

Driving Environmental Change Detection Subsystem in a Vision-Based Driver Assistance System

Chiung-Yao Fang¹, Chiou-Shann Fuh², and Sei-Wang Chen³

¹Department of Information and Computer Education
National Taiwan Normal University, Taipei, Taiwan, R. O. C.
162, Heping East. Road SEC. 1, Taipei, Taiwan, R.O.C.

²Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan, R. O. C.

³Institute of Computer Science and Information Engineering
National Taiwan Normal University, Taipei, Taiwan, R. O. C.

Abstract—We propose a computational model motivated by human cognitive processes for detecting changes of driving environments. The model consists of three major components: sensory, perceptual, and conceptual components. The proposed computational model is used as the underlying framework in which a system for detecting changes of driving environments is developed.

I. INTRODUCTION

The objective of detecting changes of driving environments is threefold. Firstly, since there are several modules involved in a driver assistance system [1], the overall performance of the entire system depends not only on the effectiveness of individual modules but on their conspiracy. Conditions of driving environments provide one of the important clues for coordinating the modules. Secondly, parametric values embedded in the modules should be updated in accordance with environmental changes, such as the changes in illumination, weather condition, vehicle speed, road surface, etc.. Finally, dramatic changes of environments during driving usually correspond to critical traffic situations. The change detection module could reduce the probability of occurrence of accidents by alerting a priori the attentions of drivers to rapid changes of driving environments.

Few researchers have discussed how to detect "changes" of environments during driving. In video sequences recorded by a camcorder mounted on a vehicle, "change" is difficult to define because everything is changing during driving. We, therefore, confine ourselves to the changes, that may be encountered while driving on an expressway. In this study, the changes under consideration include moving to the lane on the left (left-lane-cross), moving to the lane on the right (right-lane-cross), entering the expressway (expressway-entry), exiting the expressway (expressway-exit), entering a tunnel (tunnel-entry), exiting a tunnel (tunnel-exit), and approaching an overpass (overpass-ahead). The developed change detection system will handle all the above cases both in daytime and at night.

II. COMPUTATIONAL MODEL

While computers reveal excellent in computation, they are unhandy in recognition. On the contrary, human brains exhibiting impressive powers in recognition are not so good as computers in the computational ability. We believe the

features of parallel processing and distributed representation of the human brain play the important roles. Fig. 1 depicts the proposed computational model. There are three analyzers consisting of the model, referred to as the sensory, perceptual, and conceptual analyzer, respectively, in a cognitive process [3]. According to psychologists [4, 5], motion is typically analyzed and perceived earlier than form and significance. Explicitly, for a moving object its motion rather than its shape and meaning would first grasp one's attention. We hence compute first the temporal and spatial information of dynamic scenes in the sensory analyzer of the computational model. The former information contains motions of objects and the latter comprises locations of objects.

However, in the course of driving every thing is moving. How can the objects of interest be attended to among a jumble of moving objects? Intuitively, the extraneous characteristics (e.g., shapes, colors, and textures) of objects may be employed to distinguish objects. In the perceptual analyzer of the computational model, an effortful (voluntary) selectivity of attention is realized by introducing a neural module, called the spatiotemporal attentional (STA) neural network, and a long term memory (LTM), which preserves the characteristics of interesting objects. The information from the LTM will call the attention of the neural network to only the interesting objects when it is being innervated by the stimuli (i.e., the spatial and temporal information) coming from the sensory analyzer. Afterwards, the activations of the STA neurons are examined. If there is no focus of attention formed over the neurons, the system continues to process subsequent incoming images. Otherwise, the step of feature extraction is evoked to detect categorical features from the restricted image areas corresponding to the focuses of attention in the STA neural module.

