2014 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS 2014)
September 14-18, 2014, Chicago, IL, USA

Robot-Assisted Human Indoor Localization Using the Kinect Sensor and Smartphones

Chao Jiang[1], Muhammad Fahad[1], Yi Guo[1], Jie Yang[2], and Yingying Chen[1]

Abstract— **Human indoor localization was previously implemented using wireless sensor networks at the cost of sensing infrastructure deployment. Motivated by high density of smartphones in public spaces, we propose to use a robot-assisted localization system in which the low-cost Kinect sensor and smartphone-based acoustic relative ranging are used to localize moving human targets in indoor environments. An extended Kalman filter based localization algorithm is developed for real-time dynamic position estimation. We present both simulations and real robot-smartphone experiments demonstrating the performance with a localization accuracy of approximately 0.5m.**

I. INTRODUCTION

Human indoor localization has recently received increasing research attention due to many real-world applications such as location detection of medical personnel or firemen, pattern of passenger flow in airports or shopping malls. Motivated by the fact that smartphones are gradually woven into people's social life and usually a high density of them exist in public spaces, we propose an indoor localization system using smartphones and a mobile robot with low cost sensors such as the Kinect sensor. We develop new dynamic Kalman filter based indoor human localization algorithms and validate it in real robot-smartphone experiments.

A. Related Work

The astonishing development of wireless network such as radio frequency identification (RFID), WLAN and ultra-wideband has remarkably facilitated human indoor localization techniques using either specialized or minimally modified infrastructure [1]. A RFID based location sensing system was developed in [2]. By scanning the data emitted from active RFID tags, a scene analysis methods was adopted to compare the signal strength perceived from the target tag and reference tags (deployed as landmarks). The position of the target was estimated by the k-nearest neighbor algorithm with around 1m average accuracy. In [3], an online probabilistic RFID map and adaptive Kalman filter were applied to obtain localization with accuracy from 0.5m to 5m, depending on the received signal strength (RSS) noise level and the number of RSS samples collected. The authors in [4] presented the so-called RADAR localization system utilizing WLAN network. The target position was estimated by searching the signal strength map built in offline modeling phase for the closest match of signal patterns. The position accuracy was around 2m to 3m. The Horus system proposed in [5] used the probabilistic method to estimate target positions. The radio map was represented in the form of signal strength probability histogram for each access point. The average localization accuracy of 0.5m was reported. The COMPASS system [6] considered the orientation of targets in the online positioning phase, where the problem of blocking effects of human body encountered in [4] was mitigated. The average positioning error was approximately 1.65m. The use of smartphones as radio signal strength indications (RSSI) in WLAN system was studied in [7], in which the average error about 2m was reported. Overall, existing wireless indoor localization methods based on WiFi signature maps have non-negligible errors in position estimation, and high localization accuracy is usually obtained at the cost of intensive deployment of sensing infrastructure.

Localization is also a classic topic studied in navigation of autonomous mobile robots. Kalman filter based localization [8], [9], grid-based Markov localization [10] and Monte Carlo localization [11], [12] provide solutions for either local position tracking or global position estimation. A more challenging problem of simultaneous localization and mapping (SLAM) arises when the robot has no prior knowledge of the environment map [13]. Recent attention has been drawn to the cases that only relative range to the landmark can be detected. In [14], the range-only SLAM using extended Kalman filter was investigated where prior knowledge of landmark location is partially known. The authors in [15] represented the implementation of range-only SLAM on underwater vehicles. In [16], the authors proposed an initialization method where the location of landmarks were estimated using a "voting scheme", and the vehicle was driven along an optimized path to perform landmark initialization. Experiments on SLAM of mobile robots in indoor environments were presented in [17], where a wireless sensor network was deployed for either robot-to-beacon or beacon-to-beacon range measurement. The estimation error of the robot and landmark positions was reported less than 0.2m and 0.5m, respectively. Generally speaking, robot localization and SLAM offer a better estimation accuracy, but the developed algorithms rely on sensing data from expensive sensors such as laser range finders that cannot be directly extended to human indoor localization.

The work was partially supported by the National Science Foundation under Grants IIS-1218155 and CNS-1318748.

[1]Chao Jiang, Muhammad Fahad, Yi Guo and Yingying Chen are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, 07301, USA Emails: {cjiang6, mfahad, yi.guo, yingying.chen} @stevens.edu

[2]Jie Yang is with the Department of Computer Science, Florida State University, Tallahassee, FL, 32306, USA Email: jie.yang@cs.fsu.edu

B. Contribution

In this paper, we propose a novel human localization system that uses a mobile robot and smartphones to localize moving persons. The system consists of a self-localized robot tracking human targets using its onboard Kinect sensor, and a smartphone based acoustic ranging subsystem. An extended Kalman filter (EKF) based dynamic positioning algorithm is developed and integrated with the acoustic relative ranging subsystem to provide real time localization of the moving human target. Experimental results show the estimation accuracy reaching 0.5m. The main contribution of the paper is that by taking advantage of both low-cost 3D vision sensor and smartphone-based acoustic relative ranging techniques, we provide an efficient solution for indoor human localization and motion tracking without complex hardware infrastructure.

