Autonomous Localization and Mapping Using a Single Mobile Device

Tiexing Wang, Fangrong Peng and Biao Chen

***Abstract*—This paper considers the problem of simultaneous 2-D room shape reconstruction and self-localization without the requirement of any pre-established infrastructure. A mobile device equipped with co-located microphone and loudspeaker as well as internal motion sensors is used to emit acoustic pulses and collect echoes reflected by the walls. Using only first order echoes, room shape recovery and self-localization is feasible when auxiliary information is obtained using motion sensors. In particular, it is established that using echoes collected at three measurement locations and the two distances between consecutive measurement points, unique localization and mapping can be achieved provided that the three measurement points are not collinear. Practical algorithms for room shape reconstruction and self-localization in the presence of noise and higher order echoes are proposed along with experimental results to demonstrate the effectiveness of the proposed approach.**

***Index Terms*—2-D room shape recovery, self-localization, acoustic sensor, room impulse response, self-localization.**

I. INTRODUCTION

Indoor localization has become more important in recent years as numerous applications, e.g., public safety or location based services, rely on accurate indoor localization [1]. As GPS signals are severely attenuated in typical indoor environment, a number of alternative technologies have been proposed for indoor localization, e.g. those using WiFi [2]–[4], UWB signal [5]–[7], LED light [8], [9] , or some combination of the above.

These technologies inevitably require indoor geometry information. There are applications where the indoor room geometry may need to be acquired concurrently with localization. This is generally referred to as simultaneous localization and mapping (SLAM). We comment that the so-called WiFi-SLAM still requires indoor mapping information; SLAM refers to the training process that associates mapping information with the WiFi signature [10]. There are also applications where mapping itself is the ultimate goal instead of self-localization [11], [12].

For many applications where room shape reconstruction is required, acoustic based approach is arguably more suitable as rooms are often defined by dominant sound reflectors (walls). The distance measurements as measured through acoustic echoes contain rich information about the location of the measurement points as well as the room geometry. A key advantage of the acoustic based approach is that no pre-established infrastructure is needed; this is in sharp contrast with other approaches which inevitably require either deployment of anchor nodes [13], [14] or the availability of ambient WiFi signals as well as preliminary maps [10]. This unique advantage has the potential to broaden the applications of indoor mapping and localization to systems where current technologies are either unsuitable or too expensive to implement.

The most prevalent acoustic based approach is to employ a single fixed loudspeaker and a microphone array, or equivalently, a fixed loudspeaker and a mobile microphone [15]–[20]. It was shown that both the room shape and the geometry of the microphone array (or the trajectory of the mobile microphone) can be estimated by first order echoes [21]. Furthermore, bearing only SLAM can be achieved using a mobile microphone array [22].

The fact that a microphone array needs to be deployed leaves much to be desired: fully autonomous SLAM should require minimum deployment effort. Ideally, a single mobile device that moves around would autonomously reconstruct the room shape while tracking its own movement within the recovered room geometry. Indeed, room shape recovery using a single acoustic device has been addressed in the literature. It was established that any convex polygon can be reconstructed by the *entire* set of both first and second order echoes collected using a fixed device with a collocated microphone and loudspeaker [18]. However, experimental results, including that of our own, demonstrated that higher order acoustic echoes are often difficult to recover, thus the requirement of having the entire set of second order echoes makes such an approach impractical.

On the other hand, given only *grouped* first order echoes, SLAM can be achieved for a large class of convex polygon other than parallelograms [23]. This result was strengthened in [24] where it was established that parallelograms are the only convex polygons that are not recoverable via grouped first order echoes. Here "*-grouped*" means correct labeling, i.e., the correspondence between collected echoes and walls is known.

This paper makes further progress in overcoming the shortcomings of the approaches in [23], [24]. The reconstruction will again be based on first order echoes only but without the knowledge of echo labeling. To overcome the ambiguity associated with parallelograms, our approach leverages the ever expanding capability of various motion sensors embedded in latest smart phones, including accelerometer, magnetometer, and gyroscope. Those sensors are capable of measuring distance and direction information of a moving device [25]–[27]. However, existing results indicate that while distance measures have reasonable accuracy, direction measurement is often subject to large measurement error [28]. Thus our current approach only exploits the distance measurements and the key question to be addressed is how much additional information

will be needed for acoustic SLAM to be able to recover all convex polygons.

The major contribution of the paper is to establish that with three non-collinear measurement points, SLAM can be achieved for all convex polygons using *ungrouped* first order echoes provided that the distances between consecutive measurement points are known. Note that this additional information is much weaker than the knowledge of the complete geometry of the measurement - this is tantamount to knowing only two sides of a triangle which is inadequate to construct the triangle. An added advantage of this additional distance information is that it removes the need for grouped echoes, making the scheme much more widely applicable as it can accommodate a great deal of freedom in the movement of the device. Preliminary results have been reported in [29]. The present work, in addition to expanding on technical details, contains several new results including a more detailed analysis on exactly what is the minimum amount of distance information that is needed for SLAM. Specifically, it is further established that with ungrouped echoes, a single distance measure does not suffice for parallelograms. Note the subtle but important difference with that of [23], [24] in which grouped instead of ungrouped echoes are assumed.

The rest of the paper is organized as follows. Section II introduces the indoor propagation model of acoustic signals, image source model and existing results on 2-D with a single device. Theoretical guarantee of successful SLAM given distances between consecutive measurement points is provided in Section III along with a practical algorithm that handles the presence of measurement noise and higher order/spurious peaks. Experiment results are provided in Section IV followed by conclusion in Section V.

II. PROBLEM STATEMENT

A. Room Impulse Response Model

Acoustic signal propagation from a loudspeaker to a microphone in a room can be described by the room impulse response (RIR), which includes both line-of-sight (LOS) and reflected components. If the microphone and loudspeaker are much closer to each other compared to the distance between the device and the walls, we say it is a co-located device. For a co-located device at the jth measurement point denoted by O_j, the RIR is, ignoring dispersion,

$$h^{(j)}(t) = \sum_i \alpha_i^{(j)} \delta(t - \tau_i^{(j)}), \tag{1}$$

where $\alpha_i^{(j)}$'s and $\tau_i^{(j)}$'s are path gains and delays from the transmitter to the receiver, respectively. Since higher order reflective paths typically have much weaker power, $h^{(j)}(t)$ can be approximated by the first N_j+1 components including LOS and N_j reflective paths:

$$h^{(j)}(t) \approx \sum_{i=0}^{N_j} \alpha_i^{(j)} \delta(t - \tau_i^{(j)}),$$

where we assume that the N_j reflective paths contain all first order reflections and higher order ones that are detectable. Notice that for an arbitrary convex polygon, not every measurement point has first order echoes to all the walls. We refer to those measurement points can receive all first order echoes as *feasible* measurement points.

