JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 192-200
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3289

Implementation of MFCC and SVM for Voice Command Recognition as Control on Mobile Robot

Rendyansyah
Electrical Engineering Department, Faculty of Engineering, University of Sriwijaya, Palembang, Indonesia

ARTICLE INFO

Article historys:

Received: 09/09/2022
Revised: 19/09/2022
Accepted: 10/10/2022

Keywords:

Control, Mobile Robot, Voice Command

ABSTRACT

The mobile robot is a system that can move according to function and task. An example is an industrial robot taking objects using a remote control system. Robots controlled using a manual remote system are generally carried out on mobile robots. Many researchers have developed manual control methods, such as image or sound-based robot control. In this study, the mobile robot was applied in an unobstructed room and controlled using voice commands. The methods used are Mel-Frequency Cepstral Coefficients (MFCC) and Support Vector Machine (SVM). MFCC is a characteristic identification of voice command patterns such as “forward”, “backward”, “left”, “right”, and “stop.” SVM is used to recognize voice command patterns based on the value of the MFCC for each pattern. The experiment has been carried out 50 times with a success rate of 96%. Overall the robot can be controlled by voice commands with good movement.

Corresponding Author:

Rendyansyah
Department of Electrical Engineering, Faculty of Engineering, University of Sriwijaya, Palembang, Indonesia
Email: rendyansyah.unsri@gmail.com

1. INTRODUCTION

A robot is a tool that can move to help humans work. Operators in certain tasks generally control worker robots. The robot's tasks include exploring the room, investigating, and patrolling. There are two types of worker robots, namely manipulators and mobile. The robot is a substitute for the role of humans who have physical limitations. Robots that are used to navigate the room are called mobile robots. Mobile robots are usually circular with several levels. Each level is used for various placements of electrical components, sensors, microcontrollers, and also power supplies. Robots, in general, already have a control system in the processor. The control system on the robot is autonomous [1], and there is also operator intervention [2]. The use of the control system aims to regulate the movement and navigation of the robot. On the other hand, applying a manual or semi-automatic control system adjusts the system's movement so that it can monitor the results and processes properly. The manual control of a robot can use a joystick, remote, or guide with an image or voice.

A mobile robot using remote control has been carried out by Hartono et al. [3]. The remote control is in the form of an Android smartphone. Movement data is in the form of a trajectory created in the mobile application and GPS coordinates so that the mobile robot moves to follow the specified trajectory. Devana et al. designed a surveillance robot with a manual control system based on a wireless joystick [4]. The robot is equipped with a mobile application-based camera so the operator can receive information about the robot. The development of manual control has entered the stage of using image processing methods and even voice commands. There are control systems on the robot that use images

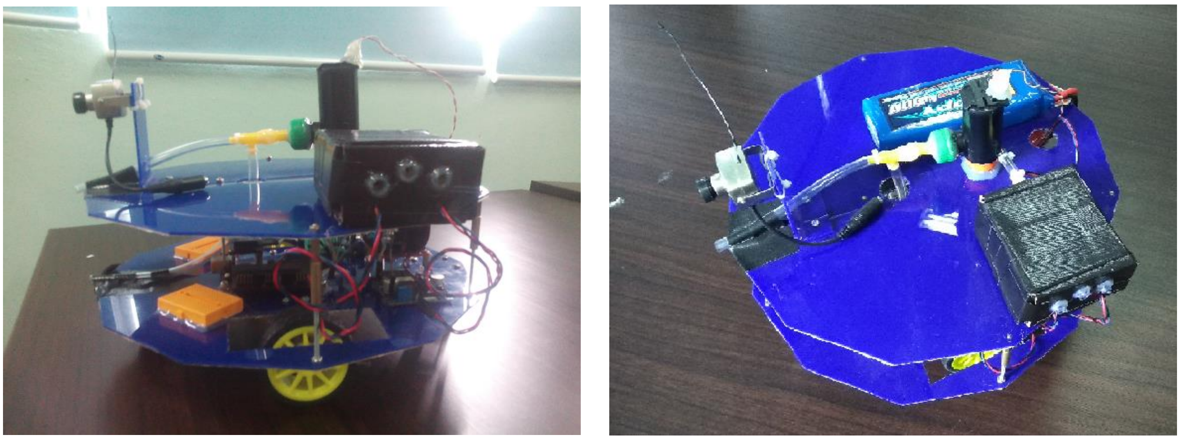
Volume 9, Issue 2, October 2022, pp. 192-200
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3289

