INTEGRATED DISPLACEMENT TRACKING AND SOUND LOCALIZATION

Parham Aarabi, Qing Hua Wang, Masoud Yeganegi

The Artificial Perception Laboratory, University of Toronto
{parham,wangq,yeganeg}@ecf.utoronto.ca

ABSTRACT

An algorithm for the robust localization of a vehicle using both displacement tracking and sound localization is proposed. The displacement tracking is performed by optical encoders that enable the turn angle and the movement distance of the vehicle to be estimated. The sound localization utilizes a speaker mounted on the vehicle and an array of 24 microphones deployed in the environment. The two modalities are integrated by modeling the displacement tracking uncertainty by a Gaussian Mixture Model (GMM) and combining it with the probability distribution obtained from the sound localization system. It is shown that the proposed integrated system results in an average localization error (at best 11cm) that is better than either modality alone.

1. INTRODUCTION

The precise and robust localization of an autonomous vehicle is often necessary for a variety of applications. These applications include finding the location of a rover on a distant planet, a robot in a home, or an automated military reconaissance vehicle. To this end, numerous localization algorithms have been proposed in the past, some of which use cameras, radio beacons, or landmarks [2, 5, 3]. Each technique has a set of corresponding advantages and disadvantages. Camera based systems, for example, suffer from their high computational complexity and failure in non-ideal situations (e.g. when an obstacle is blocking their view). Radio beacons are more reliable than camera based systems but suffer from the need for extra hardware both onboard the vehicle and in the environment.

Often, multiple data modalities can be combined to better localize and track vehicles. An example of such a system is described by [2]. The integration or fusion of multiple techniques employing different modalities can be the key to robust and accurate vehicle localizations.

In this paper, we make use of the onboard sensors of a small robotic vehicle to track the displacements of the robot in a 6m by 4m environment. While this displacement tracking results in only a rough estimate of the final location of the robot (and is affected by friction, obstacles, etc.), it is used in conjunction with a probabilistic sound localization algorithm. The goal of the combination is to improve the robustness and the accuracy of the localizations.

2. VEHICLE MOTION UNCERTAINTY MODELING

The vehicle used in this paper has optical encoders on each wheel that allow for controlled rotation and movement. When the vehicle turns and moves in any direction, however, there is always uncertainty about its final position. This uncertainty arises as a result of errors in the distance traveled as well as the angle of the initial turn.

While the actual probability distribution of these errors is a very complicated function of vehicle dynamics, a simplified model is used in this paper. It is assumed that the radial error of the vehicle Δr is linearly proportional to the distance traveled r, that the rotation error $\Delta\theta$ is linearly proportional to the angle turned θ, and that the final position error is only a combination of these two errors.

Such an error distribution corresponds to an uncertainty arc with a radius from $r - \Delta r$ to $r + \Delta r$ and an arc angle from $\theta - \Delta\theta$ to $\theta + \Delta\theta$, as shown in Figure 1. While this distribution could be modeled in a variety of direct ways, a GMM is used. The number of mixtures increase with the rotation error $\Delta\theta$ and the motion distance r, and decrease with the radial standard deviation of each mixture σ. As a result, a total of $\lceil rc_1\Delta\theta/\sigma \rceil$ radially symmetric Gaussians with a standard deviation of $\sigma = \Delta r = c_2 r$ are used, with uniform distribution along the arc using $\Delta\theta = c_3\theta$.

The values for the constants c_1, c_2, and c_3 were experimentally determined to be 1, 0.16, and 0.16 respectively, by performing several experiments with different r and θ values. These values were obtained by trial and error.

In this paper, we use $P(\mathbf{x}|r, \theta, \mathbf{x_0}, \theta_0)$ to represent the probability distribution of the vehicle location over the two-dimensional space $\mathbf{x}$ given its previous position $\mathbf{x_0}$, its previous orientation θ_0, its rotation θ, and its distance of motion r. According to the model defined in this paper, this

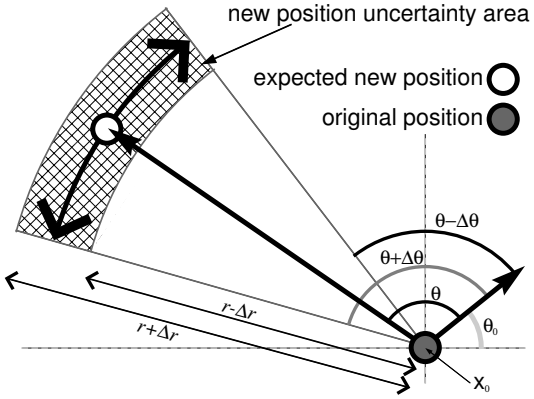

Fig. 1. The vehicle location uncertainty region due to displacement and rotation errors.

