machine learning & knowledge extraction

MDPI

Article

Category Maps Describe Driving Episodes Recorded with Event Data Recorders †

Hirokazu Madokoro *, Kazuhito Sato and Nobuhiro Shimoi

Faculty of Systems Science and Technology, Akita Prefectural University, Yurihonjo City, Akita 015-0055, Japan; ksato@akita-pu.ac.jp (K.S.); shimoi@akita-pu.ac.jp (N.S.)

* Correspondence: madokoro@akita-pu.ac.jp; Tel.: +81-18-427-2180

† This paper is partially presented on the 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, Scotland, 25–29 August 2014 and 2017 IEEE International Joint Conference on Neural Networks, Anchorage, AK, USA, 14–19 May 2017.

Received: 29 January 2018; Accepted: 8 March 2018; Published: 12 March 2018

Abstract: This study was conducted to create driving episodes using machine-learning-based algorithms that address long-term memory (LTM) and topological mapping. This paper presents a novel episodic memory model for driving safety according to traffic scenes. The model incorporates three important features: adaptive resonance theory (ART), which learns time-series features incrementally while maintaining stability and plasticity; self-organizing maps (SOMs), which represent input data as a map with topological relations using self-mapping characteristics; and counter propagation networks (CPNs), which label category maps using input features and counter signals. Category maps represent driving episode information that includes driving contexts and facial expressions. The bursting states of respective maps produce LTM created on ART as episodic memory. For a preliminary experiment using a driving simulator (DS), we measure gazes and face orientations of drivers as their internal information to create driving episodes. Moreover, we measure cognitive distraction according to effects on facial features shown in reaction to simulated near-misses. Evaluation of the experimentally obtained results show the possibility of using recorded driving episodes with image datasets obtained using an event data recorder (EDR) with two cameras. Using category maps, we visualize driving features according to driving scenes on a public road and an expressway.

Keywords: episodic memory; context; facial expressions; category maps; event data recorder; unsupervised learning

1. Introduction

Drivers adjust their focus and their behavior according to traffic conditions to maintain safety. For example, drivers carefully devote attention to pedestrians or bicycles when they drive near a school or a park. On an expressway, drivers devote attention to surrounding cars running at high speed. Further, drivers will take extra care to avoid sleepiness when driving scenes do not often change. Therefore, prediction models for ensuring safety must adjust flexibly according to traffic changes, road conditions, environments, and situations. Advanced safety knowledge, danger prediction, and situational judgment are obtained not only from personal knowledge based on experiences and memory, but also on collective intelligence in terms of experience-based stories from their family and friends, news from TV, radios, and newspapers, and lessons learned at driving schools [1]. However, existing prediction models are hindered by limitations of event-based prediction using statistical information and probability models from sensor data and its histories.

Recently, automobile manufacturers, universities, research institutes, and Internet-related service companies have been investigating automatic driving cars that have autopilot-type assistance [2,3].

For wide-range and high-precision outside sensing, such cars use stereo cameras, millimeter-wave radar, and laser range finders for autopilot systems—in limited use on expressways. Moreover, real-time sensing and processing are actualized using originally customized processing devices. The performance of outside sensing is making steady progress. It is constantly advancing. For inside sensing, existing studies have targeted drowsy-driving detection [4], inattentive driving detection [5], cognitive distraction detection [6,7], and internal state estimation from facial expressions [8]. However, compared with outdoor sensing, numerous problems remain for human sensing in terms of effective measurements for visual differences among individuals, reproduction, and time-series changes of sensing targets [9]. Moreover, outside sensing and inside sensing are handled independently. No study results have been reported for both types of sensing together.

This study was conducted to create an episodic memory model in driving scenes using an event data recorder (EDR) with two cameras used simultaneously for inside and outside sensing. Several reports in the relevant literature [10,11] have described studies of driver gaze tracking and outside sensing for the detection of traffic lane deviation and driver carelessness. Nevertheless, no reports have described studies using an EDR: a simple device for sensing outside and inside environments together. Moreover, context recognition from scene images and facial expression recognition from face images have been studied individually in computer vision, human communication, and human–machine interfaces. Our report is the first to describe a trial of both images for creating episodic memory.

Studies of behavior predictions and intention understanding are extremely active in brain sciences. According to knowledge of brain sciences, one important role for human memory with accumulation and editing is feature prediction of various events [12], especially for episodic memory [13], which combines the context for target scenes and emotions, and which has high contributions [14]. Maeno proposed an episodic memory model used for a robotic mind and an emotional system based on his originally proposed passive consciousness hypothesis [15]. He presented a basic conceptual model based on a thought experiment.

The aim of this study is the creation of a safety prediction model based on episodic memory using artificial neural networks of long-term memory (LTM) and topological mapping. This paper presents a novel episodic memory model for driving safety according to traffic scenes using machine-learning-based algorithms of three types: adaptive resonance theory (ART) networks [16], which learn time-series features incrementally with the maintenance of stability and plasticity for time-series data, self-organizing maps (SOMs) [17], which represent input data as a map with topological relations using self-mapping characteristics, and counter propagation networks (CPNs) [18], which label category maps using input features and counter signals. Category maps represent driving episodes created from driving scenes and facial expressions using synchronized images obtained from outside and inside cameras on an EDR. The bursting states of respective maps produce LTM created on ART as episodic memory.

The remainder of this paper is organized as follows. Sections 2 and 3 present our proposed feature extraction and machine-learning-based methods. Section 4 addresses facial measurements in near-misses with cognitive distractions using a driving simulator (DS). As an evaluation experiment using an actual car, Section 5 addresses a prototype model of episodic memory using category maps. We analyze the relation between visualized category maps and actual near-misses using image datasets obtained using EDRs from three cars in summer and in winter. Finally, Section 6 concludes and highlights future work. We had proposed this basic method in a preceding publication [19]. For this paper, we have improved our method to review detailed procedures, especially in Section 4.

2. Proposed Method

2.1. The Procedure

Figure 1 depicts the procedures used for our proposed method. Using an EDR with two cameras, we obtained outside and inside time-series images simultaneously. Local features on high visual

saliency regions are extracted from images obtained from an outside camera. For images obtained from an inside camera, facial expression features are extracted using Gabor wavelet descriptors after facial region detection. Extracted features are combined with our original machine-learning-based methods for representation as category maps. The following are outlines in respective algorithms.

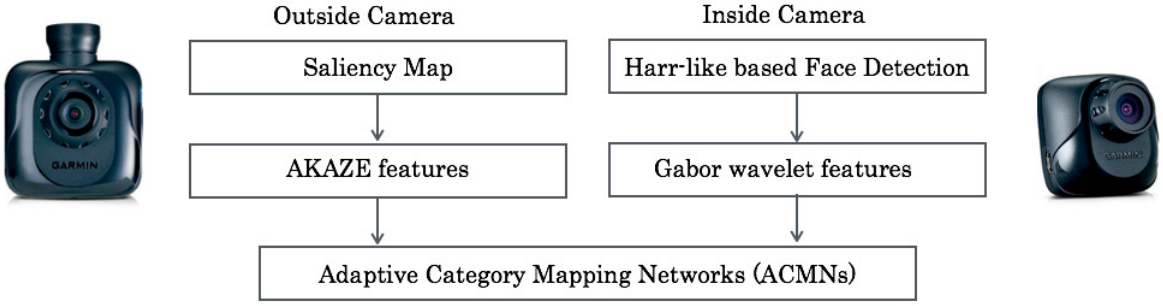

Figure 1. Procedures of our proposed method.

2.2. Saliency Maps

Visual information of various types is used for human recognition, understanding, and decisions. However, we do not use all information that we see momentarily. Humans have a mechanism to notice attentional objects or salient regions unconsciously. Itti et al. proposed saliency maps (SMs) [20] as a computational model for this attentional mechanism.

The brief procedure of SMs comprises four steps based on elemental computer vison algorithms [21–27]. The first step is low-level feature extraction using Gaussian filters after creating pyramid images in changed scales. The second step is to create component images of hue, brightness, and orientation. Subsequently, the third step is to create feature maps (FMs) that represent visual features of respective components after processing center–surround operations. The final step is to integrate SMs with a linear combination of FMs. We used high-saliency regions from SMs, although winner-take-all (WTA) competition is conducted for extracting high saliency coordinates.

2.3. AKAZE Descriptors

Gist [28] is a descriptor used to extract features from outdoor images, especially for global scenes in terms of mountains, lakes, and clouds in nature. Because of its rough granularity, it is unsuitable for describing features of driving scenes as roads and traffic signals for our target objects. As a part-based feature descriptor, scale-invariant feature transform (SIFT) [29] is used widely in computer vision studies, especially for generic object recognition. Using nonlinear scale spaces, KAZE, which is a Japanese word that means wind, descriptors [30] recorded higher performance than SIFT descriptors did. Recently, accelerated KAZE (AKAZE) descriptors [31] are specially examined not only for excellent descriptive capability but also for low real-time processing costs. The rapid processing is actualized with several fundamental technologies [32–35] used for actual applications.

2.4. Face Detection

We used the face detection algorithm proposed by Viola et al. [36] to detect a driver's face from images obtained using an inside camera. As a method to detect objects from images, numerous face detection methods have been proposed [37]. The method proposed by Viola et al. [36] was an epoch-making method for real-time video image processing using a generally available computer. Now the method is the de facto standard for face detection. Numerous improvements have been applied for the method in terms of robustness for occluded images and non-frontal faces. The application range has been expanded to tablet computers and smartphones for real-time processing with low electric power consumption [38]. The mechanism of rapid processing is based on a simple pattern features of Haar-like features [39] and cascade-connected weak classifiers [40] created by AdaBoost [41].

