W-RGB-D: Floor-Plan-Based Indoor Global Localization Using a Depth Camera and WiFi

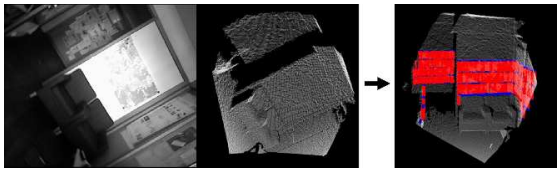
Seigo Ito Felix Endres Markus Kuderer Gian Diego Tipaldi Cyrill Stachniss Wolfram Burgard

***Abstract*— Localization approaches typically rely on an already available map to identify the position of the sensor in the environment. Such maps are usually built beforehand and often require the user to record data from the same sensor used for localization. In this paper, we relax this assumption and present a localization approach based on architectural floor plans. In general, floor plans are readily available for most man-made buildings but only represent basic architectural structures. The incomplete knowledge leads to ambiguous pose estimates. To solve this problem, we present W-RGB-D, a new method for indoor global localization based on WiFi and an RGB-D camera. We introduce a sensor model for RGB-D cameras that is suitable to be used with abstract floor plans. To resolve ambiguities during global localization, we estimate a coarse initial distribution about the sensor position using the WiFi signal strength. We evaluate our W-RGB-D localization method in indoor environments and compare its performance with RGB-D-based Monte Carlo localization. Our results demonstrate that the use of WiFi information as proposed with our approach improves the localization in terms of convergence speed and quality of the solution.**

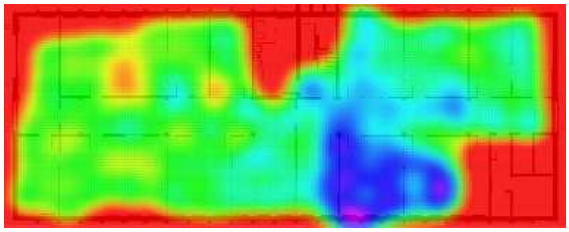
I. Introduction

Localization in indoor and thus GPS-denied environments with low-cost sensors is still an open problem. At the same time, accurate knowledge about the position of a user offers the opportunity for a variety of novel services in airports, shopping malls, conference centers or other environments where GPS is not available. One popular robust approach to localization is Monte Carlo localization (MCL), which employs a particle filter to estimate the position of the sensor [1], [2]. While this method provides a high accuracy, it generally requires a precise and pre-built map of the environment.

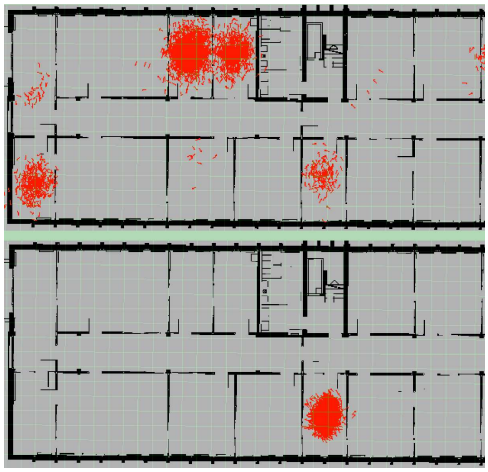
The alternative is wireless-based indoor global localization. Common technologies for wireless communication include global systems for mobile communication (GSM), ultra wideband (UWB), worldwide interoperability for microwave access (WiMAX), radio frequency identification (RFID), Bluetooth, or WiFi [3], [4], [5]. Among them, WiFi is one of the most useful signal for global localization. WiFi receivers are present in almost all devices and one can readily exploit the access points already present in the environment. A further advantage is that WiFi-based localization does not depend on an accurate model of the environment and is resilient to ambiguities due to the unique IDs provided by the access points. However, the accuracy of WiFi-based localization methods is typically lower than that of laser-based or camera-based MCL approaches.



(a) Input 1: Proposed sensor model for a floor plan: (left) grayscale image; (middle) extracted point cloud; (right) aligned point cloud and virtual measurements.



