

A Study on Precise Control of Autonomous Driving Robot by Voice Recognition

Sung-Won Jung¹, Ki-Won Sung¹, Moon-Youl Park¹, Eon-Uck Kang¹, Won-Jun Hwang¹,
Jong-Dae Won², Woo-Song Lee³, and Sung-Hyun Han⁴

¹ Department of Advanced Engineering, Kyungnam University, Changwon, 631-701, Korea, carisma3000@naver.com
(Tel: 82-55-249-2777, Fax: 82-505-999-2129)

²Youngchang Tools Co.,Ltd., Changwon, 641-847 Korea
(Tel: 82-55-273-7571, Fax: 82-55-237-7572)

³Sungsanam Deco Co.,Ltd., Changwon, 630-500 Korea
(Tel: 82-55-237-8662, Fax: 82-55-237-3680)

⁴School of Mechanical System and Automation Eng., Kyungnam University, Masan, Korea
(Tel: 82-55-249-2624, Fax: 82-505-999-2129)

Abstract— Generally, it is possible to control the motion by using information on the robot's own postures, because a type of motion and gesture produces almost the same pattern of every time. In this paper, we describe a voice recognition control system for robot system which can robustly recognize voice by adults and children in noisy environments. We evaluate the performance of robot control system in a communication robot placed in a real noisy environment. Voice is captured using a wireless microphone. To suppress interference and noise and to attenuate reverberation, we implemented a multi-channel system consisting of an outlier-robust generalized side-lobe canceller technique and a feature-space noise suppression using criteria. Voice activity periods are detected based end-point detection.

Index Terms— Robust Voice Recognition, Precise Control, Autonomous Driving Robot

I. INTRODUCTION

To make human-robot communication natural, it is necessary for the robot to recognize voice even while it is moving and performing gestures. For example, a robot's gesture is considered to play a crucial role in natural human-robot communication [1-4]. In addition, robots are expected to perform tasks by physical actions to make a presentation. If the robot can recognize human interruption voice while it is executing physical actions or making a presentation with gestures, it would make the robot more useful.

Each kind of robot motion or gesture produces almost the same noises every time it is performed. By recording the motion and gesture noises in advance, the noises are easily estimated. By using this, we introduce a new method for VRCS under robot motor noise. Our method is based on three techniques, namely, multi-condition training, maximum-likelihood linear regression[5], and missing feature theory. These methods can utilize pre-recorded noises as described later. Since each of these techniques has advantages and disadvantages, whether it is effective depends on the types of motion and gesture. Thus, just combining these three techniques would not be effective for voice recognition under noises of all types of

motion and gestures. The result of an experiment of isolated word recognition under a variety of motion and gesture noises suggested the effectiveness of this approach. In what follows, Section 2 discusses the design of voice recognition system, and Section 3 explains our method for avoid the obstacles by navigation strategy. Section 4 describes the recognition, navigation experiments, and the results, before conclusion and mentioning future work in Section 5.

II. CONTROL SCHEME

Accounting for the two problems (caused by noisy environments and differences on speaker age) described in Section I, we developed an RVR system to be robust on background noise and speakers of different ages. The first block is a front-end processing. It contains a microphone wireless transmitter. The real-time wireless microphone system for suppressing interference and noise and for attenuating reverberation consists of an outlier-robust generalized side-lobe canceller and feature-space noise suppression. In the second block, there are two decoders depending on the age of the speaker (adult or child); each decoder works using gender-dependent acoustic models. Noise-suppressed voice at the first block is recognized using these two decoders, and one hypothesis is selected based on posterior probability.

The proposed robot system has three wheels; two driven wheels fixed at both sides of the mobile robot and one castor attached at the front and rear side of the robot. The ultrasonic sensors are mounted around of the mobile robot in middle layer for the detection of obstacles with various heights. In this study, a sonar array composed of 16 ultrasonic sensors cannot be fired simultaneously due to cross talk. Instead, we adopt a scheduled firing method where sensors are activated in sequence of $\{s_1, s_{12}, s_2, s_{11} \dots\}$. The arrangement of the ultrasonic sensors in upper layer and the sensors are marked as dots in the figure. The distances e_j ($j = 1, 2, \dots, 12$) from the origin of the robot frame $\{R\}$ to obstacles detected by the sensor s_j , can be defined as $e_j = \delta_j + R_r$. Here, R_r is the radius of

the robot and the δj , is the range value measured by the sensor s_j .

