Automatic Change Detection of Driving Environments in a Vision-Based Driver Assistance System

Chiung-Yao Fang, *Associate Member, IEEE*, Sei-Wang Chen, *Senior Member, IEEE*, and Chiou-Shann Fuh

***Abstract*—Detecting critical changes of environments while driving is an important task in driver assistance systems. In this paper, a computational model motivated by human cognitive processing and selective attention is proposed for this purpose. The computational model consists of three major components, referred to as the sensory, perceptual, and conceptual analyzers. The sensory analyzer extracts temporal and spatial information from video sequences. The extracted information serves as the input stimuli to a spatiotemporal attention (STA) neural network embedded in the perceptual analyzer. If consistent stimuli repeatedly innervate the neural network, a focus of attention will be established in the network. The attention pattern associated with the focus, together with the location and direction of motion of the pattern, form what we call a categorical feature. Based on this feature, the class of the attention pattern and, in turn, the change in driving environment corresponding to the class are determined using a configurable adaptive resonance theory (CART) neural network, which is placed in the conceptual analyzer. Various changes in driving environment, both in daytime and at night, have been tested. The experimental results demonstrated the feasibilities of both the proposed computational model and the change detection system.**

***Index Terms*—Cognitive model, configurable adaptive resonance theory (CART) neural network, driver assistance system, sensory, perceptual, and conceptual analyzers, spatiotemporal attention (STA) neural network, system to detect change in driving environment.**

I. Introduction

DIVERSE methods for improving driving safety have been proposed. They can be roughly categorized into passive or active. Passive means (e.g., seat-belts, airbags, and anti-lock, braking systems), which have significantly reduced traffic fatalities, were originally introduced to diminish the degree of injury during an accident. Active means on the other hand are designed to prevent accidents in the first place. Driver assistance systems [2], [27], [30], [31], [36] are one kind of active system that are intended to bring to the attention of a driver to the potential of a dangerous situation as soon as possible.

Manuscript received January 15, 2001; revised June 5, 2002 and October 23, 2002. This work was supported by the National Science Council, R.O.C., under Contract NSC-89-2218E-003-001.

C.-Y. Fang is with the Department of Information and Computer Education, National Taiwan Normal University, Taipei, Taiwan, R.O.C. (e-mail: violet@ice.ntnu.edu.tw)

S.-W. Chen is with the Department of Computer Science and Information Engineering, National Taiwan Normal University, Taipei, Taiwan, R.O.C.

C.-S. Fuh is with the Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, R.O.C.

Digital Object Identifier 10.1109/TNN.2003.811353

Driving is indeed a sophisticated process in which three main tasks are involved: navigation, guidance, and stabilization [2]. The performances of these tasks are subject to two major factors: the temperament of drivers and the technologies of vehicles. In this study, the former factor is of prime concern. There are two kinds of influences, exterior and interior, that will affect the behavior of a driver. The external influence comes from the driver's knowledge about the driving environment, while the internal influence comes from the driver's expectations. These two influences are intrinsically related; the better the information a driver receives, the more appropriate his expectations will be.

Various sensors (e.g., infrared, multispectral, photometric, range, ladar, and ultrasonic sensors) have been utilized to facilitate human acquisition of information from the environment. In this study, we consider visual sensors (e.g., cameras and video camcorders). Vision systems have been used in many applications for detecting, tracking, monitoring, inspecting, and recognizing objects. For the purpose of driver assistance, vision systems have been exploited to detect, track, and recognize objects such as roads [6], [7], [17], [20], [23], lane markings [4], [6], [13], [16], [22], [33], [38], traffic signs [14], [29], road conditions (e.g., dry, wet, fog, freezing, and snow) [1], [39], and obstacles (e.g., pedestrians, vehicles, motorcycles and other intruders) [4], [5], [13], [28]. In this paper, a vision system for detecting critical changes in driving environment is presented.

The objective of detecting changes in driving environment is threefold. First, since a number of subsystems constitute a driver assistance system, the subsystems should be coordinated in order to achieve optimal performance of the driver assistance system. The cooperation of the subsystems, including which, when and how they are to be conducted, often depends on conditions of driving environments. Secondly, parameters embedded in the subsystems should be updated in accordance with environmental changes, such as the changes in illumination, weather, road conditions, vehicle speed, etc. Finally, unexpected changes in driving environment are often related to critical traffic situations. Early warnings of these situations to drivers, especially those who drive cargo or tanker trucks, would be highly desirable.

Few researchers have discussed how to detect environmental changes while driving. A possible reason may be that "change" is difficult to define because everything is "moving" while driving. In this paper, we confine ourselves to changes concerning driving safety which may be encountered in daytime

or at night while driving on freeways. The changes under consideration include left-lane-change (moving to the lane on the left), right-lane-change (moving to the lane on the right), freeway entry, freeway exit, tunnel entry, tunnel exit, and overpass ahead.

To detect the above changes in driving environment, a computational framework motivated by human cognitive processing and selective attention is proposed. In this framework, both temporal and spatial information are extracted from input video sequences. The extracted information serves as input stimuli to a spatiotemporal attention (STA) neural network. If consistent stimuli keep innervating the neural network, a focus of attention will be established in the network. The attention pattern associated with the focus, together with the location and direction of motion of the pattern, form what we call a categorical feature. Thereafter, based on this feature the class of the attention pattern and, in turn, the change in driving environment corresponding to the class are determined using a configurable adaptive resonance theory (CART) neural network.

We describe the proposed computational model in Section II. The STA and CART neural modules are addressed in Section III. A system to detect change in driving environment based on the computational model is developed in Section IV. The feasibility of the proposed computational model and the robustness of the developed change detection system are detailed in Section V. Finally, concluding remarks and future work are given in Section VI.

II. Computational Model

In this section, the psychophysical fundamentals of cognitive processing are briefly reviewed, followed by a discussion of selective attention. A computational model implementing cognitive processing and selective attention is then presented. This model will later be utilized to develop a system to detect change in driving environment.

A. Fundamentals of Cognitive Processing

Pattern recognition tasks generally require considering many pieces of information simultaneously. Computers, which are basically sequential machines, are not adequate for such tasks. However, human brains with their extreme degree of parallelism seem to deal with these tasks effortlessly. The effectiveness of parallelism depends on knowledge representation schemes involved in information processing. The distributed representation scheme has been widely examined, and has revealed a number of appealing characteristics [18], such as constructivity (recalling subjects from their incomplete or imperfect contents), generalization (generalizing modifications to all related subject matters), and tunability (automatically adapting to changing circumstances). The distributed representation scheme, together with parallel processing, lead to what cognitive scientists call parallel distributed processing. There are diverse processings active both within and between layers of neurons, giving rise to different levels of information processing and analysis, such as sensory, perceptual, syntactic, semantic, episodic, and action analyzes. A cognitive process is accomplished by a series of information analyzers arranged in a hierarchical order.

While information analyzers may play different roles in cognitive processing, they possess analogous constructs [21], [24]. Every analyzer consists of several layers of neurons. Neurons on the same layer are laterally connected and their links are almost always inhibitory. Neurons on different layers are vertically connected and their links are typically excitatory. A vertical link indicates the existence of a particular part-whole relationship; the synaptic strength of the link specifies the degree of the relationship. A lateral inhibitory link signifies the existence of competing relationship between two components; the strength of the link describes the degree of the competitive relationship. Both links and synaptic strengths are established through learning.

B. Selective Attention

According to Titchener [34], consciousness can be divided into center and periphery, or focus and margin. Attention is directed to the center and the focus. The question arises as to how focuses and centers of attention arise in consciousness. Everyone knows that we cannot attend to too many things at once. Indeed, it is not easy having more than one unless things or events are rather familiar. This observation suggests that there are selecting or filtering mechanisms at work in consciousness in which centers and focuses of attention are established.

Two types of selectivity of attention have been proposed [25]: involuntary (automatic) and voluntary (effortful) selectivities. The former originates from sudden and unexpected events, while the latter results from forcing attention to uninteresting things. For developing a target detection system, the latter paradigm will not be adequate for it forces the system attention to uninteresting objects rather than those of interest. In the following, we consider involuntary selectivity. For this, two principles have been introduced: filtering and amplification. The former states that unwanted things are gradually attenuated until they are filtered out from the center of attention, and the latter states that the selected events are progressively augmented and intensified until they seize the focus of attention. Both principles are readily realized with neural networks in which focuses of attention correspond to the subsets of strongly activated neurons.

