KRS KOREA ROBOTICS SOCIETY

www.robotweek.or.kr

Korea International Robot Week 2006

International Conference on Ubiquitous Robots and Ambient Intelligence 2006(URAI 2006)

Oct. 15-17, 2006 KIST / SEOUL, KOREA

* Technically Co-Sponsored by

RSJ
Robotics Society of Japan

* Co - Sponsored by

mic
Ministry of Information and Communication

ITA
Institute for Information Technology Advancement

KIST
Korea Institute of Science and Technology

ETRI
Electronics and Telecommunications Research Institute

* Supported by

KAIRA
Korea Advanced Intelligent Robot Association

SIRIUS
THE WORLD THROUGH YOUR EYES
Sirius

TE1: Navigation Algorithms

Poster Session

PO1: Poster (Sound, Vision, Sensors and Actuators, Humanoid Robots)

Multimodal Sound Source Localization for Intelligent Service Robot

Jisung Choi*, Jiyeoun Lee*, Sangbae Jeong*, Keunchang Kwak**, Suyoung Chi**, Minsoo Hahn*

*Speech and Audio Information Lab., Information and Communications Univ.
{dir81, jyle278, sangbae, mshahn}@icu.ac.kr
** Intelligent Robot Research Division, Electronics and Telecommunications Research Institute
{kwak, chisy}@etri.re.kr,

Abstract - This paper shows a multi-modal sound source localization system for intelligent service robots. A sound source localization system estimates an azimuth of the sound source with three microphones arranged on a circle while an auxiliary face detection system compensates a possible error in the azimuth estimation. The sound source localization algorithm adopted in our system is robust to microphones' gain because it uses only the time differences between signals. To make it possible, a cost function which normalizes the microphones' gain is utilized. Also, a face detection system is used to exactly detect and track the sound source. The maximum error of the sound source localization system is less than $\pm 7°$ and this error can be compensated by the face detection system.

Keywords - multimodal sound source localization, sound source localization, robust to microphones' gains, face detection

1. Introduction

In the near future, intelligent service robots will be important in human society and people will want helps of them. Intelligent service robots have wide applications. First, intelligent service robots can assist human. For example, a robot moves to a person who needs help through sound and vision information in public areas. Second, the robot can help a more human-friendly remote meeting. It can search a talking person and show him automatically.

A multimodal sound source localization (SSL) system uses sound and vision information together. With the sound system, a robot can estimate the direction of a talker. With the vision system the robot can recognize specific objects such as the speaker's face and identify the location of the detected target more precisely. Sound and vision information has some advantages. Sound information is useful when a sound source is located in the dark place and the sight of a robot is blocked. Vision information is needed when the sound source cannot be detected correctly because of some environmental noises, echoes, etc. If we use only the vision information to search a person who needs services of a robot, difficulties occur when the person doesn't exist in the visual field of the camera. Thus, both sound and vision information should play a complementary role.

For the above purpose, both the sound and the vision informations are essential to multimodal SSL systems for intelligent service robots. To integrate both informations, the algorithm includes a SSL and a face detection module. The performance is improved in comparison with using only one module. For a multimodal SSL system, our intelligent service robot firstly estimates the direction of the speaker using an SSL algorithm robust to the microphones' gain. Next, the robot picks up speaker's face in its visual field using a face detection algorithm and then the robot moves to the talker [1].

Previous SSL algorithms use only time delay information based on the simple cross correlation, the intensity difference or the phase difference [2, 3]. However, those algorithms cannot estimate correctly azimuths because their performances are influenced by the microphones' gain. Thus, we propose a rather improved algorithm robust to microphones' gain. From the proposed SSL algorithm, azimuth errors of sound sources are estimated within about $\pm 7°$. Also, we apply our algorithm to the multimodal SSL system to compensate errors in azimuth detection [4].

The organization of this paper is as follows. Section 2 introduces the multimodal SSL system. Section 3 discusses the proposed SSL algorithm and Section 4 discusses the face detection algorithm. Results and conclusions are given in Section 5 and Section 6, respectively.

2. Overall Multi modal Sound Source Localization Algorithm

Our multimodal SSL system is designed as shown in Fig1. Firstly, the SSL algorithm estimates the azimuth of the sound source. Second, the robot detects the face of a speaker by utilizing the face detection algorithm.

For SSL, short time energies of input signals received by three microphones are estimated. The sound activity detection part detects sound activity periods using the short time energy. Then our SSL algorithm calculates the time delay robust to the microphones' gain and estimates the azimuth. Using the estimated azimuth, the robot turns to the sound source direction.

Face detection uses open source a CV (Computer Vision) library developed by Intel Company. The robot