2.5. Gabor Wavelets

Visual information obtained by retinas are propagated to Visual Area 1 (V1) via the lateral geniculate nucleus (LGN) [42]. V1 comprises visual cells of two types: simple cells and complex cells. Simple cells in LNG and V1 have receptive fields that compose a visual range in response to specific stimulation. Receptive fields respond to a specific figure size, length, orientation, color, and frequency. This feature is called selective response. Hubel and Wiesel found selective response for lines from their electric physiological experiment using an anesthetized cat [43]. Selective orientation, which is one selective response, is examined specifically because similar features were achieved by Gabor wavelet filters as an engineering model [44]. Gabor wavelet filters have been used for various studies and applications in image processing and computer vision fields for enhancing specific features controlled by internal parameters.

3. Adaptive Category Mapping Networks

Figure 2 portrays the network architecture of adaptive category mapping networks (ACMNs) [45] as a learning method for visualizing time-series features on a category map. ACMNs comprise three modules: a codebook module for vector quantization of input data, a labeling module for creating labels as candidates of categories, and a mapping module for visualizing spatial relations of categories on a category map. These modules comprise self-organizing maps (SOMs) [17], adaptive resonance theory (ART) networks [16], and counter propagation networks (CPNs) [18]. Herein, SOMs and ART are unsupervised neural networks and CPNs are supervised neural networks. ACMNs actualize both learning modes for an original mechanism to create labels as candidates of categories. The following presents detailed explanations of the respective algorithms after an overview in each module is presented.

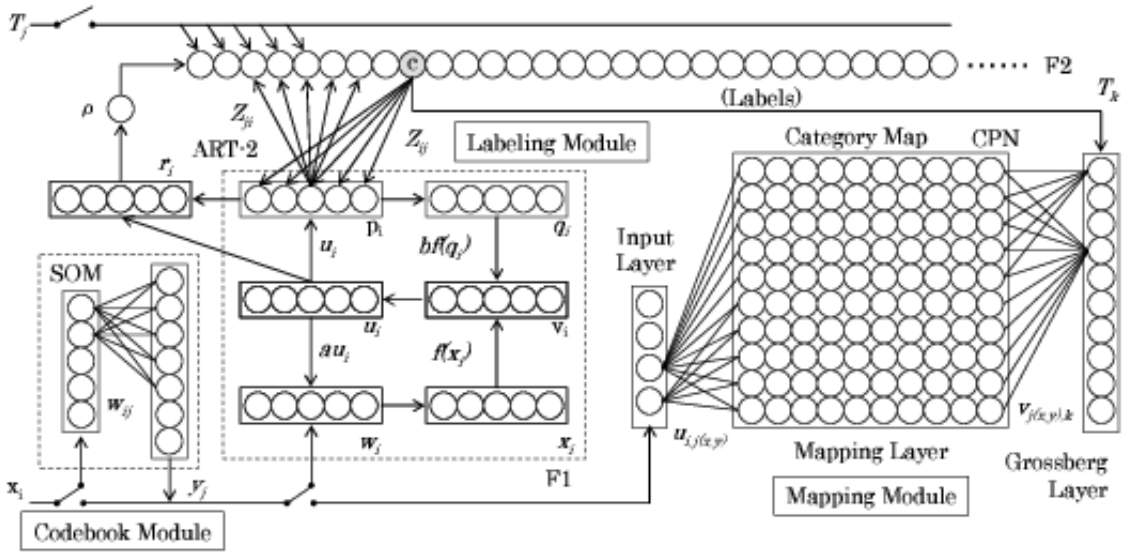

Figure 2. Overall architecture of adaptive category mapping networks (ACMNs), which comprise a codebook module, a labeling module, and a mapping module.

Input data are presented directly to the codebook module. This module is used if dimensions of input features differ among datasets. For example, dimensions are various according to the number of feature points on scale invariant feature transform (SIFT), which is used widely in generic visual object recognition as a part-based local feature. Using this module, input features are quantified to a specific dimension to represent distributions of histograms. Moreover, this module conducts vector quantization if the dimension of input features is high. For this process, data topology is preserved while

changing to a low-dimensional space. Herein, this module is not mandatory for use. This module can be passed if the dimensions of the input features are fixed for all datasets. We use this mechanism to reduce the load caused by the learning that ensues when codebooks are created and updated incrementally.

The labeling module creates candidates of categories from input features adaptively and incrementally. Based on the learning of ART, this module actualizes incremental learning while maintaining plasticity and stability. Input data are assigned to available categories if similar features are included. A new unit is assigned on F2 as a new category candidate if no similar feature is included. The labeling module actualizes incremental learning for this mechanism. For supervised or semi-supervised learning modes, teaching signals are assigned as labels for units created using this module. Unit indexes are used for candidate labels in the unsupervised learning mode.

The mapping module produces category maps with learning and mapping functions of CPNs using candidate labels of categories created from the labeling module. For this module, spatial relations among categories are visualized on category maps. Moreover, redundant labels including noise signals that occurred partially are removed using competitive learning in neighboring regions. The decision process is conducted using this module to bypass the labeling module when test datasets are presented. Herein, the module cannot learn incrementally, resembling the second layer of self-organizing incremental neural networks (SOINNs) [46]. The learning process occurs when a new dataset is presented for this module. However, this process uses training data obtained using not only candidate labels created from the labeling module but also labels in this module. This is a point of difference for standard relearning. For this mechanism, ACMNs store no training datasets for relearning. Rapid relearning is actualized using the minimum number of datasets.

3.1. Codebook Modules

For creating codebooks [47], *k*-means [48] is widely used. However, Vesanto et al. demonstrated that the clustering performance of SOMs is higher than that of *k*-means as a classic clustering method [49]. Moreover, Terashima et al. showed quantitatively that false recognition accuracy is lower when using SOMs for clustering than when using *k*-means [50]. Therefore, SOMs are used for creating the codebooks that are utilized for this module.

As a mechanism of neighborhood and competitive learning for self-mapping characteristics based on unsupervised learning, SOMs create clusters with similar input features. The SOMs network architecture comprises two layers: the input layer and the mapping layer. For the input layer, a similar number of units is assigned to the number of dimensions of input features. The mapping layer comprises units that are assigned in a low dimension. For creating codebooks, we assigned units on the mapping layer to one dimension because vector quantization is used for clustering. Learning is conducted to burst a unit on the mapping layer for input data.

The learning algorithm of SOMs is as follows. $x_i(t)$ and $w_{i,j}(t)$ respectively denote input data and weights from an input layer unit i to a mapping layer unit j at time t. Herein, I and J respectively denote the total numbers of the input layer and the mapping layer. $w_{i,j}(t)$ is initialized randomly before learning. The unit for which the Euclidean distance between $x_i(t)$ and $w_{i,j}(t)$ is the smallest is sought as the winner unit of its index c as

$$c = \underset{1 \leq j \leq J}{\operatorname{argmin}} \sqrt{\sum_{i=1}^{I} (x_i(t) - w_{i,j}(t))^2}. \tag{1}$$

As a local region for updating weights, the neighborhood region $N_c(t)$ is defined as the center of the winner unit c as

$$N_c(t) = \lfloor \mu \cdot J \cdot \left(1 - \frac{t}{O}\right) + 0.5 \rfloor. \tag{2}$$

Therein, $\mu (0 < \mu < 1.0)$ is the initial size of $N_c(t)$; O is the maximum iteration for training. Coefficient 0.5 is appended as a floor function for rounding.

Subsequently, $w_{i,j}(t)$ of $N_c(t)$ is updated to close input feature patterns.

$$w_{i,j}(t+1) = w_{i,j}(t) + \alpha(t)(x_i(t) - w_{i,j}(t)). \tag{3}$$

Therein, $\alpha(t)$ is a learning coefficient that decreases along with the progress of learning. $\alpha(0)(0 < \alpha(0) < 1.0)$ is the initial value of $\alpha(t)$. $\alpha(t)$ is defined at time t as

$$\alpha(t) = \alpha(0) \cdot \left(1 - \frac{t}{O}\right). \tag{4}$$

In the initial stage, the learning speed is higher when this rate is high. In the final stage, the learning converges while the range decreases.

For this module, the input features of I dimension are quantized into the J dimension, which is a similar dimension to the number of units on the mapping layer. The module output $y_j(t)$ is calculated as

$$y_j(t) = \sqrt{\sum_{i=1}^{I}(x_i(t) \cdot w_{i,j}(t))^2}. \tag{5}$$

This module is connected to the labeling module at the training phase. For the testing phase, this module is switched to the mapping module. Moreover, this module is passed when input features are used without creating codebooks directly.

3.2. Labeling Module

The role of this module is to create labels used for category candidates. For this study, we created this module using ART, which is a theoretical model of unsupervised neural networks to create labels adaptively and incrementally with preservation of plasticity and stability together for time-series data.

