

EchoSLAM: SIMULTANEOUS LOCALIZATION AND MAPPING WITH ACOUSTIC ECHOES

Miranda Kreković, Ivan Dokmanić, and Martin Vetterli

School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
{miranda.krekovic,ivan.dokmanic,martin.vetterli}@epfl.ch

ABSTRACT

We address the problem of jointly localizing a robot in an unknown room and estimating the room geometry from echoes. Unlike earlier work using echoes, we assume a completely autonomous setup with (near) collocated microphone and the acoustic source. We first introduce a simple, easy to analyze estimator, and prove that the sequence of room and trajectory estimates converges to the true values. Next, we approach the problem from a Bayesian point of view, and propose a more general solution which does not require any assumptions on motion and measurement model of the robot. In addition to theoretical analysis, we validate both estimators numerically.

Index Terms—Room geometry estimation, echo sorting, sound source localization, simultaneous localization and mapping.

1. INTRODUCTION

Simultaneous localization and mapping (SLAM) is a popular topic in robotics and computer vision. It is a more complex, but far more powerful alternative to localization in a known space, or mapping with a perfectly known sensing trajectory. Different flavors of SLAM are characterized by different kinds of uncertainties and sensing modalities. A common concept is that of landmarks—fixed points in space whose locations may be accessed through measurements (e.g. of range, azimuth, received power, visual information).

In this work we address SLAM based on echoes. We assume no preinstalled infrastructure in the room, and the bare minimum of sensing installed on the robot—a single omnidirectional source and a single omnidirectional receiver.

Prior work has considered visual [1, 2, 3], range-only [4, 5], and acoustic SLAM [6], as well as solutions based on multiple sensor modalities [7]. In [6], the authors propose a framework to simultaneously localize the mobile robot and multiple sound sources using a microphone array on the robot. Echoes and multipath have been used previously to do SLAM [8, 9], and for room geometry estimation in general [10, 11]. But these prior works rely on a fixed source or receiver, so that the echoes correspond to virtual beacons, or virtual landmarks, and depending on the sensing setup we may get measurements of range and/or azimuth of these virtual landmarks. In contrast, in our case there is no beacon—source and receiver are collocated on the same device. Moreover, we do not use a microphone array, rather a single microphone. Thus our landmarks (from which we get range-only measurements) are not static—they move with the robot. Our problem can then be equivalently stated as jointly reconstructing the trajectories of the robot and of the *moving* landmarks.

This work was supported by the Swiss National Science Foundation grant number 20FP-1 151073, “Inverse Problems regularized by Sparsity”.

Our contributions are as follows. We first propose an algorithm based on elementary trigonometry in order to lay down the main ideas. An additional benefit of simplicity is that we can show that the algorithm converges to the correct solution when the robot is exploring the space randomly. Next, we formulate a Bayesian solution inspired by FastSLAM [12]. We empirically observe that this more sophisticated algorithm strictly (and by a large margin) outperforms the elementary solution.

Section 2 introduces the notation, the problem setup, and the adopted image source model. In Section 3 we propose two methods for reconstructing the shape of a room from acoustic measurements. In Section 4 we numerically compare the performance of these two estimators, and we draw conclusions in Section 5.

2. PROBLEM SETUP

We assume that an omnidirectional acoustic source and a collocated omnidirectional microphone are mounted on a robot. The robot moves autonomously inside a room. At every step, the source produces a pulse, and the microphone registers the echoes. We define the room as a 2D polygon, and derive all results in 2D. The derivations can be easily extended to 3D.

