The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Search and Track Power Charge Docking Station Based on Sound Source for Autonomous Mobile Robot Applications

Ren C. Luo, Chien H. Huang, Chun Y. Huang

Abstract— **The objectives of this paper is to develop a system for an autonomous mobile robot for searching and tracking sound source based target. The target can be a power charge docking station. Finding objects or events by following a sound source direction is an intuitive response for human and animal when they cannot see the target. However, there are relative fewer discussions about sound source tracking in robotics field. We design a navigation algorithm integrated with time delay of arrival (TDOA) techniques based on cross-correlation. To confirm the practicality of our approach, we design a robot docking scenario and develop an autonomous mobile robot with two microphone pairs, a camera and a laser range finder to conduct the experiments. According to the experimental results, our approach reduces the complex localization problems to save computing power, and have an acceptable result in an indoor environment.**

I. Introduction

VISION is one of the most important sensing abilities for mobile robots. However, there are still limits for cameras, such as poor lighting condition or the interested object doesn't exist in the field of view.

In many cases, the events in which robots interested are accompanied by auditory information. For humans and animals, it is straightforward to follow the sound source until seeing the object or event. There is a product "anti-lost alarm" (Fig. 1) which fully takes advantage of above concept. "Anti-lost alarm" prevents losing children's (or important item's) location for people just by sounding to let them know its location. The idea came up with this product: why not equip our robot with acoustic sensors to find an object or event which emits sound signal?

In this article, we propose an approach for mobile robot to search target which can emit sound signal.

A. Scenario

To confirm the practicality of this idea, we design a scenario of robot finding docking station.

The experiments conducted in a domestic environment. At the beginning, the robot stands by in one room, and the docking station is outside the visual field of the robot, even in

Fig. 1. Anti-lost alarm

another room. When the robot starts to find the docking station, the docking station keeps sounding. Then the robot is able to estimate an approximate direction while searching. When the robot is close enough to see the docking station, it can starts to docking.

B. Background and Related Work

In order to find a sound source, the robot should be able to estimate the direction of the sound source, and then navigate toward it. In other words, we need to design a navigation algorithm based on sound source detection.

In robotics field, the topics of sound source detection and navigation algorithms have been studied in depth separately.

In aspect of sound source detection, the topic has been increasingly studied form 1990s. There are several famous projects about this topic such as SIG [1], HRP-2 [2].

Currently, the proposed methods can be divided into two categories. One is to collect sound wave data using microphone array and form the received data into covariance matrix, and process the matrix using beanforming theories. Typical algorithms in this category are MUSIC (Multiple Signal Classification) [3] and ESPRIT (Estimation of Signal Parameter via Rotational Invariance Techniques) [4]. The pros of these methods are high accuracy, but they cost a lot of computing power and have to measure the frequency response of each microphone in advance.

The other category is to estimate TDOA (Time Delay of Arrival) of two microphones, and calculate the sound source direction according to the geometric relation between

Manuscript received March 12, 2010.
Ren C. Luo is with the Electrical Engineering Department, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, Taiwan 106 (corresponding author to provide phone: +886-2-3366-9826; e-mail: rcluo@ntu.edu.tw).
Chien H. Huang is with the Electrical Engineering Department, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, Taiwan 106 (e-mail: r97921003@ntu.edu.tw).
Chun Y. Huang is with the Electrical Engineering Department, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, Taiwan 106 (e-mail: cyhuang@ira.ee.ntu.edu.tw).

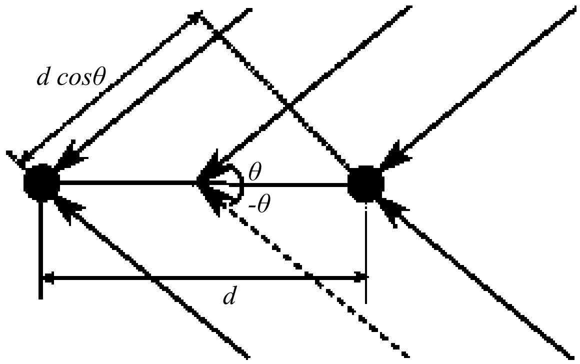

Fig. 2. Relative angle of microphone pair and sound source. Note that the estimation of TDOA cannot make out whether the sound signal comes form θ or $-\theta$.