The categorical features detected in the perceptual analyzer act serve as the input stimuli, represented as a supraliminal pattern, to a configurable adaptive resonance theory (CART) neural network in the conceptual analyzer. The input supraliminal pattern first initializes the LTM of the CART with the contents from an episodic memory which are consistent with the supraliminal pattern. We refer to this initialization of the LTM as the configurable characteristic of the CART neural module. Subliminal patterns later to be matched against the supraliminal pattern will be retrieved

from the LTM. The configurable property significantly reduces the search space of subliminal patterns. During comparing the supraliminal pattern with a subliminal pattern, if they are similar enough, the class of the supraliminal pattern is regarded as that of the subliminal pattern under consideration. The CART neural module then conducts a supervised learning in which the subliminal input in the LTM is update under the guidance of the supraliminal pattern. On the other hand, if no subliminal expectation proximate to the supraliminal pattern could be found in the LTM, an unsupervised learning for representing the supraliminal pattern in the LTM is performed. Finally, the episodic memory is updated with the current content of the LTM.

III. SPATIOTEMPORAL ATTENTIONAL NEURAL NETWORK

The STA neural module, as Fig. 2 shows, is structured as a two-layer network: one input layer and one output layer. The output layer is also referred to as the attentional layer. Neurons in this layer are arranged into a 2D array in which they are interconnected to one another. These connections are within-layer (lateral) connections and are almost always inhibitory. There are no synaptic links among input neurons; they are, however, fully connected to the output neurons. These connections are called between-layer (vertical) connections and are always excitatory.

We assume that the input neurons are organized into a 2D array as well corresponding to that of attentional neurons. In this study, both the arrays have the same size as images. Let w_{ij} denote the synaptic strength (i.e., weight) of the link between attentional neuron n_i and input neuron n_j . The strength vector of attentional neuron n_i is written as $\mathbf{w}_i = (w_{i1}, w_{i2}, \dots, w_{im})$, where m is the number of input neurons (the same as that of attentional neurons). The input to attentional neuron n_i due to input stimuli \mathbf{x} is formulated as

$$I_i^v = \mathbf{w}_i \cdot \mathbf{x} = \sum_{j=1}^m w_{ij} x_j. \quad (1)$$

The linking strengths between the input and attentional layers are defined as follows. Referring to Fig. 3, let n_j be any input neuron and n_i be the corresponding neuron in the attentional layer. Assume that a 3D Gaussian G is centered at attentional neuron n_i . Then, the linking weight w_{kj} between input neuron n_j and any attentional neuron n_k is defined as $w_{kj} = G(\mathbf{r}_{ki})$, where \mathbf{r}_{ki} is the positional vector of neuron n_k relative to neuron n_i .

The lateral interaction among attentional neurons is characterized by a "Mexican-hat" function, denoted by $M(\mathbf{r})$, where \mathbf{r} is a positional vector relative to the center of the function. The function will play an important role in focusing attentions, or equivalently clustering activations over attentional neurons. The "Mexican-hat" function is often approximated by the Laplacian of Gaussian or the difference of Gaussians. The input to attentional neuron n_i due to lateral interaction is then defined as

$$I_i^l = \sum_{k \in N_i, k \neq i} [u_{ik} M(\mathbf{r}_k - \mathbf{r}_i) a_k], \quad (2)$$

where N_i is the set of neighbors of neuron n_i , a_k is the activation of attentional neuron n_k , and u_{ik} is the synaptic

strength of the lateral connection between neurons n_i and n_k with positions \mathbf{r}_i and \mathbf{r}_k , respectively.

We finally formulate the net input to attentional neuron n_i as

$$net_i = a_i + f(-ba_i + c[I_i^v + I_i^l - \Gamma]^+), \quad (3)$$

where b and c are positive constants; the former represents the decay rate of activation a_i , and the latter weights the total input coming from the input neurons and the neighbors of attentional neuron n_i . Constant Γ serves as a threshold for preventing the effect from noise. The functions $[\cdot]^+$ and $f(\cdot)$ are defined as follows [2]:

$$[x]^+ = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}, \quad (4)$$

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ dx & \text{if } x \leq 0 \end{cases}, \quad (5)$$

where $1 > d > 0$.

After computing the net input, the actual activation of attentional neuron n_i is given by $a_i = \psi(net_i)$, where $\psi(\cdot)$ is a transfer function, which is typically simulated using a sigmoid function.

