2019 58th Annual Conference of the Society of
Instrument and Control Engineers of Japan (SICE)
September 10-13, 2019, Hiroshima, Japan

Hand Gesture Recognition System for In-car Device Control Based on Infrared Array Sensor

Shigeyuki Tateno[1], Yiwei Zhu[2] and Fanxing Meng[3†]

[1-3]Graduate School of IPS, WASEDA University, Kitakyushu, Japan
[1](E-mail: tateno@waseda.jp)
[2](E-mail: zhuyiwei@toki.waseda.jp)
[3] (E-mail: mengfanxing@akane.waseda.jp)

Abstract: Nowadays, with the development of automotive industry, more and more functions and devices are assembled in cars to improve driving experience. Meanwhile, traffic accidents are increasing in recent years. Operations of in-car device human machine interface (HMI) will cause lack of concentration, which is a major cause of the accidents. Common in-car device HMI systems are based on optics, acoustics, and so on, which are faced with environment limitations such as influenced by illumination conditions. In order to deal with these limitations, in this paper, an infrared array sensor is applied to construct a hand gesture recognition system for in-car devices control. The proposed system can overcome disadvantages of other systems and has a wider application. In the system, seven different shapes of hand and movement toward four directions are combined to achieve the aim of device operations. In data processing, convolutional neural network (CNN) is applied to realize recognition. Simulated experiments are conducted to verify the feasibility of this system.

Keywords: Infrared Array Sensor, Hand Gesture Recognition

1. INTRODUCTION

Nowadays, with the development of automotive technology, cars have been more and more functional and intelligent. In order to improve driving experience, in-car devices are commonly installed, such as radio, music player, air conditioner and so on. Usually, drivers need to operate the devices for themselves. Operation of in-car device human machine interface will cause lack of concentration on driving, which can increase the risk of accident [1]. Therefore, a novel human machine interface (HMI) control system is needed to simplify the operation of HMI and improve driving safeness [2].

Common HMI systems are based on touch (tactile), voice (acoustics) and vision (optics) [3]. However, there are limitations in these systems: normally, when people operate HMI systems based on tactile, they need to fix their eyes to their hands to correctly control devices. On the other hand, the application condition of systems based on acoustics and optical are limited: sound is interfered by environment noise, which is a serious problem especially in road conditions, while systems based on optics are heavily limited by illumination condition and cannot work at night.

This research aims at helping drivers operate in-car devices more conveniently and concentrate on driving, and also overcoming the limitations mentioned above. Therefore, a hand gesture recognition system based on infrared array sensor is purposed to control some in-car devices. Compared with other HMI systems, the purposed system is easier to operate and will cost less concentration. Moreover, it can work in various conditions such as at night, in tunnels, and on noisy roads.

In the following parts of this paper, the second section shows information of infrared array sensor applied in this research, the third section explain the hand gesture recognition method, while experiments and results are presented in the fourth section. The fifth section shows the conclusion and future work.

† Fanxing Meng is the presenter of this paper.

2. INFORMATION OF INFRARED ARRAY SENSOR

Compared with commonly used optical sensors such as camera, infrared-based sensors have advantages of less influenced by environment, which means that they are more reliable and have a wider application field. Consequently, high-resolution infrared thermographic cameras are widely adopted in military application, medical diagnosis, industrial equipment diagnosis and so on. However, high-level thermal camera is too expensive and impossible to be applied in mass production [4]. As a replacement, low-resolution thermal cameras, in other words, infrared array sensors, which have from dozens to hundreds of pixels in a grid layout, are becoming popular in human oriented researches and increasingly applied in human existence detection, fall detection, human monitoring and so on.

Some researchers use a group of infrared array sensors to monitor the movement of pedestrians by detecting human existence and create trajectories of pedestrians [5], while other researchers use LSTM (long short-term memory) and GRU (Gated recurrent unit) network to perform fall detection based on low-resolution infrared array sensor [6]. Besides, localization [7] and bed-exit detection [8] have also been attempted with infrared array sensor. However, there are few existing researches that concentrate on recognition of more complex patterns.