A number of models of selective attention are rooted in the aforementioned two principles, including Broadbent's early-selection model, Norman's late-selection model and Treisman's attenuation model. Broadbent argued that attentional selection appears only in the sensory analyzer where only selected information is passed on for further processing. In his model, information that comes through the sensory channel is selected on the basis of sensory attributes (e.g., intensity, color, texture, frequency, orientation and location). However, Norman argued that attentional selection can also take place in the semantic analyzer. In her model, attentional selection is the outcome of two input sources of stimuli: bottom-up sensory stimuli and top-down stimuli from a pertinence mechanism. This mechanism and, in turn, the Norman model illustrate both the prime effect of expectation and the effect of habituation by allowing some neurons to be activated even though they have lower activations than others. Unfortunately, the above two models cannot account for several issues regarding the effects of meaning on shadowing performance previously designed by Cherry [9] and

Spieth *et al.* [31]. In order to compensate for the shortcomings of the above two models, Treisman proposed an attenuation model in which signals are attenuated by a series of selective filters located in the sensory, perceptual, syntactic and semantic analyzers. In other words, Treisman agreed that attentional selection occurs in all analyzers. The set of neurons will form the focus of attention after all the attenuation in the analyzers.

Vision researchers have presented several models concerning visual attention [3], such as guided search [37], variable powered lens [19], spotlight [12], feature-based models [15], [26], and connectionist models [35]. These models are introduced essentially for active vision systems. Such systems include a number of distinguishing features, such as controlled eye movement, gaze orientation, foveation, selective visual feedback, as well as active sensing. In the content of active vision, visual attention has been categorized into two classes, overt and covert attentions [10]. The former externally directs the sensor toward the spatial region of interest in a scene, while the latter selects or filters visual areas from an internal representation of the scene. While our vision system is mounted in a moving vehicle, the sensing device of the system is stationary with respect to the vehicle. Therefore, only the covert attention is implemented in our system.

C. Cognitive Model

Fig. 1 depicts the proposed computational model, which captures several important aspects of cognitive processing and selective attention as addressed in the previous subsections. The model is comprised of three components, the sensory, perceptual, and conceptual analyzers. The input to the model are video sequences. Rather than detect and recognize target objects in each input image, the temporal information of moving objects is first extracted from the video sequence. However, everything appears to be moving while driving, so how can the objects of interest be attended to among a jumble of moving objects? The spatial information (e.g., shapes, colors, textures, orientation, and location) of objects then plays an important role in distinguishing among objects. In the sensory analyzer, we actually extract temporal information first and the spatial information. This is not only because, in practice, extraction of temporal information is easier than extraction of spatial information, but also because it is compatible with the fact that motion is generally analyzed and perceived earlier than form and meaning in the human cognitive system [25], [40]. The sensory process forms the early-selection stage of our computational model.

The information acquired in the sensory analyzer serves as the stimulus to a neural module, called the STA neural module, in the perceptual analyzer. The neural module implements the involuntary selectivity of attention under the guidance of the information provided by a long-term memory (LTM) in which the spatial information of target objects is preserved. The activations of the STA neurons are examined regularly during processing. If there is no focus of attention present in the network, the system repeats the above process. Otherwise, feature extraction is initiated. The categorical feature of the object appearing in the image area corresponding to the focus of attention is detected.

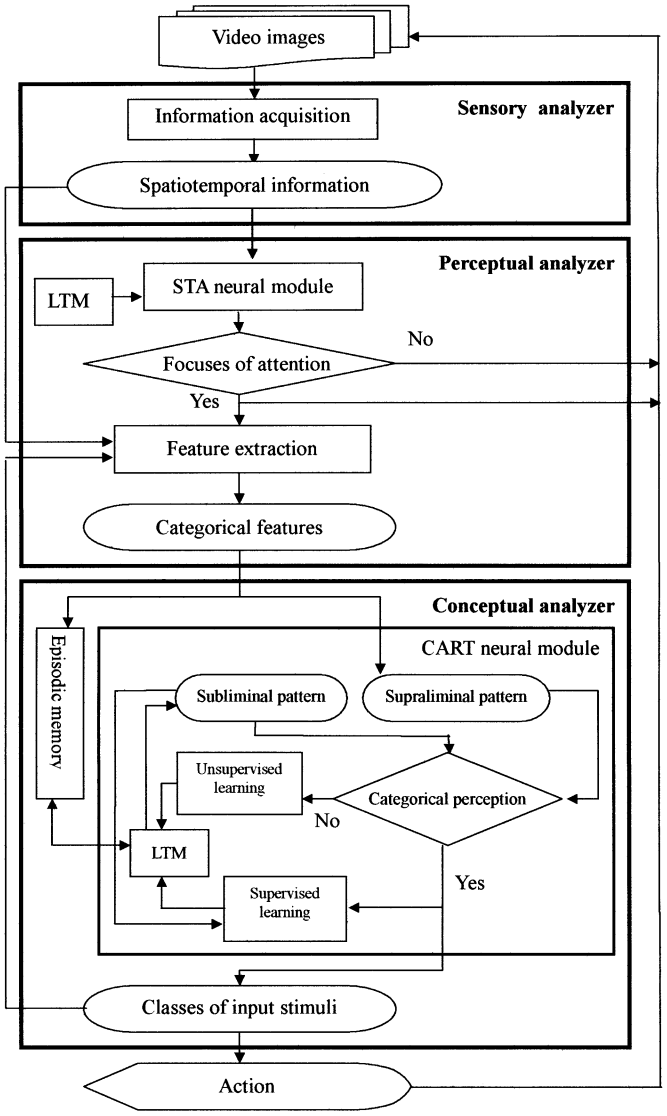


Fig. 1. Proposed computational model.

The categorical feature, represented as a one-dimensional (1-D) pattern, called a supraliminal pattern, acts as the input stimulus to a CART neural network in the conceptual analyzer. With the input supraliminal pattern, the LTM trace of the CART neural module is initialized with content (that is, a collection of subliminal examplars associated with the input supraliminal pattern) retrived from a system memory, called the episodic memory. During initialization, the configuration of the CART neural module is adjusted to fit the loaded content. We refer to the ability to adjust configuration as the configurability of the neural module. After the initialization step, the recognition of the input pattern is accomplished by comparing the input supraliminal pattern with the subliminal examplars residing in the LTM trace of the neural module. If a subliminal examplar is similar enough to the input pattern, its class is taken as that of the examplar being considered. The CART neural module next performs supervised learning to incorporate the supraliminal pattern in the subliminal examplar under consideration. On the other hand, if no subliminal examplar is similar to the input pattern, an unsupervised learning memorizes the input

supraliminal pattern as a new subliminal examplar in the LTM trace.

Thereafter, the system goes to the action stage in which a number of things are handled. These include outputting the result, updating the episodic memory, restoring the CART neural module, and inhibiting the focus of attention being considered in the STA neural module. After these, the system repeats the entire process.

III. Neural Modules

There are two neural modules which play important roles in our computational model: the STA neural module in the perceptual analyzer and the CART neural module in the conceptual analyzer. This section addresses these two modules.

A. STA Neural Network

The STA neural network, as Fig. 2 shows, is configured as a two-layer network with one input layer and one output layer. The output layer is also referred to as the attention layer. Neurons in this layer are arranged into a two–dimensional (2-D) array in which they are interconnected. These within-layer (lateral) connections are almost always inhibitory. There are no synaptic links among input neurons though they are organized into a 2-D array, too, as the attention neurons. They are fully connected to the attention neurons. These connections are called between-layer (vertical) connections and are always excitatory.

Let the size of both the arrays be the same as that of the input images. Let w_{ij} denote the weight of the link between attention neuron n_i and input neuron n_j. The weight vector of attention neuron n_i is denoted as $\mathbf{w}_i = (w_{i1}, w_{i2}, \ldots, w_{im})$, where m is the number of input neurons. The input to attention neuron n_i due to input stimuli $\mathbf{x}$ is

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

The linking weights, w_{ij}, between the input layer and the attention layer are defined as follows. Referring to Fig. 3, let n_j be any input neuron and n_i be its corresponding neuron on the attention layer. Assume that a 2-D Gaussian G is centered at attention neuron n_i. The weight w_{kj} linking the input neuron n_j with the attention neuron n_k is defined as $w_{kj} = G(\mathbf{r}_{ki})$, where $\mathbf{r}_{ki}$ is the position vector of attention neuron n_k with respect to attention neuron n_i.

