

Portable 3D Visual Sensor Based Indoor Localization on Mobile Device

Xiang He, Daniel Aloï, Jia Li

Department of Electrical and Computer Engineering

Oakland University

Rochester, MI 48309, U.S.A

xhe2@oakland.edu, aloï@oakland.edu, li4@oakland.edu

Abstract—Smart mobile device and consumer level 3D visual sensor are becoming very cheap and widely accessible in people's daily lives. Visual information based indoor localization is getting more and more attention in the computer vision research community to achieve real time and precise localization result. In this paper, we present a 6-degree-of-freedom (6-DoF) pose estimation using a portable 3D visual sensor mounted on a mobile device. A detailed 3D model and a WiFi received signal strength model of the indoor environment are constructed in the offline training phase. During the online localization, we first use WiFi signals to locate the device in a 3D sub model. The initial pose is calculated through feature matching between the online captured 2D image and the key frame images used to build the 3D model. Then we employ iterative closest point (ICP) algorithm to estimate the rigid transform between online captured 3D point cloud and local 3D point cloud. The 6-DOF pose estimation is further refined by random sample consensus (RANSAC) algorithm. We implement the system on iOS platform and the experiments carried out in an indoor environment shows promising result of our approach. To the best of our knowledge, this is the first portable 3D visual sensor based indoor localization system on mobile device.

Keywords—Smart mobile device, portable 3D visual sensor, indoor localization, 6-DoF pose estimation, iOS platform

I. INTRODUCTION

Precise indoor localization is a key component in many location-aware applications, such as navigation, interactive gaming, and merchandise advertising. People have come up with various approaches for smart mobile device (smartphone, tablet) user localization in GPS-denied indoor environment. To name a few, radio frequency identification devices (RFID), wireless fidelity (WiFi) and pedestrian dead reckoning (PDR) are the three frequently used ones. However, the RFID based system requires a large amount of tags to be installed in the indoor environment before the user can locate his own mobile device [1]. WiFi based localization techniques have the problem of signal fluctuation due to multipath fading effect in indoor environment [2]. The PDR approach suffers from the fact that the inertial sensors on the mobile device are low cost micro electromechanical system (MEMS) sensors, which has relatively low accuracy. Thus the integration drift will cause the localization error to accumulate over time, and become unacceptable in the long run [3]. Moreover, the above methods provide only location coordinates, and therefore have limits in further localization potential, such as obstacle detection, sign

recognition, and disabled assistance, etc., which are important in location-aware services.

In order to achieve the above goals simultaneously, computer vision technologies are introduced to the localization task. A majority of the existing visual information based localization approaches use 2D images to localize the mobile device's user. The main idea of 2D image-based localization is to first store the scene images, or features of these images, into a database, indexed and related to predefined geo-locations. A new query image to be localized is preprocessed with feature extraction and matched to the images in the database. The top ranking matched image is retrieved from the database. Since the retrieved image is geo-tagged, the localization task has become pose recovery between the matched images. Liang et al build an image based localization system following the above procedure [4]. The complete pipeline of their method from image retrieval (20,000 images in the database) to pose recovery takes 10-12 seconds to output a solution for a single query image on a 2.3GHz i5 laptop. The high computational cost makes it unpractical on mobile platform. Jaramillo et al develop a 6-DoF pose localization system using a monocular camera [5]. The database is constructed of a dense 3D point cloud, which can be projected to 2D to form a virtual view using the previous localized pose of the mobile device's camera. The 2D-3D point correspondences are obtained between the current captured image's 2D features and their matches on the virtual depth image (projected 3D points). This forms a perspective-n-point (PnP) problem which can be solve for the relative transformation between the current camera pose and the virtual view. Their results show that a 2D camera can be localized in a 3D model in real time. The most time consuming part of this method is the estimation of initial pose, which can take hours if it's in a large indoor environment. To improve the initialization process, Ruiz et al apply multiple sensors to estimate a coarse initial position [6]. They divide the indoor 3D model into different sub areas and refine the initial estimation with SIFT feature matching.