(b) Input 2: WiFi signal strength and WiFi signal maps expressed by a Gaussian Process. This figure shows the mean signal map of one access point.



(c) Output: Comparison between MCL (top) and W-RGB-D (bottom)

Fig. 1. Examples of inputs (top) and outputs (bottom) of the proposed algorithm.

In this paper, we address the localization problem when no previously built map of the environment is available. Our approach solely relies on already available information such as the floor plan and WiFi signal strength information. Floor plans typically are only a sketch of the actual geometric structure of the environment and typically do not

Seigo Ito is with the Information & Communication Research Division, TOYOTA Central R&D Labs., Inc., Japan. seigo@mosk.tytlabs.co.jp

The other authors are with the Department of Computer Science, University of Freiburg, Germany.

contain objects like furniture. This reduces the ability to uniquely localize the sensor, as can be seen in Fig. 1(c). In such situations, MCL requires to explore a large part of the environment to finally converge, a phenomenon that is inconvenient for applications that require to find the sensor location quickly. To resolve this problem, we introduce W-RGB-D, a novel hybrid approach based on WiFi and a RGB-D data for global localization in indoor settings. Our approach employs at its core the established Monte Carlo localization approach and consists of two main steps. First, we estimate a coarse global position using WiFi. We employ a two-dimensional WiFi model to quickly reduce the number of potential ambiguities (see Fig. 1). Second, we compute a precise estimate of the camera location based on the environment floor plan and the data acquired from the RGB-D camera. Our approach, in this way, compensates for the weakness of both sensor modalities, enabling a quick convergence combined with a good localization accuracy.

II. Related Work

A widely-used approach to indoor global localization is Monte Carlo localization [1], [2]. This algorithm employs a particle filter to estimate a distribution over the potential states of the sensor. In environments with large ambiguities, it is well-known that the particle filter shows a slow convergence and requires to move the sensor until the ambiguities have been resolved.

Another successful approach for indoor global localization is the multi-hypotheses tracking (MHT) [6], [7]. The MHT generates data association hypotheses between features of the environment (e.g., doors, and walls) and measurements. Feature-based multi-hypothesis localization [7] uses geometric constraints to prune the hypothesis tree in global localization settings. This method also shows slow convergence in ambiguous environments.

Alternative approaches employ global scan matching for indoor global localization [8], [9]. The method proposed by Tomono [8] uses geometric hashing to match an input scan with a reference scan without requiring any initial guess. The spectral scan matching (SSM) technique [9] employs the spectral matching technique [10] to obtain a coarse localization and RANSAC for a more precise alignment. Some researchers focused on image based localization approaches. The method proposed by Se *et al.* [11] represents the environment as a set of local invariant features. The authors perform global localization by matching features from the current image to the feature map. Related to that, Bennewitz *et al.* [12] propose vision-based MCL avoiding explicit data associations for robot localization of a humanoid. One of the renowned techniques in image-based localization is Fast Appearance-Based Mapping (FABMAP) [13], which employs an image retrieval system based on the bag of words approach.

WiFi signals have also been used for global localization [3], [4], [5]. The RADAR system [3] builds a detailed "radio fingerprint" of the available 802.11 access points and estimates the position of a WiFi device using the received signal strength. The PlaceLab project [4] presents a wide-area WiFi based localization system in the greater Seattle area with nearly 100% coverage. The WiFi-SLAM approach [5] address the SLAM problem with WiFi signal by employing the Gaussian Process latent variable model. These methods are resilient to environment ambiguity, however, at the cost of an accuracy of few meters at most [3].

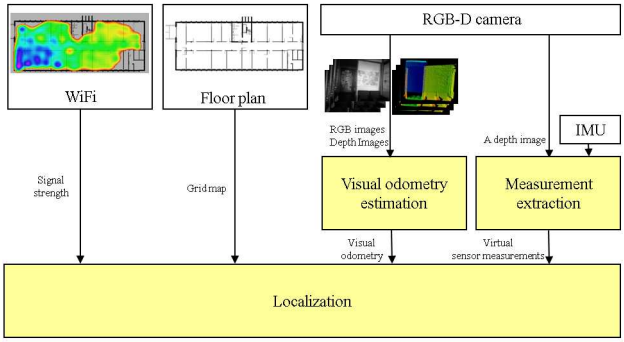


Fig. 2. Overview of the proposed approach, consisting of three main components: measurement extraction, visual odometry estimation, and localization.

