Localizing Victims Through Sound and Probabilistic Grid Maps in an Urban Search and Rescue Scenario

Holger Kenn[1] and Andreas Pfeil[2]

[1] Technologie-Zentrum Informatik, Universität Bremen, Germany
kenn@tzi.de
http://www.tzi.de/
[2] School of Engineering and Science, International University Bremen, Germany
a.pfeil@iu-bremen.de
http://www.iu-bremen.de/

Abstract. Sound source localization can be used in the Robocup Rescue Robots League as a sensor that is capable to autonomously detect victims that emit sound. Using differential time of flight measurements through energy cross-spectrum evaluation of the sound signals, the angular direction to multiple sound sources can be determined with a pair of microphones for SNRs better than -8dB. Assuming that the robot pose is known, this information is sufficient to create probabilistic occupancy grid map of the sound sources in the environment and thus localize the victims in a global map. This has been demonstrated using example measurements in an urban search and rescue scenario.

1 Introduction

Urban search and rescue robotics has established itself as a new field for teleoperated and autonomous robots in the recent years. One of the major advances of the field has been the establishment of regular competition-based benchmarks such as the Robocup Rescue Robots league. ([2],[3],[4]) As the competition is using a scoring system that is adapted to new challenges each year, the focus within the competition can be shifted slowly from teleoperated systems with a main focus on mechanical engineering and mobility issues to more advanced systems with limited autonomy up to fully autonomous systems. Currently, most teams rely on teleoperation and visual feedback through onboard cameras. In the last year, mapping of the environment has become a standard capability of many teams, but victim localization is still done by the operator by using the visual feedback from the teleoperated robot. The development of sensors capable of automatic victim identification, localization and state assessment is therefore the next step towards fully autonomous rescue robots.

This paper presents a novel approach for an inexpensive victim localization sensor based on a stereo microphone pair.

The IUB Robocup Rescue team has been competing in the Robocup Rescue Robot League competitions in Fukuoka [5] and Padova [6].

All disaster scenarios set up at these competitions contain a number of victim dummies that have human appearance and are equiped with detectable features such as movement, sound, body heat and CO_2 emission. The performance of each team is evaluated through a common scoring function that analyzes the quality of the information

A. Bredenfeld et al. (Eds.): RoboCup 2005, LNAI 4020, pp. 312–322, 2006.

gained and scores it against the number of human operators used by the team. This function thus both rewards autonomy and high-quality localization and multi-sensoric assessment of victims and their state.

During the 2002 competition, it became clear that the onboard autio sensor was useful within the competition as it was the only sensor that was capable of locating invisibly trapped or entombed victims. This finding has been one of the reasons for the extension of the robots with additional sensors with similar capabilities, such as body-heat detection through thermographic cameras.

With the introduction of the new control middleware FASTRobots [7] in the 2003 robot system, it became possible to add more sensors to the robot platform, first for non-victim related tasks such as localization and environment mapping. The generated LADAR-based map provided to be useful for robot localizaion [8]. However, as no automatic victim localizing sensor was available, the localization and identification of the victim dummies was still performed manually by the operator. In order to do this, the operator would carefully analyze all available sensors including the sound from the microphones, then note down the perceived signs of the presence and state of the victim in a paper victim sheet and then mark the victims location by using a mouse to click on the approximate position of the victim next in the LADAR map that is displayed on his operator control station screen together with the robots current position and orientation. This process is time-consuming and error-prone.

For automatic victim localization with bitmap sensors such as the visible light and thermographic cameras, computer vision based approaches may be used. However, these approaches are computationally expensive and their performance in the highly unstructured environment of a robocup rescue scenario is hard to predict.

The sensor described in this paper is capable of localizing a sound source in a global probabilistic occupancy grid map of the environment. These sources still have to be manually inspected and eventually identified as victims using the onboard cameras and other sensors but as their location is known now, their localization in the global map will be much more precise. The approach can easily be distributed over multiple robots, as no strict timing requirements for the acquisition of the sound data is required. To allow this, it is assumed that the sound sources stay at fixed positions in the environment.