In ART of various types [51], we use ART-2 [16], into which it enables input continuous values. The network of ART-2 comprises two fields: Field 1 (F1) for feature representation and Field 2 (F2) for category representation. Here, F1 comprises six sub-layers: p_i, q_i, u_i, v_i, w_i, and x_i. The sub-layers actualize short-term memory (STM), which enhances features of input data and removes noise for a filter. Here, F2 actualizes long-term memory (LTM) based on finer or coarser recognition categories. LTM is created in each unit assigned to independent labels. The j-th unit of F2 and the sub-layer p_i are connected. Top-down weights Z_{ji} and bottom-up weights Z_{ij} are included. The weights are initialized as

$$Z_{ji}(0) = 0 \tag{6}$$

$$Z_{ij}(0) = \frac{1}{(1-d)\sqrt{J}}. \tag{7}$$

Therein, J is the number of units of F2. Subsequently, input data x_i are presented to F1; the sublayers are propagated as

$$w_i(t) = x_i(t) + au_i(t-1) \tag{8}$$

$$h_i(t) = \frac{w_i(t)}{e + ||w||} \tag{9}$$

$$v_i(t) = f(h_i(t)) + bf(q_i(t-1)) \tag{10}$$

$$u_i(t) = \frac{v_i(t)}{e + ||v||} \tag{11}$$

$$q_i(t) = \frac{p_i(t)}{e + ||p||} \tag{12}$$

$$p_i(t) = \begin{cases} u_i(t) & \text{(inactive)} \\ u_i(t) + dZ_{Ji}(t) & \text{(active)} \end{cases} \tag{13}$$

$$f(x) = \begin{cases} 0 & \text{if} \quad 0 \leq x < \theta \\ x & \text{if} \quad x \geq \theta. \end{cases} \tag{14}$$

Therein, a and b respectively denote coefficients of feedback loops from u_i to w_i and from q_i to v_i. θ is a parameter to control a noise detection level in v_i. e is a coefficient to prevent zero from occurring in the denominator. Subsequently, the most active unit of its index c is searched as

$$c = \underset{1 \leq j \leq J}{\operatorname{argmax}} (\sum_j p_i(t) Z_{ij}(t)). \tag{15}$$

For c, weights are updated as

$$Z_{ci}(t+1) = Z_{ci}(t) + \alpha \cdot (p_i(t) - Z_{ci}(t)) \tag{16}$$

$$Z_{ic}(t+1) = Z_{ic}(t+1) + \alpha \cdot (p_i(t) - Z_{ic}(t)). \tag{17}$$

The vigilance threshold ρ is used to ascertain whether input data belong correctly to a category, as

$$\rho < e + ||r|| \tag{18}$$

$$r_i(t) = \frac{u_i(t) + s \cdot p_i(t)}{e + ||u|| + ||s \cdot p||} \tag{19}$$

where s is a coefficient for propagation from p_i tor_i, and d is a learning rate coefficient. Furthermore, $s \cdot \alpha / (1 - \alpha) \leq 1$ is the constraint between them. When (18) is false, the active unit is reset and is searched to a next active unit. Repeat until the range of change of F1 is sufficiently small if (18) is true. Herein, teaching signals are used for labels if ACMNs are used for supervised learning. The index c is stored as a label if ACMNs are used for unsupervised learning.

3.3. Mapping Module

For this module, category maps are created as a learning result. We built this module using CPNs, which are supervised neural networks, to classify patterns into particular categories with the functions of competitive and neighborhood learning.

The network architecture of CPNs comprises three layers: an input layer, a mapping layer, and a Grossberg layer. The input layer and mapping layer resemble those of SOMs in this module. Teaching signals are presented to the Grossberg layer. For our method, labels that are assigned for F2 on ART-2 of the labeling module are used for teaching signals. Our method actualizes automatic labeling to combine CPNs with ART.

The order of units on F2 is assigned as labels used for teaching signals in the supervised learning mode. For the semi-supervised learning mode, mixed labels that include teaching signals and those without teaching signals created from ART are mapped on the category map. For the unsupervised learning mode, labels obtained using ART are used for learning CPNs. The usage of labels differs in each learning mode. Using the intermediate representation as labels, this module performs similar learning behaviors in respective modes.

Learning results are represented as a category map on the mapping layer. Spatial relations among datasets based on similarity are visualized on a category map. ACMNs create it automatically without setting the number of categories. Moreover, redundant labels are removed through the process of competitive and neighborhood learning.

The learning algorithm of CPNs is as follows. Herein, for visualization characteristics of category maps, we set the mapping layer to a two-dimensional structure $X \times Y$ unit. We set one dimension of

the input and Grossberg layers, although they can take any structures. The numbers of units are I and K, respectively. $u_{i,j(x,y)}(t)$ is the weight from an input layer unit i to a mapping layer unit $j(x,y)$ at time t. $v_{j(x,y),k}(t)$ is the weight from a Grossberg layer unit k to a mapping layer unit $j(x,y)$ at time t. These weights are initialized randomly before learning. $x_i(t)$ represents training data to present to the input layer unit i at time t. The unit for which the Euclidean distance between $x_i(t)$ and $u_{i,j(x,y)}(t)$ is the smallest is sought as the winner unit. $c(x,y)$ is the index of the unit.

$$c(x,y) = \underset{(1,1)\leq j(x,y)\leq (X,Y)}{\operatorname{argmin}} \sqrt{\sum_{i=1}^{I}(x_i(t) - u_{i,j(x,y)}(t))^2}. \tag{20}$$

The neighborhood region $N_{(c_x,c_y)}(t)$ around $c(x,y)$ is defined as

$$N_{c(x,y)}(t) = \lfloor \mu \cdot (X,Y) \cdot \left(1 - \frac{t}{O}\right) + 0.5 \rfloor \tag{21}$$

where $\mu (0 < \mu < 1.0)$ is the initial size of the neighborhood region, and O is the maximum iteration for training. $u^i_{n,m}(t)$ of $N_{c(x,y)}(t)$ are updated to close input feature patterns using Kohonen's learning algorithm as

$$u_{i,j(x,y)}(t+1) = u_{i,j(x,y)}(t) + \alpha(t)(x_i(t) - u_{i,j(x,y)}(t)). \tag{22}$$

Subsequently, $v_{j(x,y),k}(t)$ of $N_{c(x,y)}(t)$ is updated to close teaching signal patterns using Grossberg's learning algorithm.

$$v_{j(x,y),k}(t+1) = v_{j(x,y),k}(t) + \beta(t)(T_k(t) - v_{j(x,y),k}(t)). \tag{23}$$

Therein, T_k are training signals obtained using ART-2. $\alpha(t)$ and $\beta(t)$ are learning coefficients that have decreasing values with the progress of learning. $\alpha(0)$ and $\beta(0)$ respectively denote the initial values of $\alpha(t)$ and $\beta(t)$. The learning coefficients are given as

$$\begin{bmatrix} \alpha(t) \\ \beta(t) \end{bmatrix} = \begin{bmatrix} \alpha(0) \\ \beta(0) \end{bmatrix} \cdot \left(1 - \frac{t}{O}\right). \tag{24}$$

In the initial stage, the learning is done rapidly when the efficiencies are high. In the final stage, the learning converges, although the efficiencies decrease. At the maximum number of $v_{j(x,y),k}(t)$ for the k-th Grossberg unit, category $L_k(t)$ is searched as

$$L_k(t) = \underset{(1,1)\leq j(x,y)\leq (X,Y)}{\operatorname{argmax}} v_{j(x,y),k}(t). \tag{25}$$

A category map is created after determining categories for all units. Test datasets are presented to the network that is created through learning. The mapping layer unit, which is the minimum Euclidean distance, since the test data and feature patterns are similar, is burst. Categories for these units are recognition results for CPNs.

4. Preliminary Experiment Using a Driving Simulator

4.1. Measurement Setup

The purpose of this preliminary experiment is to evaluate facial measurements for the creation of driving episodes with simulated near-misses. Figure 3 depicts our DS, which has three displays and six actuators used to move the driver's seat. The realistic feeling of this DS is higher than that of DSs with a single display and a fixed seat. We used an RGB-D camera (Xtion pro Live; ASUSTeK Computer Inc., Taipei, Taiwan) to sense the driver's face. Using this camera, depth information is obtained from

infrared dot patterns. Moreover, we used an eye tracking system (faceLAB; Seeing Machines, Canberra, Australia) for gaze motion measurements. Compared with results of an experiment using an actual car, we were able to take advanced measurements of facial features using a DS and the sensors above. Furthermore, we quantitatively evaluated drivers' biological information, especially in facial feature changes for a near-miss situation under a distracted state.

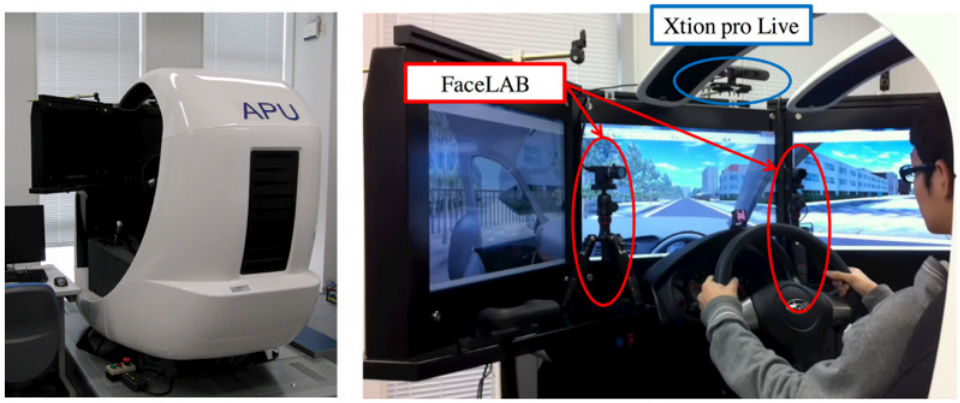

Figure 3. Driving simulator (DS) and measurement devices: (**a**) outside and (**b**) inside of the DS.

For this evaluation, we created near-miss scenarios of two cases. Both scenarios include a traffic scene at an intersection as a narrow perspective. For different patterns of a bicycle, we divide near-misses into two cases. Figure 4a depicts Case I: a bicycle runs in front of the car from the right to the left without brakes. Figure 4b depicts Case II: a bicycle runs suddenly from the left of the intersection to the right along with the car after turning in the direction in front of the car. This experimental setup, regarding details pertaining to driving scenarios, simulation environments, and near-misses, was based on our former study [52].