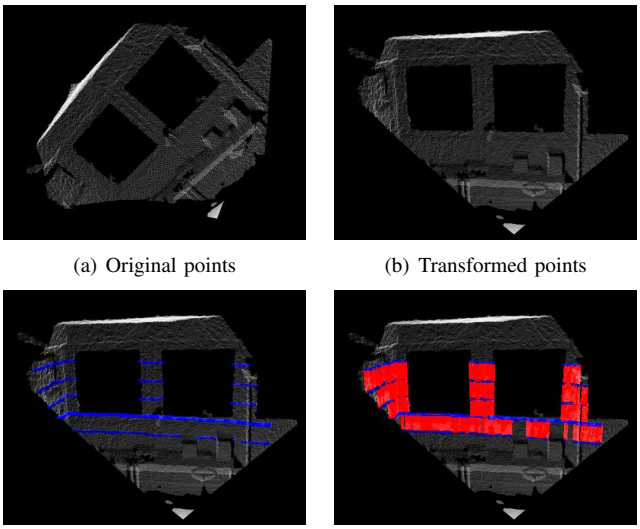
Luo *et al.* [14] also use the information of an abstract floor plan for mobile robot navigation. They extract doorplates and passage corners as landmarks to globally localize the robot. In contrast to their method, W-RGB-D localizes the robot using WiFi maps of the environment.

III. The W-RGB-D Approach

In this section we describe the main components of our proposed W-RGB-D system, which are the measurement extraction, the visual odometry estimation, and the localization. Fig. 2 depicts the major components of our system and their dependencies. The measurement extraction process described in Section III-A is responsible to extract virtual measurements from the depth data. These virtual measurements are used in a dedicated sensor model that computes the likelihood of the measurement given the floor plan at hand. The visual odometry approach presented in Section III-B is responsible for the estimation of the egomotion of the camera from image and depth data. We adopt a RANSAC-based approach to robustly estimate the motion and reject spurious association between consecutive images. Finally, we combine these components with the WiFi signal strength information within a particle filter to robustly deal with ambiguities and to achieve a localization approach that can quickly estimate the position (see Section III-C).

A. Measurement Extraction

Floor plans typically represent only the basic structure of buildings including walls, doorways, and windows. Furthermore, they are only a 2D representation of the 3D environment. To be able to correctly derive a sensor model for this kind of maps, we need to extract virtual measurements from the 3D point cloud obtained with the RGB-D sensor. In this work, we mainly focus on wall-like features and rely



(a) Original points (b) Transformed points

(c) Line extraction (d) Check of connections

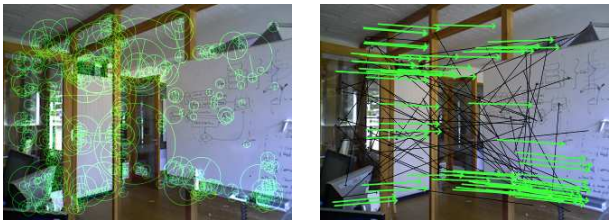
Fig. 3. Example of measurement extraction. Original point cloud in the camera frame (a). Transformed point cloud using IMU data (b). Blue points correspond to extracted virtual scans (c). Red lines indicate accepted connections (d).

on plane extraction techniques. A widely-used method to extract wall-like features from a point cloud is to employ RANSAC for plane extraction [15]. However, the RANSAC-based approach has high computational cost when the size of the point cloud is large. To reduce computation time, one can use a voxel grid filter before the plane extraction to reduce the number of points to be evaluated. This approach has the disadvantage that all points within a voxel are approximated with their centroid, reducing the accuracy of the plane extraction.

