Robot-assisted smartphone localization for human indoor tracking

Chao Jiang, Muhammad Fahad, Yi Guo *, Yingying Chen

Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, 07030, USA

HIGHLIGHTS

- A robot-assisted human indoor localization method.
- Data from Kinect vision sensor fused with smartphone-based acoustic ranging.
- Real-time dynamic position estimation and tracking of moving persons.
- Robot–smartphone cooperation without relying on wireless sensing infrastructure.

ARTICLE INFO

ABSTRACT

Smartphone-based human indoor localization was previously implemented using wireless sensor networks at the cost of sensing infrastructure deployment. Motivated by increasing research attention on location-aware human–robot interaction, we propose a robot-assisted human indoor localization scheme utilizing acoustic ranging between a self-localized mobile robot and smartphones. Data from the low-cost Kinect vision sensor are fused with smartphone-based acoustic ranging, and an extended Kalman filter based localization algorithm is developed for real-time dynamic position estimation and tracking. Real robot–smartphone experiments are performed, and performances are evaluated in various indoor environments under different environmental noises and with different human walking speed. Comparing to existing indoor smartphone localization methods, the proposed system does not rely on wireless sensing infrastructure, and has comparable localization accuracy with increased flexibility and scalability due to the mobility of the robot.

1. Introduction

Human indoor localization and tracking have 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 [1,2]. More recently, as intelligent mobile service robots are introduced into the human's life, location-aware human–robot interaction becomes popular [3,4]. Although the global positioning system (GPS) has been dominating the realm of outdoor localization applications, GPS signal transmission is prone to be blocked and distorted by buildings, which severely deteriorates its indoor performance. Accurate, robust localization technologies in indoor and other GPS-denied environments are in great demand by the bloom of indoor location-aware services and applications. In this paper, we propose a novel indoor localization method utilizing robot–smartphone cooperation. Low-cost sensors such as the Kinect sensor on robots are used together with acoustic communication techniques. A new dynamic Kalman filter based indoor human localization and tracking algorithm is developed and validated in indoor experiments.

Smartphone localization has been extensively studied utilizing prevalent WiFi-based techniques, which offer solutions for indoor positioning and localization either leveraging existing wireless access points or with a modification of infrastructure, see Section 2 for a more detailed literature review. The biggest issue of existing WiFi-signature-map based localization without expensive infrastructure deployment is in the localization accuracy of position estimates. It has been reported that significant errors in the magnitude of 6–8 m always exist for WiFi localization [5,6]. Even with recent improved statistical processing of radio signal strength [7] and advanced algorithms utilizing acoustic ranging [6], the limit of WiFi based localization by smartphones is reported to be around 1–2 m [1]. The focus of this paper is to develop accurate smartphone localization with the assistance of a mobile robot, and the goal is to achieve higher localization accuracy without the cost of intensive deployment of sensing infrastructure.

* Corresponding author.
 E-mail addresses: cjiang6@stevens.edu (C. Jiang), mfahad@stevens.edu
(M. Fahad), yi.guo@stevens.edu (Y. Guo), yingying.chen@stevens.edu (Y. Chen).

Meanwhile, localization is 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. Experiments on SLAM of mobile robots in indoor environments were presented in [15], 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 0.2 m and 0.5 m, respectively. The self-localized robot is able to navigate autonomously in a known environment and serves as a mobile beacon node which provides a global reference with respect to the environment. These reasons motivate us to deploy a robot as the mobile position reference to assist smartphone for indoor localization.

In this paper, we propose a cooperative human localization system that uses a mobile robot and smartphones to localize moving persons. The system consists of a self-localized robot that tracks 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 successful localization in different indoor environments, and performances are evaluated under various levels of environmental noises and with different human walking speed. 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 installing complex infrastructure.

The contribution of the paper is threefold. *First*, a new system configuration of robot–smartphone cooperation is proposed for robot-assisted indoor localization. *Second*, an extended Kalman filter based dynamic positioning algorithm is developed to solve the problem of localization and tracking of moving persons in real time. *Third*, the localization accuracy achieved is comparable to the current limit of WiFi-based techniques for smartphone localization without relying on wireless sensing infrastructure. This paper is a significant extension of our early conference paper [16]. Specifically, we have added comprehensive literature review in Section 2, added algorithm details in Section 4, and added extensive experimental results in different indoor environments in Section 5 of the paper.

The rest of the paper is organized as follows: Section 2 reviews relative works on human indoor localization approaches and systems. Section 3 introduces the overview of the proposed robot-assisted localization system. Section 4 describes the details of the system design and the Extended Kalman Filter based dynamic positioning algorithm. In Section 5, we present system implementation and experimental results in different indoor environments. The performance under the effect of environmental noise and walking speed is also discussed in Section 5. We conclude our work in Section 6.

2. Related work

2.1. Radio fingerprint based localization system

The astonishing development of wireless network such as active radio frequency identification (RFID), WLAN and ultra-wide-band has remarkably facilitated human indoor localization techniques using either specialized or minimally modified infrastructure [5]. A RFID based location sensing system was developed in [17]. 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 1 m average accuracy. In [18], an online probabilistic RFID map and adaptive Kalman filter were applied to obtain localization with accuracy from 0.5 m to 5 m, depending on the received signal strength (RSS) noise level and the number of RSS samples collected. The authors in [19] 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 2 m to 3 m. The Horus system proposed in [20] 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 localization accuracy of Horus at the 90th percentile of the error CDFs is about 1 m . The COMPASS system [21] considered the orientation of targets in the online positioning phase, where the problem of blocking effects of human body encountered in [19] was mitigated. The average positioning error was approximately 1.65 m. The use of smartphones as radio signal strength indications (RSSI) in WLAN system was studied in [7], in which the average error about 2 m was reported. Overall, existing wireless indoor localization methods based on WiFi signature maps needs effort on building radio map, and high localization accuracy is usually obtained at the cost of intensive deployment of sensing infrastructure.

To overcome the aforementioned limitation of infrastructure dependence, recent effort has been put into developing new techniques for smartphone localization utilizing embedded sensors and the increasing computational capability onboard of smartphones. In [22], logical localization was achieved through an ambience fingerprint map constructed by combining optical, acoustic and motion measurement from sensors on smartphones. In [23,24], the fingerprint map was built by combining users' motion in their daily activities, which was measured by accelerometers on the smartphones. In [25], the inertial sensor data from smartphones and the WiFi RSSI information were fused through a proposed Monte Carlo Markov chain algorithm for location estimation without constructing radio map beforehand. However, current smartphone-based indoor localization techniques using the off-the-shelf wireless infrastructure suffer low accuracy. Aiming at highly accurate localization, the recent work presented in [1] utilized smartphone-based acoustic ranging to estimate distances from vicinal peer-phones for the purpose of constructing a rigid graph constraint, thus to reduce the localization error to reach the accuracy level of 1 m to 2 m.

2.2. Inertial positioning system

Inertial positioning systems typically rely on the measurements from inertial measurement unit (IMU) such as accelerometers, gyroscope and magnetometers, and have been widely developed to track motion displacement of a moving agent with respect to a known initial position [26]. The maturity of Micro Electro-Mechanical Systems (MEM-S) technology enables inertial sensors to be fabricated in sufficiently tiny scale, which can be integrated into many portable devices as smartphones and tablets. The ubiquity, portability of these sensing devices make it possible to perform practical dynamic position tracking for human. The inertial positioning system was experimentally evaluated with different placement of IMU sensors on a user's body and promising results were shown for human localization [27].

Although a stand-alone inertial positioning system is able to estimate the position of a moving person independent of infrastructure deployment, the estimation error accumulates and grows rapidly over time, which makes the system reliable only in a short period of operation before the error accrues to an unacceptable level. Although methods such as zero velocity update [28] has been proposed to reduce the measurement drift, the problem of error accumulation is invertible thus affect the accuracy of estimation for inertial positioning systems. The hybrid inertial positioning systems that incorporate external measurement or map matching provide a feasible solution to long-term position tracking with inertial sensors. The environment information observed is occasionally used to correct the accrued error and maintain a bounded estimation uncertainty. The work presented in [29] developed an indoor localization system for pedestrians, where measurements from foot-mounted IMU and a detailed environment model are fed into a particle filter to perform dynamic positioning. The estimation error are bounded within 1 m. In [30], the radio signal strength measurements of WiFi or cellular signal are incorporated to provided additional constraint on the position estimation tracked by IMU sensors. Inspired by the SLAM technology in robotics, the radio signal strength map is built online instead of using a pre-surveyed radio map. In [31], the cumulative estimation error of pedestrian position developed by the inertial positioning system is corrected occasionally by relative ranging to sparsely deployed beacons in the environment, and a particle filter fuses both IMU measurements and ranging measurement to obtain position tracking with bounded uncertainties.

2.3. Cooperative localization system

The radio map based localization and the hybrid inertial positioning systems reviewed above rely more or less on existing or added wireless network infrastructure in the indoor environment, which may be difficult to implement in an unknown environment without pre-deployed wireless infrastructure. This could happen in applications such as rescue personnel, first responders, security personnel and military operation unit searching a building for people trapped [32], where infrastructure-independent positioning system is in demand.