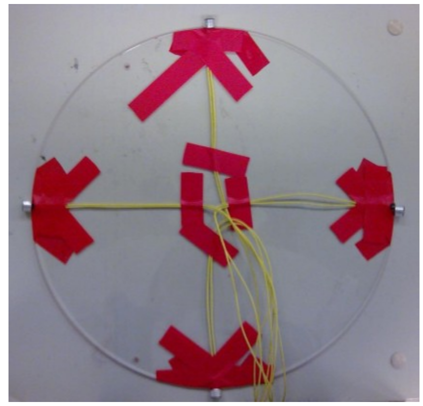

Fig. 3. Two microphone pairs.

microphones and sound source. The simplest way to estimate TDOA is to calculate the cross-correlation function between two sound wave signal that received by microphone pairs, and find the corresponding time that the maximum occurs. The method has less accuracy, but they are easy to implement. Above all, they reduce the computing power and have acceptable estimation results. Therefore, more and more robotics researches applied the method to detect sound source direction [5]-[9].

In aspect of navigation, the Bug algorithm family are well-known mobile robot navigation algorithms with proven termination conditions for unknown environments [10]. TangentBug algorithm [11] requires a robot with range sensors such a sonar or laser range finder. In an environment the wider space, TangentBug produces the shortest path because it can drive directly towards a vertex whereas other algorithms need to rely on wall-following.

However, there are relatively fewer works focusing on combining sound source detection and navigation algorithms [12]-[16].

Uchiyama, Yamamoto, Sano and Takagi [12] proposed a model reference control approach in which the robot follows a desired trajectory generated by a reference model. The approach they proposed combined sound source detection using TDOA and obstacle avoidance in aspects of robot dynamics. Our work differs in attempting to focus effort on integration of TDOA and navigation algorithm.

Huang [13] present an auditory navigation system, in which the sound localization method is based on a model of the precedence effect of the human auditory system. However, most of the work described the sound source localization and auditory system. Although there were experimental results showed that the robot could correctly localize the sound source and move toward sound source position avoiding the obstacle in between, they did not describe the navigation method.

In [14] a robot phonotaxis system is presented. The work focused on integration of biomimetic sound source localization apparatus with a taxis behavior coded in the motion description language MDLe. They emphasized the sensory-motor integration.

[15] and [16] present approaches to accomplish sound tracking. However, they use both in-room and robot-embedded microphone arrays simultaneously.

A. Paper Structure

This paper is organized as follows. Section II briefly explains the principle of sound source detection, and the TangentBug navigation method will be described in Section III. Section IV presents the algorithm that we proposed and the experimental results are shown in Section V. Finally, Section VI draws the conclusions.

II. Sound Source Detection

A. TDOA and Cross-Correlation

Assume there are a sound source and a microphone pair in the same space as Fig. 2 shows, the signal that received by the two microphones $\boldsymbol{x_1}$, $\boldsymbol{x_2}$ can be modeled as

$$\begin{cases} x_1(t) = s(t) + n_1(t) \\ x_2(t) = \alpha s(t-\tau) + n_2(t) \end{cases} \tag{1}$$

where $\boldsymbol{s}$ denotes the signal generated by the sound source, $\boldsymbol{\alpha}$ is an attenuation constant, $\boldsymbol{\tau}$ denotes the ***time-delay of arrival*** (TDOA) between microphones, and $\boldsymbol{n_1}$, $\boldsymbol{n_2}$ correspond to uncorrelated noise of each microphone. While the distance between microphones $\boldsymbol{d}$ is know and the velocity of sound wave $\boldsymbol{V}$ can be estimated, we can calculate the angle $\boldsymbol{\theta}$ by inverse cosine function

$$\theta = \cos^{-1}\left(\frac{\tau V}{d}\right) \tag{2}$$

if the delay time $\boldsymbol{\tau}$ can be obtained. The most common way to find $\boldsymbol{\tau}$ is to calculate the cross-correlation between $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ as

$$\tau = \underset{k}{\operatorname{argmax}}\, R(k) = \int_{-\infty}^{\infty} x_1(t)\, x_2(t-k) \tag{3}$$