The lateral interaction among attention neurons is characterized by a "Mexican-hat" function, denoted by $M(\mathbf{r})$ shown in Fig. 6, where $\mathbf{r}$ is a position vector originating from the center of the function. The function plays an important role in clustering activations of neurons. It is often approximated by a Laplacian of Gaussian (LOG), $\nabla^2 G(\mathbf{r})$, or a difference of Gaussians (DOG), $G_1(\mathbf{r}) - G_2(\mathbf{r})$. The input to attention neuron n_i due to lateral interaction is defined as

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

where N_i is the neighbors of neuron n_i, and u_{ik} is the weight linking neurons n_i with n_k, which have position vectors $\mathbf{r}_i$ and $\mathbf{r}_k$, respectively. Here, we assume that $u_{ik} = 1$ for all i and k. a_k is the activation of attention neuron n_k.

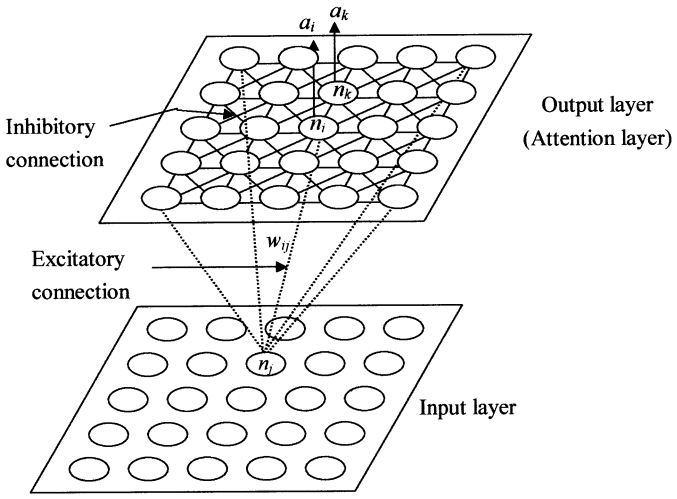


Fig. 2. STA neural network.

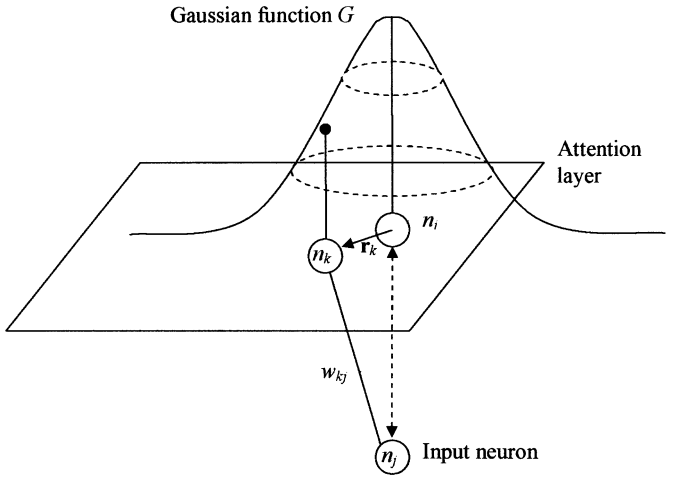


Fig. 3. Weights between input layer and attention layer.

Let the activation of attention neuron n_i be governed by

$$\dot{a}_i = A(-pa_i + qB(\mathrm{net}_i)) \tag{3}$$

where p and q are positive constants; p specifies the decay rate of a_i, and q weights the net inputs $\mathrm{net}_i = I_i^v + I_i^l - \Gamma$ in which Γ is a threshold to limit the effects of noise. Functions $A(\cdot)$ and $B(\cdot)$ are defined as

$$A(x) = \begin{cases} x, & \text{if } x > 0 \\ dx, & \text{if } x \leq 0 \end{cases} \tag{4}$$

$$B(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases} \tag{5}$$

in which $1 > d > 0$. Function $A(\cdot)$ has been called an attack function, and causes different rise and decay times for neural activations.

Considering the case in which attention neuron n_i receives a positive net input, i.e., $\mathrm{net}_i > 0$, from (5), we know $B(\mathrm{net}_i) = \mathrm{net}_i$. Substituting this result into (3), we get $\dot{a}_i = A(-pa_i + q \cdot \mathrm{net}_i)$. Suppose now that $-pa_i + q \cdot \mathrm{net}_i > 0$. According to

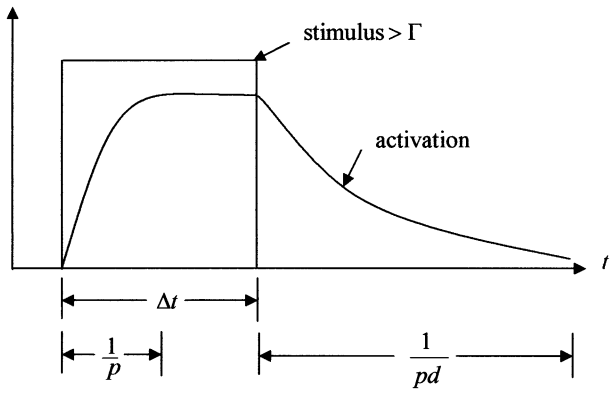


Fig. 4. Activation of an attention neuron in response to a stimulus.

(4), $\dot{a}_i = -pa_i + q \cdot \text{net}_i$. Assuming that the initial activation of neuron n_i is zero, and that the input net_i lasts from time zero to time t, the solution of the above differential equation is

$$a_i(t) = \frac{q}{p} \cdot \text{net}_i \cdot (1 - \exp(-pt)). \tag{6}$$

However, if $-pa_i + q \cdot \text{net}_i \leq 0$, then from (4) $A(-pa_i + q \cdot \text{net}_i) = d(-pa_i + q \cdot \text{net}_i)$. Substituting this result into (3), we obtain $\dot{a}_i = d(-pa_i + q \cdot \text{net}_i)$. Solving this equation for a_i, we get

$$a_i(t) = \frac{q}{p} \cdot \text{net}_i \cdot (1 - \exp(-dpt)). \tag{7}$$

The difference between (6) and (7) is in the lack of constant d in (6). Since $0 < d < 1$, the rate of rise for (7) is smaller than that for (6). Next, considering the case where the input is removed at time t', the net input to neuron n_i becomes negative, i.e., $\text{net}_i < 0$. Following the same analysis as before, we conclude that

$$a_i(t) = \frac{q}{p} \cdot \text{net}_i \cdot \exp(-p(t - t')). \tag{8}$$

Refer to Fig. 4, which shows the activation of an attention neuron in response to an input stimulus. If the net input to the neuron is greater than threshold Γ within a time interval Δt, the neuron needs about time $1/p$ to reach maximum activation, and takes about time $(1/p)d$ to decay. Since $d < 1$, the decay time is longer than the rise time.

B. CART Neural Network

The CART neural module is an ART2 neural network [8] with a configurable long-term memory (CLTM). Fig. 5 depicts the ART2 neural network. There are two subsystems comprising the network, an attentional subsystem and an orienting subsystem. The attentional subsystem is composed of two fields, an input representation field F_1 and a category representation field F_2. The CLTM is embedded in the bidirectional links between F_1 and F_2 and is reconfigurable in our neural module. F_2 contains only one layer, denoted by $\mathbf{y}$, which serves as a competitive layer. F_1 consists of six layers, denoted by $\mathbf{w}$, $\mathbf{x}$, $\mathbf{u}$, $\mathbf{v}$, $\mathbf{p}$, and $\mathbf{q}$. The orienting subsystem is composed of two components, a layer $\mathbf{r}$ and a signal generator S. Layer $\mathbf{r}$ integrates the activities from layers $\mathbf{p}$ and $\mathbf{u}$ in F_1 and sends the result to the signal generator S. It determines whether or not a reset signal is emitted to layer $\mathbf{y}$ in F_2.