Cooperative localization, in which two or more moving agents recursively estimate their positions relying on relative range measurements to peer agents or anchor nodes that provide reference positions, has attracted considerable attention recently as it relaxes the requirement of extensive infrastructure installation. Potential advantages of cooperative localization include the capability of simultaneous localization for multiple agents, increased flexibility and scalability in spatial coverage. The authors in [33] proposed a cooperative localization scheme for multiple human agents who are equipped with dual foot-mounted inertial sensors that track the motion of each person and radio devices that perform inter-agent ranging. A joint state of the position of multiple persons are estimated by a sensor fusion center that can be carried out on one of the persons. In [34], a Bayesian multi-hypothesis initialization strategy was proposed to cope with the initialization problem of Gaussian-based filter for cooperative localization. Recently, the authors in [35] addressed the problem of real-time cooperative tracking of multiple target nodes in ad hoc wireless network environments where only a small number of beacon nodes are deployed. Taking advantage of ranging between target nodes, the proposed cooperative tracking algorithm is able to concurrently estimate the position of multiple targets under low inter-node connectivity.
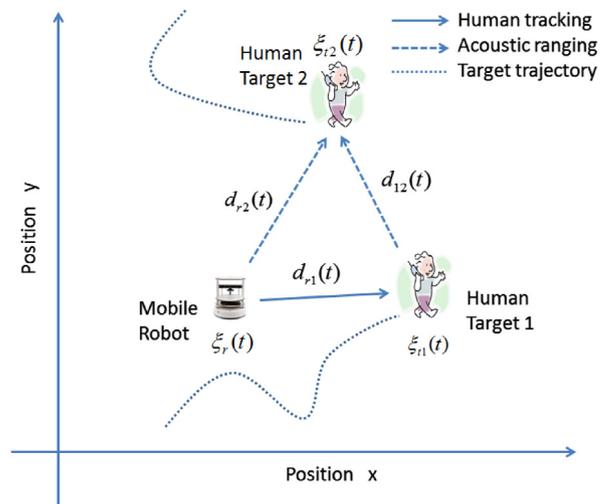
Fig. 1. Overview of robot-assisted indoor human cooperative localization scheme.

2.4. Motivation of proposed work

Motivated by the fact that social robots have been introduced into human's environments and operated alongside humans [36], we propose in this paper a novel robot-assisted human indoor localization scheme utilizing acoustic ranging between a self-localized robot and smartphones. In contrast to the systems that utilize stationary anchor nodes in wireless sensor networks, the flexibility and scalability in coverage of operation area of our proposed system is enhanced by the mobility of the robot. Also, the robot has adequate computational power to execute advanced fusing algorithms (such as the Kalman filter based method proposed in the paper) in real time, thus no central data processing units are required. Furthermore, the ranging subsystem employed in the proposed system does not need any wireless sensor network to be present in the environment, which is especially appealing to environments in search and rescue missions such as firefighter localization.

The advantages of our proposed robot-assisted localization system are twofold. *First*, we provide an alternative localization method that does not rely on the existence of WiFi infrastructure. In environments where the WiFi infrastructure is not available or it is not desirable to rely on it, such as disaster recovery sites or battlefields, our robot-assisted system provides reliable localization results. *Second*, just as smartphones have been woven into people's social life during the past decade, mobile robots are getting cheaper and smaller, and have increasingly been used in human environments. For example, robots are used in intelligent emergency systems for elderly independent living [37], and are also used to assist indoor localization of smartphones [38]. We are witnessing the beginning of an era where robots are used along with other mobile computing devices, such as smartphones and other wearable devices, to provide various services for people in their daily life. Our study provides the methodology for robot–smartphone collaboration, and contributes to the next generation of mobile computing techniques that integrate robots and other mobile devices.

3. System configuration

In this paper, 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 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 proposed localization system utilizes the mobile robot as a moving beacon, and uses range-only measurements to estimate the positions of the human targets. Comparing to traditional lateration-based positioning systems, which requires at least three beacon nodes to uniquely determine the target position in a two-dimensional space [8], the proposed system uniquely determines the targets' positions taking advantage of the mobility of the robot. The coverage of localization is also enlarged by the mobility of the robot and does not depend on fixed infrastructures as used in existing indoor localization methods.

The sensors used in the proposed system include a Kinect sensor, microphones and speakers on both the robot and the smartphones, and robot self-localization sensors (such as odometry, gyroscope, a laser-range finder or a camera that are commonly equipped with a mobile robot [39]). The expected operational distances between the robot and the human targets are within 6 m , which is the maximum detection range of the Kinect sensor. In the next section, we present the functional block diagram of the system, together with a Kalman filter based sensor fusing method to efficiently localize and track humans using the proposed system.

4. Robot-assisted human indoor localization

The proposed human indoor localization system is composed of (1) robot self-localization, (2) human-follower 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-follower 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 in this section.

4.1. 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 [40] 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 outputs the estimation of robot's current position.

Note that the sample sets of robot position estimates maintained by the Monte Carlo localization algorithm are discrete approximation of continuous belief. When the robot is localized in a known environment, the mean of the robot position estimates provides the robot's current position information. We initialize the particle filter with Gaussian distribution and compute a Gaussian
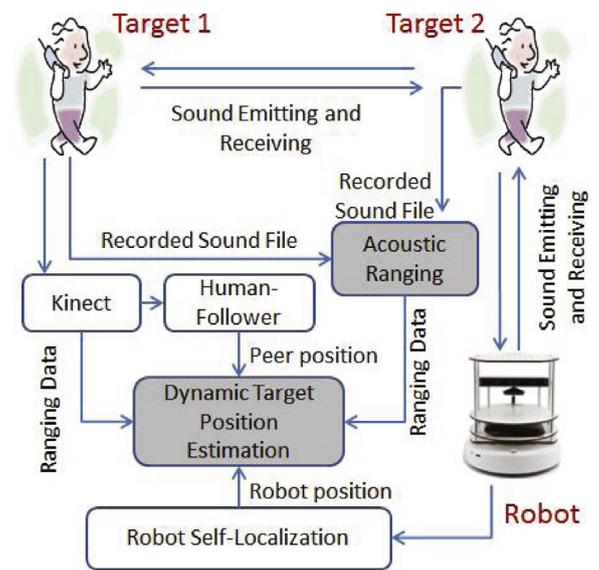
Fig. 2. Functional block diagram of the localization system, where the shaded boxes present new algorithms developed in the paper.

approximation from the discrete samples of robot position estimates using the Density Extraction approach [41]. Therefore, the mean, $\hat{\boldsymbol{\xi}}_r(t)$, and the covariance, $\boldsymbol{P}_r(t)$, of the calculated distribution of estimates are used as the output of the robot self-localization subsystem.

4.2. 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 [42] can be used for this functional module, which returns the distance from robot to Target 1, $d_{r1}(t)$, and the position of Target 1, $\boldsymbol{\xi}_{t1}(t)$.

4.3. 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 is done by time-of-arrival (TOA) method to estimate the sound travel time from one device to another [6]. However, challenges exist in preventing interference in acoustic signal, the lack of clock synchronization, and overcoming uncertainties in emitting and detection. These challenges have been tackled by designing the beep signal to reduce interference, two-way detection to avoid the need for clock synchronization, and a robust signal detection method, details of which are discussed in this section. 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.

4.3.1. Signal design for acoustic ranging

The acoustic signal referred to as the beep, consists of several evenly spaced monotonic signal bursts. The frequency, number of bursts 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 300–3400 Hz, music 50 Hz–15 kHz) is concentrated in lower frequency band, we used high frequency signal in the 16–20 kHz range, which is less susceptible to noise and easier to filter. The ranging accuracy is directly proportional to the number of signal bursts but it also adds measurement latency. We used 4 signal bursts for our experiments with beep interval of 5000 samples. We use 16 kHz 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.

4.3.2. Acoustic signal detection methods

Traditionally, correlation based method discussed in [43] has been used to detect time or arrival. However, correlation method results in larger measurement errors. In this work, we adopt the change point detection method to detect first instance of arrival of the beep signal, which was proposed in the coauthor's (Chen) previous work [6,44]. 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 calculate 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 indicates the instance of arrival of the beep signal. Under different low and high noise environments, change point detection method can achieve within 15 cm measurement accuracy in 300 cm testing distances (see Fig. 7 of [6]).

4.3.3. Multi-agent scheduling and measurement

Our acoustic ranging subsystem extends existing two-agent scheme [43] to multiple agents, i.e., three agents including the robot, Targets 1 and 2. Unlike the fixed-window method used in the recent work [6], where a fixed time window is scheduled to each agent in emitting the beep signal, we propose a new scheduling scheme that relies on an active request and acknowledgments. In our method, the robot sends the command to Targets 1 and 2 to start recording and waits for their acknowledgment. In the next step, it sends the command to Targets 1 and 2 to play the beep signal sequentially and waits for their acknowledgment once the target phone completes playing the beep signal. This active beep-signal scheduling and data acquisition cycle results in faster data processing than existing fixed time-window scheduling [6], thus results in shorter ranging measurement and decreased ranging latency, which is important in our real time human localization system.

The requests sent out by the robot and acknowledgments sent back by the targets, and the recorded sound files are sent using Transmission Control Protocol/Internet Protocol (TCP/IP) communication. The design of the TCP/IP protocol ensures lossless end-to-end data transmission even in the event of packet loss, by an acknowledgment mechanism between the sender and the receiver. Note that the accuracy of acoustic ranging is not affected by network delay or jitter as we use two way detection method in this implementation [43]. WiFi network is used in our implementation, as commonly seen in recent work (e.g. [1,43,45–47]) for acoustic ranging and localization. In the case there is no existing wireless

---

Algorithm 1 *Acoustic Ranging Algorithm*

1: $\tau = 0$
2: **repeat**
3:      Robot starts recording
4:      Robot sends *start_recording* command to Target 1 and 2
5:      Robot receives recording started ack. from Target 1 and 2
6:      Robot plays beep
7:      **for** $i = 1$ **to** 2 **do**
8:          Robot sends play beep command to Target i
9:          Robot receives beep played ack. from Target i
10:     **end for**
11:     Robot sends *stop_recording* command to Target 1 and 2
12:     Robot receives recording stopped ack. from Target 1 and 2
13:     Robot stops recording
14:     Robot retrieves recorded sound files from Target 1 and 2
15:     **if** $\tau \neq 0$ **then**
16:         **wait while:** $(\tau-1)$th thread of *range_calculation* completes

17:         update d_{r2} and d_{12}
18:     **end if**
19:     **start** τth thread of *range_calculation*
20:     $\tau = \tau + 1$
21:     **output:** d_{r2} and d_{12}
22: **until** time out

23: **Function** *range_calculation*
24: STFT of files recorded by Robot, Target 1 and 2
25: calculate S_{12} and S_{r2}, samples between beeps
26: $\Delta T_{12} = S_{12}/\text{f}$ and $\Delta T_{r2} = S_{r2}/\text{f}$
27: **return** $d_{12} = \Delta T_{12} \cdot \text{c}$ and $d_{r2} = \Delta T_{r2} \cdot \text{c}$
28: **End Function**

---

network available to the system, an ad hoc communication network can be set up between the robot and the targets to run our acoustic ranging algorithm.