The sensor applied in this research is MLX90640 far infrared thermal sensor manufactured by Melexis. This sensor has 32*24 thermopile elements that can detect infrared radiation of things in detection area and return a low-resolution thermal image which reflect temperature of targets in the area [9]. Operating temperature of this sensor is from -40°C to 85°C, which means that the sensor can work in all common conditions. Refresh rate of the sensor is programmable which varies from 0.5Hz to 64Hz, means that real-time recognition is possible for this sensor. In this research, 8Hz was applied and it can meet the experiment demand. Meanwhile, I2C compatible digital interface is offered for the data transmission. In this research, data obtained from sensor is sent to a microcomputer and then transmitted to computer through UART, where gesture recognition is performed.

3. RECOGNITION METHOD

After thermal image data is obtained from the sensor, a recognition method is performed to achieve the aim of hand gesture recognition.

The target of the system is to detect seven stable hand gestures shown in Fig. 1, and moving gestures toward four basic directions: up, down, left and right, which is shown in Fig. 2. The gestures are decided considering the balance of being easy to remember (for human) and classify (for system).

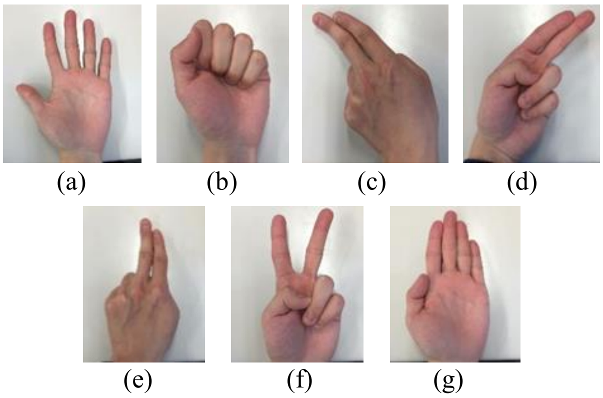


Fig. 1 Stable gestures to be recognized.

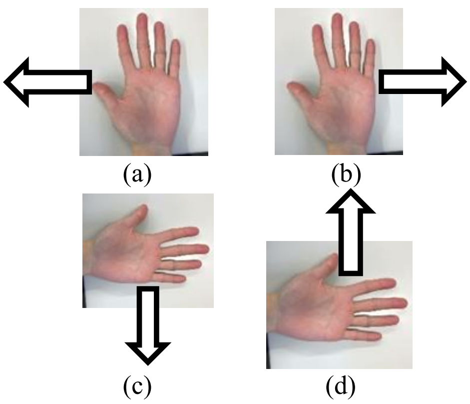


Fig. 2 Moving gestures to be recognized.

Flow chart of the proposed system is shown in Fig. 3. First, preprocessing is executed to make gesture data easier to be recognized. After that, barycenter movement detection is performed to separate stable gestures and moving gestures. When data is judged as a stable gesture, recognition of the gestures is executed using CNN. On the other hand, when significant movement is detected, moving gesture detection is applied to detect simple moving gestures in order to improve practicability of the system.

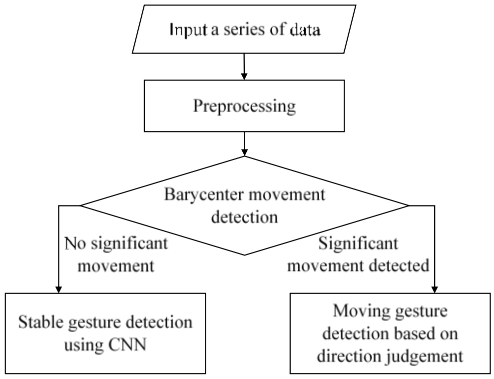


Fig. 3 Flow chart of the proposed system.

3.1 Preprocessing

In recent years, GMM is widely applied in data clustering, image segmentation and background subtraction (especially in video processing) because of its reliability in dealing with changing environments [10]. The main principle of GMM is to assume that data consists of several gaussian distributions, which are corresponding to different objects, such as environment background and target, and fit these distributions by iteratively calculate the parameters of mean and variance.