The remaining part of this paper is structured as follows: The second section gives an overview over the system setup and an introduction into the theory of sound source localization. The third section describes an experiment to estimate the performance of the obtained sound source localization and shows that the measurements obtained are close to the theoretical boundaries. These results are then used to simulate the performance of a probabilistic map based on a occupancy grid. Then, an experiment is described that uses an existing robot to gather data in a rescue scenario to produce such a probabilistic map. The last section discusses the results obtained so far.

2 System Overview and Theoretical Analysis

A typical robot system used for Robocup rescue consists of one or more mobile robots and one or more operator control stations. For the IUB Robocup Rescue system, the communication between the mobile robots and the control station is implemented through the FAST-Robots middleware[7]. Each mobile robot runs an instance of the

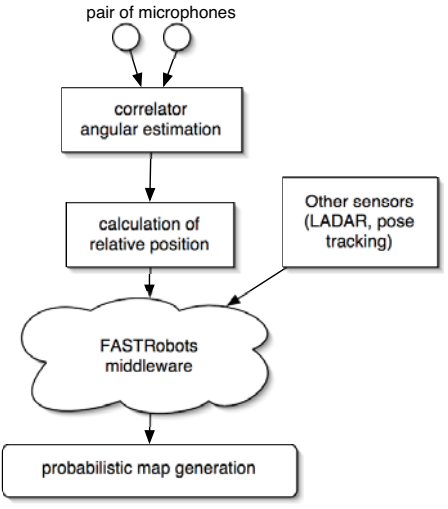

Fig. 1. An overview of the sound source localisation system

platform component of the framework that instantiates multiple sensor and actuator driver objects that communicate with the corresponding sensor and actuator hardware. The platform communicates with its counterpart, the controlstation component running on the control stations through a TCP/IP network. The controlstation visualizes sensor data coming from the platform and transmits control commands to the platform.

The sensor described in this paper can easily be accomodated by this framework. Figure 1 shows an overview of the components that are part of the sound localization sensor system.

2.1 Microphone Phase Array

The problem of sound source localization can be solved in different ways. One approach is sound source localization based on beamforming techniques, such an approach has for example been presented in [1]. However, our approach is using the cross-energy spectrum of signals recorded at microphone pairs to evaluate the sound source directions. This is computaionally less expensive for a small number of microphones and allows for easier detection of multiple sound sources. This approach is explained in the following paragraph.

A simple way to model the localisation of a sound source (sometimes also called "passive sonar") by using multiple microphones is the so-called linar microphone phase array. In this model, a number of microphones are located equidinstant along the x axis of our coordinate system. It is then possible to determine the position of a sound source in the coordinate system using differential time-of-flight measurements, i.e. the time difference for the signal of the same sound source to arrive at different microphones. This system however cannot detect the correct sign of the y coordinate, i.e. it cannot distinguish between sound sources in positive or negative y direction. When the array is limited to only two microphones, the position of the sound source can only be restricted

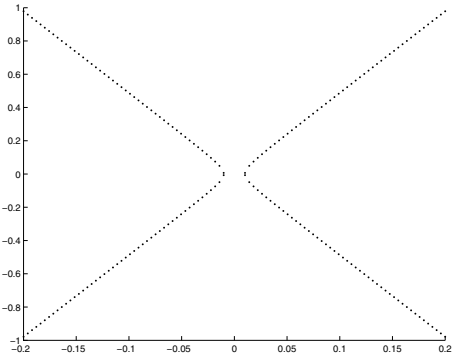

Fig. 2. Two hyperbolas indicating the possible location of the sound source for a given time-of-flight difference δt and $-\delta t$

to a hyperbola given by the path length difference of the sound signals from the source to the microphones. These hyperbolas can be approximated by their asymptotes for sound sources that are further away than the distance of the microphones, i.e. the direction to the sound source can be determined.

To determine the time delay between two incoming sound signals at the microphones, the cross-energy spectrum of the two signals is evaluated. Identical, but shifted signal portions produce peaks in this spectrum; the position of the peak is an indicator for the delay between the first and the second occurance of the same signal portion in the different signals. Should there be several distinct sound sources with different relative delays, one peak for every sound source can be detected. For the remainder of this section, it will be assumes that only a single sound source is being localized. It will be shown later that with multiple sound sources can be dealt with using probabilistic occupancy grids.

