Visual Indoor Localization with a Floor-Plan Map

Hang Chu
Dept. of ECE
Cornell University
Ithaca, NY 14850
hc772@cornell.edu

Abstract

In this report, a indoor localization method is presented. The method takes first-person view video clip as input and uses a 2D floor-plan map as prior knowledge. Motivation, methodology, elementary experimental results, and further extensions are reported and discussed. This report also serves as a brief research summary for **CS 6756**.

1 Motivation

1.1 Indoor localization

The problem of indoor localization arises as the blocking of building structure usually causes deviation to GPS system. Indoor localization has lots of potential applications and is under extensive researching by various communities. Two main approaches have been studied [1]: On one hand, wireless hotspots such as wifi accessing points are used, their unique IDs and signals are recorded and analyzed to construct an indoor signal strength map and thus enables room-wise localization; On the other hand, more complex sensors such as laser scanner, camera and IMU (inertial measurement unit) are used solely or jointly to scan the indoor room structure and localize in the same time, this technique is known as SLAM (simultaneous localization and mapping).

Both of the two approaches have various well-developed, sophisticated implementations. However, a major issue exists in practical use. For wireless ID and SLAM approaches, the construction of map (signal strength map for the former, laser range map or visual word map for the later) is inevitable. That implies the necessity of a thorough walk-through of every building before putting the localization system into service. To address this issue, a 2D floor-plan map, which is widely available and easily accessible, provides a perfect solution.

1.2 Localization with maps

Being the most universal method when it comes for a person to find a location, map-reading is certainly an useful ability, which is true for both human and machines. For this project, two specific map-related previous work inspired the major framework.

1.2.1 GPS refinement with an image and map

In this work, a method for reducing GPS error in urban environment is presented. The refinement is based on firstly the analysis of building structural lines (see Figure 1) and camera pose (Figure 2), and then a reference to the region map (Figure 3). This method is proved to produce better location than GPS generally does (Figure 4), alone with a satisfactory orientation estimation.

Though the approach in this work is fairly simple and intuitive, it still reveals the potential of utilizing a map, to be more specific clues of building structure embedded in the map, to the task of

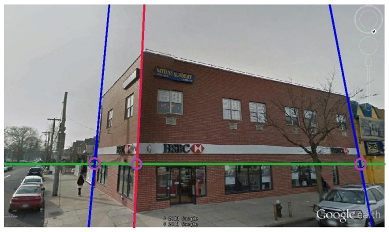

Figure 1: Vertical corner edges of building

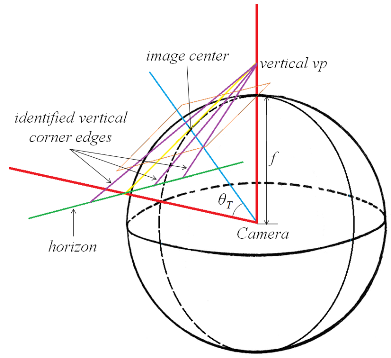

Figure 2: Computing camera tilt angle

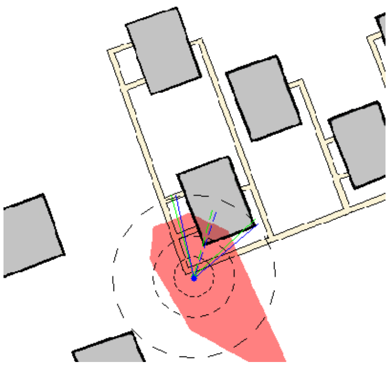

Figure 3: Localization by referring the map

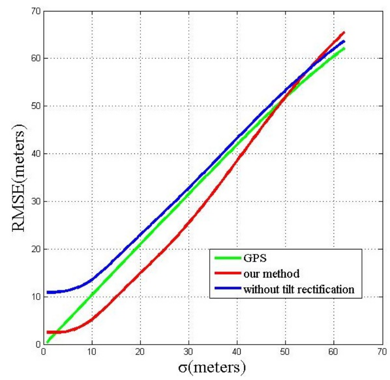

Figure 4: RMSE in location

localization. Similar to the outdoor situation, well-maintained indoor maps could be also useful. This is the work I concentrated on in the first half of the semester, it is not yet published [2].