probability distribution can be defined as:

$$P(\mathbf{x}|r,\theta,\mathbf{x_0},\theta_0) = \frac{1}{\lceil rc_1\Delta\theta/\sigma \rceil} \sum_{i=1}^{\lceil rc_1\Delta\theta/\sigma \rceil} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\|\mathbf{x}-\bar{\mathbf{x}}_i\|^2}{2\sigma^2}} \quad (1)$$

where $\bar{\mathbf{x}}_i$ is the spatial mean or center of the ith mixture. The centers are evenly distributed on an arc of radius r from $\theta - \Delta\theta$ to $\theta + \Delta\theta$.

3. ACOUSTIC VEHICLE LOCALIZATION

The modeling of the uncertainty in the vehicle location by itself does not improve the localization accuracy. In this paper, the speech signal generated by the vehicle (using its 1.6m high speaker) and recorded by an array of 24 microphones is used to provide extra information about the vehicle's location with the aim of improving the accuracy once combined with the information of the previous section. Figure 2 illustrates the setup of the environment.

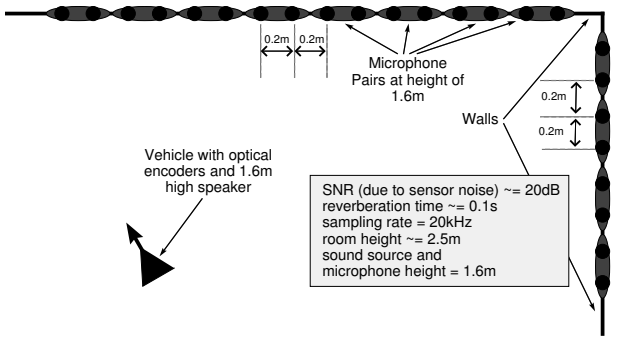

Fig. 2. Environmental setup of the microphone array.

The localization system utilizes a modified version of the SRP-PHAT algorithm [4] which uses Time Delay of Arrival (TDOA) histograms [1]. With the modified algorithm, TDOA histograms are computed for different pairs of microphones by taking repeated Phase Transforms (PHAT) [6, 4, 1] for several consecutive 20ms time-segments, with the maximizing time-delay for each PHAT being incorporated into the histogram. The maximizing PHAT time-delay τ between microphones n and m for the signal of time segment k is defined as:

$$\tau_{n,m,k} = \arg\max_{\beta} \int_{-\infty}^{\infty} \frac{X_{m,k}(\omega)\overline{X_{n,k}(\omega)}}{\left|X_{m,k}(\omega)\overline{X_{n,k}(\omega)}\right|} e^{-j\omega\beta} d\omega \quad (2)$$

where $X_{m,k}(\omega)$ and $X_{n,k}(\omega)$ are the Fourier Transforms of the kth 20ms signal segment recorded from the mth and nth microphones, respectively. In practice, each 20ms signal is obtained by sampling the continuous-time microphone signal and then windowing the samples by half-overlapped Hanning windows (and assigning an integer index number k to each segment with $k = 0, 1, 2, ...$). The frequency representation of the finite-duration and discrete-time signal is obtained by performing a Fast Fourier Transform (FFT), resulting in discrete frequency components. As a result, the integral of equation 2 is in practice a summation over the discrete FFT frequencies.

Assuming that for a given localization a total of K time-segments are available, then for microphones m and n the TDOA histogram can be defined as follows:

$$h_{m,n}(\tau) = \mathrm{hist}([\tau_{m,n,0}\ \tau_{m,n,1}\ \tau_{m,n,2}\ \ldots\ \tau_{m,n,K}], \tau) \quad (3)$$

where the $\mathrm{hist}(\mathbf{t}, \tau)$ function is a histogram operator (i.e. counting the number of TDOA estimates that fall within a finite set of preset bins) for the TDOA vector $\mathbf{t}$ and bin center τ.