For a single unidimensional gauss distribution, probability density function is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{1}$$

here μ is a mean value and σ is a standard deviation. Since a GMM method assumes that data is the sum of several gauss distributions, the probability density function under GMM is:

$$p(x) = \sum_{k=1}^{K} \alpha_k N(x|\mu_k, \sigma_k), \tag{2}$$

where $N(x|\mu_k, \sigma_k)$ is the probability density function of K-th gauss distribution, and α_k is the weight of each distribution which satisfies

$$\sum_{k=1}^{K} \alpha_k = 1. \tag{3}$$

In Fig. 4, it shows a frame of data obtained from infrared array sensor in which an experimenter is doing a hand gesture.

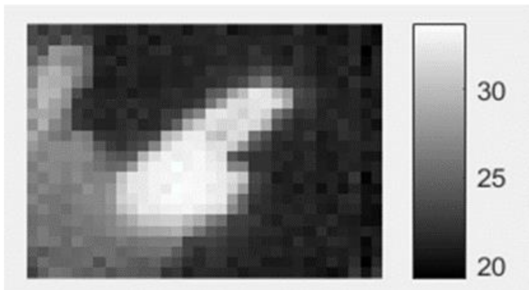


Fig. 4 Infrared image data with people doing a gesture.

The grayscale histogram of this data is shown in Fig. 5. the three main clusters, of which peaks are at around 21℃, 26℃ and 32℃, are corresponding to background, experimenter's body with clothing and hand. Therefore, a temperature threshold will be used to distinguish the area of the gesture from the other areas of the range that are relatively low in temperature.

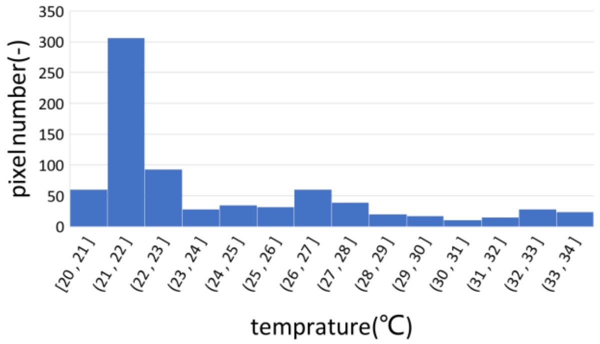


Fig. 5 Grayscale histogram of data in the above figure.

An Expectation-Maximization (EM) algorithm is applied to fit the distribution of the GMM method. In general, EM is an iterative optimization algorithm based on Maximum Likelihood Estimation algorithm that can deal with optimization problem that contain implicit variables. This algorithm is commonly used in calculating GMM distribution. In this research, the initial distribution number is decided as three, and the result of fitting is shown in Table 1. The result of data fitting in Fig. 4 is shown in Fig. 6.

Table 1 Distribution parameters of Fig. 4 fitted by EM.

parameter \ distribution	first	second	third
α	0.4965	0.3850	0.1027
μ	21.3368	25.7225	32.4441
σ	0.5519	2.4040	0.7046

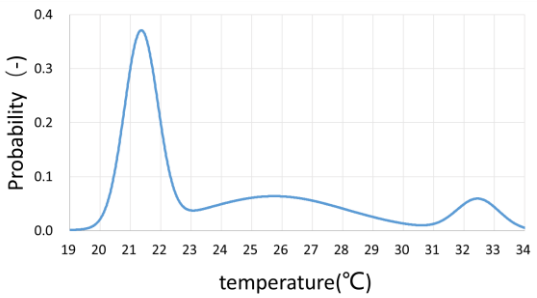


Fig. 6 Distribution of Fig. 4 fitted by EM.

In order to extract hand gesture from background, a threshold is calculated by Eq. (4)

$$Threshold = \mu_2 + (\mu_1 - \mu_2)\frac{\alpha_1}{\alpha_2+\alpha_1}, \quad (4)$$

here μ and α are the mean values and the weight values. Subscript 1 and subscript 2 represent the biggest and the second biggest value, respectively. The image data which is smaller than threshold is removed. The result is shown in Fig. 7.

