2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Using Sound Reflections to Detect Moving Entities Out of the Field of View

Nagasrikanth Kallakuri, Jani Even, Yoichi Morales, Carlos Ishi & Norihiro Hagita
Intelligent Robotics and Communication Laboratory
Advanced Telecommunications Research Institute International, Japan

***Abstract*— This paper presents a method for detecting moving entities that are in the robot's path but not in the field of view of sensors like laser scanners, cameras or ultrasonic sensors. The proposed system makes use of passive acoustic localization methods which receive information from occluded regions (at intersections or corners) because of the multipath nature of sound propagation. Contrary to the conventional sensors, this method does not require line of sight. In particular, specular reflections in the environment make it possible to detect moving entities that emit sound such as a walking person or a rolling cart. This idea was exploited for safe navigation of a mobile platform at intersections. The passive acoustic localization output is combined with a $3D$ geometric map of the environment that is precise enough to estimate sound propagation and reflection using ray casting methods. This gives the robot the ability to detect a moving entity out of the field of view of the sensors that require line of sight. Then the robot is able to re-calculate its path and waits until the detected entity is out of its path so that it is safe to move to its destination. To illustrate the performance of the proposed method, a comparison of the robot's navigation with and without the audio sensing is provided for several intersection scenarios.**

I. Introduction

When reaching an intersection, it is common to adjust our pace if we hear the footsteps of someone coming towards the intersection, even if we do not see that person. In a robotic context as illustrated in Fig.1, this would mean to use the sound information coming from the occluded region (the *blind region*) to detect an incoming person and slow down or stop until the person passes. However, as of today, robots do not use such information and usually get information only from the *visible region* since they rely on sensors that require line of sight (cameras, lasers, ultrasonic sensors). Thus, in the scenario depicted in Fig.1, the robot would proceed at normal speed towards the intersection resulting in a possible crash with the person. This problem is particularly critical for an autonomous passenger vehicle, like a robotic wheelchair, as the passenger can hear the person coming at the corridor intersection. Then the fact that the robot does not show any sign of handling the presence of an incoming person, creates a stressful situation for the passenger.

The complexity introduced by the multipath propagation of sound partly explains why the use of line of sight sensors

This research was funded by the Strategic Information and Communications R&D Promotion Programme (SCOPE) under the Ministry of Internal Affairs and Communications, Japan.
Email: {srikanth, even, yoichims, carlos, hagita}@atr.jp

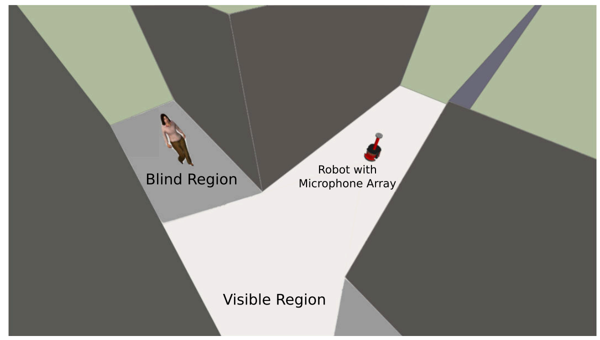

Fig. 1: Example scenario where audio sensors supplement the visual sensors for obstacle detection.

is usually preferred for detecting obstacles. However, in the proposed scenario, it is this multipath propagation of sound that gives information about the presence of an approaching person in the blind region. Using the sound reflected by the environment, we are able to localize that person.

This every day scenario is the motivation for the work presented in this paper. The goal is to give the robot the ability to detect an incoming person in the blind region. To achieve this goal, first the robot has to be able to detect the reflections and estimate their direction of arrival (DOA). This is done using a microphone array and running a steered response power (SRP) algorithm [1] which is a passive acoustic localization method. Then the robot should be able to find the sources of the received reflections. To achieve this, we use ray tracing to follow the sound propagation paths (backwards from receiver to emitter). Since we deal with sounds coming from the blind region after one or several reflections in the environment, the robot must not only know the positions of the reflectors but also its own pose (position and orientation) in the environment. With the recent advances in $3D$ Light Detection and Ranging sensors (LIDAR), it is possible to build a precise $3D$ map of the environment [2], [3] and use it to estimate the reflections. The pose of the robot is estimated by a particle filter based localization algorithm [4]. By combining these techniques, the robot is able to estimate a set of acoustic rays coming from a source in the blind region. The position of the approaching person is then estimated by considering the intersections of these rays.