In addition to the active scheduling scheme, our acoustic ranging module implements a multi-threading approach to further reduce the latency in the data acquisition cycle. We also use persistent socket connections to reduce the time latency for data communication. Taking advantage of the light CPU workload during the data acquisition cycle, the data collection cycle for the τth cycle is done simultaneously to signal processing and range calculation cycle for the $(\tau - 1)$th cycle, which results in reduced latency in comparison to the earlier work [6].

The overall acoustic ranging implementation on the robot is described in Algorithm 1. The algorithm describes our acoustic ranging scheme that runs on the robot. The robot coordinates playing of the sound beeps and recording the sounds files at the targets. The recorded sound files are retrieved by the robot and time of arrival is calculated using the method described in Section 4.3.2. The STFT of the retrieved files is first calculated. Then the numbers of samples between beeps of Target 1 and Target 2 represented by S_{12} in line 25, and of the robot and Target 2 represented by S_{r2} in line 25, are calculated. The values of S_{12} and S_{r2} are then used to calculate ΔT_{12} and ΔT_{r2}, giving the time of flight between Target 1 and Target 2, and the robot and Target 2. This time of flight is then used to calculate d_{12} and d_{r2} in line 27. Note that the parameters $f = 44.1$ kHz denotes the sampling frequency and $c = 340$ m/s denotes the nominal sound velocity. Since the beep playing by each agent is carried out in independent time slots, the extension of this scheme for more than 2 targets can be achieved without those agents interfering with each other.
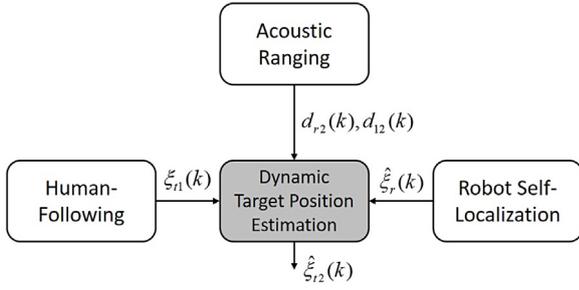
Fig. 3. The block diagram of the localization system, where the shaded box presents the dynamic target position estimation module, and the solid lines represent data flow.

4.4. 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 measurement output from the acoustic ranging, the human following, and returns the position estimate of the human Target 2. We formulate the dynamic positioning problem as follows:

Given that the robot self-localization module outputs the robot position, $\hat{\boldsymbol{\xi}}_r(k)$, the Kinect-based human following module outputs the position of Target 1, $\boldsymbol{\xi}_{t1}(k)$, and the acoustic ranging module outputs the relative distance between the robot and Target 2, $d_{r2}(k)$, Target 1 and Target 2, $d_{12}(k)$, where

$$d_{r2}(k) = \|\boldsymbol{\xi}_r(k) - \boldsymbol{\xi}_{t2}(k)\|_2, \tag{1a}$$
$$d_{12}(k) = \|\boldsymbol{\xi}_{t1}(k) - \boldsymbol{\xi}_{t2}(k)\|_2. \tag{1b}$$

our problem is to design a *dynamic target positioning algorithm* to estimate the position of Target 2, $\boldsymbol{\xi}_{t2}(k)$.

The data flow among each subsystem is shown in Fig. 3. We present the details of the dynamic target positioning algorithm in the next.

4.4.1. Motion model

The motions of involved human agents are not known a priori. We assume that the motions of the nonmaneuvering human targets are driven by white noise acceleration process and are mathematically expressed as [48]:

$$\boldsymbol{\xi}_i(k+1) = \boldsymbol{\xi}_i(k) + \Delta T \cdot \boldsymbol{v}_i(k), \tag{2a}$$
$$\boldsymbol{v}_i(k+1) = \boldsymbol{v}_i(k) + \boldsymbol{\omega}_i(k). \tag{2b}$$

where $i = 1, 2$, representing the two human targets; $\boldsymbol{\xi}_i \in \mathbb{R}^2$, $\boldsymbol{v}_i \in \mathbb{R}^2$ represent the position and velocity of Target 1 or Target 2, respectively; $\boldsymbol{\omega}_i$ is a white Gaussian noise with zero-mean and covariance $\boldsymbol{Q}_i \in \mathbb{R}^{2\times2}$, denoted as $\boldsymbol{\omega}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_i)$; ΔT is the sampling time.

Define the state vector of the system as $\boldsymbol{x}(k) = [\boldsymbol{\xi}_{t1}^T(k), \boldsymbol{v}_{t1}^T(k), \boldsymbol{\xi}_{t2}^T(k), \boldsymbol{v}_{t2}^T(k)]^T$. The system motion is re-written in the following linear discrete-time state propagation form:

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

where the state transition matrix $\boldsymbol{F}$ and the input matrix $\boldsymbol{G}$ are obtained from (2); $\boldsymbol{\omega}(k) = [\boldsymbol{\omega}_{t1}^T(k), \boldsymbol{\omega}_{t2}^T(k)]^T$ is the vector that represents a white Gaussian noise process with zero mean and covariance $\boldsymbol{Q}(k) = E[\boldsymbol{\omega}(k) \cdot \boldsymbol{\omega}(k)^T]$. We consider constant system noise covariance $\boldsymbol{Q}$ which is known a priori. Note that we assume the motion of two human targets are independent, thus the white Gaussian noise $\boldsymbol{\omega}_{t1}$ and $\boldsymbol{\omega}_{t2}$ are uncorrelated.

4.4.2. Observation model

We define the system observation vector $\boldsymbol{z}(k) = [z_1(k), z_2(k), \boldsymbol{z}_3^T(k)]^T$ received at time-step k as

$$z_1(k) = d_{r2}(k) + v_{d1}(k), \tag{4a}$$
$$z_2(k) = d_{12}(k) + v_{d2}(k), \tag{4b}$$
$$\boldsymbol{z}_3(k) = \boldsymbol{\xi}_{t1}(k) + \boldsymbol{v}_{t1}(k). \tag{4c}$$

where $z_1(k)$ and $z_2(k)$ are the measurement of relative distance between the robot and Target 2, Target 1 and Target 2, respectively, which are obtained by the acoustic ranging (discussed in Section 4.3). We assume the relative distance measurement is affected by the zero-mean Gaussian white noise $v_{d1}(k)$ and $v_{d2}(k)$, respectively, which have the same variance σ_d^2, denoted as $v_{d1}, v_{d2} \sim \mathcal{N}(0, \sigma_d^2)$. The position measurement of Target 1, $\boldsymbol{z}_3(k)$, is obtained by Kinect-based human following (discussed in Section 4.2). The position measurement is affected by the noise $\boldsymbol{v}_{t1}$ which is assumed as zero-mean Gaussian white noise with covariance $\boldsymbol{R}_{t1} \in \mathbb{R}^{2\times2}$, denoted as $\boldsymbol{v}_{t1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_{t1})$. The distance measurement noise $v_{d1}(k)$, $v_{d2}(k)$, and the position measurement noise $\boldsymbol{v}_{t1}$ are independent.

We then write the observation model in a compact form, that is

$$\boldsymbol{z}(k) = \boldsymbol{h}(\boldsymbol{x}(k)) + \boldsymbol{v}(k). \tag{5}$$

where $\boldsymbol{h}(\cdot)$ describes the observation function that is defined in (4); $\boldsymbol{v}(k) = [v_{d1}(k), v_{d2}(k), \boldsymbol{v}_{t1}(k)^T]^T$ is the vector of Gaussian noise with zero mean and covariance $\boldsymbol{R}(k) = E[\boldsymbol{v}(k) \cdot \boldsymbol{v}(k)^T]$. To derive the covariance $\boldsymbol{R}(k)$, we analyze the measurement error using the method in [49] in the next.

By linearizing $z_1(k)$ and $z_2(k)$ in (4a) and (4b), the measurement error $\tilde{z}_1(k)$ is obtained as

$$\begin{aligned} \tilde{z}_1(k) &= z_1(k) - \hat{z}_1(k), \\ &\simeq \boldsymbol{H}_{r2,1}(k) \cdot \tilde{\boldsymbol{x}}(k) + \Gamma_{r2} \cdot \boldsymbol{v}_{r2}(k). \end{aligned} \tag{6}$$

where $\boldsymbol{H}_{r2,1}(k) \in \mathbb{R}^{1\times8}$ is the first order partial derivative of $d_{r2}(k)$ with respect to the system state $\boldsymbol{x}(k)$, valuated at $\hat{\boldsymbol{x}}(k|k-1)$, and

$$\tilde{\boldsymbol{x}}(k) = \boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k|k-1),$$
$$\Gamma_{r2}(k) = \begin{bmatrix} 1 & \boldsymbol{H}_{r2,2}(k) \end{bmatrix},$$
$$\boldsymbol{v}_{r2}(k) = \begin{bmatrix} v_{d1}(k) & \tilde{\boldsymbol{\xi}}_r^T(k) \end{bmatrix}^T.$$

$\boldsymbol{H}_{r2,2}(k) \in \mathbb{R}^{1\times2}$ is the first order partial derivative of $d_{r2}(k)$ with respect to the robot position $\boldsymbol{\xi}_r(k)$, valuated at $\hat{\boldsymbol{\xi}}_r(k)$; $\tilde{\boldsymbol{\xi}}_r(k) = \boldsymbol{\xi}_r(k) - \hat{\boldsymbol{\xi}}_r(k)$ is the estimation error of robot position. Given that $v_{d1}(k)$ and $\tilde{\boldsymbol{\xi}}_r^T(k)$ are independent, the variance of the measurement error $\tilde{z}_1(k)$ is expressed as

$$\begin{aligned} \sigma_{r2}^2(k) &= \Gamma_{r2}(k) \cdot \mathbb{E}\{\boldsymbol{v}_{r2}(k) \cdot \boldsymbol{v}_{r2}^T(k)\} \cdot \Gamma_{r2}^T(k) \\ &= \boldsymbol{H}_{r2,2}(k) \cdot \mathbb{E}\{\tilde{\boldsymbol{\xi}}_r(k) \cdot \tilde{\boldsymbol{\xi}}_r^T(k)\} \cdot \boldsymbol{H}_{r2,2}^T(k) \\ &\quad + \mathbb{E}\{v_{d1}(k) \cdot v_{d1}^T(k)\} \\ &= \boldsymbol{H}_{r2,2}(k) \cdot \boldsymbol{P}_r(k) \cdot \boldsymbol{H}_{r2,2}^T(k) + \sigma_d^2. \end{aligned} \tag{7}$$

where $\boldsymbol{P}_r(k)$ is the covariance of the error in robot position estimation. One can see that the sources of noise and uncertainties that contribute to the measurement error $\tilde{z}_1(k)$ include both the uncertainty of robot position estimation and the acoustic ranging noise.