II. System Configuration

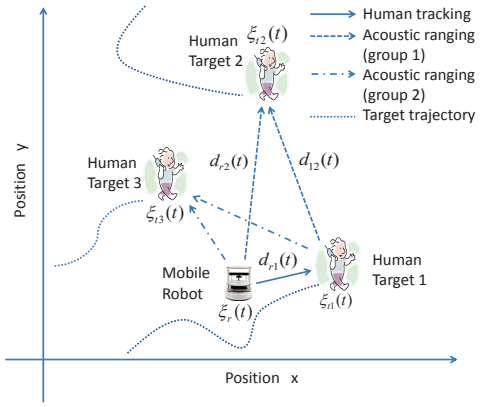


Fig. 1: Overview of robot-assisted indoor human cooperative localization scheme.

Motivated by the popularity of smartphones in public spaces, we propose a cooperative indoor localization system using smartphones and a mobile robot with low cost sensors. As shown in Fig. 1, the localization system consists of an autonomous mobile robot equipped with a Kinect 3D vision sensor and acoustic ranging devices (a microphone and a speaker), and smartphones (with microphones and speakers) with persons to be localized. The self-localized robot is able to simultaneously localize and track human targets by: 1) following the person named "Human Target 1" in the figure, and keeping certain distances from him/her using the Kinect vision sensor, 2) using the location of the robot and Human Target 1 to localize Human Target 2 or Human Target 3 utilizing acoustic ranging measurements, and 3) using the estimated location of the robot and any one of the human targets to localize any additional human targets.

The nature of the proposed localization scheme is that the robot and the Human Target 1 being followed at different time during their motion are treated as a bunch of landmarks, and the positions of other human targets can be estimated relying on acoustic range measurements between the robot and the targets. Conventional triangulation based methods need at least three beacons to uniquely localize a target in a two dimensional space [8]. We fully utilize the *dynamic* nature of the robot tracking system, and treat the position of the robot at different time as a group of beacons. The ranging between the robot, Human Target 1, and Human Target 2 at different time makes it possible to estimate the position of Human Target 2 via triangulation. Without loss of generality, we discuss localization between the robot, Human Targets 1 and 2 in Group 1, while the method is capable of localizing additional human target once the robot and one of the target positions are available. In the next section, we present our human indoor localization scheme.

III. Robot-Assisted Human Indoor Localization

The proposed human indoor localization system is composed of 1) robot self-localization, 2) human following using Kinect vision sensor, 3) acoustic relative ranging, and 4) dynamic target position estimation. A functional block diagram of the proposed localization system is shown in Fig. 2. While techniques on robot self-localization, human-following using the Kinect vision sensor are available, the main challenge of the proposed system lies in the development of an acoustic relative ranging subsystem, and a dynamic position estimation algorithm to localize the target person. Next, we describe each of the components, and then present the overall algorithm in this section.

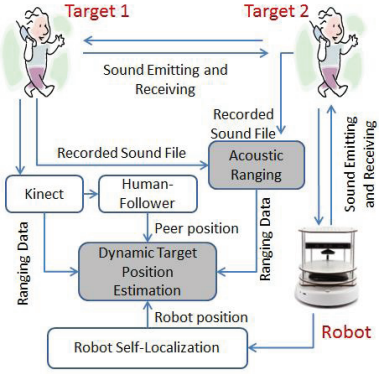


Fig. 2: Functional block diagram of the localization system.

A. Robot Self-Localization

Although localization and mapping in an unknown or partially known environment can be done using existing SLAM techniques, we focus our main attention to *human* localization, and assume known indoor environments with a prior-obtained map, which is a reasonable assumption for many indoor environments (such as shopping malls, museums, airports, or student dorms). We also assume the robot is equipped with onboard sensors and is able to localize itself in the known indoor environment. Robot localization technique such as Monte Carlo localization algorithm [18] can be used, which fuses sensory data from proprioceptive and exteroceptive sensors to estimate the pose of the robot

recursively. The robot self-localization module takes input from sensors (such as odometry, gyroscope and laser range finder), and output the robot's current position $\boldsymbol{\xi}_r(k)$.

B. Kinect-based Human Following

As mentioned earlier, we use the low-cost Kinect vision sensor to track human Target 1. The Kinect sensor detects the centroid of the moving human target and returns the relative range to it. The deviation from the reference ranging and bearing to the human target serves as the control input, which drives the robot to keep the centroid in the middle of the robot vision field and maintain a given distance to the human target. The autonomous human-following program [19] can be used for this functional module, which returns the distance from robot to Target 1, d_{r1}, and the position of Target 1, $\boldsymbol{\xi}_{t1}(k)$.

C. Acoustic Relative Ranging

The acoustic ranging subsystem consists of the robot and smartphones, which have microphones and speakers as onboard acoustic devices. Each of the robot and the target smartphones plays a pre-designed beep file in a pre-determined order, and simultaneously records the received beep files and send the files to the robot for data processing and relative distance measurement. In principle, ranging can be done by time-of-arrival (TOA) method to estimate the sound travel time from one device to another. However, challenges exist in preventing interference from different phones, the lack of clock synchronization, and overcoming uncertainties in emitting and detection. The acoustic ranging subsystem returns the relative distances between the robot and the two targets, $d_{r1}(t)$, $d_{r2}(t)$, and the distance between Targets 1 and 2, $d_{12}(t)$. Since d_{r1} is already available from the Kinect-based human following subsystem, only d_{r2} and d_{12} are returned from the acoustic ranging module. We describe the details of the proposed acoustic ranging method in the following.