In practice, the robot is able to detect any sound emitting object, not just an approaching person as in Fig.1. To illustrate the effectiveness of the proposed method, several scenarios at different intersections were investigated. In these scenarios, the behavior of the sound aware robot (proposed approach) is compared to the behavior of a robot using only laser range finders.

II. RELATED WORKS

The use of lasers [5] or RGB-D cameras [6] to detect and track moving entities by range estimation is obviously limited to the field of view. Ultra sonic sensors that have been used for robot perception for several years estimate the range of a reflecting object by measuring the time of flight of an emitted ultrasonic pulse [7]. The proposed method differs from the ultrasonic approach as we do not emit any sound but rely on passive sensing. In this sense, the reflections used by the ultrasonic sensor to determine the ranges of objects are fundamentally different from the ones we use to detect the presence of moving objects in blind regions. In particular, an ultrasonic sensor usually considers the line of sight to the object that reflects the pulse. Consequently, systems that use an active sensing approach (ultrasonic) and rely on direct field of view to the obstacle to perform robot localization [8] or obstacle detection [9] are not able to gather information about the blind region as our approach does.

In robot audition, it is common to use sound localization method on board a robot [10], [11]. The goal can be the detection of people for human robot interaction [12], human activity recognition [13] or detection of sound sources for mapping [14], [15]. However, to the best knowledge of the authors, none of the proposed approaches exploit the reflections of sound. On the contrary, while mapping, the robot movement is used to suppress the reflections from the map [14], [15].

An important issue in virtual acoustics is to model the acoustical effects of an environment. This effect is strongly related to the multipath propagation of the sound and particularly to the early reflections (the sound that reaches the listener after a few reflections). Consequently, several methods were proposed to model these early reflections. The early methods exploiting ray tracing, like [16], leading to the *image source method* [17] that computes the reflection of a sound source to a surface using a virtual sound source. The virtual source is obtained by mirroring the sound source relative to the surface. These reflections are often referred as *specular reflections*. In the remainder of this paper, we simply use the term reflection to refer to the specular reflections. However, outside of the field of virtual acoustics, these reflections are usually considered a nuisance. For example a great deal of the speech enhancement literature deals with the suppression of reflections and reverberation (see [18] and references herein).

Recently, the authors in [19], presented sound source localization method that take advantage of the reflections. In that approach, first an echo localization method is used to estimate the positions of the walls and ceiling in a meeting room. Then these positions are used to take into account the direct path of the sound and a reflection on one wall and another on the ceiling while performing a maximum likelihood based localization of the participants during a meeting. In [20], the authors use reflections on surfaces (measured beforehand) and direct paths to perform sound source localization. These methods differ from the proposed method that relies on the actual 3*D* geometric map of the environment. Moreover their aim is not to detect sources out of the field of view.

III. PROPOSED APPROACH

A. Overview

The method proposed in this paper uses a multi modal sensing approach for mapping and safe navigation. First the mapping is performed by combining the data from a 3*D* laser scanner with the odometry information from the robot. Then the resultant 3*D* map is used for audio reflection analysis during navigation of the robot. Figure 2 shows the block diagram of the proposed method. The different blocks are explained in the remainder of this section.