1.2.2 Vehicle visual localization with a map

In the work of Brubaker et al. [3], a novel approach for vehicle localization is proposed. The only required sensor is a self-motion odometry sensor (in this work a dual-vision camera that is mounted on the car roof), and the only required pre-stage database is a road map extracted from any available online city map. To do localization, motion trajectory is estimated from the odometry input, then localization solutions are proposed based on the similarity of motion trajectory and road map structures. This is shown in Figure 5, localization accuracy increases as time passes and more structural information is provided by the motion trajectory.

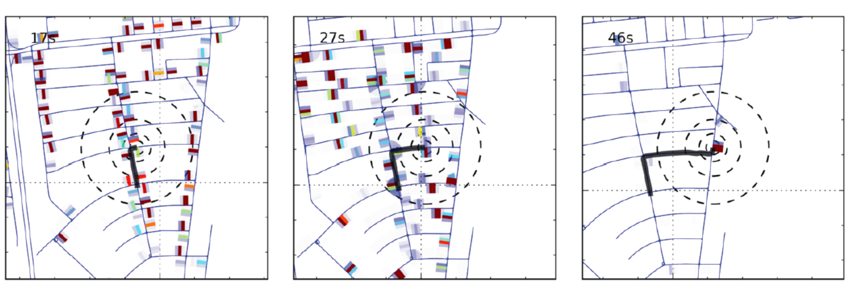

Figure 5: Brief illustration of localization technique in [3]

A similar strategy could be adopted for indoor localization. However, there exists two main differences between two problems. First, the indoor equivalent to outdoor road intersections are doors, but vehicles can only move on roads from one intersection to another, while a human can move through any wished path from one door to another. This implies that the solution space of indoor localization is much bigger than outdoor vehicle localization. Second, vehicles move steadily and the odometry sensor produces data with only 2 degree of freedom (assuming flat ground everywhere). This is not the case for pedestrian motion, whether using IMU (for example a cellphone IMU or shoe-mounted IMUs [4]) or using visual input (for example a Google glass or a waist-mounted camera [5]), the extracted motion trajectory is in 3D. This implies more noisy odometry input, thus more careful measurement calibration and more noise-robust localization algorithm are needed.

2 Approach

2.1 Localization with only kinematics

As a classic setting of indoor robot SLAM [6], the sensing system measures both *odometry* (referring to laser range measurements or visual feature matching with built visual-word-map here, might cause a little bit confusion with visual-odometry as they coincidentally chose the same word) and *kinematics* (measuring the ego-motion). In this work we want to skip the map-building part and only use daily-life sensors, so odometry is not available. The problem now becomes finding location using only kinematics (and of course a 2D floor-plan map).

Let's first not concern about noise in the measurement, which means perfect motion trajectory can be recovered from kinematics. Then finding the location simply becomes finding a location to place that trajectory on the map without hitting a wall. As shown in Figure 6, given a sufficient long trajectory and not a Manhattan-like map, satisfactory localization can be found.

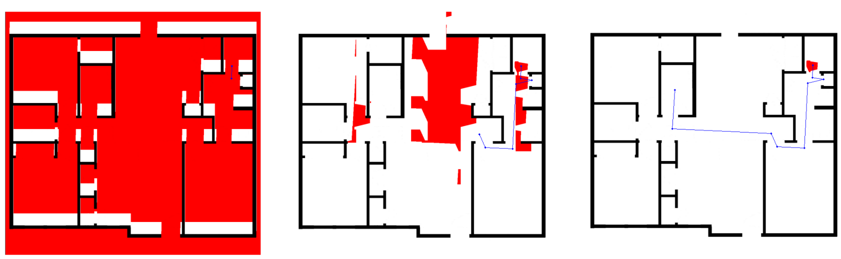

Figure 6: Localization with perfect trajectory. Blue curve shows the trajectory (in its ground-truth placement), red region shows possible placements of the *start* point of trajectory

However, perfect measurement is not possible. To deal with the noisy-input situation, the previous method can be modified into a particle-filter (also known as Monte Carlo Localization) style.