The measurement error $\tilde{z}_2(k)$ is obtained as

$$\begin{aligned} \tilde{z}_2(k) &= z_2(k) - \hat{z}_2(k) \\ &\simeq \boldsymbol{H}_{12}(k) \cdot \tilde{\boldsymbol{x}}(k) + v_{d2}(k). \end{aligned} \tag{8}$$

Algorithm 2 EKF-based Position Estimation Algorithm

1: **input**: measurements $\boldsymbol{z}(k)$
2: **if** time step $k = 0$ **then**
3:     Initialize first estimation of $\hat{\boldsymbol{x}}(0)$ with covariance $\boldsymbol{P}(0)$
4: **else**
5:     **predict**: prior state estimate $\hat{\boldsymbol{x}}(k|k-1)$ and prior covariance $\boldsymbol{P}(k|k-1)$ using (14) and (15)
6:     **if** acoustic ranging $z_1(k), z_2(k)$ updated **then**
7:         **update**: Kalman gain $\boldsymbol{K}(k)$ using (18),
8:         posterior estimate $\hat{\boldsymbol{x}}(k), \boldsymbol{P}(k)$ using (16) and (17), respectively
9:     **else**
10:         prior estimate $\hat{\boldsymbol{x}}(k|k-1) \rightarrow \hat{\boldsymbol{x}}(k), \boldsymbol{P}(k|k-1) \rightarrow \boldsymbol{P}(k)$
11:     **end if**
12: **end if**
13: **output**: posterior position estimate and associated covariance: $\hat{\boldsymbol{x}}(k), \boldsymbol{P}(k)$

where $\boldsymbol{H}_{12}(k) \in \mathbb{R}^{1 \times 8}$ is the first order partial derivative of $d_{12}(k)$ with respect to the system state $\boldsymbol{x}(k)$, valuated at $\hat{\boldsymbol{x}}(k|k-1)$. Then, the variance of the measurement error $\tilde{z}_2(k)$ is given by

$$\sigma_{12}^2(k) = \mathbb{E}\{v_{d2}(k) \cdot v_{d2}^T(k)\} = \sigma_d^2. \tag{9}$$

The measurement error $\tilde{\boldsymbol{z}}_3(k)$ is obtained as

$$\begin{aligned}\tilde{\boldsymbol{z}}_3(k) &= \boldsymbol{z}_3(k) - \hat{\boldsymbol{z}}_3(k) \\ &= \tilde{\boldsymbol{\xi}}_{t1}(k) + \boldsymbol{v}_{t1}(k).\end{aligned} \tag{10}$$

where $\tilde{\boldsymbol{\xi}}_{t1}(k) = \boldsymbol{\xi}_{t1}(k) - \hat{\boldsymbol{\xi}}_{t1}(k|k-1)$. Then, the covariance of the measurement error $\tilde{\boldsymbol{z}}_3(k)$ is given by

$$\begin{aligned}\boldsymbol{R}_{t1}(k) &= \mathbb{E}\{\boldsymbol{v}_{t1}(k) \cdot \boldsymbol{v}_{t1}^T(k)\} \\ &= \boldsymbol{P}_r(k) + \boldsymbol{R}_f(k).\end{aligned} \tag{11}$$

where $\boldsymbol{R}_f(k)$ denotes the covariance of the Gaussian white noise corresponding to the relative position measurement. One can see from the equation that the uncertainty of Target 1 position measurement depends on both the uncertainty of robot position estimation, $\boldsymbol{P}_r(k)$, and the uncertainty corresponding to the relative position measurement with respect to the robot, $\boldsymbol{R}_f(k)$. Note that the estimation error of robot position and relative position measurement from the Kinect sensor are independent.

Due to the common component of measurement uncertainty originating from the robot position estimation in (7) and (11), the measurement error of $z_1(k)$ and $\boldsymbol{z}_3(k)$ are correlated, and the correlation is given as

$$\begin{aligned}\boldsymbol{R}_{r2,t1}(k) &= \Gamma_{r2}(k) \cdot \mathbb{E}\{\boldsymbol{v}_{r2}(k) \cdot \boldsymbol{v}_{t1}^T(k)\} \\ &= \Gamma_{r2}(k) \cdot \begin{bmatrix} \boldsymbol{0}_{2\times 1} & \boldsymbol{P}_r(k) \end{bmatrix}^T \\ &= \boldsymbol{H}_{r2,2}(k) \cdot \boldsymbol{P}_r(k).\end{aligned} \tag{12}$$

The results of (7), (9), (11) and (12) allows for the evaluation of the 4×4 covariance matrix $\boldsymbol{R}(k)$ of the measurement error in (5), which can be written as

$$\boldsymbol{R}(k) = \begin{bmatrix} \sigma_{r2}^2(k) & 0 & \boldsymbol{R}_{r2,t1}(k) \\ 0 & \sigma_{12}^2(k) & \boldsymbol{0}_{1\times 2} \\ \boldsymbol{R}_{r2,t1}^T(k) & \boldsymbol{0}_{2\times 1} & \boldsymbol{R}_{t1}(k) \end{bmatrix}. \tag{13}$$

4.4.3. 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 position estimate of the Target 1, $\hat{\boldsymbol{\xi}}_{t1}$ and the Target 2, $\hat{\boldsymbol{\xi}}_{t2}$. The algorithm is described below.

- **Prediction**

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

$$\hat{\boldsymbol{x}}(k|k-1) = \boldsymbol{F} \cdot \hat{\boldsymbol{x}}(k-1) \tag{14}$$

$$\hat{\boldsymbol{P}}(k|k-1) = \boldsymbol{F} \cdot \boldsymbol{P}(k-1) \cdot \boldsymbol{F}^T + \boldsymbol{G} \cdot \boldsymbol{Q} \cdot \boldsymbol{G}^T \tag{15}$$

where the $\hat{\boldsymbol{x}}(k-1)$ and $\boldsymbol{P}(k-1)$ 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:

$$\hat{\boldsymbol{x}}(k) = \hat{\boldsymbol{x}}(k|k-1) + \boldsymbol{K}(k) \cdot [\boldsymbol{z}(k) - \boldsymbol{h}(\hat{\boldsymbol{x}}(k|k-1))] \tag{16}$$

$$\boldsymbol{P}(k) = (\boldsymbol{I} - \boldsymbol{K}(k) \cdot \boldsymbol{H}(k)) \cdot \boldsymbol{P}(k|k-1) \tag{17}$$

where

$$\boldsymbol{K}(k) = \boldsymbol{P}(k|k-1) \cdot \boldsymbol{H}^T(k) \cdot [\boldsymbol{H}(k) \cdot \boldsymbol{P}(k|k-1) \cdot \boldsymbol{H}^T(k) + \boldsymbol{R}(k)]^{-1} \tag{18}$$

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

$$\boldsymbol{H}_{d,[ij]}(k) = \frac{\partial \boldsymbol{h}_i}{\partial \boldsymbol{x}_j}|_{\hat{\boldsymbol{x}}(k|k-1)}, i = 1, 2; j = 1, \ldots, 8 \tag{19}$$

$$\boldsymbol{H}_p = \begin{bmatrix} \boldsymbol{I}_{2\times 2} & \boldsymbol{0}_{2\times 6} \end{bmatrix} \tag{20}$$

Given the range-only measurement at each time-step, the probability density representing the possible position of Target 2 is multi-modal. However, with the prior target position estimation predicted from the previous step, the multi-modal density can be approximated as a uni-modal Gaussian distribution. Therefore, the target location is *uniquely* determined in real time by the recursive prediction and update process in the above EKF algorithm. The implementation of the proposed EKF-based algorithm is described in Algorithm 2.

5. Experimental validation

To validate the proposed indoor human localization system and the dynamic positioning algorithm, we conducted experiments in two different indoor environments, namely, a lab/office environment and an atrium environment on the campus of Stevens Institute of Technology. In this section, we present the experimental setup, testing results, and performance evaluation under different environmental and operating conditions.

5.1. Experimental setup

Mobile Robot: As shown in Fig. 4(a), the experiment platform includes a Pioneer 3AT mobile robot equipped with a SICK laser range finder, a Kinect 3D sensor and an on-board ASUS X201-DH01 laptop which has an Intel® Celeron® 847 Dual Core Processor. The Kinect generates 57° laser scans horizontally at the rate of 30 Hz. Odometry is also equipped to detect the linear and angular displacement. The robot is operated on Robot Operating System (ROS), where Algorithm 2 was implemented.