Denote by $s(t)$ the emitted signal at the speaker. Then the received signal at the microphone for the jth measurement point is

$$r^{(j)}(t) = s(t) * h^{(j)}(t) + \omega(t), \tag{2}$$

where $*$ denotes linear convolution and $\omega(t)$ is the additive noise. Ideally, the delays can be recovered from the received signal $r^{(j)}(t)$ if $s(t)$ behaves like a Dirac delta function [17]. However, this requires a wideband acoustic signal along with a wideband acoustic channel, including that of the microphone receiver. A more practical alternative is to emit $s(t)$ with a desired auto-correlation function that is *peaky* and then implement a correlator at the microphone:

$$m^{(j)}(t) = r^{(j)}(t) * s(t). \tag{3}$$

Thus, the first and dominant peak of $m^{(j)}(t)$ corresponds to the LOS components, while the remaining peaks correspond to reflective components. The time difference of arrival (TDOA) in reference to the LOS component can be used for estimating the delays of different reflective paths. A simple peak-detection method will be introduced in Section V.A, where the chirp signal is used for $s(t)$ because of its nice auto-correlation property.

Define a column vector

$$\tilde{\mathbf{r}}_j = \left\{ \frac{(\tau_i^{(j)} - \tau_0^{(j)})c}{2} \right\}_{i=1}^{N_j}, \tag{4}$$

where c is the speed of sound and $\tau_i^{(j)}$ is the arrival time of the ith path with $\tau_0^{(j)}$ corresponding to the LOS component. Then $\tilde{\mathbf{r}}_j$ contains all the distances between the device and the walls, along with some higher order terms.

B. Image Source Model

With the image source model [15], reflections within a constrained space can be viewed as free space LOS propagations from virtual sources to the receiver. Let the coordinate of O_j be denoted by $\mathbf{o}_j$. As show in Fig. 1, the first order image source of O_j with respect to the ith wall is

$$\tilde{\mathbf{o}}_{j,i} = 2\langle \mathbf{p}_i - \mathbf{o}_j, \mathbf{n}_i \rangle \mathbf{n}_i + \mathbf{o}_j,$$

where $\mathbf{p}_i$ is any point on the ith wall, $\mathbf{n}_i$ is the outward norm vector of the ith wall and $\langle \mathbf{x}, \mathbf{y} \rangle$ denotes the inner product between $\mathbf{x}$ and $\mathbf{y}$. Let $r_{j,i}$ be the distance between O_j and the ith wall, then

$$r_{j,i} = \frac{1}{2} ||\tilde{\mathbf{o}}_{j,i} - \mathbf{o}_j||_2. \tag{5}$$

Moreover, the second order image source of O_j with respect to the ith and the kth wall is

$$\tilde{\mathbf{o}}_{j,ik} = 2\langle \mathbf{p}_k - \tilde{\mathbf{o}}_{j,i}, \mathbf{n}_k \rangle \mathbf{n}_k + \tilde{\mathbf{o}}_{j,i}.$$

Similarly, we denote by $r_{j,ik}$ the half distance between $\mathbf{o}_j$ and $\tilde{\mathbf{o}}_{j,ik}$. Following similar steps, higher order image sources can be represented by lower order image sources. Then all

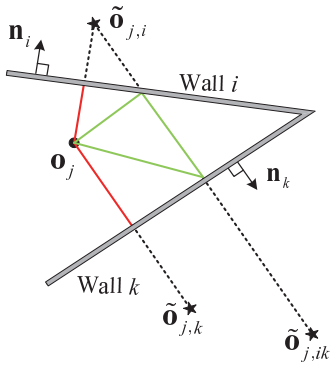


Fig. 1: The image source model: $\tilde{\mathbf{o}}_{j,i}$ and $\tilde{\mathbf{o}}_{j,k}$ are first-order image sources with respect to the ith and kth wall and $\tilde{\mathbf{o}}_{j,ik}$ is the second-order image source with respect to the ith and kth wall in the stated order.

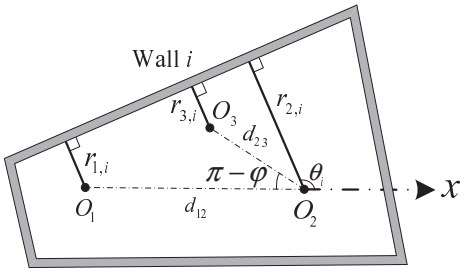


Fig. 2: A mobile device is employed to achieve SLAM. The mobile device emits signal and collects echoes at O_1, O_2 and O_3 successively. The distances between the consecutive measurement points are d_{12} and d_{23}.

the elements of $\tilde{\mathbf{r}}_j$ can be represented by the real source and image sources. For the rest of the paper, the term *echo* is used to refer to either the delay $\tau_i^{(j)}$ or the corresponding elements of $\tilde{\mathbf{r}}_j$ if no ambiguity occurs.

C. Two Extreme Cases

The most benign case is when the location of the measurement points are known, or equivalently, the distance between pairwise measurement points are given [30]. In this case, only room shape reconstruction is of interest and the problem becomes trivial, at least in the noiseless case. It amounts to finding common tangent lines of circles centered at three non-collinear measurement points.

The other extreme is when the reconstruction is free of any geometry information of the measurement points. In this case, both room shape and self-localization are of interest. This was first investigated in [23] where it was established that a large class of convex polygons can be reconstructed by *grouped* first order echoes and, subsequently, the coordinates of measurement points can be also estimated. An important exception is parallelograms and it was shown in [23] that unique reconstruction of parallelograms is impossible using first-order echoes alone. The result was later strengthened in [24] where it was proved that all convex polygons except parallelogram can be reconstructed subject to the usual rotation and reflection ambiguities.

III. SLAM WITH KNOWN PATH LENGTHS

A. SLAM with Two Path Lengths

Consider a convex planar K-polygon. As shown in Fig.2, a mobile device with co-located microphone and loudspeaker emits pulses and receives echoes at $\{O_j\}_{j=1}^3$. Without loss of generality, we assume that O_1 is the origin, O_2 lies on the x-axis, and O_3 lies above the x-axis. Let $\varphi = (\pi - \angle O_1O_2O_3) \in (0, \pi)$ and the lengths of O_1O_2 and O_2O_3 be denoted by d_{12} and d_{23}, respectively.[1] Suppose the mobile device is capable of measuring its path length when moving from one place to another, i.e. d_{12} and d_{23} are known. Our goal is to simultaneously determine the room shape and the coordinate of O_3 using first-order echoes.