such as hand gestures [5] and eye tracking [6]. On the other hand, the implementation of speech recognition also supports robot control [7]. Azargoshasb et al. use voice command recognition to control the robot [8]. The commands given are "go", "stop", "left", "right", "forward", "backward", etc. The feature recognition technique they use is the Mel-Frequency Cepstral Coefficients (MFCC) method. Pattern recognition and speech characteristics are usually done using artificial intelligence or intelligent systems.

The application of intelligent systems is a compliment to controlling robots. The intelligent system is a method for creating a system that can make decisions based on the logic embedded in the microcontroller. Intelligent systems consist of methods such as Fuzzy Logic, Artificial Neural networks, Data Mining, Probabilistic, Reasoning, etc. On the other hand, there is a method similar to the Artificial Neural Network, namely the Support Vector Machine (SVM) method. SVM is supervised machine learning in data clustering as well as pattern recognition. Our previous research also implemented SVM for gas identification based on mobile robots [9, 10]. Therefore, this study developed a voice command-based mobile robot control method by applying the MFCC and SVM patterns. In several related experiments, SVM has similar accuracy to Artificial Neural networks and other clustering methods [11, 12] and is a good learning process for nonlinear data using kernel functions [13]. The motivation for developing this research is based on implementing a mobile application or computer-based patrol robot for investigation. The implementation of voice commands provides convenience in real-time teleoperation.

2. RESEARCH METHOD

The mobile robot is one type of robot that can move to help humans work. Mobile robots generally have wheels to move from one point to another. In this research, we develop a mobile robot connected to a computer. This mobile robot consists of a microcontroller module, an actuator, and Bluetooth communication media. Figure 1 shows the mobile robot used in this experiment. The mobile robot is a circle with a diameter of 40 cm with two levels. The first level is for the placement of the microcontroller and driver module, on the second level is for the placement of batteries and other electrical modules. This robot moves based on voice commands given by the operator.



(a) (b)

Figure 1. Mobile robot, (a) front view, and (b) top view

The operator gives voice commands through a microphone connected to a computer. The voice commands include "Forward", "Backward", "Left", "Right", and "Stop". The MFCC and SVM process the command on the computer. The results of the decision from the SVM are then transferred to the microcontroller via Bluetooth communication. The microcontroller processes the digital data and transmits the signal to the motor driver to drive the robot. A schematic illustration of the diagram system in the mobile robot is shown in Figure 2.

Volume 9, Issue 2, October 2022, pp. 192-200
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3289