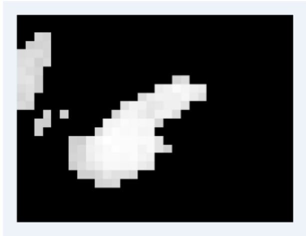


Fig. 7 Infrared image data after removing background.

3.2 Stable hand gesture recognition algorithm based on CNN

Recently, CNN has performed excellently in the field of image recognition, and it has been widely applied in industry [11]. Therefore, it is expected that CNN will show a preferable performance in thermal image recognition.

In this research, a CNN is structured to recognize seven hand gestures shown in Fig. 1 and conditions if there is no gesture performed.

The network consists of two groups of convolutional layers, the Relu layer and pooling layers, two fully connected layers with dropout and result output layers (softmax and classification layers). The structure of CNN is shown in Fig. 8.

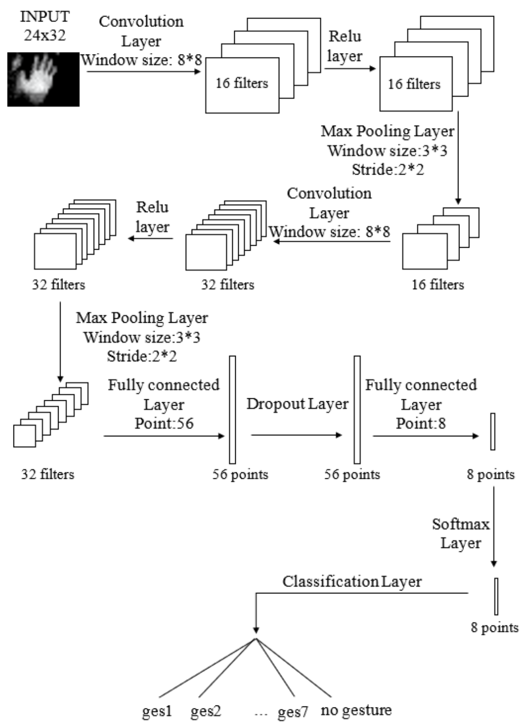


Fig. 8 Structure of the constructed CNN.

After the preprocessing, image data whose size is 24*32, is input into the network. Firstly, there is a convolutional layer with 16 filters and window size of each filter is 8*8. The number of filters is decided according to complexity of the target to be recognized and the number of classes. After that, a Relu layer is applied to finish the work of activation. Then, a max pooling layer with window size of 3*3 and stride of 2*2 for compression is used. Above are the basic structure of CNN, which repeats once again. Higher filter number is used to obtain more hidden features and improve recognition accuracy. After the convolutional part, fully connected layer is attached whose size is 56 points. Here 50% dropout is applied for this layer to avoid overfitting. Finally, fully connected layer of eight points (corresponding to seven gestures and no gesture) is connected to the output part of the softmax layer and the classification layer, which output the classification result.

3.3 Simplified movement detection

In the proposed system, in-car devices to be controlled should be addable, whereas users can select device and finish operation all by gesture. When only stable gestures are assembled, for each device, a gesture should be chosen to switch to it, which will make the system difficult to use when the number of devices grows. Therefore, a simplified moving gesture detection function is applied to finish this part of work.

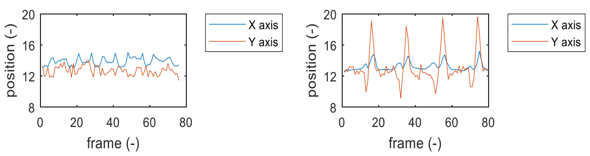
The moving gesture detection part aims to detect movements of hand to four basic directions: up, down, left and right from no movement and slight movement such as doing a stable gesture.