1) Acoustic Ranging Signal Design: The acoustic signal, referred to as beep, consists of several evenly spaced, monotonic signal beeps. The frequency, number of beeps and spacing between them in the ranging signal directly impact the accuracy, measurement latency, noise susceptibility and intrusiveness to humans, so these parameters are carefully selected for the design of the beep in our system. As most of background noise (e.g. human conversation 300Hz - 3400Hz, music 50Hz-15kHz) is concentrated in lower frequency band, we use high frequency signal in the 16-20kHz range, which is less susceptible to noise and easier to filter. The ranging accuracy is directly proportional to the number of beeps but adds measurement latency. We used 4 beeps for our experiments with beep interval of 5000 samples. We use 16kHz signal in our experiments since human ears are less sensitive to higher frequency signals, which makes our beep signal less intrusive. Considering the sampling frequency of 44.1 kHz (used in our experiments) and nominal sound velocity of 340 m/s, the minimum resolution for acoustic relative ranging is 0.77 cm.

2) Acoustic Signal Detection Methods: Traditional correlation method discussed in [20] has larger measurement errors. We adopt the change point detection method to detect first instance of arrival of the beep signal, which was proposed in two coauthors' (Yang and Chen) previous work [21], [22]. The method requires the beep signal to be uniformly distributed in a narrow frequency band, and ensures better measurement accuracy over traditional correlation based method. Specifically, we first take the Short Time Fourier Transform (STFT) of the acquired signal to filter out the low frequency background noise and extract the signal at the beep frequency band. The first instance of strong deviation from the normal background noise is detected next, which indicates the instance of arrival of the beep signal. Under different low and high noise environments, change point detection method can achieve within 15cm measurement accuracy in 300cm testing distances (see Figure 7 of [22]).

3) Multi-agent Scheduling and Measurement: Our acoustic ranging subsystem extends existing two-agent scheme [20] to multiple agents, *i.e.*, three agents including the robot and Targets 1 and 2. Unlike the fixed-window method used in the recent work [22], where a fixed time window is scheduled to each agent in emitting the beep signal, we propose a new beep-signal scheduling scheme that relies on an active request and acknowledgement method. In our method, the robot sends the command to Targets 1 and 2 to start recording and waits for their acknowledgement. In the next step, the robot sends the command to Targets 1 and 2 to play the beep signal sequentially and waits for their acknowledgement once the target phone completes playing the beep signal file. The multi-agent scheduling scheme is illustrated in Fig. 3. The active beep-signal scheduling and data acquisition cycle results in faster data processing than existing fixed time-window scheduling [22], thus results in shorter ranging measurement and decreased ranging latency, which is important in our real time human localization system.

D. Dynamic Target Position Estimation

A key component of the localization system is the dynamic positioning algorithm running on the robot to determine the target phone position. We propose an extended Kalman filter (EKF) based position estimation algorithm, which fuses the sensing data from acoustic ranging, Kinect range measurement, and robot self-localization, and returns the position estimate of the human Target 2. We formulate the dynamic position problem as follows:

Denote the positions of the robot, Target 1, and Target 2 by $\boldsymbol{\xi}_r(t)$, $\boldsymbol{\xi}_{t1}(t)$, and $\boldsymbol{\xi}_{t2}(t)$, respectively. The distances between the robot and Target 1, the robot and Target 2, Target 1 and Target 2 are represented by $d_{r1}(t)$, $d_{r2}(t)$, and $d_{12}(t)$, respectively, as shown in Fig. 1. Our problem is to *design a dynamic positioning algorithm to determine the position of Target 2, $\boldsymbol{\xi}_{t2}(t)$, using the observed relative ranging information, $d_{r1}(t)$, $d_{r2}(t)$, $d_{12}(t)$, and the robot self-localization information.*

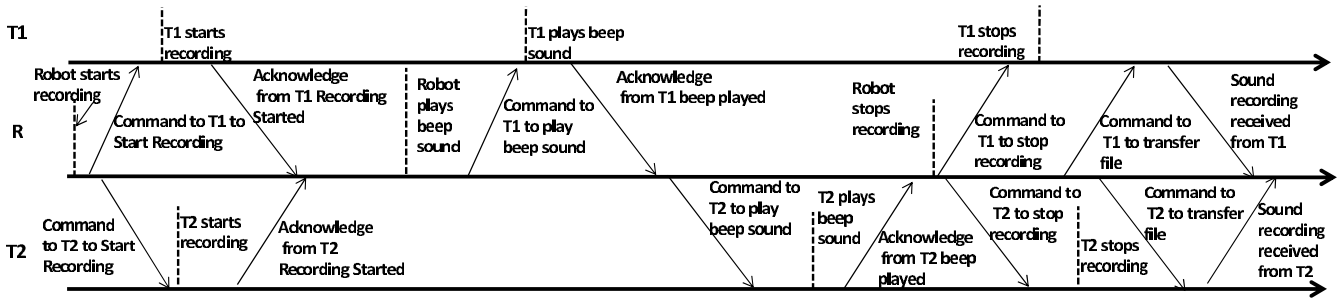


Fig. 3: Multi-agent data acquisition. R: Robot, T1: Target 1, T2: Target 2.

