In-Vehicle Device Control System by Hand Posture Recognition with Movement Detection Using Infrared Array Sensor

Shigeyuki Tateno*, Yiwei Zhu*, and Fanxing Meng*

Abstract: Nowadays, with the development of automotive driving technologies, more and more functions and devices with control systems based on tactile, optical, and acoustic sensors are assembled into cars. However, these systems are faced with environmental limitations such as environmental noise and illumination conditions. Moreover, operations of these systems will cause lack of concentration on driving, which is a major cause of car accidents. In order to overcome these limitations, in this paper, an infrared array sensor is applied to construct a hand posture recognition system for in-vehicle device control. In the system, 10 kinds of target hand postures and posture movements toward four directions are combined to achieve the aim of the device selection and operations. The input images are separated into images with objects and without objects. Then, images in which object appears in boundary areas as well as blurred images are removed to improve the accuracy of the system. A convolutional neural network is applied as a classifier in order to realize the recognition of the 10 target hand postures and non-target postures for the in-vehicle device selection. After that, a detection method of the posture movement directions is applied for the device operations. Both indoor and in-vehicle experiments are conducted to verify the robustness of this system, and the results show that the proposed system can overcome the disadvantages of other systems and has a wide application with high accuracy.

Key Words: human machine interface, infrared array sensor, hand posture recognition.

1. Introduction

Nowadays, with the development of automotive driving technologies, cars have been more functional and intelligent. In order to improve driving experience, many kinds of in-vehicle devices are commonly installed, such as radios, music players, and air conditioners. Usually, drivers need to operate the devices by themselves. However, the operations of traditional in-vehicle human machine interface devices sometimes cause lack of concentration on driving, which will increase the risk of car accidents [1]. Therefore, a novel human machine interface (HMI) control system is needed to change the operating methods of HMI as well as improve driving safeness [2].

Common in-vehicle HMI systems are based on touch (tactile), voice (acoustic), and vision (optical) [3]. However, there exist some limitations in these systems. When people operate the systems based on tactile such as haptic control and touchpads, they need to focus their attention on the device screen, which causes visual and cognitive distractions [4]. On the other hand, suitable application conditions of the systems based on acoustic and visual devices are highly required. Although in-vehicle voice-controlled infotainment systems are becoming increasingly common, the problem that the accuracy degrades when the background noise presents has not been solved. As pointed out by N. Sokol in 2017 [5], both noise-sensitive and noise-robust simulated voice-controlled infotainment systems in automobiles are compared. It is found that these two kinds of systems are perceived to be less useful and satisfying in driving performance. Meanwhile, the systems based on vision are heavily limited under the weak illumination and the non-uniform illumination conditions. In 2018, the research of S. Anant et al. [2] showed that although the vision-based hand gesture recognition system can reach 91% classification accuracy in non-uniform illumination environment, this system cannot work in dark environment. Therefore, it is necessary to use infrared sensors to recognize gestures in complete darkness.

Compared with commonly used sensors such as cameras and microphones, infrared-based sensors have their own advantages [6]. Some researchers use a group of infrared array sensors to monitor the movement of pedestrians by detecting human existence and create trajectories of pedestrians [7], while other researchers use a long short-term memory and gated recurrent unit models to construct a fall detection system based on infrared array sensors [8]. Besides, location [9] and bed-exit detection [10] have also been attempted with infrared array sensors. These studies show that infrared-based sensors have advantages of less influenced by environmental conditions in human activity detections.

