Driver Status Monitoring System in Autonomous Driving Era

Driver Status Estimation with Time-series Deep Learning

HYUGA Tadashi, KINOSHITA Koichi, NISHIYUKI Kenta and HASEGAWA Yuki

We propose a novel driver's status estimation system based on driver's behavior in autonomous driving scene. In future years, drivers are expected to be responsible for driving and monitoring the surrounding environment even in autonomous driving. To monitor the driver status is very important to ensure that he/she stays proper condition during driving. We define new criteria to measure the adequateness level of the driver during autonomous driving, and realize real-time driver status monitoring based on the criteria. It will contribute to decrease the risk of traffic accidents.

To measure driver's status precisely, we utilize two features: 1st one is driver's facial information, i.e. face direction, gaze direction and eye open-close level, and 2nd one is driver's behavior taken by image sequence. These features are concatenated and fed into a pre-trained recurrent neural network. By using the state-of-the-art neural network technology, the driver's status can be estimated with high accuracy.

1. Introduction

While the movement toward the practical realization of autonomous driving is accelerating, autonomous driving in particular environments such as on expressways is expected to be realized by around 2020[1]. However, autonomous driving is still far from complete realization, and thus it is said that "level 2 autonomous driving," which realizes partial autonomous driving under the responsibility of the driver will become mainstream for the time being among the autonomous driving levels shown in Table 1, and drivers need to monitor whether the driving behavior is appropriate during autonomous driving. According to the results of a survey, this phase will continue for the time being[2].

Thus, we developed driver status monitoring technology for estimating whether a driver is concentrating on driving, focusing attention on the status of the driver during autonomous driving. In this paper, we define three new criteria based on actual driving behavior as the criteria for evaluating the driver's concentration level, and also propose technology for sensing the driver's concentration level that sequentially outputs the results of evaluating the three criteria by inputting a time series of images of the driver.

Contact : *HYUGA Tadashi* tadashi.hyuga@omron.com

Table 1 Autonomous driving levels advocated by SAE in the United States

Level 0	No automation
Level 1	Driver assistance (ADAS)
Level 2	Partial driving automation
Level 3	Conditional driving automation
Level 4	High driving automation
Level 5	Full driving automation

2. Background

2.1 Social trend of autonomous driving

The revision of laws toward autonomous driving is still being debated. In the discussion at the United Nations toward the revision of the international standards for automatic steering (commonly known as "R79"), autonomous vehicles are expected to be institutionalized to satisfy the following requirements[3]:

- A warning is issued to the driver at least four seconds before the system reaches its functional limit
- The driver is monitored constantly to check whether he/she is concentrating on driving, and a warning is issued in the event of a driver becoming drowsy
- If the driver fails to respond to the warning, control for minimizing the danger is performed automatically

From the above, it can be said that monitoring driver status is a function which will be necessary in the future, and in particular, an autonomous driving system needs to check whether the driver can respond to the warning properly.

2.2 Conventional driver status monitoring system

Many conventional driver status monitoring systems detect whether a driver is driving normally while operating a vehicle manually basically using only a single criterion. For example, driver drowsiness detection focusing on the opening/closing of the eyelids and inattentive driving detection focusing on the direction of the face are being put into practical use[4]. However, these technologies can only detect whether a driver is watching ahead carefully during manual control, and it is difficult to detect various behaviors which may be performed during autonomous driving. In addition, a technology for detecting drowsiness by measuring the pulse wave with a device worn in a driver's ear has recently been developed[5] and it has enabled the determination of whether the driver is sleepy, but it is still difficult to respond to various dangers which could possibly occur at the time of awakening. At the same time, wearing a device could strain the psyche of a driver during driving, and thus a contactless sensor is preferred.

There is also a technology which provides labels to human behaviors to recognize which behavior corresponds to which label based on a time series of images[6] [7], and applying this technology to driver status monitoring is under consideration. However, since this method only recognizes predetermined behaviors, a deterioration in identification performance is assumed if various behaviors which could be performed during autonomous driving are covered. Thus, it is considered difficult to apply conventional methods to autonomous driving because they cannot respond to the diversified behaviors of drivers.

2.3 Proposal of monitoring concentration level utilizing Deep Learning

The authors[8] proposed a driver status monitoring technology which determines the possibility of the driver's return to driving when switching the mode from autonomous driving to manual driving. This technology determines how long it would take for a driver to return to driving when switching the driving mode and transmits the information to the vehicle's control system so as to switch the driving mode smoothly depending on the time required. However, although this technology can cope with assumed mode switches, the responses to emergency situations which could occur owing to incomplete awareness of the environment surrounding the autonomous vehicle are not assumed. Therefore, for the technology we propose in this paper, we targeted realizing new criteria for both emergency responses and assumed switches.