- Initialize location states $x_0^{(i)}$ using an uniform distribution, with equal weights $w_0^{(i)}$.
- For each step, measure ego-motion in 2D as m_t.
  - Sample $x_t^{(i)}$ according to $x_{t-1}^{(i)}$ with weights $w_{t-1}^{(i)}$.
  - Update states $x_t^{(i)} \sim p(x'|m_t, x)$.
  - Set $w_t^{(i)}$ as 0 if $x_t^{(i)}$ locates in wall region, C otherwise. Normalize $w_t^{(i)}$.

2.2 Motion trajectory estimation

There are multiple options for measuring the motion trajectory of a walking person, the most accurate one is using a camera. The technique for computing relative motions between frames is known as visual odometry [7]. As the key for solving it is essentially an epipolar geometry problem, visual odometry can be seen as a special case of general structure-from-motion. I start with the well-developed normalized 8-point RANSAC [8] and 5-point RANSAC algorithm [9]. Figure 8 shows

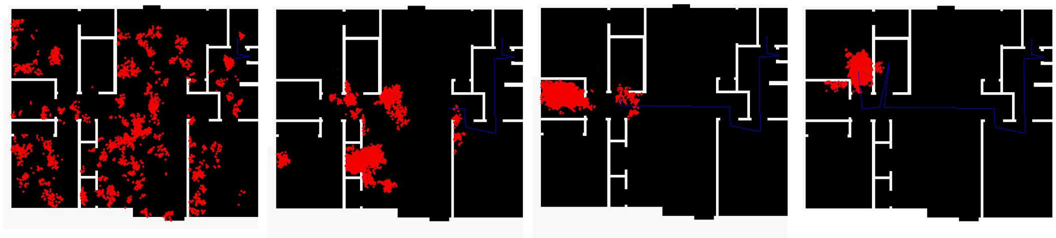

Figure 7: Localization with noisy trajectory. Blue curve shows the trajectory (in its ground-truth placement), red dots show location estimations of particle filter of the *current* location

sample frames of a testing video sequence taken in the Ward Lab, Figure 9 shows the results of mentioned algorithms. Both algorithms are unable to recover a highly accurate trajectory, but they both succeed in capturing the overall trend and fragmental little motions.

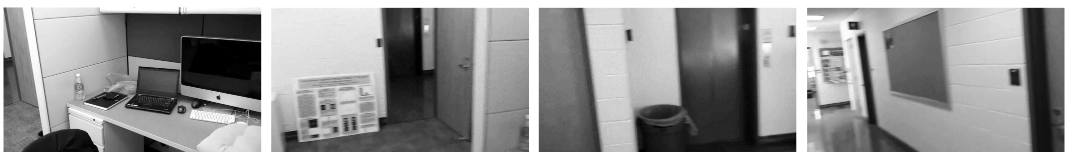

Figure 8: Sample frames of a 9-second video sequence of Ward Lab

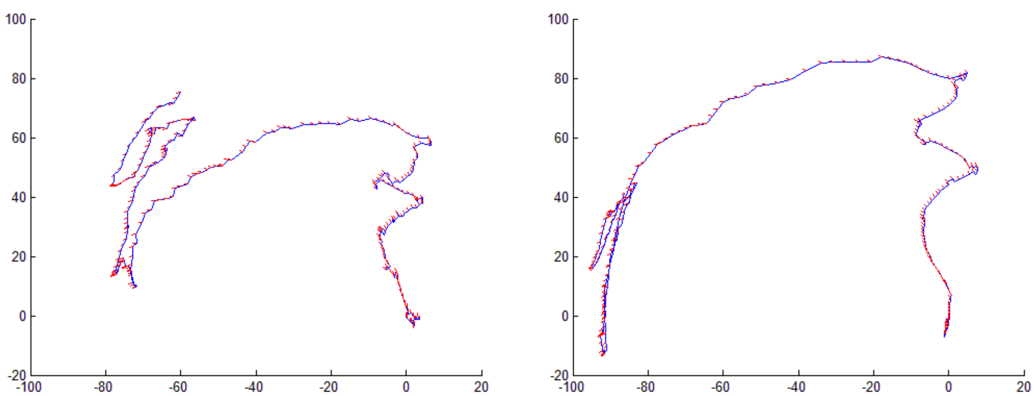

Figure 9: Trajectories of normalized 8-point RANSAC (left) and 5-point RANSAC (right), showing only the X-Z plane, assuming the first frame looking at positive-Z direction. Red needles show the estimated camera orientation.