1) Motion Model: The motions of involved human agents are not known a priori. Under the assumption that the sampling rate is relatively high compared to the moving speed of the persons, it is fair to consider that their motions is driven by white noise acceleration process that can be mathematically expressed as $\ddot{\boldsymbol{\xi}} = \boldsymbol{\omega}(t)$. In other words, the velocity of the human agents are Wiener processes, changing over time as white noise $\boldsymbol{\omega}(t) \sim \boldsymbol{N(0, Q)}$ with zero-mean and variance $\boldsymbol{Q}$. As the robot always follows a human agent, the motion of robot is consequently considered as white noise acceleration process consistent with the motion of the human agent. Specifically, the kinematic equations that are used to predict the state from white Gaussian noise $\boldsymbol{\omega}_i(t)$ are expressed as follows:

$$\ddot{\boldsymbol{\xi}}_r = \boldsymbol{\omega}_r(t), \quad \ddot{\boldsymbol{\xi}}_{t1} = \boldsymbol{\omega}_{t1}(t), \quad \ddot{\boldsymbol{\xi}}_{t2} = \boldsymbol{\omega}_{t2}(t) \tag{1}$$

where $\boldsymbol{\xi}_r \in \mathbb{R}^2, \boldsymbol{\xi}_{t1} \in \mathbb{R}^2, \boldsymbol{\xi}_{t2} \in \mathbb{R}^2$ represent the position of the robot, target 1, and Target 2, respectively. The corresponding discrete-time process model is in the form:

$$\begin{aligned} \boldsymbol{\xi}(k+1) &= \boldsymbol{\xi}(k) + \Delta T \boldsymbol{v}(k) \\ \boldsymbol{v}(k+1) &= \boldsymbol{v}(k) + \boldsymbol{\omega}(k) \end{aligned} \tag{2}$$

where $\boldsymbol{\xi}(k) = [\boldsymbol{\xi}_r(k), \boldsymbol{\xi}_{t1}(k), \boldsymbol{\xi}_{t2}(k)]^T$, $\boldsymbol{v}(k) = [\boldsymbol{v}_r(k), \boldsymbol{v}_{t1}(k), \boldsymbol{v}_{t2}(k)]^T$, $\boldsymbol{\omega}(k) = [\boldsymbol{\omega}_r(k), \boldsymbol{\omega}_{t1}(k), \boldsymbol{\omega}_{t2}(k)]^T$, and ΔT is the sampling period. Denote the state vector of the system $\boldsymbol{x} = [\boldsymbol{\xi}_r(k), \boldsymbol{\xi}_{t1}(k), \boldsymbol{\xi}_{t2}(k), \boldsymbol{v}_r(k), \boldsymbol{v}_{t1}(k), \boldsymbol{v}_{t2}(k)]^T$. The system motion is re-written in the folloiwng linear discrete-time state propagation form:

$$\boldsymbol{x}(k+1) = \boldsymbol{F}\boldsymbol{x}(k) + \boldsymbol{G}\boldsymbol{\omega}(k) \tag{3}$$

where $\boldsymbol{F}$ is the state transition matrix that can be obtained from system equation (2), $\boldsymbol{\omega}(k)$ is the vector that represents a white Gaussian noise process with zero mean and covariance $\boldsymbol{Q} = E[\boldsymbol{\omega}(k) \cdot \boldsymbol{\omega}(k)^T]$.

2) Observation Model: We denote the system observation vector received at time-step k as:

$$\boldsymbol{z}(k) = [\boldsymbol{d}^T(k), \boldsymbol{\xi}_r(k), \boldsymbol{\xi}_{t1}(k)]^T \tag{4}$$

where $\boldsymbol{\xi}_r(k)$ is the position of the robot determined from robot self-localization (discussed in Section III-A), $\boldsymbol{\xi}_{t1}(k)$ is the position of Target 1 obtained by Kinect-based human following (discussed in Section III-B), and $\boldsymbol{d}(k) = [d_{r1}(k), d_{r2}(k), d_{12}(k)]^T$ is the vector of relative distances between the robot and targets obtained by acoustic range measurement (discussed in Section III-C) and Kinect-based human following, that is,

$$\begin{aligned} d_{r1} &= \|\boldsymbol{\xi}_r - \boldsymbol{\xi}_{t1}\| = \sqrt{(\boldsymbol{\xi}_r - \boldsymbol{\xi}_{t1})^T \cdot (\boldsymbol{\xi}_r - \boldsymbol{\xi}_{t1})}, \\ d_{r2} &= \|\boldsymbol{\xi}_r - \boldsymbol{\xi}_{t2}\| = \sqrt{(\boldsymbol{\xi}_r - \boldsymbol{\xi}_{t2})^T \cdot (\boldsymbol{\xi}_r - \boldsymbol{\xi}_{t2})}, \\ d_{12} &= \|\boldsymbol{\xi}_{t1} - \boldsymbol{\xi}_{t2}\| = \sqrt{(\boldsymbol{\xi}_{t1} - \boldsymbol{\xi}_{t2})^T \cdot (\boldsymbol{\xi}_{t1} - \boldsymbol{\xi}_{t2})}. \end{aligned} \tag{5}$$

We have the observation model as:

$$\boldsymbol{z}(k) = \boldsymbol{h}(\boldsymbol{x}(k)) + \boldsymbol{\nu}(k) \tag{6}$$

where $\boldsymbol{h}(\cdot)$ describes the observation function that relates the measurement to the system states; $\boldsymbol{\nu}(k)$ is the vector of a white Gaussian noise with zero mean and covariance $\boldsymbol{R} = E[\boldsymbol{\nu}(k) \cdot \boldsymbol{\nu}(k)^T]$.