Let i be an input pattern. It propagates back and forth within the F_1 field until all its layers stablize. The activities of the layers within F_1 are

$$\begin{aligned} w_i &= i_i + au_i, & x_i &= \frac{w_i}{e + \|\mathbf{w}\|} \\ v_i &= f(x_i) + bf(q_i), & u_i &= \frac{v_i}{e + \|\mathbf{v}\|} \\ p_i &= u_i + \sum_j g(y_j) z_{ji}, & q_i &= \frac{p_i}{e + \|\mathbf{p}\|} \end{aligned} \tag{9}$$

where a and b are positive constants, and e is a small positive value. Function $f(\cdot)$ here performs a contrast enhancement to the input pattern and is defined by

$$f(x) = \begin{cases} 0, & \text{if } 0 \leq x \leq \theta \\ x, & \text{if } x > \theta \end{cases} \tag{10}$$

in which θ is a positive constant less than one. Function g is the transfer function of $\mathbf{y}$ neurons and is given by

$$g(y_i) = \begin{cases} d, & T_J = \max\limits_k \{T_k\} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

where J indicates the winner on layer $\mathbf{y}$ and d is a constant between zero and one. T_k is the net input from layer $\mathbf{p}$ to the kth $\mathbf{y}$ neuron, $T_k = \sum_i p_i z_{ik}$, in which z_{ik} is the weight from the ith $\mathbf{p}$ neuron to the kth $\mathbf{y}$ neuron.

Once the layers in F_1 have stablized, the activity of layer $\mathbf{p}$ is transmitted to layer $\mathbf{y}$ in F_2. The neurons on layer $\mathbf{y}$ compete with one another for the signal. If no neuron wins the competition, some uncommitted neuron on layer $\mathbf{y}$ will be selected for encoding the input pattern as a new prototype through unsupervised learning. Learning takes place by revising both the weights of the top-down and bottom-up connections between the uncommitted neuron and layer $\mathbf{p}$. Let j denote the uncommitted neuron which is selected. Its bottom–up z_{ij} and top-down z_{ji} weights are updated according to

$$z_{ij} = z_{ji} = \frac{u_i}{1 - d}. \tag{12}$$

On the other hand, if one of the $\mathbf{y}$ neurons wins the competition, the neuron initiates the corresponding prototype encoded on the top-down connections. The prototype and the input patterns are matched with each other within F_1. Matching proceeds until all layers in F_1 calm down. The activities of layers $\mathbf{u}$ and $\mathbf{p}$ are then forwarded to layer $\mathbf{r}$ in which the activities of $\mathbf{u}$ and $\mathbf{p}$ are integrated according to

$$r_i = \frac{u_i + cp_i}{e + \|\mathbf{u}\| + \|c\mathbf{p}\|} \tag{13}$$

where c is a constant subject to the constraint that $cd/(1-d) \leq 1$. The condition for triggering a reset signal is determined by $\rho/(e + \|\mathbf{r}\|) \leq 1$, where ρ is called the vigilance parameter. If a reset signal is emitted, the currently activated neuron on $\mathbf{y}$ is inhibited and the process is repeated; otherwise, the input pattern is regarded as being in the category represented by the neuron.

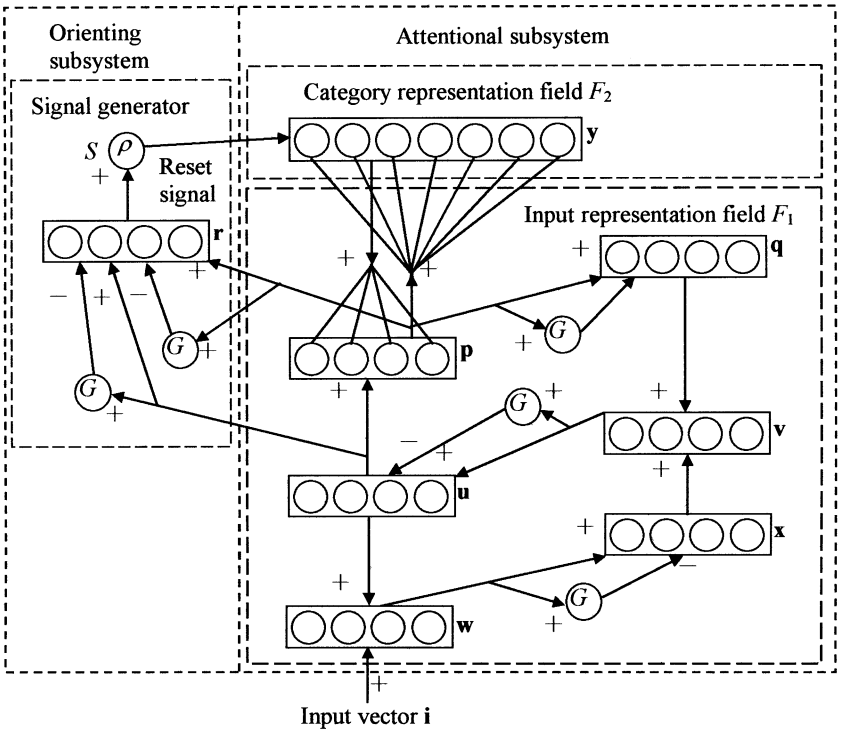


Fig. 5. Architecture of the ART2 neural network.

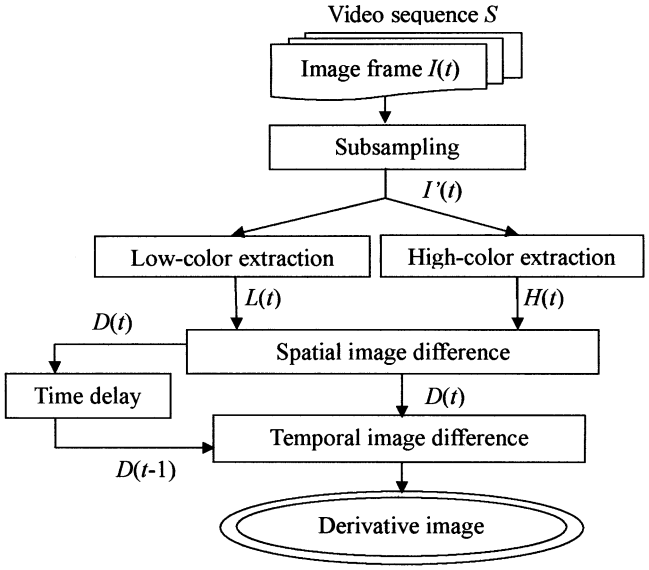


Fig. 6. Flowchart of information acquisition.

IV. System to Detect Change in Driving Environment

A system based on the presented computational model for detecting changes in driving environment is developed in this section. The system consists of three components, referred to as the sensory, perceptual, and conceptual components, corresponding to the three analyzers of the proposed computational model.

A. Sensory Component

The input data to the system are color video sequences acquired using a camcorder mounted in the windshield of a moving vehicle. In the sensory component, temporal and spatial information of dynamic scenes is extracted from the input video sequences. The input video sequences are typically unstable because of movements and vibrations of the vehicle. Image stabilization techniques [11] are commonly used to attenuate the effect of nonsmooth motions of vehicles in video sequences. However, such techniques are generally time expensive and are not adequate for real-time applications. We thus turn to another approach to compensating for image instability.

