹ thuật thị giác máy tính để xây dựng bản đồ môi trường, xác định vị trí và phát hiện các vật cản có trong môi trường. Các phương pháp đề xuất cho phép robot nhanh chóng biết được vị trí trên bản đồ. Một giải thuật mạnh phát hiện và nhận dạng được áp dụng nhằm giúp cho Robot tránh những vật cảntĩnh và động trong môi trường thực hiện trong thời gian thực. Để trợ giúp người khiếm thị định hướng trong môi trường, một phương pháp tìm đường tối ưu được đề xuất nhằm tìm ra con đường ngắn nhất và an toàn cho việc di chuyển, dẫn hướng cho Robot. Để tương tác giữa người và robot, chúng tôi thiết kế và phát triển một giao diện trên màn hình cảm ứng, sử dụng các độ rung mẫu của thiết bị điện thoại thông minh. Hệ thống này được đánh giá trên một số học sinh khiếm thị tại các môi trường khác nhau. Các kết quả thử nghiệm xác nhận rằng hệ thống đề xuất là khả thi để triển khai trong các ứng dụng thực tế.

Index terms

Visually impaired Robot Small Pervasive Environments.

1. Introduction

In today's society, daily activities of visually impaired or blind people, hereinafter visually impaired people, have to face to many difficulties because of their inability to perceive the world around them. Their moving in a small environment often takes very longtime training. To be able to navigate in a new and a larger environment such as a building, the visually

(1)International Research Institute MICA - Hanoi University of Science and Technology.
(2)Hung Yen University of Technology and Education.

Fig. 1. The functional components of the system navigational aids visually impaired people using mobile robot

impaired people require assistances of a sighted person. Therefore, one of the most important issues in aid for visually impaired people aims to develop a navigational aids system.

The support systems generally focus to resolve two main issues: *1) Assisting visually impaired people to understand surrounding environment including objects; 2) Helping visually impaired people to locate the position of themselves*. Several research or commercialized products utilize ultrasonic sensors such as iNavBelt [26], GuideCane [3], [4]. Other systems such as EyeRing [19] developed by MIT's Media Lab uses finger-won device to translate images into aural signals. Despite of many efforts to improve communication methods, these products still have inherent limitations, such as limited resources or poor interaction with the user. Recently, research in computer vision, machine learning, mobile devices have opened up tremendous opportunities for the development of systems to support the visually impaired people.

In this paper we present a completely new system supporting navigation for the visually impaired people in small pervasive environments using robot. The advantages of the proposed system are that it is fully automatic, easy to communicate as well as feedback navigation information in real time. The proposed system is presented in Fig. *1*.

In order to achieve the above functions, the system uses image sensor mounted on a mobile robot. Vision data support to locate robot's position and to find the shortest path from a current point to destination. The robot interacts with visually impaired people using a smart-phone device. We evaluate a system with several visually impaired pupils. The experimental results show that the proposed system can lead to move to any pre-defined destination on a plan floor of a hallway.

The remaining of this paper is organized as follows: Section 2 presents related works on navigation assistance for visually impaired people. Section 3 describes in detail the components of the proposed system. Section 4 presents the evaluation methodology and experimental results. Finally, we conclude the work and give some direction to develop the system in the future.

2. Related Works

Developing localization and navigation assistance tools for visually impaired people have been received many intention in the autonomous robotics community [6]. Most of the works focus on finding efficient localization solutions based on positioning data from different sensors such as GPS, laser, Radio Frequency Identification (RFID), vision or the fusion of several of them. Loomis et al. [17] surveyed efficiency of GPS-based navigation systems supporting visually impaired people. The GPS-based systems share similar problems: low accuracy in urban-environments (localization accuracy is limited to approximately 20 m), signal loss due to multi-path effect or line-of-sight restrictions due to the presence of buildings or even foliage. Kulyukin et al. [15] proposed a system based on Radio Frequency Identification (RFID) for aiding the navigation of visually impaired people in indoor environments. The system requires the design of a dense network of location identifiers. Helal et al. [13] proposed a wireless pedestrian navigation system. They integrated several signals such as voiced, wireless networks, Geographic Information System (GIS) and GPS to provide the visually impaired people with an optimized route. Recent advanced techniques in computer vision offer substantial improvements with respect to localization and navigation services in known or unknown environments. The vision-based approaches offer not only safe navigation, but also provide a very rich and valuable description of the environment. For example, [2] develops an application named LocateIt, which helps blind people locate objects in indoor environments.

In [30], ShelfScanner is a real-time grocery detector that allows online detection of items on a shopping list. With respect to visual mapping and localization, Alcantarilla [9] utilizes well-known techniques such as Simultaneous Localization and Mapping (SLAM) and Structure from Motion (SfM) to create a 3-D Map of an indoor environment. He then utilizes visual descriptors (such as Gauge- Speeded Up Robust Features, G-SURF) to mark local coordinates on the constructed 3-D map. Instead of building a prior 3-D map, Lui et al. [16] utilize a pre-captured reference sequence of the environment. Given a new query sequence, their system attempts to find the corresponding set of indices in the reference video. Some wearable applications based on visual SLAM have also been proposed. Pradeep et al. [23] present a head-mounted stereo-vision platform for detecting obstacles in the path and warn subjects about their presence. They incorporate visual odometry and feature based metric-topological SLAM. Murali et al. [18] estimate the user's location relative to the crosswalks in the current traffic intersection. They develop a vision-based smartphone system for providing guidance to blind and visually impaired travelers at traffic intersections. The system of Murali et al. in [18] requires supplemental images from Google Map services, therefore its applicability is limited to outdoor travel.

It is clear from these works that a SLAM-based approach is ideally suited to the task of guiding the visually impaired, because SLAM combines the two key elements required for a user friendly and widely applicable system: map building and self-location. However, the complexity of the map building task varies in function of environment size. In some case, a map can be acquired from the visual sensor, but in other cases, the map is such that it must be constructed from other sensor modalities such as GPS, WIFI [5]. Furthermore, matching a current view to a position on the created map seems to be the hardest problem in many works [1], [10]. Important work towards appearance-based place recognition has been conducted in [27] which borrowed ideas from text retrieval systems and introduced the concept of the so called visual vocabulary. The idea was later extended to vocabulary trees by [21], allowing to efficiently use large vocabularies [25] demonstrated city-scale place recognition using these tree structures.

Recently, Maddern et al. report an improvement to the robustness of FAB-Map by incorporating odometric information into the place recognition process. [28] Propose BRIEF-GIST, a very simplistic appearance-based place recognition system based on the BRIEF descriptor. BRIEF-GIST is much easier to implement and its performance is comparable to FAB-MAP. In our point of view, an incremental map is able to support us in improving matching results. Therefore, different from the systems mentioned above, we attempt to create a rich map as good as possible through many trials. When new observations arrive, these new observations must be locally and globally consistent with the previously constructed map. To this end we employ the the loop closure algorithms from [5], [20]. Furthermore, we pay significant attention to the creation of the visual dictionary. We deploy the GIST features [22], a holistic representation of the natural scenes. Selection of the most representative frames helps to construct a robust visual dictionary of the environment.

Different from above works, in this paper, we focus on resolving the issues of the navigation such as positioning the robot in the environment, detecting and recognizing obstacles that warning for visually impaired people along the way. In term of the end-user, the proposed system needs to be easy and friendly for using.

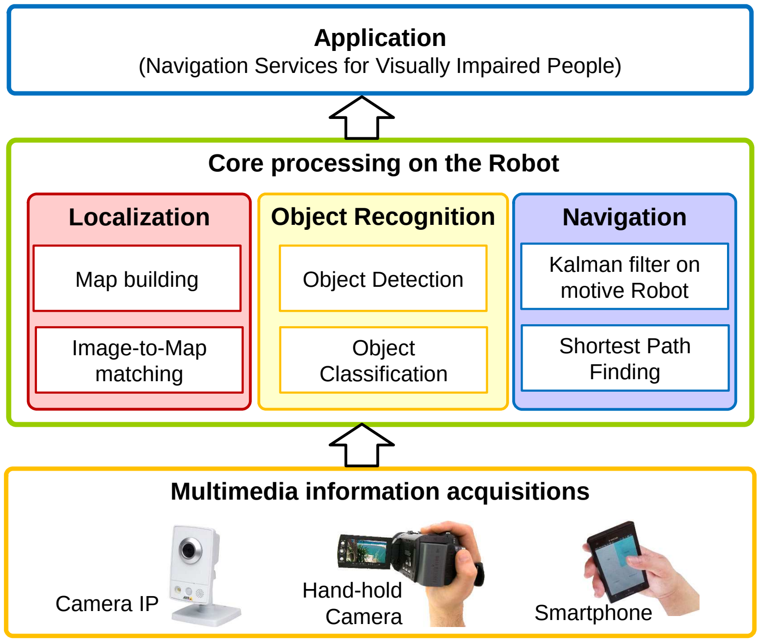


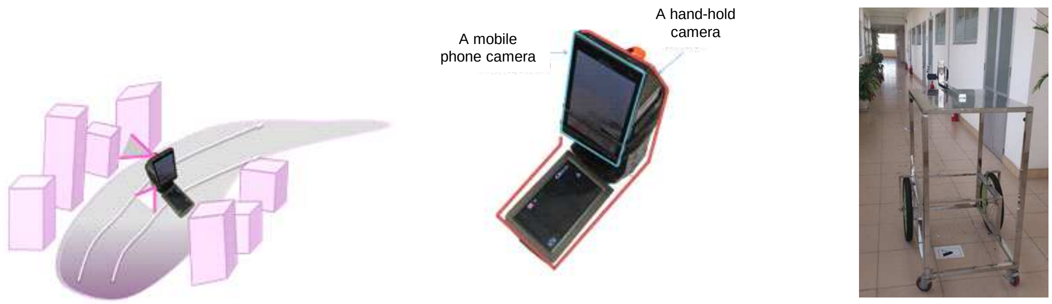
Fig. 2. The framework of the proposed system