3 Future works

As this is still an ongoing research project, only elementary framework and examples are shown in the previous section. This section we will focus on possible improvements and extensions that can be added to the framework.

3.1 More reliable and faster localization

The main problem in localization now is that it does not consider the orientation of trajectory placement. If we incorporate the initial orientation, the solution space becomes much larger. In classic robot localization, this is not a problem because laser range measurement and visual-feature matching significantly reduces the possible solution. In our situation as this not available, the only solution is to increase particle numbers so it covers the whole initial solution space well. That indicates a tradeoff between efficiency and reliability of the algorithm: more particles gives better coverage of the solution space but also makes the algorithm slower.

A possible solution is to adaptively change particle numbers. Particles form clusters as time passes, by adding a clusterness measurement, the algorithm is able to start with a large number of particles and reduce the number while it is running. Besides, mixture-monte-carlo process also offers a possible solution.

3.2 Obtaining better trajectory measurement

3.2.1 Improving accuracy of visual odometry

Three techniques could be used for producing more accurate motion trajectory. First is applying a filtering process such as the extended-kalman-filter, this might reduce the influence of noise while requires little computational resource. Second is applying a pose-graph optimization [10], which measures relative motion of any pairs of frames in a sliding window. Third is applying bundle adjustment, where the optimization focus on visual features rather than just on a frame-level. The last to techniques are both essentially a non-linear least square problem and are proved solvable efficiently, their drawback is that they require much more feature matching operations, which would increase computational resource demand significantly.

3.2.2 Other problems in visual odometry

Besides improve the accuracy of visual odometry, there exist two other problems, naturally due to the specific setting of this task. The first problem is how to deal with the dimension reduction from a 3D trajectory to a 2D one, to do this precisely the estimation of gravity direction will be needed. A possible solution is to investigate structural lines of walls and doors, which is normally vertical and helpful for estimating gravity direction. The second problem is reduce the interference of other moving object to motion estimation, though the RANSAC process already handles this issue, it might not be enough for an indoor scenario where a walking person is more likely to generate feature points than walls or stationary furniture. This is also a problem that needs to be solved for practical use of the entire framework.

[1] Oliver Woodman & and Robert Harle. (2008) Pedestrian localisation for indoor environments. *International Conference on Ubiquitous Computing*.

[2] Hang Chu, Andrew Gallagher & Tsuhan Chen. (2013) Efficient GPS refinement and camera orientation estimation from a single image and a 2D map. *CVPR submission*.

[3] Marcus A. Brubaker, Andreas Geiger & Raquel Urtasun. (2013) Lost! leveraging the crowd for probabilistic visual self-localization. *CVPR*.

[4] John-Olof Nilsson, Isaac Skog, Peter Handel, & K.V.S. Hari. (2012). Foot-mounted INS for everybody-an open-source embedded implementation. *Position Location and Navigation Symposium*.

[5] Jwu-Sheng Hu, Chin-Yuan Tseng, Ming-Yuan Chen & Kuan-Chun Sun. (2013) IMU-assisted monocular visual odometry including the human walking model for wearable applications. *ICRA*.

[6] Sebastian Thrun, Dieter Fox, Wolfram Burgard & Frank Dellaert. (2001) Robust Monte Carlo localization for mobile robots. *Artificial Intelligence*.

[7] Davide Scaramuzza, & Friedrich Fraundorfer. (2011) Visual odometry tutorial. *Robotics and Automation Magazine*.

[8] Richard Hartley, & Andrew Zisserman. (2000) *Multiple view geometry in computer vision* Cambridge.

[9] David Nister. (2004) An efficient solution to the five-point relative pose problem. *TPAMI*.

[10] Edwin Olson, John Leonard & Seth Teller. (2006) Fast iterative alignment of pose graphs with poor initial estimates. *ICRA*.