Therefore, when considering the constraints of actual in-vehicle environment like noise, illumination conditions, and car vibration [11], infrared array sensors are more suitable for the in-vehicle device control. In this paper, a hand posture recognition system based on an infrared sensor is proposed to realize the in-vehicle devices operations. The working mechanism of the proposed system has been partially described in the prior conference publication [12]. However, the accuracy of the system proposed in the conference paper is 87.5%, which is not high enough for actual applications. According to the image data analysis, thermal images in which postures are in the boundary area or unclear will highly influence the classification results. Therefore, in this paper, two new methods of separating

* Graduate School of Information, Production and Systems, Waseda University, Kitakyushu 808–0135, Japan
E-mail: tateno@waseda.jp, zhuyiwei@toki.waseda.jp, mengfanxing@akane.waseda.jp
(Received October 25, 2019)
(Revised January 15, 2020)

JCMSI 0003/201303–0148 © 2019 SICE
images in the boundary area and blurred images, respectively, are proposed to improve the accuracy of the system. Ten target postures are proposed to be recognized in this system. In addition, the class of non-target postures is added to avoid the influence of postures that do not belong to the target postures. To evaluate these proposed methods, more hand postures are designed, and both indoor and in-vehicle experiments are conducted. Moreover, the performances of different classifiers on processing postures image data are compared.

2. Overall Description of System

2.1 Infrared Array Sensor

The sensor applied in this research is an MLX90640 (55° × 35°) far-infrared thermal sensor manufactured by Melexis. This sensor has 32 × 24 thermopile elements that can detect infrared radiation of objects in the detection area and return a low-resolution thermal image which reflects objects temperatures in the area [13]. Operating temperature of this sensor is from −40°C to 85°C, which means that the sensor can work in almost all common in-vehicle conditions. The refresh rate of the sensor is programmable from 0.5 Hz to 64 Hz, which means that the real-time recognition is possible. In this research, up to 16 Hz is applied, and it can meet the experiment demand. Data obtained from the sensor is sent to a microcontroller and transmitted through Wi-Fi.

2.2 Hand Postures and Moving Direction

In order to control in-vehicle devices, in this paper, two processes are set, the static hand posture recognition process and the moving posture detection process. The former is used to select a specific device by detecting static postures shown in Fig. 1, and the latter is used to operate the selected device by detecting the movements of the postures in four directions shown in Fig. 2. The postures are decided by considering the balance of being easy to remember (for human) and classify (for system).

2.3 System Flow Chart

The flow chart of the proposed system is shown in Fig. 3. After a series of thermal images is obtained from the infrared array sensor, image pre-processing is executed firstly to separate the thermal images into background-only images and object images, and the background of the object image is removed to extract a binary object image. Then, static posture detection is applied to remove those images in which objects appear in the boundary area or objects are blurred due to fast hand movements. These images are defined as no-posture images. After these steps, the remaining images in which clear objects appear in the middle area are regarded as static hand posture images. In order to recognize some other postures that do not belong to the target postures, the class of non-target posture is defined. When considering the actual in-vehicle implementation, the target hand posture images along with the non-target posture images are both classified by a convolutional neural network (CNN). The classification result is composed of 11 categories: non-target postures and 10 different kinds of target hand postures. Finally, moving hand posture detection is applied to detect the moving direction of the target hand postures to improve practicability of this system.

3. Methodology

3.1 Image Pre-Processing

After collecting thermal images from the sensor, the image pre-processing step is executed to separate the obtained thermal images into background-only images and object images. The background of the object image is removed to extract a binary object image as well.

In recent years, the Gaussian mixture model (GMM) is widely applied in data clustering, image segmentation, and background subtraction, especially in video processing, because of its reliability in changing environment [14]. The main principle of GMM is to assume that data consists of several Gaussian distributions, which are corresponding to different objects, such as environment backgrounds and detecting targets. It fits these distributions by the iterative calculation of the parameters for a mean and a variance. For a single unidimensional Gaussian distribution, the probability density function is
f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},
\tag{1}
where \(\mu\) is the mean value and \(\sigma\) is the standard deviation. Therefore, 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 Gaussian distribution, and \(\alpha_k\) is the weight of each distribution which satisfies
\[\sum_{k=1}^{K} \alpha_k = 1.\]
\tag{3}

An expectation-maximization (EM) algorithm is applied to fit distributions of the GMM method. In general, the EM is an iterative optimization algorithm based on the maximum likelihood estimation that can solve optimization problems with implicit variables. A frame of image data while a hand posture is performed is shown in Fig. 4.