Image source model. In a multipath environment, a microphone records both the direct path of the sound and its reflections from the walls. In the image source (IS) model [13, 14] we replace the reflections from the walls with signals produced by image sources—mirror images of the real sources across the corresponding walls—as shown in Fig. 1. For a first-order echo and the k th wall, described by the unit normal \mathbf{n}_k and any wall point \mathbf{p}_k , the image source $\tilde{\mathbf{s}}_k$ of the real source \mathbf{r} is computed as $\tilde{\mathbf{s}}_k = \mathbf{r} + 2\langle \mathbf{p}_k - \mathbf{r}, \mathbf{n}_k \rangle \mathbf{n}_k$. The sound propagation is then described by a family of room impulse responses where each RIR is idealized as a train of Dirac delta impulses produced by the real and image sources, and recorded by the microphone at position \mathbf{r}_n , $h_n(t) = \sum_{k \geq 0} a_k \phi(t - \tau_{n,k})$. a_k are the received magnitudes that depend on the wall absorption coefficients and the distance of the image source from the microphone. The propagation time $\tau_{n,k}$, also known as the time of arrival (TOA), is proportional to the distance between the microphone \mathbf{r}_n and the source $\tilde{\mathbf{s}}_{n,k}$:

$$\tau_{n,k} = \frac{\|\tilde{\mathbf{s}}_{n,k} - \mathbf{r}_n\|}{c}, \quad (1)$$

where c is the speed of sound.

The case of collocated microphone and source is illustrated in Fig. 2. We focus on the k th wall of the room and explain how to localize it with reference to the figure. The measurements consist of TOAs extracted from RIRs, $\tau_{n,k}$, and the robot motion commands for each step, \mathbf{v}_n . Since the source and the microphone are collocated, it is not possible to discriminate between translated, rotated

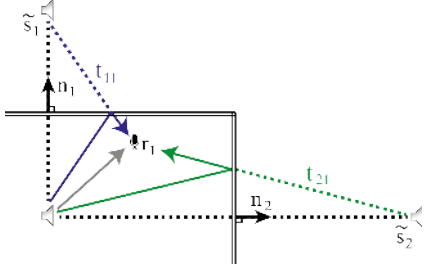


Fig. 1. Illustration of the image source model for first-order reflections.

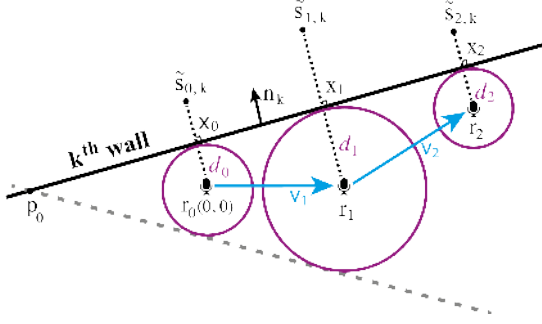


Fig. 2. Setup with collocated source and microphone mounted on a robot. The robot makes steps and obtains measurements. Distances from the wall, extracted from RIRs are shown in purple, and motion vectors in blue. The illustration presents the unique position of the wall assuming noiseless measurements.

and reflected variants of the room about the robot. We resolve this ambiguity by fixing some degrees of freedom—the initial robot’s position \mathbf{r}_0 indicates the origin, and we set the orientation of the first robot’s step to 0° . Then, we can calculate robot’s position at any step n , $\mathbf{r}_n = \sum_{i=0}^n \mathbf{v}_i$.

Proposition 1. Assuming ideal, noiseless measurements, we can uniquely determine the wall line after three measurements as (2):

$$\mathbf{x}_n = \mathbf{r}_n + \frac{d_n^2}{\|\mathbf{q}_n\|^2} \mathbf{q}_n \pm \frac{d_n \sqrt{\|\mathbf{q}_n\|^2 - d_n^2}}{\|\mathbf{q}_n\|^2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{q}_n, \quad (2)$$

where
$$\mathbf{p}_n = \frac{d_n}{d_n - d_{n+1}} \mathbf{r}_{n+1} - \frac{d_{n+1}}{d_n - d_{n+1}} \mathbf{r}_n, \quad (3)$$

and $\mathbf{q}_n = \mathbf{p}_n - \mathbf{r}_n$, \mathbf{r}_n is the robot’s position at step n , and d_n is its distance from the wall, $d_n = c\tau_n/2$.