3. The criteria for evaluating driver's status in autonomous driving

From the above perspective, we define three new criteria for sensing the concentration level on driving during autonomous driving as follows: Eyes-on/off, Readiness-high/mid/low and Seated-on/off. These criteria are closely related to actual driving behaviors such as "cognition," "judgment" and "operation." Fig. 1 shows their relationships.

3.1 Eyes-on/off

This criterion is for checking whether a driver is monitoring the travel motion constantly. The state where a driver is checking the traveling direction and the state where a driver is performing a brief checking operation required when driving a vehicle such as checking an instrument or looking at a mirror are defined as "Eyes-on," and other driver behaviors such as the state where a driver is looking at a smartphone, book, or vehicle navigation system or keeps his/her eyes closed are defined as "Eyes-off."

3.2 Readiness-high/mid/low

The driver's readiness for driving is output in three stages. The state where a driver is wakeful and is not performing any behavior which is unrelated to driving is defined as "Readiness-high," the state where a driver is performing a behavior which is unrelated to driving but can return to driving through a simple procedure after receiving a warning from the system is defined as "Readiness-mid," and the state where a driver has difficulty in driving such as when he/she is sleeping is defined as "Readiness-low."

3.3 Seated-on/off

This criterion is for determining whether a driver can perform a driving act, based on whether he/she is seated in the driver's seat. The state where a driver is seated in the driver's seat is defined as "Seated-on," and the state where a driver is not seated in the driver's seat is defined as "Seated-off." This criterion is defined because it is possible that the driver's awareness about monitoring will decline and he/she will fail to prepare for a driving act as autonomous driving becomes more advanced, although it is difficult to assume during manual control.

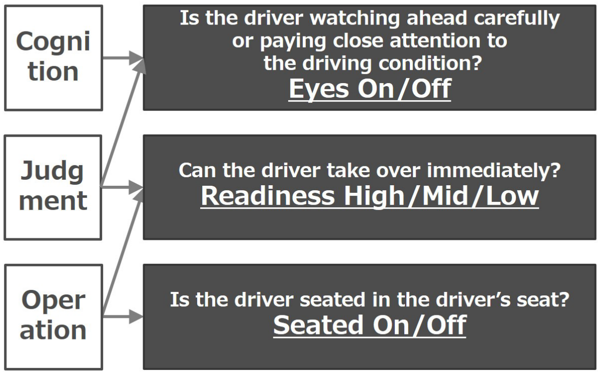
Cogni tion
Judg ment
Oper ation
Is the driver watching ahead carefully or paying close attention to the driving condition?
Eyes On/Off
Can the driver take over immediately?
Readiness High/Mid/Low
Is the driver seated in the driver's seat?
Seated On/Off

Fig. 1 The relationship among the three criteria for the concentration level on driving and driving acts

4. Sensing of the concentration level on driving during autonomous driving

This chapter describes the process flow of the proposed method for discriminating the three criteria (Fig. 2). The discriminator roughly consists of three phases. First, the face image sensing technology is applied to the image sequence input from the near-infrared camera to obtain regional face information. At the same time, the Convolutional Neural Network (hereinafter referred to as "CNN")9) is used to obtain the features corresponding to the driver's posture. Then, these outputs are integrated to recognize the transition of the driver status which varies from hour to hour, using the Long Short-Term Memory (hereinafter referred to as "LSTM")10) which is a recurrent neural network (hereinafter referred to as "RNN"). Although the network configuration combining CNN with RNN has been well-researched3), we combined face image sensing to achieve high precision and speed. The details are as follows:

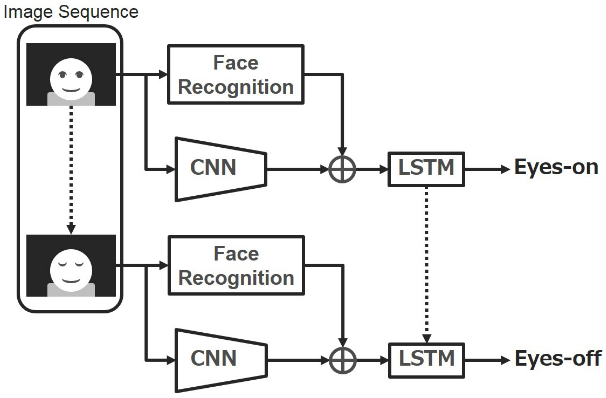

Fig. 2 The schematic view of the network for sensing the concentration level on driving