In the following, we address some behaviors of the entire neural module. Suppose that a point stimulus is fed to a neuron on the input layer. The point stimulus laid on the input neuron then innervates the attentional layer through vertical links. Instead of equal stimulation, the attentional neurons are activated by the stimulus with different degrees of innervation governed by a Gaussian function. In other words, the point stimulus on the input layer will become a Gaussian spread stimulus before arriving the attentional layer. However, the lateral interactions, controlled by a Mexican-hat function, among attentional neurons cluster the activations of the neurons. The above spread and clustering mechanisms of the neural module are analogous to the morphological operations of dilation and erosion, respectively. There are several advantages associated with the combinations of these two operations. Of which, the most interesting one may be the compensation for imperfect input data.

Next, consider the temporal behavior of the neural module. Refer to Fig. 2. Suppose that a moving point stimulus is applied to attentional neuron n_i at time t and late moves to its neighboring neuron n_k at time $t+1$. Assume that the activation of attentional neuron n_i is $a_i(t-1) = 0$ at time $t-1$. From Eqs. (1) - (3), the inputs to the neuron due to the stimulus and lateral interactions are $I_i^v(t) (> \Gamma)$ and $I_i^l(t) (=0)$, respectively, and as a consequence the net input to attentional neuron n_i is $net_i(t) = A(c[I_i^v(t) - \Gamma]^+)$. This leads its activation to $a_i(t) = \psi(net_i(t))$. At this moment, the activation $a_k(t)$ of attentional neuron n_k remains zero. At time $t+1$, since the point stimulus moves to attentional neuron n_k , the activation of neuron n_i begins to decay while the activation of neuron n_k is arising. Since the decay rate is smaller than the arising rate, if the point stimulus keeps moving it will leave a linear trace on the attentional layer. In

practice, surface-like traces are created that preserve both the shapes and paths of stimuli.

IV. CHANGE DETECTION OF DRIVING ENVIRONMENTS

A. Sensory Component

The input data to the change detection system are color image sequences acquired using a camcorder mounted on a moving vehicle. The video sequences are generally unstable in view of vibrations of the vehicle while driving. We take care of the issue of image in stability implicitly in the steps during extracting information from video sequences.

Fig. 4 depicts the flowchart of information acquisition from a video sequence S . Let $I(t)$ be its t th image frame. Image $I(t)$ is first subsampled (1/4 in practice) in order to reduce the processing time. There is a by-product of this subsampling step; that is, both noise and image instability will somehow be suppressed. Let $I'(t)$ be the subsampled image. Next, a low-intensity image, $L(t)$, and a high-intensity image, $H(t)$, are calculated by $L(t) = \min(I'(t), L(t-1))$ and $H(t) = \max(I'(t), H(t-1))$, respectively, where $L(0) = I'(0) = H(0)$. The \min and \max operations in the above equations are conducted on the (pixel, color component)-wise basis. The low-intensity and high-intensity images retain the minimum and maximum pixel values of the video sequence up to time t , respectively. Afterwards, the difference image $D(t)$ is calculated according to $D(t) = I_H(t) - I_L(t)$, where I_H and I_L are the intensity components of images H and L , respectively. The difference image highlights the image areas where relatively static objects with respect to the viewer should stay if no environmental change occurs thereafter. It is these image areas that the sensory component can tolerate both drift of the vehicle during driving and the instability of the input video sequence. Finally, the derivative image $D'(t)$ is computed from $D'(t) = |D(t) - D(t-1)|$. Once the relatively static objects begin to move out their image areas of stay, traces of the objects will appear in the derivative image. The derivative image preserves both spatial and temporal information of the objects.