From Fig. 2, it is straightforward to show that

$$(r_{2,i} - r_{1,i}) + d_{12}\cos\theta_i = 0, \tag{6}$$

$$d_{23}\cos(\theta_i - \varphi) + (r_{3,i} - r_{2,i}) = 0. \tag{7}$$

B. Ideal Case

Let $\mathbf{r}_j = \{r_{j,i}\}_{i=1}^K$ be a column vector with its entries defined in (5). We assume for now that, the one-to-one mapping $f_j : \tilde{\mathbf{r}}_j \mapsto \mathbf{r}_j$ is known for all j's. In other words, $r_{j,i}$'s have been correctly chosen from $\tilde{\mathbf{r}}_j$ for $j = 1, 2, 3$ and $i = 1, \ldots, K$. For the rest of the paper, we say that the received echoes are *grouped* if echoes are correctly labeled. The remaining problem is to determine the uniqueness of θ_i's and φ given (6) and (7).

Define $\alpha_{ii'} = -\frac{r_{2,i} - r_{1,i'}}{d_{12}}$ and $\beta_{ii'} = -\frac{r_{3,i'} - r_{2,i}}{d_{23}}$. For simplicity we denote α_{ii} and β_{ii} by α_i and β_i, respectively. Given grouped echoes and Eqs. (6) and (7), we have

$$\theta_i = \pm\arccos\alpha_i \quad \text{and} \quad \theta_i - \varphi = \pm\arccos\beta_i, \tag{8}$$

There are four possible sign combinations for a given i,

$$\theta_i = \arccos\alpha_i \quad \text{and} \quad \theta_i - \varphi = \arccos\beta_i \tag{9}$$

$$\theta_i = \arccos\alpha_i \quad \text{and} \quad \theta_i - \varphi = -\arccos\beta_i \tag{10}$$

$$\theta_i = -\arccos\alpha_i \quad \text{and} \quad \theta_i - \varphi = \arccos\beta_i \tag{11}$$

$$\theta_i = -\arccos\alpha_i \quad \text{and} \quad \theta_i - \varphi = -\arccos\beta_i. \tag{12}$$

Lemma III.1. *Suppose O_j $(j = 1, 2, 3)$ are feasible and not collinear. Given grouped first order echoes, with probability* 1*, there exist exactly two sign combinations such that* (6) *and* (7) *hold simultaneously for all i if φ and the direction of both $\overrightarrow{O_1O_2}$ and $\overrightarrow{O_2O_3}$ are randomly chosen. The two possible sign combinations have opposite signs for φ and all θ_i's and correspond to reflection of each other in terms of recovered room shapes.*

[1] If $\varphi \in (0, 2\pi)$, i.e. we do not have control of where to place O_3, then the reconstruction is subject to reflection ambiguity (c.f. Theorem III.3).

Proof: Assume without loss of generality that the ground truth of the polygon is (9) for all $i \in \{1, \ldots, K\}$. Note that (9) implies that (12) holds for $\theta_i' = -\theta_i$ and $\varphi' = -\varphi < 0$ for all i, i.e., they correspond to reflections of each other.

Suppose multiple sign combinations hold for a wall. Without loss of generality, let $i = 1$. From (9) we have

$$\varphi = \arccos\alpha_1 - \arccos\beta_1. \tag{13}$$

Assume that one of the following equations also holds,

$$\varphi = -\arccos\alpha_1 - \arccos\beta_1, \tag{14}$$

$$\varphi = \arccos\alpha_1 + \arccos\beta_1, \tag{15}$$

$$\varphi = -\arccos\alpha_1 + \arccos\beta_1. \tag{16}$$

Then we have the following three cases

1) If (13) and (14) hold, we must have $\theta_1 = 0$ which implies that O_1O_2 is perpendicular to the first wall, and $\varphi = -\arccos\beta_1$.
2) If (13) and (15) hold, we must have $\arccos\beta_1 = 0$, which implies that O_2O_3 is perpendicular to the first wall.
3) If (13) and (16) hold, we must have $\varphi = 0$, which contradicts the assumption that O_1, O_2 and O_3 are not collinear.

Given that the three measurement points are randomly chosen, and, subsequently, φ, $\overrightarrow{O_1O_2}$ and $\overrightarrow{O_2O_3}$ are random, the first two cases do not occur with probability one.

If a subset of (10)-(12) holds for i and i' simultaneously, then we must have $(\theta_i, \theta_{i'}) \in \{\theta_i = 0, \theta_i = \varphi, \varphi = 0\} \times \{\theta_{i'} = 0, \theta_{i'} = \varphi, \varphi = 0\}$, which again, do not occur due to randomly chosen measurement points. Similarly, it can be shown that for more than 2 walls, (9) would imply none of (10)-(12) holds for all walls. ∎

C. Echo Labeling

Since echoes may arrive in different orders at different O_j's and $\tilde{\mathbf{r}}_j$ contains higher order echoes if $N_j > K$, f_j is usually unknown. We say the received echoes are *ungrouped* if f_j is unknown for some j. Thus given $\tilde{\mathbf{r}}_j$, our task is to first determine the mapping f_j, i.e., label the echoes, followed by estimation of θ_i's and φ.

Lemma III.2. *With ungrouped echoes, any mapping f_j' that differs from the correct mapping f_j will result, with probability* 1*, the following two possible cases*

1) *there exists no solution to* (6) *and* (7) *given no parallel edges, or*
2) *the reconstructed room shape has larger dimension with respect to parallel edges.*

Proof: We illustrate the proof by considering the case $K = 4$. The result can be easily extended to $K = 3$ and $K > 4$.

