Place Recognition for Indoor Navigation

Senior Honors Thesis 2010 – 2011
School of Computer Science

Extended Abstract

Samreen Anjum
samreen@cmu.edu
Carnegie Mellon University in Qatar

Advisors:

Brett Browning
brettb@cs.cmu.edu

M. Bernardine Dias
mbdias@cs.cmu.edu

Carnegie Mellon University/Carnegie Mellon Qatar/Robotics Institute

1 Introduction

Navigating inside buildings is often a straightforward task for sighted people but can be quite a challenging task for the visually impaired. Unless the environment is familiar, the visually impaired are generally dependent on others for directions, and thus suffer loss of independence and privacy. There have been several approaches to develop indoor blind navigation systems to overcome these challenges, such as Barcode detectors [1], RFIDs [2], Wi-Fi and Sensors [3-5]. However, in most of these approaches there exist challenges such as high cost of deployment, scalability, and lack of user orientation information. In this thesis, we aim to develop a system that uses place recognition to address these limitations. The fundamental goal of place recognition is to identify a location based on its visual appearance, an example of which is a shown in fig 1. This has applications to topological map building, navigation and loop closure in a mapping system [6-10] and is often deeply related to the task of image retrieval. This thesis seeks to build upon and enhance state-of-the-art place recognition techniques to create a robust and portable indoor blind navigation system that will allow the visually impaired to obtain directions to their destinations.

The specific research goal of this thesis is:

To develop and evaluate relevant place recognition algorithms that can be combined with an intelligent path planning algorithm on a portable device to enable independent navigation for the visually impaired in GPS-denied indoor environments.

Our approach focuses on separating the problem into two parts: mapping and localization. The mapping, or training phase, occurs offline and consists of building a map of the environment by a sighted person taking images that are stored for use in the localization phase. In operation, the user takes images, for instance, using a smart phone, and these images are used to register, or *localize*, the location of the visually impaired user in the map. After localization, the system will find the shortest path to the named destination (e.g. the "cafeteria"), and guide the user along the path. A successful navigation system must first ensure that the user has been localized effectively. Hence, place recognition is a vital component of indoor blind navigation and is the primary focus of this thesis work. The navigation component of this system is a simple implementation of a relevant path planning algorithm.

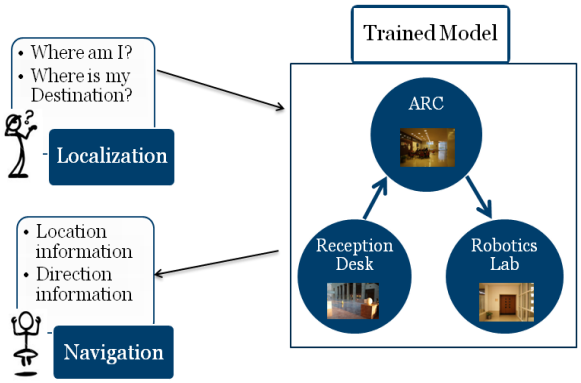

Fig 1: An example of the blind navigation system

The following sections describe the approach in detail, and demonstrate some experiments and the results obtained.

2 Approach

As mentioned earlier, the approach is divided into two main phases: mapping and localization. In this mapping phase, we use the bag of words approach [9] that has been developed and used in many other works. In the localization phase, we build a baseline system that generates a similarity vector by comparing the query image with the images in the trained system. In addition to the baseline system, we have performed four different validation tests to reject the false positives. The rest of this section describes the mapping and localization phase in detail.

2.1 Place Recognition: Mapping Phase

The training phase is an offline stage and consists of two significant components:

1. Building a dictionary of visual words based on the images,
2. Creating a map of the building and tagging the images with its appropriate locations.

The entire process is graphically illustrated in fig 3.