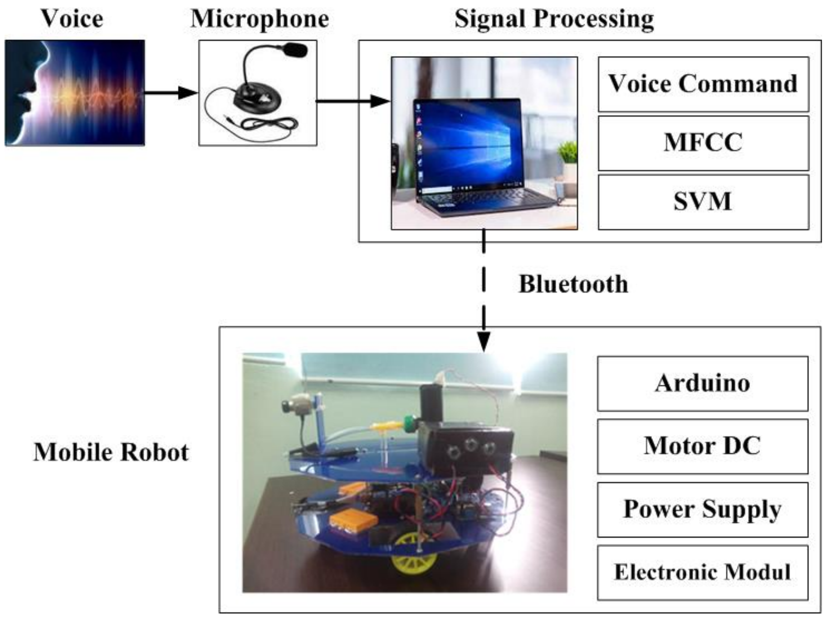


Figure 2. Diagram of a voice command-based mobile robot

The operator's voice is recorded on the computer for 2 seconds, and pre-processing is carried out to take the sound detection part. The detection of sound in the form of signal waves can be seen in Figure 3. The sound signal is processed digitally with a sampling frequency of Fs = 11000 Hz. The voice signal needs to be processed to find information in the form of features. The feature extraction method uses Mel-Frequency Cepstral Coefficients (MFCC) in this research.

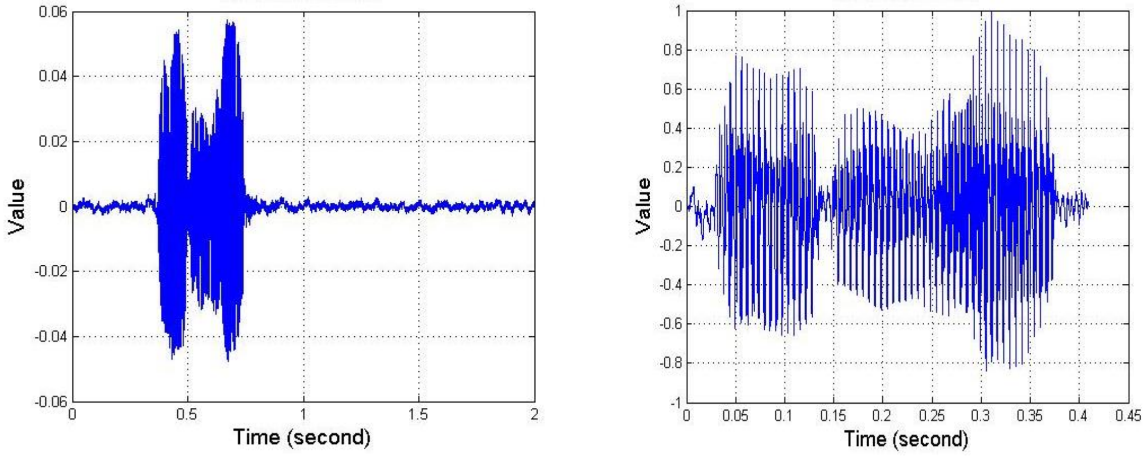
The MFCC method is one of the methods used in recognizing voice command patterns or voice signals. The stages of the MFCC are as follows [14]:

Stage 1. Detect sound using a microphone. The voice signal is recorded (see Figure 3. a) and cropped for voice data only (see Figure 3. b).

Stage 2. Counting the number of frames from the sample signal with (1).

$$nFrame = \frac{Signal\,length}{Frame\,length} \quad (1)$$

The frame length used is 256.



(a) (b)

Figure 3. Signal of voice, (a) forward command, and (b) voice detection

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 192-200
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3289

Stage 3. Perform the Hamming Window process or a process called windowing. This process uses (2).

$$Y[n] = X(n) * W(n) \quad 0 \le n \le N-1$$
$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \tag{2}$$

Description for variable Y[n] as signal output, X(n) signal input, W(n) Hamming window process, and N number of samples per frame.

Stage 4. The Fast Fourier Transform (FFT) method processes the signal into a frequency response.