Fig. 6 depicts the flowchart of information acquisition from a color video sequence S. Let $I(t)$ denote its tth image frame. First, $I(t)$ is subsampled every two pixels in order to reduce the processing time. Let $I'(t)$ specify the subsampled image and $\mathbf{D}$ be its 2-D domain manifold. Suppose $R'(t)$, $G'(t)$, and $B'(t)$ are the color components of $I'(t)$, i.e., $I'(t) = (R'(t), G'(t), B'(t))$. A low-color image, $L(t) = (R^l(t), G^l(t), B^l(t))$, is then computed by the following. For all $(i,j) \in \mathbf{D}$

$$R^l_{(i,j)}(t) = \min\{R'_{(i,j)}(t), R^l_{(i,j)}(t-1)\}$$
$$G^l_{(i,j)}(t) = \min\{G'_{(i,j)}(t), G^l_{(i,j)}(t-1)\}$$
$$B^l_{(i,j)}(t) = \min\{B'_{(i,j)}(t), B^l_{(i,j)}(t-1)\}.$$

Likewise, a high-color image, $H(t) = (R^h(t), G^h(t), B^h(t))$, is computed by

$$R^h_{(i,j)}(t) = \max\{R'_{(i,j)}(t), R^h_{(i,j)}(t-1)\}$$
$$G^h_{(i,j)}(t) = \max\{G'_{(i,j)}(t), G^h_{(i,j)}(t-1)\}$$
$$B^h_{(i,j)}(t) = \max\{B'_{(i,j)}(t), B^h_{(i,j)}(t-1)\}.$$

Here, $L(0) = H(0) = I'(0)$. Suppose that $I^h(t)$ and $I^l(t)$ are the intensity components of color images $H(t)$ and $L(t)$, respectively. The intensity value $I_{(i,j)}$ of a pixel (i,j) with

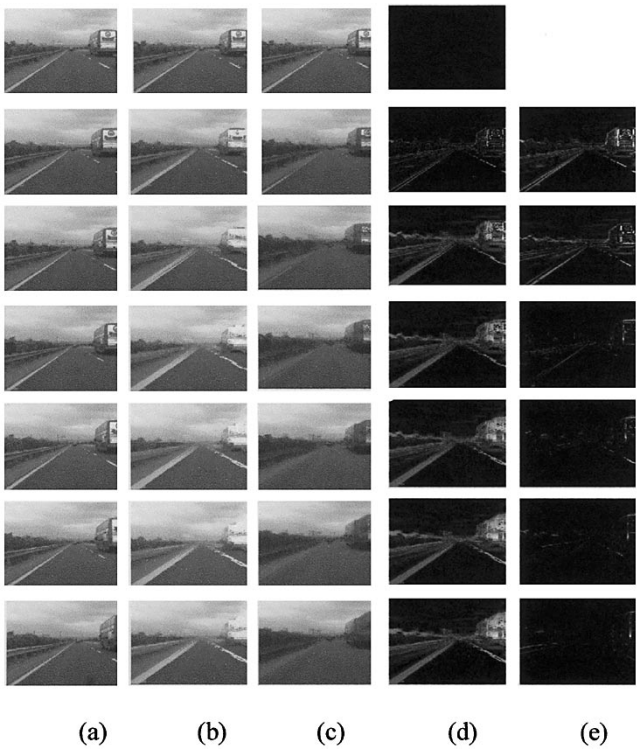


Fig. 7. Example illustrating the steps of the sensory component.

$(R_{(i,j)}, G_{(i,j)}, B_{(i,j)})$ color values is determined simply by $I_{(i,j)} = (R_{(i,j)} + G_{(i,j)} + B_{(i,j)})/3$. A spatial difference image $D(t)$ is then computed by $D_{(i,j)}(t) = I^h_{(i,j)}(t) - I^l_{(i,j)}(t)$ for all $(i,j) \in \mathbf{D}$. The difference image will highlight objects which are relatively stationary with respect to the viewer.

See the example shown in Fig. 7. The first column of the figure displays a portion of the input video sequence. In this example, the observing car was driven in the left-most lane and a bus was traveling in the next lane to the right. The second and third columns of the figure display the high-color and low-color image sequences, respectively. A high-color $H(t)$ (low-color $L(t)$) image at time t preserves the maximum (minimum) color values of the input video sequence up to time t. In the fourth column of the figure, the difference images computed from the high-color and low-color images are displayed. In this sequence, the objects having relatively small movements with respect to the viewer are highlighted.

Having computed spatial difference images $D(t)$, we calculate temporal difference (derivative) images $D'(t)$ from $D'_{(i,j)}(t) = \left|D_{(i,j)}(t) - D_{(i,j)}(t-1)\right|$ for all $(i,j) \in \mathbf{D}$. Turning to the previous example of Fig. 7, the sequence of derivative images are depicted in the last column of the figure. In this sequence, the white regions in the first few images graduately diminish with time. This is because the highlighted areas in the difference images will cancel each other out during subtractions of successive difference images. The weakened white regions in the derivative images which originate from image instability will not be able to activate the neural networks in the subsequent system components. We use this method to effectively get rid of the influence of image instability.

Recall that the highlighted regions in the difference images correspond to objects which are relatively stationary with respect to the viewer. Such objects can only produce small entries in the derivative images. See Fig. 8 for another example of right-lane-change, in which the input video sequence is shown in the top row and the resultant derivative images are shown in the middle row. In this example, the observing car is moving from the left to the right lane so that lane markings keep moving from right to left in the input video sequence. This creates distinct white regions in the derivative images. Note here that the sequence of derivative images preserves both spatial information (e.g., shape and location) of objects and their temporal information (e.g., direction of motion and speed).

Our sensory component actually simulates the vision systems of some lower animals, e.g., frogs and snakes. In their visual systems, stationary objects are invisible, while both shapes and motions of moving objects are easily perceived.

B. Perceptual Component

The derivative images resulting from the sensory component serve as input stimuli to the STA neural module in the perceptual component. If the input derivative images contain only small values (see the example in Fig. 7) or include large values which are randomly distributed across the sequence of images, significant attention patterns (patterns of neural activations over the attention layer) will not be activated. In practice, we examine each derivative image before feeding it to the STA neural module. Those derivative images containing small values are ignored so as to save processing time.

A focus of attention can be established only if an attention pattern becomes strong enough. Refer to the previous example of right-lane-change illustrated in Fig. 8. The attention maps (i.e., snapshots of neural activations over the attention layer) are displayed in the bottom row. At the beginning, a very weak pattern near the lower right corner of the attention map is created when the vehicle starts changing lane. The pattern gets larger and stronger as the process of lane change continues. Finally, the pattern reaches its maximum at which time the vehicle completes its lane change. Subsequently, if the vehicle stays in the current lane, the pattern slowly decays.

Different changes in driving environment will generate different attention patterns. Fig. 9 depicts the prototypical attention patterns under consideration, each corresponding to a specific change in driving environment. It is common, however, that patterns extracted from attention maps may not normally be so perfect. We have to depend on other features for achieving reliable performance. Hence, two additional features are introduced, locations and directions of motion of attention patterns. As can be seen in Fig. 9, patterns may appear at different locations in an attention map. Since precise locations of patterns are difficult to determine, we qualitatively define five types of locations (labeled 0, 1, 2, 3, 4) in Fig. 10(a). In a similar vein, five types of motion directions (labeled 0–4) are defined in Fig. 10(b). The pattern, its location and direction of motion form a feature vector, referred to as a categorical feature, to be used in the conceptual component.

C. Conceptual Component

The categorical features received from the perceptual component consist of both the qualitative (location and direction of mo-

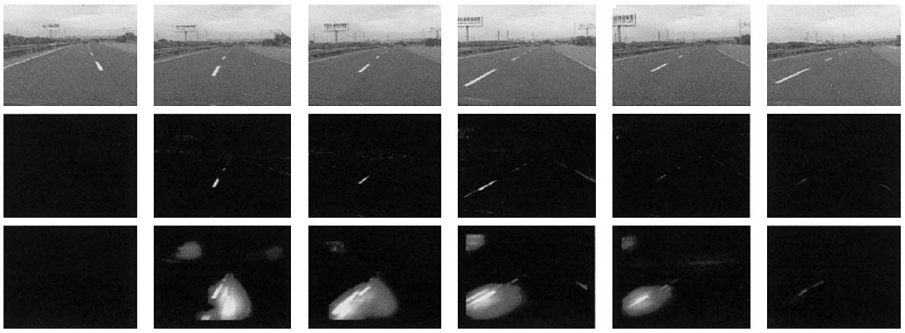


Fig. 8. Example of right-lane-change.

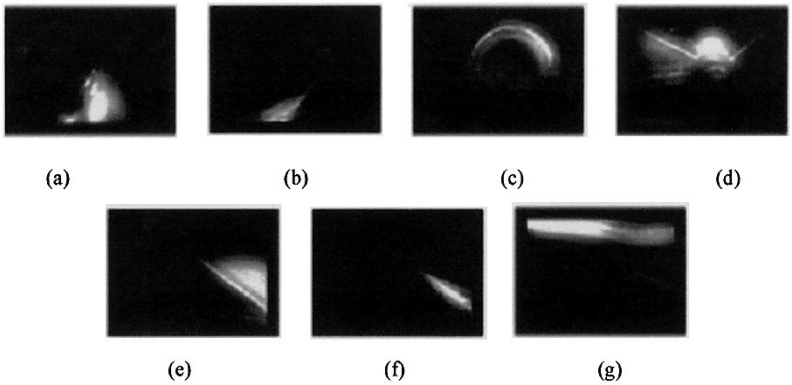


Fig. 9. Prototypical attention patterns for the driving environmental changes under consideration.