Building a dictionary of visual words, which is shown as step 3 in fig 3, involves data collection, feature extraction and clustering of the entire set of features based on their similarity. For data collection, several images of the testing environment are collected to train the model. To train the system with these images, we first extract significant features from these images to facilitate identification. There are several approaches to perform feature extraction such as SIFT, SURF, and MSER [11-13]. We use 128-dimensional SIFT (Scale-Invariant Feature Transform) descriptors [11], an example of

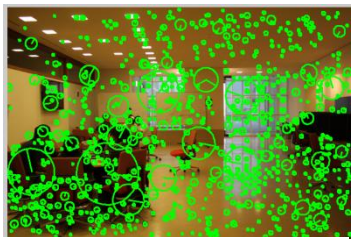

Fig 2: SIFT frames

which is shown in fig 2, for feature extraction as these descriptors are invariant to rotation, luminosity and scale. Our final step is to use the flat *k*-means clustering algorithm to cluster these features of all the images into *k* different clusters, and the cluster centers form the dictionary of *k* visual words, where k = {5000, 10,000}.

Our next step is to represent these images in a vector space model using a bag of words approach. The first step in this process is to represent each image in terms of the size of their features in each cluster. In other words, for each image *I* in the training data set, a *k*-sized histogram vector is generated; where each element *w* denotes the number of SIFT features from image *I* present in cluster *w*. We then apply tf-idf (term frequency – inverse document frequency) weighting [16] to these vectors to down-weight the common features in the entire dataset and increase the weight of those that are extremely significant. These normalized tf-idf vectors for each image form the respective image descriptors.

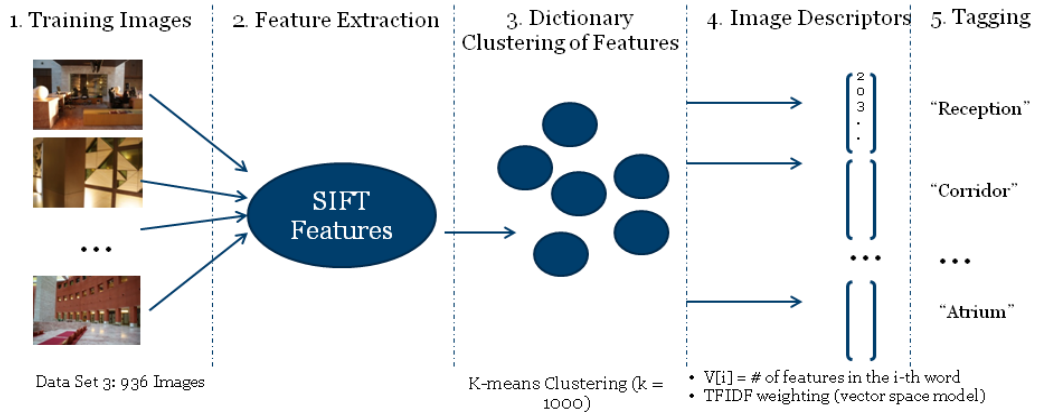

Fig 3: Graphical Representation of Phase 1 - Training Phase

In addition to collecting and storing images, we must train the model with the physical map of the building. In this stage, the official map of the building is represented as a graph where each node corresponds to a location in the map, and an edge between two nodes signifies that they are connected in the physical map. Each image in the training data set is then manually tagged to its appropriate location in the map.

At the end of this phase, we have a trained model that consists of various tagged images, and the map of the building. We now proceed to Phase 2 where we use this trained model to test the query images for image retrieval as well as image labeling.

2.2 Place Recognition: Localization Phase

This testing or localizing phase is an online stage where the visually impaired user will collect an image of his/her surrounding, and query the trained model for registration and localization against the map. To perform localization, we build upon the baseline system that generates a similarity vector indicating how similar the query image is to each image in the trained model. Additional verification tests are performed on the top-ranked results, and finally the graph of the building is used for labeling the image to a location (localization).

To develop a baseline system, we first represent the query image as an image descriptor using the visual dictionary, similar to the process described in the training phase. Then, we compare this query image descriptor with all the image descriptors in the trained model using the cosine distance metric,

$$similarity = \frac{Q.A_i}{|Q||A_i|}$$

where Q is the query descriptor and A_i is a trained image descriptor. This generates a ranking or similarity vector that illustrates the similarity of the query image with each image in the trained model. The higher the ranking, the closer the image is to the query image. When the results have been retrieved, the next step is to tag the query image with the location on the map of the building. To tag the query image, we consider the majority of the location tags of the results generated in the previous step. These labels had been manually assigned in the training phase.

