様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ３０ 年　 ６ 月 １５ 日現在

機関番号：９４３０１
研究種目：若手研究(B)
研究期間：2016～2017
課題番号：１６Ｋ１６１３５

研究課題名（和文）Using mobility to localize sound sources in complex environments

研究課題名（英文）Using mobility to localize sound sources in complex environments

研究代表者
Ｅｖｅｎ　Ｊａｎｉ（Even, Jani）

株式会社国際電気通信基礎技術研究所・石黒浩特別研究所・研究員

研究者番号：４０５８６２２８

交付決定額（研究期間全体）：（直接経費）　　2,300,000円

研究成果の概要（和文）：移動ロボットを使って音環境の3次元地図を作る技術を開発した。この技術はマイクロホンアレーと音信号処理をロボットの自己位置推定と組み合わせる。この3次元音地図構築はリアルタイム処理ができる。
音の反射を使って音源の位置を推定する技術も開発した。音反射がある時に見えない音源を検知できる。環境の3次元地図を使って音反射を3次元レーキャスティングで音源まで広めて音源の位置を推定する。

研究成果の概要（英文）：A technique to map sound sources using a mobile robot equipped with a microphone array was developed. The spatial distribution of the sound is exploited to determine the position of the sound sources in the environment. For this purpose, a quad-tree based sampling technique has been introduced.
By using the geometric information known about the environment, the path of reflected sounds are estimated and used to detect sound sources that are out of the field of view.
A way to modulate the sound perception inside and outside the field of view has been implemented in order to prioritize sound sources that are not visible.

研究分野：Electrical Engineering

キーワード：sound localization　sound mapping　Field of view

１．研究開始当初の背景

(1) This research deals with the use of microphone array and sound localization algorithms on-board mobile robots. In particular, the use of locomotion to localize sound sources that are not initially in the field of view.

(2) Thus, the current work tries to integrate sound localization results in the framework used by mobile robot for mapping, self-localization and motion planning.

２．研究の目的

(1) The goal of the research is to enable the localization of non-visible sound sources by using locomotion.

(2) An important point is also the integration of the sound information in the environment representation conventionally used in the mobile robot frameworks. The improvement of audio information mapping is one of these task.

３．研究の方法

(1) A first necessity for performing the task was to create a reliable mobile platform to carry the microphone array. The most important is the integration of the audio processing in the software stack used for mapping, localization and navigation. After experimenting with our own shared memory based framework, we switch to the Robot Operating System (ROS). Thus the principal investigator developed a complete audio stack in ROS that perform sound localization.

(2) Using the developed audio stack, it was possible to integrate the sound localization with the robot self-localization framework and build sound map of large area. Namely, the robot is able to autonomously navigate the environment and create a map of the sound sources. One particularity of our approach is the use of acoustic ray casting for associating detected sound directions to structures in the environment.

(3) Sound sources that are out of the field of view are still perceived because of the propagation of reflected and diffracted sound that reach the microphone array. Knowing the geometry of the environment, it is possible to gather some information on the potential position of a non-visible sound source. Thus being able to harvest the information from reflected sound is an important capacity for our system.

(4) When considering multimodal sensing, in particular sound and vision. The perception of the sound is influenced by our field of view. In particular, sound information from sources out of the field of view is more informative than the one from sound source we can see. Consequently, we studied the coupling of field of view movement and sound perception.

４．研究成果

(1) The development of a performant audio processing stack for ROS is an important result as it is a solid base for conducting further research in mobile audio. It should be noted that there currently few frameworks that are used for multichannel audio on robots. But these frameworks did not provide the functionality and modularity that we needed. In particular, great care was given to use the same coordinate transform engine as ROS (the gray boxes in fig.1) in order to integrate the spatial information of the sound in the robot framework in an efficient and flexible manner. Two approaches for storing the spatial information of the sound were developed: An octree structure containing sound information in voxels or some augmented point cloud representation (fig. 1 is showing that particular stack).

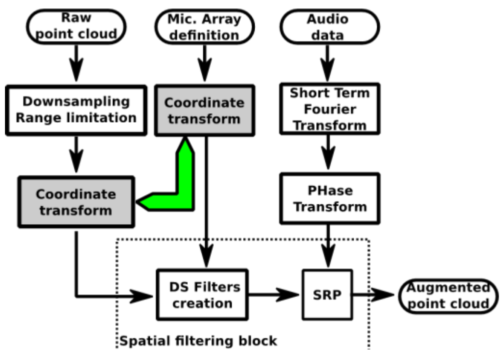

Figure 1: Audio stack diagram

(2) Using the proposed audio stack, the sound localization and especially the spatial information about the sound sources was integrated with the robot mapping framework. By moving from our own shared memory based system to the ROS based system, it was possible to create larger maps and effectively deals with sound sources that were originally out of the field of view. Moreover, the sound information is now in 3D and attached to the geometric structure of the environment. Figure 2 shows one of such sound maps.

The color scale from blue to red indicates the probability that the structure at that location emits sound (structures that are not probed yet appear in grey). Thus the islands of red color indicates sound sources. This map was built using the octree approach. Figure 3 (left) presents a sound map built with the point cloud approach. The fridge is emitting sound during the mapping (note that for the point cloud based mapping the color scale is from red to blue).