In order to process the signal from the microphones, it is sampled with the highest possible time resolution that the hardware offers. For the standard audio interfaces of PCs, this is typically 48khz, i.e. 20.8 microseconds between two samples. This consequentially is the shortest delay Δt_{min} that the system can distinguish. Together with the speed of sound c, this results in a quantization of the differential distance measurements into $c\Delta t_{min}$.

As the distance difference δ is quantized, the angular resolution of the microphone pair detector significantly differs for different angular areas. Angles near the direction of the normal vector, i.e. "in front" of the microphone pair, can be measured with a fine resolution and angles in the direction of the line connecting the microphone pair, i.e. outside of the 'focal region' can only be measured with a high uncertainty.

2.2 Angular Resolution

The resolution of the localization is strongly depending on the resolution of δ. Given a sampling frequency f_s and the speed of sound c, the maximal time difference of k

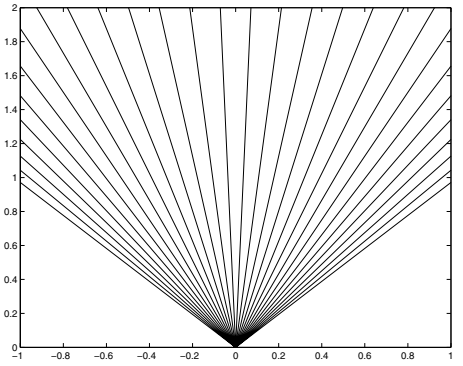

Fig. 3. The angular sectors that can be distinguished by one microphone pair 10cm apart, using 48kHz sampling frequency

samples is reached, when a signal is coming from the x-axis outside of the microphone pair. It is:

$$k = \frac{m \cdot f_s}{c} \tag{1}$$

δ varies from $-k$ to k samples.

Evaluating the angles for all possible k, the half-circle in front of the microphone pair can be divided into different zones, a sample resolution is shown in figure 3.

With several angular measurements from different positions, the position of the sound source can be determined through triangulation.

Assuming that the sound source is immobile, these angular measurements can as well be done sequentially by one robot only. The robot needs to measure at one point, move for a precise distance and measure the angle again. Using both angles and the base length, triangulation can be performed.

Assuming that the current pose of the robot is known, the system has sufficient information to create a probabilistic occupancy grid map [12] of the environment in a world coordinate system. Unlike the occupancy grid map used for robot navigation[8], this map does not contain information about the probability of cells being occupied by obstacles but with the probability of cells being the location of a sound source.

This type of map has been chosen over other approaches to probabilitstic mapping ([9],[10],[11] or see [13] for an excellent overview of the topic) as we assume that it is hard to extract features from the sensory input that could be redetected in the future. Moreover, the location of sound sources will not provide much structure as the location of walls in an office environment would give us. As this sensor is not intended for robot self-localization but only for sound source localization, it is assumed that an accurate estimation of the current pose of the robot is provided by other means. Note that this information could be provided through other means of probabilistic mapping and localization such as SLAM[11], but this mapping would then use other sensors such as LADAR.

The probabilistic map building algorithm is implemented in a straight-forward way: For every grid cell a value is calculated that represents its change in probability of being a sound source based on the current sensor data and this value is added to the current value stored for the grid cell.

The calculation of this change in probability does not only depend on the current sensor value but also on the properties of the sensor, i.e. on a sensor model. Here we assume that the sensor only gives good information for sound sources that are neither too faint (i.e. far away) nor are outside of the focus area where angular information is unreliable. Information concerning these areas is ignored.

If a sound source is located within the focus area of the sensor, its signal energy level is compared against a threshold T. If there is no signal higher than the treshold, the angular area reaching from the robot to a constant maximum reliability distance D is considered free of sound sources and every cell that has its centerpoint in this angular area receives a negative probabiltity change $-\Delta$. If a sound source is detected in the angular area, every cell receives a positive probability change Δ. The probability values in the cells are then updated and limited to reasonable positive and negative maxima P_{MAX} and P_{MIN}.

Initially, all cells are initialized with a value of 0 that corresponds to a maximum of uncertainty for this cell, we neither know that it is a sound source nor we know that it is one.