The grayscale histogram of the image data is shown in Fig. 5. It is obvious that the temperature values are categorized into three clusters, of which peaks are at around 21°C, 27°C, and 32°C. In this case, the number of Gaussian distributions is supposed to be three.

After applying the EM algorithm to fit the GMM method, the result of image data distribution is shown in Fig. 6.

Finally, a temperature threshold is used to distinguish the area of the hand posture from the other areas of the thermal image that are relatively low in temperature values. This threshold is calculated by Eq. (4):
\[\text{Threshold} = \mu_2 + (\mu_1 - \mu_2) \frac{\alpha_1}{\alpha_2 + \alpha_1},\]
\tag{4}
where \(\mu_1\) and \(\mu_2\) are the mean values of the first and the second Gaussian distributions, and also \(\alpha_1\) and \(\alpha_2\) are weight values of the first and the second Gaussian distributions. The image data values which are less than the threshold are removed. Therefore, the result of the object detection is composed of the binarized image data, which is shown in Fig. 7.

### 3.2 Static Posture Detection

The static posture detection is used to separate the posture images from object images after the image pre-processing step, which means the object images are divided into two classes: posture images and no-posture images. The static posture detection is composed of two detecting methods: boundary detection and stop detection.

The boundary detection is applied to remove those images in which object appears in the boundary area. During the process of making postures introduced in detail in Section 3.1, it sometimes appears that only a part of hand enters the detecting area of the sensor. This situation causes the incompleteness of the obtained postures, which leads to misjudgments of the posture.

In the boundary detection, when the object is detected in the boundary area of the image as shown in Fig. 8 (a), this kind of images will be categorized into the class of no-posture. On the other hand, when the object is near the image center as shown in Fig. 8 (b), this image will be considered as a kind of posture images.

The stop detection is applied to obtain the images of the static
In order to avoid overfitting. Finally, a fully connected layer of 11 points is connected to the output part of the softmax layer and the classification layer, which output the classification result.

In order to compare the performance in processing image data between different classifiers, a random forest (RF) classifier [16] is chosen since the RF can be used for image recognition, and it is quite popular in the machine learning field recently [17]. In 2015, O. Sangjun et al. proposed a hand gesture recognition system using the RF as the classifier [18]. Therefore, in this paper, an RF is chosen as a classifier for comparison. The feature extraction method is a normalized pixel difference (NPD) method. The NPD features are computed as the ratio of value difference to sum between two pixel values, which is inspired by the Weber fraction [19]. The NPD features are calculated by Eq. (6):

\[ f(x, y) = \frac{x - y}{x + y}, \]  

where \( x \) and \( y \) are the temperature values of the two pixels, \( f(x, y) \) is the NPD feature value which is bounded in \( -1, 0, 1 \), as shown in Eq. (7). This feature represents the local intensity difference of the thermal images. After the calculation, the NPD features of static posture images are input into the RF for the hand posture recognition.

\[ f(x, y) = \begin{cases} 
-1 & \text{if } x = 0, y \neq 0, \\
0 & \text{if } x = 0, y = 0, \\
1 & \text{if } x \neq 0, y = 0.
\end{cases} \]  

### 3.4 Moving Posture Detection

The following part is the principle of how to determine the moving hand posture direction. First, sequential 10 frames of static hand posture image are obtained, and then, the barycenter of each frame is calculated by Eqs. (8) and (9):

\[ x_{center} = \frac{\sum a_{x} x}{\sum a_{x}}, \]  

\[ y_{center} = \frac{\sum a_{y} y}{\sum a_{y}}, \]  

where \((x_{center}, y_{center})\) is the coordinate of the barycenter in this frame, and \(a_{x/y}\) is the temperature value on the point \((x, y)\).

The moving hand posture detection is based on the movement of the barycenter. Compared with the static hand posture, when the hand posture movement appears in a series of data, there will be a rapid and significant change on the position of the barycenter. In Fig. 10, it shows the movement of the barycenter in a series of images on the X-axis and Y-axis.