3) EKF-Based Dynamic Positioning Algorithm: The EKF-based dynamic positioning algorithm takes the input from the observation vector $\boldsymbol{z}(k)$, by the steps of initialization, prediction, and updating, it returns the estimate of the Target 2 position, $\hat{\boldsymbol{\xi}}_{t2}$. The algorithm is described below.

- **Prediction**

With the given motion model, the system process is propagated by the following equation:

$$\hat{\boldsymbol{x}}(k+1|k) = \boldsymbol{F} \cdot \hat{\boldsymbol{x}}(k) \tag{7}$$

$$\hat{\boldsymbol{P}}(k+1|k) = \boldsymbol{F} \cdot \boldsymbol{P}(k) \cdot \boldsymbol{F}^T + \boldsymbol{G} \cdot \boldsymbol{Q} \cdot \boldsymbol{G}^T \tag{8}$$

where the $\hat{\boldsymbol{x}}(k)$ and $\boldsymbol{P}(k)$ are the posterior state estimates and associated covariance matrix, respectively.

- **Update**

After an observation is taken, the posterior estimates of the state vector and covariance are updated by:

$$\hat{\boldsymbol{x}}(k+1) = \hat{\boldsymbol{x}}(k+1|k) + \boldsymbol{K} \cdot [\boldsymbol{z}(k+1) - \boldsymbol{h}(\hat{\boldsymbol{x}}(k+1))] \tag{9}$$

$$\boldsymbol{P}(k+1) = (\boldsymbol{I} - \boldsymbol{K} \cdot \boldsymbol{H}(k)) \cdot \boldsymbol{P}(k+1|k) \tag{10}$$

where

$$\boldsymbol{K} = \boldsymbol{P}(k+1|k) \cdot \boldsymbol{H}^T(k) \cdot [\boldsymbol{H}(k) \cdot \boldsymbol{P}(k+1|k) \cdot \boldsymbol{H}^T(k) + \boldsymbol{R}]^{-1} \tag{11}$$

is the Kalman gain, and $\boldsymbol{H}$ is the Jacobian matrix of observation function $\boldsymbol{h}(\cdot)$ valuated at current prior state $\hat{\boldsymbol{x}}(k+1|k)$, that is $\boldsymbol{H}(k) = [\boldsymbol{H}_d(k), \boldsymbol{H}_p(k)]^T$, where

$$\boldsymbol{H}_{d,[ij]}(k) = \frac{\partial \boldsymbol{h}_i}{\partial \boldsymbol{x}_j} \mid_{\hat{\boldsymbol{x}}(k+1|k)}, i = 1, ..., 3; j = 1, ..., 12 \tag{12}$$

$$\boldsymbol{H}_p = \begin{bmatrix} \boldsymbol{I}_{4\times4} & \boldsymbol{0}_{4\times8} \end{bmatrix} \tag{13}$$

The system noise covariance $\boldsymbol{Q}$ of the motion model (3) and the measurement noise covariance $\boldsymbol{R}$ of the observation model (6) are carefully selected. We consider constant system noise covariance $\boldsymbol{Q}$ which is known a priori. We consider different measurement noise covariance in $\boldsymbol{R}$ due to different noises and uncertainties associated with different devices (such as different sensors used for acoustic ranging, robot self-localization, and human following), which can be obtained offline by running the corresponding subsystems. That is,

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_d & 0 & 0 \\ 0 & \boldsymbol{R}_r & 0 \\ 0 & 0 & \boldsymbol{R}_{t1} \end{bmatrix} \tag{14}$$

where $\boldsymbol{R}_d \in \mathbb{R}^{3\times3}$, $\boldsymbol{R}_r(k) \in \mathbb{R}^{2\times2}$ and $\boldsymbol{R}_{t1}(k) \in \mathbb{R}^{2\times2}$ denote the covariances of the range measurement noise, the robot position estimate, and Target 1 position estimate, respectively.

Remark 1: Given the range-only measurement at each time-step, the probability density representing the possible position of Target 2 is multimodal that has two peaks. However, with the prior target position estimation predicted from the previous step, the multimodal density can be approximated as a unimodal Gaussian distribution. Therefore, the target location is *uniquely* determined real time by the recursive prediction and update process in the above EKF algorithm.

E. Overall Robot-Assisted Localization Algorithm

After describing each of the component subsystem above, we can now present our overall human indoor localization algorithm. Since different sensors return data at different frequencies, we select a common sampling rate for data fusing. For example, the robot self-localization and the Kinect human following subsystem return data at a frequency of $30Hz$, but the acoustic ranging subsystem is slow and returns relative ranging measurement about every 5 seconds due to latency related in beep signal detection and transmission as discussed in Section III-C. To address the multi-rate fusion problem, the proposed localization algorithm fills in prior estimate generated from the previous measurement and the motion model before the acoustic ranging subsystem returns ranging measurement data. The overall algorithm is described in Algorithm 1.

