LOCALIZING A MOBILE ROBOT WITH INTRINSIC NOISE

Brian F. Allen, Flavien Picon, Sébastien Dalibard, Nadia Magnenat-Thalmann, Daniel Thalmann

LOCALIZING A MOBILE ROBOT WITH INTRINSIC NOISE

B. F. Allen, F. Picon, S. Dalibard, N. Magnenat-Thalmann and D. Thalmann

BeingThere Centre, Institute for Media Innovation
Nanyang Technological University
Singapore

ABSTRACT

Robot localization is a key barrier to providing natural interaction between 3D virtual characters, human users and mobile robots. Knowing where the robot is, relative to a known world-frame, is essential to directed gestures, gazes and expressions between the robot and the other real and virtual participants in a telepresence system. The intrinsic noise of robots is a flexible and robust, yet under-examined, source for localization information. Sounds that likely localize the robot are identified and separated from background noises using a support vector machine. The resulting sound-bearing data is combined with robot odometry using a particle filter. Experiments conducted in a noisy office environment show substantial improvement over odometry alone.

Index Terms — Robot Localization, Intrinsic Noise, Particle Filter, Support Vector Machine, Mixed-Reality, Integrated Interaction

1. INTRODUCTION

Several new technologies hold out the promise of integrated, situated human-autonomous agent interaction. Life-like virtual humans can be displayed on 3D auto-stereoscopic LCD displays to appear within the viewer's interaction space. Mobile humanoid robots are able to walk within human environments, directly interacting with both users and manipulable objects. Such mobile robots and 3D virtual humans will exist explicitly within the 3D space of the human viewers. The regions of 3D space that human users can directly touch, see and hear are the same regions of 3D space that the virtual humans and mobile robots can pointed to, look at and manipulate. With a common reference frame and spatial context, interactions between controlled agents (virtual human or mobile robot) are seamlessly integrated into the real-world environment of the user. We refer to this specific kind of mixed-reality as integrated interaction environments (IIE). Fig. 1 pictures an IIE within a relaxed office setting, complete with a mobile humanoid robot and a virtual human character on the display.

A key difficulty with realizing an IIE is determining precisely the locations of humans and robots. This problem is exacerbated by two practical factors common to home and commercial settings. First, environments are generally dynamic, cluttered and visually complex, making localization algorithms based solely on vision hard to use. Second, physical properties of floor surfaces, including rigidity and friction, vary broadly across types of floors, and floor type may differ even within a single IIE. This limits the accuracy of odometry methods that could be used to localize robots. Classic approaches to localization depend primarily on robot odometry and ego-centric vision. Robot odometry measures

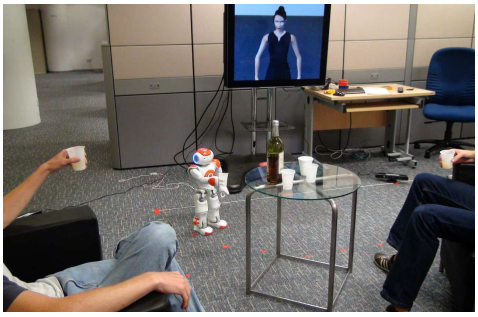

Figure 1. Scenario example image

changes to the robot's internal state (e.g., joint angles) to determine the robot's change in world-space position. Odometry alone is subject to drift and errors due to unmeasured quantities, such as the slip between a robot's foot and the floor. Odometry is less reliable when the robot is walking on unusual, complex or uncertain surfaces, such as home or office carpeting. Ego-centric vision is perhaps the most common mean of robot localization [1], and has found significant success in semi-structured environments such as autonomous driving [2]. However, vision's effectiveness is greatly reduced in practical environments that may contain reflecting objects or confounding colors or textures.

In this paper, we introduce a localization system designed to mitigate these real-world impediments. The key contribution is the observation that operating mobile robots emit significant, identifiable motor noise, and that this motor noise provides usable information about the location of the robot. This intrinsic noise is shown to be, on its own, sufficient to localize an unmodified Aldebaran Nao humanoid robot within a noisy office setting. Further, we show that inexpensive, commodity microphone arrays, such as the Microsoft Kinect are sufficient for practical implementations. While this paper considers localization using only intrinsic noise, we speculate that intrinsic noise could also augment vision-based localization or be used as an alternative in circumstances when vision fails, such as when the line-of-sight is partially or totally occluded or in particularly visually complex environments.