In Fig. 10 (a), the moving hand posture “up” is performed for four times, while the same static hand posture is made for four times shown in Fig. 10 (b) as a comparison. It is clear that the hand posture movement causes a more obvious change of the barycenter. In this research, a sliding window method with 10 frames is applied to detect the biggest continuous change of the barycenter. In the previous work, a sliding window method with 10 frames was applied to detect the biggest continuous change of the barycenter. Therefore, the movement of hand posture and the moving direction can be detected at the same time.

### 4. Experiment and Results

#### 4.1 Experiment Environment Settings

The typical in-vehicle experiment environment can be seen in Fig. 11. The common constraints of actual in-vehicle environment like car vibration, road conditions, car speed, and weather
Fig. 10 Movement of barycenter of moving hand posture and static hand posture.

Fig. 11 In-vehicle experiment condition.

are considered. More details about in-vehicle experiments are introduced in Section 4.4.

Since it is not so convenient to conduct in-vehicle experiments for many experimenters, in order to obtain more data, the simulated indoor environment is set. The indoor experimental environment can be seen in Fig. 12 (a). An experimenter sits in a specific seat, and the sensor is placed on a tripod. The heights of chair and tripod are fixed as 40 cm and 70 cm. This installation of the experiment environment aims at simulating a real driving seat. On the top side, as Fig. 12 (b) shows, the sensor is placed 40 cm away from the midline of the experimenter’s body. In this distance, the experimenter raises his hand spontaneously. In order to reduce the individual difference, the distance between the hand and the sensor is set from 20 cm to 40 cm. The sensor is installed at a fixed angle of 30° from the horizontal, as shown in Figs. 13 (a) and (b). The reason why the angle is applied is that when people raise their hand spontaneously, their hand tends to incline forward. This installation of the sensor can improve the accuracy of the system and make users easy to use.

In the static hand posture recognition experiment, an experimenter repeats the following steps: raising his hand beside the sensor to make a specified hand posture without looking at the sensor, moving the hand towards the sensor and keeping the
posture for two seconds, finally withdrawing his arm. During this process, all the frames of data are recorded. This process is repeated for 10 times. After the image pre-processing and the static posture detection, 150 to 200 static posture images are obtained for each posture from each experimenter. In this experiment, 20 experimenters’ data in the indoor environment and five experimenters’ data in the in-vehicle environment are collected.

In the moving posture direction detection experiment, an experimenter moves his hand from the center area of the sensor to the boundary area of the sensor for one second. In this experiment, the hand moves through the center of the sensor’s detection area within an angle range considering the actual in-vehicle environment. During this process, all the frames of data are recorded. This process is repeated for 10 times in four directions. In this experiment, 20 experimenters’ data in the indoor environment and five experimenters’ data in the in-vehicle environment are also obtained.

4.2 Algorithm Parameter Settings

In the image pre-processing step, the GMM model is used. In order to decide the number of Gaussian distributions in the thermal images, the data distributions of 1000 images from five experimenters’ data are analyzed. The results of the distribution number are shown in Table 1.

According to the results shown in Table 1, in this paper, the number of the Gaussian distributions is set to three since it appears most frequently in the image analyses.

After that, the temperature threshold is calculated by Eq. (4). The distribution of temperature thresholds in 1000 images is shown in Fig. 14.

According to the result of Fig. 14, it can be concluded that approximately 80% of temperature threshold values are concentrated at 26°C to 27°C, which is the environment temperature. Therefore, with the method of threshold calculation, the background can be correctly removed from the object images in this environment.

4.3 Result Comparisons of Hand Posture Recognition with and without Static Posture Detection between Different Classifiers in the Indoor Environment

In order to confirm the effectiveness of the proposed method, the results of applying the static posture detection between the CNN and the RF are shown in the tables below.

In these analyses, 100 images are randomly selected for each posture from each experimenter. Therefore, 1100 images are obtained from one experimenter and totally 22000 images are used. In order to verify the capability of this system to analyze individual difference, cross validations are conducted. The data of 19 experimenters is used for training and one experimenters’ data is used for test. This procedure is repeated for 20 times with different experimenter’s data for test. The classification results with and without static posture detection between the CNN and the RF are shown in Tables 2 to 5.