3. The Proposed Approaches

3.1. The overall architecture

We propose a general architecture as shown in Fig. 2 including three main modules:

- **Localization**: Two tasks will be carried out: building environment map and constructing environmental dictionary. Here, we propose using FAB-MAP algorithm [5] to serve for matching the current view with the ones in the database. This module is presented in detail in Section 3.2.
- **Obstacle detection and recognition**: We aim to detect and recognize static obstacles *{Potted plants, Trash, and Extinguisher}* and dynamic one *{Human}* that the visually impaired person could meet on the path. Thanks to this module, the visually impaired people can avoid obstacles or be informed about it. Details are described in Section 3.3.
- **Navigation and Human-Machine Interaction**: We solve two problems: Path finding in Environment and human-machine interaction. A powerful algorithm *"Shortest Path"* to path finding in environment and the direction of travel of the robot is developed. The exact driver of robot movement is improved using Kalman filter. Finally, an application is built on the Android OS platform allowing the visually impaired person to interact with the system. The interaction is very easy and commode using vibration or touch screen. This module is described in more detail in Section 3.4.



(a) (b) (c)

Fig. 3. (a) A schematic view of the visual data collection scheme; (b) The proposed imaging acquisition system in which a mobile phone camera is attached on rear of a hand-hold camera; (c) The image acquisition system attached on a wheel vehicle.

(a) Route-Camera setting (b) Creating markers on the route (c) Zoom-in around a position A

Fig. 4. The collection databases on Road

3.2. Localization

We design a compact imaging acquisition system to capture simultaneously scenes and routes in the indoor environments. A schematic view of the data collection platform is shown in Fig. *3*(a). The proposed acquisition system has two cameras. One camera captures scenes around the environment. The second one aims at capturing road on the travels. The camera setting is shown in Fig. *3*(b). These cameras are mounted on a vehicle, as shown in Fig. *3*(c). The details of the collected images are described in the experiments. The vehicle will be only used during the off-line phase to build the map of the environment and capture scene images. Using a vehicle in the offline phase has the advantage that it avoids the vibration of the camera system. As a consequence, it allows a more accurate reconstruction of the route. To build route of the travel, we utilize a visual odometry method proposed by Van Hamme et al. [11]. The method is based on the tracking of ground plane features. Particularly, it is designed to take into account the uncertainty on the vehicle motion as well as uncertainty on the extracted features.

Our system configures the acquisition camera so that it is perpendicular to the ground plane, as shown in Fig. *4*(a). 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 between the two frames. Due to the floor characteristic of the corridor environment, the number of feature points detected by the original work [11] is quite limited and leads to a very poor reconstruction of the travel. To solve this issue, we manually placed additional markers over the whole journey as shown in Fig. *4*(b-c). In future work, the odometry method should be adapted to better work in case of sparse feature distribution.

According to [11], the authors use undistorted pinhole camera model to project 3D world coordinates to 2D image coordinates. A 2-D image point $x = [x\ y\ \omega]^T$ is given from a 3-D point $X = [X\ Y\ Z\ 1]^T$ by:

$$x = C[R|t]X \tag{1}$$

where C is triangular intrinsic camera matrix; $[R|t]$ is the rotation matrix R that aligns the world axes with the camera axes, augmented by the 3D translation vector t between their origins. Eq. (1) suggests an inverse transformation that projects image points onto the world ground plane. While measure translation vector t is part of the extrinsic calibration process, estimating rotation matrix R with heading, pitch and roll are key challenge to overcome. They are well-known issues for visual odometry and image stitching techniques. These issues usually are solved by estimating correspondences between the features of consecutive frames. Once the feature correspondences have been established, we can reconstruct the trajectory of the vehicle between the two frames. In the Fig. *5* shown as the central idea of Van Hamme et al. method [11] is the back projection of image features to the world ground plane, and the uncertainties associated with this back projection. The parameters in the rotation matrix are estimated in view of uncertainty models: Perspective Uncertainty Tetragons (PUT) and Motion Uncertainty Tetragon (MUT). For each detected feature (such as Harris corners), PUT is estimated by each of the four combinations of extremely pitch and roll values. The MUT of a feature position is calculated by displacing the feature along four circle segments, representing the four extremely combinations of possible. A few number of PUT and MUT detected on a homogenous ground plane, where-as Fig. *6*(b) shows an increasing number of the detected PUT and MUT on the ground plane with markers. Once a lot of inliers are established, it is more precisely constructing the route in the Fig. *6*

The places visited along the trajectory of interest will be stored in a condensed visual representation. This visual representation preferably needs to be easy to adapt to our specific indoor context and efficient at distinguishing scenes. To meet these goals, we involve the FAB-MAP technique [11] which was recently demonstrated to be successful at matching places in routes over a long period time. It is a probabilistic appearance-based approach to place recognition. Each time an image is taken, its visual descriptors are detected and extracted. In our system, we utilize SURF extractors and descriptors for creating a visual vocabulary dictionary.

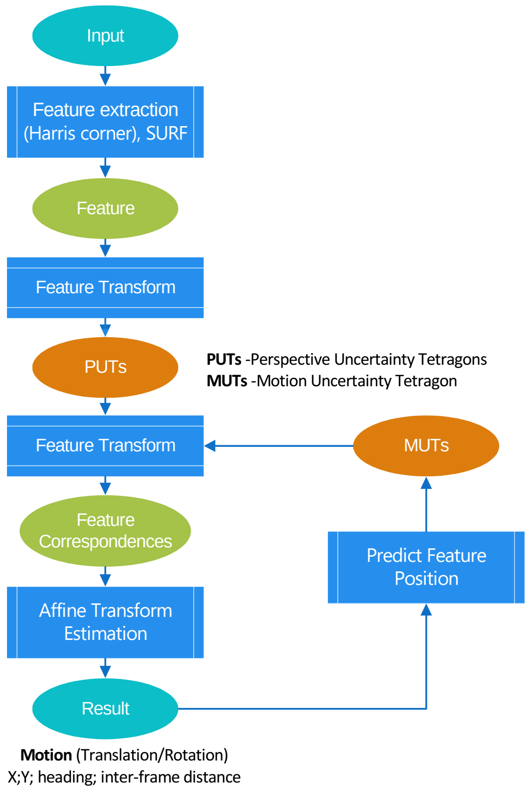
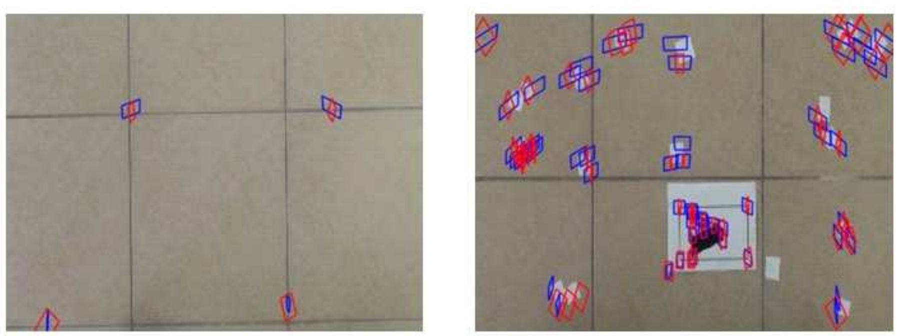


Fig. 5. The algorithm is based on the tracking of ground plane feature



(a) (b)

Fig. 6. The algorithm is based on the tracking of ground plane feature

A Chow Liu tree is used to approximate the probability distribution over these visual words and the correlations between them. Fig. 7(a)-(b) shows the extracted features and visual words to build the visual dictionary. Beyond the conventional place recognition approaches that simply compare image similarity between two visual descriptors. FAB-MAP examines co-occurrence of visual words for the same subject in the world. For example, Fig. 7(c) shows that for several windows, some visual words co-appearances are present. Consequently, the distinct scenes are learnt from visual training data. For updating new places, we incorporate