Refer to Fig. 5. Column (a) of the figure shows the first seven images of a video sequence. The images are arranged from top to bottom in sequence. Columns (b) and (c) display the high-intensity and low-intensity image sequences, respectively. These two sequences are used to calculate the difference image sequence D shown in column (d). There are several bright areas in the difference images which are created by relatively static objects. The areas turn into bands with time if no significant environmental change occurs. Column (e) shows the derivative image sequence D' . The entries of derivative images indicate the degrees of environmental change. In this example, no environmental change is detected because of small entries of derivative images. Refer to another example shown in Fig. 6 where only the input video sequence and the associated derivative sequence are displayed. In this example, the vehicle moves from the right to the left lanes. The lane marks keep moving

from left to right in the video sequence and produce large entries in certain areas of the derivative images.

B. Perceptual Component

The derivative images computed in the sensory component are transmitted to the perceptual component. There are two major modules in the perceptual component: a STA neural module and a feature extraction module. The derivative images from the sensory component are first fed into the neural module. If there is no environmental change, the derivative images contain small entries (see the example in Fig. 5). These entries will not be able to activate the attentional neurons in the STA neural module (due to the threshold Γ in Eq. (3)). However, if large entries appear in the derivative images, some pattern of activation over the attentional neurons can be generated. The pattern then grows with time if the environmental change continues. The pattern eventually reaches its maximum and forms focuses of attention in the attentional layer of the neural module.

Refer to Fig. 7 in which an example of right-lane-cross is exhibited. In the figure, only the input video sequence and the associated attentional sequence of the STA neural module are displayed. See the attentional sequence; a pattern of activation is generated over the attentional layer when the vehicle begins to change its lanes from the left to the right one. The pattern grows keeping pace with the actions of lane change. Eventually, the pattern reaches its maximum. At this moment, the vehicle has completed its lane change actually. Afterwards, the pattern gradually decays and finally it disappears if no environmental change happens again during this period.

Different types of environmental changes will lead to distinct patterns of activation over the attentional layer. To see this, let us consider the cases of left-lane-cross and right-lane-cross. In the former case, the left lane marker shifts from left to right in the video sequence, causing a moving stimulus sequentially innervating the attentional layer from its left to right sides. A pattern of activation is thus created over the attentional layer. Since the neurons in the left part of the attention layer activate and decay earlier than those in the right part of the layer, the activations of neurons in the left part of the pattern are smaller than those in the right part. Conversely, in the case of right-lane-cross, the activations of neurons in the right part are smaller than those in the left part. Fig. 8 shows the patterns of activation of different types of environmental changes.

C. Conceptual Component

The categorical feature extracted in the perceptual component serves as a supraliminal feature to be fed into the CART neural module in the conceptual component. The CART neural module is actually an ART2 neural network [2] with a configurable long term memory (CLTM).

The long term memory of the ART2 neural network is disposed to the bidirectional links between the category and the input representation fields in the attentional subsystem. Rather than fixed, the links and the associated components are reconfigurable. Note that there are several subsystems

(e.g., road sign, obstacle, traffic condition, and environmental change detection subsystems) comprising a driver assistance system. They share a common storage, referred to as the episodic memory in this study, for knowledge representation. The subliminal features of the environmental change detection subsystem under consideration will be loaded from the episodic memory into the long term memory of the neural network. The links and the associated components of the long term memory have to be configured accordingly.

The episodic memory of the driver assistance system has to be refreshed with the content of the LTM of the CART neural module each time when the environmental change detection subsystem is to be terminated. Thus, the latest subliminal features will be preserved for later uses. Moreover, the CART neural module can be exploited by the other subsystems, which may need it.

V. EXPERIMENTAL RESULTS

In the experiments, video sequences containing all the change conditions we discuss above are applied to our system. Each sequence is converted into about two hundred digital images by computer software (Adobe Premiere 5.1). The size of each image is reduced to 160×120 by subsampling; the time between two successive images is 0.2 seconds. In this section we show only a portion of the experimental results, including one sequence recorded at night.