However, there are several cases where the ranking vector provides false indications or requires refinement. To address this issue, we perform four additional verification tests:

1. **Homography Test** [14]: Although the ranking vector provides a good indication of how similar images are, it does not consider the spatial configuration of the SIFT features. For example, the matches returned could include flipped images as shown in fig 4. In this test, we perform image verification by using the RANSAC algorithm on the initial feature correspondences between the query image and the results obtained from the ranking vector. We extract these correspondences using a feature matching algorithm. Then an optimal homography is obtained for the correspondences and the number of inliers are counted. The results that are below a certain threshold of inliers are rejected.

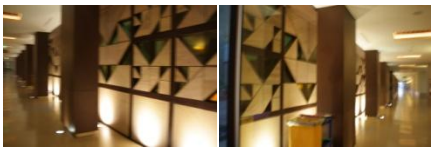

Fig 4: Flipped images of the corridor in CMU-Q

2. **Query Expansion Test** [15]: This is an image retrieval algorithm where an image is queried, and the results are generated. Of these results, the top 2-5 choices are picked, and queried recursively as single images. The results are stored, and the most common results in the cumulative set of results are chosen as the final output.
3. **Physical Graph Test:** In this test, we consider the location tags of all the results retrieved, and reject those images with tags that are in the minority. This test helps us to verify the tagging process for certain areas of the building that look similar to each other.
4. **Sequential Images Test:** This approach helps tackle the problem of tagging common areas, such as notice boards, plain walls, etc. In this process, we consider a sequence of images instead of just one query image. The user is requested to provide a query image as well as 2-3 neighboring images in the physical map. Finally, the system assigns a tag to the initial query image based on the tags obtained from the query image and these additional images.

In this thesis, we test and analyze these four different approaches independently and in different combinations. The system performance on image retrieval and image labeling is evaluated inside the Carnegie Mellon University Qatar (CMU-Q) building, and the results are recorded.

2.3 Navigation Advice

After successful localization, the next task is to direct the user to his/her destination. As mentioned earlier, the user will be prompted for a destination after localization. Once the user enters the destination into the system, an intelligent path planning algorithm will compute the path from the current location to the destination in the physical map. In this work, we use the Djikstra's path algorithm [17] to generate a path from the user location to the destination. This path is then narrated back to the user as approximate number of steps to facilitate navigation.

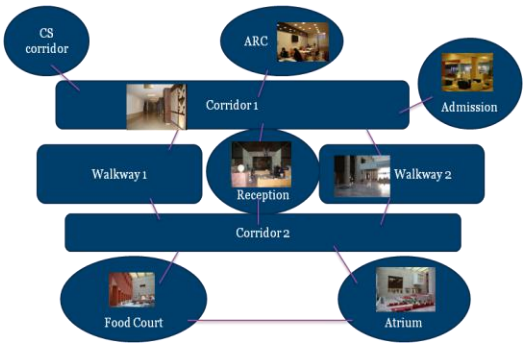

Fig 5: A graphical representation of CMUQ's first floor

3 Experiments & Results

We have tested our approach in the Carnegie Mellon University Qatar building. We collected a total of 1167 indoor images of the first floor to train the system. The total number of features extracted was 924,551 and the time taken to create of dictionary of these features was about 1 hour. The results obtained for different query images are shown in fig 6.