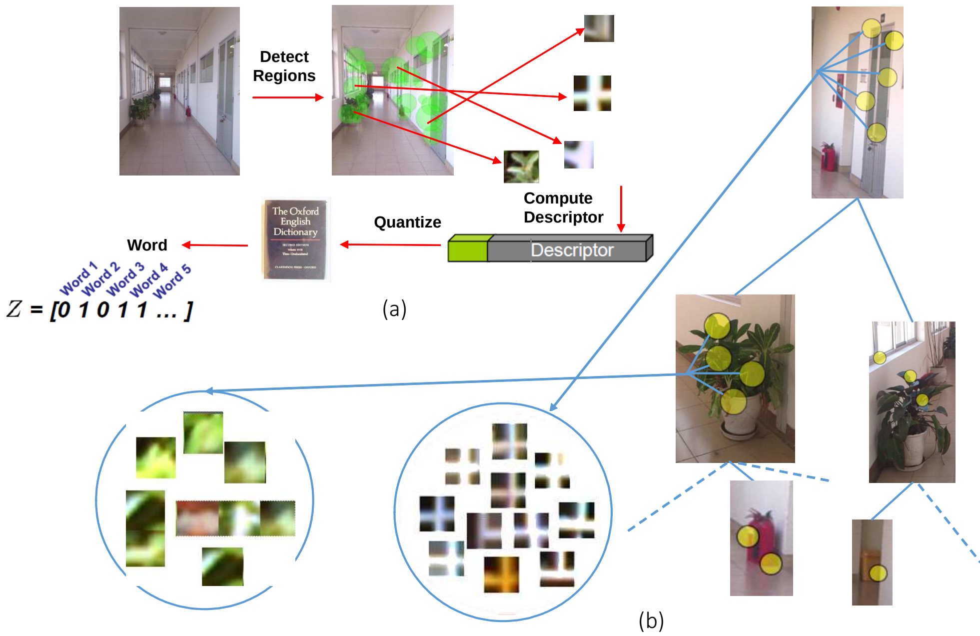


Fig. 7. FAB-MAP algorithm to learn places: (a) SURF features are extracted from image sequences and visual words defined from SURF extractors; (b) Co-occurrence of visual words corresponding to same object

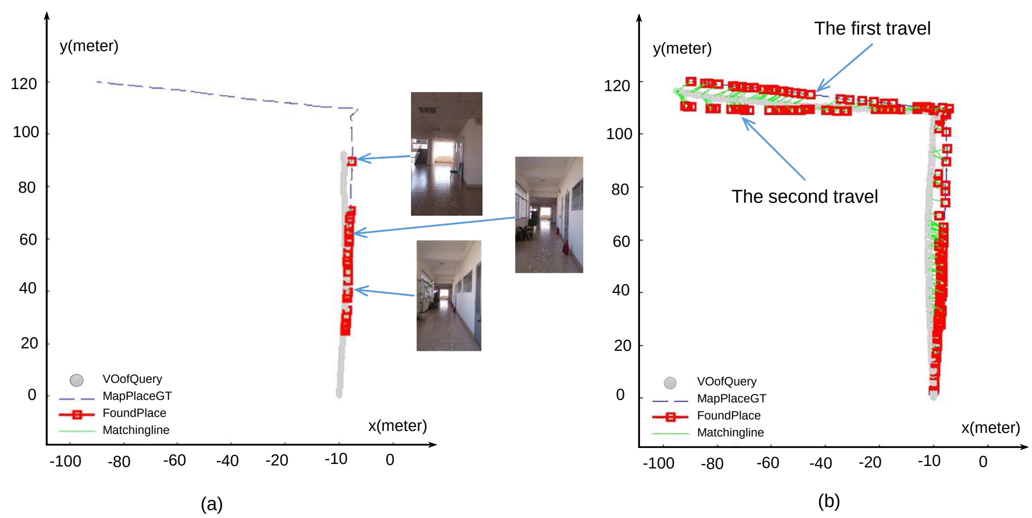


Fig. 8. (a) The places are learnt and their corresponding positions are shown in the constructed map data; (b) Many new places are updated after second trial

captured images through several trials. For each new trial, we compare the images with the previously 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. *8*(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. *8*(b).

In related works [11], [9] report that FAB-MAP obtains reasonable results for place recog-

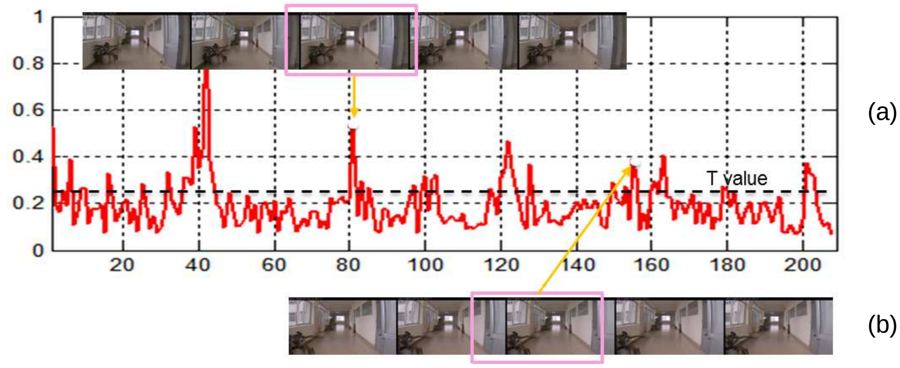


Fig. 9. (a) Dissimilarity between two consecutive frames. A threshold value T = 0.25 is preselected; (b) Two examples shows the selected key frames and their neighbor frames.

nition over long travels in term of both precisions and recall measurements. However, those experiments were implemented in outdoor environments which usually contain discriminate scenes. The original FAB-MAP [9] still has unresolved problems in discriminating scenes to define visual dictionary. This issue affects the results of FAB-MAP when we deploy it in indoor environments, where scenes are continuous and not clearly distinct.

Therefore, a pre-processing step is proposed to handle these issues. Given a set of scene images $S = \{I_1, I_2..I_n\}$ we learn key frames from S by evaluating inter-frame similarity. A feature vector F_i is extracted for each image I_i. In this work, the GIST feature [2] is utilized to build F_i. GIST presents a brief observation or a report at the first glance of a scene that summarizes the quintessential characteristics of an image. Feature vector F_i contains 512 responses which are extracted from an equivalent of model of GIST proposed in [16]. A Euclidean distance D_i between two consecutive frames is calculated to measure dissimilarity. Fig. *9*(a) shows distance D_i of a sequence including 200 frames. The key-frame then is selected by comparing D_i with a pre-determined threshold value T. Examples of selecting two key-frames are shown in Fig. *9*(b).

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:

$$p\left(L_i|Z^k\right) = \frac{p\left(Z_k|L_i\right)p\left(L_i|Z^{k-1}\right)}{p\left(Z_k|Z^{k-1}\right)} \qquad (2)$$

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 as learnt in the training data. In our system, a Li is matched at a place $k*$ when $argmax(p(Z_k|L_i))$ is large enough (through a pre-determined threshold $T = 0.9$). The Fig. *10* shows an example of the matching procedure.

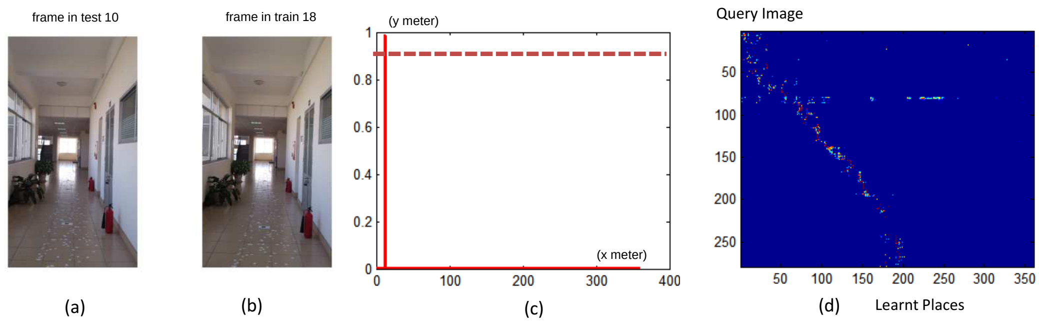


Fig. 10. (a) Given a current observation, (b) the best matching place. (c) The probability $p(L_i|Z_k)$ calculated with each location k among $K = 350$ learnt places. (d) Confusion matrix of the matching places with a sequence of collected images $(290 frames)$

Table 1. Result of the matching places (FAB-MAP algorithms)without and with Scene discriminations

Travels	**Without scene discrimination**		**With scene discrimination**	
	Precision	*Recall*	*Precision*	*Recall*
L2	12%	90%	67%	82%
L3	36%	85%	**74%**	**88 %**

Given an observation as shown in Fig. *10*(a), the best matching place is found at $placeID = 12$. The probability $p(L_i|Z_k)$ is shown in Fig. *10*(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. *10*(d). This example shows that we can resolve most places in a testing phase.

The collected images in $L\#2$ and $L\#3$ travels are utilized for the evaluations. Visually, some matching places results from $L\#3$ travel are shown in Fig. *11*. Two demonstrations are shown in Fig. *11 (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.

We collect data in 4 times (each time equals to one trial). To build the visual dictionary in off-line phase, we have used images collected from $L\#1$ trial. By experience, we set the size of the dictionary to 1300. We then use the images collected from $L\#4$ trial to learn places along the travel. In total, $K = 140$ places have been learnt. The visual dictionary and descriptors of these places are stored in XML files. The collected images in $L\#2$ and $L\#3$ travels are utilized for the evaluations.