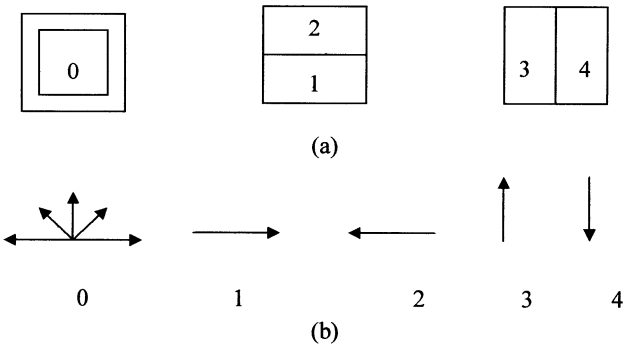


Fig. 10. Qualitative types of location and direction of motion.

tion) and quantitative (attention pattern) information of moving objects. We use the qualitative information of location and direction of motion to quickly weed out inappropriate prototypical patterns (Fig. 9) from further processing. The surviving prototypical patterns are then retrieved from the episodic memory of the system and are installed in the LTM trace of the CART neural module. In this neural module, the input attention pattern is matched with the prototypical patterns. If the input pattern is recognized as one of the prototypical patterns, the class of the prototypical pattern is temporarily preserved. Only when three successive results turn out to be the same is the final outcome confirmed and reported through a visual interface (see Fig. 11) as well as an audio speaker (not shown), which are both attached to the same computer. In this case, the selected prototypical pattern is updated by means of supervised learning. However, if the input pattern cannot find a match from among the prototypical patterns, the input pattern is retained through unsupervised learning. The retained pattern will later be discarded or be regarded as a new prototypical pattern by an off-line method. In the latter case, the class of the new prototypical pattern is assigned by hand.

V. Experiments

The input data to our system was acquired using a video recorder mounted in the windshield of a vehicle and while driving on freeways. Each video sequence was downsampled to a frame rate of 5 Hz to reduce the processing load on the computer. This frame rate is also fast enough for a driver to respond to any of the environmental changes considered in this paper. Furthermore, the size of each input image (320×240 pixels) was reduced to 160×120 pixels by uniformly subsampling so as to further reduce the processing time.

A number of video sequences have been collected for our experiments. The sequences are categorized into seven classes, referred to as the right-lane-change, left-lane-change, tunnel entry, tunnel exit, freeway entry, freeway exit, and overpass ahead. Each class is further divided into two groups, termed the "day" and "night" groups. Since the cases of lane change have already been discussed in the previous sections, in this section we discuss the other cases of environmental change.

A. Experimental Results

Fig. 12 presents the experimental results for tunnel entry, tunnel exit and overpass ahead, and Fig. 13 gives the results for freeway entry, freeway exit, and right-lane-change at night. In each example, only a portion of the input video sequence and

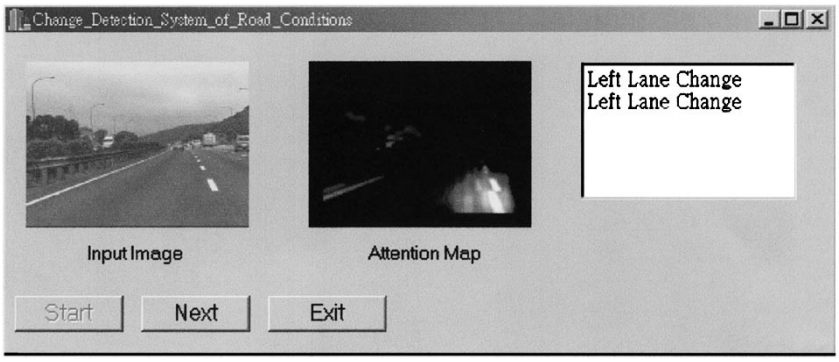


Fig. 11. Visual interface for reporting the results of change detection in driving environment.

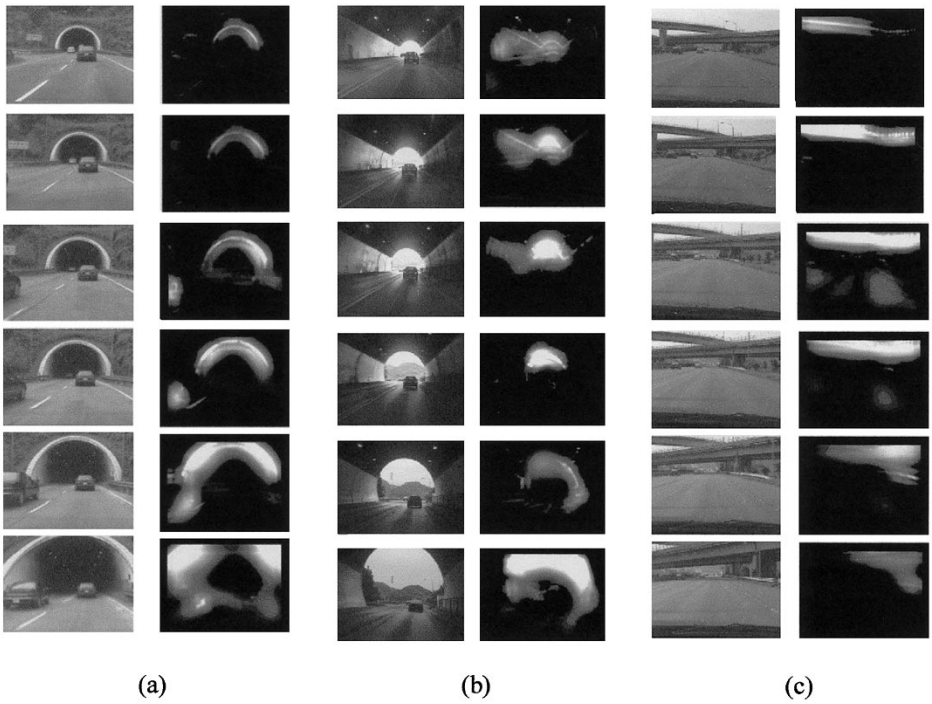


Fig. 12. Experimental results for tunnel entry, tunnel exit, and overpass ahead.

the associated attention maps are shown. In all cases, recognition results are reported through a visual interface (see Fig. 11) and an audio speaker.

Refer to the example of tunnel entry shown in Fig. 12(a). In the input video sequence, the figure of the tunnel entry has a relatively dark interior surrounded by a ring-like white border. Such a figure has caused a rainbow-like pattern on the attention map. The pattern became larger and brighter as the vehicle approached the tunnel. Once the pattern becomes well developed, its current location and historical direction of motion form a categorical feature. Based on this feature, a tentative decision on the type of environmental change is made by the system. Only when three successive decisions are all the same, the final decision (i.e., the type of environmental change) is reported.

Fig. 12(b) shows an example of tunnel exit, which has a picture nearly opposite to the tunnel entry, that is, a bright interior surrounded by a large portion of dark area. Such a figure creates a sunrise-like pattern on the attention map. Our system successfully discriminated between tunnel entries and exits based on their attention patterns. This is applicable for both the day and night groups. However, the figures and attention patterns of tunnel entries in daytime are similar to those of tunnel exits at night, and vice versa. Therefore, we need strategies to distinguish between the cases of day and night. A heuristic is introduced which states that a tunnel exit should come after a tunnel entry. A more reliable method may be to incorporate an illumination assessment technique in the system.