Proof. The vector between the image source and the robot’s position is given by

$$\mathbf{y}_n = \tilde{\mathbf{s}}_n - \mathbf{r}_n = 2\langle \mathbf{p}_n - \mathbf{r}_n, \mathbf{n} \rangle \mathbf{n} \quad (4)$$

for all n . We observe: *i*) the direction of \mathbf{y}_n is perpendicular to the wall, and *ii*) the length of \mathbf{y}_n equals twice the distance between the robot and the wall, denoted d_n . The only line that satisfies both conditions for all n is the common tangent of the circles with centroids at \mathbf{r}_n and radii d_n given by (2). \square

As Fig. 2 shows, having only two measurements gives two solutions for the wall—there are two common tangents of two circles such that the centroids are on the same side of the tangents. They are shown as a thick black line and a dashed gray line in the figure. Assuming that the robot does not move in parallel with the wall, three measurements are sufficient to get a unique solution.

3. EchoSLAM

In practice, time measurements are imperfect. RIRs contain peaks that are introduced by various sources of noise and that do not correspond to any image source. It may be challenging to distinguish real echoes from spurious peaks. Moreover, peaks are unlabelled—we do not know which echo corresponds to which wall. It can happen that echoes produced by second-order image sources arrive before echoes produced by first-order image sources. We address both the problem of extracting correct impulses, and the problem of matching them with the corresponding walls. In addition to noise in acoustic measurements, we also assume noise in the robot’s movements.

3.1. Echo labelling

Echo labelling solves the problem of matching the echoes with corresponding walls.

Uniqueness claim. Given the propagation times of the first-order echoes, we can almost always correctly assign them to corresponding walls. The claim is based on the fact that in every two consecutive steps robot’s real positions, along with its image sources, define isosceles trapezoids with sides of the same length, equal to the length of the robot’s step. This is illustrated in Fig. 3. We assume that the length of the step $\|\mathbf{v}_n\|$ is known up to some uncertainty, and we claim that there is only one way to arrange the given propagation times, $\tau_{n-1,k}$ and $\tau_{n,k}$, to obtain such isosceles trapezoids.

Practical algorithm. In a real RIR, it is challenging to detect impulses that belong to first-order echoes only. A possible strategy is to extract a larger number of echo candidates from the recordings, and then group them using a combinatorial search based on the above criterion. Although combinatorial, this can be executed fast for typical q and K , and we can speed up using various heuristics. An example of TOA measurements for one simulated trajectory is illustrated in Fig. 4. The above steps are summarized in Algorithm 1, where we assume that the propagation times of echoes are stored in the matrix \mathbf{U} .

3.2. Robot’s localization and room reconstruction

We adopt a probabilistic model of the uncertainties in the robot’s perception and motion. At time n , we seek to calculate a posterior of the robot’s positions \mathbf{r}_n along with the parameters of the walls θ ,

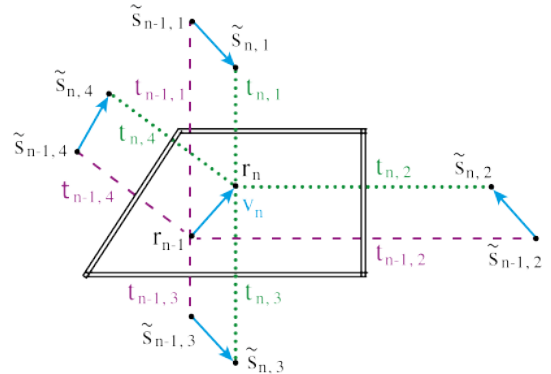


Fig. 3. Image source model for first-order echoes, and collocated microphone and source. Sound rays at the measurement $n - 1$ are shown in purple (dashed line) and sound rays at the measurement n in green (dots). For every two consecutive steps, the robot’s real positions, along with its image sources, define isosceles trapezoids with sides of the same length—the length of the robot’s step.

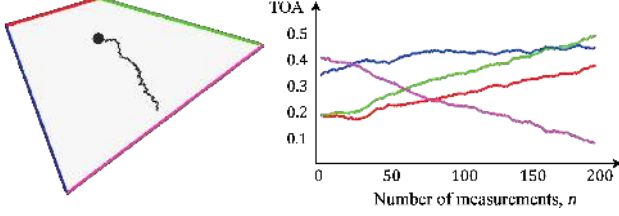


Fig. 4. *Left:* An example of the robot’s trajectory in a room with four walls. The trajectory is shown in gray, and the initial position of the robot is marked with a circle. *Right:* TOA measurements of echoes for 200 robot’s steps. Correspondence between the echo and the wall is visualized with the same color.