Unfortunately, naive use of microphone arrays to locate the robot can fail in the presence of situational noise, for example, conversations, passing footsteps, cell phone rings, opening and closing of doors, and office printers. In order for the intrinsic noise of the robot to be useful for localization, a method is needed to distinguish nuisance sounds from the useful intrinsic robot noise. Further, even when nuisance sounds are removed, the accuracy and precision of commodity microphone arrays are poor, leading to erratic and noisy position estimates. Previous approaches to using sound for robot localization have used either specific sounds

emitted by the robot [3], or considered the related but distinct problem of using robot-mounted microphones for localizing environmental sounds [4].

Our approach uses two techniques from machine learning to address the two key problems of situational noise and sound localization sensor inaccuracy. First, as described in section 2, segments of recorded sound are classified based on a small number of signature features. Second, sounds that are likely to reveal the robot's location are provided as sensor data to a sequential Monte Carlo filter, as discussed in section 3. To validate the proposed method, an IIE was setup in a natural, noisy office environment, and the accuracy of intrinsic-noise localization was compared to simple odometry as reported in section 4. Although the proposed method uses only intrinsic-noise, it is found to be significantly more accurate than odometry, and could serve both to augment better-established vision-based techniques, and as a backup for real-world situations were vision is confounded.

2. SOUND PROCESSING AND CLASSIFICATION

A sound localization device, such as the Kinect, outputs the estimated bearing of a perceived sound in its own frame, i.e. a bearing of 0 corresponds to a sound emitted in front of the sensor. A key difficulty with effective sound localization is distinguishing sounds that inform about the robot position from the broad class of sounds that do not. Different IIEs will certainly be subject to different confounding sounds, and different mobile robots will emit different intrinsic noise. Experimental recordings are used to build a statistical model and a classification module capable of distinguishing robot intrinsic noise from ambient sounds. This data-driven approach requires a training phase for each robot-environment pairing, but also allows the flexibility of learning.

To record the sample sounds, the robot walks in-place for a short period of time. It is placed by hand at an actual bearing θ in the sensor frame. During this period, a sample of typical confounding sounds are made, such as human speech or cell-phone rings. The sound localization sensor outputs an observed bearing $\hat{\theta}$. θ and $\hat{\theta}$ are recorded for each sound sample, at 60 Hz. For this particular experiment, the recordings are repeated at nine separate actual bearings, shown in Fig. 2, for approximately one minute at each bearing, for a total of approximately 36,000 samples. The fast Fourier transform is applied to each sample and the intensities are aggregated across 500 Hz-wide bands evenly from 0 Hz to 8000 Hz. The aggregated intensity corresponding to the j^{th} bandwidth is noted I^j. For the system to ignore properly extraneous environmental noises, each sound sample is classified using its frequency signature $(I^j)_j$. Classification is done using a support vector machine (SVM) [5]. Samples were annotated according to the correspondence between the measured bearing and the robot position, i.e. $\hat{\theta} \sim \theta$. The feature vector for each sound sample is the vector of sixteen intensities I^j scaled to within $(-1, 1)$. A binary classification of:

$$a = \begin{cases} 0 & \text{if } |\hat{\theta} - \theta| \leq \frac{\pi}{24} \\ 1 & \text{otherwise} \end{cases} \qquad (1)$$

was assigned to each scaled sample I^j for training the SVM. Five-fold cross-validation gave a 76.1% classification accuracy of samples. Classification of individual samples is fast enough to match the sensor rate. As the system is based on a probabilistic approach we use the probability of each sample, given by SVM, to modify the confidence of the associated bearing. Such an approach also reduces the effects of mis-classification.

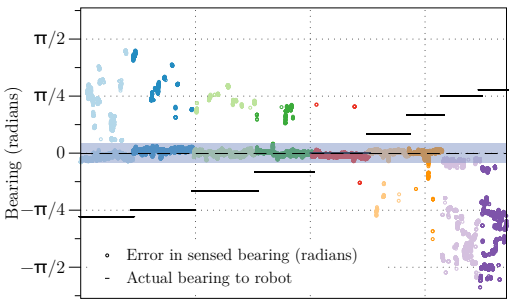

Figure 2. Recorded error in bearing angle (circles) for each actual bearing angle (black lines) between robot and microphone array. Each color represents a different actual bearing. Kinect is more accurate for sounds coming in front of it (bearings close to 0), and less accurate at the ends of its range ($\pm\,\pi/4$).