4.1 Image input using a near-infrared camera

We adopted a near-infrared camera in order to stably obtain images of a driver in a vehicle. General RGB cameras are usable depending on the sunshine conditions, but are not preferred because the influence of incident direct sunlight is significant in the daytime and a light is required for RGB, as well as owing to the fact that the face pattern obtained during the daytime changes during the nighttime. The camera we adopted this time can obtain stable face patterns in the daytime as well as at night with a near-infrared LED mounted on the camera unit, because the light illuminates the facial area constantly, even though it is invisible to the naked eye (Fig. 3).

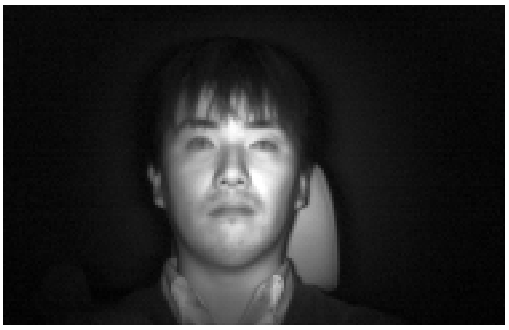

Fig. 3 An image of a person taken by a near-infrared camera (Permission to use the image was confirmed based on a letter of consent)

4.2 Face Recognition

Face image sensing refers to a function which detects a facial area from an image to output various information associated with the face. We are using technology called "OKAO Vision" developed by OMRON as a base, and utilizing the following functions for sensing the concentration level on driving:

- Face detection
- Facial landmark detection
- Facial landmark detection
- Eyelid opening/closing identification
- Gaze estimation

Although these respective technologies respond only to the image of a face taken by a conventional RGB camera, making them responsive to the images taken by a near-infrared camera enables the features of a driver to be obtained at high speed and with high precision.

4.3 CNN

Unlike conventional fully connected networks, CNN is a network with enhanced robustness against image deformation which has a structure with many layers, including the convolution layer where a small region filter and image obtained through learning are convoluted to perform computation, as well as the pooling layer, where the image obtained in the convolution layer is compressed according to a predetermined rule. Although CNN itself has long been practiced, it became the instigator of the current Deep Learning boom which updated the state of the art of various image recognition benchmark tests through the recent proposal of the learning method which improved generalizing performance.

4.4 LSTM

LSTM (Fig. 4) uses the time series data as input and regards the intermediate output of the previous frame in addition to the information of the current frame as input in order to obtain the results of recognizing the predetermined single frame. In addition, it also has an internal memory called a "cell" to compute the weight of the input of the current frame relative to the output based on this value. This weight has behaviors which are preset through prior learning and is known to enable longer memory than conventional RNN.

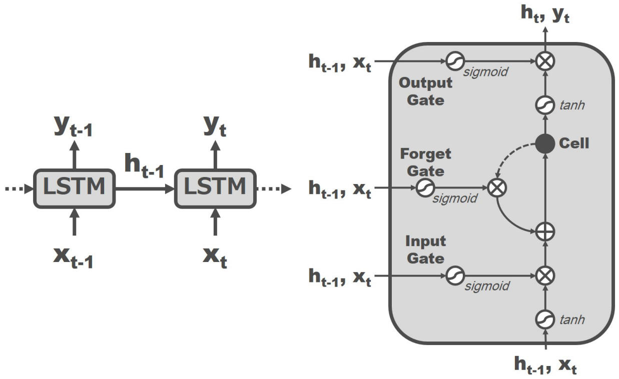

Fig. 4 The schematic view of LSTM

4.5 Construction of learning data set

For the network shown in Fig. 2, the respective parameters are determined using the time series Deep Learning. We identified the behaviors which a driver could perform during autonomous driving and then selected typical patterns from among them to utilize them for learning. Tables 2, 3 and 4 show examples of behaviors related to the respective criteria.

We used the sequence of images taken while 100 subjects were performing the respective behaviors according to our instructions. Among these images, the data of 50 subjects were used as the data for learning and the remaining data were used for evaluation. In the evaluation, only a single behavior was performed for one video, the recognition results of the respective frames used for recognition were evaluated, and the level of the possibility of returning to driving and the level of correct answers, which were decided by a majority vote, were compared. The image size, angle of view and frame rate of the camera were 720 × 480, approximately 10 degrees and 30 fps, respectively. We installed this camera in front of the driver's seat to record driver behaviors for a certain period of time and collect the data of a total of two million frames. For the learning and evaluation of the neural network, we used Chainer[11], which is a learning framework provided by PreferredNetworks.