In this paper we propose a different approach depicted in Fig. 3 with low computational cost. We first correct the orientation of the raw point cloud using an inertial measurement unit (IMU) 3(b). Then, we extract horizontal scans at different heights as potential virtual measurements based on the point cloud. For each point of a horizontal scan, we connect it to the scans above and below if the vertical gradient is below a certain threshold. We accept points connected with at least one other level in the virtual measurement and project them down onto the 2D plane of the floor plan. We call this approach *sensor measurement extraction* (SME).

B. Visual Odometry

For visual odometry, we use the technique developed by Endres *et al.* [16]. Their approach employs the visual image of an RGB-D camera to detect keypoints and extract their respective descriptors. It matches the keypoints extracted from the current image with the ones of previously recorded images according to their descriptors. Finally, they employ RANSAC to reject false correspondences and compute the camera transformation between the two frames in closed form [17]. Combined with the depth information, this allows us to register the point clouds in a common coordinate



(a) Extracted SIFT keypoint (b) Estimated visual odometry

Fig. 4. Example of keypoint extraction and visual odometry estimation. Circles show the extracted keypoints (a). Thick green arrows indicate the estimated transformations of the inliers, and thin black arrows correspond to the transformations of the outliers (b).

system. Fig. 4 shows an example of the keypoint extraction and visual odometry estimation.

C. Localization

To perform global localization, we employ the Monte Carlo localization algorithm (MCL) [2] together with an initialization strategy based on the WiFi signal. The MCL recursively estimates the posterior of the robot's pose as follows:

$$p(x_t \mid z_{1:t}, u_{0:t-1}) \propto$$
$$p(z_t \mid x_t) \int_{x'} p(x_t \mid x', u_{t-1}) p(x' \mid z_{1:t-1}, u_{0:t-2}) dx',$$

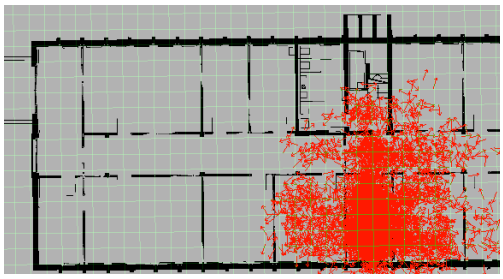
where $u_{0:t-1}$ is the sequence of motion command executed by the robot. The motion model $p(x_t \mid x', u_{t-1})$ denotes the probability of the robot's state x_t given it executes u_{t-1} in the previous state x'. In our case, no real odometry is available and we rely on the result of the visual odometry module for the motion update step. The observation model $p(z_t \mid x_t)$ denotes the likelihood of the observation z_t given the pose x_t. The MCL approximates the belief of the robot with a set of samples called particles and updates it iteratively by sampling those particles from the motion model and compute a weight according to the observation model. Particles are then resampled according to this weight and the process iterates.

A widely-used approach for the initialization of particles is to randomly generate initial particles uniformly throughout the environment (see Fig. 5(a)). However, the number of particles depends on the size of the environment and many particles are required in large environments. Furthermore, having a uniform prior in ambiguous environments may result in slow or even incorrect convergence of the filter.

To speed up the initial global localization phase and to resolve ambiguities in global localization, we propose an initialization based on WiFi (see Fig. 5(b)). To realize this and to cope with the underlying uncertainties, we rely on Gaussian Processes to represent a signal strength map.

Instead of uniformly sampling from the entire state space, we bias the sampling process to regions of the environment with consistent signal strength. To do so, we discretize the environment and compute a weight w for each discretized

(a) Particle filter initialization using a uniform distribution.



(b) Particle filter initialization using the WiFi signal.