Suppose again that the ground truth is (9) for all i. We first consider parallelograms and exclude odd higher order echoes resulting from a pair of parallel walls. The distances between O_j $(j = 1, 2, 3)$ and the four walls satisfy

$$r_{1,1} + r_{1,2} = r_{2,1} + r_{2,2} = r_{3,1} + r_{3,2} = a, \tag{17}$$

$$r_{1,3} + r_{1,4} = r_{2,3} + r_{2,4} = r_{3,3} + r_{3,4} = b. \tag{18}$$

One can see that for some f_j''s, pairs of $\{\alpha_{ii'}, \beta_{ii'}\}$ $(i, i' \in \{1, 2, 3, 4\})$ are related to each other. Consider for example the f_j''s resulting in $\{\alpha_{12}, \alpha_{21}, \alpha_{34}, \alpha_{43}\}$ and $\{\beta_{12}, \beta_{21}, \beta_{34}, \beta_{43}\}$. Since $\alpha_{12} + \alpha_{21} = 0$, $\alpha_{34} + \alpha_{43} = 0$, $\beta_{12} + \beta_{21} = 0$ and $\beta_{34} + \beta_{43} = 0$, we have

$$\arccos(\alpha_{21}) = \pi \pm \arccos(\alpha_{12}),$$

$$\arccos(\alpha_{43}) = \pi \pm \arccos(\alpha_{34}),$$

$$\arccos(\beta_{21}) = \pi \pm \arccos(\beta_{12}),$$

$$\arccos(\beta_{43}) = \pi \pm \arccos(\beta_{34}).$$

Thus (8) reduces to two equations

$$\varphi = \pm\arccos(\alpha_{12}) \pm \arccos(\beta_{12}),$$

$$\varphi = \pm\arccos(\alpha_{34}) \pm \arccos(\beta_{34}).$$

With probability 1, these two equations do not hold simultaneously as α_{12}, β_{12} are independent of α_{34}, β_{34} due to randomly chosen measurement points. Other $f_j'(\neq f_j)$'s always have at least two equations with independent choice of α and β. Hence no solution can be found for those instances.

Suppose f_j''s are chosen such that we have $\alpha_{ii'}$ and $\beta_{ii''}$ $(i \neq i',\ i \neq i'')$. For rooms with no more than one pair of parallel walls, only echoes chosen according to f_j's satisfy (9) for all i. This is because for those rooms, at least one of (17) and (18) does not hold. Thus some $\alpha_{ii'}$'s and $\beta_{ii''}$'s are not related since $r_{1i'}$, r_{2i} and $r_{3i''}$ are randomly chosen from $\tilde{\mathbf{r}}_1$, $\tilde{\mathbf{r}}_2$ and $\tilde{\mathbf{r}}_3$, respectively.

Given parallel edges, however, higher order echoes may also satisfy (6) and (7). For instance, as shown in Fig. 3, suppose that walls 1 and 3 are parallel. Then it is easy to verify that

$$r_{j,131} - r_{j',131} = r_{j,1} - r_{j',1},$$

$$r_{j,313} - r_{j',313} = r_{j,3} - r_{j',3},$$

where $j \neq j'$. Hence, (6) and (7) provide the same $\cos\theta_1$, $\cos\theta_3$, $\cos(\theta_1 - \varphi)$ and $\cos(\theta_3 - \varphi)$ if $r_{j,1}$ and $r_{j,3}$ are replaced by $r_{j,131}$ and $r_{j,313}$, respectively. By Lemma III.1, the third order echoes resulting from a pair of parallel edges lead to a larger room with the same norm vectors. Exactly the same argument applies to odd higher order echoes from a pair of parallel edges. Therefore, Lemma III.2 is proved. ∎

Remark 1: The ambiguities resulting from parallel edges can be easily eliminated if we always choose SLAM result with the smallest room size.

Given Lemma III.1 and Lemma III.2, we have the following result on the identifiability of any convex polygonal room by using only first order echoes.

Theorem III.3. *With probability* 1*, SLAM can be achieved subject to reflection ambiguity given any convex planar K-polygon, by using the first order echoes received at three random points in the feasible region, with known d_{12} and d_{23} and unknown $\varphi \in (0, 2\pi)$.*

Remark 2: Both the room shape and the coordinate of O_3 are subject to reflection ambiguity for $\varphi \in (0, 2\pi)$. If, however, we can limit $\varphi \in (0, \pi)$, SLAM will be free of such ambiguity.

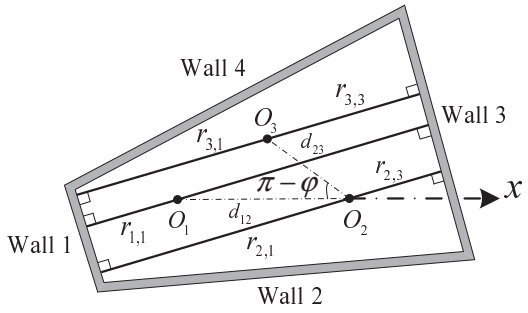


Fig. 3: A room with a pair of parallel edges. Here wall 1 and 3 are parallel.

Remark 3: In reality, it is inevitable to collect reflections from the ceiling and the floor. However, by Theorem III.3, if distances corresponding to these echoes are included, no polygon can be recovered provided that the trajectory of the device lies in a plane that is perpendicular to the walls.

D. A Practical Algorithm

In a real acoustic system, $m^{(j)}(t)$'s in (3) are inevitably corrupted by measurement noise leading to corrupted measurement of $\tilde{\mathbf{r}}_j$. Let the corrupted version of $\tilde{\mathbf{r}}_j$ be denoted by $\hat{\mathbf{r}}_j$. Two issues arise. First, given f_j's, φ obtained by (8) for different i's are not necessarily identical. The second issue is the possibility that the computed cosine values in (6) may have absolute value exceeding 1. For the former, we propose a heuristic scheme of choosing the echo and sign combination that yield the smallest variance of the estimated φ's across different i's. Notice that in the noiseless case with perfect echo measurements, the variance of the estimated φ's across different i's is 0 if the correct echo and sign combination is selected while all others will have non-zero (potentially large variance). For the latter, define a feasible $\cos\theta_i$ as

$$\cos\theta_i = \begin{cases} 1, & \text{if } 1 \le -\frac{\hat{r}_{2,i}-\hat{r}_{1,i}}{d_{12}} < 1+\epsilon \\ -\frac{\hat{r}_{2,i}-\hat{r}_{1,i}}{d_{12}}, & \text{if } -1 < -\frac{\hat{r}_{2,i}-\hat{r}_{1,i}}{d_{12}} < 1 \\ -1, & \text{if } -1-\epsilon < -\frac{\hat{r}_{2,i}-\hat{r}_{1,i}}{d_{12}} \le -1 \end{cases},$$

where $\epsilon > 0$ is a tuning parameter determined by the noise level. Feasible $\cos(\theta_i - \varphi)$ can be similarly defined. The echo combination is said to be infeasible if either $|\frac{\hat{r}_{2,i}-\hat{r}_{1,i}}{d_{12}}| > 1+\epsilon$ or $|\frac{\hat{r}_{3,i}-\hat{r}_{2,i}}{d_{23}}| > 1+\epsilon$. Only those feasible θ_i's and φ will be used in computing the variance of the estimated φ.