The results from Tables 2 to 4 can draw the conclusion that whether the static posture detection is applied or not, the performance of the CNN is much better than that of the RF. From the results shown in Table 4 and Table 5, both the CNN and the RF perform better when the static posture detection is applied, which means this method is effective indeed. The reason is that in the static posture detection, those images in which postures are in the boundary area or blurred are categorized into the non-

| Table 1 The number of distributions in the thermal images. |
|-----------------------------|-----------------------------|-----------------------------|
| Number of distributions    | Number of images            | Percentage                  |
| 2                           | 74                          | 7.4%                        |
| 3                           | 834                         | 83.4%                       |
| 4                           | 92                          | 9.2%                        |
| Total                       | 1000                        | 100.0%                      |

| Fig. 14 Distribution of the temperature thresholds. |
Results of hand posture recognition with the static posture detection by the RF.

| Real | p0 | p1 | p2 | p3 | p4 | p5 | p6 | p7 | p8 | p9 | p10 | Accuracy |
|------|----|----|----|----|----|----|----|----|----|----|-----|-----------|
| p0   | 1968 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 98.4% |
| p1   | 38 | 175 | 84 | 0 | 105 | 0 | 20 | 0 | 2 | 0 | 0 | 87.6% |
| p2   | 25 | 0 | 1688 | 86 | 30 | 55 | 0 | 60 | 0 | 0 | 0 | 84.0% |
| p3   | 3 | 0 | 52 | 1826 | 0 | 36 | 0 | 33 | 18 | 0 | 32 | 91.3% |
| p4   | 82 | 82 | 70 | 0 | 590 | 0 | 49 | 109 | 0 | 6 | 29 | 79.7% |
| p5   | 102 | 118 | 0 | 72 | 54 | 1642 | 0 | 0 | 36 | 0 | 0 | 82.1% |
| p6   | 0 | 0 | 44 | 0 | 0 | 74 | 1802 | 72 | 0 | 0 | 8 | 90.1% |
| p7   | 8 | 0 | 108 | 0 | 50 | 66 | 1688 | 0 | 0 | 80 | 0 | 84.4% |
| p8   | 28 | 0 | 120 | 0 | 30 | 52 | 96 | 0 | 1580 | 104 | 0 | 79.0% |
| p9   | 42 | 0 | 0 | 0 | 41 | 0 | 0 | 16 | 112 | 1745 | 46 | 87.2% |
| p10  | 26 | 0 | 41 | 0 | 0 | 74 | 0 | 38 | 58 | 60 | 1703 | 85.2% |

Results of the hand postures with boundary detection in the in-vehicle environment by the CNN on a cloudy day.

| Real | p0 | p1 | p2 | p3 | p4 | p5 | p6 | p7 | p8 | p9 | p10 | Accuracy |
|------|----|----|----|----|----|----|----|----|----|----|-----|-----------|
| p0   | 496 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 99.2% |
| p1   | 0 | 477 | 0 | 0 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 95.4% |
| p2   | 10 | 0 | 469 | 0 | 13 | 0 | 0 | 0 | 8 | 0 | 0 | 93.8% |
| p3   | 3 | 0 | 486 | 9 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 97.2% |
| p4   | 0 | 0 | 35 | 465 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 93.0% |
| p5   | 18 | 0 | 0 | 14 | 0 | 431 | 0 | 0 | 16 | 21 | 0 | 86.2% |
| p6   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100.0% |
| p7   | 29 | 0 | 0 | 0 | 0 | 0 | 0 | 468 | 0 | 0 | 3 | 93.6% |
| p8   | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 461 | 17 | 10 | 92.2% |
| p9   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 491 | 0 | 0 | 98.2% |
| p10  | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 0 | 8 | 466 | 93.2% |

Results of the hand posture with the boundary detection in the in-vehicle environment by the CNN on a sunny day.