Stage 5. Calculating Mel frequency using (3). The Mel-frequency scale is a linear frequency below 1000 Hz and logarithmic above 1000 Hz.

$$F(Mel) = \left[2595 * \log 10\left(1 + \frac{f}{700}\right)\right] \tag{3}$$

Stage 6. Converting log Mel spectrum into the time domain using Discrete Cosine Transform (DCT) as shown in (4). Determine the coefficient value of the MFCC from n = 1 to k.

$$\tilde{C}_n = \sum_{k=1}^{K} \left(\log \tilde{S}_k\right) \cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{K}\right] \tag{4}$$

In this study, 20 coefficient values from the MFCC were used. These coefficient values become input for the pattern recognition process using the Support Vector Machine (SVM). SVM is implemented for voice command pattern recognition. SVM is a supervised machine learning technique that can cluster data linearly and non-linearly. SVM uses kernel functions to create a hyperplane to separate data by class. In general, SVM is used to separate two classes, so SVM has evolved to be able to separate multi-class cases. This study used the SVM one vs others technique to recognize voice commands such as “forward”, “backward”, “left”, “right”, and “stop”. Figure 4 shows a diagram of the SVM one vs others technique for recognizing patterns of voice commands.

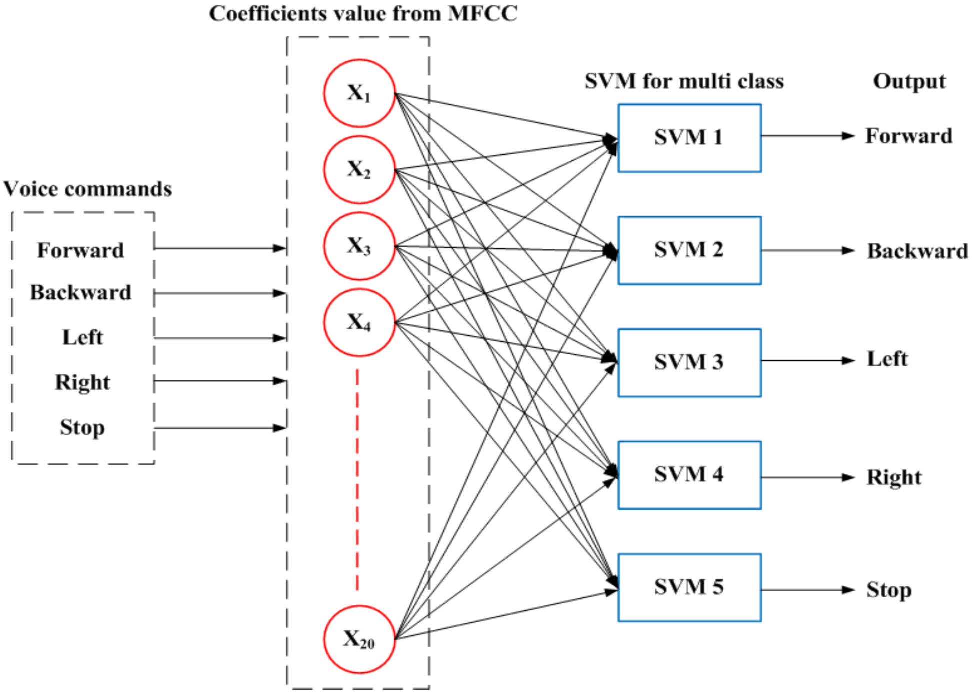


Figure 4. Diagram SVM one vs others

SVM training stage by following the sequential algorithm [9, 15] as follows:

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 192-200
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3289

1. Initialize the value $\alpha_i = 0$.
2. Calculating the D_{ij} matrix, see (5).

$$D_{ij} = y_i y_j \left(K \left\langle x_i, x_j \right\rangle + \lambda^2 \right) \tag{5}$$

for $i, j = 1$ to l. l is the number of data, and the parameter λ is a constant.
3. Each pattern to be clustered is calculated using (6), where the training data is from i = 1 to l.

$$\begin{aligned} E_i &= \sum_{j=1}^{l} \alpha_j D_{ij} \\ \delta\alpha_i &= min\left\{ max\left[\gamma\left(1 - E_i\right), -\alpha_i \right], C - \alpha_i \right\} \\ \alpha_i &= \alpha_i + \delta\alpha_i \end{aligned} \tag{6}$$

The parameter γ is the learning rate.
4. If the training process looks convergent, the process is stopped. Conversely, if it has not converged, repeat the process in (6). The indicator is convergent if the rate of change in value is low or close to zero.