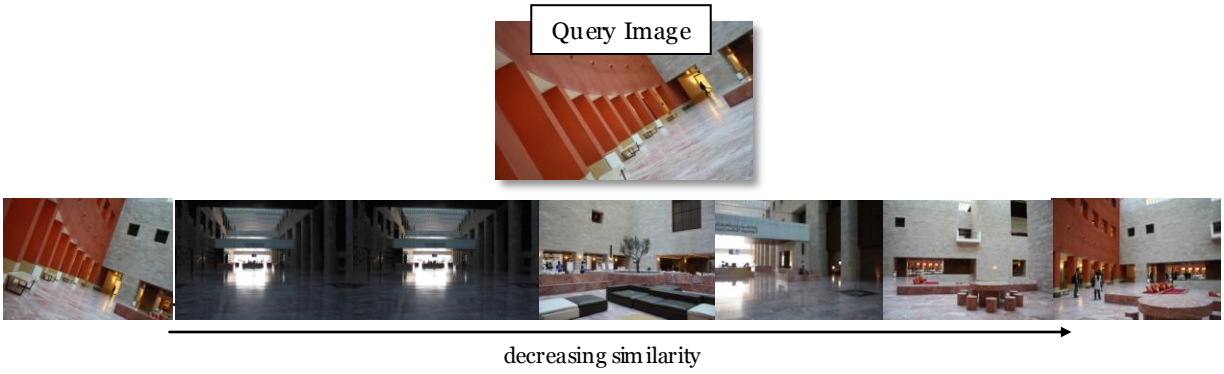

Fig 6(i): ***Ranking Vector Results****:*
Top Image: Query Image. Bottom Images: Top 7 images obtained from the ranking vector

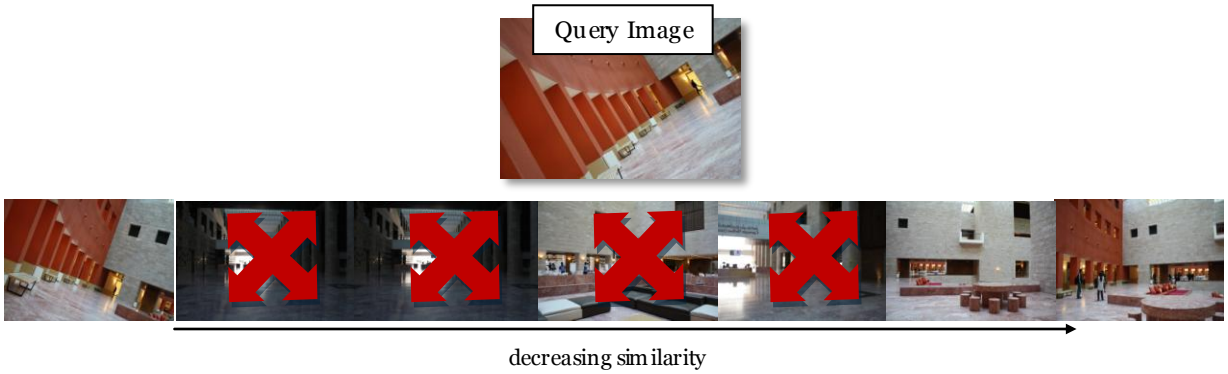

Fig 6(ii): ***Homography Test Results:***
Top Image: Query Image. Bottom Images: Images obtained from the ranking vector and the rejected by the homography test