As the number of walls for the room is not known in prior, the proposed algorithm needs to first reconstruct some room shapes with $K = 3, \ldots, N$ walls. Then the desired room shape is the feasible one with the largest number of walls. In order to reconstruct a room shape with K walls, the number of echo combinations that need to be exhausted is

$$\binom{N_1}{K}\binom{N_2}{K}\binom{N_3}{K}(K!)^2.$$

For simplicity assume that $N = N_1 = N_2 = N_3$. Let V_{th} be the threshold of the variance. The corresponding algorithm is summarized as Algorithm 1.

Algorithm 1 Reconstruct convex polygon given distances between consecutive measurement points

Set $K = 3$ and V_{th}.
if $K \le N$ **then**
  Set $V_K = \inf$ and the stored polygon with K walls be empty.
  for $n = 1 : \left(\binom{N}{K}\right)^3 (K!)^2$ **do**
    Based on the nth echo combination, choose K elements from $\hat{\mathbf{r}}_1$, $\hat{\mathbf{r}}_2$, $\hat{\mathbf{r}}_3$, respectively.
    Compute $\cos\theta_i$'s and $\cos(\theta_i - \varphi)$ for $i = 1, \ldots, K$.
    if $\cos\theta_i$'s and $\cos(\theta_i - \varphi)$ are feasible **then**
      Compute Var$[\varphi]$ for different sign combinations and keep the one with the smallest Var$[\varphi]$.
      if Var$[\varphi] < V_K$ and the room shape does not fully cover the stored one with K walls **then**
        Keep the echo and sign combination and set $V_K = $ Var$[\varphi]$ for K.
      end if
    end if
  end for
  $K = K + 1$.
else
  Keep the SLAM results the largest K such that $V_K < V_{th}$.
end if

E. SLAM with One Path Length

Now that we have established that two distances between three consecutive measurement points are sufficient to overcome the drawback of using first order echoes alone, a natural question is what would be the least amount of information that is required to achieve SLAM for any convex polygons. Specifically we examine the case where only one distance between a pair of measurement points is known. We show that for a parallelogram, there exist multiple rooms satisfying (6) and (19) in this case, thus the answer is negative, i.e. a single distance measurement is insufficient for SLAM with ungrouped first order echoes.

Without loss of generality, assume d_{12} is known but d_{23} is not. As shown in Fig. 4, let O_1 be the origin, O_2 be on the x-axis and $O_3(x_3, y_3)$ $(y_3 \neq 0)$ is unknown. We also assume that the direction of $\overrightarrow{O_1O_2}$ with respect to the desired room is unknown. By geometry, we have (6) and

$$(r_{3,i} - r_{1,i}) + x_3\cos\theta_i + y_3\sin\theta_i = 0. \tag{19}$$

Eq. (19) can also be rewritten in a matrix form

$$\mathbf{A}[x_3, y_3]^T = \mathbf{b}, \tag{20}$$

where

$$\mathbf{A} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ \vdots & \vdots \\ \cos\theta_K & \sin\theta_K \end{bmatrix},$$

and

$$\mathbf{b} = [-(r_{3,1} - r_{1,1}), \ldots, -(r_{3,K} - r_{1,K})]^T.$$

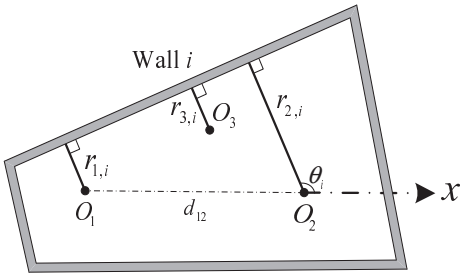


Fig. 4: A mobile device is employed to measure the geometry of a room. The mobile device collects echoes at O_1, O_2 and O_3 successively. Only the distances between O_1 and O_2 (d_{12}) is known.

The ground truth of a parallelogram is assumed to be

$$\mathbf{A} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ \cos\theta_2 & \sin\theta_2 \\ \cos\theta_3 & \sin\theta_3 \\ \cos\theta_4 & \sin\theta_4 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} -(r_{3,1} - r_{1,1}) \\ -(r_{3,2} - r_{1,2}) \\ -(r_{3,3} - r_{1,3}) \\ -(r_{3,4} - r_{1,4}) \end{bmatrix},$$

where

$$r_{1,1} + r_{1,3} = r_{2,1} + r_{2,3} = r_{3,1} + r_{3,3},$$
$$r_{1,2} + r_{1,4} = r_{2,2} + r_{2,4} = r_{3,2} + r_{3,4}.$$

Let

$$\mathbf{A}' = \begin{bmatrix} \cos\theta_{13} & \sin\theta_{13} \\ \cos\theta_{24} & \sin\theta_{24} \\ \cos\theta_{31} & \sin\theta_{31} \\ \cos\theta_{42} & \sin\theta_{42} \end{bmatrix} \mathbf{b}' = \begin{bmatrix} -(r_{3,1} - r_{1,3}) \\ -(r_{3,2} - r_{1,4}) \\ -(r_{3,3} - r_{1,1}) \\ -(r_{3,4} - r_{1,2}) \end{bmatrix}.$$

Then

$$\cos\theta_{13} + \cos\theta_{31} = 0 \quad \text{and} \quad \cos\theta_{24} + \cos\theta_{42} = 0.$$

Moreover, since $\sin\theta = \pm\sqrt{1 - \cos^2\theta}$,

$$\sin\theta_{13} + \sin\theta_{31} = 0 \quad \text{and} \quad \sin\theta_{24} + \sin\theta_{42} = 0$$

can hold if we manipulate the sign of square root properly.

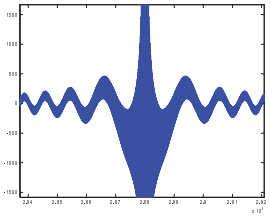
Then $\text{rank}(\mathbf{A}') = \text{rank}([\mathbf{A}', \mathbf{b}']) = 2$. Thus a room shape and the coordinate of O_3 different from the ground truth and its reflection also satisfy both (6) and (19).

IV. Experimental results

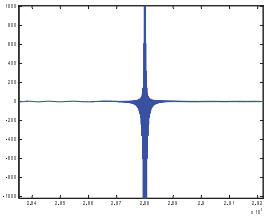
A. Experiment Setup

We describe in the following some preliminary experimental results. Enormous challenges exist to conduct a truly autonomous SLAM. Chief among them are: the search space (number of combinations) is extremely large - using for example, some modest numbers, e.g. $K = 4$ and $N_1 = N_2 = N_3 = 8$, the number of echo combinations exceeds 10^7, combining with the sign combinations the search space is in the billions; the measurement of motion sensors is still subject to large errors and some robustification of the reconstruction algorithm will need to be investigated if the true motion sensor measurements are used. The purpose of the experimental design is thus to demonstrate the feasibility of the proposed scheme in an idealized situation with a certain degree of human intervention to alleviate the above challenges.