Table 2 Examples of Eyes-on/off behaviors

On	Driving Watching ahead Leaning back against the window
Off	Driving inattentively Using a smartphone Dropping off to sleep

Table 3 Examples of Readiness-high/mid/low behaviors

High	Driving Watching ahead Checking an instrument temporarily
Mid	Eating and drinking Using a smartphone Talking on the phone
Low	Dropping off to sleep Putting the head down Feeling panicked

Table 4 Examples of Seated-on/off behaviors

On	Above driver behaviors
Off	None (Not seated in the driver's seat)

4.6 Evaluation

Almost the same quantity of data as the learning data was used as evaluation data. At that time, the accuracy rates of the respective criteria were as shown in Fig. 5.

Table 5 The accuracy rates of the respective criteria

Eyes	95.4%
Readiness	94.8%
Seated	99.0%

5. Studies toward practical realization

5.1 Construction of a large database covering the issues in the field

To build the above technologies, we constructed a database recording various driver behaviors required for learning and evaluation. Since it is currently difficult to obtain the data of realistic driver behaviors during autonomous driving, we constructed the database by taking images in various situations such as by (1) utilizing a driving simulator, (2) taking images at the passenger's seat while a vehicle was running, and (3) taking images of a driver during manual control.

The inappropriate driver behaviors and abnormal conditions shown in Tables 2 and 3 were recorded using a driving simulator, because it was dangerous to conduct them in a moving vehicle (Fig. 5). We instructed the subjects to conduct various actions while the view which could be seen during driving was displayed on the monitor in the front so as to create an environment which was close to the actual environment during driving. In addition, a passenger in the passenger seat of a moving vehicle performs behaviors which are close to driver behaviors during autonomous driving in terms of monitoring the driving conditions. However, since the passenger is not actually responsible for driving, it is highly possible that the transition of the level of the passenger's concentration on driving during autonomous driving is distant from that of the driver's. In the case of a driver during driving, it is possible to take images while the level of the driver's concentration on driving is highest because he/she is responsible for driving. However, there is a disadvantage that it is difficult to obtain the data when his/her concentration level is low. We judged that combining these data covers diversified driver conditions.

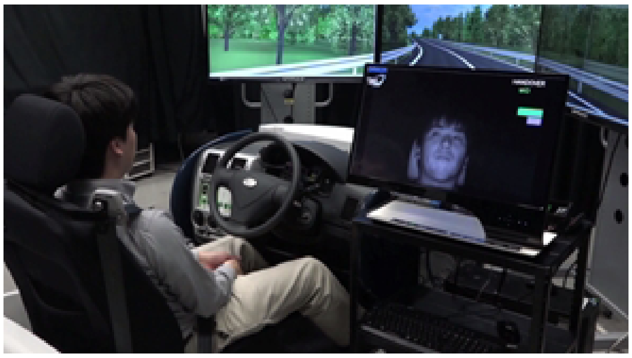

Fig. 5 Driving simulator (Permission to use the image was confirmed based on a letter of consent)

5.2 Enhancement of the environment resistance of face image sensing

In the face image sensing field, technology for estimating the position of a face in an image is called "face detection," which generally learns unique light and dark patterns shown on a face to output similar patterns shown in the image as the face. In this case, it is difficult to detect a face image with significantly changed patterns such as a side face and a face wearing a mask (hereinafter referred to as "masked face"). Thus, we extracted side face and masked face samples from the above database and combined them with existing face detection learning samples so as to enable the detection of side and masked faces with high precision, as is the case with frontal faces. In addition, identifying a face wearing a mask as a masked face enables the performance of subsequent processes such as landmark detection in response to a masked face. Fig. 6 and 7 show the processing result examples.

