2012 12th International Conference on Control, Automation and Systems
Oct. 17-21, 2012 in ICC, Jeju Island, Korea

An Intelligent Control of Mobile Robot Based on Voice Command

Byoung-Kyun Shim[1], Kwang-wook Kang[1], Woo-Song Lee[2], Jong-Baem Won[3] and Sung-Hyun Han[4]

[1] Division of Advanced Eng., Graduate School, Kyungnam University, Masan, Korea
(Tel : +82-55-249-2777; E-mail: sbk1210@gmail.com)
[2]Kwangjin Precision Co.,Ltd., Masan, Korea
(Tel : +82-55-232-5211; E-mail: elflws@nate.com)
[3]Smec Co., Ltd., Changwon, Korea
(Tel : +82-55-250-4800; E-mail : epia13@esmec.com)
[4]Division of Mechanical System and Automation Eng., Kyungnam University, Masan, Korea
(Tel : +82-55-249-2624; E-mail: shhan@kyungnam.ac.kr)

Abstract: In general, it is possible to estimate the noise by using information on the robot's own motions and postures, because a type of motion and gesture produces almost the same pattern of noise every time. In this paper, we describe an voice recognition control system for robot(VRCS) system which can robustly recognize voice by adults and children in noisy environments. We evaluate the VRCS 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 MMSE criteria. Voice activity periods are detected using GMM-based end-point detection

Key words: Robust voice recognition, Side-lobe canceller, navigation system

1. 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 [1-9]. In addition, robots are expected to perform tasks by physical actions [12] to make a presentation [10]. 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 (MLLR) [5], and missing feature theory (MFT) [7]. 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.

2. 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 to both 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 (RGSC) and feature-space noise suppression (MMSE). MMSE noise suppression is applied after RGSC to reduce the residual noise at the RGSC output. After that, the voice activity period detected by the GMM-based end-point detection (GMM-EPD) is transferred to the second block. 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 following sub-sections describe each module of our VRCS system.

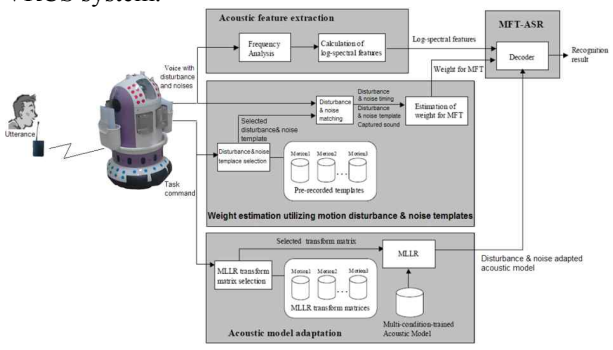


Fig 1. Block diagram of the proposed method

acoustic features is estimated correctly. One of the main issues in applying them to VRCS is how to estimate the reliability of input acoustic features correctly. Because the *signal-to-noise ratio (SNR)* and the distortion of input acoustic features are usually unknown, the reliability of the input acoustic features cannot be estimated. So, the SS-based method is not suitable for the robot. In addition, the size of a total system tends to be large. This means that the number of parameters for the system increases and more computational power is required by the system. Because the room and computational power a robot can use are limited, they are hard problems when being applied to a robot. Therefore, we focus on single channel approaches in this paper. Consequently, we use multi-condition acoustic model training, MLLR, and MFT.

2.1. Obstacle detection and local map

The VRCS 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 [9] where sensors are activated in sequence of $\{s_1, s_{12}, s_2, s_{11} \ldots\}$. 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,...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 δ_j, 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 *0'*. the distance value $se'_j = \| se'_j \|$ from the origin of the frame $\{R\}'$ to the obstacle and angle $s\varphi'$ can be calculated by Eq.(1) . Here, ||.|| 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:

$$Se_n \leftarrow Se'_j, n = INT\left(\frac{s\varphi'_j}{\overline{\varphi}}\right) + \frac{N}{2}; j = 1,2,\ldots,N \quad (2)$$

Where ← and *INT* denote the updating operation and integer operation, respectively. Here, se_n, denotes the distance value of n th 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 e_j = (j = 1,2 ,... *12*) , the new local map is partially updated as follows :

$se_j \leftarrow e_j, j$ = 1,2 .... *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.