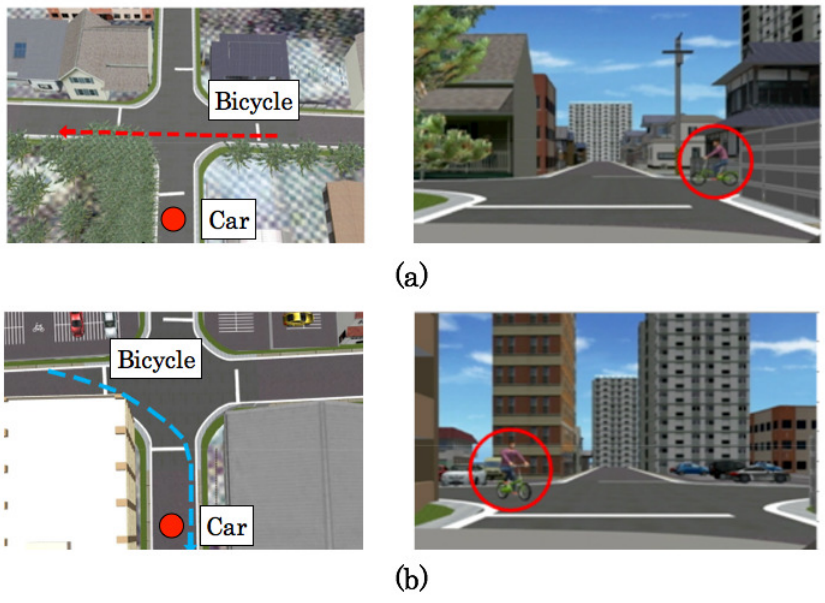

Figure 4. Simulated near-miss scenarios: (**a**) Case I and (**b**) Case II.

For creating simulated cognitive distractions, we used questions that comprise the multiplication of single-digit numbers based on studies by Suenaga et al. [53] and Abe et al. [54]. We provided it to all subjects in 3 s intervals as information delivered vocally from a speaker. All subjects answered them verbally. We recorded answers using a microphone to calculate the correct rate. As simulation parameters, we set the weather to fine in the daytime, which provides high visibility. The subjects consisted of two women, Subjects A and B, and 10 men, Subjects C–L. All subjects were university students who had been licensed drivers for up to four years. We selected a subject for evaluation using questionnaires of driving characteristics because this study was conducted to create an individual episodic model.

4.2. Driving Characteristics

For measuring driving characteristics in advance, we used two questionnaires: a driving style questionnaire (DSQ) [55] and a workload sensitivity questionnaire (WSQ) [56] by the Research Institute of Human Engineering for Quality Life. Quantitatively, the DSQ and WSQ respectively measure driving styles comprising driving attitudes, desires, and cognitive and driving burdens.

The first, the DSQ, comprises 18 questions classified into nine categories: (1) skill confidence, (2) negative factors, (3) hasty tendencies, (4) methodical tendencies, (5) preparation for a traffic light, (6) your car as a status symbol, (7) unsteadiness, (8) worry, and (9) false discovery. The respective questions are scored in four steps. The second, the WSQ, comprises 38 questions classified into 10 categories: (1) traffic condition posture, (2) road environmental posture, (3) hindrance of driving concentration, (4) decrease in body activity, (5) driving pace inhibition, (6) affliction, (7) driving path recognition and searching, (8) interior, (9) controls and operations, and (10) driving position. The respective questions are scored in five steps. Higher points scores are interpreted as showing high driving sensitivity for each measurement category.

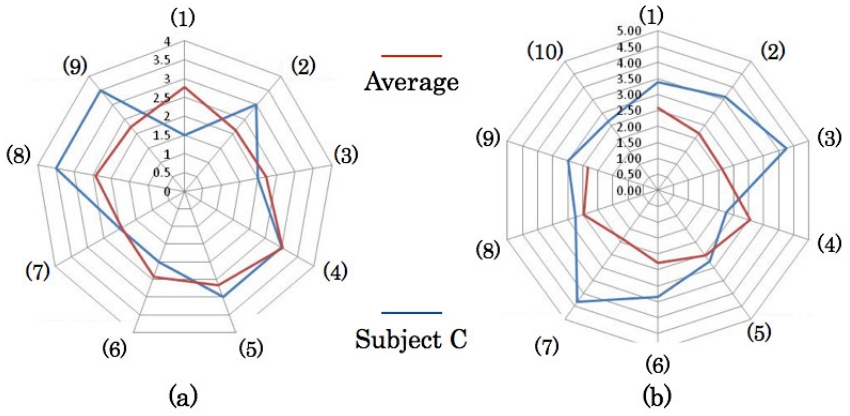

Figure 5. Subjective evaluation results: (**a**) The driving style questionnaire (DSQ) and (**b**) the workload sensitivity questionnaire (WSQ).

Figure 5 depicts measurement results obtained for the DSQ and the WSQ. Among all subjects, Subject C shows a salient tendency of worry and false discovery in the DSQ. Therefore, we analyzed Subject C further in detail.

4.3. Measurement Results of Gaze and Face Orientation

Figure 6 depicts distribution results of gaze and face orientations for Subject C when the subject encountered a near-miss. The measurement range is from entry into an intersection to the termination of a right turn after stopping in front of a traffic sign. Results of Cases I and II reveal that the distribution complexity of face orientations is higher than that of gazes.

Face orientations were varied because the driver moved his neck to check the non-visible intersection and the bicycle that ran from the right to the left in Case I. In contrast, face orientation changes to the right were slight in Case II because the bicycle that appeared suddenly from the left of the intersection passed through the same road as that used by cars. Face movements expanded upward and downward in Case II because the bicycle did not pass through the intersection, similar to Case I.

Using a DS, advanced and diverse information from drivers' faces can be measured without risk of a crash. However, it is still a challenging task to measure steady information from a driver on an actual car using a device such as FaceLAB. For experimentally obtained results, we obtained a tendency of diverse information of face orientations similar to that of gazes.

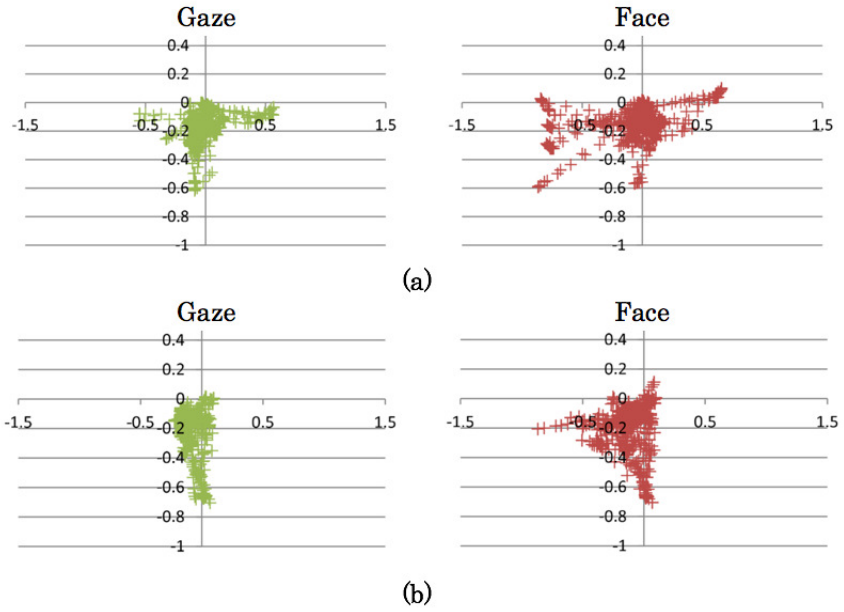

Figure 6. Distribution of gaze and face orientations of Subject C for near-misses of Subject C: (**a**) Case I and (**b**) Case II.

4.4. Relation between Near-Miss and Cognitive Distraction

Figure 7 depicts time-series changes of face directions of Subject C for the near-miss scenario in Case II under the status of normal driving with a cognitive distraction [57] using a calculation task. Vertical and horizontal axes respectively portray the normalized DS display sizes and time t. Positional values on the horizontal axes maintain values of less than zero because the subject watched the lower half of the display. The result in Figure 7a shows that the face movements were slight for the near-miss. In contrast, Figure 7b, which is a result under the calculation task, depicts wide and rapid movements of face orientations after a slight delay of responses. The subject moved their head lower unnaturally while turning left. However, it is difficult to observed responses repeatedly because most behaviors are mere single instances. For this study, we aimed to record these features as driving episodes because driving behaviors are actual measurement results.

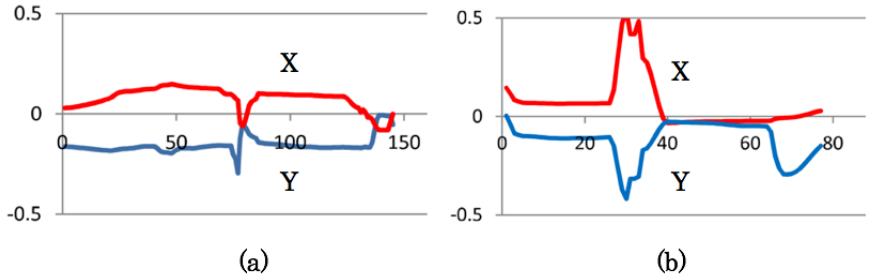

Figure 7. Time-series changes of face orientations of Subject C for a near-miss scenario in Case II: (**a**) normal driving and (**b**) cognitive distraction.