A local map is introduced to record the sensory information provided by the 16 sonar sensors with respect to the mobile robot frame $\{R\}$. Sector map defined locally at the current mobile robot frame is introduced. Then, the obstacle position vector se_j with respect to the frame $\{R\}'$ can be calculated by

$$Se_j' = \begin{bmatrix} \cos \delta\theta & \sin \delta\theta & 0 & -\sin \delta\theta / \rho_p \\ -\sin \delta\theta & \cos \delta\theta & 0 & (1 - \cos \delta\theta) / \rho_p \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where se_j denotes the obstacle position vector defined at the frame $\{R\}$. Hence, when the mobile robot is located at a point O' , the distance value $se_j' = \|se_j'\|$ from the origin of the frame $\{R\}'$ to the obstacle and angle $s\phi_j'$ can be calculated by Eq.(1). Here, $\|\cdot\|$ denotes Euclidean norm.

The local map defined at the frame $\{R\}'$ is newly constructed by using the previous local map defined at the frame $\{R\}$ as follows:

$$Se_n \leftarrow Se_j', n = INT(\frac{s\phi_j'}{\varphi}) + \frac{N}{2}; j = 1, 2, \dots, N \quad (2)$$

Where \leftarrow and INT denote the updating operation and integer operation, respectively. Here, se_n , denotes the distance value of n sector and N represents the number of the sector. If the range values obtained by sensors when the mobile robot is located at a point O' are $ej = (j = 1, 2, \dots, 12)$, the new local map is partially updated as follows:

$se_j \leftarrow ej, j = 1, 2, \dots, 12$. The maximum range of the sonar sensor is set to be $\delta_{max} = \delta_{max} - R_r$. Any return range which is larger than is ignored.

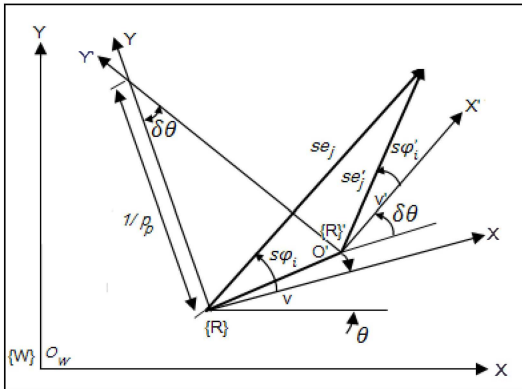


Fig. 1 The coordinate transformation for updating the local map

The primitive behaviors may be divided as follows: goal-seeking behavior, ball-following behavior, keep-away behavior, free space explorer and emergency stop, etc. The output of a primitive behavior is defined by the vector.

$$u(t) = (v(t), \Delta\theta(t))^T = (v(t), w(t), Tms)^T \quad (3)$$

where t and Tms denote the time step and the sampling time, respectively. Here, T denotes the transpose and $\omega(t)$ denotes the angular velocity of the robot.

We will divide the primitive behaviors into two basic: avoidance behavior and goal-seeking behavior.

The avoidance behavior is used to avoid the obstacles irrespective of the goal position, while the goal-seeking behavior is used to seek the goal position irrespective of obstacle location. Design of each behavior proceeds in following sequences;

(A) fuzzification of the input/output variables, (B) rule base construction through reinforcement learning, (C) reasoning process, (D) defuzzification of output variables.

In order for the mobile robot to arrive at the goal position without colliding with obstacles, we must control the mobile robot motion in consideration of the obstacle position $X_{oi} = (x_{oi}, y_{oi})$, the mobile robot position $X = (x, y)$ and its heading angle θ with respect to the world coordinate frame $\{W\}$ shown in Fig. 1.