Now, a given location $\mathbf{x}$ has a set of TDOAs corresponding to each microphone pair. We can pre-calculate the TDOA $\Omega_{m,n}(\mathbf{x})$ between microphones m and n corresponding to position $\mathbf{x}$ using $\Omega_{m,n}(\mathbf{x}) = (\|\mathbf{x_m} - \mathbf{x}\| - \|\mathbf{x_n} - \mathbf{x}\|)/\nu$, where $\mathbf{x_m}$ and $\mathbf{x_n}$ are the spatial locations of the mth and nth microphones, respectively, and ν is the speed of sound in air (approximately 345m/s).

In order to compute the likelihood of a speaker at position $\mathbf{x}$, we sum up the histogram values at the TDOAs corresponding to $\mathbf{x}$ using $\psi(\mathbf{x}) = \sum_m \sum_n h_{m,n}(\Omega_{m,n}(\mathbf{x}))$, where $\psi(\mathbf{x})$ is a spatial likelihood function (SLF) representing the likelihood of a speaker at each point in space. By normalizing the SLF according to $f(\mathbf{x}) = \psi(\mathbf{x})/\sum_{\mathbf{u}} \psi(\mathbf{u})$ we obtain a pseudo probability distribution representative of the probability of the speech source being at location $\mathbf{x}$ given $\mathbf{X}$ and Φ (i.e. $P(\mathbf{x}|\mathbf{X}, \Phi)$) where $\mathbf{X}$ represents the entire data collected from all microphones for each localization and Φ is the event that the signal-to-noise ratio (SNR) is high enough such that the sound localization data would be correct. Using this, and by observing that if the SNR is not

strong enough then the sound localization data would not be providing any new information (i.e. $P(\mathbf{x}|\mathbf{X}, \overline{\Phi}) \approx P(\mathbf{x})$) we can define $P(\mathbf{x}|\mathbf{X})$ as follows:

$$P(\mathbf{x}|\mathbf{X}) = f(\mathbf{x}) \cdot P(\Phi) + P(\mathbf{x}) \cdot (1 - P(\Phi)) \quad (4)$$

For the described sound localization process, only microphone pairs that were 60cm apart or less were used to form pairs. It was experimentally determined that for greater inter-microphone distances, the localization accuracy improvements would not be significant. Furthermore, it was experimentally determined that the accuracy of the proposed sound localization algorithm was slightly better than the SRP-PHAT technique, and as a result the former algorithm was chosen for implementation.

4. INTEGRATION OF DISPLACEMENT TRACKING AND ACOUSTIC LOCALIZATION

In order to combine the ideas of the previous two sections, we need to merge the two probability distributions $P(\mathbf{x}|\mathbf{X})$ and $P(\mathbf{x}|r, \theta, \mathbf{x_0}, \theta_0)$ to obtain $P(\mathbf{x}|r, \theta, \mathbf{x_0}, \theta_0, \mathbf{X})$. The merged distribution can be obtained using the assumption that, given $\mathbf{x}$, the two localization methods are independent (i.e. $P(\mathbf{X}, r, \theta, \mathbf{x_0}, \theta_0|\mathbf{x}) = P(r, \theta, \mathbf{x_0}, \theta_0|\mathbf{x})P(\mathbf{X}|\mathbf{x})$) and can be simplified using equation 4, as shown below:

$$\begin{aligned} &P(\mathbf{x}|r, \theta, \mathbf{x_0}, \theta_0, \mathbf{X}) \\ &\quad = P(\mathbf{x}|\mathbf{X})P(\mathbf{x}|r, \theta, \mathbf{x_0}, \theta_0)\tfrac{P(\mathbf{X})P(r,\theta,\mathbf{x_0},\theta_0)}{P(\mathbf{x})P(r,\theta,\mathbf{x_0},\theta_0,\mathbf{X})} \\ &\quad = (f(\mathbf{x}) + \alpha)\, P(\mathbf{x}|r, \theta, \mathbf{x_0}, \theta_0)\beta \end{aligned} \quad (5)$$

where $\alpha = P(\mathbf{x})\frac{(1-P(\Phi))}{P(\Phi)}$ is a constant that only depends on the physical parameters of the environment and the microphone arrays (and independent of $\mathbf{x}$ assuming $P(\mathbf{x})$ is constant) and $\beta = \frac{P(\Phi)P(\mathbf{X})P(r,\theta,\mathbf{x_0},\theta_0)}{P(\mathbf{x})P(r,\theta,\mathbf{x_0},\theta_0,\mathbf{X})}$ is a positive constant independent of $\mathbf{x}$ (again assuming that $P(\mathbf{x})$ is constant). Since we only care about the relative values of the merged probability distribution, any positive constant scaling of it can be ignored. Hence, in practice, β can be ignored and α can be precomputed for a given room/application. As a result, our overall spatial likelihood function (OSLF) $\Psi(\mathbf{x})$ can be defined as:

$$\Psi(\mathbf{x}) = (f(\mathbf{x}) + \alpha)\, P(\mathbf{x}|r, \theta, \mathbf{x_0}, \theta_0) \quad (6)$$

which is the fusion equation used in this paper.