In practice, training the STA neural network is unnecessary, because its output reflects the present situation. Fig. 9 shows the experimental results of the tunnel-entry sequence (the values of the parameters in Eqs. (3) and (5) are as follows: $b = 0.5$; $c = 1$; $d = 0.5$). Once our system has already detected an environmental change and confirmed the change, these two images are refreshed for the following detection. The shape of a tunnel and the darkness of its interior give clues to its existence. Since the image of the tunnel gets bigger as it gets closer while driving, this change is reflected in the attentional patterns, and then the focus of attention moves to the boundary of the tunnel. Once this condition has been detected, our system should prepare to operate in a dim driving environment by switching illumination parameters.

Fig. 10 illustrates an expressway-entry condition. This condition can be detected because of the special marking on the pavement at an expressway entry. The attentional patterns in this change condition are different from those of a right-lane-cross condition. The focus of attention never moves to the left. Detecting this change condition can alert the driver to pay attention to his right-hand side, because some other vehicles may be entering the expressway from that direction.

Fig. 11 shows the right-lane-cross condition at night. Both Figures 11 and 7 show similar attentional patterns because both have the same change condition. The lane markers are clearly seen because of the street light and the vehicle's headlights. Note that the dark background at night somewhat decreases the effects of noise compared to the light background in the daytime, so it actually seems easier to detect lane changes at night.

Although ART2 can be an unsupervised-learning system, in our experiments we train the ART2 first for classifying the known change conditions. In the learning stage we select two or three attentional patterns of each change condition as the training patterns. Since each ART2 neural network converges within two iterations, it spends little time on training. In our system once a sequence is input to the STA, all the attentional patterns forming the focus of attention are applied to the ART2 neural network. These attentional patterns, no matter whether or not they have been learned by the ART2, all should be classified. A few attentional patterns are classified incorrectly, but they do not affect the accuracy of our system. Our system outputs the results only when three successive attentional patterns are classified into the same class. Through this step, our system can detect change conditions correctly and robustly.

Using a 700 MHz Pentium III CPU, our system can process one image in 0.5 seconds, which is slower than real time (0.2 seconds). In the study, we did not use any strategy to improve the computational efficiency. Hence, we believe the processing time can be greatly reduced by additional effort.

VI. CONCLUSIONS AND FUTURE WORK

An environmental change detection system for expressway driving was developed in this paper. We may extend the system to manage different types of environmental changes in addition to those encountered in expressways. However, this will lead to an issue about multiple environmental changes. The system may confuse with the situations in which several environmental changes occur at the same time. More robust strategies are still desired to resolve complex conditions.

A general purpose computational model motivated by human cognitive processing has been proposed in this study. Besides the environmental change detection system, other detection systems for road signs, traffic signals, obstacles, and weather conditions can also be developed on the basis of the proposed model. Most of detecting tasks can be achieved by means of pattern classification. For the objective of recognition, the current computational model has to be extended.

ACKNOWLEDGEMENT

This work was supported by the National Science Council, R. O. C. (NSC-89-2218-E-003 -001).

REFERENCE

- [1] T. Bachmann and S. Bujnoch, "Connected Drive--Driver assistance systems of the future," <http://195.30.248.73/mercatorpark/pdf/connectedDrive.pdf>, 2001.
- [2] G. A. Carpenter and S. Grossberg, *Pattern Recognition by Self-Organizing Neural Networks*, Ed., The MIT Press, Cambridge, Massachusetts, pp. 397-423, 1991.
- [3] J. Konorski, *Integrative Activity of the Brain*, Chicago: University of Chicago Press, 1967.