Smartphones: The LG® E960 smartphones with Android operating system (OS), shown in Fig. 4(b), are used by the human targets for emitting and recording beep signals. The microphones of the smartphones respond to the designed beep sound with 16 kHz
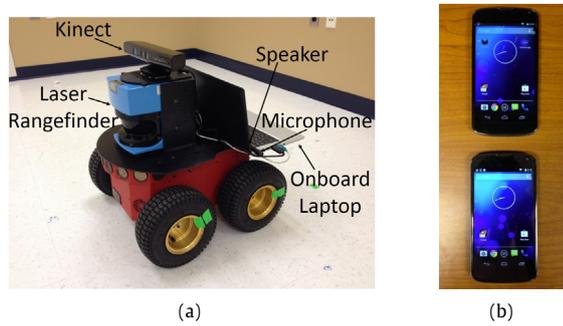
Fig. 4. Experimental testbeds: (a) Pioneer 3AT autonomous mobile robot with SICK LMS200 laser range finder and Kinect; (b) LG E960 smartphones with Android operating system.
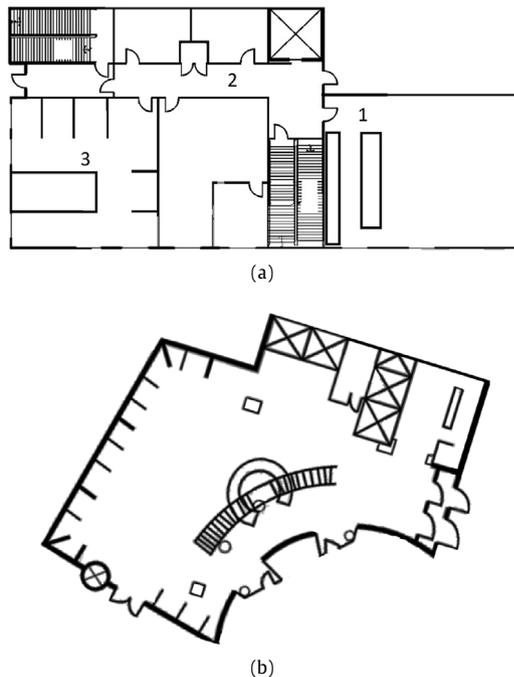
Fig. 5. The floor plan: (a) The lab/office environment: 1—laboratory, 2—corridor, and 3—office; (b) The atrium environment in the Babbio Center of the Stevens campus.

frequency. We used 4 beeps with an interval of 5000 samples in the beep signal.

Implementation: The software programs for system implementation are developed on both the robot laptop and the smartphones. 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. Program for estimation algorithm is developed in C++ and wrapped as a ROS package. Data between different platforms (ROS and Android OS) is exchanged using sockets over wireless network to ensure compatibility between different programs.

Experimental Environments:

(1) The Lab/Office Environment: The first experiment takes place in an indoor campus building with a laboratory, an office and corridor connected between them. The total size is about 69 by 33 square feet, as shown in Fig. 5(a). The laboratory room has an open testing area and shelves and tables located in one corner of the room. The office room is furnished with cubicles and dividers
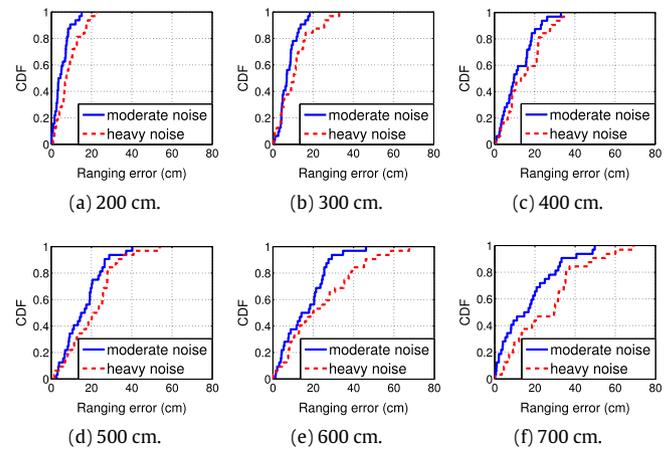
Fig. 6. Cumulative density function (CDF) of ranging errors under moderate and heavy noises in the lab/office environment.

that are 1.6 m in height. The corridor is 1.8 m in width, connecting the laboratory room and the office room.

(2) The Atrium Environment: As shown in Fig. 5(b), the atrium is approximately 400 square meters with an open area, a 1 m high round table, a stair, and a few pillars scattered in the area. It is located in the Babbio center of the Stevens campus.

5.2. Evaluation of the acoustic ranging and robot self-localization subsystems

Before we tested Algorithm 2 for the complete human indoor localization system, we first tested each subsystem shown in Fig. 2. Particularly, we evaluated two subsystems, the acoustic ranging and the robot self-localization subsystems, for its performances under different environmental and operating conditions to characterize the uncertainties and choose appropriate noise covariance matrices, which may significantly affect the performance of the overall system.

5.2.1. The acoustic ranging subsystem

We carried out experiments on the acoustic ranging subsystem to obtain the statistical noise characteristics of ranging measurement error in the lab/office and the atrium environments under different environmental noise levels. The statistical results obtained were used for selecting the variance of acoustic ranging noise σ_d^2. Specifically, in each environment, the experiments were conducted under 50 dB (moderate) and 70 dB (heavy) environmental noise. Two smartphones were placed apart at a distance varying from 200 cm to 700 cm. For each separation distance, 32 trials are performed and the results were analyzed to obtain the statistical results of measurement accuracy.

Fig. 6 shows the cumulative density function (CDF) of ranging errors under different environmental noise levels in the lab/office environment. It shows that the level of background noise slightly affects the ranging accuracy when the two smartphones were put apart by 700 cm or less. As shown in Fig. 6(d), for example, the 90th percentile error are less than 30 cm under both background noise levels, with two smartphones 500 cm apart. Better results are achieved in the case of smaller separation distance, as shown in Fig. 6(a) to 6(c). Fig. 7 illustrates the comparison of median and 90th percentile ranging errors between the lab/office and the atrium environments. It shows that the atrium environment has a slightly increased estimation error, which is caused by the higher noise levels in the atrium environment comparing to the quieter lab/office environment.
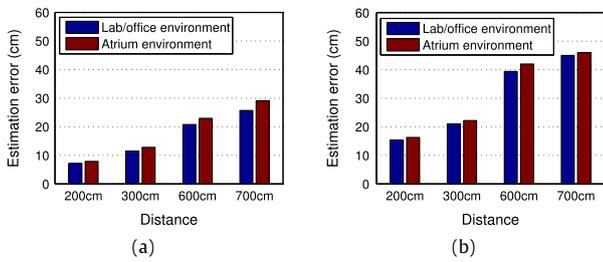
Fig. 7. Comparison of ranging errors between the lab/office and the atrium environments: (a) Median error; (b) 90th percentile error.

From the statistical results, we found that the standard deviation of ranging errors are less than 0.1 m under the moderate noise level and 0.15 m under the heavy noise level in lab/office environment when the separation distance is below 400 cm. Accordingly, in our experiments for the lab/office environment, the variance of ranging errors σ_d^2 in was set to be $\sigma_d^2 = 0.1^2$ m^2 and 0.15^2 m^2 under the moderate and heavy noises, respectively. Considering the larger distance between the robot and targets during the experiments in the atrium environment, the covariance σ_d^2 was hence set to be 0.25^2 m^2. The ranging error covariance was selected in order to obtain superior performance of the filtering algorithm discussed in Section 4.4.3. Inappropriate selection of the parameter could affect the convergence rate and optimality in position estimate.

5.2.2. The robot self-localization subsystem

We performed robot autonomous navigation experiments in both the lab/office and the atrium environments, and recorded the time evolution of robot position estimation. Specifically, during the experiment, the robot started navigation at an approximately known initial position with respect to a map which was built in advance and moved at a speed of 0.4 m/s, and returned to the initial position. The robot self-localization subsystem returns in real time the robot's estimated position and the associated co-variance, $\boldsymbol{P}_r(k)$, which reflects the uncertainty of the robot localization algorithm discussed in Section 4.1. The standard deviation of robot position estimate were calculated by taking square root of the covariance. Fig. 8(a) and 8(b) show the time evolution of standard deviation of robot position estimate. It can be seen that the estimated uncertainties in both positions x and y converge after 20 s. As shown by the red dashed lines in the inset subfigures, the standard deviations in both positions x and y are upper bounded by 0.1 m for the lab/office environment and 0.18 m for the atrium environment. The temporal average of the standard deviation after 20 s is around 0.06 m for the lab/office environment and 0.1 m for atrium environment. These experimental results show that our robot self-localization subsystem has a good accuracy. We further plot the autocorrelation functions of the estimated uncertainty in Fig. 8(c) and 8(d) for the lab/office environment and the atrium environment, respectively. We can see from the figures that the robot localization uncertainty can be considered uncorrelated.

It is worth noting that in many practical problems, Gaussians are *robust* estimators, and EKFs have been applied with great success to state estimation problems that the underlying assumptions are not strictly followed [41].

5.3. Overall performance of the localization system

5.3.1. Indoor lab/office environment

In this subsection, we present our experimental results of the overall localization system in the indoor lab/office environment

as shown in Fig. 5(a). The trajectories of the two human targets are pre-selected starting from the laboratory room to the office room going through the corridor. To obtain the ground truth for performance evaluation, we marked 19 waypoints on the floor for the persons (*i.e.*, Targets 1 and 2) to follow. We started each trial of the experiments by taking the first acoustic ranging measurement at the start position of each person, and the robot follows Target 1 as both persons move along the marked waypoints at an average speed of 0.4 m/s. We run the human localization algorithm on the robot in the ROS environment.