The following is the principle of moving gesture detection. First, a series of image data frame is obtained from the sensor, then, normalization is implemented. After that, barycenter of each image is calculated by Eqs. (5), and (6),

$$x_{center} = \frac{\sum_{x,y} a_{x,y} x}{\sum_{x,y} a_{x,y}}, \tag{5}$$

$$y_{center} = \frac{\sum_{x,y} a_{x,y} y}{\sum_{x,y} a_{x,y}}, \tag{6}$$

where (x_{center}, y_{center}) is the coordinate of the barycenter in this frame of image, and $a_{x,y}$ is the temperature value on point (x, y). The moving gesture recognition is based on the movement of barycenter. Compared with the no gesture and the stable gesture, when there is a moving gesture in a series of data, there will appear a rapid and significant change on position of the barycenter. In Fig. 9 it shows the movement of barycenter in a series of images on X axis and Y axis.



(a) Moving gesture. (b) Stable gesture.

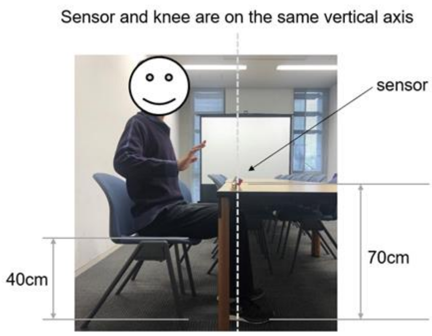
Fig. 9 Movement of barycenter of moving gesture and stable gesture.

In Fig. 9(a), the moving gesture "up" is performed for five times, while the stable gesture in Fig. 9(b) is performed for five times as a comparison. It is clear that the moving gesture cause a more obvious change of the barycenter than the stable gesture. In this research, a sliding window is applied to detect the biggest continuous change of the barycenter in a series of the image. When the change is above the given threshold, it is detected as the moving gesture and moving direction can also be recognized at the same time.

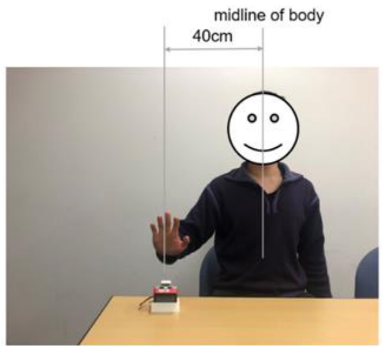
4. EXPERIMENTS AND RESULTS

4.1 Experiment environment and method

The experiments are performed both in in-door condition which is a simulation of in-car environment and in the car environment. The experiment environment can be seen in Fig. 10 (a). The experimenter seats in a specified seat, and the sensor is placed on the table. The height of chair and table is fixed as 40 cm and 70 cm. The sensor and experimenter's knees are on the same vertical axis. The method of installing experiment environment aim at simulating real driving seat. On the front side, as Fig. 10 (b) shows, the sensor is placed 40 cm away from midline of the experimenter's body. In this distance, an experimenter raises his or her hand spontaneously, basically, the sensor is right against the experimenter's hand. And the distance between the hand and the sensor is from 15 cm to 20 cm. The sensor is installed in a fixed angle of 30° from the horizontal, as shown in Fig. 11.



(a) Shown in side.



(b) Shown in front.

Fig. 10 Experiment condition.

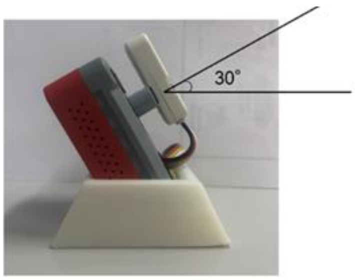


Fig. 11 Installation of sensor.

The reason why an angle is applied is that when people rise their hand spontaneously, their hands tend to incline forward. This installation way of the sensor can improve the accuracy of the system and make users easy to use.

In the stable hand gesture recognition experiment, experimenters repeat the follow things: raise his or her hand beside the sensor making specified hand gesture, without looking at the sensor, and withdraw his arm. During this process, all the frame of data is recorded. After that, frames on which a hand gesture appears is collected as stable gesture data.

In the moving hand gesture detection experiment, for each moving gesture, an experimenter raises his or her hand at start side and move to end side. For example, in moving gesture "up", an experimenter raise hand to a low position and raise up till upper position. In this period, the hand moves through central of sensor's detection area. Speed of hand's movement is changeable according to the sensor's sampling rate. In practice, one second is a typical time needed.