Next, see the example of overpass ahead shown in Fig. 12(c). This class of environmental change can easily be identified based on its particular locations of attention patterns. Referring to Fig. 9, the prototypical attention patterns can actually be divided into four groups based on the locations of the patterns. For an input pattern, our system quickly narrows down to a small set of candidate prototypical patterns for testing according to the location of the input pattern. The direction of motion attribute is utilized whenever the candidate set contains more than one pattern. In general, less than three patterns are retained in the candidate set for examination. The categorical

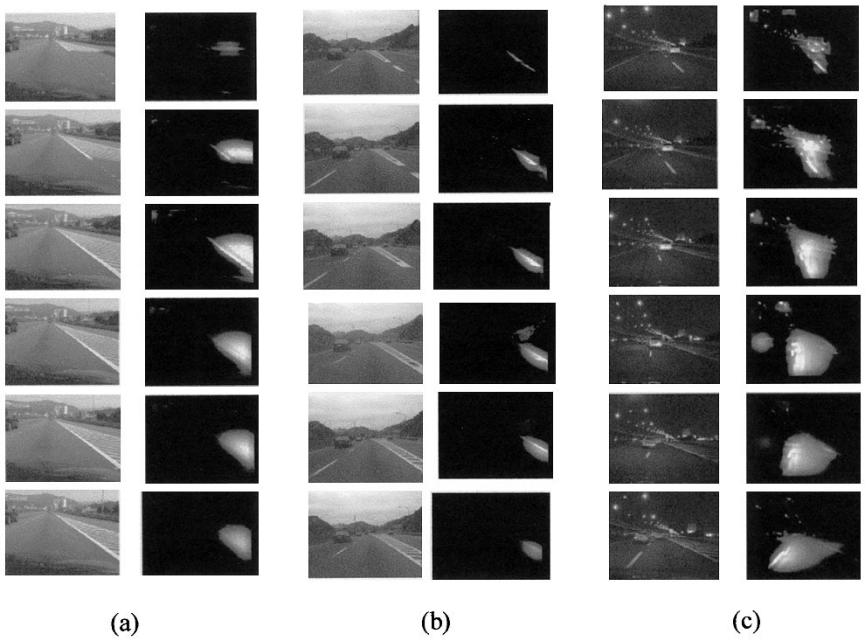


Fig. 13. Experimental results for freeway entry, freeway exit, and right-lane-change at night.

Fig. 14. Lane change on a curved road.

features of overpasses extracted in daytime and at night are almost the same. However, overpasses will only be detected at night when the side of an overpass is illuminated by outdoor lighting because the light from vehicles' headlights do not reach overpasses.

Turn to Fig. 13(a) and (b) for examples of freeway entry and exit. This pair of environmental changes are easily separated from the others based on the attribute of location. However, unlike right and left lane changes, freeway entry and exit cannot be distinguished based on their directions of motion. A pattern test is required to distinguish between them. Referring again to Fig. 9, the attention patterns for freeway entry and exit are different in both size and distribution of neural activation. The pattern for entry has a larger size than that for exit. Furthermore, the ridge of distribution of activation appears around the lower border for the entry pattern while it appears around the medial axis for the exit pattern. The CART neural module has discriminated these two types of patterns reliably. The attention patterns for entry and exit generated at night are similar to those produced in daytime.

See the example of right-lane-change at night shown in Fig. 13(c). Compare the attention patterns of this example with those depicted in the bottom row of Fig. 8 which were generated by a right-lane-change in daytime. Ignoring noisy patterns, these two groups of patterns are visually comparable to each other. Experimental results also showed the consistency between the two groups of attention patterns in their attributes of location and direction of motion. Indeed, we have observed that a better performance was achieved during night than during day because many background objects are invisible at night.

B. Discussion

Here we discuss curved roads, varying illumination conditions, shadows, passing cars, rain, snow, faded lane markings, and multiple environmental changes. Referring to Fig. 14, our system did detect a right-lane-change (see the patterns at the bottoms of the attention maps) on a curved road. Recall that the frame rate of each input video sequence has been reduced to 5 Hz, so the time interval between successive input images is 0.2 s. During such a period, a vehicle moving at 60 mi/h will travel a distance of about 0.0033 mi. Even on a sharply curved road, lane markings, including those near the vehicle, could shift only a few pixels between successive images. Such a shift is, in practice, comparable to those resulting from image instability.

Our system seemed to manage different illumination conditions (e.g., sunny, cloudy, foggy, and dusty days) well. Recall that spatial difference images are computed from the input video sequence at an early stage of processing. These images encode contrasts of brightness in scenes. Different illumination conditions actually lead to different degrees of contrast for spatial

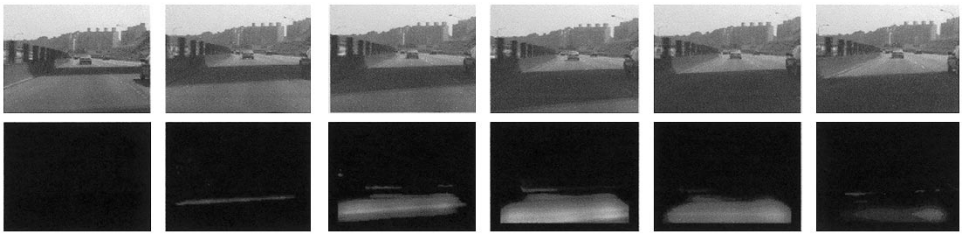


Fig. 15. Shadow cast by an overpass.

difference images. The degrees are also preserved in the derivative images which are obtained from the temporal difference between successive spatial difference images. The STA neurons receiving derivative images as stimuli will be activated if consistent stimuli (i.e., spatiotemporal contrasts of brightness) repeatedly innervate the neurons. Accordingly, low contrast will take longer than high contrast to activate the STA neurons.

Shadows do not degrade the performance of the system in detecting environmental changes. Considering an example of lane change, the amount of contrast between lane markings and pavements within the shadow is different from that outside the shadow. As already mentioned, different contrast affects only the detection time, not the result. However, shadows themselves can actually affect the attention of the system. See the example shown in Fig. 15. A shadow cast by an overpass is present in the video sequence. Since there is a relative motion between the shadow and the moving car, the shadow produces a series of activation patterns over the attention maps (see the bottom row of Fig. 15). Compare this series of patterns with that of Fig. 12(c) which was generated by an overpass. The patterns in these two groups look similar but our system distinguished between them based on locations of patterns. We ignore shadows because they do not influence driving safety. Likewise, cars in front of the observing car with speeds different from the observing vehicle's will attract the system's attention too. We leave shadows, passing cars, and several others (e.g., trees, road signs, construction cones, traffic signals, etc.) to the study of obstacle detection.

When raining or snowing, the regular movement of the windshield wiper can attract the attention of the system. It may not be a good solution to simply move the camera in front of the wiper because water or snow will accumulate on the camera lens. However, if water or snow cause only blurring of the image, our system may be able to overcome this because only regular stimuli can activate the STA neurons. On the other hand, since the movements of wipers are extremely regular, we may get rid of their effects based on the unique prototypical pattern generated by wipers. Similarly, if worn lane markings still produce steady stimuli for the STA neural module, our system will not be hampered by them. Finally, if multiple environmental changes occur simultaneously, the activated attention patterns will get mixed up. Our current system cannot handle such situations.

VI. Conclusion and Future Work

A computational model motivated by cognitive processing and selective attention was proposed. The current model consists of three analyzers, each dedicated to a specific task. Additional analyzers may be introduced to increase the ability of the model to manage sophisticated tasks. Furthermore, individual analyzers can also be reinforced in both function, to increase the robustness of the model, and configuration, to enhance efficiency of processing.

Based on the proposed computational model, a system for detecting changes in driving environment was developed. The system is composed of three components, referred to as the sensory, perceptual, and conceptual components, corresponding to the three analyzers of the cognitive model. The first component simulates visual systems of lower animals. The second realizes an involuntary selectivity of attention. The third implements a model-based recognition process.

The current system can handle seven types of change in driving environment both in daytime and at night, using the attributes of attention pattern and its location and direction of motion, which together constitute a categorical feature. In principle, the system is readily extended to manage additional kinds of environmental changes by introducing the associated categorical features into the system. In the mean time, the set of attributes comprising a categorical feature should also be reconsidered. A right set of attributes increases the ability of the system to discriminate among a large set of possible environmental changes. However, a large set of recognizable environmental changes increases the chance of a simultaneous occurrence of multiple changes. Our system is not yet at the stage of development where it can deal with that. This and several mentioned above as well as in the last section will form the topics for further research. In addition, several subsystems (e.g., road sign, obstacle, traffic condition detection systems, etc.) of driver assistance systems should be build on expanded computational models.