The Table *1* shows the precision and recall with $L\#2$ and $L\#3$ travels with/without scene

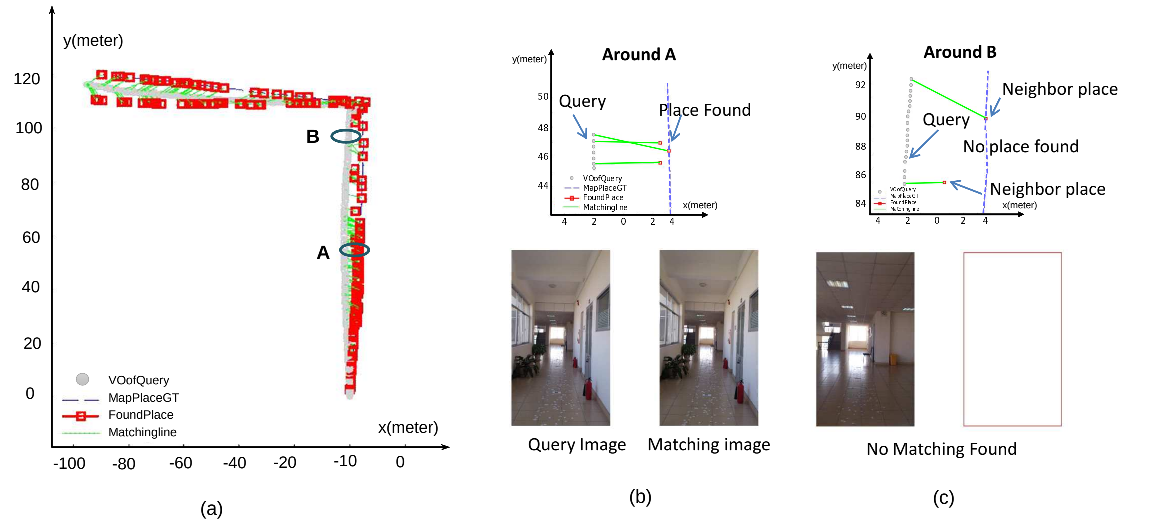


Fig. 11. (a) Results of the matching image-to-map with $L\#3$ trial. Two positions around A and B are given; (b)-(c): Current view is on the left panel (query image); matching is on the right panel. Upper panel is a zoom-in around corresponding positions.

discrimination step. For learning places (using original FAB-MAP, without scene discrimination), the recall of $L\#3$ travel is clearly higher than $L\#2$. The main reason is that some "new" places which were not learnt from $L\#4$ are able to update after $L\#2$ running. Therefore, more *"found"* places are ensured with $L\#3$ travel. Table *1* also shows the efficiency of the scene discriminations step, the performances of image-to-map matching obviously increasing and stable for precisions measurement with scene discrimination step, whereas high confidence of the recalls is still consistent.

3.3. Obstacle detection and recognition

Various vision-based approaches, which utilize different types of low level features and classifiers have been presented in the literature. In this section, we evaluate three different approaches for object classifications. This study is to select the best one which could lead to a reliable solution for object classifications. The framework for object detection and recognition is shown Fig. *12*.

Three methods for object detection are studied:

- **Haar and Adaboost** [29]
- **HoG and SVM** [7]
- **GIST and k-NN** [24], [22]

As aforementioned, these methods are studied due to their effectiveness in classification problem. For each method, we proposed to recognize each object by a learning a binary classifier. At classification phase, sliding window technique is used to scan the whole image;

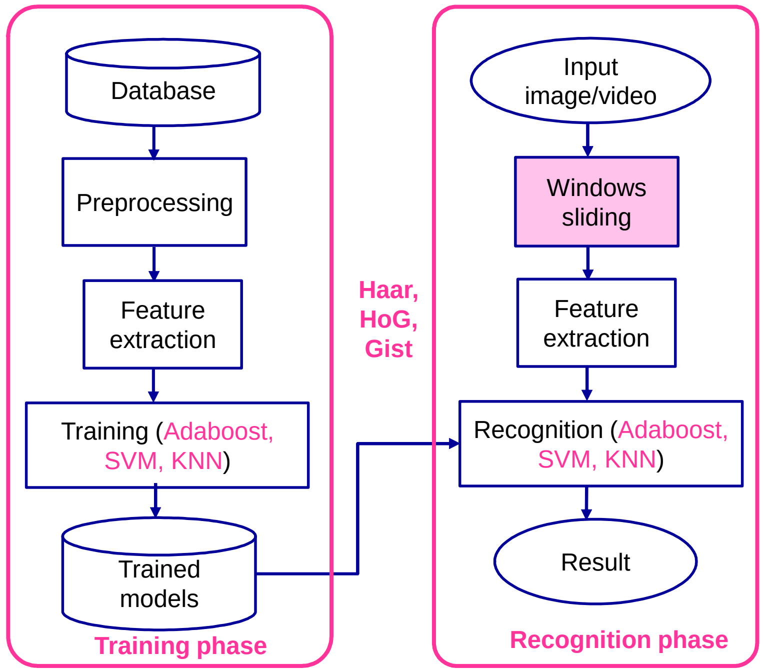


Fig. 12. Generic framework of object recognition for comparative study

each window candidate will be passed through feature extraction module then the computed descriptor will be passed into the corresponding binary classifier as shown Fig. *12*. For details of each method, the readers are invited to read the original papers.

In this work, we are interested to detect and recognize four classes of obstacles: {*Potted plant, Trash, Extinguisher, and Human*}. For training and testing detection and recognition methods, we have built a dataset containing 2104 images. The resolution of images is 600x400 pixels. Each object class has 526 images under daylight condition in a corridor of a build. This dataset is very challenge because objects are taken under different views point and distances. Some examples are presented in the Fig. *13*. All images in the database are annotated manually and organized in the directory. We divide the database into 2 parts: 504 images for training and 1600 images for testing.

To evaluate detection and recognition algorithms, there are many measures such as Recall, Precision, and Accuracy [8]. In our context, as we know the distribution of positive and negative examples (the ratio between positive and negative is 1/4) so we propose to evaluate our system by Precision criterion, which is defined as follows:

$$Precision = \frac{tp}{tp + fp} \tag{3}$$

The following table gives the precision for each object class and average precision. There methods run on a computer with the following configurations *(CHIP Intel(R) Core(TM) i5-2520M CPU @ 3.2 GHz x 2, RAM 8GB)*. We can observe that the human detection rate

Fig. 13. Some images in our database

[!t]

Table 2. Three rounds data results

No	Object Classes	Precision
1	Potted plants	78.16%
2	Trash	58.22%
3	Extinguisher	69.00%
4	Human	90.00%
	Average	**73.85%**

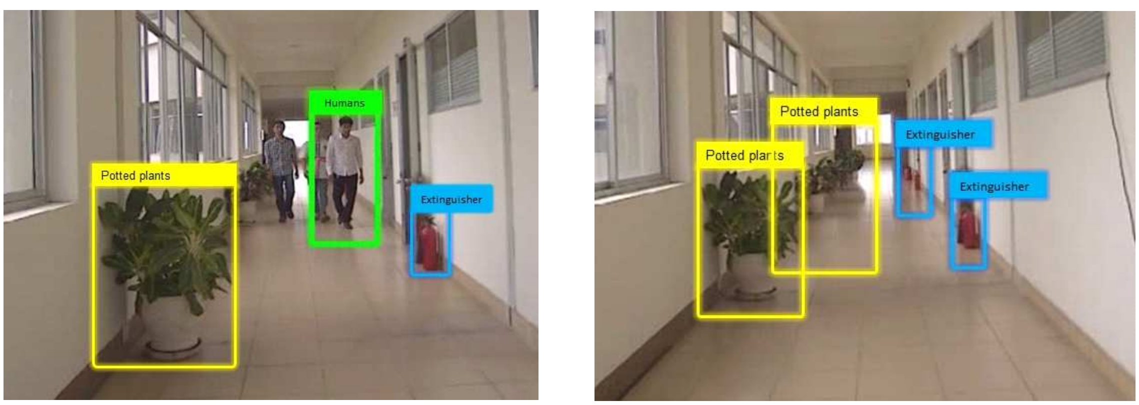


Fig. 14. Results of obstacle detection

is very high (90%) while other classes detection rate is smaller, mostly with $Trash$ class because many object with rectangular appearance could be classified as $Trash$. Fig. *14* show some examples of test results. For moving obstacles (human class), the system could be able to detect human at minimum distance 2m.

3.4. Navigation and Human-Machine Interaction

3.4.1. Improvement of path finding with path pruning and Kalman filter: Path-planning is a popular research topic which could be solved using some variant of IDA [14], based on the well-known A* [12]. Most of the solutions approach the environment in uniformly shaped grids [12] and making use of Voronoid subtraction/addition as well as geometric computation

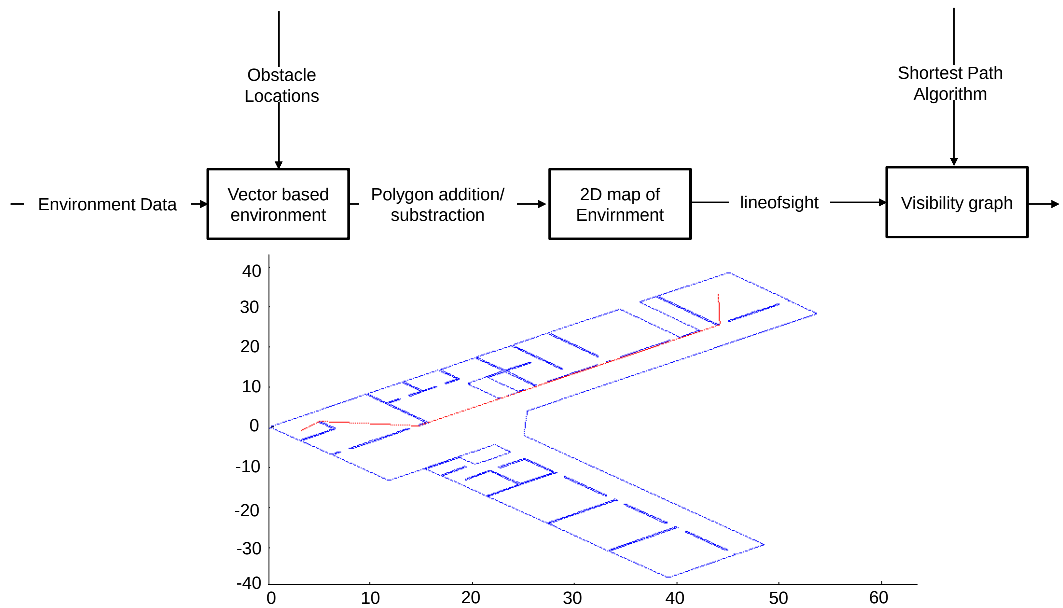


Fig. 15. Finding shortest possible path from 2 arbitrary locations in environment

such as visibility graph in order to prevent collision. Our solution models the environment in vector-based form with geometric computation of visibility graph. The environment in this case will be considered in 2D space with projection of objects such as walls, doors, robots, obstacles in to 2D floor plan. As a result, the map of environment at any given time is a polygon with holes as obstacles. The optimal path will be an optimal-path for the robot to move inside the polygon without crossing holes.