Experiments were conducted during the evening and daytime, and recorded for performance evaluation as shown in Figs. 9 and 10 for evenings testing with moderate noise level and daytime testing with heavy noise level, respectively. Figs. 9(a) and 10(a) show the estimated positions of Targets 1 and 2 together with the ground truth trajectories marked by the waypoints. Figs. 9(b) and 10(b) illustrate estimation errors versus time. It can be seen from the results that the proposed positioning algorithm is able to track the motion of the moving persons, and the algorithm takes about 15 to 18 s (roughly 5 to 6 updating iterations) to converge without the prior knowledge of target initial positions. Steady state estimation error is calculated by taking the average of estimation errors for Targets 1 and 2 after 18 s (when the algorithm converges). Calculated from Fig. 9(b), the steady-state estimation errors in the evening are 0.48 m for Target 1, and 0.76 m for Target 2. In comparison, calculated from Fig. 10(b), the average errors in the daytime are 0.42 m for Target 1, and 0.87 m for Target 2. We can see that better localization accuracy was obtained in the evening due to the noise effect, because background noises from the environment and pedestrians during the daytime degrade the acoustic ranging accuracy comparing to the evening setup.

Furthermore, we characterize the position estimation uncertainty of Target 2 using the square root of the determinant of estimation covariance matrix, $\sqrt{det(\boldsymbol{P}_{t2})}$, associated with Target 2, that is, $\boldsymbol{P}_{t2}$ is the submatrix of the covariance matrix $\boldsymbol{P}$ defined in (15) in Section 4.4.3, and $\boldsymbol{P} \in \mathbb{R}^{8 \times 8}$ has the diagonal entries of $\{\boldsymbol{P}_{t1}, \boldsymbol{P}_{v,t1}, \boldsymbol{P}_{t2}, \boldsymbol{P}_{v,t2}\}$ with the first two submatrix elements denoting the position and velocity covariance matrices for Target 1, respectively, and the last two submatrix elements representing the position and velocity covariance matrix for the robot Target 2, respectively. The temporal propagation of the estimated localization uncertainty for Target 2 are shown in Figs. 9(c) and 10(c). The steady-state estimation uncertainty during daytime is slightly larger than that during evening.

5.3.2. Indoor atrium environment

In this subsection, we present the experiment performed in the atrium environment of the Babbio Center on Stevens campus. Similar to the lab/office environment presented in the last subsection, the trajectories of human targets were pre-selected along 12 waypoints, which are used as ground-truth positions in performance evaluation. The process of each trial of experiments were the same as described in the previous subsection for the lab/office environment, and the experiment result is shown in Fig. 11, where Fig. 11(a) and 11(b) illustrate the temporal propagation of the estimation errors. With unknown initial position, the estimation errors converge after 12 s. (about 4 iterations) for both Targets 1 and 2. The steady-state error, obtained by taking the average of estimation errors after 12 s (when the estimates converge), are 0.51 m for Target 1 and 1.37 m for Target 2, respectively. The temporal propagation of estimated localization uncertainty, $\sqrt{det(\boldsymbol{P}_{t2})}$, for Target 2, is shown in Fig. 11(c), which shows that the estimated uncertainty settles around 0.25 m^2 after 15 s.

Comparing the experimental results obtained in the lab/office environment and the atrium environment, we found the performance of the system in the lab/office environment is slightly better than that of the atrium environment. This can be explained
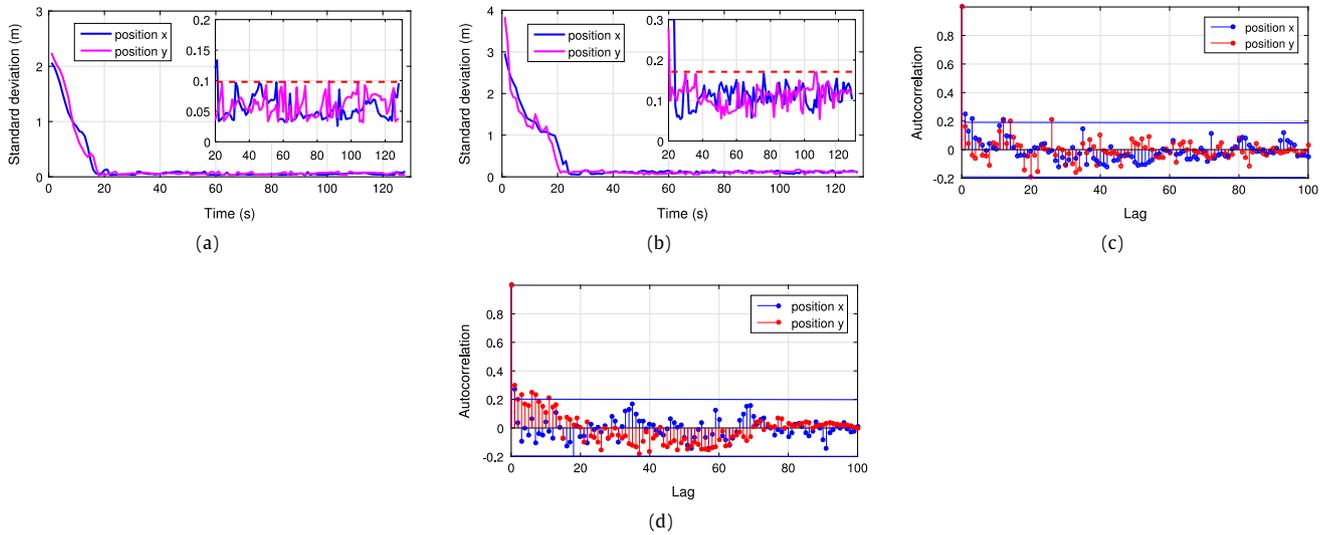
Fig. 8. Robot position estimation: temporal evolution of standard deviation of robot position estimate in (a) the lab/office environment; (b) the atrium environment; (c) the autocorrelation function of the estimated uncertainty shown in (a); (d) the autocorrelation function of the estimated uncertainty shown in (b). The red dashed lines in the inset figures of (a) and (b) correspond to the upper bound of the standard deviation that are used as the noise characterization of robot self-localization.
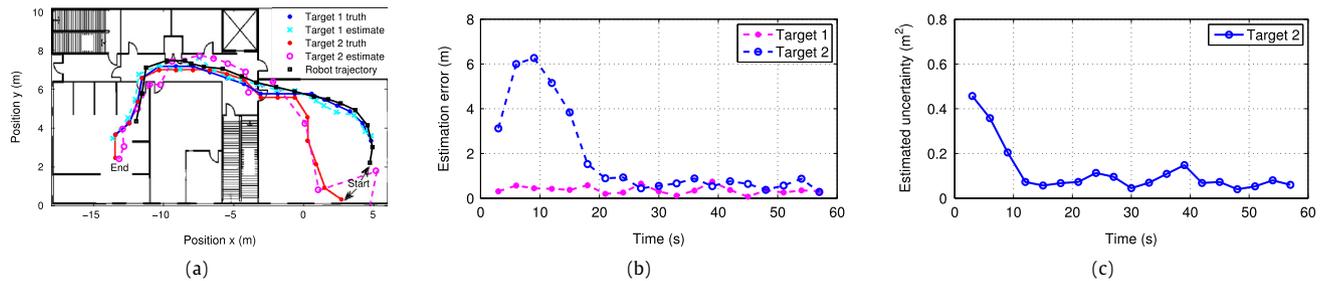
Fig. 9. Human localization experiment in the lab/office environment under moderate noise level (during evenings): (a) True and estimated trajectories; (b) Temporal propagation of estimation errors; (c) Estimated localization uncertainty, *i.e.*, the square root of the determinant of estimation covariance matrix, $\sqrt{det(\boldsymbol{P}_{t2})}(m^2)$, for position estimate of Target 2.
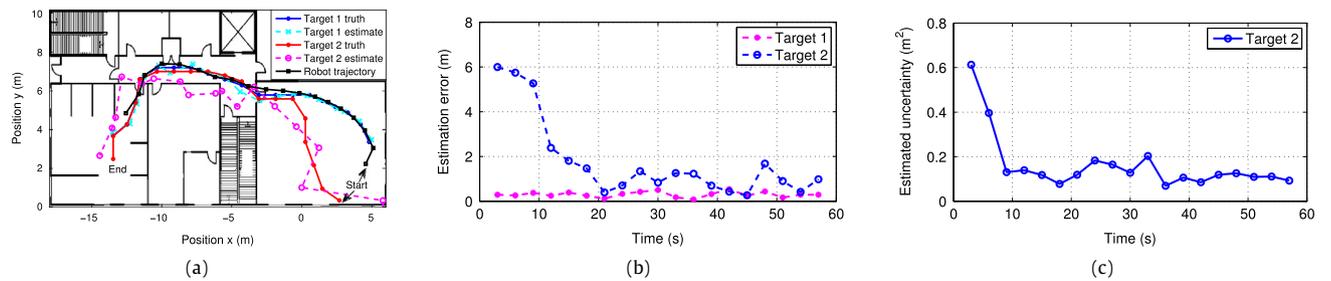
Fig. 10. Human localization experiment in the lab/office environment under heavy noise level (during daytime): (a) True and estimated trajectories; (b) Temporal propagation of estimation errors; (c) Estimated localization uncertainty, *i.e.*, the square root of the determinant of estimation covariance matrix, $\sqrt{det(\boldsymbol{P}_{t2})}(m^2)$, for position estimate of Target 2.

from two aspects of the environmental effects. First, the atrium environment has a larger open area. Correspondingly, the relative distances between the robot and the persons (targets) can increases up to 6.5 m during the experiment, comparing to the maximum relative distance of 4 m in the lab/office environment. As demonstrated in Fig. 7, the 90th percentile ranging error increases to around 0.4 m as the distance increase to 6 m or 7 m. Hence larger acoustic ranging error caused by the distance between the robot and targets contributes to the overall localization performance deterioration. Second, the atrium environment has sparse fixtures and a huge glass wall on one side, comparing to the lab/office environment that has a nearby wall or doors or other landmarks to sense all the time. During the robot self-localization process, the robot frequently updates and corrects its position estimate using existing references such as pillars and tables, so lack of environmental fixtures affects robot self-localization accuracy.