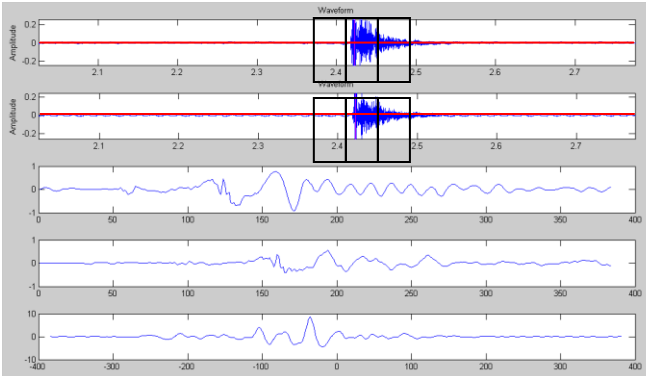

Fig. 4. Integration of VAD and cross-correlation to estimate sound source direction. The first and second rows are data collected from one microphone pair, and the average volumes are denote by red lines. The third and forth rows are the chosen frames (black rectangles) to process cross-correlation, and we can observe a time delay between them. The last row shows the result of cross-correlation, a maximum occurs at index -33.

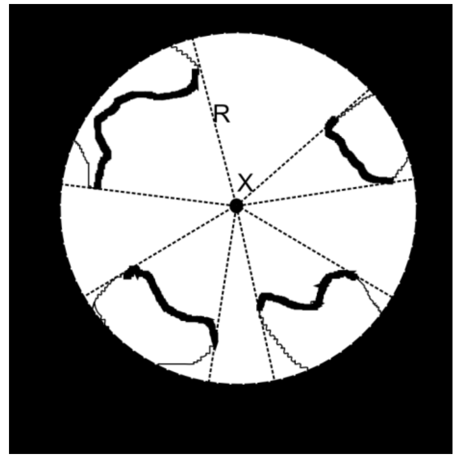

Fig. 5. A LTG example. Robot (X) equipped with a range sensor with a detection range R. The visible obstacle boundaries (bold) are modeled as thin walls.

In practice, the number of samples in a sound wave frame is limited, and the analog waveform is converted to digital data points, thus (3) should be modified as

$$\tau = \arg\max_{k} R[k] = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} x_1[t]\,x_2[t-k] \qquad (4)$$

where N denotes the number of samples in a wave frame. However, one microphone pair is not enough to get exact direction of sound source. In our work, two microphone pairs (Fig. 3) are used to find the exact direction of sound source.

B. Voice Activity Detection

However, if the sound source always emits sound signal, people will feel uncomfortable. In our work, the sound source plays a role as direction beacon; it should only sound when the robot acquires its location.

In order to detect the timing that a ***meaningful*** signal arrived, it is necessary to combine ***voice activity detection*** (VAD) mechanism with cross-correlation method. The topics about VAD have already been deeply discussed in the research field of speech recognition. It is usually used to monitor the speaking activity of incoming sound signal. We implemented a simple VAD skill that the robot just checks the volume magnitude in time domain. The way we integrated VAD and TDOA is described below.

1) Record a wave file, calculate the average volume and determine a threshold p according to the average volume (in our work, p is ten times of average volume).
2) Segment the wave file into subframes, and calculate the volume of each subframe.
3) Finally, search from the first subframe, if there is a subframe whose volume greater than p, combine the one and the previous and next ones into a new frame to process cross-correlation. If the subframes whose volumes exceed p are different between microphones, choose the earlier one.

The data processed in each procedure are demonstrated in Fig. 4.

III. TangentBug-Inspired Algorithm

We designed a simple navigation algorithm which is similar to a well-known robot navigation algorithm ***TangentBug*** [9].

The TangentBug algorithm usually produces a shortest path among Bug algorithm family when robots equipped with range sensors such as laser range finder or sonar. One primary feature of TangentBug is using local-tangent-graph (LTG), by which the algorithm can find a locally optional direction during the motion towards the target.

We adopted LTG in our navigation algorithm (Fig. 5). With a given target, robot can find a temporary target based on LTG according to obstacle distribution. The inputs of the algorithm are as follows.

1) The robot location $\mathbf{x}_{robot}$ and robot radius **R**.
2) The target location $\mathbf{x}_{target}$.
3) The LTG based on current range sensor data.

The algorithm is demonstrated in Alg. 2 and Fig. 6.

Alg. 2. Navigation Algorithm

1. If the **target is reached**, terminate the procedure.
2. Draw a virtual rectangle with height the segment joins $\mathbf{x}_{robot}$ and $\mathbf{x}_{target}$, and with width 2**R**.
3. If there is no obstacle inside the rectangle, move toward the given target. Terminate the procedure.
4. If there are any obstacles inside the rectangle, find the endpoints of their both sides. Select the endpoint that has a minor angle formed with $\mathbf{x}_{robot}$ and $\mathbf{x}_{target}$ as epS.
5. Draw a virtual line L containing and epS, and find a point which is apart 1.5R from epS on the normal line that passes epS to L. Set the point as a temporary target and move toward it. Repeat 1.