Fig. 5. Two initialization methods in the particle filter. The initialization method based on uniform distribution puts particles over the whole state space (a), while WiFi based initialization quickly resolves the ambiguity (b).

position x as follows

$$w = \prod_{i=1}^{n} \overline{w_i}, \tag{1}$$

where n is the number of currently observed access points. The weight $\overline{w_i}$ corresponding to a received signal strength s_i of access point i is given by

$$\overline{w_i} = P(s_i; \mu_{gp}(x, D_i), \Sigma_{gp}(x, D_i)), \tag{2}$$

where

$$P(s_i; \mu_{gp}, \Sigma_{gp}) = \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma_{gp}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(s_i - \mu_{gp})\Sigma_{gp}{}^{-1}(s_i - \mu_{gp})\right), \tag{3}$$

is the probability density function of the received signal strength s_i, expressed in terms of the mean $\mu_{gp}(x, D_i)$ and covariance $\Sigma_{gp}(x, D_i)$ of the Gaussian Process [18]. Fig. 6 shows an example of a mean signal map. The term D_i is the training data of WiFi signals and k is the number of dimension. We sample the particles by first sampling the discretized location according to the weight w and then by uniformly sampling in the discretized cell.

In the current implementation of our approach, we use the signal strength only for particle initialization and for checking whether the robot has been kidnapped and whenever the robot stands still. This is because the signal strength measurements typically have a varying time delay, most likely caused by the WiFi driver. This makes the integration of the signal strength too inaccurate when the robot moves quickly.

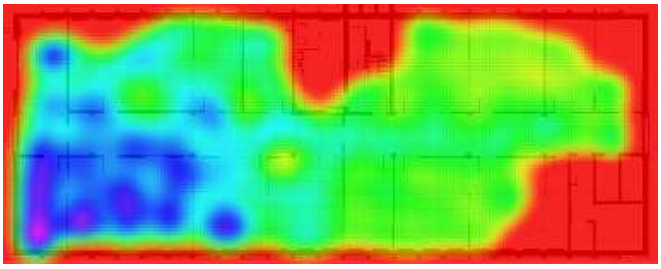


Fig. 6. A mean signal map of an access point expressed by the Gaussian Process. The color gradient depicts the strength of the WiFi signal.

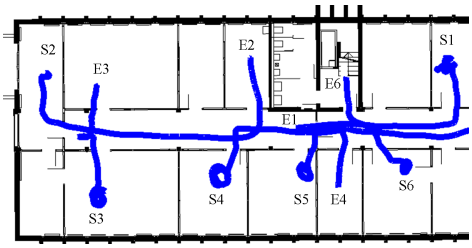


Fig. 7. Floor plan of the environment. The blue solid lines show the trajectories taken in the experiments. The points S indicate the starting positions and the points E the final positions of each trial.

IV. Experiments

We evaluated the approach presented in this paper in a set of real-world experiments. The experimental evaluation is designed to demonstrate that (a) information about WiFi signal strength can boost the performance of global localization in ambiguous maps, that (b) the measurement extraction algorithm is more efficient than RANSAC and is suitable for real-time localization, and (c) our approach does not require a map built with an RGB-D SLAM system and rather allows us to localize an RGB-D camera given only an architectural floor plan of the environment.

We perform the evaluation on three sets of experiments. In the first set, we evaluate the benefits of a WiFi bootstrap in terms of convergence speed. To do so, we compare the convergence of the particle filter starting with a uniform distribution with and without the WiFi initialization proposed in this paper. The second set of experiments evaluates the runtime of our measurement extraction algorithm and compares it to an alternative approach based on the point cloud library (PCL). In the third set of experiments, we evaluate the accuracy of the proposed method with respect to the ground truth, which we obtained by employing MCL and a laser range finder. Fig. 7 shows the floor plan of the experimental environment and the ground-truth trajectories.

The system used to carry out the experiments consists of an ASUS Xtion Pro Live and an iPhone 4S with a gyroscope and an accelerometer. For our implementation, we modified the adaptive Monte Carlo localization algorithm in ROS [19].

A. Global Localization with WiFi Maps

In this experiment, we investigate the advantage of the proposed global localization with WiFi initialization, "W-RGB-D", which is the initialization strategy described in

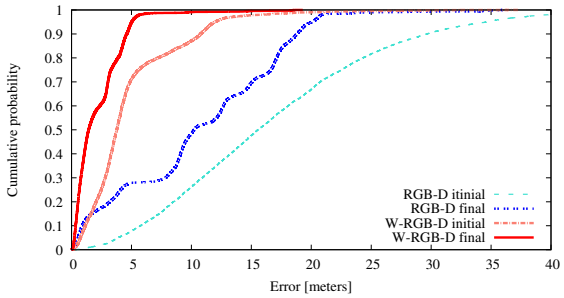


Fig. 8. Cumulative histogram of the translational error of the particles in localization for the different approaches. The initialization with WiFi information prevents particles to become stuck in local minima.