Inspired by previous works, we develop an indoor localization algorithm on mobile platform. A portable 3D visual sensor is mounted onboard to do localization in a prebuilt 3D point cloud. During the offline training phase, we not only model the indoor environment as a 3D point cloud, but also apply Gaussian process (GP) regression to model the WiFi received signal strength (RSS) dataset, which will be used in initial pose estimation. Once we have the 3D model and GP model of the indoor environment, we are ready for online localization. After the coarse estimation of the initial pose

using WiFi RSS data, the user location is narrowed down within a sub area of the 3D model. We further calculate the pose of the mobile device by matching features between online captured images with key frame images of the prebuilt 3D model. Then we can consecutively estimate the pose of the device by solving the rigid transform between online captured 3D point cloud and local 3D model using ICP algorithm [7]. Moreover, RANSAC algorithm [8] is applied to improve the 6-DoF pose estimation accuracy.

Specifically, we make the following contributions: We built an indoor localization system on iOS platform based on 3D visual sensor and WiFi RSS. To the best of our knowledge, our iOS App is the first application attempting to achieve 6-DoF pose estimation in an indoor environment modeled as a 3D point cloud and GP based WiFi RSS model. We believe this app will boost various location based services around it.

The rest of the paper is organized as follows. Section II explains the details of our indoor localization system, including offline training phase and online localization phase. The experimental results in typical indoor environment are presented in Section III. We conclude our work in section IV with a discussion of future improvement.

II. SYSTEM SETUP

This section describes the design of our proposed indoor localization method. Figure 1 presents an overview of the system workflow. During the offline training phase, color images, depth images and WiFi RSS are recorded in predefined survey points. By manually or automatically aligning the data in each survey point, we are able to generate a detailed 3D model of the indoor environment. The scanned WiFi RSS values and corresponding BSSIDs of the access points are used to train a GP model. Based on the survey point locations, we further divide the indoor environment into different sub areas. During the online localization phase, we first determine an initial coarse estimation that indicates a sub area where the mobile device user seems to be, using the WiFi RSS captured in the air. Next, making use of the image captured by the camera, we perform a matching process against the key frame images in the 3D sub model to determine the 6-DoF pose of the mobile device. Finally, we can continuously update the device pose by finding the 3D points correspondences between online captured 3D point cloud and local 3D model using ICP algorithm. In the following subsections, we will explain the details of the offline training phase and online localization phase.

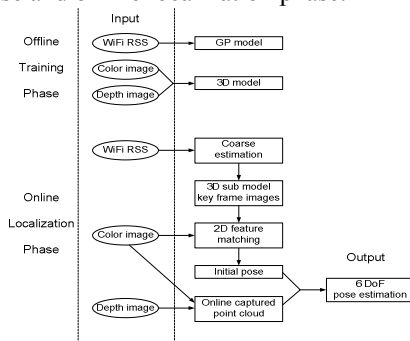


Figure 1: System workflow

A. Offline Training Phase

First of all, our localization method requires a detailed 3D point cloud with color and texture information. The color and texture information are usually captured by cameras, while the depth information are usually obtained using range sensors. In order to build a color 3D model, we need to fuse the color images with depth images effectively. There are mainly two methods to achieve this goal. One is to use RGB-D device such as Kinect or PrimeSense sensor. The RGB-D device is able to register the color images with the depth images. Then we can apply ICP algorithm to automatically align individual, consecutive local 3D point cloud to generate a complete 3D model. This approach is fast and easy to implement. The drawback for this approach is that the RGB-D sensor can only provide depth information up to a very limited range (around 5m). And its depth estimates are pretty noisy compare to LiDAR. Therefore, we use it to model a small indoor room area by placing it in the middle of the room and rotating around itself to scan the entire room. An example of a room model is shown in Figure 2.