We use a laptop as a microphone and a HTC M8 phone as our loudspeaker. As the loudspeaker of the cell phone is not omnidirectional and is power limited, we place the speaker of the cell phone towards each wall to ensure the corresponding first order echo is strong enough. Note that the microphone will record both first order echoes and some higher order ones. A chirp signal linearly sweeping from 30Hz to 8kHz is emitted by the cell phone. The sample rate at the receiver is $f_s = 96$kHz. It has been shown in [31], [32] that if the input chirp signal is correlated with its windowed version, the output may resemble a delta function, which is desirable for better delay resolution. Our simulation indicates that correlating the



(a) Transmitted signal convolves with itself



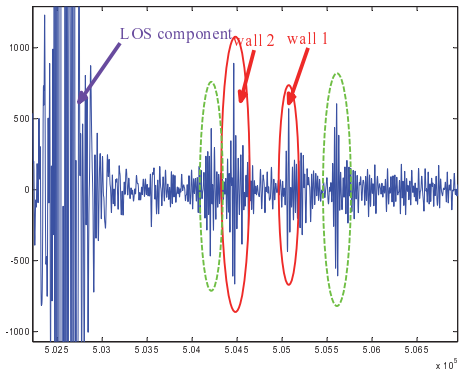
(b) Transmitted signal convolves with its windowed version

Fig. 5: Comparison of convolution result. The maximum values of the two convolution result are set to be identical.

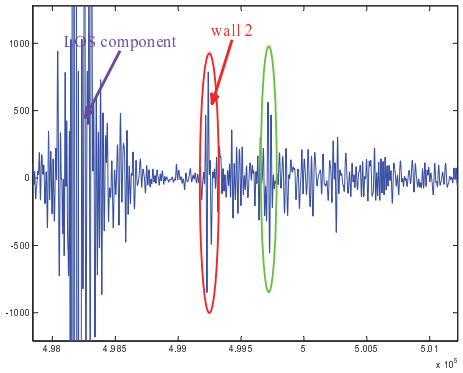
received signals with its triangularly windowed version outperforms the correlator using the original one. The comparison is shown in Fig. 5.

Fig. 6 is a sample path of the correlator output collected in the room where this experiment is conducted. In Fig. 6, peaks marked with red ellipse are desired while those with green ellipse correspond to noise, the ceiling, the floor, higher order echoes or other spurious sources. In our experiment, we use $|m^{(j)}(t)|$ rather than $m^{(j)}(t)$ since the true peaks may be either positive or negative. Local maxima of $|m^{(j)}(t)|$ corresponding to Fig. 6 are shown in Fig. 7.

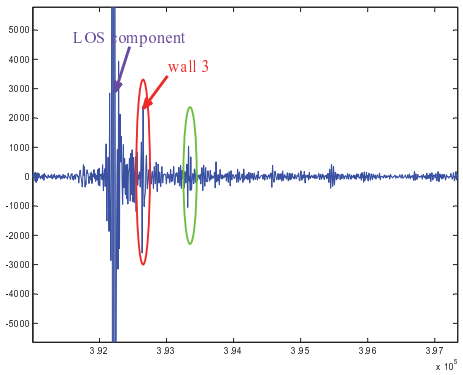
A heuristic way to detect peaks, summarized in Algorithm 3, is to check the slope of each local maxima. Three requirements are needed for the proposed algorithm: 1) the minimum distance between the device and the walls is no less than d_{min}, 2) the minimum TDOA of two detected consecutive echoes is no less than Δt, 3) the maximum candidate distance corresponding to detected peaks is no more than d_{max}. The reason for the requirements is as follows: 1) since the correlation property of the chirp signal is not ideal and the power of the LOS component is much larger than that of reflective components, the distance between the device and the walls should be large enough such that the peaks corresponding to reflective components are not overshadowed by the LOS component, 2) as the power of reflective paths decays rapidly, it is reasonable to restrict the detectable echoes within certain distances which depends on the power of loudspeaker. Given $d_{min} = 0.6$m, $d_{max} = 6.5$m and $\Delta t = \frac{0.5\text{m}}{c}$, where



(a) Correlator output at O_1 towards the first wall



(b) Correlator output at O_2 towards the second wall



(c) Correlator output at O_3 towards the third wall

Fig. 6: Sample of correlator output: Peaks with solid red ellipses correspond to walls while peaks with dash green ellipses correspond to either noise or higher order echoes

$c = 346$m/s, the detection results are marked by arrows in Fig. 7. We can see that the desired peaks are always detected. In order to detect as less false peaks as possible, one possible modification is to apply a tapering threshold which decreases as t increases.

B. SLAM Results

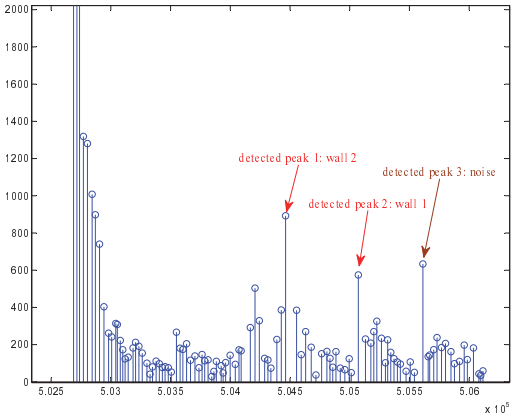
Echoes are collected at O_j $(j = 1, \ldots, 4)$ and $d_{j,j+1}$ $(j = 1, 2, 3)$ are measured by tape measure. The proposed peak detection algorithm is used to estimate the candidate distances from received signals. Note that the number of detected peaks are much larger than the number of first order echoes. Heuristics are used to remove peaks (e.g. those of small magnitudes) - otherwise, checking all combinations of echoes become computationally prohibitive. The proposed algorithm for SLAM is verified by experiment at O_1, O_2, O_3 and O_2, O_3, O_4. Given O_2, O_3, O_4, we assume that O_2 is the origin and O_3 lies on the x-axis. Even if some elements of $\mathbf{r}_j$ have measurement errors up to 10cm, SLAM is accomplished with small error of both the room shape and the coordinates of O_3 and O_4 with only unlabeled first-order echoes. In the presence of higher order echoes, the proposed algorithm may perform poorly and ambiguity may occur when the variance of φ is the only criterion used to determine f_j's. With noisy measurement, it is possible that the incorrect echo combination may yield feasible θ_i and φ with variance smaller than that of the correct echo combination. Furthermore, an interesting phenomenon is that sometimes the proposed algorithm is unable to provide the correct room shape, but the estimate of φ is always close to the true value. This means that better echo labeling approach is needed for robust SLAM. As most rooms are regular, we add a heuristic constraint: all the angles of two adjacent walls are between 50° and 130°. The comparison between the SLAM result and the ground truth is illustrated in Fig. 8. The candidate distances are obtained by the peak detection algorithm. Note that the coordinate system in Fig. 8(b) is a rotation of that in Fig. 8(a) by 135° counterclockwise. The SLAM results shown in the two figures are rotational images of each other. Experimental result indicate that heuristic constraints such as the above can largely eliminate incorrect combinations.