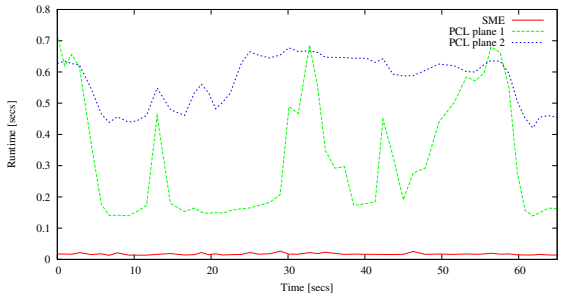


Fig. 9. Runtime of the sensor model extraction for SME and PCL plane.

Section III-C. To this end, we compare it to a particle filter without WiFi initialization, "RGB-D". In both cases, we initialize the filter by drawing particles from a uniform probability density over the three-dimensional state space (position and orientation) of the camera in the environment.

We performed the experiments at known positions in 10 different rooms. To evaluate the convergence speed, we first evaluated the error just after the initialization. Then, we turned our hand-held device three times on the spot, by 360 degrees for each turn, and evaluated the localization error of the final estimate.

Fig. 8 shows a cumulative histogram of the error in localization using the two different approaches in all experiments. For each distance d, it depicts the ratio of particles with an translational error below d. For "RGB-D", the cyan dashed line shows this cumulative distribution for the initialization and the blue double-dashed line the final localization. The pink chain line and the red solid line show the respective distributions for initialization and final localization of our W-RGB-D approach. The figure clearly shows the effect of the ambiguity of the map. Without the wireless information, simply rotating on the spot is not enough to disambiguate the camera location. This is evident when comparing Fig. 5 and Fig. 1. The camera is located in the middle room, which has about the same dimension as five other rooms. After turning, the system was correctly able to infer that the camera is located in the middle of a room of that size. In case of WiFi initialization, we were able to infer exactly in which room it was located, thanks to the prior information of the WiFi signal.

The differences in convergence revealed by this experiment show an important issue for global localization. If the system cannot uniquely localize the camera, the camera has to be moved through rooms and corridors until the system can localize its position uniquely, which would be time-consuming and tedious.

B. Runtime Evaluation

This experiment is designed to analyze the computational demands of the measurement extraction algorithm proposed in this paper. We compare our measurement extraction method to the plane extraction approach implemented in point cloud library [15] as this is an alternative way to compute a virtual measurement for our problem. We use the TUM RGB-D dataset [20] for the runtime evaluation.

The PCL plane extraction method uses a downsampled cloud of the points using a voxel grid filter with leaves of 3 cm side length. It then extracts several planes using RANSAC until 70% of points are covered by one of the planes. Finally, it projects the points belonging to the extracted planes from the 3D space to the 2D space. We call this approach PCL plane 1. In addition to this, we use the organized multi plane segmentation method in the point cloud library. We call this approach PCL plane 2.

Fig. 9 shows the runtime of SME (see Section III-A) and PCL plane 1 and 2. The runtime of PCL plane varies largely, while our approach is faster and shows less variation in execution time.