This SVM algorithm produces an optimal Lagrange α value for clustered data, where the input data is converted to dimensional features using the Radial Basis Function kernel function as in (7). For is a kernel parameter with a value determined by the programmer, this study uses the kernel parameter 1000 [9, 13].

$$K\left(\vec{x}, \vec{y}\right) = \exp\left(-\gamma \left\| \vec{x} - \vec{y} \right\|^2 \right) \tag{7}$$

The testing stage follows the following rules:
1. Determine the number of classes in each SVM, namely clusters +1 and -1.
2. Solve the input data into feature space using the Radial Basis Function kernel.
3. Calculate the last decision-making function using (8).

$$\begin{aligned} f(x) &= sign\left(\sum_{i \in SV} \alpha_i y_i K\left(x_i, x\right) + \alpha_i y_i \lambda^2 \right) \\ sign(k) &= \begin{cases} -1 & ; \quad k < 0 \\ 1 & ; \quad k \geq 0 \end{cases} \end{aligned} \tag{8}$$

The value of SV is the number of support vectors, and sign(k) is used to determine the classification.

3. RESULTS AND DISCUSSION

The mobile robot experiment was carried out in the Robotics and Control Engineering Laboratory, Faculty of Engineering, UNSRI. First, an operator gives voice commands through the microphone installed on the computer. In the computer, MFCC and SVM programs are used to process voice signals into commands in controlling the robot's movement. The recognized voice commands are "forward", "backward", "left", "right", and "stop" commands. The MFCC program extracts the characteristics of the voice signal. The coefficient values of the MFCC are input into the SVM learning process. Figure 5 shows the pattern form of the MFCC for voice commands. Each voice command forms a unique pattern recognized by SVM learning. Each voice command takes as many as 10 sample data, so the total becomes 50. The SVM one vs others method (see Fig. 4) has five SVM blocks for each voice command cluster. SVM is a cluster for two classes, namely classes +1 and -1. Therefore each SVM block compares the clustered data against the other data.

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 192-200
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3289