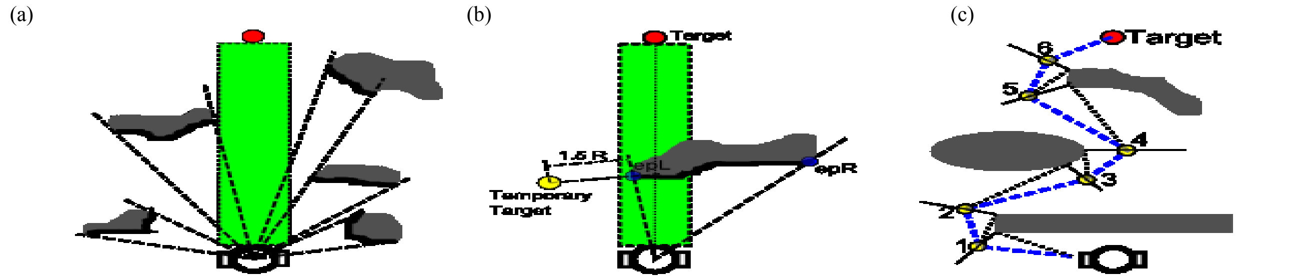

Fig. 6. The navigation algorithm. (a) In situations that no obstacles exit inside the virtual rectangle, move toward given target. (b) There is an obstacle inside the rectangle, the robot find a temporary target to move forward. The end points at both side of the obstacle are noted as epL and epR, and epL is chosen as a refernce point epS since the angle formed by $\mathbf{x}_{target}$, $\mathbf{x}_{robot}$ and epL is smaller than epR does. The temporary target is set at the normal line to L pass by epL, and the distance between epL and the temporary target is 1.5R as our parameter. (c) The path produced by our robot navigating to a given target by applying this algorithm.

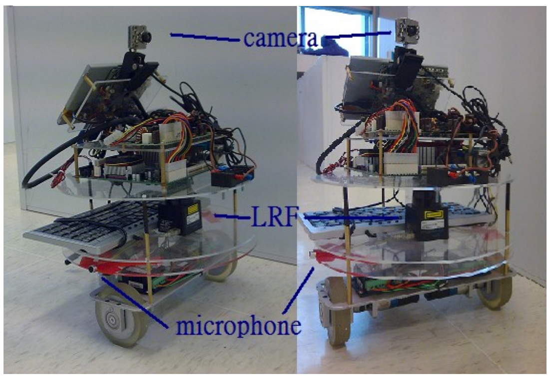

Fig. 7. Robot architecture.

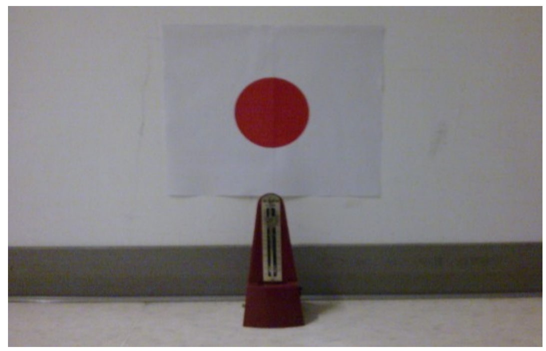

Fig. 8. The docking station.

IV. Proposed Approach and System Architecture

To integrate sound source detection with our navigation algorithm to achieve sound source searching, it is an intuitive idea to set a point in the determined sound source direction as the target in our navigation algorithm. Practically, some problems should be concern.

First, TDOA method can only provide direction information of the target, but not distance information. However, the navigation algorithm needs a point target, not only direction. Furthermore, a target searching task requires a termination condition that aware the robot to stop searching (i.e. the robot arrives the target point). If sound source detection is the only way for a robot to find a target, the robot will never know how close it is to the goal.

Second, many external factor (e.g., footsteps) will cause inaccurate result that estimated using TDOA. If the environment is too noisy, TDOA sometimes fails.

Above problems imply that a robot can not complete a target searching task only relies on acoustic sensors, i.e., microphone pairs. As a result, we equipped a camera on our robot, and designed a simple graphic pattern nearby docking station for recognition.