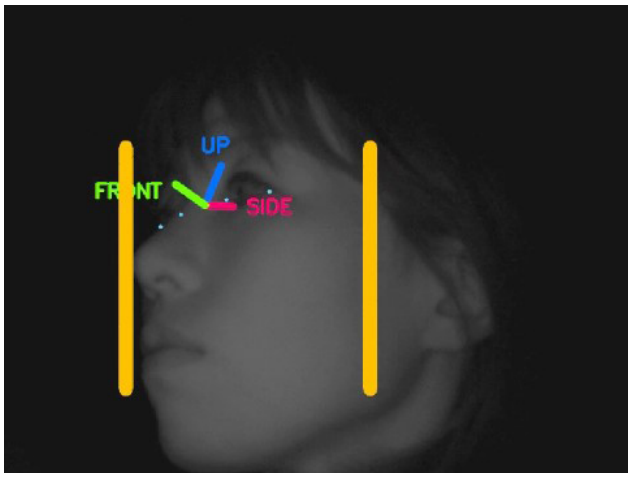

Fig. 6 Side face detection results (Permission to use the image was confirmed based on a letter of consent)

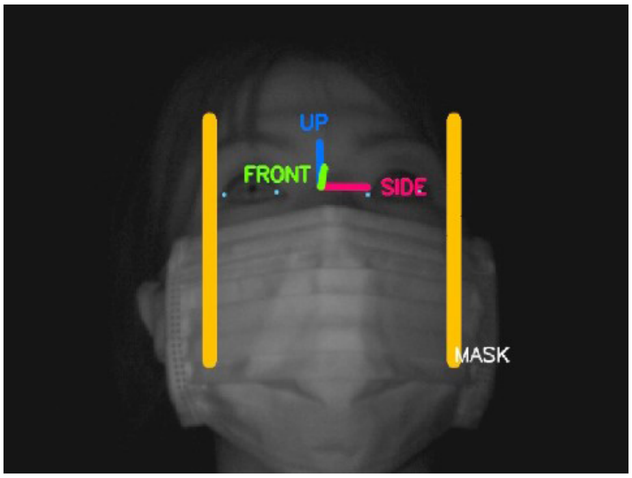

Fig. 7 Masked face detection results (Permission to use the image was confirmed based on a letter of consent)

5.3 The efficiency of output by sharing of a network

When thinking about practical realization, the computation cost and memory consumption need to be considered. To construct optimized neural networks separately for outputting the three proposed criteria, it simply takes three times as much as it takes to process a single network. Therefore, we introduced a system under which up to intermediate layers of the network are communalized to output three outputs in parallel at the final output stage (Fig. 8). The problem in this case is that the label needs to be revised because three outputs are learned simultaneously and a difficulty arises in the computation of correct answer information and loss. Specifically, a rule that "The computation of loss is not performed for Eyes and Readiness in the event of an image with no one seated (Seated-off)" needs to be added.

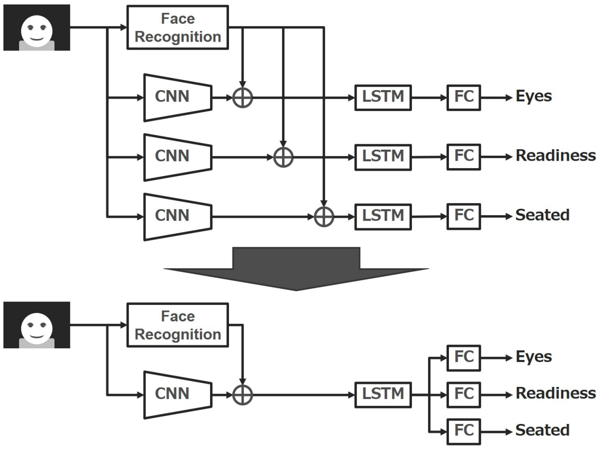

Fig. 8 Network communalizing

As a result of conducting learning based on the correct answer information in which the label was revised, there was no significant performance change before and after. Therefore, we adopted the communalized model.

5.4 Study on speeding up

When establishing the proposed method in a system with insufficient computational resources, installing a large-scale Neural Network creates difficulty in terms of both computation cost and memory consumption. In general, compressing the image to be input reduces the quantity of the required elements in the network significantly to solve the above issue. However, simply compressing an image causes a loss of information on the driver's face in the image. Since technologies such as drowsiness detection and inattentive driving detection were proposed, it means that sensing a face enables the information which is required to estimate driver's status to be obtained.

Thus, decreasing only the resolution of movement features and using the image resolution for face image sensing which is the same resolution for the original image enables the highly precise estimation even for low-resolution images by using high-level information about the driver which can be estimated from the face. In addition, it is expected that decreasing the resolution reduces the number of network parameters significantly as well as increases the processing speed. In the experiment conducted by the authors[5] where learning and evaluation were performed with low resolution, it was confirmed that a deterioration of only about 5 pt occurred even if the resolution was decreased from 720 × 480 to 24 × 18. Therefore, the method we propose in this paper also requires the resolution to be compressed to 24 × 18.