4.2 Result of moving gesture detection experiment

In order to confirm that the system is able to detect moving gestures from stable gestures, moving gesture detection experiment is performed. In this experiment, two groups of data are collected. In the first group, experimenter repeat doing stable gestures for 100 times. Result of this group of experiment shows that no slight movement is detected as moving gesture, which means that the system will not misjudge other movements as moving gestures. In the second group of experiment, experimenter performed each of the four above mentioned moving gesture (up, down, left, right) for 25 times. Detection result of this group is shown in Table 2

Table 2 Result of moving gesture detection.

detect / real	up	down	left	right	none
up	23	2	0	0	0
down	0	24	0	0	1
left	0	0	25	0	0
right	0	0	0	25	0

As table shows, in the 100 times of gestures, two times of gestures are misjudged as other gestures, while one is undetected. Totally, accuracy of this experiment is 97%. This experiment shows that the system can successfully detect moving gesture with a high accuracy and low misjudgment rate.

4.3Result of stable gesture recognition experiment

In the experiment of the stable gesture recognition, five experimenters' data is collected. For each experimenter, 100 frames of images are collected for each gesture. In order to verify the capability of this system to deal with individual difference, cross experiment is conducted. Use four experimenter's data for training and one experimenter's data for test, and repeat this procedure for five times with different experimenter's data for test. The result is shown in Table 3.

Table 3 Result of stable gesture recognition with experimenters' data not trained.

detected / real	g1	g2	g3	g4	g5	g6	g7
g1	419	0	0	31	7	4	39
g2	0	422	41	0	9	0	28
g3	1	21	478	0	0	0	0
g4	38	0	0	462	0	0	0
g5	8	1	0	0	456	34	1
g6	3	1	0	0	99	397	0
g7	29	0	0	23	6	14	428

Totally, in 3500 gestures, 3062 gestures are correctly recognized, the accuracy is 87.5%. For each gesture, the accuracy is shown in Table 4. For each experimenter, the accuracy is shown in Table 5.

Table 4 Result of stable gesture recognition divided by gesture.

gesture	accuracy
g1	83.8%
g2	84.4%
g3	95.6%
g4	92.4%
g5	91.2%
g6	79.4%
g7	85.6%

Table 5 Result of stable gesture recognition divided by experimenters.

experimenter	accuracy
e1	97.7%
e2	98.3%
e3	82.1%
e4	75.0%
e5	84.4%

It is obvious that g6 is lower than the other gestures and the accuracy of experimenter 5 (e5) is also lower than other experimenters. Approximately 20% of the amount is recognized as g5. It could be considered that the similarity between g5 and g6 is the main reason for this result. Moreover, the pixels' number collected from the e5 is about only 80% of the other experimenters, and that should be caused by that the size of his hands is much less than the others. Therefore, the size of the hands also has an impact on the results. In subsequent experiments, a large amount of data is required for training to improve accuracy.

4.4 Result of stable gesture recognition experiment in the car environment

In the experiment of stable gesture recognition in the car environment, 100 frames of images are collected for each gesture, and the result is shown in Table 6. Totally, in 700 gestures, the accuracy is 88.7%. As shown in Tables 3 and 6, there is no significant difference between the results of the actual experiment in the car and the results of the simulation environment. Therefore, it can be concluded that the sensor can adapt well to the in-car environment and work properly.

Table 6 Result of stable gesture recognition in the car environment.

detected / real	g1	g2	g3	g4	g5	g6	g7
g1	83	3	0	0	0	4	10
g2	13	87	0	0	0	0	0
g3	0	0	92	0	0	8	0
g4	10	0	8	82	0	0	0
g5	6	0	0	0	94	0	0
g6	5	0	0	1	0	92	2
g7	9	0	0	0	0	0	91

4.5 Result of stable gesture recognition experiment in the dark environment

In order to test the effect of lightness on the experiment, the experimenter will be in a completely dark environment during the experiment. Meanwhile, because of no significant difference between the results of the actual experiment in the car and the results of the simulation environment, the experiment will be conducted indoors. For each gesture the accuracy is shown in Table 7.

