An Experimental Study of Mobile Device Localization

by

Ernesto Alejandro Ramirez

Submitted to the
Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degree of
Bachelor of Science in Mechanical Engineering

at the
Massachusetts Institute of Technology

June 2015

© 2015 Massachusetts Institute of Technology. All rights reserved.
An Experimental Study of Mobile Device Localization

by

Ernesto Alejandro Ramirez

Submitted to the Department of Mechanical Engineering on May 18, 2015 in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Mechanical Engineering

ABSTRACT

The objective of this experiment was to generate a dataset to be used for benchmarking text-based SLAM algorithms. A Google Tango and Android application were used and developed specifically for this experiment. The Tango logged visual-inertial-odometry, images, and depth readings at a rate of 1 hz each. The dataset was taken over 2 floors of MIT’s New House dormitory and was taken over the span of half an hour, covers about 3 miles of distance, and is 270 megabytes in size. The dataset taken and the Android application developed for the taking the data have been made publicly available for reuse.

Thesis Supervisor: John J. Leonard
Tile: Samuel C. Collins Professor of Mechanical and Ocean Engineering
Table of Contents

Abstract 3
Acknowledgements 4
Table of Contents 5
List of Figures 6
List of Tables 6
1. Introduction 7
2. Related Work 7
3. Implementation 9
   A. Hardware 9
   B. Software 9
   C. Data Set Overview 9
4. Results 10
   A. Accessing Application and Data Files 10
   B. Data Set Comparisons 10
   C. Data Set Processing and Visualization 11
   D. Application and Dataset Improvements 12
5. Conclusion 12
6. References 12
6. Appendix A: Floor Plan of New House 14
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1:</td>
<td>Image of different AprilTag families</td>
<td>8</td>
</tr>
<tr>
<td>Figure 2:</td>
<td>Maps generated from the Victoria Park and MIT Stata Center Datasets</td>
<td>9</td>
</tr>
<tr>
<td>Figure 3:</td>
<td>Image of MIT’s New House dormitory</td>
<td>10</td>
</tr>
<tr>
<td>Figure 4:</td>
<td>Map of path taken along dataset taken (VIO only and including feature detection)</td>
<td>11</td>
</tr>
<tr>
<td>Figure 5:</td>
<td>Example of a positive AprilTag detection.</td>
<td>11</td>
</tr>
</tbody>
</table>
1. Introduction

Simultaneous Localization and Mapping (SLAM) is the problem of constructing and updating an unknown/known map of the environment while localizing where an agent’s location is at the same time. The SLAM problem lends itself to giving robots the ability to autonomously navigate their surroundings. While range-finding, cameras, and a wide variety of sensors are used to attempt to solve this problem, few algorithms take advantage of text to build a semantic map for autonomous navigation. A recent paper has shown the effects of combining text-spotting and SLAM to be positive.\(^{1,8}\)

In order to evaluate and compare the new sets of algorithms that might arise from this bridging, a data set to benchmark algorithms must be made publicly available. This thesis describes a data set taken in halls of the New House dormitory at the Massachusetts Institute of Technology. It focuses on mobile devices localization, specifically phone and tablet. The data was taken with a Google Tango device, which provides an integrated visual-inertial-odometry (VIO) estimate as well as depth sensor output. The landmarks used for data collection contained text and April Tags.

2. Related Work

A. AprilTags

Developed by Edwin Olson, AprilTags are fiducial markers that provide 6 degree of freedom (DOF) localization based on features from an image.\(^{3}\) They have about 12 bits of information making it beneficial to phone/tablet computation because they have limited processing speeds. This payload allows the tags to be detected in a variety of lighting, poses, and locations.
AprilTags can also serve as an artificial landmark making them useful in evaluating SLAM algorithms. An example of different sets of AprilTags are shown below in Figure 1.

![AprilTag examples](Image)

**Figure 1:** Three examples of different AprilTag groups (36, 25, and 16 from left to right) with varying hamming distances (11, 9, and 5 from left to right). Image credit: AprilTag website. 