It can easily be seen that the occupancy grid can solve the triangulation from two different robot poses provided that the sound source is within the detection range from both poses. If the robot is in the first pose A, it will increase the probability value of all cells between its location and its detection range in the direction of the sound source. All other cells within the detection range will receive a decrease in probability value. After a number of sensor readings are analyzed, the probability value for all cells between the robots current position and the sound source will converge to a value of P_{MAX} and all other cells of the grid will either remain 0 or will converge to P_{MIN}. If the robot is now moved to a pose B and if the sound source is still in the detection range of the robot, it will further increase the probability value in all cells in between of the current position of the robot and the sound source and it will decrease the probability value for all cells that are not in the direction of the sound source, thus the probability value of all cells in the proximity of the sound source will remain at P_{MAX} and all other cells will either converge to P_{MIN} or remain 0.

Unfortunately, a sector that has received a positive probability from pose A and is not in the detection range from pose B will remain with P_{MAX} probability value. However, this value is misleading as it only depends on a single measurement and therefore is not a true triangulated value. These sectors would lead to false positives, i.e. the detection of a sound source when there is none. In order to eliminate these false positives, additional measures have to be taken. A true triangulation consists of two measurements that use different angular directions to establish the triangulation. To distinguish true triangulations from false positives, the robot taking the measurement and incrementing the probability value in a cell additionally computes an angular sector ID in world coordinates. This angular sector ID is an integer that numbers the angular sectors of the semicircle from 0 to AS_{MAX} so that every direction gets a distinct ID. If a robot finds a different sector ID in the grid cell it is about to increment, it sets a flag in the cell indicating that it contains the result of a true triangulation.

This algorithm uses a number of parameters. The parameters that specify the size of the distinguished angular areas are determined by the geometric properties and the

sampling frequency of the sensor. The treshold energy T and the reliability distance D parameters are dependent on the properties of the transmission system formed by the sound sources to be detected, the transmission medium and the microphones. The Parameters Δ, P_{MAX} and P_{MIN} determine the number of iterations that are needed for convergence. Additionally, the model described here assignes the same probability value increase to all grid cells in a sector. This does not reflect the real probabilities as the sector becomes wider when the cells are further away from the sensor. Consequently an individual cell that is further away should receive a linearily lower probability increase than a cell that is close to the sensor, but the simulations have shown that for rather small angular sensors, a fixed value is a reasonable approximation.

3 Experimental Results

The perfomance of the whole system was evaluated using a combination of simulations and measurements.

First, the performance of the angular detector in the presence of (white-gaussian) background noise was simulated and it was concluded that the system is rather immune to this kind of disturbances, provided that the SNR at the receiver is above $-8dB$.[1] Then, the predicted angular resolution of the sensor was verified in an experiment. In this experiment, a mobile sound source and the microphone pair were set up on a desk. In this setting, all angular sectors could be detected correctly. (see [14]).

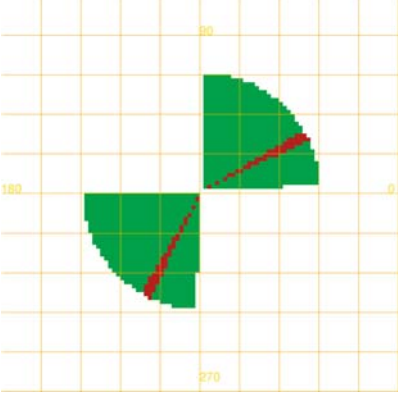

Fig. 4. One of the zones that can be distinguished by the sensor

3.1 Probabilistic Mapping

In order to to estimate the performance of the sensor in a probabilistic grid map, a sensor model has been derived from the data gained so far. The sensor model has a number of

[1] Altough white-gaussian background noise is far from realistic, it is a good test pattern for this situationas it has minimal cross-correlation, any correlated noise would result in additional sound sources being detected with only minor impact on the detection of the primary source.