5.4. Performance impact factors and discussion

We have conducted experiments under different environmental and operating conditions. The performances are evaluated and discussed in this subsection.
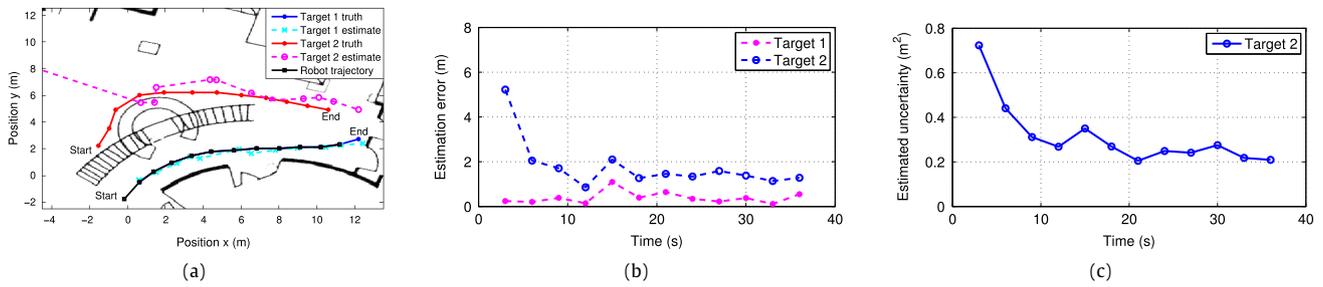
Fig. 11. Human localization experiment in the atrium environment: (a) True and estimated trajectories; (b) Temporal propagation of estimation errors; (c) Estimated localization uncertainty, *i.e.*, the square root of the determinant of estimation covariance matrix, $\sqrt{det(\boldsymbol{P}_{t2})}(m^2)$, for position estimate of Target 2.
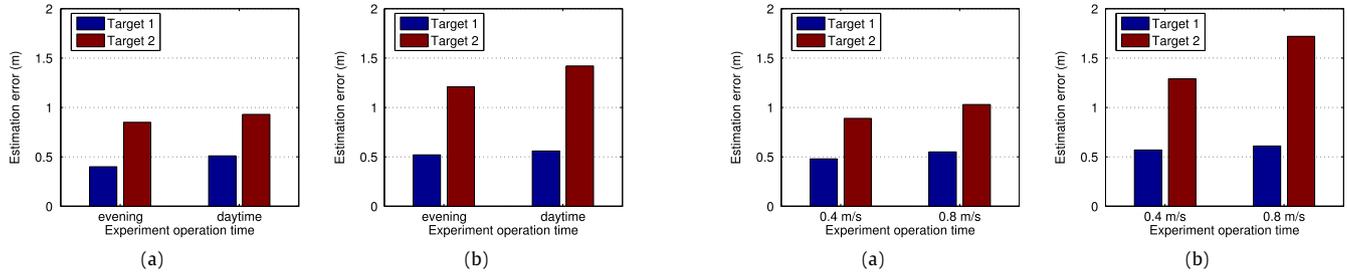
Fig. 12. Estimation accuracy under different environmental noise levels: (a) median error; (b) 90th percentile error.

Fig. 13. Estimation accuracy under different walking speed: (a) median error; (b) 90th percentile error.

5.4.1. Effect of environmental noise

Environmental noise is an important factor that affects the estimation accuracy of the acoustic ranging subsystem. To evaluate the performance of the proposed system, we performed experiments during the daytime and evenings with different noise levels. While the environment noise is moderate in the evening, which is around 50 dB, the experiments could experience 70 dB noise during the daytime due to human conversation, people walking by, air conditioning units, and other types of noises.

The experiments were repeated 10 times during daytime and 10 times during evenings in the lab/office environment to obtain the statistics of estimation accuracy. The steady state estimation error of each trial of the experiments was used to calculate the median and 90th percentile error. Fig. 12 shows the comparison of the estimation accuracy between different environmental noise levels for both Target 1 and Target 2. As shown in Fig. 12(a), the median error of Target 1 and Target 2 are 0.43 m and 0.85 m under the moderate noise level (evening), respectively, in comparison to 0.51 m and 0.93 m under the heavy noise level (daytime), respectively. The 90th percentile error of Target 1 and Target 2, as shown in Fig. 12(b), are 0.52 m and 1.21 m under the moderate noise level, respectively, in comparison to 0.56 m and 1.42 m under the heavy noise level, respectively. The experimental results show that the estimation accuracy for Target 2 degrades by approximately 17% as the noise level increases, which is due to sensitivity of the acoustic ranging subsystem to environmental noises. In contrast, the estimation error for Target 1 varies insignificantly because position estimation of Target 1 mainly relies on the Kinect vision sensor measurement, which is not sensitive to environmental noises.

5.4.2. Effect of walking speed

To study the effect of human walking speed on the estimation accuracy, we compare the experimental results obtained under the cases of person strolling at a speed of 0.4 m/s and walking at a normal speed of 0.8 m/s in average, at each of which Target 1 and Target 2 follow the same pre-selected trajectories. Due to the delay of acoustic ranging signal processing, the acoustic ranging measurement becomes spatially sparse as the person's walking speed increases. We repeat the experiment 10 times for each walking speed, and calculate the median and 90th percentile error of estimation error in the same way as described in Section 5.4 A. Fig. 13 shows the performance comparison between two walking speeds under the same moderate environmental noise level. Specifically, the median estimation error of Target 2 at the speed of 0.4 m/s and 0.8 m/s are 0.89 m and 1.03 m, respectively. The effect of walking speed on the estimation accuracy is negligible for Target 1 because the Kinect vision sensor returns measurement data fast enough, which is less sensitive to the human walking speed.

Table 1 summarizes the estimation accuracy of Target 1 and Target 2 under the effect of both walking speed and environmental noise level. We can see that the median estimation accuracy for Target 1 is between 0.43 m and 0.55 m, and ranges from 0.85 m to 1.12 m for Target 2.

5.4.3. Non-line-of-sight issue

Line of sight (LoS) between transmitter and receiver pair ensures correct measurement result for acoustic ranging. If the path between the transmitter and receiver pair is partially or completely obstructed, acoustic ranging is not able to return accurate results. In our experiments, the student room in the lab/office environment is a clustered area that may cause non-line-of-sight (NLoS) problem if the targets and robot are obstructed by the cubicle dividers. The impact of NLoS on acoustic ranging has been analyzed by adjusting the distance between two smartphones divided by cubicle divider from 0.1 m to 1 m. We observed that the returned range value is much higher than the true value and the error raises from 0.5 m to 5 m with an increasing separation distance. This may result from the attenuation of beep sound as it travels across the divider. Also, the beep sound received may have traveled through multiple paths that bounce off the walls or ceiling, which equivalently produce longer distance between the transmitter and receiver.

Table 1
Median and 90 percentile estimation error versus both environmental noise and walking speed.

	Target 1		Target 2	
	Walking speed 0.4 m/s	Walking speed 0.8 m/s	Walking speed 0.4 m/s	Walking speed 0.8 m/s
Moderate noise level (evening)	0.43 m (0.52 m)	0.54 m (0.58 m)	0.85 m (1.21 m)	0.98 m (1.72 m)
Heavy noise level (daytime)	0.51 m (0.56 m)	0.55 m (0.62 m)	0.93 m (1.42 m)	1.12 m (1.87 m)

* The values inside and outside the bracket represent median and 90th percentile errors, respectively.

Although the proposed system may return false estimation result without LoS ranging measurement, we found from the experiments that the localization algorithm is able to converge again from incorrect position estimate which is caused by erroneous acoustic ranging result, if the targets and robot regain LoS for acoustic ranging. The proposed system allows occasional interruption of NLoS and requires several update iterations to converge after LOS is regained. Also, the NLoS issue can be addressed by taking advantage of the proposed cooperative localization scheme, that is, if the original LoS ranging is lost, the ranging from other nearby person peers or robots, to which the LoS transmission is established, can be used for position estimation.

5.4.4. Multipath propagation effect

In acoustic ranging, smartphones may receive the beep signals transmitted through multiple paths in addition to line-of-sight path as the sound might be reflected by obstacles. Multipath propagation most likely occurs in narrow space, for example, the corridor and student room in the lab environment. The presence of multipath propagation basically degrades ranging accuracy. However, the acoustic ranging adopted in our system use changepoint detection method which identifies the first strong signal that deviates from noise [50], which makes the system robust to multipath propagation effect.

6. Conclusion

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 measurements from both the Kinect 3D vision sensor and smartphone-based acoustic ranging, so that the target positions can be iteratively estimated. Experiments were conducted using a Pioneer 3AT mobile robot and two LG smartphones, which showed that the positioning algorithm was able to locate and track moving human targets in different indoor environments. The median estimation accuracy ranges from 0.43 m to 1.12 m under different environmental noise levels and with different human walking speed. The localization performance is comparable to most indoor localization methods for moving targets using WiFi signature maps without the cost of deploying intensive sensing infrastructure.

Acknowledgments

References

[1] H. Liu, J. Yang, S. Sidhom, Y. Wang, Y. Chen, F. Ye, Accurate WiFi based localization for smartphones using peer assistance, IEEE Trans. Mob. Comput. 13 (10) (2014) 2199–2214.

[2] B. Zhou, Q. Li, Q. Mao, W. Tu, X. Zhang, Activity sequence-based indoor pedestrian localization using smartphones, IEEE Trans. Hum. Mach. Syst. 45 (5) (2015) 562–574.

[3] D. Macharet, D. Florencio, Learning how to increase the chance of human-robot engagement, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013, pp. 2173–2179.

[4] C. Zhu, W. Sheng, Wearable sensor-based hand gesture and daily activity recognition for robot-assisted living, IEEE Trans. Syst. Man Cybern. A 41 (3) (2011) 569–573.