5. Evaluation Experiment Using an Event Data Recorder

5.1. Experimental Setup

EDRs are used widely not only for crash recording but also for the recording of driving scenes [58]. For this study, we used an EDR (GDR45DJ; Garmin Ltd., Schaffhausen, Switzerland) that comprises a front camera and a back camera connected by a USB cable. Synchronized time-series images are readily obtainable using EDR as a cost-effective system. Table 1 presents major specifications of the EDR. The 132° diagonal and 120° horizontal viewing range allows for an expansive capture of the driving scene.

Table 1. Major specifications of the event data recorder (EDR).

Size (front)	82 × 66 × 42 mm
Size (rear)	46 × 46 × 46 mm
Weight (front)	122 g
Weight (rear)	37 g
Imaging device	CMOS
Resolution	3 million pixel
Frame rate	30 fps
View angle	diagonal 132° (horizontal 120°)
Focal length	F2.0
Embedded sensors	Gyroscope, GPS, and TLS

Figure 8 depicts some obtained sample images. We installed a back camera on a dashboard in a car to take images of the driver's face, although its normal usage is rear-view monitoring. This installation provides captured face images from the lower side, as depicted in Figure 8b. The EDR has a function of saving a route using a global positioning system (GPS). One can check a route on an online map using an original tool provided by the manufacturer. This tool is used for a photographic navigation system using geotagging information that is included with each image.

We obtained image datasets from three cars. The obtaining period comprises two seasons: summer (July–August in 2015) and winter (January–February in 2016). The obtained area was in the Akita prefecture, which is an area of heavy snowfall in Japan. Therefore, our datasets include images obtained on snowy roads. Moreover, for extracting daily episodes from their daily car life, we obtained repeated data from similar routes for a commuter trip.

The video files of this EDR were saved automatically as short video clips with maximum sizes of 255 MB, which corresponds approximately to a five-minute video clip for the two camera mode.

For this experiment, category maps were created in each video clip. We changed the sampling rate from 30 to 3 fps because high rate images include numerous mutually resemblant parts.

Figure 8. Sample images obtained using EDR: (**a**) front camera and (**b**) back camera.

5.2. Feature Extraction Results

Figure 9 depicts extracted scene features for an image obtained from the front camera. Figure 9b depicts an extraction result of AKAZE features from the original image depicted in Figure 9a. Feature points with orientations and scales are distributed on the road and over the traffic sign that is viewed by drivers. Figure 9c depicts high saliency regions extracted using SMs. Combined with Figure 9b,c, Figure 9d depicts an extraction result of AKAZE features in high saliency regions for a binary mask image. Green lines are boundaries of highly salient regions that include traffic signs and white lines.

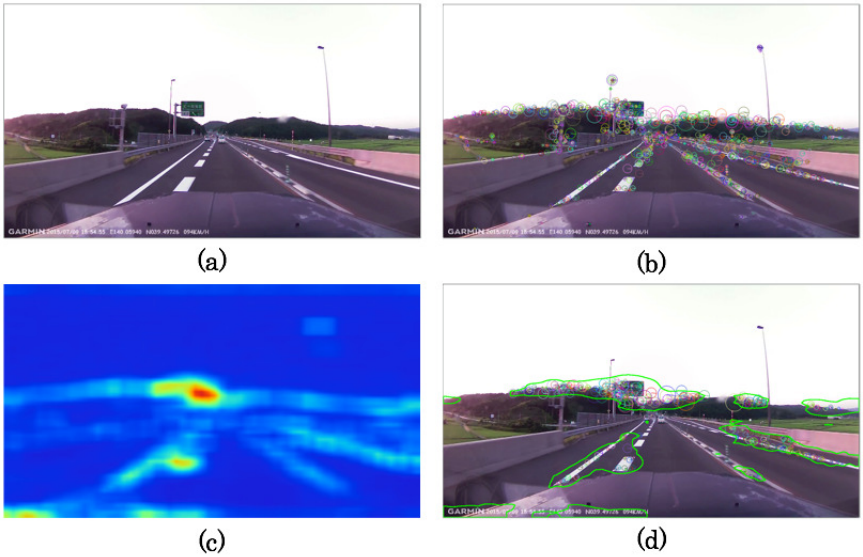

Figure 9. Extracted scene features: (**a**) original image, (**b**) all detected AKAZE features, (**c**) SM, and (**d**) AKAZE features of saliency regions.

Figure 10 depicts face extraction results for images obtained from the back camera. Wide variation of face orientations or brightness leads to detection failure. However, drivers are fixed in their seat by a seat belt. Therefore, the variation of a face size and its position is less than that of normal face detection applications. Before calculating a gap of size s and the center (c_x, c_y) between the current region of interest (RoI) at time t and the former RoI at $t-1$, our method corrects the RoI using the former RoI if the gap exceeds thresholds that were set in advance. Figure 10a,b respectively depict a successful example and a corrected example. The red RoI signifies that the former RoI was used for correction.

Figure 10. Face detection results: (**a**) detected face and (**b**) undetected and using the $(t-1)$-th RoI.

Figure 11 depicts four orientation GW images used for input features. We created codebooks for reducing fixed input vectors from features of all images. Figure 12 depicts codebooks for images in Figure 8. The numbers of AKAZE features differ among images, although the descriptor comprises 61 dimensions. Using codebooks, features are integrated to a fixed dimension. In Figure 12a,b, we set the mapping layer on SOMs to 128 units that engender the codebook dimension. We set the mapping layer on SOMs to 128 units that engender the codebook dimension. The GW features after downsampling comprise 900 dimensions of 30×30 patches. We reduced the dimensions using codebooks equivalent to the AKAZE dimension. Herein, Figure 12c depicts codebooks created using *k*-means [48] for inside images as a visual comparison.

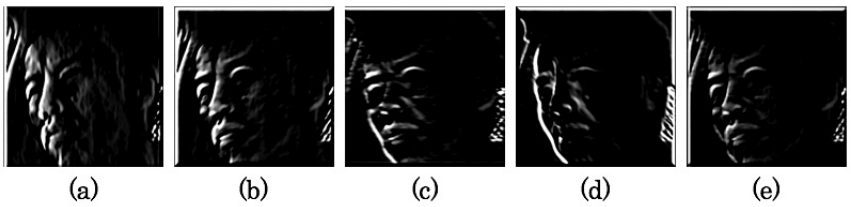

Figure 11. Four orientation features of GW: (**a**) 0°, (**b**) 45°, (**c**) 90°, (**d**) 135°, and (**e**) combined.

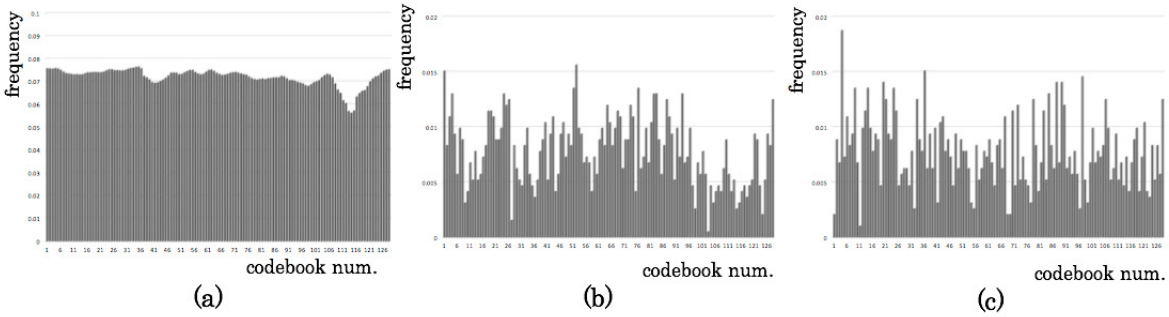

Figure 12. Three 128-dimensional codebooks created using (**a**) self-organizing maps (SOMs) for outside images, (**b**) SOMs for inside images, and (**c**) *k*-means for inside images.

5.3. Classification Granularity

Figure 13 depicts the numbers of ART labels and CPN categories according to changes in ρ from the 0.9950 to 0.9980 step by 0.001. Although ART labels are increased monotonically, CPN categories are increased with variation. The expanded difference between both numbers shows that the compression of CPNs works suitably for the redundance of labels created from ART. ρ most effectively determines classification granularity. For this study, we set ρ to 0.9970, which is the average of values of 0.9965 and 0.9975, as a steady range before expanding the gap separating ART labels and CPN categories.

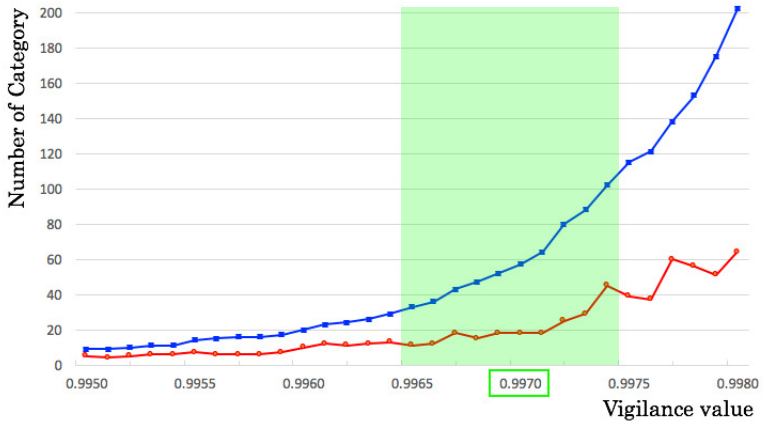

Figure 13. The relation between ρ and the numbers of adaptive resonance theory (ART) labels and counter propagation network (CPN) categories.