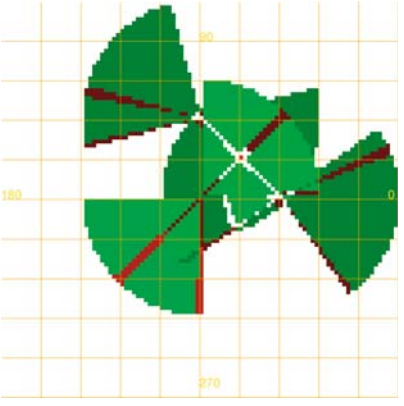

Fig. 5. A simulation of three sensor readings of a single sound source with a simple sensor model

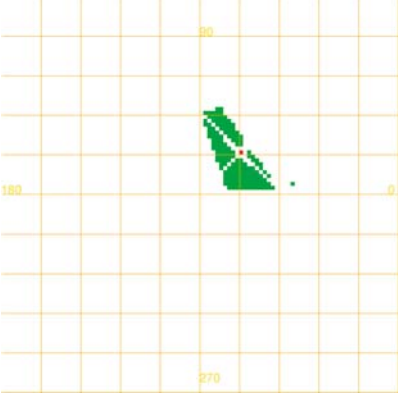

Fig. 6. A simulation of three sensor readings of a single sound source with a better sensor model using a touchcount filter

different zones that can be distinguished, each zone consisting of two angular sectors in positive and negative y direction as shown in Figure 4. To produce this figure, a sensor in the origin with a normal orientation of 45 degrees and 16 distinguishable angular sectors and a sound source at position x=2/y=1 was simulated. As the sensor cannot distinguish the exact position of the sensor in the zone, the probability of the presence in the zone is uniformly increased (red areas) and the probability of it not being in any other zone is uniformily decreased (green areas). The sensor is assumed to have a fixed range and for sources that are further away, it is assumed that the source is lost in the background noise, so it will not be detected. From the simulation result, it can be seen that a single sensor measurement is quite ambiguous.

In Figure 5, the simulation results for a sound source from three different sensor positions are shown. In this case, the sound source at position x=1/y=1 is clearly indicated with a positive probability. However, there are other parts of the map that receive

positive probability. This occurs due to the fact that these areas are only covered by a single sensor reading, so the probability is increased by the sensor model of that one sensor reading, but is never decreased by the model of another sensor reading. This sensor model is formally correct, as there could be indeed three independent sources that are each only detectable by a single sensor. However, it is much more likely that only a single source creates the sensor readings. Therefore, we add an additional counter to each cell of the probabilistic map that counts how many sensor readings have contributed to the final value of the cell. By comparing this value against a threshold and filtering the result by this, we obtain the simulation result shown in Figure 6. Here the sound source can clearly be distinguished as the single point with positive probability that remains.

3.2 Full Sytem Experiment

To test the performance of the mapping system under the intended working conditions, a robot was placed in a scenario with one sound source present (see Figure 7). It was driven into 4 different poses and the perceived sound recorded. The data was analyzed and fed to the mapping algorithm yielding the map shown in Figure 8. The position of the robot were obtained by measuring the marks after the robot had been removed, thus eliminating errors created by the self-localization system of the robot.

It is important to note that, due to a design difference in the sensor system, this robot is using a sensor base length of $m = 20cm$, which is wider than in the systems used for simulation and thus leads to 57 distinguishable sectors. Using this high number of sectors did not improve results, so the angular resolution was then reduced again by joining several neighboring sectors to produce results comparable to the simulation. Additionally, the sensor range in the sensor model was increased as it was found that the attenuation model in simulation was overestimating, so sound sources further away could be detected by the sensor.

On Figure 8, it can be seen that the region detected as sound source is much wider than in the theoretical simulation. This can be explained by the fact that the robot position is further away from the sound source than in the simulated case and that the

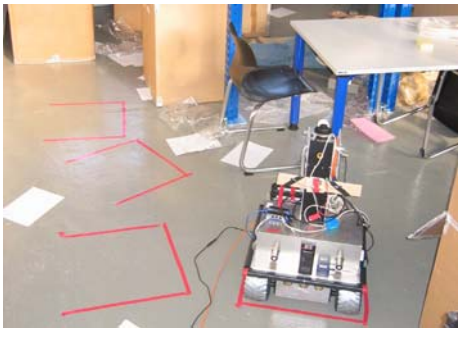

Fig. 7. An image of the scenario. The robot pose is marked with red tape, the open side of the rectangle being the back of the robot.