3. PROBABILISTIC LOCALIZATION

The goal of localization is to estimate the position of the robot $\mathbf{x}^k = [x, y, \phi]^T$ at time $t = k$. Each i^{th} sound sensor registered in the system provides an estimate of the bearing $\hat{\theta}_i$ from that sensor to the robot. Each of these bearings are imprecise, due to noise in the microphone array time-delay estimation, and uncertain, due to the uncontrolled nature of the robot's intrinsic sounds and the inherent randomness of the environment. The total sensor information at any one time is $z = [\hat{\theta}_i, a_i]_i$. Fig. 3 shows a theoretical example of a robot localized by two sound sensors.

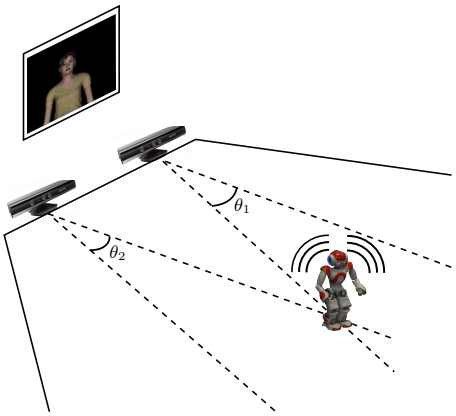

Figure 3. Schematic experimental setup. Intrinsic noise from the robot is localized by two Kinects.

The sequential Monte Carlo approaches, such as the particle filter [6], have proven extremely effective for robot localization. These methods discretize the probability distribution of belief about the robot's current position into a set of weighted hypotheses or samples. Belief about the robot's position is updated by a recursive Bayes filter, estimating the posterior distribution of positions conditioned on sensor data up to time k, $z_{1:k}$.

The recursive process entails a prediction step, where the model of the robot's motion $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is used to update the prior probability distribution of robot state $p(\mathbf{x}_{k-1}|z_{1:k-1})$ as

$$p(\mathbf{x}_k|z_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|z_{1:k-1})d\mathbf{x}_{k-1}. \qquad (2)$$

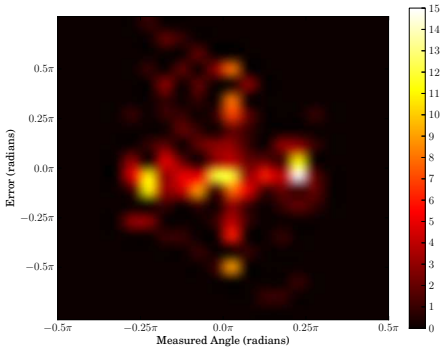

Figure 4. Histogram of the signed error in bearing reading vs the measured sound angle. Note the increased error around the zero angle and the slightly asymmetric form.

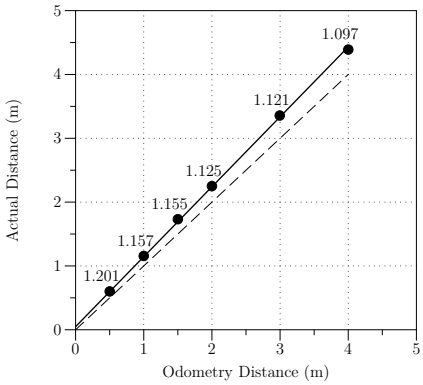

Figure 5. Linear robot odometry vs distance actually travelled.

In practice, the model of robot motion is conditioned on the commanded motion of the robot and the model of this error, as described in detail in section 3.1. With each iteration, new sensor data z_k for time k is measured and used to calculate the posterior in the update step, with a post-normalization step,

$$p(\mathbf{x}_k|z_{1:k}) \propto p(z_k|\mathbf{x}_k)p(\mathbf{x}_k|z_{1:k-1}), \tag{3}$$

The particle filter discretizes the probability distribution of robot state as $S_j^k = [\mathbf{x}_j^k, w_j^k]$ for time $t = k$ and with j as the particle index. In each time-step, prediction and update are applied to each particle. The final estimate of robot position for each time-step is a robust mean of a windowed average around the highest-weighted particle.