For example, consider the localization of the vehicle in Figure 3. The probability distribution for the vehicle location derived from the acoustic localization is shown in Figure 3(top left), the distribution from the GMM uncertainty modeling in Figure 3(top right), and the combined distribution using $\alpha = 0$ in Figure 3(bottom left). The probability distribution peaks for each technique are illustrated in Figure 3(bottom right).

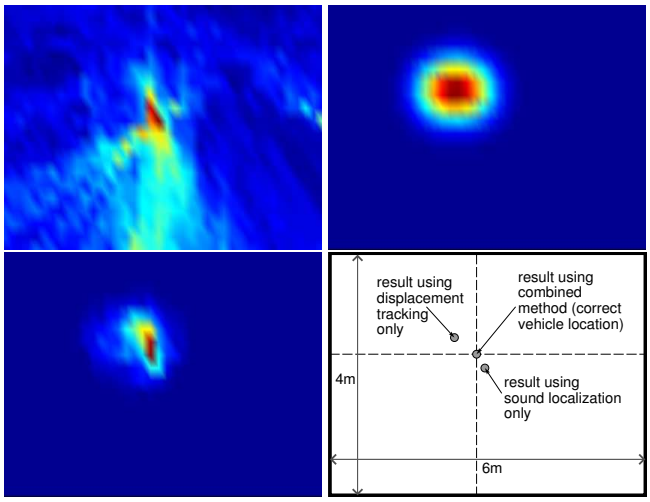

Fig. 3. Localization example with acoustic source location probability distribution (top left), displacement tracking modeled using GMMs (top right), combined SLF (accourding to equation 6) (bottom left), and the environment setup (bottom right). Here $\theta = 180^o$ and r =2m.

5. EXPERIMENTAL RESULTS

The parameter α was analyzed in a series of experiments in which the vehicle made a rotation of 180^o followed by a forward movement of 2m, as shown in Figure 4. The vehicle then produced a pre-recorded sound from its speakers which triggered its localization using the microphone array. The resulting acoustic spatial likelihood function, obtained from the array, was then integrated with the displacement tracking probability distribution of the vehicle, obtained from knowledge about the initial position and direction as well as the displacement distance and turn angle. The peak of the combined OSLF was then selected as the estimated vehicle position, and was compared to the position estimates obtained by either displacement tracking or sound localization alone.

Three scenarios were tested, with 10 trials per scenario (the resulting errors were averaged over all 10 trials). In the first scenario (Figure 4(top left)), a 2s speech signal was produced by the vehicle with no competing noise source (only sensor noise was present). The combined algorithm had an error of about 11cm (at $\alpha = 0$) compared to 16cm for sound localization only and 56cm for displacement tracking alone (Figure 4(bottom left)).

In the second scenario (Figure 4(top middle)), a person with an intense voice spoke at a location far away from the final position of the vehicle (the signal-to-noise ratio of the vehicle sound to the speaker was approximately -20dB). At $\alpha = 0$, the combined technique had an error of 26cm compared to 68cm for displacement tracking only and 274cm