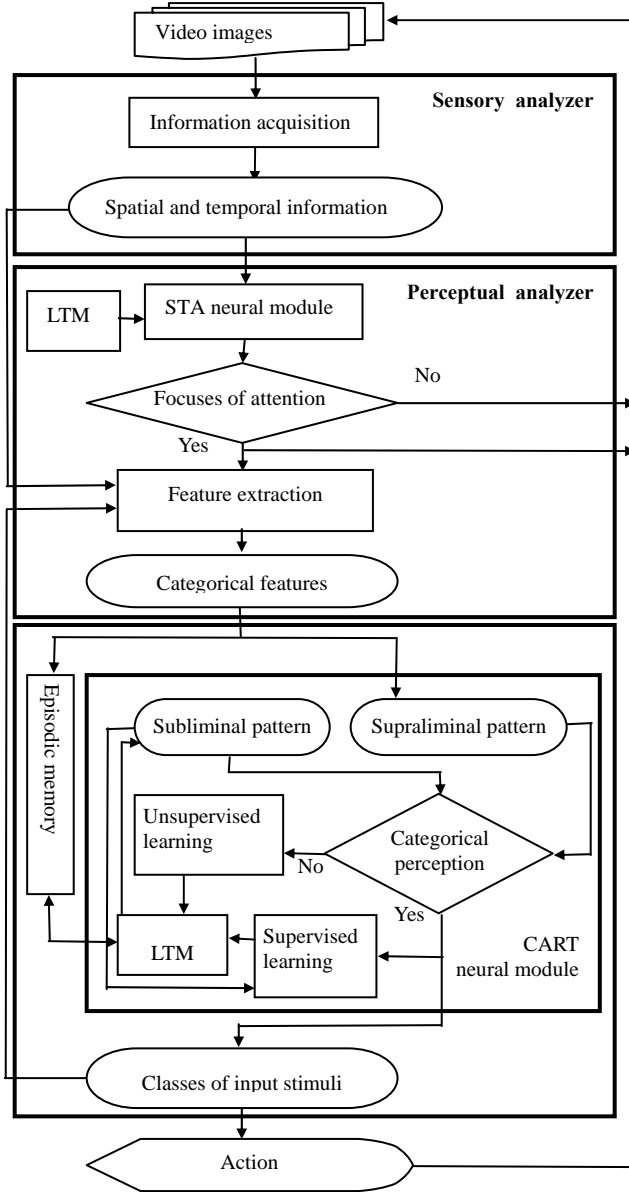


Figure 1: The proposed computational framework.

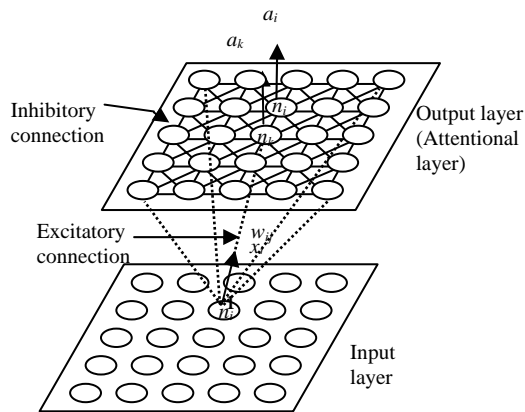


Figure 2: STA neural network.

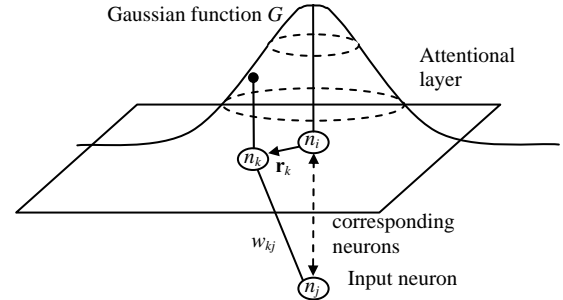


Figure 3: The linking strengths between the input and the attentional layers.

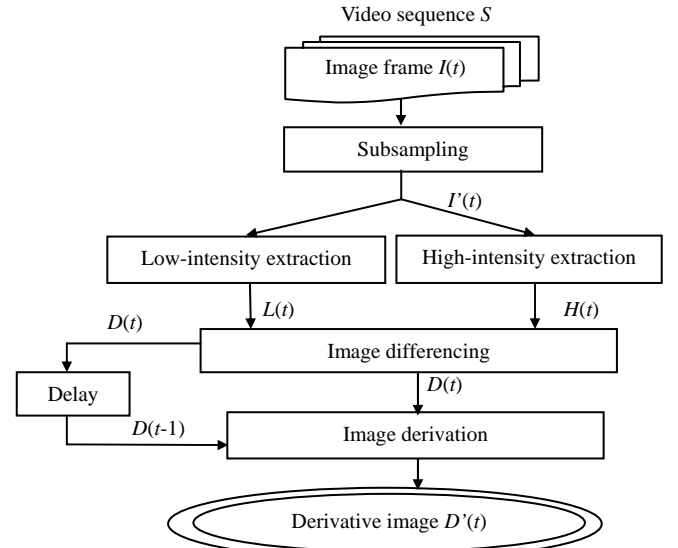


Figure 4: Flowchart of spatial and temporal information acquisition.