A. Robot Hardware

To accomplish the task of searching the docking station, we equipped our robot with two microphone pairs, a laser range finder (LRF) and a camera as Fig. 7 shows. As above mentioned, the microphone pairs are used to collect sound wave for TDOA, laser range finder to detect obstacles for navigation and camera to figure out the docking station when robot is close enough to it.

B. Docking Station

The docking station is designed to generate sound for robot to estimate its location. For testing purpose, we use a metronome to replace a real docking station to reduce the cost. On the other hand, a red filled circle pattern is stick above the metronome for vision recognition. The simulated docking station is shown in Fig. 8.

C. Searching Procedure

The searching task is composed of two stages. In stage 1, the robot acquires a sound signal form docking station. After receiving a sound signal, robot estimates the sound source direction using TDOA. Then the robot faces to the sound source direction and finds the target using camera. If the target is detected, just moves toward it; if not, enter stage 2. In stage 2, the robot determines a target point at the direction of the sound source according to the sensory data of laser range finder. As already explained, robot can find a "temporary target" according to the distribution of obstacles and target using our approach. After robot arrives the temporary target, repeat stage 1. The overall procedure is described in Alg. 3.

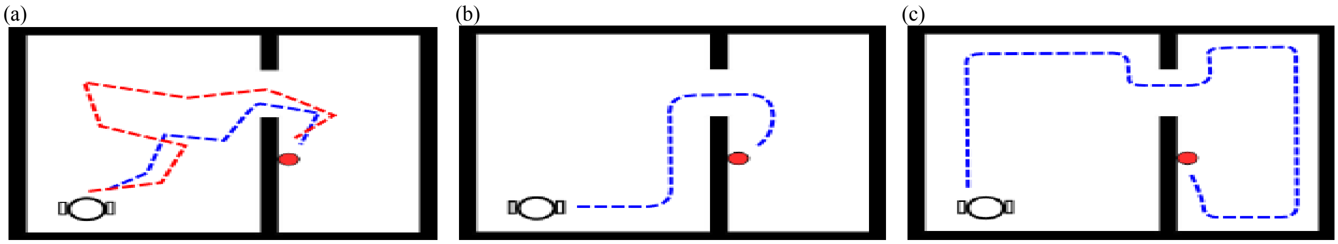

Fig. 9. Experimental results. (a) shows the paths produced by our searching approach based on sound source detection and TangentBug-inspired navigation algorithm. Since the result of TDOA sometimes affected by external factors such as noise, the paths are different in every round. The red dash line shows the TDOA failed several times during searching the sounding target. However, the robot can still find the correct path in the end as long as the rest of TDOA results are acceptable. (b) and (c) show the paths produced by applying wall-following navigation by using laser range finder and detect the target using only camera. If the target is outside the visual field of the robot in the beginning, the length of searching path heavily depends on the initial direction of wall-following.

Alg. 3 Searching Procedure

1. Acquire a sound signal, estimate the target direction *tD*.
2. Face to *tD* and detect using camera. If target is found, move toward it. Terminate the procedure.
3. If target is not found, set a point at the *tD* as target point *tP*. According to the distribution and target point position, find a *tP*.
4. Move to *tP*, and repeat 1.

Note that in every round of the procedure, robot can roughly estimate the target direction by sound source detection, and determine a temporary target using our approach. **The robot will always find a *relative* position between itself and targets (or temporary target) by the procedure. Thus the robot doesn't need to perform localization task when searching the docking station.**

V. Experimental Result

The experiment is designed to confirm that the proposed approach successfully reduce the searching work in our scenario. To design a control group that the robot uses only range sensors and camera to accomplish the searching task, we used wall-following as navigation strategy (the reason is described in next section) while finding the target. Fig. 9 shows the paths produce by the robot during searching docking station, and (a) is the result of our sound-aid approach, while (b) and (c) are results produced by wall-following method with different initial direction guess respectively.

It is clear that if the robot uses only LRF and camera to find the docking station with wall-following navigation, the result will strongly depend on the initial guess of the direction. A bad initial guess of direction will cause a longer path for robot to find the target.

Fig. 9(a) shows that the robot always find a shorter path than (c) using the proposed approach. Although sound source detection based on TDOA can't provide an accurate target direction, it is still acceptable that the estimated direction can guide the robot toward a roughly correct path until camera can detect the target.