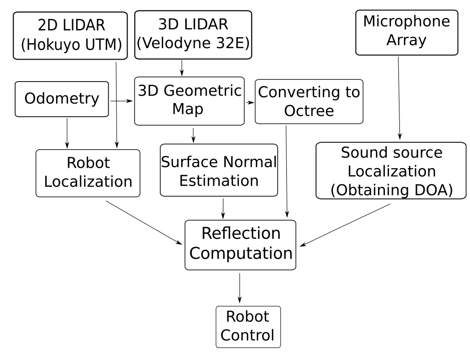

Fig. 2: System Framework.

In this paper, the audio modality is used to detect moving objects out of the field of view of the robot at intersections. The assumptions are that these moving objects are emitting sounds and that these sounds reflect in the environment.

Based on the sound propagation property, the reflected sounds reach the microphone array on board the robot. Consequently, at intersections, the on board audio localization estimates the direction of arrival (DOA) of these incoming reflected sounds.

Knowing the DOA of an incoming sound and the pose of the robot, it is possible to use ray tracing to find the point of the environment from which the received sound originated. If we are able to determine the properties of the reflector, particularly the normal to the reflective surface, it is then possible to compute the DOA (at the reflection point) of the sound that created the reflection. Iterating with ray tracing and reflection computation procedure, an *audio ray* can be determined from the robot position to the sound source, see Fig.5.

Note however that with one ray it is not possible to determine the number of reflections and the exact position of the sound source. For this reason, the position of the sound source is estimated by computing several audio rays and determining their intersection. This requires that the audio scan presents several local maxima each corresponding to a different reflection of the same sound source. However, this condition is not always met in practice and the proposed method has to take into account this indeterminacy.

In addition to the multipath propagation created by the reflections, sound is also attenuated and scattered. In this paper, we do not consider diffraction or scattering. Because of diffraction, the lower frequencies of the sound of an incoming moving object may curve around the corner of an intersection. However, in our experiments, clear occurrences of diffraction were not observed and the local maxima detected by the SRP method corresponded to reflections.

B. Sound source localization

The DOA of the sounds that are impinging the microphone array are estimated by a SRP algorithm (see [1] and references there-in). In this paper, the SRP method is based on a delay and sum beamformer. The processing is done in the frequency domain after applying a short time Fourier transform (STFT) to the observed signals sampled at 48kHz (the analysis window is 25 ms long and the shift of the window is 10 ms).

After the STFT, at the sampled time t, the observed signals from the Q microphones at the frequency f are denoted by $U_1(f,t), \cdots, U_Q(f,t)$. The phase transform is applied to the frequency domain signals prior to processing the beamforming by taking $U_i(f,t) := U_i(f,t)/|U_i(f,t)|$ as it results in an SRP that is less sensitive to reverberation [1].

Because the geometry of the microphone array is precisely known, it is possible to steer the array using spatial beam forming to estimate the sound from a spatial location (described by the spherical coordinates $\{\rho, \theta, \phi\}$ in the array's frame of reference). The beamforming output at the frequency f is denoted by

$$S(f,t,\{\rho,\theta,\phi\}) = \frac{1}{Q}\sum_{q=1}^{Q} e^{-2j\pi f\tau(\{\rho,\theta,\phi\},q)} U_q(f,t), \qquad (1)$$

where $\tau(\{\rho,\theta,\phi\},q)$ is the delay at the qth microphone for the location $\{\rho,\theta,\phi\}$ (the microphone 1 is the reference $\tau(\rho,\theta,\phi,1) = 0$).

The SRP algorithm estimates the powers of the beamformer output for a set of N locations. We assume the far field condition holds (ρ large compared to the array aperture). Thus a set of N locations is defined by the angles $\{\theta_n, \phi_n\}$ with $n \in [1,N]$. The search grid, see Fig.3, is composed of equilateral triangles (the angle between two summits is approximately 3^o). The power is computed in the frequency band $[1000, 5000]$ Hz corresponding to the discrete frequencies f_{1000} and f_{5000} by taking

$$P(t,\{\theta_n,\phi_n\}) = \frac{1}{\#F}\sum_{f=f_{1000}}^{f_{5000}} |S(f,t,\{\theta_n,\phi_n\})|^2$$