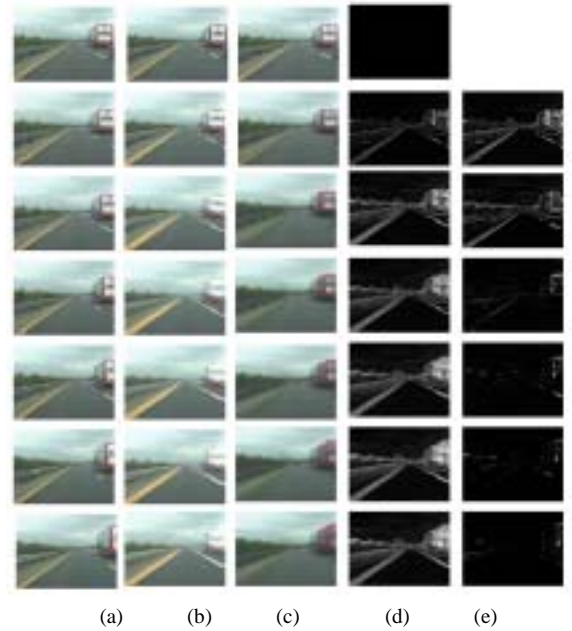


Figure 5: Example illustrating the steps of the sensory component. (a) Input image sequence I . (b) High-intensity image sequence H . (c) Low-intensity image sequence L . (d) Difference image sequence D . (e) Derivative image sequence D' .



Figure 6: Example of left-lane-cross. The upper row is the original image sequence, the lower one is the derivative image sequence S .

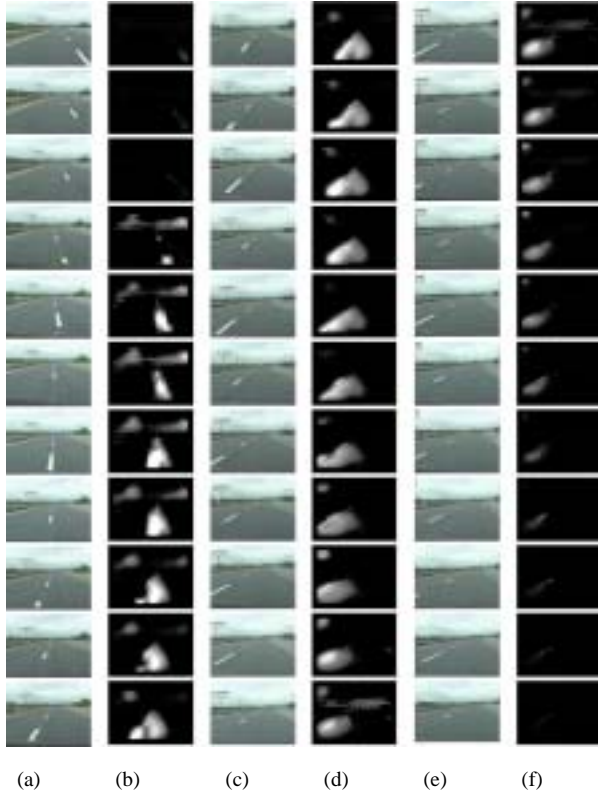


Figure 7: Example of right-lane-cross. Columns (a), (c), and (e) are the input image sequence I , and columns (b), (d), and (f) show the associated attentive sequence of the STA neural module.