VI. Discussion

Theoretically, if a robot wants to perform searching works in an unknown environment with only range sensors and camera, it is unavoidable to perform SLAM (simultaneously localization and mapping) at the same time, or the robot will probably get lost and just fool around. In a domestic environment, one chance to achieve the goal is applying wall-following as navigation method while searching, so we choose wall-following method as control group.

On the other hand, some times the TDOA estimation will fail due to external factors (i.e. footsteps, people talking). However, assume there are several TDOA estimations failure during searching, the robot can still find the correct path in the end if the rest TDOA estimation results are acceptable, as the red dash line shown in Fig. 9(a).

As mentioned, several researches address similar works [10]-[14]. However, most of them focus on the TDOA models or mathematical robot dynamic constrains while we propose a simple, practical, and inexpensive method to combined TDOA and robot navigation.

There are still researchers working on integration of robot navigation algorithm with different sensors just as we did in this paper. One typical research topic is integration of robot navigation with olfaction and vision [17]-[19]. The feature of gas is somewhat similar to sound – it can travel from one room to another, and the gradient of gas provides direction information of the gas source. No doubt olfaction provides significant information in many conditions, especially in fire accidences, but we believe auditory information plays a more essential role in our daily life.

In the scenarios of robot docking, modern researches tend to combine vision and other sensors such as laser range finder, IrDA (infrared data association), RFID to localize the docking station [20]-[21]. However, most of these works assume the docking station is inside the visual field of the robot in the beginning of docking process. Our method provides an alternative way to search the docking station before seeing it, which should be useful for these works.

TABLE I
COMPARISON OF SLAM BASED SEARCHING METHOD AND PROPOSED SOUND BASED METHOD

	SLAM and Vision Based Searching	Sound Based Searching
Computational Power Consumption	High	Low
Length of Searching Path	Could be very long, very short or in between, depends strongly on environments and path planning algorithms.	Short
Implementation	Hard. Localization and Mapping issues are very complex.	Simple
Effects of Obstacles	Obvious. Laser range finder, camera and sonar cannot detect the target behind obstacles.	Not Obvious. Sound can penetrate obstacles.
Detect Direction	Restricted. Robot should be in correct direction to detect vision pattern.	Not Restricted Robot can detect sound in any direction.

VII. COMPARISON

In comparison with traditional target searching method based on SLAM and vision, our method is superior in several aspects. First, sound signal is omni-directional, which means a robot can detect sound signal without facing to the sound source or being in front of it. Second, sound signal can penetrate obstacles, thus a robot can detect it even there are something between the robot and sound source. Third, searching a target by sound can avoid complex localization issues and save computational power. Finally, although sound signal cannot provide an accurate direction during searching, acceptable TDOA results are still very helpful for searching the target. Relatively, traditional SLAM methods are hard to find good initial directions to find the target. Table I compares our method with traditional SLAM and vision method.

VIII. CONCLUSIONS

In this article, a searching method based on sound source detection and TangentBug-inspired navigation is proposed. We designed a scenario of robot finding docking station to demonstrate the searching process and confirm its practicality. The proposed approach provides a pretty good solution for target searching task in several aspects. In comparison with traditional SLAM approaches, our method avoids the complicated localization issue and saves computational power and searching time. Although the target should be able to generate sound signal and the robot should have acoustic sensor, it is still inexpensive and straightforward.

REFERENCES

[1] K. Nakadai, K. Hidai, H. Mizoguchi, H. G. Okuno, H. Kitano, "Real-Time Auditory and Visual Multiple-Object Tracking for Humanoids," in *17th International Joint Conference on Artificial Intelligence*, Seattle, Washington, USA, 2001, pp. 1425–1436.

[2] F. Asano, Y. Motomura, H. Asoh, T. Yoshimura, N. Ichimura, S. Nakamura, "Fusion of Audio and Video Information for Detecting Speech Events," in Proc. 6th International Conf. on Information Fusion, Cairns, Queensland, 2003, pp. 386–293.

[3] R. O. Schmidt, "Multiple Emitter Location and Parameter Estimation," *IEEE Trans. Antennas Propagat.* , vol. AP-34, pp. 276–280, 1986.

[4] R. Roy, T. Kailath, "ESPRIT-Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[5] J. M. Valin, F. Michaud, J. Rouat, D. Letourneau, "Robust Sound Source Localization using a Microphone Array on a Mobile Robot," *IEEE/RSJ Int.Conf. on Intelligent Robots and Systems*, vol. 2, pp. 1228–1233, October 2003.