where $\#F$ is the number of discrete frequencies in the band. Time averaging is applied to combine $L = 10$ STFT frames

$$J_n(k) = \frac{1}{L}\sum_{t=k-L}^{k} P(t,\{\theta_n,\phi_n\})$$

where k is the index corresponding to the blocks of L frames. Namely the kth audio scan is a set of N angles $\{\theta_n, \phi_n\}$ with their associated power $J_n(k)$. The frequency of the audio scan is 10 Hz ($L = 10$ STFT window with a shift of 10 ms). The DOAs of the signals impinging the microphone array are obtained by performing a local maxima search in the audio scan. For example, the audio scan in Fig.3 clearly exhibits two local maxima.

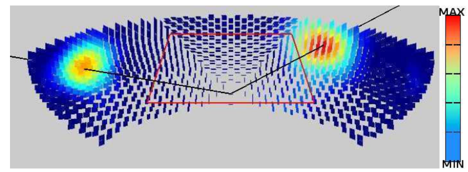

Fig. 3: Audio scan and ray casting at local maxima

C. Building 3D geometric map of the environment

One of the novel ideas in the proposed approach is the use of actual geometric map of the environment to perform audio reflection analysis. In order to obtain a 3-Dimensional geometric map of the environment, a Simultaneous Localization and Mapping (SLAM) based approach was used. SLAM based mapping techniques have been well studied, matured and implemented. To obtain the 3D Map a Velodyne 32E Laser range finder (LRF) was used to obtain the range information, a Pioneer robot platform equipped with wheel encoders was used to obtain the odometry data. The robotic platform used for mapping the environment is shown in Fig.6. To build the map, the robot was navigated in the environment with a joystick while the odometry and the 3D laser sensor information were recorded. Then we used an iterative closest point (ICP) based SLAM to correct the trajectory of the robot. To align the laser sensor scans the *3DToolkit* library framework [2], [3] was used. With the resulting aligned scans an occupancy grid map was created [21], [22]. The cell resolution of the map is set to $5cmX5cmX5cm$.

After obtaining the 3D map by using the ICP based SLAM, this map was converted into an octree using the *Octomap* framework [23]. The 2D Geometric map and the corresponding octomap for one of the test environments is shown in Fig.4.

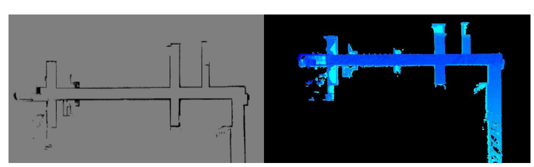

Fig. 4: $2D$ Environmental geometric map (left) and the corresponding $3D$ Octomap (right).

D. Robot localization

We used a particle filter based approach to localize the robot with a weighted set of 100 particles. We used the ray casting approach likelihood model [4] to compute the particle weights. Particle re-sampling was regularly performed, and the robot pose is given by the average weight of the particles. The map update depends on the state of the particle dispersion and the matching of the laser scans.

E. Ray Tracing in the 3D Geometric Map and reflection computation

As explained in Section III-B the beam forming was carried out using the SRP algorithm. Since the direction of interest is in front of the robot, the audio scan was limited to the front side of the robot. The audio scan is limited in yaw θ_n and pitch ϕ_n based on the requirement. We consider the scan with in $\theta_{\max}$ and $\theta_{\min}$ in yaw and $\phi_{\max}$ and $\phi_{\min}$ in pitch in the robot centric coordinate system. The SRP gives the audio power distribution in the microphone array referential frame. This data is transformed to the robot referential using the pose of the array with respect to the robot. The exact pose (position and orientation) of the robot in the geometric map is obtained from the particle filter based localization algorithm as explained in Section III-D. Combing the pose of the robot and the result of the SRP process we get the audio power distribution in the global reference in the vicinity of the robot.