Algorithm 2 Peak detection algorithm

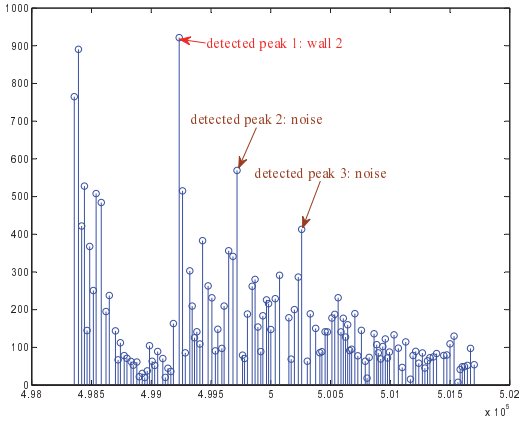
1: find LOS peak $(t_0^{(j)}, m_0^{(j)})$.
2: find local maxima of $|m^{(j)}(t)|$ appearing from $t_0^{(j)} + t_{min}$ to $t_0^{(j)} + t_{max}$.
3: find all peaks that are *peaky* and store them in M
4: set $P = \emptyset$
5: **if** **then**$|P| < |M|$
6: **if** there exist peaks in M whose locations are "close" to any peak in P **then**
7: remove those peaks from M.
8: **else**
9: add the peak with the largest magnitude of M to P.
10: **end if**
11: **end if**

V. Conclusion

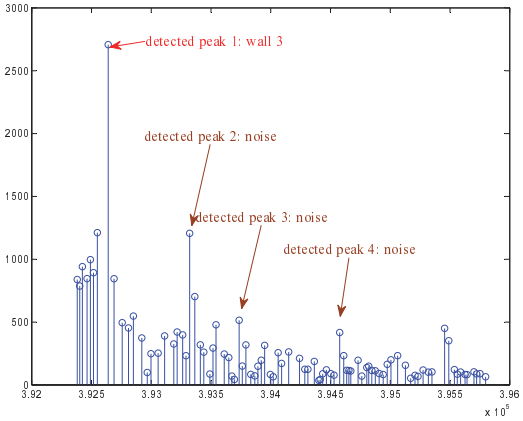
This work makes progress in acoustic SLAM using a single mobile device with unlabeled first order echoes. Theoretical guarantee of 2-D SLAM is established when two path lengths corresponding to three consecutive measurement points are available. Conversely, it was also shown that with only a single distance measurement, 2-d SLAM with unlabeled first order



(a) Peaks detected from correlator output at O_1 towards the first wall

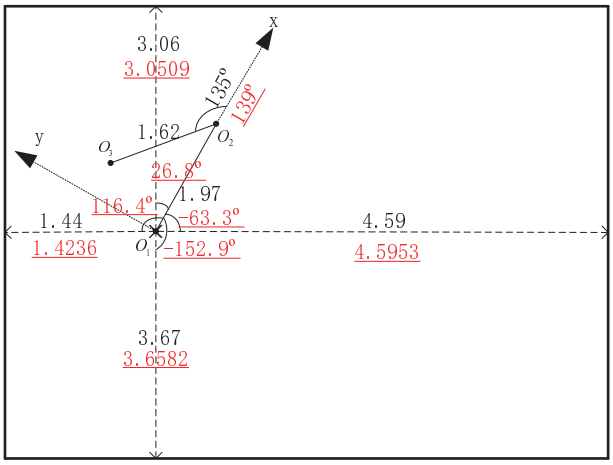


(b) Peaks detected from correlator output at O_2 towards the second wall

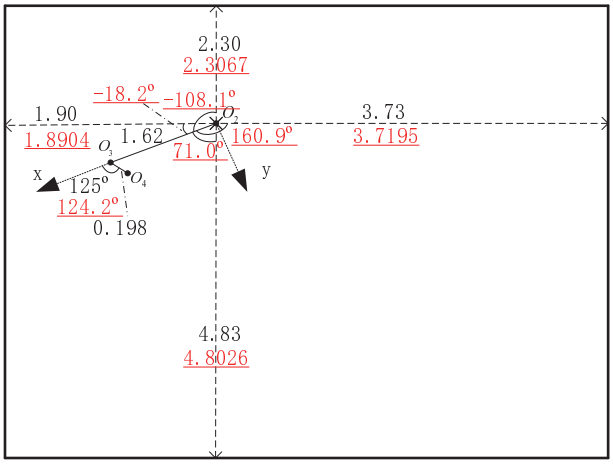


(c) Peaks detected from correlator output at O_3 towards the third wall

Fig. 7: Illustration of the performance of the proposed peak detection algorithm.



(a) SLAM via echoes collected at O_1, O_2 and O_3



(b) SLAM via echoes collected at O_2, O_3 and O_4

Fig. 8: Comparison between the ground truth (black) and experiment result (red underlined)

TABLE I: Feasibility of SLAM with unlabeled first order echoes and different geometry knowledge

geometry knowledge	any convex polygon
d_{12}, d_{23}, d_{13}	Yes
d_{12}, d_{23}	Yes
d_{12}	No
none	No

echoes is not possible for all convex polygons. The result is summarized in Table I.

While theoretical guarantee can be established for the noiseless case, the proposed algorithm needs to be enhanced to ensure a fully autonomous 2-D SLAM. Two particular issues that need to be further addressed include the robustness with respect to measurement noise and the computational complexity when a large number of peaks are detected at each measurement location.

REFERENCES

[1] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007.

[2] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, "Indoor localization without the pain," in *Proc. 16th Annu. Int. Conf. Mobile Computing, Networking (MobiCom)*, Chicago, IL, USA, Oct. 2010, pp. 173–184.