In an average map of few thousands vertices, the optimal-path itself is not a trouble for real time calculation. However, as the environment is dynamic (with moving object), we need to update all the time the optimal path which causes most trouble in the process. We have applied a method to reduce the complexity which composes of two main parts: 1) vertices pruning and 2) environment fragmentation.

As shown in Fig. *15*, the number of vertices of a visibility graph is very huge, vertices pruning aims to reduce the number of considered vertices in the environment and consequently improve the performance. One important feature of line-of-sight scheme is that we could be benefit from vertices properties especially the convexity of vertices. An easy observation is that only concave vertices of boundary polygon and convex vertices of holes would be important in searching for light-of-sight, other vertices would be easily covered by these vertices. Thus, by scanning and collecting only concave vertices for boundary polygon and convex vertices of holes in an linear search, we could eliminate as good as $\frac{1}{4}$ of the vertices *(in average)*. This would greatly enhance the process of building visibility graph. With these techniques, the system then could generate visibility graph of environment and run any traditional shortest path algorithm to find the optimal path. The process includes collecting environment information

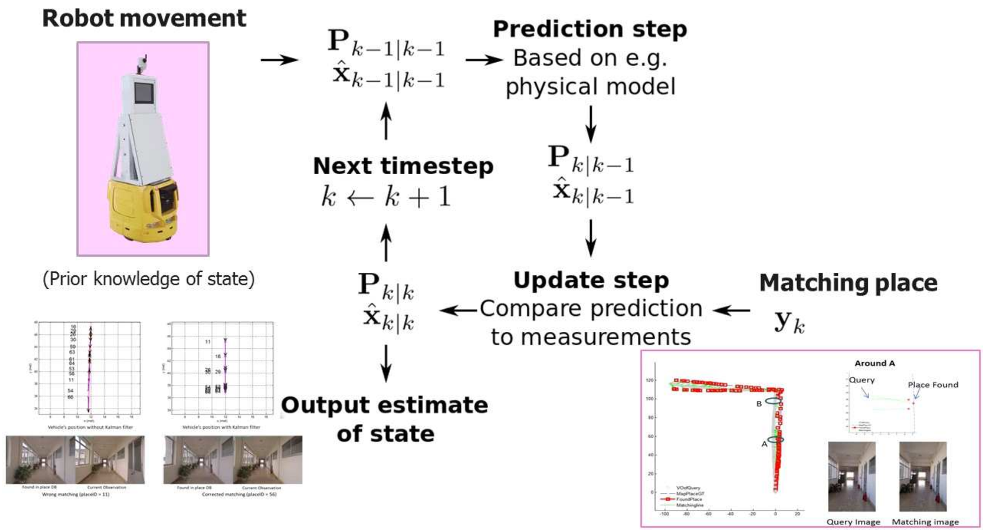


Fig. 16. *Kalman filter model predicted location for robot*

from model, generating 2D floor plan as well as visibility graph, finding shortest path at given time t. This process is repeated during navigation if any moving obstacle is detected.

In our context, the observations of the robot are images captured over time, which are then converted to coordinates *(x, y, z)* in a predefined coordinate system using above matching procedure. However, in indoor environment, the scene does not always change significantly. Consecutive scenes could repeat when the robot moves. Therefore, the performance of image matching is not good. Sometimes, a current observation could be matched with a very far *forward / backward* image that makes incorrect localization of the robot. To overcome this problem, we propose to use a *Kalman filter* to correct the position of the robot from observation as shown Fig. *16*. A Kalman filter is one of the most popular techniques to improve SLAM results.

In our context, we suppose that the robot moves in a flat plane, so the z coordinate of the robot is constant then we can ignore it. The state vector of the robot at a given time k is simply presented by its coordinates and velocity in two directions x and y. *Observation vector* is defined at each time where the image matching is found, the position of the robot could be estimated. We use this information as observation in Kalman filter. *State transition model* F_k allows to predict the state vector at time $k+1$:

$$x_{k+1} = F_k * x_k + \omega_k \tag{4}$$

where ω_k is process noise, which is assumed to follow a normal distribution with co-variance $Q_k : \omega_k \sim N(0, Q_k)$. Observation model H_k maps the true state space into the observed space:

$$z_k = H_k * x_k + v_k \tag{5}$$

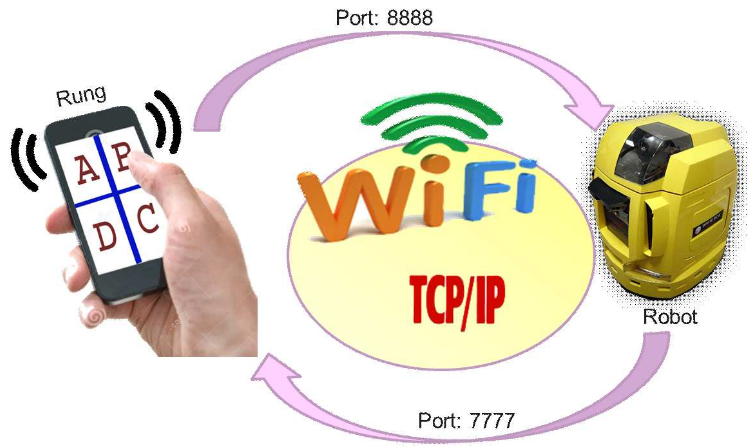


Fig. 17. *Communication model using TCP/IP between human and robot*

In our case: $H = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ where v_k is observation noise which is assumed to be *zero* mean Gaussian white noise with covariance $R_k : v_k \sim N(0, R_k)$. The control of robot movement will be based on the results of path finding.

3.4.2. Human-Machine Interaction: For communicate with the robot, we propose to use the smartphone as a wireless keyboard to send control commands to the robot and receive feedback commands from the robot (though vibration sensor).

- **Smartphone**: Connects to the robot via *Wi-Fi* local network using *TCP/IP* protocol. The interactive interface on touch screen of smartphone is divided into four regions corresponding to the four predefined positions. The smartphone sends messages to Robot on port 8888 and receives messages from Robot on port 7777. Once Robot and smartphone are connected together, Robot will send a message to trigger the vibration sensor in order to help the human navigating in environment.
- **Robot**: Receives command messages from smartphone, activates all services such as localization, path finding then feedback to the human's smartphone a message every $100ms$. This message provides information about the type of pre-defined vibrations belonging to the set of controls *turn left, turn right, go straight, and stop*. To understand the vibrational frequency, visually impaired people have to be trained in a duration of time.

3.4.3. System integration: Each modules presented above has been developed separately. In order to build the whole system, we have to integrate these modules. Fig. *18* shows how these modules are integrated in the system.

- **Step 1**: The positioning module will automatically determine the coordinates of the

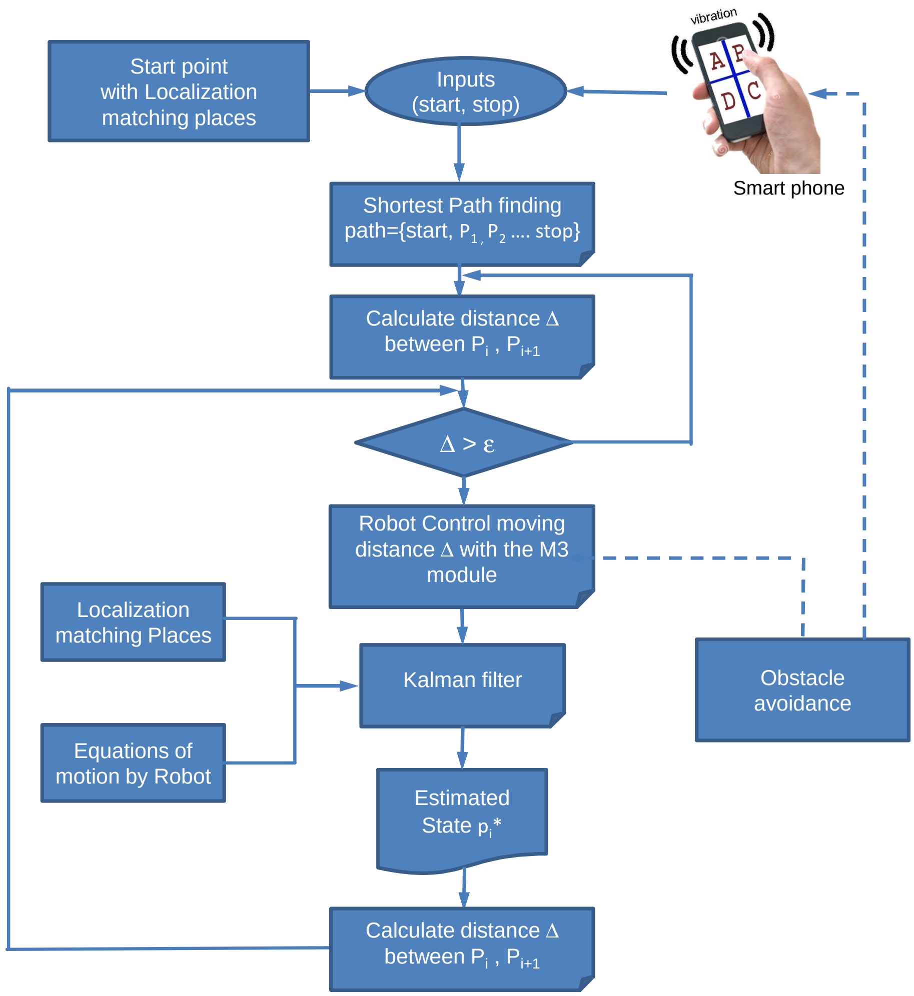


Fig. 18. *Integrated modules for robot navigation system*

starting point (start point) of Robot. At the same time, the system *"listens"* to receive user commands sent from smartphone.