| Real | p0 | p1 | p2 | p3 | p4 | p5 | p6 | p7 | p8 | p9 | p10 | Accuracy |
|------|----|----|----|----|----|----|----|----|----|----|-----|-----------|
| p0   | 500 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100.0% |
| p1   | 0 | 472 | 5 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 94.4% |
| p2   | 8 | 0 | 462 | 5 | 0 | 0 | 13 | 0 | 0 | 0 | 12 | 92.4% |
| p3   | 28 | 0 | 0 | 0 | 0 | 0 | 0 | 446 | 26 | 0 | 0 | 89.2% |
| p4   | 7 | 0 | 16 | 0 | 453 | 9 | 0 | 0 | 0 | 0 | 0 | 95.0% |
| p5   | 8 | 0 | 11 | 0 | 6 | 475 | 0 | 0 | 0 | 0 | 0 | 95.0% |
| p6   | 0 | 0 | 0 | 0 | 0 | 0 | 500 | 0 | 0 | 0 | 0 | 100.0% |
| p7   | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 471 | 5 | 16 | 94.2% |
| p8   | 0 | 0 | 0 | 0 | 0 | 0 | 32 | 8 | 446 | 16 | 0 | 88.8% |
| p9   | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 460 | 5 | 92.0% |
| p10  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100.0% |

Results of moving posture detection experiment.

In the moving posture detection experiments, different detect angle ranges are set to compare the results of moving posture in each environment, and the averages of the pixel number in each posture image are calculated. The results are shown in Table 7.

These results indicate that there is no obvious difference between the posture images obtained from the in-vehicle environment and the simulated indoor environment. The reason why the posture images obtained from different environments are close is that the resolution of this sensor is relatively low (32×24). Therefore, the vibration caused by the road condition and the engine of the car hardly ever influences the quality of the posture images obtained from the sensor. Even if there are some blurred images caused by the car vibration, these images will be categorized into no-posture class in the static posture detection step. Therefore, it can be concluded that the system can adapt to the car vibration and work properly.

4.4.2 Influence of environment sunlight

The working mechanism of an infrared array sensor is constructing the thermal images by measuring infrared radiation. However, the composition of sunlight contains infrared components. Therefore, it is necessary to detect the influence of sunlight in the in-vehicle environment.

The experiments are conducted in two weather situations: a bright sunny day and a cloudy day. Five experimenters perform 10 target postures each time, and this procedure is repeated for 10 times in both situations. After that, 100 image frames of each posture in both situations are randomly selected to construct the test dataset, while the training dataset is composed of 20 different experimenters’ image data from the simulated indoor experiments. The results are shown in Table 8.

According to the classification result shown in Table 6 and Table 8, the average accuracy is 94.7% on the cloudy day and 94.2% on the bright sunny day in the in-vehicle experiments. The results show that the sunlight has little influence on the performance of the sensor in the actual in-vehicle environment. Therefore, this system can adapt to different sunlight situations.

4.5 Result of Moving Posture Detection Experiment

In the moving posture detection experiments, different detect angle ranges are set to compare the results of moving posture in each environment, and the averages of the pixel number in each posture image are calculated. The results are shown in Table 7.
Table 9 Results of the moving posture detection in the different angle ranges.

| Angle ranges | Correctly classified | Misclassified | No-movement |
|--------------|---------------------|---------------|-------------|
| ±15°         | 86.2%               | 2.3%          | 11.5%       |
| ±30°         | 92.9%               | 2.9%          | 4.2%        |
| ±45°         | 90.3%               | 9.7%          | 0.0%        |

Table 10 Results of the moving posture detection in the indoor experiment.

| Real     | up | down | left | right | none | Classification accuracy |
|----------|----|------|------|-------|------|-------------------------|
| up       | 1819 | 0 | 36 | 56 | 89 | 95.4% |
| down     | 0 | 1830 | 37 | 28 | 105 | 96.8% |
| left     | 19 | 24 | 1885 | 0 | 72 | 97.9% |
| right    | 20 | 16 | 0 | 1899 | 65 | 98.2% |

Table 11 Results of the moving posture detection in the in-vehicle experiment.

| Real     | up | down | left | right | none | Classification accuracy |
|----------|----|------|------|-------|------|-------------------------|
| up       | 452 | 0 | 18 | 9 | 21 | 94.6% |
| down     | 0 | 463 | 8 | 15 | 14 | 95.4% |
| left     | 6 | 14 | 474 | 0 | 6 | 96.0% |
| right    | 5 | 3 | 0 | 484 | 8 | 98.4% |

detection in different situations.