Algorithm 1 Human Indoor Localization Algorithm

1: **repeat**
2: // *robot self localization*
3: **input**: odometry $\boldsymbol{z}_p(k)$ and gyro $\boldsymbol{z}_g(k)$
4: call function $robot\_pose\_ekf \Leftarrow$ Algorithm in Section III-A
5: **output**: $\hat{\boldsymbol{\xi}}_r(k)$
6: // *human follower*
7: **input**: Kinect scan $\boldsymbol{z}_k(k)$, $\hat{\boldsymbol{\xi}}_r(k)$
8: call function $human\_follower \Leftarrow$ Algorithm in Section III-B
9: **output**: $\hat{\boldsymbol{\xi}}_{t1}(k)$ and $d_{r1}(k)$
10: // *acoustic ranging*
11: **input**: recorded sound files
12: **if** all sound files are received **then**
13: call function $acoustic\_ranging \Leftarrow$ Algorithm in Section III-C
14: **else**
15: go to Line 19
16: **end if**
17: **output**: $d_{12}(k)$, $d_{r2}(k)$
18: // *EKF-based position estimation*
19: Initialize first estimation of $\hat{\boldsymbol{x}}(0)$ with covariance $\boldsymbol{P}(0)$
20: **predict**: prior state estimate $\hat{\boldsymbol{x}}(k|k-1)$ and prior covariance $\boldsymbol{P}(k|k-1)$ using equation (7) and (8)
21: **input**: position $\hat{\boldsymbol{\xi}}_r(k)$ and $\hat{\boldsymbol{\xi}}_{t1}(k)$
22: **if** relative ranging $\boldsymbol{d}(k)$ is received **then**
23: **update**: Kalman gain $\boldsymbol{K}$ using equation (11)
24: posterior estimate $\hat{\boldsymbol{x}}(k)$, $\boldsymbol{P}(k)$ using equation (9) and (10)
25: **else**
26: prior estimate $\hat{\boldsymbol{x}}(k|k-1) \rightarrow \hat{\boldsymbol{x}}(k)$, $\boldsymbol{P}(k|k-1) \rightarrow \boldsymbol{P}(k)$
27: **end if**
28: **return**: $\hat{\boldsymbol{x}}(k)$, $\boldsymbol{P}(k)$
29: next time step $k \leftarrow k+1$
30: **output**: $\hat{\boldsymbol{\xi}}_{t2}(k)$
31: **until** time out

IV. EXPERIMENTAL VALIDATION

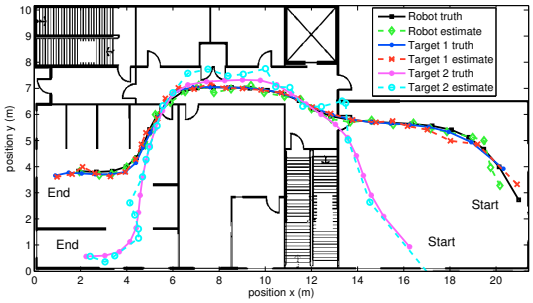
We present both Matlab simulation and real robot experiments in this section.

A. Computer Simulations

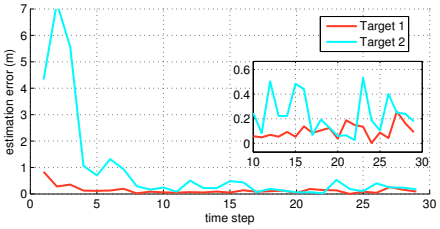
1) Simulation Setup: We simulated Algorithm 1 using the motion model (3) and the observation model (6), and assuming that the output from robot self-localization, Kinect human-follower, and acoustic ranging modules return true values with added noises. The process noise $\boldsymbol{\omega}(k)$ and observation noise $\boldsymbol{\nu}(k)$ are zero-mean with covariance $\boldsymbol{Q} = (0.5m)^2\boldsymbol{I}_6$ and $\boldsymbol{R} = (0.1m)^2\boldsymbol{I}_7$, respectively, where $\boldsymbol{I}_6$ and $\boldsymbol{I}_7$ are the identity matrix of six-dimension and seven-dimension, respectively. The true initial position of the robot and human targets are selected as $\boldsymbol{\xi}(0) = \begin{bmatrix} 21 & 3 & 20 & 4 & 16 & 1 \end{bmatrix}^T$. Human Targets 1 and 2 are set to move in a approximately $10m \times 21.4m$ space with a linear velocity of $0.8m/s$ in average. We assume that the first belief of the position estimate is initialized as $\hat{\boldsymbol{\xi}}(0) = \begin{bmatrix} 5 & 15 & 16 & 5 & 5 & -10 \end{bmatrix}^T$, while the covariance matrix is given as $\boldsymbol{P}(0) = 10\boldsymbol{I}_{12}$. The velocity estimate of the three agents are initialized as $\hat{\boldsymbol{v}}(0) = 0.6\boldsymbol{I}_6$.