5.4. Created Category Maps

First, we created individual category maps using images of driving scenes and facial expressions separately. We set the size of category maps to $50 \times 50 = 2500$ units to ensure a sufficient mapping space related to the number of input images. Figure 14 depicts category maps running on a public road. The category maps are represented using color temperatures from blue to red that correspond to low and high temperatures, respectively. Using color temperature, one can confirm the order of categories according to the distribution. The vertical bar beside the category maps shows the indexes of color temperature that are divided by the number of categories. Based on neighborhood and competitive learning, ART labels that are removed by CPNs are not included in the indexes of color temperature. Figure 14 depicts category maps that comprise 14 categories for outside images and nine categories for inside images. According to the color temperature distribution, Figure 14a,b respectively depict a superior distribution for the first half and the last half of categories.

Subsequently, Figure 15 depicts category maps for a dataset driving on an expressway. Outside images and inside images were divided, respectively, into 13 categories and 8 categories. Similar to the results for a public road described above, the number of categories for outside images is greater than that of inside images. The categories of facial expressions are distributed throughout the map, although initial categories are distributed on the left and right regions of the map.

Figure 16 depicts category maps created using time-series images obtained from both outside and inside cameras. The codebook dimension of the combined input is double that of respective inputs. The numbers of categories on a public road and an expressway respectively represent eight and seven categories. The color temperature shows that the first half labels occupied the majority distribution. Compared with category maps in Figures 14 and 15, the effect of driving scenes is greater than that of facial expressions. We consider that this gap is caused by the difference of diverse codebooks depicted in Figure 12.

Figure 17 depicts time-series category changes mapped as a color bar. The color map corresponds to the color on the category map in Figure 16a. The color bar presents results of categories that correspond to the color temperature. The upper images respectively correspond to the 100th, 300th, 500th, and 700th frames. The brief route is depicted in an online map on the right side, which comprises the following. First, the car ran through a residential area to a main street after starting from a parking lot. Then, the car stopped at a traffic light before an intersection. After a few minutes, the car moved forward again. The zone corresponding to the high-temperature color shows the state stopped in front of a traffic light. The high-temperature color categories are present between the 400th and 500th frames, which correspond to driving scenes in heavy traffic after going to a main road.

Figure 18 depicts time-series category mapping results obtained while driving on an expressway. The category changes are compared to those of a case of a public road with monotonous driving scenes. The new category was created around the 620th frame because, at that point, the car ran into a tunnel. Subsequently, category changes were slight with momentary returns to existing categories.

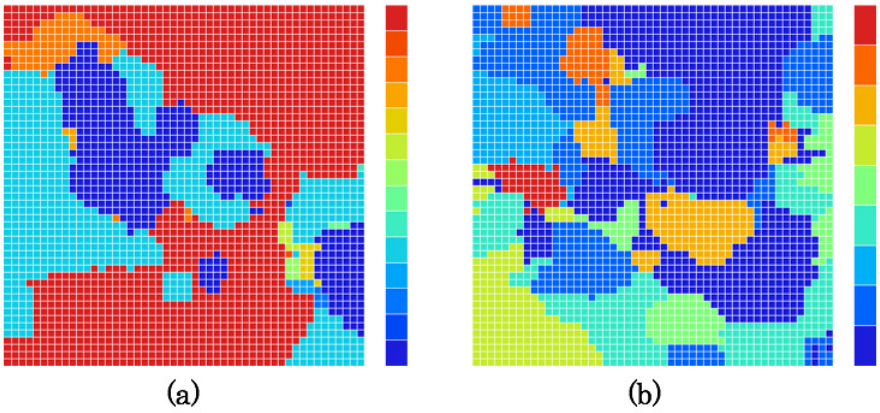

Figure 14. Category maps on a public road: (**a**) outside and (**b**) inside images.

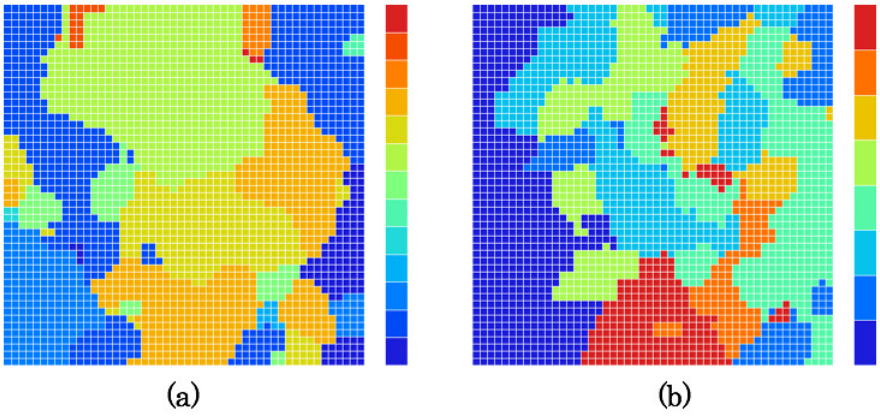

Figure 15. Category maps on expressway: (**a**) outside and (**b**) inside images.

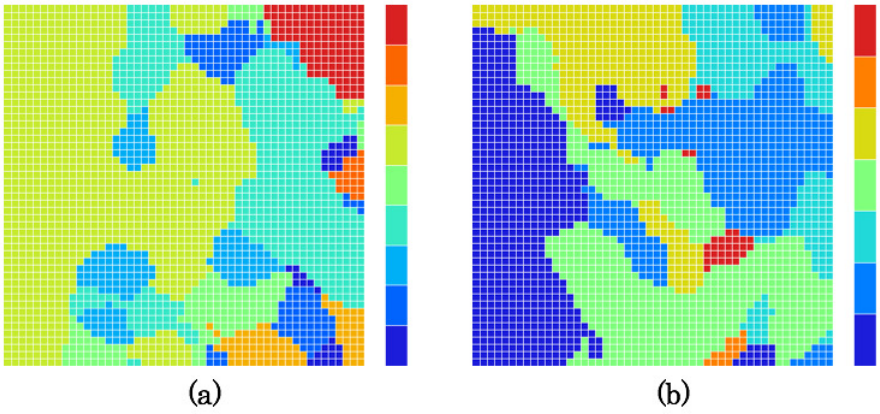

Figure 16. Category maps of driving scenes and facial expressions inputted together: (**a**) a public road and (**b**) an expressway.

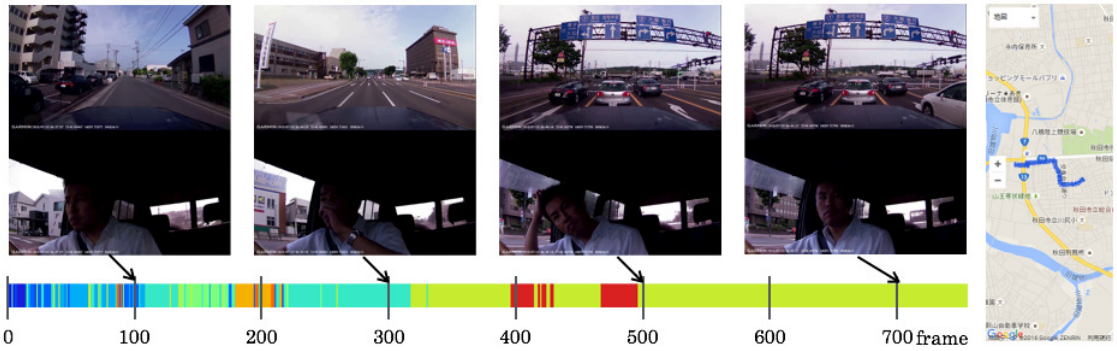

Figure 17. Time-series classification results obtained on a public road.

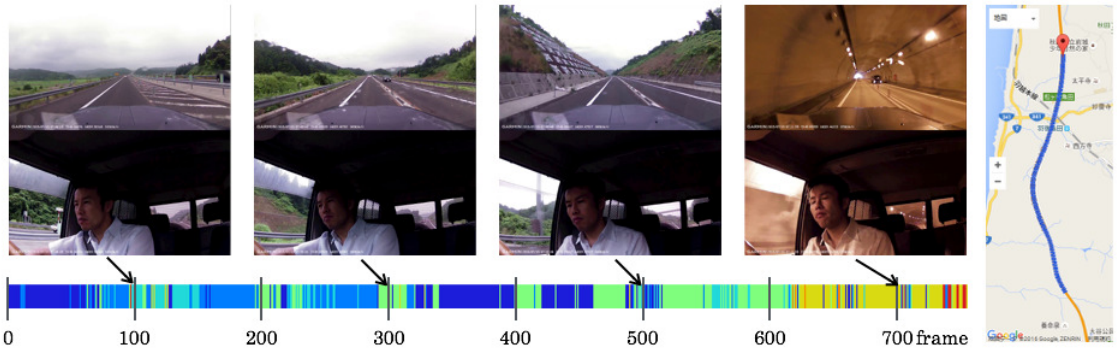

Figure 18. Time-series classification results on an expressway.

5.5. Driving Episodes with Near-Misses

During the data acquisition from three cars, we obtained two actual near-misses as shown in Table 2. Case I occurred during a summer evening. A compact car driven by an elderly man rushed suddenly out from a parking lot at a pachinko parlor. A crash was avoided by strong braking with fully locked tires because the car had no anti-lock braking system (ABS). The compact car passed from east to west. We infer that the elderly driver did not notice the car because of the bright sun in the evening. However, the actual reason was unclear; we did not hear any explanation from the driver. Figure 19 depicts salient signals of the accelerator embedded in the EDR. The red arrow indicates the position of harsh braking related to the near-miss.