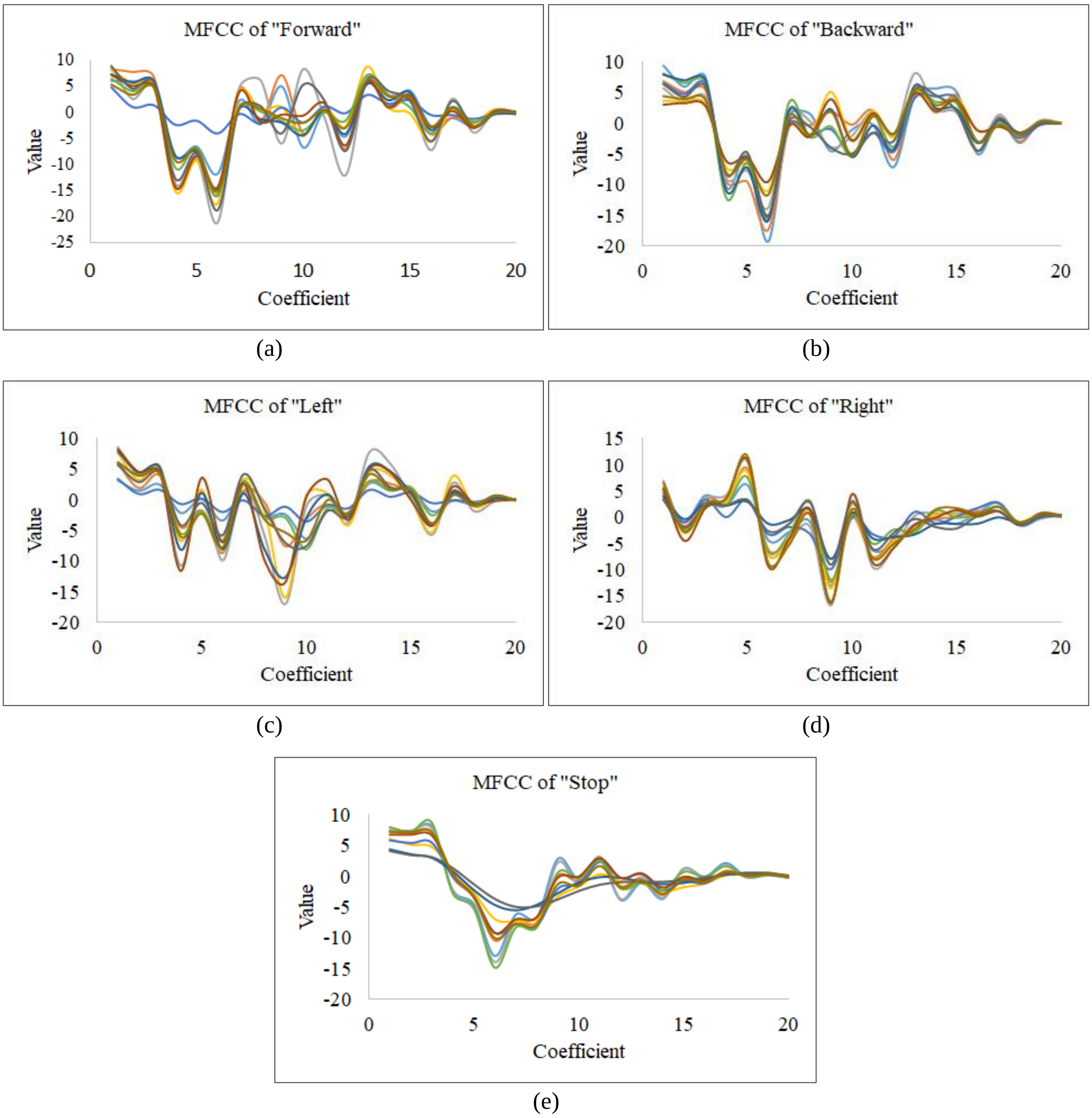


Figure 5. MFCC pattern of voice commands, (a) Forward, (b) Backward, (c) Left, (d) Right, and (e) Stop

Based on Figure 5, the MFCC pattern of this voice command has certain characteristics to the pattern of the voice command. SVM can classify the pattern of MFCC-based voice commands, and a good training process can be seen as shown in Figure 6. Training on SVM reaches convergence at the 100th epoch, and there is no change in the value of the support vector. The results of SVM learning are used to control the movement of the mobile robot based on voice commands. The mobile robot is placed in a free arena, and the robot is active according to orders given by the operator. Figure 7 illustrates an experiment on a mobile robot when it moves based on voice commands. Experiments on the movement of the robot for "forward", "backward", "left", "right", and "stop" carried out movements ten times each. The total number of experiments is 50 times.

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 192-200
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3289