Algorithm 1 Algorithm for echo labelling

Input: number of echo candidates, $q \in \mathbb{N}$,
RIR from the n th measurement,
length of the robot’s step, $\|\mathbf{v}_n\| \in \mathbb{R}$,
labelled TOAs for the $(n-1)$ st measurement, $\mathbf{U}_{n-1} \in \mathbb{R}^K$

Output: labelled TOAs for the n th measurement at n , $\mathbf{U}_n \in \mathbb{R}^K$

- 1: $\mathbf{T} \leftarrow$ select TOAs of q candidate echoes from RIR
- 2: **for** all subsets τ of \mathbf{T} with K el. and each permutation $\pi(\tau)$ **do**
- 3: **if** one can construct K isosceles trapezoids as in Fig. 3, given the lengths of the bases τ , tops \mathbf{U}_{n-1} , and sides, $\|\mathbf{v}_n\|$ **then**
- 4: **return** $\mathbf{U}_n \leftarrow \pi(\tau)$
- 5: **end if**
- 6: **end for**

given all the motion vectors and TOA measurements up to time n , $p(\mathbf{r}_n, \theta | \mathbf{v}^n, d^n)$. The superscript n refers to a set of variables from step 1 to step n .

In the rest of the paper we assume to know the echo labelling and estimate each wall independently. We study two estimators: The first one is simply the mean of the independent estimators at each robot’s step, which by the law of the large numbers converges to the expectation $\mathbb{E}[p(\theta | \mathbf{v}^n, d^n)]$. The second one is based on Bayes filtering, where we simultaneously update the distribution of the wall parameters and with the distributions of the robot’s positions. The room geometry and the robot’s trajectory are both estimated using the maximum a posteriori (MAP) rule. The reason for introducing the first estimator is to show that a simple estimator converges to the correct solution as the number of measurements grows. Even though we have no formal proof of the same property for the second estimator, we compare it with the first one and empirically observe that it achieves correct solution, while the convergence rate significantly improves.

3.2.1. Estimation of the wall

For the first estimator we assume that TOA measurements are noiseless and that errors in robot’s steps have Gaussian distributions. The wall is modelled as a line with slope $\tan(\theta)$ and offset b . Since the first robot’s position, \mathbf{r}_0 , and the first measurement, d_0 , are known, the only unknown parameter is the slope.

Proposition 1. Let us define each robot’s step by its length and orientation, $\mathbf{v}_n(x_n, \phi_n)$, where both variables have Gaussian distributions

$$x_n \sim \mathcal{N}(\mu_{x_n}, \sigma_x^2), \quad \phi_n \sim \mathcal{N}(\mu_{\phi_n}, \sigma_\phi^2). \quad (5)$$

Assume that $\mu_{x_n} = \mu_x$ for every n . We define the estimator of θ for each measurement n as:

$$\begin{aligned} \hat{\theta}_n &= \mu_{\phi_n} + \arcsin\left(\frac{d_{n+1} - d_n}{\mu_x}\right) \\ &= \mu_{\phi_n} + \arcsin\left(\frac{x_n \sin(\theta - \phi_n)}{\mu_x}\right), \end{aligned} \quad (6)$$

and the final estimate of θ after N measurements as:

$$\hat{\theta}^N = \frac{1}{N} \sum_{n=1}^N \hat{\theta}_n. \quad (7)$$

Then, $\hat{\theta}^N$ is unbiased when μ_{ϕ_n} is uniformly distributed on the circle.