[6] Y. Sasaki, Y. Tamai, S. Kagami, H. Mizoguchi, "2D Sound Source Localization on a Mobile Robot with a Concentric Microphone Array," *IEEE Int.Conf. on Systems, Man and Cybernetics*, vol. 4, pp. 3528–3533, October 2005.

[7] B. Kwon, G. Kim, Y. Park, "Sound Source Localization Methods with Considering of Microphone Placement in Robot Platform", *IEEE International Symposium on Robot and Human Interactive Communication*, pp. 127–130, August 2007.

[8] R. C. Luo, C. T. Liao, and S. C. Lin, "Multi-Sensor Fusion for Reduced Uncertainty in Autonomous Mobile Robot Docking and Recharging," *IEEE/RSJ Int.Conf. on Intelligent Robots and System* 2009.

[9] R. C. Luo, W. H. Cheng, "Combined 3D Sound Source Localization with Stereo Vision for Intelligent Human-Robot Interactions of Service Robots," *IEEE Workshop on Advanced Robotics and its Social Impacts*, Tokyo, Japan, Nov. 2009

[10] J. Ng, T. Braunl, "Performance Comparison of Bug Navigation Algorithms," Journal of Intelligent and Robotics Systems. Springer Netherlands, vol. 50, No. 1, pp. 73–84, September 2007.

[11] I. Kamon, E. Rivlin, "A New Range-Sensor Based Globally Convergent Navigation Algorithm for Mobile Robots," in *Conf. 2002 IEEE Int. Conf. Roboitcs and Automation,* pp. 429–435.

[12] N. Uchiyama, A. Yamamoto, S. Sano, S. Takagi, "Sound Source Tracking and Obstacle Avoidance for a Mobile Robot," in *2009 IEEE Int. Conf. Emerging Technologies and Factory Automation),* pp. 1–5.

[13] J. Huang, T. Supaongprapa, I. Terakura, F. Wang, N. Ohnishi, N. Sugie , "A Model-Based Sound Localization System and Its Application to Robot Navigation," Robotics and Autonomous Systems, vol. 27, pp. 199–209, 1999.

[14] S. B. Andersson, A. A. Handzel, V. Shah, P. Krishnaprasad, "Robot Phonotaxis with Dynamic Sound-Source Localization," in *Conf. 2004 IEEE Int. Conf. Robotics and Automation,* pp. 4833–4838.

[15] K. Nakadai, H. Nakajima, M. Murase, H.G. Okuno, Y. Hasegawa, H. Tsujino, "Real-Time Tracking of Multiple Sound Source by Integration of In-Room and Robot-Embedded Microphone Arrays," in *Conf. 2006 IEEE/RSJ Int. Conf. Intelligent Robotics and Systems,* pp. 852–859.

[16] K. Nakadai, H. Nakajima, M. Murase, S. Kaijiri, K. Yamada, T. Nakamura, Y. Hasegawa, H.G. Okuno, H. Tsujino, "Robust Tracking of Multiple Sound Source by Spatial Integration of Room and Robot Microphone Arrays," in *Conf. 2006 IEEE Int. Conf. Acoustics, Speech and Signal Processing,* pp. IV–IV.

[17] G. Sandini, G. Lucarini, M. Varoli, "Gradient Driven Self-Organizing Systems," *Proc. IEEE/RSJ International Conference on Intelligent Robots Systems*, 1994, pp. 429–432.

[18] Y. Kuwana, I. Shimoyama, H. Miura, "Steering Control of a Mobile Robot Using Insect Antennae," *Porc. IEEE/RSJ International Conference on Intelligent Robots Systems*, 1995, pp. 530–535.

[19] H. Ishida, T. Ushiku, S. Toyama, "Mobile Robot Path Planning Using Vision and Olfaction to Search for a Gas Source," *Sensors,2005 IEEE*, pp. 1–4.

[20] D. Amarasinghe, G. K. Mann, R.G. Gosine, "Vision-Based Hybrid Control Strategy for Autonomous Docking of a Mobile Robot," *Proc. IEEE Int.Conf. on Control Applications* 2005, pp.1600–1605.

[21] R. C. Luo, C. T. Liao, K. C. Lin, "Vision-Based Docking for Automatic Security Robot Power Recharging," *Advanced Robotics and its Social Impacts, 2005. IEEE Workshop on*, pp. 214–219.