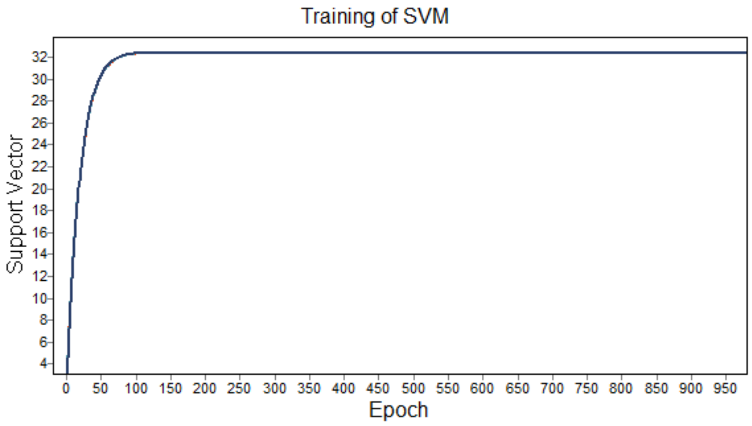


Figure 6. Training of SVM for voice command recognition

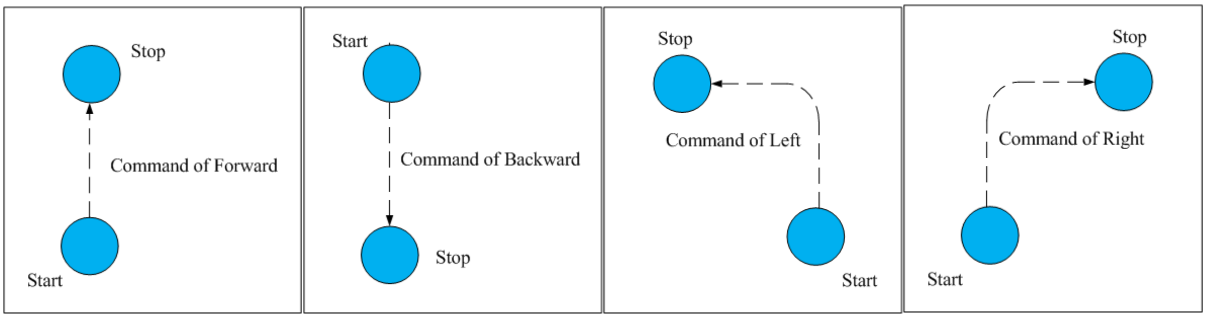


Figure 7. The illustration of the robot's movement when given a voice command, (a) forward, (b) backward, (c) left, and (d) right and stop

The results of the mobile robot experiment with voice commands are summarized in Table 1. For the forward and backward commands, each has failed once. The failure of the forward and backward movement is due to an almost similar pattern. In Figure 7. a, the robot is in the enabled position, and the operator gives a forward command. Within ± 3 seconds, the robot responds to the command to move forward continuously. The operator can give a stop command, and the robot responds to a stop. On the backward command, the robot returns to its original position (see Figure 7. b). On the left and right commands (see Figures 7. c and 7.d), the robot adjusts left or right navigation by deactivating one of the motors.

Table 1. The Experiment of recognizing voice commands to robot's responses

Command	Experiments	Success
Forward	10 times	9
Backward	10 times	9
Left	10 times	10
Right	10 times	10
Stop	10 times	10
Rate		96%

The results of this experiment show that the mobile robot can respond well to voice commands given by the operator. Applying the MFCC and SVM methods can give the robot system the ability to recognize voice command patterns. The experiments in Table 1 show that the robot can move according to voice commands with a success rate of 96%.

JURNAL ECOTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 192-200
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3289

4. CONCLUSION

The mobile robot is a type of robot that can navigate to ease human tasks. This research has developed a mobile robot for voice command-based navigation. Voices from the operator are "forward", "backward", "left", "right" and "stop" commands. The pattern recognition method for voice commands uses MFCC, and 20 coefficients are created. The value of these coefficients becomes input data for the learning process using SVM. SVM can classify voice command patterns with one vs others technique. The experiments for each command are ten times, so a total of 50 experiments. There are two failures because there is a similar pattern for the forward and backward commands. The experimental results show that the system can recognize voice command patterns with success reaching 96%, and the robot can respond well to commands within a waiting time interval of ± 3 seconds.

REFERENCES

[1] D. Hutabarat, D. Purwanto, M. Rivai, and H. Hutomo, "Lidar-based Obstacle Avoidance for the Autonomous Mobile Robot," in *12th International Conference on Information & Communication Technology and System (CTS)*, 2019, pp. 197–202.