Sketch of the proof. The ratio x_n/μ_x is distributed as $\mathcal{N}(1, \sigma_x^2/\mu_x^2)$ for all n . Since it does not depend on n , so we delay taking the corresponding expectation. We introduce an auxiliary Gaussian random variable, $\theta - \phi_n \sim \mathcal{N}(\theta - \mu_{\phi_n}, \sigma_\phi^2)$. One can verify that the bias of $\hat{\theta}_n$ depends only on the parameter μ_{ϕ_n} . Therefore, we rewrite it as $\hat{\theta}_n = \theta + f(\mu_{\phi_n})$, where $f(\cdot)$ is a periodic with zero mean over the period. Then we observe that

$$\begin{aligned} \mathbb{E}(\hat{\theta}^N) &= \mathbb{E}\left(\frac{1}{N} \sum_{n=1}^N (\theta + f(\mu_{\phi_n}))\right) \\ &= \theta + \frac{1}{N} \sum_{n=1}^N \mathbb{E}(f(\mu_{\phi_n})), \end{aligned} \quad (8)$$

and the uniform distribution of μ_{ϕ_n} on the interval $[0, 2\pi]$ provides that $\mathbb{E}(f(\mu_{\phi_n})) = 0 \forall n$, so that $\mathbb{E}(\hat{\theta}^N) = \theta$. \square

The physical meaning is as follows: the estimator is positively biased if the robot walks towards the wall $\mathbb{E}(\hat{\theta}^N) - \theta \geq 0$, and negatively biased if the robot walks away from the wall $\mathbb{E}(\hat{\theta}^N) - \theta \leq 0$. These biases cancel for a robot that picks its direction at random. As we assume that the robot performs a random walk, the values of μ_{ϕ_n} are uniformly distributed on the circle, and one can verify from the graph that the function has zero mean. Thus $\mathbb{E}(f(\mu_{\phi_n}))$ is zero. By the law of the large numbers, the sequence of estimates $\hat{\theta}_n$ converges to the real value θ .

3.2.2. Bayes filtering

The second approach, based on Bayes filtering, is not bound to a limited parametric subset of distributions for motion or measurement models. Similar to [12], and the vast number of the papers on SLAM that adopt this methodology, the approach is based on a conditional independence property of the problem—knowledge of the robot’s position renders the individual TOA measurements independent. Therefore, knowledge of the exact position of one wall tells us nothing about the other walls, when the robot’s location is known. These conditional independences imply the following factorization:

$$p(\mathbf{r}_n, \theta | \mathbf{v}^n, d^n) = p(\mathbf{r}_n | \mathbf{v}^n, d^n) \prod_{k=1}^K p(\theta_k | \mathbf{r}_n, \mathbf{v}^n, d_k^n)$$

The algorithm that simultaneously localizes the robot and estimates the walls consists of three steps and is summarized in Algorithm 2.

Algorithm 2 Bayes filtering

Input: TOAs for every n , $n \leq N$
robot's steps for every n , \mathbf{v}_n

Output: estimation of the angle of the wall after N steps, $\hat{\theta}^N$
estimation of the robot's positions for every n , $\hat{\mathbf{r}}_n$

1: **for** every measurement n **do**

2: Predict the robot's position for the n th measurement based on the motion model:

$$p(\mathbf{r}_n | \mathbf{v}^n, d^{n-1}) = \int p(\mathbf{r}_n | \mathbf{v}_n, \mathbf{r}_{n-1}) p(\mathbf{r}_{n-1} | \mathbf{v}^n, d^{n-1}) d\mathbf{r}_{n-1}$$

3: Assume the measurement model $p(d_n | \mathbf{v}^n, \mathbf{r}_{n-1}, d^{n-1}, \theta) = p(d_n | \mathbf{r}_n, \theta)$, and update the wall's parameter θ :

$$p(\theta | \mathbf{r}^n, d^n, \mathbf{v}^n) \sim p(d_n | \theta, \mathbf{r}_n) p(\theta | \mathbf{r}^{n-1}, d^{n-1}, \mathbf{v}^{n-1})$$

4: Update the robot's position in order to incorporate the last measurement:

$$p(\mathbf{r}_n | \mathbf{v}^n, d^n) \sim p(d_n | \theta, \mathbf{r}_n) p(\mathbf{r}_n | \mathbf{v}^n, d^{n-1})$$