Table 2. Actual near-misses.

Case	Near-Miss	Situation	Season
I	rushing out	evening	summer
II	slip	snow	winter

The dataset was classified into nine categories. In the first stage, categories were created throughout the map. In the middle stage, categories were created at the bottom of the map. The ninth category, created as the final category, is distributed on the upper right corner independently. Figure 20 depicts time-series classification results. The near-miss occurred between the 550th and 560th frames. The white circle depicted in Figure 21a corresponds to burst units related to the near-miss. Categories were changed around category boundaries. Existing categories were used for a transition without the creation of new categories for this near-miss.

Case II was a slip accident on an icy road. The car reduced its speed immediately before it reached a corner. However, the brakes did not work sufficiently, despite the fact that the driver applied full

pressure to their brake pedals. The car went into a space covered with snow, and the ABS functioned automatically. There was no salient value of the accelerator. Figure 21 depicts a category map of Case II. All six categories, a smaller number of categories compared to Case I, were distributed to several regions in small clusters. This distribution tendency indicates that the driving scenes differed partially but were similar globally. Figure 22 depicts the time-series classification results. Feature changes were salient because categories were mixed in a complex manner, with wide color temperature gaps. The near-miss occurred between the 390th and 400th frames. The category transition was active among existing categories. No new category was generated for the near-miss.

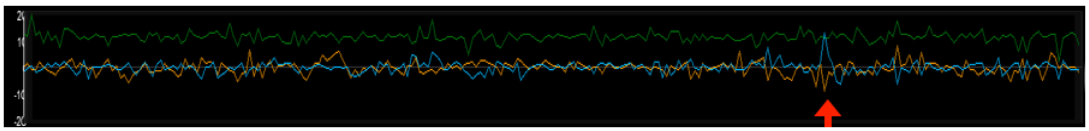

Figure 19. Output signals from the acceleration sensor. The red arrow indicates the position of the brakes during the near-miss of Case I.

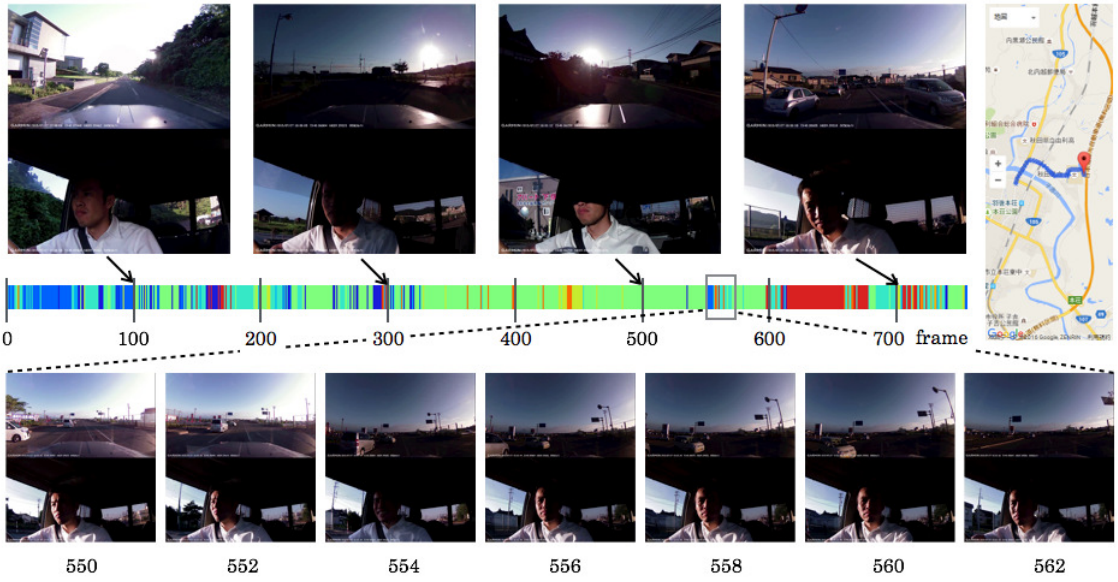

Figure 20. Time-series category transition results in Case I.

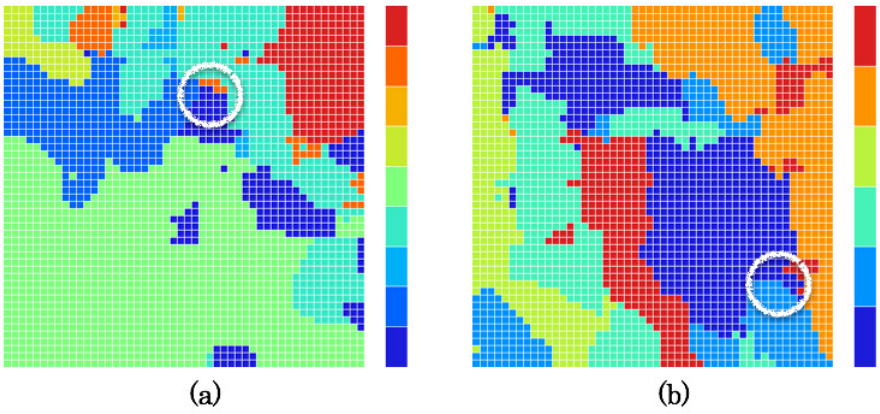

Figure 21. Category maps: (**a**) Case I and (**b**) Case II. White circles show burst units for the near-misses.

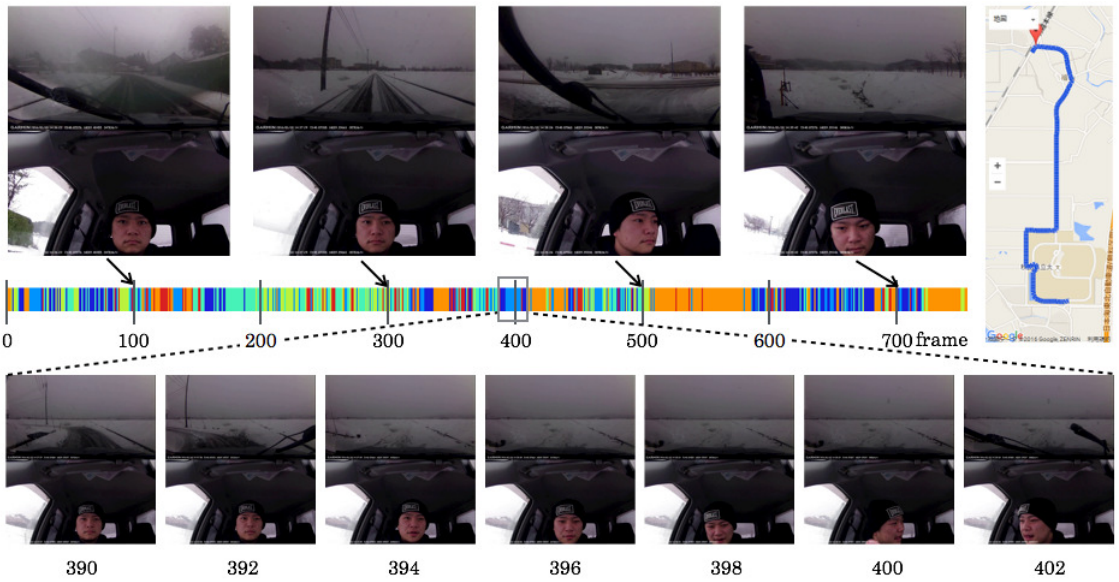

Figure 22. Time-series category transition results in Case II.

6. Conclusions

This study was undertaken to present driving episodes using machine-learning-based algorithms that address LTM and topological mapping. To this end, for preliminary experimentation using a DS, we measured the gazes and face orientations of drivers. Moreover, we measured the effects of facial features for cognitive distraction using simulated near-misses. Results show the possibility of recording driving episodes using image datasets obtained using an EDR with two cameras. Using category maps, we visualized driving features according to the driving scenes on a public road and an expressway. Moreover, we created original datasets that include near-misses and here describe the position of these near-misses on category maps.

In future studies, we will apply our methods to unknown driving environments and integrate both the collective intelligence and the experiential knowledge of numerous drivers. Moreover, we will develop a new interface that can sense actual responses and use that information to support recognition, judgment, and safety evaluations for elderly drivers.

Acknowledgments: This study was supported by Takata Foundation Research Grants and Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 17K00384.