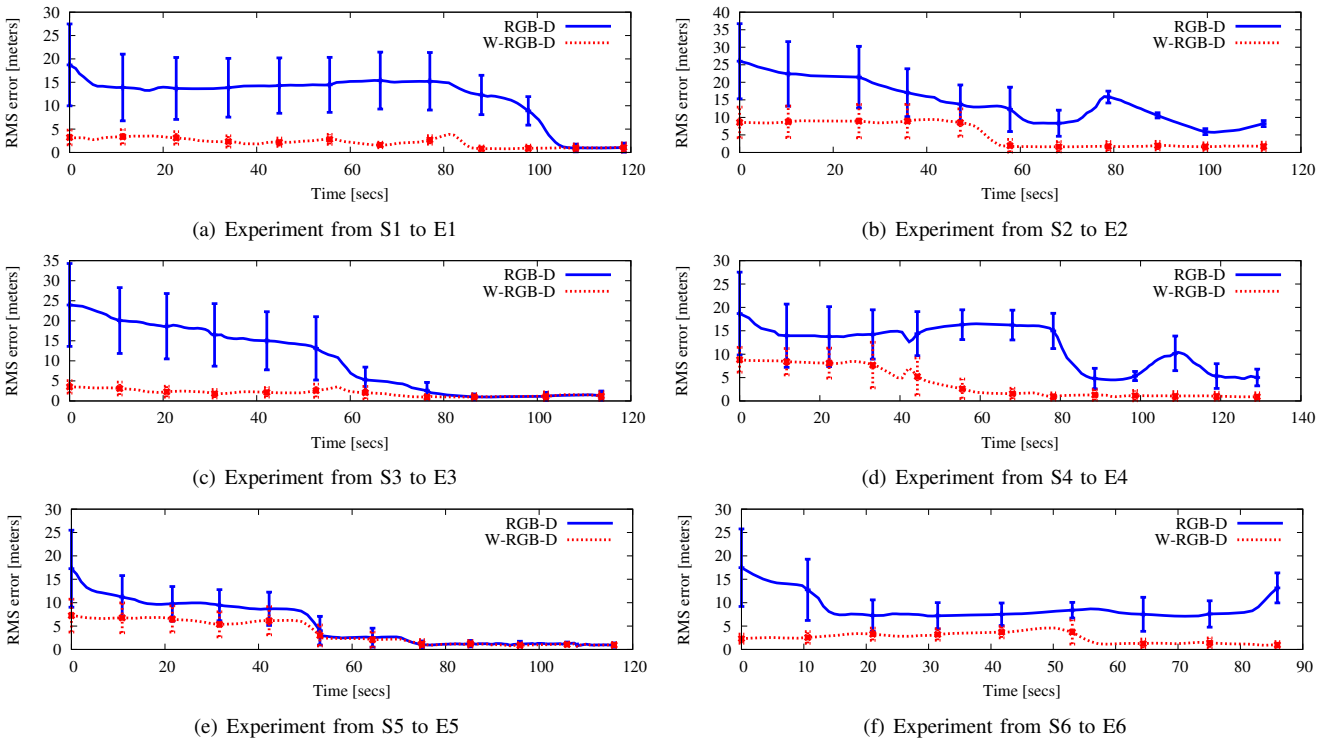
C. Localization Accuracy

To analyze the accuracy of our method, we use a robotic wheelchair [21] equipped with a 2D laser range finder. We mounted the hand-held system on the wheelchair and manually measured its displacement with respect to the laser range finder. As ground-truth, we considered the results of Monte Carlo localization using the 2D laser scanner mounted on the wheelchair. Fig. 7 indicates the ground-truth trajectories, where the points denoted with S refer to the start positions and the points denoted with E to the end positions of each trajectory taken by the wheelchair.

Fig. 10 depicts the results of our experiments for both localization strategies. The solid line in each graph shows the results of RGB-D and the dashed line shows the results of W-RGB-D. Both lines depict the root mean square (RMS) error and the bars show the standard deviation to illustrate the distribution of the localization error. The figure shows three important aspects. First, localization based on RGB-D sensors using floor plans is possible and also accurate. Second, even without WiFi information, when the approach converges to the correct solution, the localization error is close to that of WiFi localization. Third, the WiFi information in fact increases convergence speed and convergence ratio.