For ray tracing, in the first step all the local maxima values above a particular threshold were considered. The yaw and pitch angles of each of the local maxima are obtained. Rays are casted in the $3D$ geometric map from the precise position of the microphone array in the global coordinate system. An image showing the yaw and pitch range considered and the local maxima occurrence during test runs is shown in Fig. 3.

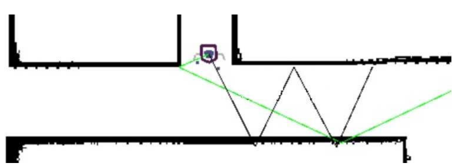

Fig. 5: Ray casting in the geometric map (the actual ray casting is performed in the $3D$ map).

The ray casting was performed in the map which was represented in an octree format. As mentioned above, octomap framework was used for operations and manipulations involving the octree maps. For a given coordinate in the map (*origin* for the ray casting) and a direction vector ($\vec{d}$), a ray can be casted from the origin and propagated in the direction of $\vec{d}$ until the ray hits an occupied cell in the map, as shown in Fig.5. The cell which is hit by the ray is considered as *obstacle*. Once the output of ray casting (*obstacle*) is obtained, the ray needs to be reflected from the obstacle in order to ray trace further. The normal direction at the obstacle was obtained from a predetermined lookup table made by off line surface normal estimation. For all the points in the 3D geometric map, the normal directions are calculated before hand in-order to aid the real time performance of the system. The normal vectors are estimated using the point cloud library [24] which uses the Principal Component Analysis (PCA) of a covariance matrix created from the nearest neighbors of the point where the surface normal is being estimated.

The direction of the reflected ray ($\vec{r}$) is obtained from the incidence ray ($\vec{i}$) propagating from *origin* to *obstacle* and the surface normal vector ($\vec{n}$) by taking $\vec{r} = \vec{i} - 2\,(\,\vec{n}\,.\,\vec{i}\,)\,\vec{n}$

Now, the *obstacle* becomes the new *origin* for the next level of ray casting and the direction vector is the defined by the reflection angle obtained $\vec{r}$. Again the process of ray casting is repeated to obtain where the new ray from the *obstacle* hits the map. This process is continued up to a predefined number of reflections from walls. In our case we have considered ray tracing until 4 reflections owing to the error accumulation over the number of reflections.

This ray tracing in the $3D$ geometric map gives an estimate of how the sound from the source traveled before reaching the microphone array. Assuming that there is only one active moving source, all the rays from the robot will have one common point of intersection which, ideally, should represent the source of the audio received at the microphone array on the robot.

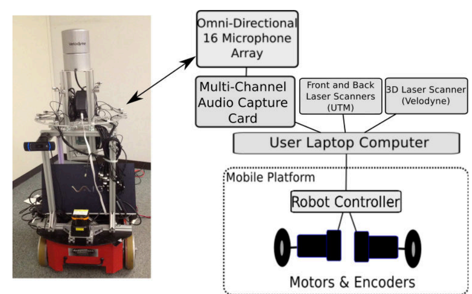

Fig. 6: Robot platform used for experiments.

The rays from different local maxima orientation (representing different reflections of the sound from the source) are traced down up to a predefined number of reflections. Let us assume each audio scan gives X local maxima with value $J_n(k)$ greater than p. A ray is traced in each of these X maxima directions. We consider these X rays in pairs and calculate the points of intersections of each pair. Since the audio received at the microphone array is generated from the point of contact of feet of the person and the ground, ideally all the X rays are supposed to have one common point of intersection at this point. Practically, there may not be a proper point of intersection of these rays, since they are projected in $3D$ and the reflecting surfaces present irregularities. But, these rays must pass close to each other. So we consider the closest points of the ray pairs and plot them in the geometric map. These points of intersection gives the direction from which the source is coming. Using this information, the robot understands the situation and performs corresponding action for safe navigation.