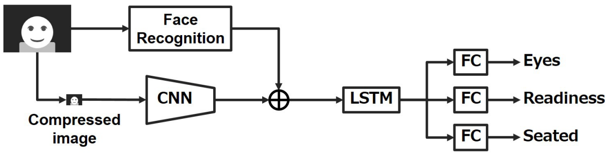

Fig. 9 The network configuration utilizing a compressed image

6. Future perspective

Our future perspective is to target deeply sensing driver status by combining an image taken by a camera, biological information and information on the surrounding traffic. Since image sensing technology enables contactless measurement, it is possible to recognize expressed phenomena such as the driver's expression and behavior without placing a burden on the driver, but it is difficult to measure internal conditions. For example, even if a driver is facing the front but thinking about something and insufficiently monitoring the vehicle-running condition, "Eyes-on" is output in this sensing of the concentration level on driving. In addition, appropriate driving behaviors which cannot be detected from images such as safety confirmation when turning right or left or when passing a vehicle are not covered.

We consider that utilizing biological information reduces such driving risks significantly. For example, the detection of the signs of drowsiness in combination with pulse measurement and facial expression estimation enables a warning to be given before a driver becomes sleepy to take over the driving in a safe manner. In addition, detecting a change in the driver's health to recommend that he or she take a rest adjusts the burden on the driver.

Furthermore, fusion with external perimeter monitoring sensors which have been actively developed in recent years enables the object that a driver is looking at to be recognized. This makes it possible to enhance the accuracy of the determination of whether the driving act is appropriate as well as to give feedback for safer driving such as the provision of traffic information to a driver while driving.

7. Conclusion

In this paper, aiming at the realization of safe and smooth autonomous driving, we proposed driver status monitoring technology for estimating whether a driver can be responsible for driving by using an image sequence as an input and focusing on the driver status during autonomous driving. In the proposed method, we could confirm that the combination of CNN with LSTM, as well as the use of not only images, but also facial image sensing results, enabled high-precision identification. In the future, we will develop deeper estimation of driver status by enhancing the performance further as well as adding biological information and information obtained from surround monitoring sensors, etc.

References

1) Cabinet Secretariat It Strategy Division. ITS concept and roadmap 2017 between public and private sectors. http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20170530/roadmap.pdf

2) Yano Research Institute. 2017 Automobile Technology Perspectives by 2030. 2017.

3) Ministry of Land, Infrastructure, Transport and Tourism. https://www8.cao.go.jp/cstp/gaiyo/sip/iinkai/jidousoukou_26/siryo26-5.pdf. 2016.

4) DENSO Corporation. "Driver Status Monitor". https://www.denso.com/global/en/products-and-services/information-and-safety/pick-up/dsm/

5) Fujitsu Ltd. "FEELythm". http://www.fmworld.net/biz/uware/ve31/

6) J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko and T. Darrell: "Long-term recurrent convolutional networks for visual recognition and description", CoRR, abs/1411.4389, (2014).

7) A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar and L. Fei-Fei: "Large-scale video classification with convolutional neural networks", CVPR (2014).

8) Hyuga, et al. "Estimation of the driver's readiness level with Time-series Deep Learning", Meeting on Image Recognition and Understanding (MIRU), 2017.

9) Y. Lecun, L. Bottou, Y. Bengio and P. Haffner: "Gradient-based learning applied to document recognition", Proceedings of the IEEE, 86, 11, 1998, p. 2278-2324.

10) S. Hochreiter and J. Schmidhuber: "Long short-term memory", Neural Comput., 9, 8, 1997, p. 1735-1780.

11) Preferred Networks. "Chainer: A exible framework of neural networks". http://chainer.org/

About the Authors

HYUGA Tadashi

Sensing Technology Research Center
Technology And Intellectual Property H.Q.
Specialty: Image processing and Pattern Recognition
Affiliated Academic Society: IEEE

KINOSHITA Koichi

Sensing Technology Research Center
Technology And Intellectual Property H.Q.
Specialty: Image processing
Affiliated Academic Society: IEICE, IPSJ

NISHIYUKI Kenta

Sensing Technology Research Center
Technology And Intellectual Property H.Q.
Specialty: Image processing and Pattern Recognition

HASEGAWA Yuki

Sensing Technology Research Center
Technology And Intellectual Property H.Q.
Specialty: Image processing
Affiliated Academic Society: JSPE

The names of products in the text may be trademarks of each company.