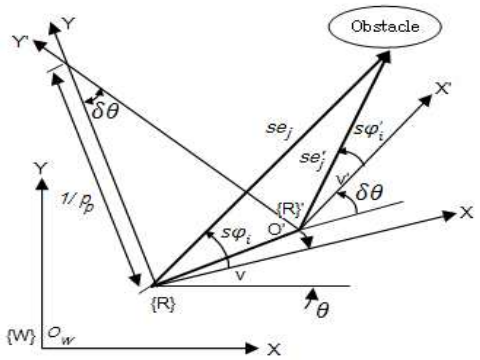


Fig. 2 The coordinate transformation for updating the local map

2.2. Design of Goal-seeking Behavior and Avoidance Behavior

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 T_{ms} 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. 5.

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

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

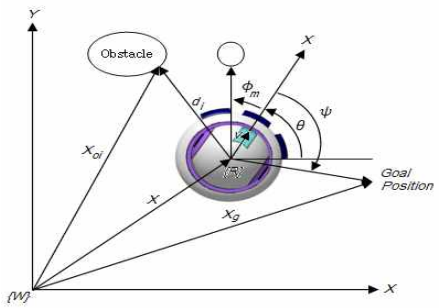


Fig. 3 The coordinate frames and control variables

As shown in Fig 3. $\phi_m(-\pi \le \phi_m \le \pi)$ denotes the orientation of a sector with the shortest range. We choose the input variables for avoidance behavior as ϕ_m and $d_i = \|X_{0i} - X\|, (i = 1,2,3,4)$ for goal-seeking behavior as heading angle difference ψ and distance to goal $z = \|X_g - X\|$. The input linguistic variables d_i, ψ, 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 .

2.3. Fuzzy Decision Maker

New method of selecting an appropriate behavior has been proposed among many primitive behaviors by using a *fuzzy* decision maker. Let $u = (u_1, u_2, u_3, ...u_n)$ be a set of motion commands resulting from the primitive behaviors and $\widetilde{G} = (\widetilde{G}_1, \widetilde{G}_2, \widetilde{G}_3, ...\widetilde{G}_m)$ a set of fuzzy goals by which the suitability of a behavior is judged. The *j-th* fuzzy goal $\widetilde{G}_j$ is characterized by their membership functions $\mu_G(u)$. In what follows, the tilde sign(~) representing the fuzzy sets will be dropped for notational simplicity. However, there are some cases where some goals are of greater importance than others. In such cases, fuzzy decision function *D* might be expressed as the intersection of the goals with the weighting coefficients reflecting the relative importance of the constituent terms. The problem is then to determine one of alternatives $u_i (i = 1,2,3,...n)$ with the highest degree of suitability with respect to all relevant goals $G_j (j = 1,2,3,...m)$ .To this end, the fuzzy set decision *D* in discrete space is defined by

$$D = \begin{bmatrix} \{u_i, \min\{\mu_G(u_i).w_j\}\} \\ i = 1,2...n; j = 1,2...m \end{bmatrix} \quad (5)$$

where $\sum_{j=1}^{m} w_i = 1.0$. Here, the coefficient w_j denotes the importance of goal G_j . The optimal motion command is defined as the output with the highest degree of membership in *D*.