3.1. Robot Motion Model

The model of the robot's motion is determined empirically, and is specific to the particular robot (here, the Aldebaran Nao) and the floor surface conditions. Motion commands of linear velocity v and angular rate of change in heading ω are communicated to the robot in coordinates relative to its current direction of facing ϕ. The motion model updates hypothesis positions over timestep Δt according to

$$\begin{pmatrix} x_k \\ y_k \\ \phi_k \end{pmatrix} = \begin{pmatrix} x_{k-1} + \bar{v}\Delta t \cos(\phi_{k-1}) \\ y_{k-1} + \bar{v}\Delta t \sin(\phi_{k-1}) \\ \phi_{k-1} + \bar{\omega}\Delta t \end{pmatrix} \tag{4}$$

with

$$\begin{pmatrix} \bar{v} \\ \bar{\omega} \end{pmatrix} = \begin{pmatrix} v + \mathcal{N}(l_\mu, l_{\sigma^2}) \\ \bar{\omega} + \mathcal{N}(r_\mu, r_{\sigma^2}) \end{pmatrix} \tag{5}$$

where $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution and d_μ, d_{σ^2} define the linear noise distribution and r_μ, r_{σ^2} define the angular.

3.2. Sensor Model

The role of the sensor model within the particle filter is to determine the probability of the observed sensor data z_k at time k given some hypothetical robot state $\mathbf{x}$,

$$p(z_k|\mathbf{x}_k) = \prod_i (1 - a_i^k) + a_i^k(\text{erf}(\theta_i^k + \frac{w}{2}) - \text{erf}(\theta_i^k - \frac{w}{2})) \tag{6}$$

for fixed sample width w and with error function erf from $\mathcal{N}(s_\mu, s_\sigma^2)$.

4. EXPERIMENTAL VALIDATION

4.1. Empirical Sensor Model

The accuracy and precision of the Kinect's sound-bearing sensing were determined experimentally, using the average values from two different Kinect devices. From the data described in section 2, the best-fit parameters of the assumed normal distribution were a mean s_μ of 0.145 rad and a standard deviation of 0.15 rad. A histogram of the error in measured bearing is shown in Fig. 4.

4.2. Empirical Robot Motion Model

In order to estimate the error in linear and angular motion, the robot was assigned different walking tasks, and for each we recorded the robot command, the odometry measures, and real measures taken using a meter and a protractor. The robot commands were divided in two groups: first walking straight covering 0.5, 1.0, 1.5, 2, 3 and 4 meters, and second walking one meter after turning on the spot of -135, -90, -45, 45, 90, 135 and 180 degrees.

The best-fit Gaussian noise model for the distance travelled by the robot in this environment was $\begin{pmatrix} d_\mu \\ d_{\sigma^2} \end{pmatrix} = \begin{pmatrix} 1.125 \\ 0.075 \end{pmatrix}$ and the best fit for rotational error, $\begin{pmatrix} r_\mu \\ r_{\sigma^2} \end{pmatrix} = \begin{pmatrix} 0.15v \\ 0.05 \end{pmatrix}$.

Fig. 5 illustrates the measured values for distance error, and shows the high quality of the linear fit. Note also that angular noise depends on the commanded rate of travel.

4.3. Spatial Setup

Our final layout for the walking experiment is similar to Fig. 3. Even though two Kinects are theoretically sufficient to localize a robot navigating on a 2D plane, we have added a third Kinect for better precision. The coordinates, 2D position and orientation (x, y, α) of the Kinects in a fixed world-frame are: $(-1.0, 0.0, 0.0)$, $(2.0, 0.0, \frac{\pi}{4})$ and $(0.0, 4.0, \pi)$. We are covering, with satisfying results a rough area of 3 by 3 meters.