2) Simulation Results: Fig. 4a shows the true and estimated trajectories of the robot and targets in our laboratory environment with hallway and a student room. We can see that the positioning algorithm is able to track the motion of two moving persons, and the position estimation gradually

converges with a small error. Fig. 4b shows the estimation error over time, where the estimation error is calculated using the Euclidean distance between the actual and estimated positions. It can be seen that after the algorithm converges, the median and 90% errors for Target 1 are $0.12m$ and $0.2m$, and for Target 2 are $0.35m$ and $0.8m$, respectively.



(a)



(b)

Fig. 4: Simulation results: (a) The true and estimated trajectories of the robot and human targets; (b) Estimation error over time steps.

B. Real Robot-Smartphone Experiments

1) Experimental Setup: The test platform is a Turtlebot mobile robot equipped with a Kinect 3D sensor and an on-board ASUS Eee PC 1215N laptop which has an Intel®Atom™D525 Dual Core Processor. The Kinect generates 180° laser scans horizontally at the rate of $30Hz$. Odometry and EVAL-ADXRS620Z 300°/sec yaw rate gyroscope are also equipped to detect the linear and angular displacement. The robot is operated on Robot Operating System (ROS), where Algorithm 1 was implemented. The HTC EVO smartphones with Android operating system (OS) are used by the human targets for emitting and recording beep signals. The microphones of the smartphones respond to the designed beep sound with $16kHz$ frequency. We used 4 beeps with an interval of 5000 samples in the beep signal. The Android application for acoustic data acquisition on smartphones is developed in Java. Program developed in C++ has been used for data acquisition, scheduling control and processing of acquired sound files for ranging calculation on the robot. Data between different platforms (ROS and Android OS) is exchanged using sockets over WiFi to ensure compatibility between different programs. The experiments are conducted in a 30 by 24 square feet lab area as shown in Fig. 5, which is the rightmost area of the floor plan shown in Fig. 4a. 10 positions are marked as the ground truth along each trajectory of the robot and human targets. Human Target 1 and Target 2 move at an average speed of $0.09m/s$ and $0.12m/s$, respectively.

To characterize the uncertainties in the acoustic ranging measurement, we use experimental data presented in two co-authors' previous work [22], where ranging errors under different environments (lab, train station, mall, airport) are shown (in Figure 7 of [22]). It was shown that the ranging errors within 3m testing distances in lab environments for HTC phone are under 0.15m. In our experiments, the covariance parameter $\boldsymbol{R}_d$ in equation (14) are set to be $\boldsymbol{R}_d = (0.1m)^2\boldsymbol{I}_3$ and $(0.2m)^2\boldsymbol{I}_3$ in two experiments, respectively. The other two covariance parameters in (14) are chosen as $\boldsymbol{R}_r = \boldsymbol{R}_{t1} = (0.2m)^2\boldsymbol{I}_2$.

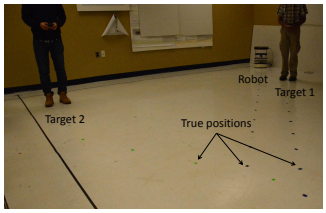


Fig. 5: Real-robot experimental setup.

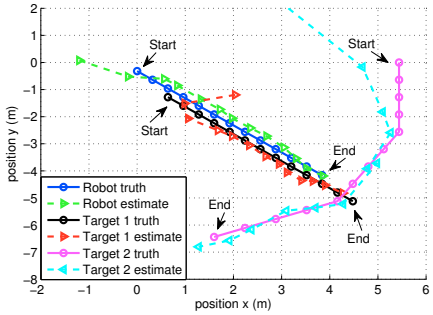
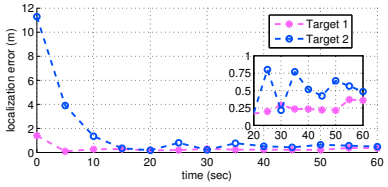


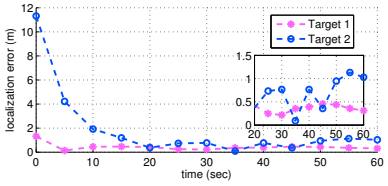
Fig. 6: Experimental results of true and estimated trajectories using the proposed Algorithm 1. Noise covariance is chosen as: $\boldsymbol{R}_d = (0.1m)^2\boldsymbol{I}_3, \boldsymbol{R}_r = \boldsymbol{R}_{t1} = (0.2m)^2\boldsymbol{I}_2$.

2) Experimental Results: Fig. 6 illustrates results from the experiments, where the estimated trajectories of the robot, Target 1, and Target 2 approach true positions as time evolves. Fig. 7 shows temporal propagation of localization errors under different covariances of range measurement $\boldsymbol{R}_d$. It can be seen that the average localization error is around $0.25m$ for Target 1 and $0.5m$ for Target 2 when acoustic ranging covariance is $\boldsymbol{R}_d = (0.1m)^2\boldsymbol{I}_3$. When the acoustic ranging uncertainty increases to $\boldsymbol{R}_d = (0.2m)^2\boldsymbol{I}_3$, the average localization error for Target 2 is around $0.85m$.