- **Step 2**: The shortest path finding module is enabled. The returned result is trajectory composing of a set of intermediate points $\{path = p_1, p_2, p_3, .., p_N\}$ where p_1 is the *start point* and the *stop point* p_N;
- **Step 3**: For each pair of adjacent points in the path, the system computes the distance $D = dist(p_i, p_i + 1)$;
- **Step 4**: The system verity the condition $D > e$ where e is the *minimum* threshold that Robot can move. In this system, we set $e = 20cm$;
- **Step 5**: If the condition is not satisfied, the system backs to Step 4. Otherwise, it goes to Step 6;
- **Step 6**: Move the robot a distance D. Specifically,
    - ◇ *Step 6.1*: When the robot moves, it observes the environment. At a current position

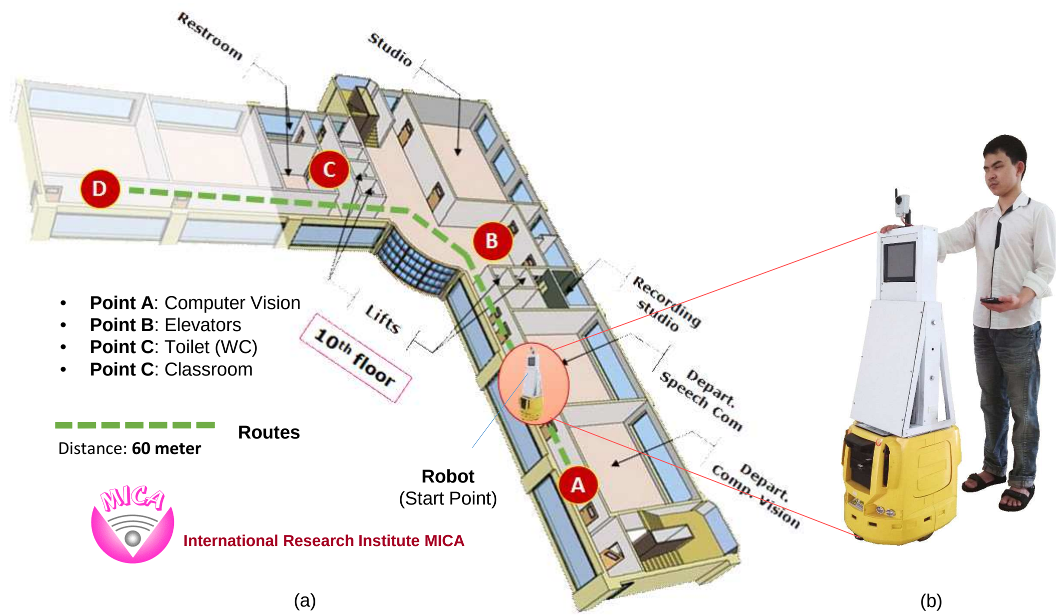


Fig. 19. (a) A 3-D map of the evaluation environment; (b) the blind clinging to Robot navigation

of the robot, we collect one image every a second. Module localization is activated to locate the current position of the robot. Simultaneous equations of motion of the robot is set up with a fixed velocity v. In experiment, we set v = 200mm/s to match the walking speed of visually impaired people. This information is used to formulate the models of Kalman Filter to predict the actual location of the Human-robot (position p_i*).

◇ *Step 6.2*: Compute the distance $D = dist(p_i*, p_{i+1})$.

◇ *Step 6.3*: During the whole process, the module of obstacle detection is activated. A detected true positive will stop the robot movement.

◇ *Step 6.4*: Repeat *Step 4* to check the *D* and *e*.

4. Experimental Results

4.1. Evaluation Environments

We evaluate the proposed method in a corridor environment of a building, where is 10^{th} floor of *International Research Institute MICA-Hanoi University of Science and Technology (HUST)* as shown Fig. *19*. We collect data in four times (trials), as described in Table *3*.

4.2. Evaluate the localization of starting point

As mentioned previously, we will use a robot to help blind person navigating in the environment from any position on the map to a predefined destination point. To this aim, we have to localize the robot at the current time. In this section, we will examine if the proposed

Table 3. Four 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

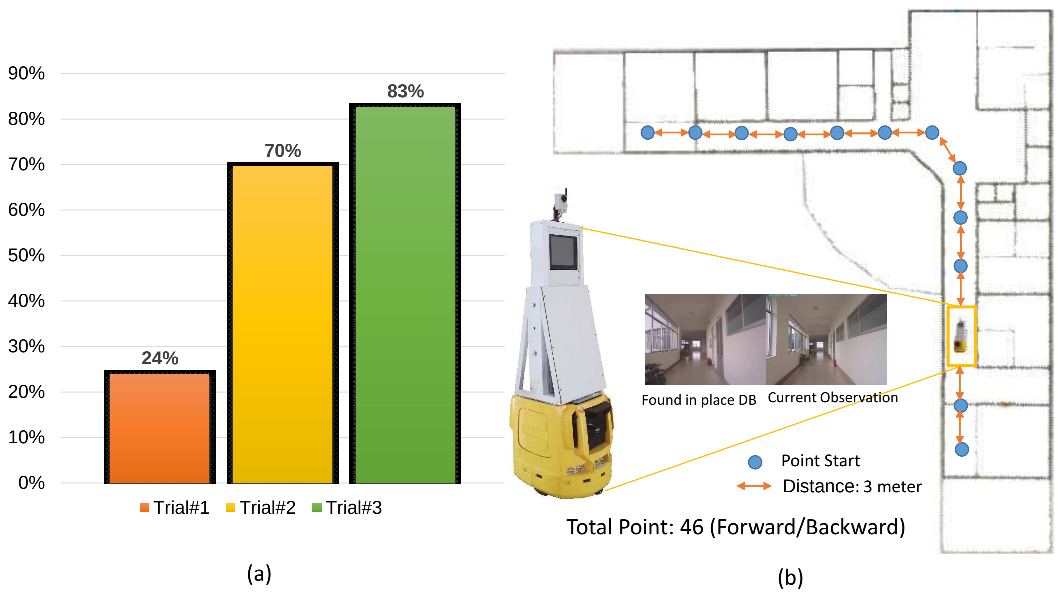


Fig. 20. The performance of localization of starting point of Robot. (a) Evaluation chart prediction starting point. (b) The starting point in the environment

localization method is good enough to identify starting position of the robot. Following round-trip along the corridor of total length about $60m$, every $3m$, we take a point and consider it as the starting point of the robot.

This procedure is repeated at 3 different times (Testing #3- morning, Testing #2- afternoon, Testing #1- evening). For each testing time, the total number of sampled *point is 46*. To identify the position, we take 100 consecutive images then apply the image matching as presented in section 3-C. We determine the most repeatable matched location when its reputation is large enough (larger than 70% in our experiment). The robot is then considered as being localized at that position. The Fig. *20* shows the performances with 3 trials in term of recall. The best result among 3 trials is 83%. These results show that environment conditions have a relatively big impact to the experimental results. They also suggest us for further improvement so that the proposed algorithms are solid to the lighting conditions.

4.3. Evaluate the role of Kalman filter

Given the starting point of the robot and the destination point, based on the environment information, we could determine the shortest path to follow.