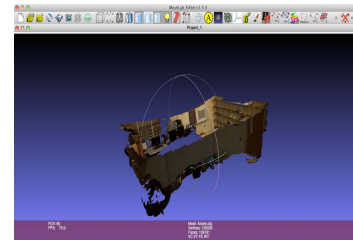


Figure 2: 3D model of a small room

In order to model a large indoor area such as corridor, we choose to fuse the data from camera and LiDAR. Figure 3 shows the assembly of our LiDAR-camera data acquisition system. It includes a 2D line-scan LiDAR, a servo, and a 2D digital camera. The LiDAR is installed on the servo and the camera is rigidly mounted on top of the LiDAR. The whole system is mounted on a push cart for stop-and-go scanning. [9] explains the detail of this scanning system. A brief 3D modeling procedure is given in Figure 4. We first project the 3D point cloud to 2D space based on pinhole camera model. Then we colorize the 2D image with LiDAR intensity value to form 2D LiDAR intensity image. Next, a fiducial target (checkerboard) based feature extraction is performed for both RGB camera image and LiDAR intensity image. RANSAC algorithm is applied here to refine the geometric constraints detection process. By making use of the geometric constraints from the checkerboard pattern, we can find the extrinsic transformation matrices between the camera and the LiDAR. After the extrinsic calibration, we use the transformation matrices to register the color images with the LiDAR intensity images. Finally, the 2D points with color information are back project to 3D space to generate color 3D point cloud. By manually aligning the 3D point clouds in different survey points, we build a 3D model of a large corridor area, as shown in Figure 5. At the same time, the 3D model is partitioned based on the survey point locations.

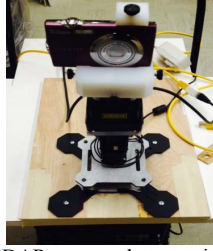


Figure 3: LiDAR-camera data acquisition system

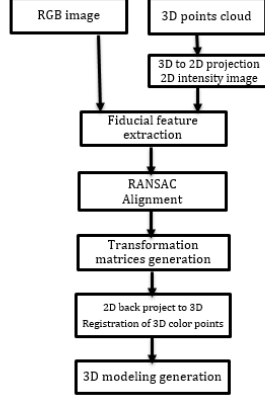


Figure 4: Flow chart for 3D modeling procedure.

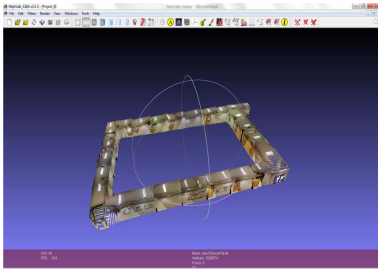


Figure 5: 3D model of a large corridor area

Besides building 3D model of the indoor environment, we also record WiFi RSS data in each survey point to train a Gaussian process model. The detailed procedure of Gaussian process regression can be found in [2] [10].

B. Online Localization Phase

After all the preparations are done in the offline training phase, we are ready for online localization. During the initialization step, we scan the surrounding WiFi RSS values and corresponding BSSIDs. We compare them with the offline built GP model to coarsely determine the user location. This initial estimation tells which partition the user locates in the 3D model. We then perform SURF [11] feature matching between online captured image with key frame images of the 3D sub model. The result is sorted based on the matching distance in feature descriptor space. The nearest neighbor is selected as the closest key frame image to the captured image. Since we have the 3D coordinates of the key frame image's feature points (given in its registered depth image), and the 2D coordinates of the captured image's corresponding feature points (specified in the image), these n feature points form a

classic Perspective n Point (PnP) problem. By solving the PnP problem [12], we get the transformation matrix between the captured image and the key frame image. Therefore, we can calculate the initial 6-DoF pose of the user's mobile device $Pose_{t_0}$ by simply multiply the global pose of the key frame image $Pose_{keyframe}$ with the transformation matrix $TForm_{t_0}$.