According to the results shown in Table 9, considering both the accuracy and the no-movement number, the detect angle range can be set to ±30° in this system.

The moving posture detection experiments are conducted in both the indoor environment with 8000 data from 20 experimenters and the in-vehicle environment with 2000 data from five experimenters. The results are shown in Table 10 and Table 11.

As shown in Tables 10 and 11, the classification accuracy is calculated from the sum of the correct classified movement number and the no-movement number divided by the total data number. Moreover, the average classification accuracy of each movement in the indoor experiment is 97.1%, and the average classification accuracy in the in-vehicle experiment is 96.1%. Therefore, it can be concluded that the moving posture detection method can adapt to the actual in-vehicle environment.

5. Conclusion and Future Work

In this paper, an in-vehicle device control system by hand posture recognition with movement detection using an infrared array sensor is proposed to improve the in-vehicle HMI operation and help drivers concentrate more on driving.

In this system, both 10 hand postures and their moving forms towards four directions are applied to be recognized. The thermal image data collected from the infrared sensor is processed by the GMM to separate the background images and object images in the image pre-processing step. After the static posture detection, the object images are divided into posture images and no-posture images. In addition, the class of non-target postures is added to avoid the influence of postures that are not included in the targeted 10 postures. Both non-target postures and the target postures dataset are classified by the CNN. After the static posture image is judged as one kind of target postures, moving posture detection is applied to detect the moving direction of the posture in order to improve the practicability of the system.

In the experiment results part, the hand posture recognition results with and without static posture detection between different classifiers in the indoor environment are compared. The comparison results confirm the effectiveness of the static posture detection, and the CNN is more suitable than the RF in this system. Then, the CNN model trained from indoor environment data is used to test the data from in-vehicle environment in the static hand posture recognition. Two experiments in the in-vehicle environment are conducted considering the interference of sunlight and car vibration. In the indoor experiments, the average accuracy of 20 experimenters is 95.3%. In the in-vehicle experiments, the average accuracy of five experimenters is 94.7%. Moreover, the results of the moving posture detection in both the indoor and the in-vehicle environments indicate the accuracy of 97.1% and 96.1%, respectively. Therefore, this hand posture recognition system is robust to be used for the in-vehicle device control with high accuracy.

However, some work still needs to be done in the future. One of them is that the influence of different seasons needs to be considered since the temperature of the environment may affect the reliability of the infrared sensor. What is more, more classifiers need to be compared to choose the better performing one.