5: Estimate $\hat{\theta}^N$ and $\hat{\mathbf{r}}_n$ using MAP:

$$\hat{\theta}^N = \arg \max_{\theta} p(\theta | \mathbf{r}^n, d^n, \mathbf{v}^n), \hat{\mathbf{r}}_n = \arg \max_{\mathbf{r}^n, d^n, \mathbf{v}^n} p(\mathbf{r}_n | \mathbf{v}^n, d^n)$$

6: **end for**

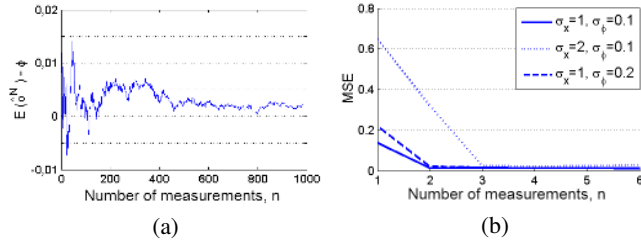


Fig. 5. Estimation of the wall's parameter θ . (a) Dependence of $\mathbb{E}(\hat{\theta}^N) - \theta$ on the number of steps for the first estimator. (b) Dependence of the MSE on the number of steps for the second estimator.

The underlying idea of the second step is that all midpoints between the robot \mathbf{r}_n and the image source $\tilde{\mathbf{s}}_n$ lie on the same line, and that line is perpendicular to every line connecting the robot \mathbf{r}_n and the image source $\tilde{\mathbf{s}}_n, \forall n$. The line has the same direction as the wall, $\tan \theta$.

4. NUMERICAL SIMULATIONS

We validated both approaches numerically. In Fig. 5(a) we verify the unbiasedness result from Section 3.2.1 empirically, and show that $\mathbb{E}(\hat{\theta}^N) - \theta \rightarrow 0$ as N grows. For the second algorithm we assumed the same model as for the first estimator, for which we have proved the convergence, and computed the mean squared error (MSE) in every step. The result is given in Fig. 5(b) (the means of the Gaussian distributions are $\mu_x = 30$, and $\mu_{\phi_n} \sim \mathcal{U}[0, 2\pi]$). Faster convergence of the second estimator shows that the room estimation significantly improves when we exploit the information about the possible robot's positions at each step and localize the robot simultaneously, rather than observe the steps independently as in the first approach.

Figures 6 - 8 relate to the Algorithm 2. Distributions of the room geometry at steps $n = 2, 4, 6$ and 8 are illustrated in Fig. 6; and MAP estimates for $n = 4, 6$ and 8 are shown in Fig. 7. The estimator recovers the room geometry with high accuracy after a small number

of steps. Distributions of the robot's positions and its image sources are shown in Fig 8.

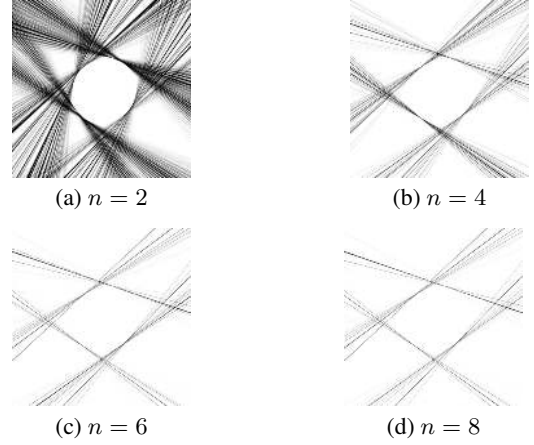


Fig. 6. Sampled distribution of the room geometry estimated with the Algorithm 2 at different steps. The lightness level of the lines presents the probability of the estimation: darker line—higher probability that the line corresponds to the wall.

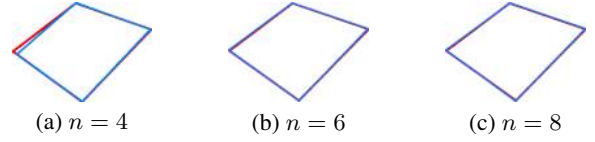


Fig. 7. Estimates of the room geometry at different steps, and the actual room geometry are shown in blue and red, respectively.