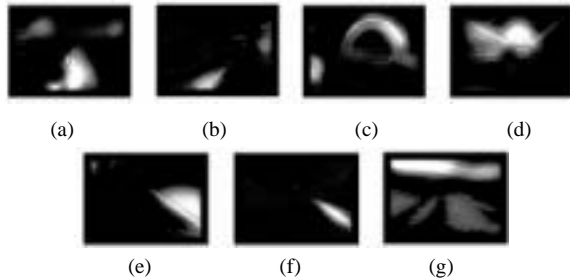


Figure 8: The patterns of activation of different types of environmental changes. (a) Left-lane-cross. (b) Right-lane-cross. (c) Tunnel-entry. (d) Tunnel-exit. (e) Expressway-entry. (f) Expressway-exit. (g) Overpass-ahead.

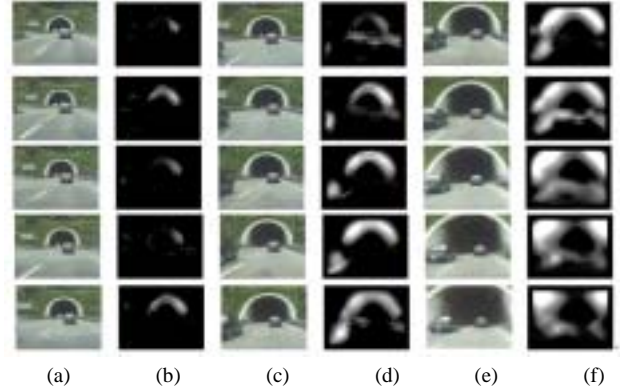


Figure 9: Example of tunnel-entry. Columns (a), (c), and (e) are the input image sequence I , and columns (b), (d), and (f) show the output sequence of the STA neural network.

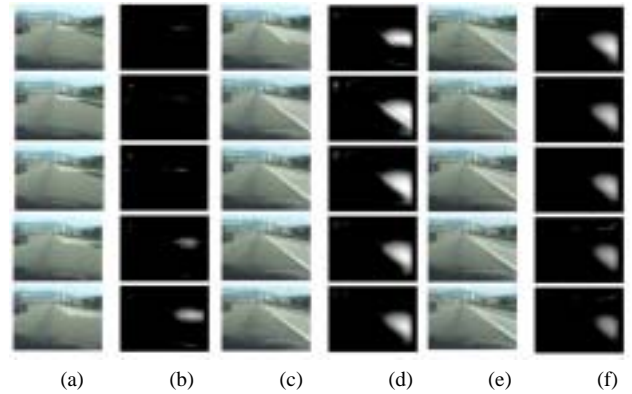


Figure 10: Example of expressway-entry. Columns (a), (c), and (e) are the input image sequence I , and columns (b), (d), and (f) show the output sequence of the STA neural network.

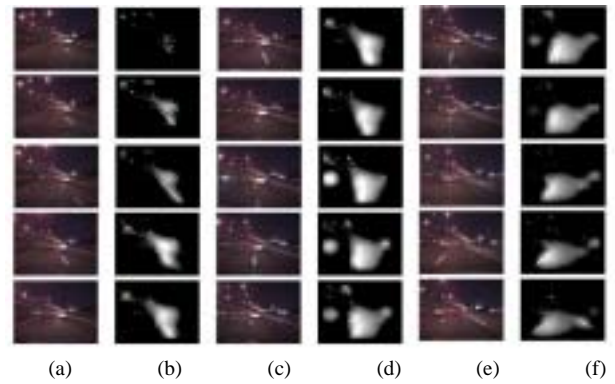


Figure 11: Example of right-lane-cross at night. Columns (a), (c), and (e) are the input image sequence I , and columns (b), (d), and (f) show the output sequence of the STA neural network.

- [4] C. Martindale, *Cognitive Psychology, A Neural-Network Approach*, Brooks/Cole Publishing Company, Pacific Grove, California, pp. 95-116, 1991.
- [5] S. M. Zeki, "Functional Organization of a Visual Area in the Posterior Bank of the Superior Temporal Sulcus of the Rhesus Monkey," *Journal of Physiology*, 236, pp. 549-573, 1974.