When the mobile robot is located at *O*, as shown in Fig. 4, the repulsive and attractive potentials at the point can be calculated, respectively. Suppose that the mobile robot moves from a point *O* to a point *O'* along the path shown in the figure by the output vector of each behavior. The two potentials at a point *O'* then can be calculated, respectively. Using the above potentials, the differences between potential values at *O* and *O'* are calculated. Thus, when the output of each behavior is applied to the mobile robot, the changes of the repulsive and attractive potentials can be calculated by

$c_{rep}(u_i) = \Delta E_{rep}(u_i)$ and $c_{att}(u_i) = \Delta E_{att}(u_i)$, respectively.

The robot motion is controlled by its linear velocity *v* and rotational velocity *w*. In order to control the mobile robot, the reference posture $P_r(t) = (x_r, y_r, \theta_r)^T$ and the current posture $P_C(t) = (x_C, y_C, \theta_C)^T$ shown in Fig. 7 are used. The reference posture is calculated by the reference velocity (v_r, w_r) which is determined by the output of a behavior selected by fuzzy decision maker. If the output of the selected behavior is $u = (v, \Delta\theta)$, the velocity (v_r, w_r) is defined as $(v, \Delta\theta)$. In order for mobile robot to have the reference velocity at the reference position, the velocities of two wheels must be controlled.

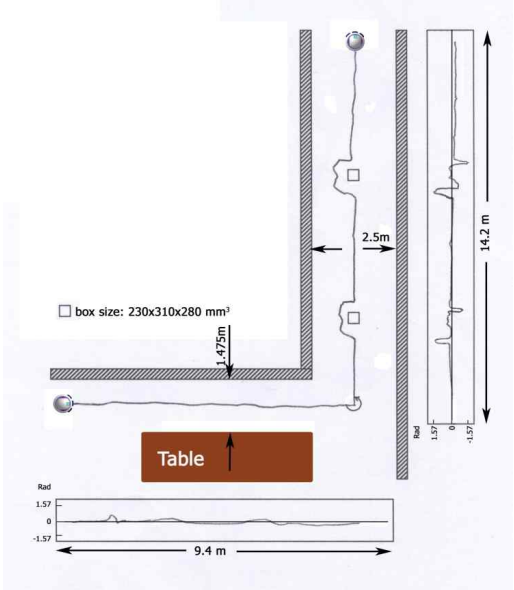
The purpose of this tracking controller is to make the error posture converge to *0*. To achieve this, target velocities are calculated by using the error posture and reference velocities.

3. EXPERIMENTS AND RESULT

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. 8 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.

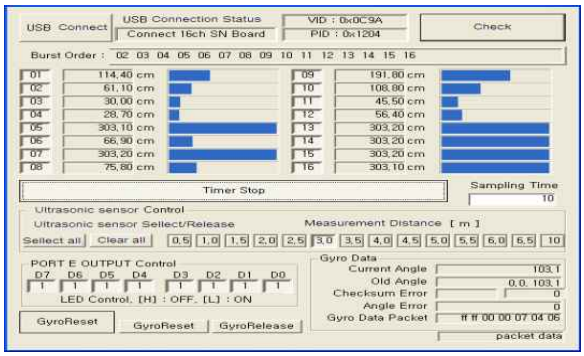
Table 1. The parameter values used for experiments

$D_r = D_l = 0.15m$	*Encoder = 512x4*
$R_r = 0.4m$	K_p=0.8, K_i=0.23, K_d=0.04
v_{max}=0.55m/s	a_{max}=0.2m/sec^2
$\overline{\varphi} = 11.5^0$	c_1=-0.01N.m
c_2=-0.03N.m	E_1=0.9N.m
T_{ms}=60msec	R_{max} = 6.5m



(a) Indoor environment

(b) Voice recognition interface



(c) Obstacle detection using ultrasonic sensors

Fig 4. The indoor environment for voice recognition and navigation experiments

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.

5. 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.

REFERENCES

[1] 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.

[2] 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.

[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] 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.

[5] 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.

[6] J.-M. Valin, J. Rouat, and F. Michaud, "Enhanced robot audition based on microphone array source separation with post-filter," in Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004.