Table 7 Result of stable gesture recognition in the dark environment.

detected / real	g1	g2	g3	g4	g5	g6	g7
g1	95	0	0	4	0	1	0
g2	4	94	0	2	0	0	0
g3	0	1	96	2	0	1	0
g4	10	0	8	82	0	0	0
g5	2	0	0	0	98	0	0
g6	6	0	0	2	0	92	0
g7	13	9	0	0	0	2	76

As shown in Table 7, the accuracy is 90.4% in 700 gestures. The results show that the sensor performs better in a completely dark environment, and at the same time, it means that the sensor can be used normally at night. As for the reason why the accuracy of gesture 7 is much lower than others, it is considered to be caused by the similarity between g1 and g7.

5. CONCLUSION

In this paper, a hand gesture recognition system based on infrared array sensor is proposed to control in-car devices in order to simplify in-car device HMI operation and help drivers concentrate on driving, which can improve driving safety. In this system, seven stable gestures and four simple moving gestures are applied to be recognized. GMM algorithm is used to realize background subtraction, while CNN is chosen as the recognition method. Also, a simplified moving gesture detection method is proposed to detect the movement of hand toward four detections. In-door simulate experiments are performed to test the performance of this system. For stable hand gesture recognition, the total accuracy is 87.5% without previous training. Moving gesture detection achieve the accuracy of 97%.

In the following research, more gestures will be applied in the system, while moving gestures can be combined with stable gestures to make the system more convenient to use. Meanwhile, algorithm should be improved to increase accuracy. Besides, real in-car experiments need to be performed to verify the capability of this system, while the influence of season and clothing should be considered to improve the reliability of this system.

REFERENCES

[1] Allert S. Knappe, “Do in-car devices affect experienced users' driving performance?”, *Journal of international association of traffic and safety sciences*, Vol. 39, Issue 1, pp. 72-78, 2015.

[2] Shalini Anant, “Safe Driving using Vision-based Hand Gesture Recognition System in Non-uniform Illumination Conditions”, *Journal of ICT Research and Applications*, Vol. 12, No. 2, pp. 154-167, 2018.

[3] Shalini Anant, “Sensor-Based Hand Gesture Control System for Safe Driving”, *Jour of Adv Research in Dynamical & Control Systems*, Vol. 10, 03-Special Issue, pp. 690-698, 2018.

[4] H. Mohammadmoradi, "Measuring People-Flow through Doorways Using Easy-to-Install IR Array Sensors," *13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, Ottawa, pp. 35-43, 2017.

[5] Kerridge Jon, “Using Low-Cost Infrared Detectors to Monitor Movement of Pedestrians: Initial Findings”, *Transportation Research Record*. Vol. 1878. pp. 11-18, 2004.

[6] Xiuyi Fan, “Fall Detection with Unobtrusive Infrared Array Sensors”, *Multisensor Fusion and Integration in the Wake of Big Data, Deep Learning and Cyber Physical System*, pp 253-267, 2017

[7] Y. Kobiyama, "Recognition of Frequently Appeared Locations/Activities Based on Infrared Sensor Array," *3rd IEEE International Conference on Cybernetics (CYBCONF)*, Exeter, pp. 1-5, 2017

[8] Chiu, Sheng-Yang, “A Convolutional Neural Networks Approach with Infrared Array Sensor for Bed-Exit Detection”, *International Conference on System Science and Engineering (ICSSE)*. pp. 1-6, 2018

[9] S. Pleshkova, “Development of system model for audio visual control of mobile robots with voice and gesture commands”, *2017 40th International Spring Seminar on Electronics Technology (ISSE)*, Sofia, pp. 1-4, 2017.

[10] Horng-Horng “Regularized background adaptation: a novel learning rate control scheme for Gaussian mixture modeling”, *IEEE Transactions on Image Processing*, Vol. 20, Issue 3, pp. 822-836, 2011

[11] R Arandjelovic, “NetVLAD: CNN Architecture for Weakly Supervised Place Recognition”, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5297-5307, 2016