4.4. Results

As a test scenario the robot had to perform a walking task and reach a pre-define goal, starting from a fixed configuration. Fig. 6 shows one trial of this experiment. During the walk, the system uses information from the sound localization and the robot odometry to update an estimate of robot position and heading. The robot used this estimate in a feedback loop to constantly walk towards

Figure 6. Representation of robot key position while traveling to the goal point

the goal. The robot was always placed on the same starting configuration $(1.0, 3.0, \pi)$ and aimed at reaching the 2D goal point $(0.0, 1.5)$. No prior knowledge about the robot initial position was given to the probabilistic localization engine. The robot started by walking straight for 5 s, to bootstrap the localization engine. Then, the walk control feedback-loop was running at a frequency of 1 Hz. The walk was stopped when the estimate of the robot distance to the goal was less than 0.2 m.

trial	time(s)	Position Estimate Error(m)	Distance to goal(m)
1	62	0.142	0.17
2	32	0.165	0.18
3	45	0.14	0.184
4	84	0.21	0.178
5	35	0.21	0.161
avg	**51.6**	**0.173**	**0.174**
Reference	35	1.06	1.05

Table 1. Recordings from 6 robot trials.

As a reference trial, we put the robot at the starting configuration and made it walk towards the target using the same walking strategy and only odometry for localization during the walk. In this trial, perfect knowledge of the robot world-frame initial configuration was assumed. Table 1 shows the average results of five trials of the experiment and the results of the reference trial. We have indicated the total walking time, the error in the robot position estimate when the robot stopped walking and the distance between the robot and the goal at the end of the walk. The error in the position estimate is due to sensor noise and small errors in the environment setup (microphone positions and orientations). The final distance to the goal is always less than 0.2 m, which is the required precision in accomplishing the task. On the other hand, the reference walk motion, without using sound for localization, ended up at around 1 m from the goal. Some trials required the robot to walk for a long time, because the localization engine was unable to localize the robot with sufficient precision for some time. Note that the robot needs at least 20 s at full walking speed to travel the required distance. In the trial presented in Fig. 6, we can clearly see the different stages of the robot control. First the robot started by walking straight for 5 seconds. Once this bootstrapping phase is over, the system takes over and makes the robot turn right towards the goal. Finally, the sampled images of the robot were evenly time spaced, so the visible jamming of robot at the end of the trajectory highlights our choice of slowing down the robot pace when getting closer to the goal.

5. CONCLUSION AND FUTURE WORK

In this paper, we have presented a localization system designed for integrated interaction environments. Our system is based on intrinsic robot noise tracking, and uses inexpensive and easily available microphone arrays. To perform effective sound localization, we have evaluated and implemented different methods from signal processing and robotics literature: first, we have used a support vector machine to identify sound signals coming from the robot; second, we have performed probabilistic localization using a particle filter. Both methods are well suited for handling noisy data. Our system has been shown to work well for tracking a small humanoid robot within a natural, noisy office environment. We were able to overcome the limitations of odometry-based localization and the robot could reach navigation targets a few meters away with a precision of about 15 cm.

This work is a first step towards a general IIE, and suffers from limitations that will be addressed in future work. First, in this paper, we have focused on intrinsic sound data and proved that it is useful for robot localization. In near future, we plan to add vision based tracking of the robot and robot ego–centric vision data to our probabilistic localization framework. In free and static environments, vision will give precise results, while sound would take over in case of visual obstructions, or within dynamic or visually complex environments. Second, for now, we have used our system in a single office environment. Sensor models for odometry and microphone arrays are dependent of this particular setting. In future work, we will evaluate how these models have to be adapted in different rooms, with different configurations (e.g. furniture, people), that could lead to sound obstructions or echos.

Finally, our long term plan is to have humans, virtual humans and robots interacting within an IIE. For this, we will use our system to track humans and various robots at the same time. We will have to evaluate the robustness and scalability of our probabilistic localization framework in these more challenging environments.

6. ACKNOWLEDGEMENTS

This research, which is carried out at BeingThere Centre, is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

7. REFERENCES

[1] JJ Leonard, "Mobile robot localization by tracking geometric beacons," *Robotics and Automation*, 1991. 1

[2] S Thrun, W Burgard, and D Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents series)*, The MIT Press, Aug. 2005. 1

[3] EB Martinson and F Dellaert, "Marco Polo Localization," in *Proceedings of IEEE International Conference on Robotics and Automation.*, pp. 1960–1965, 2003. 2

[4] R Munguia, "Single sound source SLAM," *Progress in Pattern Recognition*, 2008. 2

[5] VN Vapnik, "Statistical Learning Theory," *Wiley-Interscience*, 1998. 2

[6] D Fox, S Thrun, and W Burgard, "Particle filters for mobile robot localization," *Sequential Monte Carlo Methods in Practice*, pp. 401–428, 2001. 2