3) Performance Discussion: From the experimental results, the human localization error reaches $0.5m$ for a person (Target 2) moving at an average speed of 0.12m/s. The localization accuracy is better than most WiFi-based indoor localization methods reported that depend on sensing infrastructure deployment. Compared to the recent smartphone-based indoor human localization work [22], we achieve better localization accuracy and provide *dynamic* position estimation, while [22] focuses on stationary smartphone



(a)



(b)

Fig. 7: Localization error over time with the covariance of range measurement $\boldsymbol{R}_d = (0.1m)^2\boldsymbol{I}_3$ in (a) and $\boldsymbol{R}_d = (0.2m)^2\boldsymbol{I}_3$ in (b).

localization. We expect that the localization performance is improvable considering the following issues:

a) The accuracy of target positioning largely depends on the result of robot self-localization. In our presented experiments, the robot is localized by fusing sensor data from the odometry and gyroscope, which has non-negligible localization error in large areas. We will use other sensors such as a laser range finder for robot self-localization in future experiments.

b) The acoustic ranging subsystem has significant delays. In our presented experiments, the sound collection takes 3s, data transmission takes 1s roughly, and range calculation takes 1.5s, which results in about 5.5s latency in range measurement. In our future work, we will optimize the ranging subsystem and parameters to decrease the latency, and improve the dynamic position algorithm to deal with the latency more efficiently.

V. CONCLUSIONS

In this paper, we developed a cooperative human indoor localization system utilizing a self-localized mobile robot and smartphones. An EKF-based dynamic localization algorithm was developed to fuse distance-only measurements from both the Kinect 3D vision sensor and smartphone-based acoustic ranging, so that the targets position can be recursively estimated. Experiments were conducted using a Turtlebot and two HTC smartphones, which showed that the positioning algorithm was able to locate and track moving human targets with a localization error around $0.5m$. The localization performance is comparable to recent indoor localization methods using WiFi signature maps, without the cost of deploying intensive sensing infrastructure. In the future work, extensive experiments will be conducted to evaluate system performances in various environments.

REFERENCES

[1] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 6, pp. 1067–1080, 2007.

[2] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: indoor location sensing using active RFID," *Wireless Networks*, vol. 10, no. 6, pp. 701–710, 2004.

[3] A. Bekkali, H. Sanson, and M. Matsumoto, "RFID indoor positioning based on probabilistic RFID map and Kalman filtering," in *Third IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, pp. 21–21, 2007.

[4] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, pp. 775–784, 2000.

[5] M. Youssef and A. Agrawala, "The horus WLAN location determination system," in *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services*, pp. 205–218, 2005.

[6] T. King, S. Kopf, T. Haenselmann, C. Lubberger, and W. Effelsberg, "Compass: A probabilistic indoor positioning system based on 802.11 and digital compasses," in *Proceedings of the 1st International Workshop on Wireless Network testbeds, Experimental Evaluation & Characterization*, pp. 34–40, 2006.

[7] E. Martin, O. Vinyals, G. Friedland, and R. Bajcsy, "Precise indoor localization using smart phones," in *Proceedings of the International Conference on Multimedia*, pp. 787–790, 2010.

[8] J. J. Leonard and H. F. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons," *IEEE Transactions on Robotics and Automation*, vol. 7, no. 3, pp. 376–382, 1991.

[9] P. Jensfelt and S. Kristensen, "Active global localization for a mobile robot using multiple hypothesis tracking," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 5, pp. 748–760, 2001.

[10] W. Burgard, D. Fox, D. Hennig, and T. Schmidt, "Estimating the absolute position of a mobile robot using position probability grids," in *Proceedings of the National Conference on Artificial Intelligence*, pp. 896–901, 1996.

[11] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte Carlo localization for mobile robots," in *IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1322–1328, 1999.

[12] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust Monte Carlo localization for mobile robots," *Artificial Intelligence*, vol. 128, no. 1, pp. 99–141, 2001.

[13] M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 229–241, 2001.

[14] G. Kantor and S. Singh, "Preliminary results in range-only localization and mapping," in *IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1818–1823, 2002.

[15] P. Newman and J. Leonard, "Pure range-only sub-sea SLAM," in *IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1921–1926, 2003.

[16] E. Olson, J. J. Leonard, and S. Teller, "Robust range-only beacon localization," *IEEE Journal of Oceanic Engineering*, vol. 31, no. 4, pp. 949–958, 2006.

[17] J. Djugash, S. Singh, G. Kantor, and W. Zhang, "Range-only SLAM for robots operating cooperatively with sensor networks," in *IEEE International Conference on Robotics and Automation*, pp. 2078–2084, 2006.

[18] D. Fox, "Adapting the sample size in particle filters through KLD-sampling," *The International Journal of Robotics Research*, vol. 22, no. 12, pp. 985–1003, 2003.

[19] J. Santos, "Turtlebot follower tutorials." `http://ros.org/wiki/turtlebot_follower/Tutorials/Demo`. [Online; accessed 5-Feb-2014].

[20] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "Beepbeep: a high accuracy acoustic ranging system using cots mobile devices," in *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems*, pp. 1–14, 2007.

[21] J. Yang, S. Sidhom, G. Chandrasekaran, T. Vu, H. Liu, N. Cecan, Y. Chen, M. Gruteser, and R. P. Martin, "Detecting driver phone use leveraging car speakers," in *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*, pp. 97–108, 2011.

[22] H. Liu, Y. Gan, J. Yang, S. Sidhom, Y. Wang, Y. Chen, and F. Ye, "Push the limit of WiFi based localization for smartphones," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, pp. 305–316, 2012.