IV. Evaluation

A. Experimental setup

For experimental validation of the proposed approach, we selected corridor intersections where the LRFs on board the robot are blind beyond the turning at the intersection. Two different types of corridor intersections - one a 'T' junction and the other 'X' junction were considered. All the walls at the intersections are made of concrete and are mostly smooth planes. The ceiling is made of a porcelain material. The ambient noise at the location of experiment is approximately $42dB$ due to a centralized air conditioner system. Since our intention was to observe sounds generated by foot steps of people and by cart wheels we considered the range of angles for yaw and pitch as $\theta_{\min} = -75^\circ$ to $\theta_{\max} = 75^\circ$ in yaw and $\phi_{\min} = -45^\circ$ to $\phi_{\max} = 15^\circ$ in pitch in the robot centric coordinate system. The threshold value for local maxima was set to 28 on an exemplary basis.

The platform shown in the Fig.6 is used for the experiments.

The scenario for the experiments is explained as follows.
1) The robot plans a path from the start position to the destination using a global planner.
2) The robot starts moving along the planned path at $0.5m/s$.
3) When the robot is near the corridor (approximately 4-5 meters from the intersection), it starts using the audio data to identify obstacles beyond the line of sight.
4) If the SRP is higher than the threshold value of 28, the robot slows down to $0.2m/s$.
5) When the robot is about 2.5 meters from the corridor intersection, if the SRP still presents high values, the reflection analysis is performed to obtain the direction of the moving object.
6) The robot analyzes to see if the moving obstacle might come in its planned path. In such a case, it waits till the moving object reaches a position which is safe for the robot to proceed. Otherwise, it continues towards destination at $0.5m/s$.

B. Experimental results

The proposed approach helped the robot navigate safely to its goal by avoiding any collisions in all test runs. The experiments were conducted with walking people and people pushing an indoor cart. As explained in Section IV-A, the velocity of the robot is the quantity of interest in the proposed scenarios. The robot displayed intelligent behavior by slowing down when it listens to some reflections of sound from corridor branches. If a moving entity is identified in the direction of goal point, the robot stops and waits until the entity crosses the intersection.

Fig.7 (a) to (f) present six different runs covering three scenarios. The trajectory of the robot is color coded with the velocity and important velocity changes are indicated on the figures. The first scenario was with no moving entity coming towards the intersection of corridor, the second was with a person walking from left to right in the corridor in the robot referential and third scenario with a person pushing an indoor cart from right to left. Apart from the results shown, the experiment was repeated in different combinations of directions and the moving entity being a walking person or a cart and the robot performed as expected in all the cases. Also, the results are found to be mostly similar in both 'T' type intersection and 'X' type intersection of the corridors.

C. Discussion

The robot successfully navigated to the goal position by anticipating the moving entities that might suddenly appear at the corridor intersection. The robot was able to figure out the branch from which a moving entity was coming towards the intersection. In this paper, the focus is on performing the reflection analysis in real time and controlling the robot. To achieve this goal, a relatively coarse search grid was used for the SRP method ($\approx 3^o$). Consequently the precision of the audio ray casting is degraded after a few reflections (as each reflection doubles the angular error $6^o, 12^o, 24^o, \cdots$). As a result, the estimation of the incoming entity's position is not so accurate. However, the precision used in all the proposed scenarios was sufficient for the robot to behave as planned. Namely, the approaching entities were successfully detected in the branch of the corridor they were actually in.

The noise from the robotic platform is approximately in the range of $52dB - 60dB$ depending on the speed at which it moves. But this noise is mainly from the left and right side of the robot at the height of wheels where the motors are present. Since our audio scan is limited to the front side of the robot, this ego-noise is mostly ignored.

The proposed approach proved to be quite robust to environmental noise as in the T corridor, a strong fan noise was emitted by an electric appliance situated in the robot branch. To increase the robustness of the proposed method, a possible extension would be to use an *audio map* as the ones generated in [14], [15] that gives the positions of the fixed noise sources in the environment. Note that the use of such an *audio map* would be necessary if there is a fixed sound source beyond the intersection of the corridors to avoid a dead lock condition where the robot waits indefinitely to pass through the intersection.