**B. Victoria Park and MIT Stata Center data sets**

Taken fifteen years ago, the Victoria Park data set was taken and made publicly available through the internet. The Victoria Park data set features an omni-camera, IMU, and GPS information and was focused on outdoor vehicle applications in operating in a large environment. The data was taken over a 300 m by 300 m square area in Victoria Park, Australia which can be seen in Figure 2. The MIT Stata Center data features multiple cameras, depth range, an IMU, and laser scanner information. The data set was taken over almost two years, covers 42 km of the building, and comprises 2.3 TB which allows it to make dense maps like in Figure 2.2,9

The two data sets show how different the information density taken from the environment can be. With the real-time localization on phones and tablets in mind, the data chosen to be looked at should be just rich enough to achieve localization while remaining computationally tractable.
3. Implementation

A. Hardware

All data was recorded on the Google Tango Tablet (development version). The Tango features a NVIDIA Tegra K1 w/ 192 CUDA cores, a 4 MP 2μm RGB-IR pixel sensor, an on-board accelerometer and gyroscope.

B. Software

The Google Tango runs on an Android OS so an application was developed specifically to collect data. The application saves images, VIO, and depth to local memory at a designated rate of 1 hz each.

C. Dataset Overview

The dataset was taken over 2 floors of MIT’s New House dormitory, seen in Figure 3. The data was taken over the span of an half an hour, covers about 1 miles of distance, and is 270 megabytes in size. The area’s that were mapped can be seen in appendix A. Forty posters with a unique ArriTag and word were placed around the dormitory to place provide unique features
throughout the building. An example of the poster can be seen in Figure 3. The dataset contains three forms of data: VIO as .txt files, images of the path taken in .jpg format, and depth frames in .pcd format. In the end, there were 570 files for each data type respectively.

![Poster Example](image)

**Figure 3:** Left: Picture of MIT's New House dormitory. Right: An Example of a poster placed around New House with the unique word and AprilTag.

### 4. Results

**A. Accessing Application and Data Files:**

The collected data set can be found at: [https://github.com/ernramirez92/newhousedataset](https://github.com/ernramirez92/newhousedataset).

The application used to collect the data can be found at: [https://github.com/ernramirez92/tango_tablet_logger_app](https://github.com/ernramirez92/tango_tablet_logger_app).

**B. Data Set Comparisons:**

The dataset presented in this paper is intended to mimic the sensor capabilities of consumer smartphones and tablets to perform localization in text rich environments. As opposed to the rich multi-sensor MIT Stata Center dataset, the dataset presented is focused more developing new algorithms to extract the most information from limited sensors. The Victoria Park and Intel datasets were important for LIDAR-based SLAM algorithm evaluation and comparisons. The goal of the New House dataset is to provide something similar for text-based SLAM algorithms.
C. Data Set Processing and Visualization

Two types of localization capable with the data taken are shown below. The first is a pure VIO localization of the path taken can be seen in the left half of Figure 4. The AprilTag locations, shown as colored squares, and the corridors drift on each pass. The second type involves using iSAM and AprilTags as landmarks which can be seen on the right half of Figure 4. The camera images taken were used to detect the AprilTags and an example of a detection is seen in Figure 5.

Figure 4: SLAM results: The orange lines are the trajectory estimates for the two cases: Left: VIO only. Right: SLAM with AprilTag detections.

Figure 5: Example of an AprilTag detection. The AprilTag is surrounded in a magenta box when detected.

The improvement in localization can be seen when comparing the top half of both graphs. The inclusion of iSAM eliminates most of drift where landmarks are abundant. However, this is
not the case in the lower half of the graph where there are less and farther spaced AprilTags. The real-time visualization of the two graphs can be seen that at:

http://youtu.be/5yT_27bsjxM.

D. Application and Dataset Improvements:

There are updates that the application used to log the data and dataset could have to improve their and each other’s quality. Working with Google Tango while still in development caused problems at times for the application. Pictures will often unexpectedly take a blue hue that can sometimes be removed by resetting the Tango. Removing this would allow the pictures to better represent the environment. Another improvement could be made by also saving the fisheye camera images as another source of data. Increasing the length of the dataset and including more loops would also make the dataset more valuable for algorithmic testing purposes.

5. Conclusion

The objective of this experiment was to generate a dataset to be used for benchmarking text-based SLAM algorithms. A Google Tango and Android application were used and developed specifically for this experiment. The dataset taken and the application are publicly available for reuse. Future work includes improving quality of pictures taken, capturing fisheye image to supplement the other data, and increasing the length and number of loops in the dataset.

6. References

1. Hsueh-Cheng Wang and Chelsea Finn et al. ,“Bridging Text Spotting and SLAM with Junction Features”,


Appendix A

Mapped area is shaded in gray

Floor 1