In order to avoid the increase in the dimension of input space, the distance values d_i , ($i = 1, 2, 3, 4$) are defined by

$$\begin{aligned} d_1 &= \min(se_1, se_2, se_3) \\ d_2 &= \min(se_4, se_5, se_6) \\ d_3 &= \min(se_7, se_8, se_9) \\ d_4 &= \min(se_{10}, se_{11}, se_{12}) \end{aligned} \quad (4)$$

$\phi_m (-\pi \leq \phi_m \leq \pi)$ denotes the orientation of a sector with the shortest range. We choose the input variables for avoidance behavior as ϕ_m and

$d_i = \|X_{oi} - X\|, (i = 1, 2, 3, 4)$ for goal-seeking behavior as heading angle difference ψ and distance to goal $z = \|X_g - X\|$. The input linguistic variables $d_i, \psi,$

ϕ_m and z are expressed by linguistic values (VN, NR, FR), (NB, NM, ZZ, PS, PM, PB), (LT, CT, RT) and (VN, NR, FR, VF), respectively. Their membership functions are expressed.

III. EXPERIMENTS

All the parameters used in the navigation experiments are given in Table 1. The mobile robot has the maximum travel speed of 0.52 m/s and the maximum steering rate of 1.854 rad/sec. Experiments are performed in an indoor with the first experiment for voice recognition without objects and second experiment for both of them: voice recognition and obstacles avoidance. The first experimental space is approximately 9.4m by 1.475m wide, and the second experimental space is approximately 14.2m by 2.5m wide. Fig. 1 shows a sketch of the top view of the room with the object drew in the box shape. Since this environment is too simple to

test the performance of the overall system, several polygon obstacles were randomly placed in the path of the mobile robot navigation. The mobile robot was initially located at the origin of the world coordinate frame $\{W\}$ and the goal of the first experiment was that robot goes straight until

someone speaks “turn left”. Then, the second experiment was active, respectively.

Through a series of the navigation experiments, it was observed that the heading angle error is a serious problem to mobile robot navigation relying on dead reckoning. The large heading angle error almost resulted from the wheels’ slippage when the mobile robot changes its direction. Even if the wheel slippage occurs, the true position and heading angle of the mobile robot could be updated by two beacon pairs and consequently the mobile robot could arrive at the given goal position while avoiding the obstacles.

IV. CONCLUSIONS

We have proposed the integration of robust voice recognition and navigation system capable of performing autonomous navigation in unknown environments. In order to evaluate the performance of the overall system, a number of experiments have been undertaken in various environments. The experimental results show that the mobile robot with the complete voice recognition and navigation system can arrive at the goal position according to the desire of speaker even if the wheel slip occurs. From the developed of voice recognition and navigation system, it was observed that the mobile robot can successfully arrive at the desired position through the unknown environments without colliding with obstacles.

ACKNOWLEDGMENT

This work was supported (National Robotics Research Center for Robot Intelligence Technology, Kyungnam University) by Ministry of Knowledge Economy under Human Resources Development Program for Convergence Robot Specialists.

REFERENCES

- [1] S. Yamamoto, K. Nakadai, H. Tsujino, and H. Okuno, “Assessment of general applicability of robot audition system by recognizing three simultaneous speeches.” in Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2004). IEEE and RSJ, 2001, p. to appear.
- [2] P. Renevey, R. Vetter, and J. Kraus, “Robust speech recognition using missing feature theory and vector quantization,” in Proc. Of Eurospeech-2001. ESCA, 2001, pp. 1107–1110.
- [3] J. Barker, M. Cooke, and P. Green, “Robust asr based on clean speech models: An evaluation of missing data techniques for connected digit recognition in noise,” in Proc. of Eurospeech-2001. ESCA, 2001, pp. 213–216.
- [4] R. P. Lippmann, E. A. Martin, and D. B. Paul, “Multi-style training for robust isolated-word speech recognition,” in Proc. of ICASSP-87. IEEE, 1987, pp. 705–708.
- [5] M. Blanchet, J. Boudy, and P. Lockwood, “Environment adaptation for speech recognition in noise,” in Proc. of EUSIPCO-92, vol. VI, 1992, pp. 391–394.