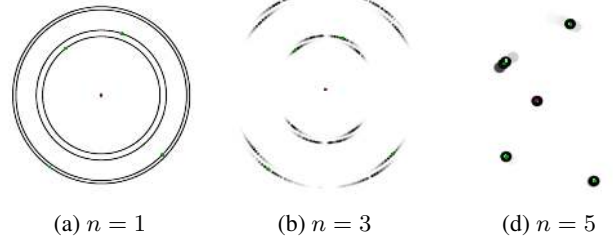


Fig. 8. Estimates of the robot's real positions and its image sources.

5. CONCLUSION

We proposed an algorithm for simultaneous localization and mapping based on multipath propagation inside a room. Our sensing setup is rudimentary—we assumed a single omnidirectional sound source and a single omnidirectional microphone collocated on a robot, and no preinstalled infrastructure in the room. We demonstrated that the measurement of distances between the robot and the walls are sufficient to develop an algorithm that estimates robot's trajectory precisely, and recovers the geometry of the room.

Ongoing research includes building a generic range-only framework for simultaneous localization and mapping, as we believe that current solutions do not fully exploit the geometry of such setups.

6. REFERENCES

- [1] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, *MonoSLAM: Real-time single camera SLAM*, in IEEE Transaction on Pattern Analysis and Machine Intelligence, pp. 1052 - 1067, June 2007
- [2] B. Clipp, L. Jongwoo, J. M. Frahm, and M. Pollefeys, *Parallel, real-time visual SLAM*, in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3961 - 3968, Oct. 2010
- [3] M. Blösch, S. Weiss, D. Scaramuzza, and R. Siegwart, *Vision based MAV navigation in unknown and unstructured environments*, in 2010 IEEE International Conference on Robotics and Automation (ICRA), pp. 21 - 28, May 2010
- [4] J. L. Blanco, J. A. Fernandez-Madrigal, and J. Gonzalez, *Efficient probabilistic Range-Only SLAM*, in IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1017 - 1022, Sept. 2008
- [5] J. Djughash, S. Singh, G. Kantor, and W. Zhang, *Range-only SLAM for robots operating cooperatively with sensor networks*, in Proceedings IEEE International Conference on Robotics and Automation, pp. 2078 - 2084, May 2006
- [6] J. S. Hu, C. Y. Chan, C. K. Wang, M. T. Lee, and C. Y. Kuo, *Simultaneous localization of a mobile robot and multiple sound sources using a microphone array*, in IEEE International Conference on Advanced Robotics, vol. 25, no. 1-2, pp. 135 - 152, 2011.
- [7] C. Brunner, T. Peynot, and T. Vidal-Calleja, *Combining multiple sensor modalities for a localisation robust to smoke*, in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2489 - 2496, Sept. 2011.
- [8] E. Leitinger, P. Meissner, M. Lafer, and K. Witrisal, *Simultaneous localization and mapping using multipath channel information*, in IEEE ICC 2015 Workshop on Advances in Network Localization and Navigation (ANLN), June 2015
- [9] C. Gentner, and T. Jost, *Indoor Positioning using time difference of arrival between multipath components*, in IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1 - 10, Oct. 2013
- [10] I. Dokmanic, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli, *Acoustic echoes reveal room shape*, in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 321 - 324, 2011
- [11] F. Antonacci, J. Filos, M. R. P. Thomas, E. A. P. Habets, A. Sarti, P. A. Naylor, and S. Tubaro, *emphInference of room geometry from acoustic impulse responses*, in IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 10, pp. 2683 - 2695, July 2012.
- [12] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, *Fast-SLAM: A factored solution to the simultaneous localization and mapping problem*, in Proceedings of the AAAI National Conference on Artificial Intelligence, pp. 593 - 598, 2002
- [13] J. B. Allen and D. A. Berkley, *Image method for efficiently simulating small-room acoustics*, in Journal of the Acoustic Society of America, vol. 65, no. 4, pp. 943-950, 1979.
- [14] J. Borish, *Extension of the Image Model To Arbitrary Polyhedra*, in Journal of the Acoustic Society of America, vol. 75, no. 6, pp. 1827-1836, 1984.