Once the initial pose of the device is known, we perform pose estimation using the portable 3D visual sensor on the device. The procedure runs as follows: At a given time step $t > 0$, we have an online captured 3D point cloud $OnlinePCloud_t$. Meanwhile, a local point cloud $LocalPCloud_{t-1}$ is generated from the previous pose $Pose_{t-1}$, computed at time $t-1$ using the prebuild 3D model. Both these point clouds $OnlinePCloud_t$ and $LocalPCloud_{t-1}$ have their own color image ($OnlineColorI_t$, $LocalColorI_{t-1}$) and depth image ($OnlineDepthI_t$, $LocalDepthI_{t-1}$). We perform SURF feature matching between $OnlineColorI_t$ and $LocalColorI_{t-1}$. The matched feature points' 3D locations are acquired from the corresponding depth images ($OnlineDepthI_t$, $LocalDepthI_{t-1}$). Thus, the pose estimation problem has become finding the rigid transform between two sets of 3D points, which can easily be solved with ICP algorithm. Figure 6 illustrates the pipeline of this process. Additionally, we employ RANSAC algorithm to improve the feature matching process. An example is given in Figure 7.

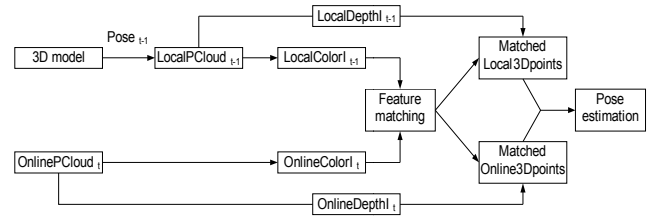


Figure 6: 6DoF pose estimation pipeline

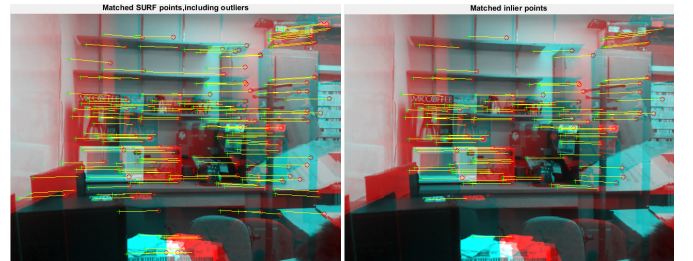


Figure 7: SURF feature matching with and without RANSAC

III. IMPLEMENTATION ON IOS PLATFORM & EXPERIMENTAL ANALYSIS

We realize the proposed indoor localization system on iOS platform and build an app to test the system performance. Thanks to Occipital Inc, they have developed a 3D structure



Figure 8: Portable 3D sensor mounted on iPad

sensor [13] which can be mounted on iPad (shown in Figure 8) to capture depth image and register it with color image captured by the original camera on the iPad, and generate 3D point cloud of the environment using their provided Application Program Interface (API). The test bed of our experiment includes a large corridor area and a small room area. Figure 9 shows a snapshot of our indoor localization app.

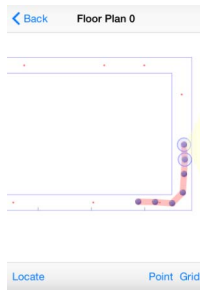


Figure 9: Snapshot of indoor localization app in a corridor area

We compare pairwise error between the ground truth pose of the mobile device and its estimated pose in the 3D model. The error metric is based on [14]. It computes the relative pose between the reference pose and the estimated pose:

$$\varepsilon(x_{1:T}) = \sum_{t=1}^T \|x_t \ominus x_t^*\|^2 \quad (1)$$

where $x_{1:T}$ is the estimated poses from time 1 to T , $x_{1:T}^*$ is the ground truth poses, and $\ominus$ stands for the inverse motion composition operator.