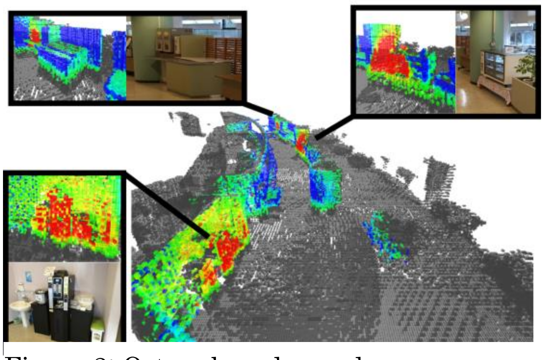

Figure 2: Octree based sound map

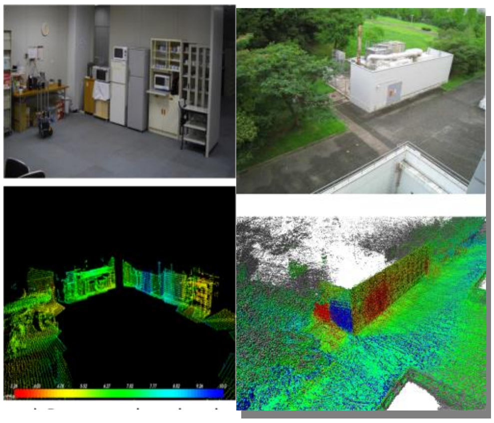

Figure 3: Point cloud based sound map (left) and outdoor mapping with octree (right)

(3) In addition to indoors sound mapping, the mapping of structures outdoors was also investigated. In Fig. 3 (right), the part of the building from which sound was emitted is localized by the proposed approach. However, the main focus was on the indoors mapping as the localization of sound sources out of the line of sight is especially important inside buildings.

(4) To deal with sound sources that are not visible, the system uses movement to approach their location. In order to decide the direction of the movement, the perceived reflected sound is exploited. Ray casting is applied to trace back received sound to their origin source. The ray casting procedure estimates the specular reflection of the sound on the structures, see fig.4 (bottom). For this purpose, it is necessary to obtain a 3D model of the environment accurate enough to allow surface estimation and normal to surface estimation. Here again, the integration of the sound stack with the conventional mapping framework used in robotics proved to be a good approach. In the current state, the use of normal estimated on local neighborhoods is the most efficient as it requires few preprocessing and enable the computation of reflected sound on all structures. To avoid using structures with a poor normal estimation, some index describing how well the local neighborhood could be fitted by a plane was introduced. Figure 4 (top) shows the result of the normal estimation for a T shape corridor. The uniform red and blue colors for the three components of the normal show that the wall are flat. Note that the index on the right is lower at edges where two walls connect and the normal is not well defined. The sound rays intersect and these intersections are potential sound source positions that indicate a possible direction to move and investigate.

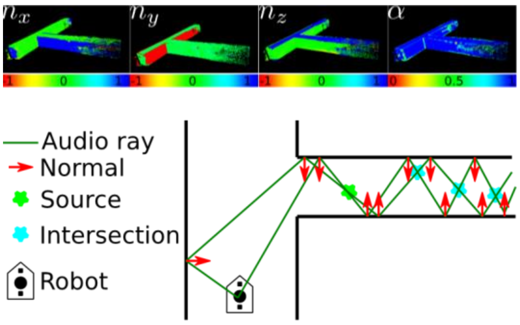

Figure 4: T shape corridor and reflections

(5) A model of the field of view interaction with the sound perception was introduced. The goal was to give more importance to sounds detected out of the field of view and implement a short memory of sound events. Figure 5 shows the masking of the environment that takes into account the field of view. Sounds within the field of view are masked. That masking varies slowly when the robot moves as illustrated on the top right in order to not forget what was previously seen. The masking influences which direction the robot should face (the white cross in the bottom) by giving less importance to what can be

directly seen. Thus the modified sound localization could be coupled with a visual detection method in the future. The vision would manage detection in the field of view while the audio would care of the surrounding. A difficulty with this model is the setting of the parameter with respect to the robot motion speed.

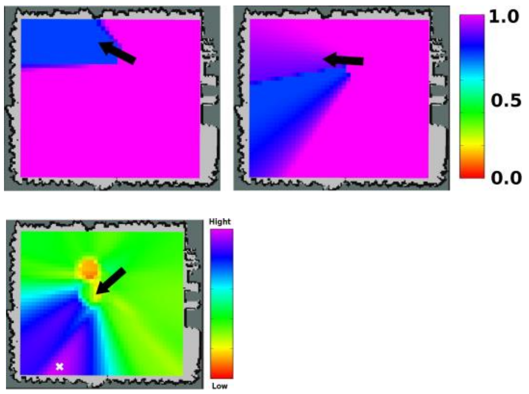

Figure 5; Field of view masking

(6) Combining the different parts of the problem that have been investigated, the detection of sound sources in a complex environment with no line of sight is possible but the path followed by the robot is not optimal. Further research has to be conducted on the prediction of sound source positions from perceived reflections.

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　1　件）
①Jani Even, Jonas Furrer, Yoichi Morales, Carlos Toshinori Ishi, Norihiro Hagita, “Probabilistic 3-D Mapping of Sound-Emitting Structures Based on Acoustic Ray Casting”, IEEE Trans. Robotics 33(2), 2017, 333-345.
DOI:10.1109/TRO.2016.2630053

〔学会発表〕（計　0　件）

〔図書〕（計　0　件）

〔産業財産権〕

○出願状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者
Even, Jani（ Even, Jani ）
Advanced Telecommunications Research Institute International, Hiroshi Ishiguro Laboratories, Researcher.
研究者番号：40586228

(2)研究分担者
（　　　）

研究者番号：

(3)連携研究者
（　　　）

研究者番号：

(4)研究協力者
Morales Yoichi（Morales, Yoichi）
Furrer Jonas（Furrer, Jonas）
Ishi Carlos（Ishi, Carlos）