[2] Azeta, Bolu, Hinvi, Abioye, Anakhu, and Onwordi, "An Android-Based Mobile Robot for Monitoring and Surveillance," in *2nd International Conference on Sustainable Materials Processing and Manufacturing*, 2019, pp. 1129–1134.

[3] R. Hartono, T. N. Nizar, I. Robani, and D. A. Jatmiko, "Motion and Navigation Control System of a Mobile Robot as A Prototype of An Autonomous Vehicle," in *IOP Conf. Series: Materials Science and Engineering*, 2020, pp. 1–7.

[4] M. Devana, T. Dewi, N. L. Husni, P. Risma, and Y. Oktarina, "Desain Robot Pengintai Segala Medan Dengan Kendali Wireless PS 2," *J. Appl. Smart Electr. Netw. Syst.*, vol. 2, no. 2, pp. 64–70, 2021.

[5] T. Dewi, Amperawan, P. Risma, and Y. Oktarina, "Finger Cue for Mobile Robot Motion Control," *Comput. Eng. Appl.*, vol. 9, no. 1, pp. 39–48, 2020.

[6] C. Carreto, D. Gego, and L. Figueiredo, "An Eye-gaze Tracking System for Teleoperation of a Mobile Robot," *J. Inf. Syst. Eng. Manag.*, vol. 3, no. 2, pp. 1–9, 2018.

[7] T. Q. Nguyen, P. Nauth, and S. Sharan, "Control of Autonomous Mobile Robot using Voice Command," in *Proceedings of the ARW & OAGM Workshop 2019*, 2019, pp. 123–127.

[8] S. Azargoshasb, A. H. Korayem, and S. Tabibian, "A Voice Command Detection system for controlling the movement of SCOUT Robot," in *Proceedings of the 6th RSI International Conference on Robotics and Mechatronics*, 2018, pp. 326–330.

[9] Rendyansyah, A. P. P. Prasetyo, and K. Exaudi, "Implementasi Electronic Nose dan Support Vector Machine pada Aplikasi Olfactory Mobile Robot dalam Mengenali Gas," *J. Nas. Tek. Elektro*, vol. 7, no. 1, pp. 69–79, 2018.

[10] Rendyansyah, M. Rivai, and D. Purwanto, "Olfactory Arm Mobile Robot for Object Inspection Based on Fuzzy Logic and Support Vector Machine," in *IOP Conf. Series: Journal of Physics: Conf. Series*, 2019, pp. 1–9.

[11] N. S. A. Yasmin, N. A. Wahab, A. N. Anuar, and M. Bob, "Performance Comparison of SVM and ANN for Aerobic Granular Sludge," *Bull. Electr. Eng. Informatics*, vol. 8, no. 4, pp. 1392–1401, 2019.

[12] D. R. Rizvi, I. Nissar, S. Masood, M. Ahmed, and F. Ahmad, "An LSTM-based Deep learning model for voice-based detection of Parkinson's disease," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 5s, pp. 337–343, 2020.

[13] W. Caesarendra and M. Irfan, "Classification Method of Hand Gestures Based on Support Vector

JURNAL ECTIPE
Electronic, Control, Telecommunication, Information, and Power Engineering
Volume 9, Issue 2, October 2022, pp. 192-200
ISSN 2355-5068 ; e-ISSN 2622-4852
DOI: 10.33019/jurnalecotipe.v9i2.3289

Machine," *Comput. Eng. Appl.*, vol. 7, no. 3, pp. 179–190, 2018.

[14] A. Salahul, C. Setianingsih, M. Nasrun, and M. A. Murti, "Speech Recognition Implementation using MFCC and DTW Algorithm for Home Automation," in *Proceeding of the Electrical Engineering Computer Science and Informatics*, 2020, pp. 78–85.

[15] S. Vijayakumar and S. Wu, "Sequential Support Vector Classifiers and Regression," *Proc. Int. Conf. Soft Comput.*, vol. 619, pp. 610–619, 1999.