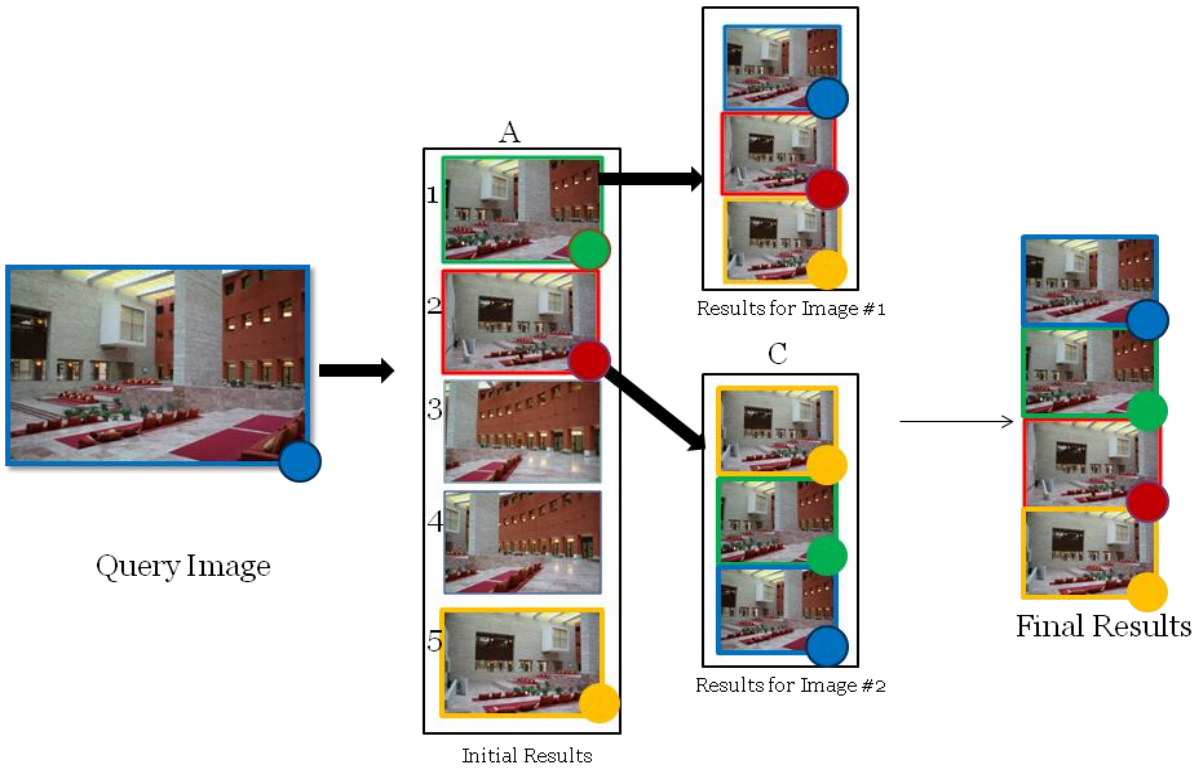

Fig 6(iii): ***Query Expansion Test Results:***
The image with the blue icon is the query image. The box with label 'A' contains the initial results obtained after querying this image. From the initial results, we pick the top two images – Image #1 and #2, and repeat the query process on each to generate their results. The boxes labeled as 'B' and 'C contain the results obtained from querying Image #1 and #2 respectively. The similar colored icons represent identical images retrieved in the process. The most common images retrieved after the recursive queries form the final results.