Author Contributions: Hirokazu Madokoro and Kazuhito Sato conceived and designed the experiments; Nobuhiro Shimoi performed the experiments; Hirokazu Madokoro and Kazuhito Sato analyzed the data; Nobuhiro Shimoi contributed reagents/materials/analysis tools; Hirokazu Madokoroo wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

1. Woolley, A.W.; Chabris, C.F.; Pentland, A.; Hashmi, N.; Malone, T.W. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science* **2010**, *330*, 686–688.
2. *The Pathway to Driverless Cars: A Detailed Review of Regulations for Automated Vehicle Technologies*; Department for Transport: London, UK, 2015.
3. Anderson, J.M.; Kalra, N.; Stanley, K.D.; Sorensen, P.; Samaras, C.; Oluwatola, O.A. *Autonomous Vehicle Technology: A Guide for Policymakers*; RAND Corporation: Santa Monica, CA, USA, 2014.
4. Obinata, G. Nap Sign Detection during Driving Automobiles. *J. Jpn. Soc. Mech. Eng.* **2013**, *116*, 774–777. (In Japanese)
5. Uchida, N.; Hu, Z.; Yoshitomi, H.; Dong, Y. Facial Feature Points Extraction for Driver Monitoring System with Depth Camera. *Tech. Rep. Inst. Image Inf. Telev. Eng.* **2013**, *37*, 9–12. (In Japanese)

6. Hirayama, T.; Mase, K.; Takeda, K. Timing Analysis of Driver Gaze under Cognitive Distraction toward Peripheral Vehicle Behavior. In Proceedings of the 26th Annual Conference of the Japanese Society for Artificial Intelligence, Yamaguchi, Japan, 12–15 June 2012; pp. 1–4. (In Japanese)
7. Terada, Y.; Morikawa, K. Technology for Estimation of Driver: Distracted State with Electroencephalogram. *Panasonic Tech. J.* **2011**, *57*, 73–75. (In Japanese)
8. Matsunaga, J. Facial Expression Recognition Technology to understand the State of Driver. *J. Automot. Eng.* **2015**, *69*, 94–97. (In Japanese)
9. Dong, Y.; Hu, Z.; Uchimura, K.; Murayama, N. Driver Inattention Monitoring System for Intelligent Vehicles: A Review. *IEEE Trans. Intell. Transp. Syst.* **2011**, *12*, 596–614.
10. Apostoloff, N.; Zelinsky, A. Vision In and Out of Vehicles: Integrated Driver and Road Scene Monitoring. *Int. J. Robot. Res.* **2004**, *23*, 513–538.
11. Fletcher, L.; Zelinsky, A. Driver Inattention Detection based on Eye Gaze—Road Event Correlation. *Int. J. Robot. Res.* **2009**, *28*, 774–801.
12. Atance, C.M.; O'Neill, D.K. Episodic future thinking. *Trends Cogn. Sci.* **2001**, *5*, 533–539.
13. Tulving, E. Episodic Memory: From Mind to Brain. *Annu. Rev. Psychol.* **2002**, *53*, 1–25.
14. Schacter, D.L.; Benoit, R.G.; Brigard, F.D.; Szpunar, K.K. Episodic Future Thinking and Episodic Counterfactual Thinking: Intersections between Memory and Decisions. *Neurobiol. Learn. Mem.* **2015**, *117*, 14–21.
15. Maeno, T. How to Make a Conscious Robot: Fundamental Idea based on Passive Consciousness Model. *J. Robot. Soc. Jpn.* **2005**, *23*, 51–62. (In Japanese)
16. Carpenter, G.A.; Grossberg, S. ART 2: Stable Self-Organization of Pattern Recognition Codes for Analog Input Patterns. *Aied Opt.* **1987**, *26*, 4919–4930.
17. Kohonen, T. *Self-Organizing Maps*; Springer Series in Information Sciences; Springer: Berlin, Germany, 1995.
18. Nielsen, R.H. Counterpropagation networks. *Aied Opt.* **1987**, *26*, 4979–4983.
19. Madokoro, H.; Sato, K.; Nakasho, K.; Shimoi, N. Adaptive Learning Based Driving Episode Description on Category Maps. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 3138–3145.
20. Itti, L.; Koch, C.; Niebur, E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259.
21. Otsu, N. An Automatic Threshold Selection Method Based on Discriminant and Least Squares Criteria. *Trans. Inst. Electron. Commun. Eng. Jpn.* **1980**, *J63-D*, 349–356. (In Japanese)
22. Hou, X.; Zhang, L. Saliency Detection: A Spectral Residual Approach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
23. Rensink, R. Seeing, Sensing, and Scrutinizing. *Vis. Res.* **2000**, *40*, 1469–1487.
24. Rensink, R.A.; O'Regan, J.K.; Clark, J. To See or not to See: The Need for Attention to Perceive Changes in Scenes. *Psychol. Rev.* **1997**, *8*, 368–373.
25. Rensink, R.; Enns, J. Preemption Effects in Visual Search: Evidence for Low-Level Grouping. *Psychol. Rev.* **1995**, *102*, 101–130.
26. Treisman, A.; Gelade, G. A Feature-Integration Theory of Attention. *Cogn. Psychol.* **1980**, *12*, 97–136.
27. Wolfe, J. Guided Search 2.0: A Revised Model of Guided Search. *Psychon. Bull. Rev.* **1994**, *1*, 202–238.
28. Oliva, A.; Torralba, A. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *Int. J. Comput. Vis.* **2001**, *42*, 145–175.
29. Lowe, D.G. Object Recognition from Local Scale-Invariant Features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; pp. 1150–1157.
30. Alcantarilla, P.F.; Bartoli, A.; Davison, A.J. KAZE Features. *Lect. Notes Comput. Sci.* **2012**, *7577*, 214–227.
31. Alcantarilla, P.F.; Nuevo, J.; Bartoli, A. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. In Proceedings of the 24th British Machine Vision Conference, Bristol, UK, 9–13 September 2013.
32. Scharr, H. *Optimal Operators in Digital Image Processing*; Heidelberg University: Heidelberg, Germany, 2000.
33. Weickert, J.; Scharr, H. A Scheme for Coherence-Enhancing Diffusion Filtering with Optimized Rotation Invariance. *J. Vis. Commun. Image Represent.* **2002**, *13*, 103–118.
34. Yang, X.; Cheng, K.T. LDB: An Ultra-Fast Feature for Scalable Augmented Reality. In Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Atlanta, GA, USA, 5–8 November 2012; pp. 49–57.

35. Perona, P.; Malik, J. Scale-Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 1651–1686.
36. Viola, P.; Jones, M. Rapid Object Detection Using a Boosted Cascade of Simple Features. *Comput. Vis. Pattern Recognit.* **2001**, *1*, 511–518.
37. Rowley, H.; Baluja, S.; Kanade, T. Neural Network-Based Face Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 22–38.
38. Papageorgiou, C.; Oren, M.; Poggio, T. A General FrameWork for Object Detection. In Proceedings of the Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), Bombay, India, 4–7 January 1998.
39. Haar, A. Zur Theorie der Orthogonalen Funktionensysteme. *Math. Ann.* **1910**, *69*, 331–371. (In German)
40. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Aication to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139.
41. Schapire, R.E.; Freund, Y. *Boosting: Foundations and Algorithms*; The MIT Press: Cambridge, MA, USA, 2012.
42. Amari, S. *Encyclopedia of Brain Sciences*; Asakurashoten Press: Tokyo, Japan, 2000. (In Japanese)
43. Hubel, D.H.; Wiesel, T.N. Functional Architecture of Macaque Monkey Visual Cortex. *Proc. R. Soc. B* **1978**, *198*, 1–59.
44. Lee, T.S. Image representation using 2D Gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.* **1996**, *18*, 959–971.
45. Madokoro, H.; Shimoi, N.; Sato, K. Adaptive Category Mang Networks for All-Mode Topological Feature Learning Used for Mobile Robot Vision. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Edinburgh, UK, 25–29 August 2014; pp. 678–683.
46. Sudo, A.; Sato, A.; Hasegawa, O. Associative Memory for Online Learning in Noisy Environments Using Self-Organizing Incremental Neural Network. *IEEE Trans. Neural Netw.* **2009**, *20*, 964–972.
47. Csurka, G.; Dance, C.R.; Fan, L.; Willamowski, J.; Bray, C. Visual Categorization with Bags of Keypoints. In Proceedings of the ECCV Workshop Statistical Learning Computer Vision, Prague, Czech Republic, 16 May 2004; pp. 1–22.
48. McQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Statistics*; University of California Press: Berkeley, CA, USA, 1967; pp. 281–297.
49. Vesanto, J.; Alhoniemi, E. Clustering of the Self-Organizing Map. *IEEE Trans. Neural Netw.* **2000**, *11*, 586–600.
50. Terashima, M.; Shiratani, F.; Yamamoto, K. Unsupervised Cluster Segmentation Method Using Data Density Histogram on Self-Organizing Feature Map. *Trans. Inst. Electron. Inf. Commun. Eng.* **1996**, *J79-D-II*, 1280–1290. (In Japanese)
51. Carpenter, G.A.; Grossberg, S. *Pattern Recognition by Self-Organizing Neural Networks*; The MIT Press: Cambridge, MA, USA, 1991.
52. Sato, K.; Ito, M.; Madokoro, H.; Kadowaki, S. Driver Body Information Analysis for Distraction State Detection. In Proceedings of the IEEE International Conference on Vehicular Electronics and Safety (ICEVS2015), Yokohama, Japan, 5–7 November 2015.
53. Suenaga, O.; Nakamura, Y.; Liu, X. Fundamental Study of Recovery Time from External Information Processing while Driving Car. *Jpn. Ergon. Soc. Ergon.* **2015**, *51*, 62–70. (In Japanese)
54. Abe, K.; Miyatake, H.; Oguri, K. Induction and Biosignal Evaluation of Tunnel Vision Driving Caused by Sub-Task. *Trans. Inst. Electron. Inf. Commun. Eng. A* **2008**, *J91-A*, 87–94.
55. Ishibahi, M. *HQL Manual of Driving Style Questionnaires*; Research Institute of Human Engineering for Quality Life: Tsukuba, Japan, 2003. (In Japanese)
56. Ishibahi, M. *HQL Manual of Driving Workload Sensibility Questionnaires*; Research Institute of Human Engineering for Quality Life: Tsukuba, Japan, 2003. (In Japanese)
57. Regan, M.A.; Lee, J.D.; Young, K.L. *Driver Distraction: Theory, Effects, and Mitigation*; CRC Press: Boca Raton, FL, USA, 2008.
58. Hynd, D.; McCarthy, M. *Study on the Benefits Resulting from the Installation of Event Data Recorders*; Project Report PRP707; Transport Research Laboratory: Wokingham, UK, 2014.