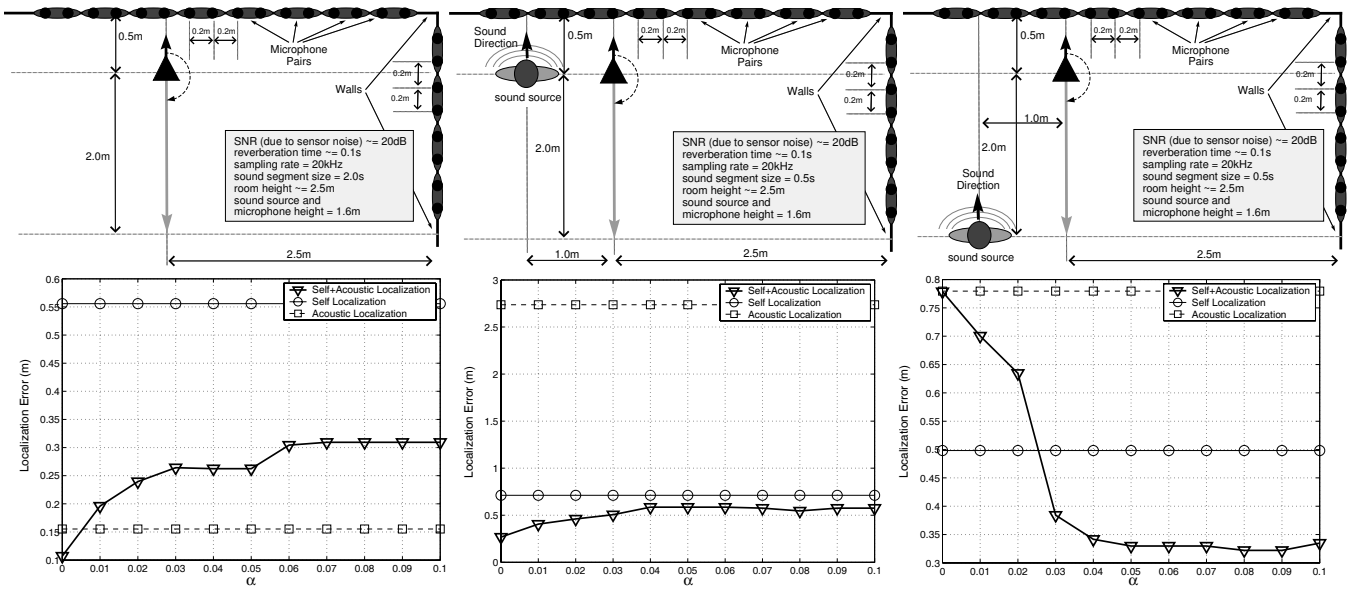

Fig. 4. The effect of varying alpha values on localization error with no noise present (left), noise present far away from the Gaussian mixture centers (middle), and noise present close to the Gaussian mixture centers (right).

for sound localization only (Figure 4(bottom middle)). The reason that, in this noisy case, integration with $\alpha = 0$ results in the lowest localization error is that by multiplying the two probability distributions, the sound localization distribution is ignored except in the vicinity of the Gaussian mixture centers. As a result, a noise source far away from these centers would not result in any significant error in the combined algorithm for all values of α.

The third and final scenario (Figure 4(top right)) involved a person intensely speaking close to the final location of the vehicle (thereby being in the vicinity of the Gaussian mixture centers). In this case, as shown in Figure 4(bottom right), the noise source does significantly increase the error of the combined algorithm for $\alpha < 0.03$. For $\alpha \geq 0.03$, however, the combined technique has a much lower average error than either the sound localization or the displacement tracking alone.

6. CONCLUSIONS

An integrated vehicle localization system using displacement tracking and sound localization was proposed. The proposed technique was tested in a real, noisy, and reverberant environment using three different scenarios. In light of the three experiments, and given that the location or presence of the noise source cannot be controlled, a value of 0.03 was chosen for α for the previously described environment. This value may not be optimal for every situation, but it does provide robust results in all scenarios.

The accuracy of the proposed technique is similiar to that of other vehicle localization techniques. For example, the system of [5] which utilized a radio transponder had an average localization error that was at best 23cm, compared to at best 10cm for [3], and at best 11cm for [2] which fused the results of localization with cameras and lasers.

7. REFERENCES

[1] P. Aarabi. The fusion of distributed microphone arrays for sound localization. *EURASIP JASP Spec. Iss. on Sensor Networks*, 2003:4:338–347, 2003.

[2] Clerentin et al. Cooperation between two omnidirectional perception systems for mobile robot localization. In *IEEE Inter. Conf. on Intel. Robots and Systems*, volume 2, pages 1499–1504, October 2000.

[3] Dellaert et al. Using the condensation algorithm for robust, vision-based mobile robot localization. In *IEEE CVPR'99*, volume 2, pages 588–594, 1999.

[4] DiBiase et al. Robust localization in reverberent rooms. *Brandstein and Ward (eds.), Microphone Arrays:Signal Proc. Techniques and Applications*, 2001.

[5] Kantor et al. Preliminary results in range-only localization and mapping. In *IEEE Inter. Conf. on Robot. and Automat.*, volume 2, pages 1818–1823, May 2002.

[6] C. H. Knapp and G. Carter. The generalized correlation method for estimation of time delay. *IEEE Trans. on ASSP*, ASSP-24(4):320–327, August 1976.