V. CONCLUSIONS

In this paper, we proposed a method based on the analysis of reflected sounds to enhance the perception of a mobile robot at corridor intersections. In particular, by using the audio modality, the robot is able to detect an incoming entity in the blind region of its visual and laser sensors. Using this framework, a safer navigation was made possible by anticipating the motion of that entity. The robot could reach the destination safely utilizing the information about the moving entity given by the audio reflection analysis. Recent advances in $3D$ sensing and mapping techniques made it easy to combine the geometry of the environment with audio information to perform reflectivity analysis.

The focus of the future research is to precisely localize the position and to obtain the velocity of the incoming entities while they are still in the blind region of the line of sight sensors. Also, it is left for future work, the evaluation of the system towards different types of environment configurations and types of surfaces. Also, the integration of the system towards an existing sound source mapping framework [15] to suppress environmental noise can be explored. Finally, it is of our interest to test the limits of the approach running on-line on a mobile robot to build a multi-modal human tracking

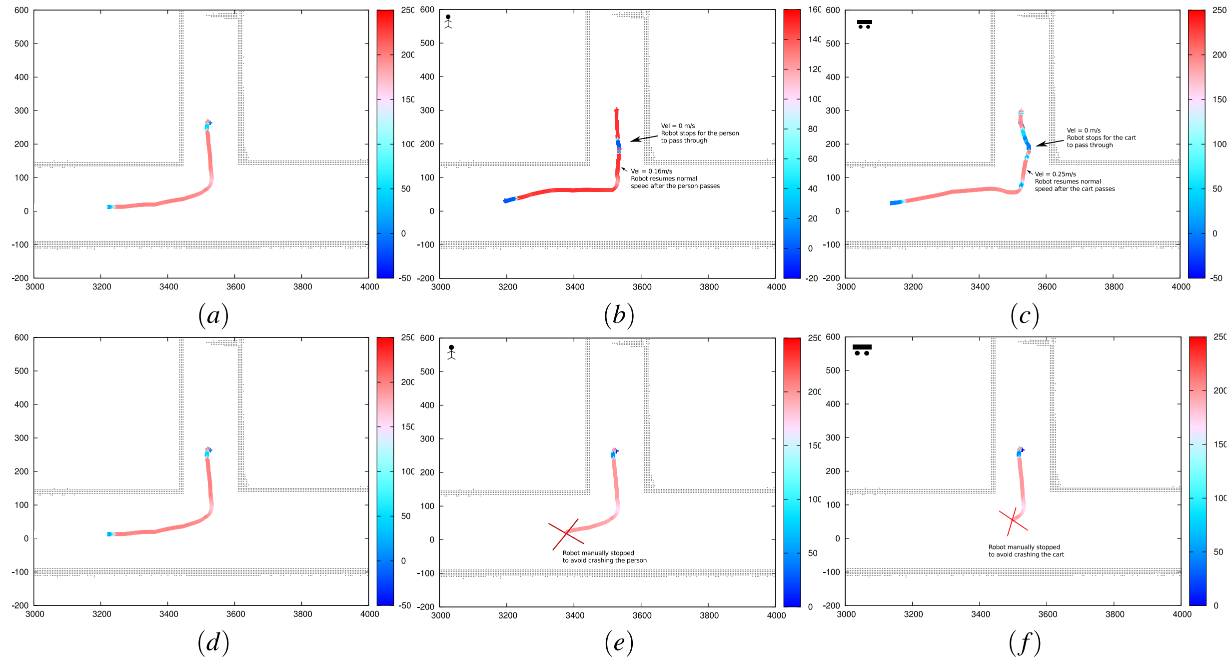

Fig. 7: Comparison of robot navigation with (top row) and without (bottom row) audio reflection analysis. The trajectory of the robot is color coded with velocity at that location. Without any active sources (left column), with a person walking from left to right (middle column), and with a cart moving from right to left (right column). X-axis and Y-axis represent global location in Cms. The start point is $(3500, 300)$ and goal point is $(3200, 0)$. The color coding is based on velocity values in mm/sec.

system combining data from LIDAR, microphone array and an environmental map.