References

[1] A.S. Knapper, M.P. Hagenzieker, and K.A. Brookhuis: Do in-car devices affect experienced users’ driving performance, Journal of International Association of Traffic and Safety Sciences, Vol. 39, No. 1, pp. 72–78, 2015.
[2] S. Anant and S. Veni: 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] S. Anant and S. Veni: Sensor-based hand gesture control system for safe driving, Journal of Advanced Research in Dynamical & Control Systems, Vol. 10, No. 3, pp. 690–698, 2018.
[4] P. Molchanov, S. Gupta, K. Kim, and K. Pulfi: Multi-sensor system for driver’s hand-gesture recognition, Proc. IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) 2015, pp. 1–7, 2015.
[5] N. Sokol: The effects of noise-robustness of in-car voice-controlled systems on user perceptions and driving behavior, M.S. thesis, University of Toronto, 2017.
[6] J. Kerridge, A. Armitage, D. Binnie, L. Lei, and N. Sumpter: Using low-cost infrared detectors to monitor movement of pedestrians: Initial findings, Journal of the Transportation Research Record, Vol. 1878, No. 1, pp. 11–18, 2004.
[7] H. Mohammadmoradi, S. Munir, O. GnaWal, and C. Shelton: Measuring people-flow through doorways using easy-to-install IR array sensors, Proc. International Conference on Distributed Computing in Sensor Systems (DCOSS) 2017, pp. 35–43, 2017.
[8] X. Fan, H. Zhang, C. Leung, and Z. Shen: Fall detection with unobtrusive infrared array sensors, Proc. Multisensor Fusion and Integration in the Wake of Big Data, Deep Learning and Cyber Physical System, pp. 253–267, 2017.
[9] Y. Kobiyama, Q. Zhao, R. Ota, and S. Ichimura: Recognition of frequently appeared locations/activities based on infrared sensor array, Proc. IEEE International Conference on Cybernetics (CybCONF) 2017, pp. 1–5, 2017.
[10] S. Chiu, J. Hsieh, C. Hsu, and C. Chiu: A convolutional neural networks approach with infrared array sensor for bed-exit detection, Proc. International Conference on System Science and
Engineering (ICSSE) 2018, pp. 1–6, 2018.

[11] F. Parada-Loira, E. González-Agulla, and J.L. Alba-Castro: Hand gestures to control infotainment equipment in cars, Proc. IEEE Intelligent Vehicles Symposium (IV) 2014, pp. 1–6, 2014.

[12] S. Tateno, Y. Zhu, and F. Meng: Hand gesture recognition system for in-car device control based on infrared array sensor, Proc. SICE Annual Conference 2019, pp. 701–706, 2019.

[13] S. Pleshkova and Z. Zahariev: Development of system model for audio visual control of mobile robots with voice and gesture commands, Proc. 40th International Spring Seminar on Electronics Technology (ISSE) 2017, pp. 1–4, 2017.

[14] H. Lin, J. Chuang, and T. Liu: Regularized background adaptation: A novel learning rate control scheme for Gaussian mixture modeling, IEEE Transactions on Image Processing, Vol. 20, No. 3, pp. 822–836, 2011.

[15] R. Arandjelović, P. Gronat, A. Torii, T. Pajdla, and J. Sivic: NetVLAD: CNN architecture for weakly supervised place recognition, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, pp. 5297–5307, 2016.

[16] T. Li, J. Leng, L. Kong, S. Guo, G. Bai, and K. Wang: DCNR: deep cube CNN with random forest for hyperspectral image classification, Journal of Multimedia Tools and Applications, Vol. 78, No. 3, pp. 3411–3433, 2019.

[17] A. Gautam and B. Raman: Segmentation of ischemic stroke lesion from 3D MR images using random forest, Journal of Multimedia Tools and Applications, Vol. 78, No. 6, pp. 6559–6579, 2019.

[18] O. Sangjun, R. Mallipeddi, and M. Lee: Real time hand gesture recognition using random forest and linear discriminant analysis, Proc. 3rd International Conference on Human-Agent Interaction 2015, pp. 279–282, 2015.

[19] S. Liao, A.K. Jain, and S.Z. Li: A fast and accurate unconstrained face detector, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 38, No. 2, pp. 211–223, 2016.

Shigeyuki Tateno (Member)

He received his B.S., M.S., and Ph.D. degrees from Kyushu University, Japan, in 1990, 1992, and 2001, respectively. In 2003, he joined Waseda University, Japan, where he is currently an Associate Professor of Graduate School of Information, Production and Systems. His research interests include wireless network systems.

Yiwei Zhu

He received his B.Eng. degree from Xidian University, China, in 2017. In the same year, he joined Waseda University, where he is a student of Graduate School of Information, Production and Systems. His research interests include human machine interface systems.

Fanxing Meng

He received his B.Eng. degree from Beijing Institute of Technology, China, in 2018. In the same year, he joined Waseda University, where he is currently a student of Graduate School of Information, Production and Systems. His research interests include human machine interface systems.