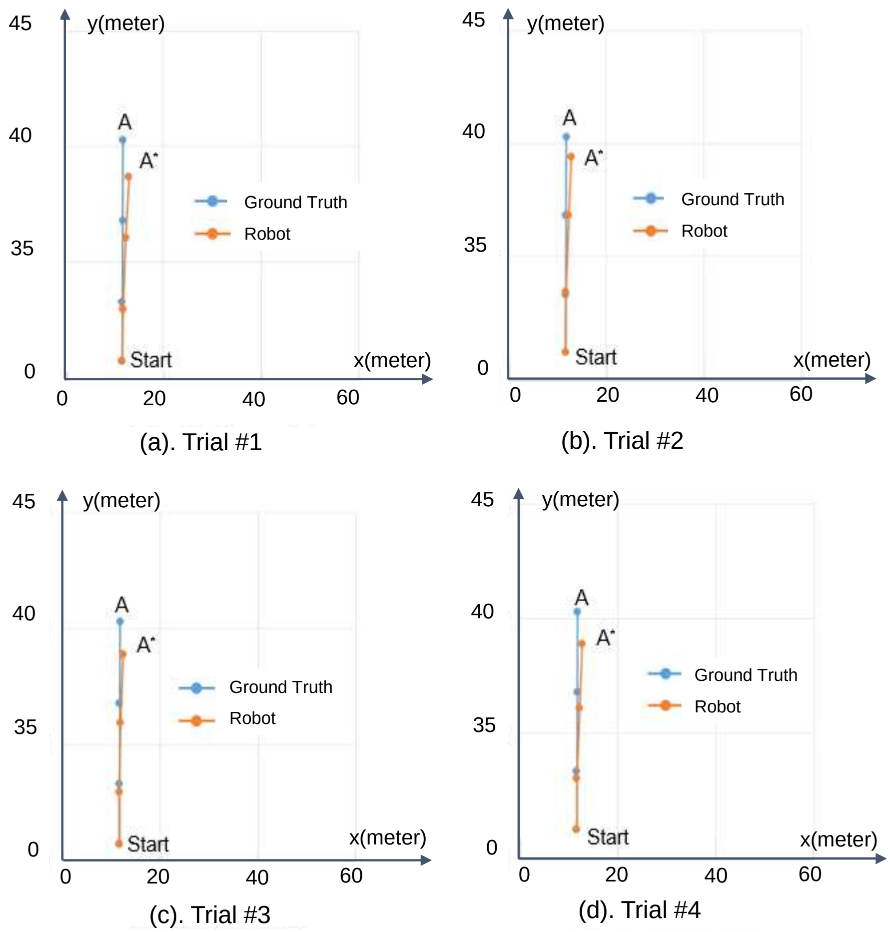


Fig. 21. Comparison of real positions of the robot to the groundtruth ones

However, when we make the robot moves, dues to the mechanical errors, the robot cannot attend exactly the positions provided by drive/control module. The Fig. *21* shows the real positions of the robot that drifts away the ground truth ones. Because of these reasons, the robot cannot reach the desired position. We propose to correct the drift with the use of image matching based localization in combination with Kalman filter (see Section 3.4).

We conducted four experiments in each the robot moves follows a straight road of *length 30 m at velocity about 200 mm / second* in the morning. To show the effectiveness of Kalman filter, Fig. *22* demonstrates navigation data without and with using Kalman filter.
Using only the place recognition results, the directions supporting navigation services are obviously uncontrolled. Some matching places (show in numbers) are misses and in the wrong order in this case. The main reason is the erroneous matching of some places *(e.g. place ID = 11, shown in bottom panel)*. By using a Kalman Filter, directions supporting navigation services are correctly ordered. To evaluate the localization accuracy in case of use / nonuse of Kalman filter, we measure median and average error of estimated position with

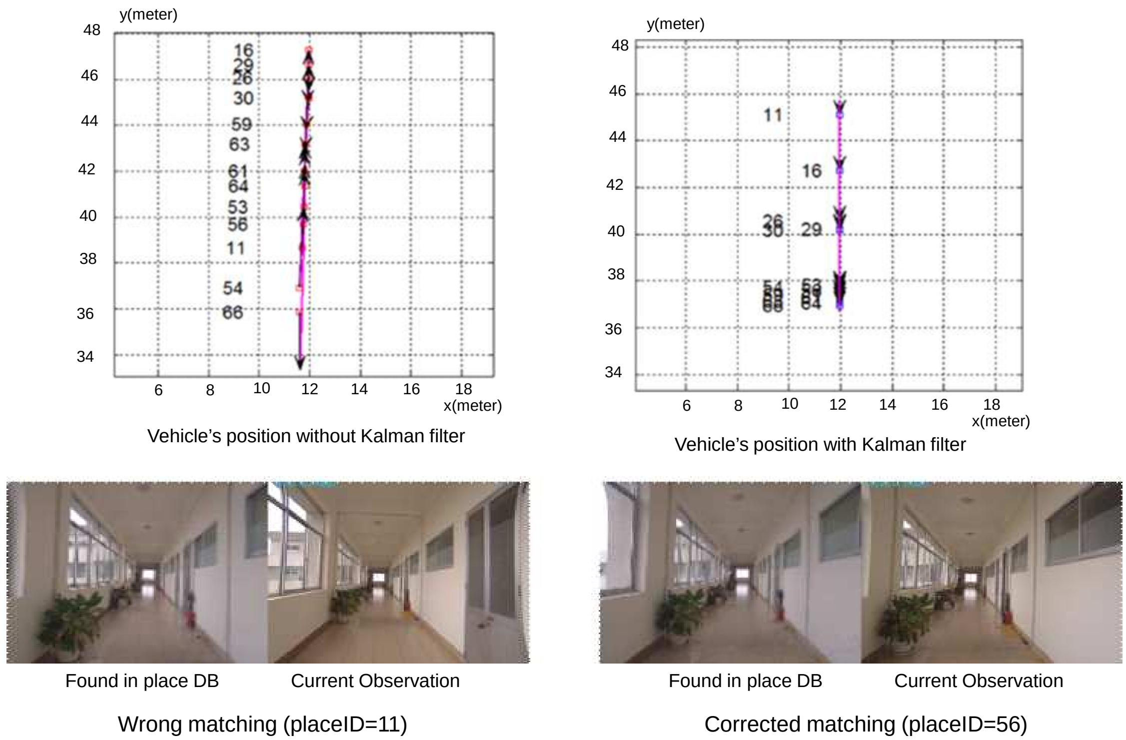


Fig. 22. Vehicle moving without/with Kalman Filter. Top row: Left panel: vehicle positions on the map using only results of the matching image-to-map procedures. The arrows show directions to guide vehicle. Numbers on left of each red box show placeID *of the current observation. Right panel: positions of the vehicle are updated using Kalman filter. Bottom row: Left panel: This result shows wrong direction to vehicle. Right panel: is a good matching with Kalman filter.*

Table 4. Average error (in meter)

Step	L#1		L#2		L#3		L#4		**Average**	
	avg	*std*	*avg*	*std*	*avg*	*std*	*avg*	std	avg	std
Matching	0.6	0.4	0.6	0.6	1.0	1.0	1.3	0.7	**0.9**	**0.7**
NoMatching	0.6	0.5	0.7	0.4	1.6	2.0	2.3	3.2	**1.4**	**1.6**

respect to the ground truth one. The evaluation results are presented in the table below.

The best accuracy is obtained with the 2^{nd} trial $(\Delta \sim 0.4meter)$. We investigate in more detail the deviation at each position on the trajectory. Matching and NoMatching (L#2, L#3) as shown in Fig. *23*. Consequently, if we do not utilize the Kalman Filter and imaging matching, the error rate is $\Delta \sim 1.2m$ high; whereas, by using Kalman filter, the localization error rate significantly reduces to $\Delta \sim 0.4m$.

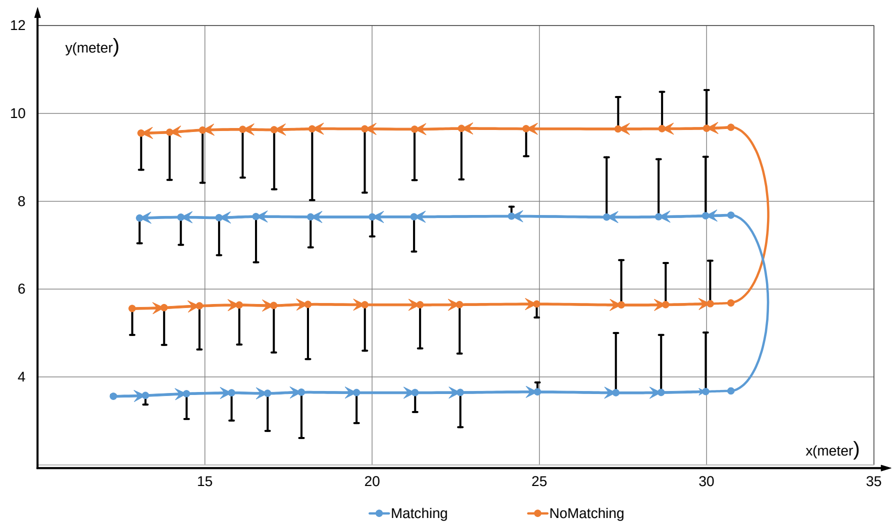


Fig. 23. Compare the results of locate error between Matching and NoMatching

4.4. Navigational aids to visually impaired people in Small Pervasive Environments using robot

To show the feasibility of using robot to help blind people navigating in environment, we have asked a blind pupil to follow the robot as shown Fig. *24*. The human-machine interaction is carried out through the mobile phone. According to guidance by vibration sensor mobile smartphone, he could go whole travel in the corridor environment. We obtain average performances of the matching $\sim 88\%$ recall, detection obstacles $\sim 73.85\%$ precision, localization $\sim 66\%$ recall, locate error $\triangle \sim 0.4m$. These results are acceptable to deploy the system in reality.

5. Conclusions

In this paper, we proposed a system for assisting the visually impaired people moving in small pervasive environments. The proposed method was deployed on a mobile robot. The proposed system consisted of a series of module: environmental mapping, positioning, optimal path finding, human-machine interaction. To build such modules, we utilized start-of-the-art techniques of Visual Odometry, Fast-appearance based mapping, Kalman filters, and shortest Path. These modules have been integrated successfully on robot. Empirical evaluation on a real blind user has been carried out in the corridor of a building. The result is very promising which motivate us to deploy such a system in reality. For future work, we continue to improve the accuracy of localization and test with more users in different environments such as inside a library, museum, a public hall, and so on.