References

[1] M. Brandstein and H. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," in *ICASSP 1997*, 1997, pp. 375–378.

[2] D. Borrmann et al., "The Efficient Extension of Globally Consistent Scan Matching to 6 DoF," in *Proceedings of 3DPVT '08*, Atlanta, USA, 2008, pp. 29–36.

[3] slam6d, "Slam6d - simultaneous localization and mapping with 6 dof," Retrieved December May, 20 2011 from http://www.openslam.org/slam6d.html, 2011.

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

[5] J. H. Lee et al., "People tracking using a robot in motion with laser range finder." Ieee, 2006, pp. 2936–2942.

[6] L. Spinello and K. O. Arras, "People detection in rgb-d data." in *Proc. of The International Conference on Intelligent Robots and Systems (IROS)*, 2011.

[7] T. Yata, A. Ohya, and S. Yuta, "A fast and accurate sonar-ring sensor for a mobile robot," in *Proceedings of the 1999 IEEE International Conference on Robotics and Automation*, May 1999, pp. 630–636.

[8] J. H. Lim and J. J. Leonard, "Mobile robot relocation from echolocation constraints," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1035–1041, 2000.

[9] I. Abdul Aziz, S. Mahamad, M. Mehat, and N. Samiha, "Blind echolocation using ultrasonic sensors," in *Information Technology, 2008. ITSim 2008. International Symposium on*, vol. 4, Aug., pp. 1–7.

[10] K. Nakadai et al., "An open source software system for robot audition hark and its evalation," in *IEEE-RAS International Conference on Humanoid Robots*, 2008, pp. 561–566.

[11] A. Badali et al., "Evaluating real-time audio localization algorithms for artificial audition on mobile robots," in *Proceedings of IROS 2009*, 2009, pp. 2033–2038.

[12] J.-M. Valin et al., "Enhanced robot audition based on microphone array source separation with post-filter," in *in Proceedings of IROS 2004*, vol. 3, sept.-2 oct. 2004, pp. 2123 – 2128 vol.3.

[13] J. Stork et al., "Audio-based human activity recognition using non-markovian ensemble voting." in *Proc. of RoMan 2012*, 2012.

[14] E. Martinson and A. C. Schultz, "Auditory evidence grids." in *IROS*. IEEE, 2006, pp. 1139–1144.

[15] N. Kallakuri, J. Even, Y. Morales, C. Ishi, and N. Hagita, "Probabilistic approach for building auditory maps with a mobile microphone array," in *To appear in Proceedings of 2013 IEEE International Conference on Robotics and Automation, ICRA 2013*, 2013, pp. –.

[16] U. Krockstadt, "Calculating the acoustical room response by the use of a ray tracing technique," *J. Sound and Vibrations*, vol. 8, no. 18, pp. 118–125, 1968.

[17] J. Allen and D. Berkley, "Image method for efficiently simulating small room acoustics," *J. of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[18] J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*. Springer-Verlag, 2005.

[19] F. Ribeiro, D. Ba, C. Zhang, and F. D., "Turning enemies into friends: using relections to improve sound source localization," in *Proceedings of 2010 IEEE International Conference on Multimedia & Expo, ICME 2010*, 2010, pp. 731–736.

[20] C. Ishi, J. Even, and N. Hagita, "Using multiple microphone arrays and reflections for 3d localization of sound sources." Ieee, 2013, pp. –.

[21] H. Moravec and A. E. Elfes, "High resolution maps from wide angle sonar," in *Proceedings of the 1985 IEEE International Conference on Robotics and Automation*, March 1985, pp. 116–121.

[22] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46–57, June 1989.

[23] A. Hornung et al., "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, 2013, software available at http://octomap.github.com.

[24] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9-13 2011.