V. Conclusions and Future Work

In this paper, we presented W-RGB-D, a novel approach for indoor global localization. The main contributions of this paper are the descriptions and the analysis of an integrated system based on RGB-D localization and WiFi localization



(a) Experiment from S1 to E1

(b) Experiment from S2 to E2

(c) Experiment from S3 to E3

(d) Experiment from S4 to E4

(e) Experiment from S5 to E5

(f) Experiment from S6 to E6

Fig. 10. Comparison of the convergence speed and convergence ratio of RGB-D and W-RGB-D in the situations depicted in Fig. 7.

for quick convergence of particles in an ambiguous environment. Our approach is most useful during global localization especially when a robot or the user carrying the sensor platform is not supposed to move much. Given that small RGB-D cameras for tablet computers or smartphones will soon be available to the public at a low price, W-RGB-D will allow people to quickly and accurately localize themselves in indoor environments. Possible extensions of this work include the incorporation of an image-based localization approach. For example, a bag of words approach can estimate new proposal distributions for the particle filter and resolve localization ambiguity similarly as a WiFi-based method.

REFERENCES

[1] F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte Carlo localization for mobile robots, *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 1999.

[2] S. Thrun, W. Burgard, and D. Fox. Probabilistic Robotics, *The MIT Press*, 2005.

[3] P. Bahl, and V.N. Padmanabhan. RADAR: An In-Building RF-based User Location and Tracking System, *Proc. of the IEEE Infocom*, pp.775-784, 2000.

[4] A. Lamarca, Y. Chawathe, S. Consolvo, et al. Place Lab: Device Positioning Using Radio Beacons in the Wild, *Proc. of the Third Int. Conf. on Pervasive Computing*, 2004.

[5] B. Ferris, D. Fox, N. Lawrence. WiFi-SLAM Using Gaussian Process Latent Variable Models, *Proc. of the 20th International Joint International Joint Conference on. Artificial Intelligence*, 2007.

[6] P. Jensfelt, and S. Kristensen. Active Global Localization for a Mobile Robot Using Multiple Hypothesis Tracking, *IEEE Transactions on Robotics and Automation*, Vo.17, No.5, 2001.

[7] K. Arras, J. Castellanos, M. Schilt, and R. Siegwart. Feature-based multi-hypothesis localization and tracking using geometric constraints, *Robotics and Autonomous Systems*, vol. 44, pp. 41-53, 2003.

[8] M. Tomono. A scan matching method using Euclidean invariant signature for global localization and map building, *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pp.866-871, 2004.

[9] S. Park, and S. K. Park. Spectral Scan Matching and Its Application to Global Localization for Mobile Robots, *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2010.

[10] M. Leordeau and M. Hebert. A Spectral Technique for Correspondence Problems using Pairwise Constraints, *Proc. of the Int. Conf. on Computer Vision*, pp. 1482-1489, 2005.

[11] S. Se, D. G. Lowe, and J. J. Little. Vision-Based Global Localization and Mapping for Mobile Robots, *IEEE Transactions on robotics*, vol. 21, No.3, pp. 364-375, 2005

[12] M. Bennewitz, C. Stachniss, W. Burgard, and S. Behnke. Metric localization with scale-invariant visual features using a single perspective camera. *European Robotics Symposium*, pages 143–157, 2006.

[13] M. Cummins and P. Newman. Highly scalable appearance-only SLAM - FAB-MAP 2.0, *Robotics: Science and Systems*, 2009.

[14] R. C. Luo, Y. Lin, and C. Kao. Autonomous Mobile Robot Navigations and Localization Based on Floor Plan Map Information and Sensory Fusion Approach, *Proc. of the IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, 2010

[15] R. B. Rusu, and S. Cousins. 3D is here: Point Cloud Library (PCL), *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2011.

[16] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard. An Evaluation of the RGB-D SLAM System, *Proc. of the IEEE Int. Conf. on Robotics and Automation*, 2012.

[17] Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (13), 1991.

[18] C. E. Rasmussen, and C. K. I. Williams. Gaussian Process for machine learning, *MIT Press*,2005.

[19] ROS amcl, http://www.ros.org/wiki/amcl

[20] J. Sturm, N. Engelhard, F. Endres, W. Burgard and D. Cremers. A Benchmark for the Evaluation of RGB-D SLAM Systems. *Proc. of the International Conference on Intelligent Robot Systems*, 2012

[21] E. Demeester, E. Vander Poorten, A. Huntemann, et al. Robotic ADaptation to Humans Adapting to Robots: Overview of the FP7 project RADHAR, *International Conference on Systems and Computer Science*, 2012