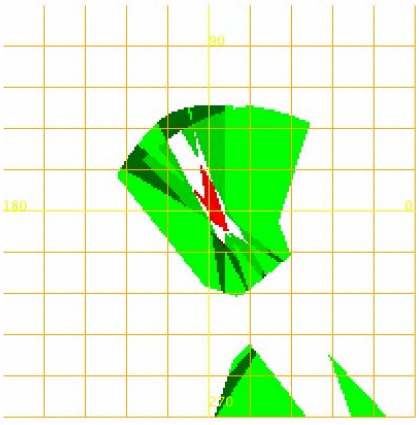

Fig. 8. The map created from real sensor readings. There is one source present and measurements have been taken from 4 different poses.

robots were all measuring the sound source from similar positions, this leads to a fuzzy and elongated localization. Additionally, it was found that the angular resolution of the sensor and the resolution of the grid map are important parameters to be tuned. In this case, a map resolution of 5 centimeters per grid cell and an angular resolution of 7.5 degrees produced the best results, for finer angular resolutions an more coarse grids, the sensor beams would not overlap sufficiently to allow for a good detection.

4 Conclusion

The problem of automatic victim localization in RoboCupRescue has been presented. A solution using microphones mounted on mobile robots and differential time-of-flight measurements of sound has been simulated and its accuracy shown to be sufficient in a simple experiment. A mapping algorithm using occupancy grids has been presented based on the experimental finding and it has been shown in simulation that is able to localize a sound source in a global map.

The next step will be the implementation of the sensor on a robot of the IUB Robocup Rescue team and the comparision of the simulation results with the real performance of the sensor. This will be an improvement over current victim localization techniques that are entirely based on human operators.

Acknowledgment

The measurements for this work have been made at the North-European Robocup Urban Search and Rescue Arena at the International University Bremen. The authors espechially thank Prof. Birk at IUB for his support.

References

1. Mattos, L., Grant, E.:Passive Sonar Applications: Target Tracking and Navigation of an Autonomous Robot. Proceedings of the IEEE International Conference on Robotics and Automation. IEEE Press, 2004.
2. Kitano, H., Tadokoro, S.: Robocup rescue. a grand challenge for multiagent and intelligent systems. AI Magazine 22 (2001) 39-52
3. Takahashi, T., Tadokoro: Working with robots in disasters. IEEE Robotics and Automation Magazine 9 (2002) 34-39
4. Osuka, K., Murphy, R., Schultz, A.: Usar competitions for physically situated robots. IEEE Robotics and Automation Magazine 9 (2002) 26-33
5. A. Birk, H. Kenn et al., The IUB 2002 Smallsize League Team, Gal Kaminka, Pedro U. Lima and Raul Rojas (Eds), RoboCup 2002: Robot Soccer World Cup VI,LNAI, Springer, 2002
6. A. Birk, S. Carpin and H. Kenn, The IUB 2003 Rescue Robot Team RoboCup 2003: Robot Soccer World Cup VII, LNAI, Springer, 2003
7. H. Kenn, S. Carpin et al., FAST-Robots: a rapid-prototyping framework for intelligent mobile robotics, Artificial Intelligence and Applications (AIA 2003), ACTA Press, 2003
8. S. Carpin, H. Kenn and A. Birk, Autonomous Mapping in the Real Robots Rescue League, RoboCup 2003: Robot Soccer World Cup VII, LNAI, Springer, 2003
9. McLachlan, G., Krishnan, T.: The EM Algorithm and Extensions. Wiley-Interscience, 1996
10. Kalman, R.: A new approach to linear filtering and prediction problems. Transactions of ASME. Journal of Basic Engineering 83, 1960
11. Dissanayake, G., Newman, P., Clark, S., Durrant-Whyte, H., , Csorba., M.: A solution to the simultaneous localisation and map building (slam) problem. IEEE Transactions of Robotics and Automation 17, 229-241, 2001
12. Moravec, H.P.: Sensor fusion in certainty grids for mobile robots. AI Magazine, 1988
13. Thrun, S.: Robot mapping: a survey. Technical Report CMU-CS-02-111, Carnegie Mellon University, 2002
14. Kenn, H., Pfeil, A.: A sound source localization sensor using probabilistic occupancy grid maps. Proceedings of the Mechatronics and Robotics Conference 2004, IEEE Press 2004