The error $\mathcal{E}$ includes a rotation matrix component $\mathcal{E}_{rot}$ and a translation vector component $\mathcal{E}_t$. We define the total pose estimation error e_t as the size of the translation error:

$$e_t = \|\mathcal{E}_t\| \quad (2)$$

and the total angular error as the principal angle of the rotation matrix error:

$$e_{rot} = |\cos^{-1}(0.5\text{tr}(\mathcal{E}_{rot}) - 1)| \quad (3)$$

We measure the error in 30 waypoints, the average error is around 10cm (translational) and 8 degree (rotational). Meanwhile, we are aware that this metric is not sufficient to determine the accuracy of the app under all circumstances. A

delicate motion-capture system should be used to track the iPad camera during the quantitative error analysis.

IV. CONCLUSION & FUTURE WORK

In this paper, we have presented an indoor pose localization system on mobile platform using portable 3D visual sensor. We first reduce the time consuming feature matching process by using WiFi RSS model to narrow down the search space from the whole 3D model to a small partition. Since we have a portable 3D sensor instead of just a 2D camera, we can bypass the 3D/2D or 2D/3D projection process and directly compare the online captured 3D point cloud with the local 3D model to find the pose estimation. As far as we know, this is the first attempt to use a portable 3D visual sensor to localize the mobile device in a 3D point cloud. The experiment carried out in the indoor environment encourages us to keep on working on this system.

In the future, we are looking to fully implement the system on mobile platform and test it in real time. And we will try to combine the system with more features like assistive technology to provide various location-based services to people with special needs.

REFERENCES

- [1] G. Jin, X. Lu, MS. Park. "An indoor localization mechanism using active RFID tag" in *Proceedings of International Conference on Sensor Networks, Ubiquitous and Trustworthy Computing*, 2006
- [2] X. He, S. Badiei, D. Aloï, J. Li. "WiFi iLocate: WiFi based indoor localization for smartphone" in *Proceedings of Wireless Telecommunication Symposium*, August 2014.
- [3] Fan Li, Chunshui Zhao, Guanzhong Ding, Jian Gong, Chenxing Liu, Feng Zhao. "A reliable and accurate indoor localization method using phone inertial sensors" in *Proceedings of ACM Conference on Ubiquitous Computing*, 2012
- [4] J. Liang, N. Corso, E. Turner, A. Zakhori. "Image based localization in indoor environments" in *Proceedings of Computing for Geospatial Research and Application*, 2013
- [5] C. Jaramillo, I. Dryanovski, R. Valenti, J. Xiao. "6-DoF Pose Localization in 3D Point-Cloud Dense Maps Using a Monocular Camera" in *Proceedings of IEEE International Conference on Robotics and Biomimetics*, 2013
- [6] A. Ruiz-Ruiz, P. Lopez-de-Teruel, O. Canovas. "A multisensor LBS using SIFT-based 3D models" in *Proceedings of International Conference on Indoor Positioning and Indoor Navigation*, 2012
- [7] Besl, Paul J.; N.D. McKay. "A Method for Registration of 3-D Shapes". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1990
- [8] Martin A. Fischler, Robert C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography" in *Comm. of the ACM* 24 (6): 381-395
- [9] J. Li, X. He, J. Li. "2D LiDAR and Camera Fusion in 3D Modeling of Indoor Environment" in *Proceedings of IEEE National Aerospace & Electronics Conference*, 2015
- [10] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006
- [11] H. Bay, A. Ess, T. Tuytelaars, and L. Vangool. "Speeded-Up Robust Features (SURF)" in *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, Jun. 2008.
- [12] V. Lepetit, F. Moreno-Noguer and P. Fua. "EPnP: An Accurate O(n) Solution to the PnP Problem" in *International Journal Of Computer Vision*, vol. 81, p. 155-166, 2009.
- [13] www.structure.io
- [14] F. Endres, J. Hess, D. Cremers, and N. Engelhard. "An Evaluation of the RGB-D SLAM System," in *Proceedings of International Conference on Robotics and Automation*, vol. 1, no. c. Ieee, May 2012, pp. 1691–1696.