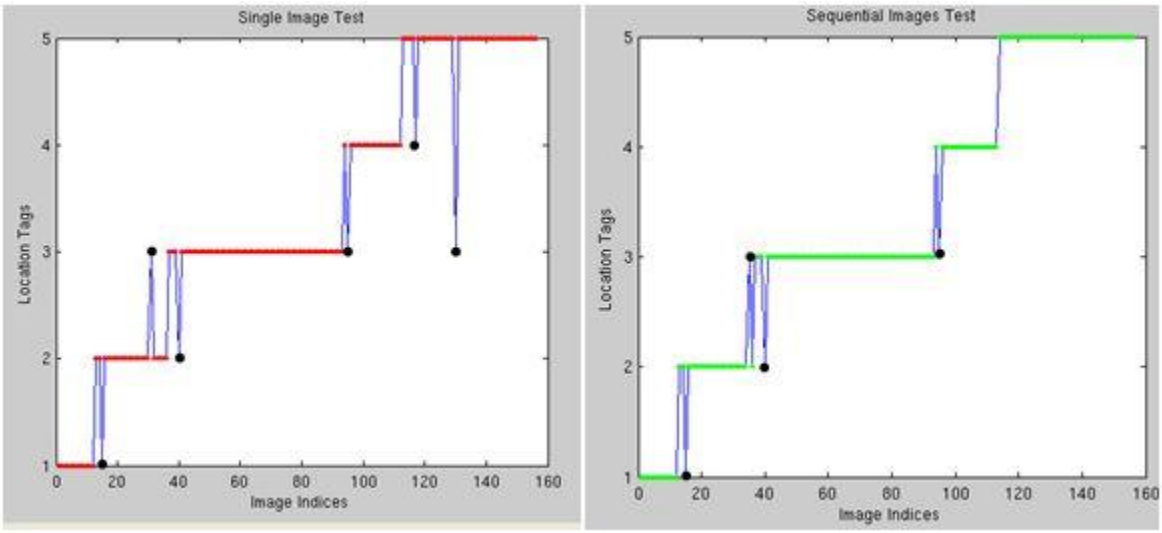

Fig 6 (iv): ***Image Tagging Results:***
This is an example of Image tagging along a path from the Reception Desk to the Academic Resource Centre at CMU-Q. The black dots represent the images that are tagged incorrectly. Graph A shows the results obtained by testing single images along the path. Graph B demonstrates the improved results upon consideration of two neighboring images in the sequence to tag a single image.

4 Conclusions

In this thesis, we present a place recognition based approach to enable the visually impaired to navigate independently in indoor environments. The approach is divided into two phases: the mapping or training phase, and the localization or testing phase. In the mapping phase, a trained model is developed which is trained with the images as well as the map of the building. In the localization phase, an image is queried and the trained model is used to obtain its location in the map. We accomplish this, we used built a similarity vector by comparing the query image to the images trained in the mapping phase. Then, we applied four additional tests – homography, query expansion, physical map, and sequential images, to validate the results given by the similarity vector. After localization, a simple path planning algorithm computes the path to the destination and provides it to the user. This work has been evaluated on the images and map of the Carnegie Mellon University Qatar building. The final thesis will include analysis to illustrate the performance of image retrieval and image labeling in the localization phase.

5 Acknowledgements

I would like to acknowledge Ameer Abdulsalam, Hatem Alismail, Peter Hansen, Samira Islam, and Ermine Teves for their help and contributions to this work.

6 References

1. J. Coughlan, R. Manduchi, and H. Shen, "Cell phone-based wayfinding for the visually impaired," Proc. IMV 2006, 2006.
2. S. Willis and S. Helal, "RFID information grid for blind navigation and wayfinding," Ninth IEEE International Symposium on Wearable Computers, 2005. Proceedings, 2005, pp. 34–37.
3. J.A. Hesch and S.I. Roumeliotis, "Design and Analysis of a Portable Indoor Localization Aid for the Visually Impaired," Jun. 2010.
4. A. Hub, J. Diepstraten, and T. Ertl, "Design and development of an indoor navigation and object identification system for the blind," ACM SIGACCESS Accessibility and Computing, 2003, pp. 147–152.
5. L. Ran, S. Helal, and S. Moore, "Drishti: an integrated indoor/outdoor blind navigation system and service," 2004.
6. A. Kawewong, N. Tongprasit, S. Tangruamsub, and O. Hasegawa, "Online and Incremental Appearance-based SLAM in Highly Dynamic Environments," The International Journal of Robotics Research, 2010.
7. I. Ulrich and I. Nourbakhsh, "Appearance-based place recognition for topological localization," IEEE International Conference on Robotics and Automation, 2000, pp. 1023–1029.
8. M. Cummins and P. Newman, "Probabilistic appearance based navigation and loop closing," 2007 IEEE International Conference on Robotics and Automation, 2007, pp. 2042–2048.
9. I. Posner, D. Schroeter, and P. Newman, "Using scene similarity for place labelling," Experimental Robotics, 2008, pp. 85–98.
10. M. Cummins and P. Newman, "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance," The International Journal of Robotics Research, vol. 27, Jun. 2008, pp. 647 -665.
11. J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, 2004, pp. 761–767.
12. H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: speeded-up robust features," *9th European Conference on Computer vision*, 2008, pp. 346–359.
13. D.G. Lowe, "Object recognition from local scale-invariant features," *iccv*, 1999, p. 1150.
14. R. Hartley and A. Zisserman, *Multiple view geometry*, Cambridge university press, 2000.
15. O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval," 2007.
16. C.D. Manning, P. Raghavan, H. Schütze, and E. Corporation, *Introduction to information retrieval*, Cambridge University Press, 2008
17. E.W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, 1959, pp. 269–271.