[5] H. Liu, H. Darabi, P. Banerjee, J. Liu, Survey of wireless indoor positioning techniques and systems, IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev. 37 (6) (2007) 1067–1080.

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

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

[8] J.J. Leonard, H.F. Durrant-Whyte, Mobile robot localization by tracking geometric beacons, IEEE Trans. Robot. Autom. 7 (3) (1991) 376–382.

[9] P. Jensfelt, S. Kristensen, Active global localization for a mobile robot using multiple hypothesis tracking, IEEE Trans. Robot. Autom. 17 (5) (2001) 748–760.

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

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

[12] S. Thrun, D. Fox, W. Burgard, F. Dellaert, Robust Monte Carlo localization for mobile robots, Artificial Intelligence 128 (1) (2001) 99–141.

[13] M.G. Dissanayake, P. Newman, S. Clark, H.F. Durrant-Whyte, M. Csorba, A solution to the simultaneous localization and map building (SLAM) problem, IEEE Trans. Robot. Autom. 17 (3) (2001) 229–241.

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

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

[16] C. Jiang, M. Fahad, Y. Guo, J. Yang, Y. Chen, Robot-assisted human indoor localization using the kinect sensor and smartphones, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014, pp. 4083–4089.

[17] L.M. Ni, Y. Liu, Y.C. Lau, A.P. Patil, LANDMARC: Indoor location sensing using active RFID, Wirel. Netw. 10 (6) (2004) 701–710.

[18] A. Bekkali, H. Sanson, M. Matsumoto, RFID indoor positioning based on probabilistic RFID map and Kalman filtering, in: IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, 2007, pp. 21–28.

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

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

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

[22] M. Azizyan, I. Constandache, R. Roy Choudhury, Surroundsense: Mobile phone localization via ambience fingerprinting, in: Proceedings of the Annual International Conference on Mobile Computing and Networking, 2009, pp. 261–272.

[23] C. Wu, Z. Yang, Y. Liu, W. Xi, WILL: Wireless indoor localization without site survey, IEEE Trans. Parallel Distrib. Syst. 24 (4) (2013) 839–848.

[24] C. Wu, Z. Yang, Y. Liu, Smartphones based crowdsourcing for indoor localization, IEEE Trans. Mob. Comput. 14 (2) (2015) 444–457.

[25] Y. Liu, Q. Wang, J. Liu, T. Wark, MCMC-based indoor localization with a smart phone and sparse WiFi access points, in: IEEE International Conference on Pervasive Computing and Communications Workshops, 2012, pp. 247–252.

[26] K. Al Nuaimi, H. Kamel, A survey of indoor positioning systems and algorithms, in: International Conference on Innovations in Information Technology, 2011, pp. 185–190.

[27] U. Steinhoff, B. Schiele, Dead reckoning from the pocket-an experimental study, in: IEEE International Conference on Pervasive Computing and Communications, 2010, pp. 162–170.

[28] E. Foxlin, Pedestrian tracking with shoe-mounted inertial sensors, IEEE Comput. Graph. Appl. 25 (6) (2005) 38–46.

[29] O. Woodman, R. Harle, Pedestrian localisation for indoor environments, in: Proceedings of the International Conference on Ubiquitous Computing, 2008, pp. 114–123.

[30] R. Faragher, C. Sarno, M. Newman, Opportunistic radio SLAM for indoor navigation using smartphone sensors, in: IEEE/ION Position Location and Navigation Symposium, 2012, pp. 120–128.

[31] Y. Jin, W.-S. Soh, M. Motani, W.-C. Wong, A robust indoor pedestrian tracking system with sparse infrastructure support, IEEE Trans. Mob. Comput. 12 (7) (2013) 1392–1403.

[32] J. Rantakokko, J. Rydell, P. Stromback, P. Handel, J. Callmer, D. Tornqvist, F. Gustafsson, M. Gruden, Accurate and reliable soldier and first responder indoor positioning: Multisensor systems and cooperative localization, IEEE Wirel. Commun. 18 (2) (2011) 10–18.

[33] J.-O. Nilsson, D. Zachariah, I. Skog, P. Handel, Cooperative localization by dual foot-mounted inertial sensors and inter-agent ranging, EURASIP J. Adv. Signal Process. 2013 (1) (2013) 164–180. http://dx.doi.org/10.1186/1687-6180-2013-164.

[34] J.-O. Nilsson, P. Handel, Recursive Bayesian initialization of localization based on ranging and dead reckoning, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013, pp. 1399–1404.

[35] T. Sathyan, M. Hedley, Fast and accurate cooperative tracking in wireless networks, IEEE Trans. Mob. Comput. 12 (9) (2013) 1801–1813.

[36] I. Leite, C. Martinho, A. Paiva, Social robots for long-term interaction: A survey, Int. J. Soc. Robot. (2013) 1–18.

[37] L. Wu, J. Lu, T. Zhang, J. Gong, Robot-assisted intelligent emergency system for individual elderly independent living, in: IEEE Global Humanitarian Technology Conference, 2016, pp. 628–633.

[38] C. Qiu, M.W. Mutka, AirLoc: Mobile robots assisted indoor localization, in: IEEE International Conference on Mobile Ad Hoc and Sensor Systems, 2015, pp. 407–415.

[39] R. Siegwart, I.R. Nourbakhsh, D. Scaramuzza, Introduction to Autonomous Mobile Robots, second ed., The MIT Press, 2011.

[40] D. Fox, Adapting the sample size in particle filters through KLD-sampling, Int. J. Robot. Res. 22 (12) (2003) 985–1003.

[41] S. Thrun, W. Burgard, D. Fox, Probabilistic Robotics, MIT press, 2005.

[42] J. Santos, Turtlebot follower tutorials. [Online] http://ros.org/wiki/turtlebot_follower/Tutorials/Demo. (Accessed 5-Feb-2014).

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

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

[45] B. Xu, G. Sun, R. Yu, Z. Yang, High-accuracy tdoa-based localization without time synchronization, IEEE Trans. Parallel Distrib. Syst. 24 (8) (2013) 1567–1576.

[46] P. Misra, S.S. Kanhere, S. Jha, W. Hu, Sparse representation based acoustic rangefinders: From sensor platforms to mobile devices, IEEE Commun. Mag. 53 (1) (2015) 249–257.

[47] S.-M. Moosavi-Dezfooli, Y.-A. Pignolet, D. Dzung, Simultaneous acoustic localization of multiple smartphones with Euclidean distance matrices, 2015. arXiv preprint arXiv:1512.05103.

[48] E. Brookner, Tracking and Kalman Filtering Made Easy, Wiley Online Library, 1998.

[49] A.I. Mourikis, S.I. Roumeliotis, Performance analysis of multirobot cooperative localization, IEEE Trans. Robot. 22 (4) (2006) 666–681.

[50] J. Yang, S. Sidhom, G. Chandrasekaran, T. Vu, H. Liu, N. Cecan, Y. Chen, M. Gruteser, R.P. Martin, Sensing driver phone use with acoustic ranging through car speakers, IEEE Trans. Mob. Comput. 11 (9) (2012) 1426–1440.

Chao Jiang received the B.E. degree in measuring and control technology and instrumentation from Chongqing University, Chongqing, China. He is currently working toward the Ph.D. degree in electrical engineering at Stevens Institute of Technology, Hoboken, NJ, USA. His research interests include cooperative distributed localization, human-aware robot navigation, human–robot interaction, and learning-based optimal control.

Muhammad Fahad received his undergraduate degree in electrical engineering from University of Engineering and Technology, Lahore, Pakistan. He received his Masters degree in electrical engineering from Stevens Institute of Technology, Hoboken, NJ. He is currently a Ph.D. candidate for the Robotics and Automation Laboratory, at Stevens Institute of Technology, Hoboken, NJ. His research interests include cooperative distributed localization, multivariable autonomous systems, and autonomous mobile robot teams.

Yi Guo received her B.S. and M.S. from Xi'an University of Technology, Xi'an, China, in 1992 and 1995, respectively, and Ph.D. from the University of Sydney, Australia, in 1999, all in electrical engineering. She was a postdoctoral research fellow at Oak Ridge National Laboratory from 2000 to 2002, and a visiting assistant professor at the University of Central Florida, Orlando, FL, USA, from 2002 to 2005. She joined the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, in 2005, where she is currently a Professor. Her main research interests include autonomous mobile robotics, distributed sensor networks, and nonlinear control systems. She has published more than 100 peer-reviewed journal and conference papers, authored one book (Wiley 2017) and edited one book (Springer 2013). She currently serves on the editorial boards of IEEE Robotics and Automation Magazine, IEEE Access, and IEEE/ASME Transactions on Mechatronics. She has been involved in various conference programs and organizing committees including serving as the Finance Chair of IEEE International Conference on Robotics and Automation (ICRA) in 2015 and the Local Arrangement Co-Chair of ICRA in 2006.

Yingying (Jennifer) Chen is a Professor in the Department of Electrical and Computer Engineering at Stevens Institute of Technology. Her research interests include cyber security and privacy, Internet of Things, smart healthcare and mobile computing and sensing. She has published over 100 journals and referred conference papers in these areas. She received her Ph.D. degree in Computer Science from Rutgers University. Prior to joining Stevens, she was with Alcatel-Lucent. She is the recipient of the NSF CAREER Award and Google Faculty Research Award. She also received NJ Inventors Hall of Fame Innovator Award. She is the recipient of the Best Paper Awards from ACM AsiaCCCS 2016, IEEE CNS 2014 and ACM MobiCom 2011. She also received the IEEE Outstanding Contribution Award from IEEE New Jersey Coast Section each year 2005–2009. Her research has been reported in numerous media outlets including MIT Technology Review, Fox News Channel, Wall Street Journal, and National Public Radio. She serves on the editorial boards of IEEE Transactions on Mobile Computing (IEEE TMC), IEEE Transactions on Wireless Communications (IEEE TWireless), and IEEE Network Magazine.