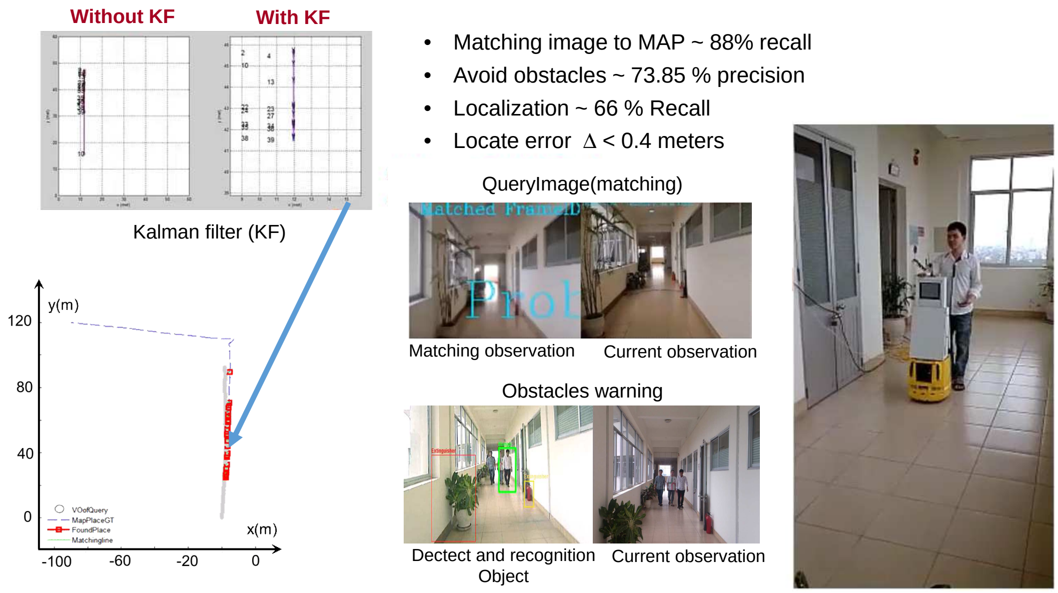


Fig. 24. A blind pupil follows Robot in the corridor of a building. The places on his travel are marked on the constructed map as red rectangles

Acknowledgment

This work was supported by the Research Grant from Vietnam's National Foundation for Science and Technology Development (NAFOSTED), No. FWO.102.2013.08.

References

[1] Tim Bailey und H. Durrant-Whyte: *Simultaneous localization and mapping (SLAM): part II*; *Robotics Automation Magazine, IEEE*; 13(3), S. 108–117; Sept 2006.

[2] J.P. Bigham, C. Jayant 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*; S. 65–72; June 2010.

[3] Johann Borenstein und Iwan Ulrich: *The guidecane-a computerized travel aid for the active guidance of blind pedestrians*; in *Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on*; Bd. 2; S. 1283–1288. IEEE; 1997.

[4] C Chow und C Liu: *Approximating discrete probability distributions with dependence trees*; *Information Theory, IEEE Transactions on*; 14(3), S. 462–467; 1968.

[5] Mark Cummins und Paul Newman: *FAB-MAP: Probabilistic localization and Mapping in the space of Apperance*; *International Journal of Robotics Research*; 27(6), S. 647–665; 2008.

[6] Dimitrios Dakopoulos und Nikolaos G Bourbakis: *Wearable obstacle avoidance electronic travel aids for blind: a survey*; *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*; 40(1), S. 25–35; 2010.

[7] Navneet Dalal und Bill Triggs: *Histograms of oriented gradients for human detection*; in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*; Bd. 1; S. 886–893. IEEE; 2005.

[8] M Everingham, L Van Gool et al.: *The PASCAL visual object classes challenge 2009*; in *2th PASCAL Challenge Workshop*; 2009.

[9] Pablo Fernández Alcantarilla: *Vision based localization: from humanoid robots to visually impaired people*; 2011.

[10] F. Fraundorfer und D. Scaramuzza: *Visual Odometry : Part II: Matching, Robustness, Optimization, and Applications*; *Robotics Automation Magazine, IEEE*; 19(2), S. 78–90; June 2012.

[11] David Hamme und Peter Veelaert: *Robust visual odometry using uncertainty models*; *In Proceeding of the Advances Concepts for Intelligent Vision Systems*; 6915, S. 1–12; 2011.

[12] Peter E Hart, Nils J Nilsson und Bertram Raphael: *A formal basis for the heuristic determination of minimum cost paths*; *Systems Science and Cybernetics, IEEE Transactions on*; 4(2), S. 100–107; 1968.

[13] Abdelsalam Helal, Steven Edwin Moore und Balaji Ramachandran: *Drishti: An integrated navigation system for visually impaired and disabled*; in *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*; S. 149–156. IEEE; 2001.

[14] Richard E Korf: *Iterative-deepening-A: an optimal admissible tree search*; in *Proceedings of the 9th international joint conference on Artificial intelligence-Volume 2*; S. 1034–1036. Morgan Kaufmann Publishers Inc.; 1985.

[15] Vladimir Kulyukin, Chaitanya Gharpure et al.: *RFID in robot-assisted indoor navigation for the visually impaired*; in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*; Bd. 2; S. 1979–1984. IEEE; 2004.

[16] Jason J Liu, Cody Phillips und Kostas 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*; S. 23–30. IEEE; 2010.

[17] Jack M Loomis, Reginald D Golledge und Roberta L Klatzky: *GPS-based navigation systems for the visually impaired.*; 2001.

[18] Vidya N Murali und James M Coughlan: *Smartphone-based crosswalk detection and localization for visually impaired pedestrians*; in *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*; S. 1–7. IEEE; 2013.

[19] Suranga Nanayakkara, Roy Shilkrot et al.: *EyeRing: a finger-worn input device for seamless interactions with our surroundings*; in *Proceedings of the 4th Augmented Human International Conference*; S. 13–20. ACM; 2013.

[20] Paul Newman und Kin Ho: *SLAM-loop closing with visually salient features*; in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*; S. 635–642. IEEE; 2005.

[21] David Nister und Henrik Stewenius: *Scalable recognition with a vocabulary tree*; in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*; Bd. 2; S. 2161–2168. IEEE; 2006.

[22] Aude Oliva und Antonio Torralba: *Modeling the shape of the scene: A holistic representation of the spatial envelope*; *International journal of computer vision*; 42(3), S. 145–175; 2001.

[23] Vivek Pradeep, Gerard Medioni und James Weiland: *Robot vision for the visually impaired*; in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*; S. 15–22. IEEE; 2010.

[24] A. Quattoni und A. Torralba: *Recognizing indoor scenes*; S. 413–420; June 2009.

[25] G. Schindler, M. Brown und R. Szeliski: *City-Scale Location Recognition*; in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*; S. 1–7; June 2007.

[26] S. Shoval, J. Borenstein und Y. Koren: *Auditory guidance with the Navbelt-a computerized travel aid for the blind*; *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*; 28(3), S. 459–467; 1998.

[27] J. Sivic und A. Zisserman: *Video Google: a text retrieval approach to object matching in videos*; in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*; Bd. 2; S. 1470–1477; Oct 2003.

[28] Niko Sunderhauf und Peter Protzel: *Brief-gist-closing the loop by simple means*; in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*; S. 1234–1241. IEEE; 2011.

[29] Paul Viola und Michael J Jones: *Robust real-time face detection*; *International journal of computer vision*; 57(2), S. 137–154; 2004.

[30] Tess Winlock, Eric Christiansen und Serge Belongie: *Toward real-time grocery detection for the visually impaired*; in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*; S. 49–56. IEEE; 2010.

Manuscript received 01-07-2014; accepted 07-11-2014.

■

Nguyen Quoc Hung received the Master degrees in computer science from Thai Nguyen University of information and communication technology in 2010. He is currently a PhD student of Hanoi University of Science and Technology. His research interests includes methods for acquiring, processing, analyzing in order to understand images and human-robot interaction.
E-mail: Quoc-Hung.Nguyen@mica.edu.vn

Vu Hai received B.E. degree in Electronic and Telecommunications in 1999 and M.E. in Information Processing and Communication in 2002, both from Hanoi University of Technology. He received Ph.D. in Computer Science from Graduate School of Information Science and Technology, Osaka University, 2009. He join MICA International Research Institute from 2012. He am interested in computer vision, medical imaging techniques, mainly video capsule endoscopy analysis for diagnostic assistance.
E-mail: hai.vu@mica.edu.vn

Tran Thanh Hai graduated in Information Technology from Hanoi University of Science and Technology. She is currently lecturer/researcher at Computer Vision group, International centre MICA, Hanoi University of Science and Technology. Her main research interests are visual object recognition, video understanding, and human-robot interaction.
E-mail: Thanh-Hai.Tran@mica.edu.vn

Nguyen Quang Hoan received the University degree the former Soviet Union (1973) Former head of the IT department and TT Internet and Library Academy of Telecommunications Technology (1998-2010). Chairman of Planning and Investment and Deputy Tr. IT department DHSPKT Hung Yen from 2010-Present. His research interests includes methods for Artificial intelligence (especially artificial neural networks), Computer Hardware, Modern control and system stability.
E-mail: quanghoanptit@yahoo.com.vn