[3] C. H. Lim, Y. Wan, B. P. Ng, and C. M. S. See, "A real-time indoor wifi localization system utilizing smart antennas," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 618–622, May 2007.

[4] E. Martin, O. Vinyals, G. Friedland, and R. Bajcsy, "Precise indoor localization using smart phones," in *Proc. 18th ACM Int. Conf. Multimedia*, Firenze, Italy, Oct. 2010, pp. 787–790.

[5] M. A. Stelios, A. D. Nick, M. T. Effie, K. M. Dimitris, and S. C. A. Thomopoulos, "An indoor localization platform for ambient assisted living using uwb," in *Proc. 6th Int. Conf. Advances in Mobile Computing, Multimedia (MoMM)*, Linz, Austria, Nov. 2008, pp. 178–182.

[6] J. Schroeder, S. Galler, K. Kyamakya, and T. Kaiser, "Three-dimensional indoor localization in non line of sight uwb channels," in *Proc. IEEE Int. Conf. Ultra-Wideband*, Singapore, Sept. 2007, pp. 89–93.

[7] Y. Zhou, C. L. Law, Y. L. Guan, and F. Chin, "Indoor elliptical localization based on asynchronous uwb range measurement," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 1, pp. 248–257, Jan. 2011.

[8] S. Y. Jung, S. Hann, and C. S. Park, "Tdoa-based optical wireless indoor localization using led ceiling lamps," *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, pp. 1592–1597, Nov. 2011.

[9] A. M. Vegni and M. Biagi, "An indoor localization algorithm in a small-cell led-based lighting system," in *Proc. Int. Conf. Indoor Positioning and Indoor Navigation (IPIN)*, Sydney, Australia, Nov. 2012, pp. 1–7.

[10] J. Huang, D. Millman, M. Quigley, D. Stavens, S. Thrun, and A. Aggarwal, "Efficient, generalized indoor wifi graphslam," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, Shanghai, China, May 2011, pp. 1038–1043.

[11] A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Trans. Audio, Speech, Language Processing*, vol. 21, no. 2, pp. 439–443, Feb. 2013.

[12] Y. Kuang, S. Burgess, A. Torstensson, and K. Åström, "A complete characterization and solution to the microphone position self-calibration problem," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Vancouver, BC, Canada, May 2013, pp. 3875–3879.

[13] J. L. Blanco, J. A. Fernandez-Madrigal, and J. Gonzalez, "Efficient probabilistic range-only slam," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots, Syst.*, Nice, France, Sep. 2008, pp. 1017–1022.

[14] J. Djugash, S. Singh, G. Kantor, and W. Zhang, "Range-only slam for robots operating cooperatively with sensor networks," in *Proc. IEEE Int. Conf. on Robotics, Automation (ICRA)*, Orlando, FL, USA, May 2006, pp. 2078–2084.

[15] I. Dokmanić, R. Parhizkar, A. Walther, Y.M. Lu, and M. Vetterli, "Acoustic echoes reveal room shape," *Proc. Nat. Academy of Sci.*, vol. 110, no. 30, pp. 12186–12191, 2013.

[16] S. Tervo and T. Tossavainen, "3d room geometry estimation from measured impulse responses," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 513–516.

[17] F. Antonacci, J. Filos, M.R.P. Thomas, E.A.P. Habets, A. Sarti, P.A. Naylor, and S. Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 20, no. 10, pp. 2683–2695, Dec. 2012.

[18] I. Dokmanić, Y.M. Lu, and M. Vetterli, "Can one hear the shape of a room: The 2-d polygonal case," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 321–324.

[19] D. Ba, F. Ribeiro, C. Zhang, and D. Florencio, "L1 regularized room modeling with compact microphone arrays," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP)*, Dallas, TX, USA, Mar. 2010, pp. 157–160.

[20] M. Crocco, A. Trucco, V. Murino, and A. Del Bue, "Towards fully uncalibrated room reconstruction with sound," in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO)*, Lisbon, Potugal, Sept. 2014, pp. 910–914.

[21] I. Dokmanić, L. Daudet, and M. Vetterli, "From acoustic room reconstruction to slam," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 6345–6349.

[22] C. Evers, A. H. Moore, and P. A. Naylor, "Acoustic simultaneous localization and mapping (a-slam) of a moving microphone array and its surrounding speakers," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 6–10.

[23] F. Peng, T. Wang, and B. Chen, "Room shape reconstruction with a single mobile acoustic sensor," in *Proc. IEEE Int. Conf. Signal, Inform. Process. (GlobalSIP)*, Orlando, FL, USA, pp. 1116–1120.

[24] M. Krekovic, I. Dokmanic, and M. Vetterli, "Look, no beacons! optimal all-in-one echoslam," *arXiv preprint*, 2016.

[25] W. Kang, S. Nam, Y. Han, and S. Lee, "Improved heading estimation for smartphone-based indoor positioning systems," in *Proc. IEEE 23rd Int. Symp. Personal, Indoor, Mobile Radio Commun. (PIMRC)*, Sydney, Australia, Sep. 2012, pp. 2449–2453.

[26] F. Li, C. Zhao, G. Ding, J. Gong, C. Liu, and F. Zhao, "A reliable and accurate indoor localization method using phone inertial sensors," in *Proc. ACM Conf. Ubiquitous Computing (UbiComp)*, Pittsburgh, PA, USA, Sep. 2012, pp. 421–430.

[27] N. Roy, H. Wang, and R. Roy Choudhury, "I am a smartphone and i can tell my user's walking direction," in *Proc. 12th Annu. Int. Conf. Mobile Syst., Applicat., Services (MobiSys)*, Bretton Woods, NH, USA, June 2014, pp. 329–342.

[28] L. Zhang, K. Liu, Y. Jiang, X. Y. Li, Y. Liu, P. Yang, and Z. Li, "Montage: Combine frames with movement continuity for realtime multi-user tracking," *IEEE Trans. Mobile Computing*, vol. PP, no. 99, pp. 1–1, 2016.

[29] T. Wang, F. Peng, and B. Chen, "First order echo based room shape recovery using a single mobile device," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 21–25.

[30] M. Kreković, I. Dokmanić, and M. Vetterli, "Echoslam: Simultaneous localization and mapping with acoustic echoes," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 11–15.

[31] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Eng. Soc. Conv. 108*, Paris, France, Feb. 2000.

[32] G. Stan, J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *J. Audio Eng. Soc.*, vol. 50, no. 4, pp. 249–262, 2002.