References

[1] T. Bachmann, Hoppstock, and K. Naab, "On-board detection of friction between type and road as an example of a driver assistance system," presented at the 6th Intelligent Transport Systems World Congr., Toronto, ON, Canada, 1999, Paper 2141.

[2] T. Bachmann and S. Bujnoch. (2001) CONNECTED DRIVE–Driver Assistance Systems of the Future. [Online]. Available: http://195.30.248.73/mercatorpark/pdf/connectedDrive.pdf

[3] G. Backer, B. Mertsching, and M. Bollmann, "Data- and model-driven gaze control for an active-vision system," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 1415–1429, Dec. 2001.

[4] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Trans. Image Processing*, vol. 7, pp. 62–81, Jan. 1998.

[5] M. Bertozzi, A. Fascioli, and A. Broggi, "Performance analysis of a low-cost solution to vision-based obstacle detection," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, 1999, pp. 350–355.

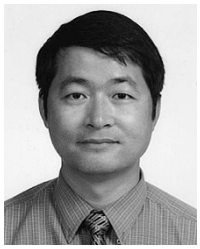
[6] A. Broggi, "Robust real-time lane and road detection in critical shadow condition," in *Proc. IEEE Int. Symp. Computer Vision*, Nov. 1995, pp. 353–358.
[7] A. Broggi and S. Berte, "Vision-based road detection in automotive systems: A real-time expectation-driven approach," *J. Artif. Intell. Res.*, vol. 3, pp. 325–348, Dec. 1995.
[8] G. A. Carpenter and S. Grossberg, *Pattern Recognition by Self-Organizing Neural Networks*. Cambridge, MA: MIT Press, 1991, pp. 397–423.
[9] E. C. Cherry, "Some experiments on the recognition of speech with one and with two ears," *J. Acoust. Soc. Amer.*, vol. 25, pp. 975–979, 1953.
[10] J. Ditterich, T. Eggert, and A. Straube, "The role of the attention focus in the visual information processing underlying saccadic adaptation," *Vis. Res.*, vol. 40, pp. 1125–1134, 2000.
[11] Z. Duric and A. Rosenfeld, "Image sequence stabilization in real time," *Real-Time Imaging*, vol. 2, pp. 271–284, 1996.
[12] S. Ernst, C. Stiller, J. Goldbeck, and C. Roessig, "Camera calibration for lane and obstacle detection," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, 1999, pp. 356–361.
[13] A. de la Escalera and L. Moreno, "Road traffic sign detection and classification," *IEEE Trans. Ind. Electron.*, vol. 44, pp. 847–859, Dec. 1997.
[14] G. J. Geifing, H. Janben, and H. Mallot, "Saccadic object recognition with an active vision system," in *Proc. 10th European Conf. Artificial Intelligence*, 1992, pp. 803–805.
[15] J. Goldbeck and B. Huertgen, "Lane detection and tracking by video sensors," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, 1999, pp. 74–79.
[16] A. Guiducci, "Parametric model of the perspective projection of a road with applications to lane keeping and 3D road reconstruction," *Computer Vision and Image Understanding*, vol. 73, pp. 414–427, 1999.
[17] G. E. Hinton, J. L. McClelland, and D. E. Rumelhart, "Distributed representation," in *Parallel Distributed Processing, Explorations in the Microstructure of Cognition*, D. E. Rumelhart and J. L. McClell, Eds., 1987, vol. 1, pp. 77–109.
[18] A. Hurlbert and T. Poggio, "Do computers need attention," *Nature*, vol. 321, p. 12, 1986.
[19] K. Kluge and C. Thorpe, "The YARF system for vision-based road following," *Math. Comput. Modeling*, vol. 22, pp. 213–233, 1995.
[20] J. Konorski, *Integrative Activity of the Brain*. Chicago, IL: Univ. Chicago Press, 1967.
[21] H. S. Lai and H. C. Yung, "Lane detection by orientation and length discrimination," *IEEE Trans. Syst., Man, Cybern. B*, vol. 30, pp. 539–548, Aug. 2000.
[22] W. Li, X. Jiang, and Y. Wang, "Road recognition for vision navigation of an autonomous vehicle by fuzzy reasoning," *Fuzzy Sets Syst.*, vol. 93, pp. 275–280, 1998.
[23] C. Martindale, *Cognition and Consciousness*. Pacific Grove, CA: Brooks/Cole, 1981.
[24] ——, *Cognitive Psychology, A Neural-Network Approach*. Pacific Grove, CA: Brooks/Cole, 1991, pp. 95–116.
[25] R. Milanese, "Detecting Salient Regions in an Image: From Biological Evidence to Computer Implementation," Ph.D. dissertation, Univ. Geneva, Geneva, Switzerland, 1993.
[26] K. Naab and G. Reichart, "Driver assistance systems for lateral and longitudinal vehicle guidance—Heading control and active cruise control," in *Proc. AVEC*, 1994, pp. 449–454.
[27] D. Nair and J. K. Aggarwal, "Moving obstacle detection from a navigating robot," *IEEE Trans. Robot. Automat.*, vol. 14, pp. 404–416, June 1998.
[28] G. Piccioli, E. D. Micheli, P. Parodi, and M. Campani, "A robust method for road sign detection and recognition," *Image Vision Comput.*, vol. 14, pp. 208–223, 1996.
[29] G. Reichart, R. Haller, and K. Naab, "Toward future driver assistance systems," *Automotive Technol. Int.*, pp. 25–29, 1995.
[30] M. Schraut, K. Naab, and T. Bachmann, "BMW's driver assistance concept for integrated longitudinal support," presented at the 7th Intelligent Transport Systems World Congr., Turin, Italy, 2000, Paper 2121.
[31] W. Spieth, J. F. Curtis, and J. C. Webster, "Responding to one of two simultaneous messages," *J. Acoust. Soc. Amer.*, vol. 26, pp. 391–396, 1954.
[32] H. D. Tagare, K. Toyama, and J. G. Wang, "A maximum-likelihood strategy for directing attention during visual search," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 490–500, May 2001.
[33] A. Takahashi, Y. Ninomiya, M. Ohta, and K. Tange, "A robust lane detection using real-time voting processor," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, 1999, pp. 577–580.
[34] E. B. Titchener, *A Text-Book of Psychology*. New York: Macmillan, 1910.
[35] J. K. Tsotsos, W. W. S. M. Culhane, Y. Lai, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," *Artificial Intell.*, vol. 78, no. 1-2, pp. 507–545, 1995.
[36] P. Venhovens, J. Bernasch, J. Lowenau, H. Rieker, and M. Schraut, "The application of advanced vehicle navigation in BMW driver assistance systems," presented at the SAE Congress, Detroit, MI, 1999.
[37] J. M. Wolfe, "Guided search 2.0: A revised model of visual search," *Psychonomic Bull. and Rev.*, vol. 1, no. 2, pp. 202–238, 1994.
[38] S. M. Wong and M. Xie, "Lane geometry detection for the guidance of smart vehicle," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, 1999, pp. 925–928.
[39] M. Yamada, K. Ueda, I. Horiba, and N. Sugie, "Discrimination of the road condition toward understanding of vehicle driving environments," *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, pp. 20–24, 1999.
[40] S. M. Zeki, "Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey," *J. Physiol.*, vol. 236, pp. 549–573, 1974.

Chiung-Yao Fang (M'96–A'97) received the B.Sc. and M.Sc. degrees in information and computer education from National Taiwan Normal University, Taipei, Taiwan, R.O.C., in 1992 and 1994, respectively

She is currently an Instructor of the Department of Information and Computer Education, National Taiwan Normal University. Her areas of research interest include neural networks, pattern recognition, and computer vision.

Sei-Wang Chen (S'85–M'85–SM'97) received the B.Sc. degree in atmospheric and space physics and the M.Sc. degree in geophysics from National Central University, Taipei, Taiwan, R.O.C., in 1974 and 1976, respectively, and the M.Sc. and Ph.D. degrees in computer science from Michigan State University, East Lansing, in 1985 and 1989, respectively.

From 1977 to 1983, he worked as a Research Assistant in the Computer Center, Central Weather Bureau, R.O.C. In 1990, he was a Researcher with the Advanced Technology Center, Computer and Communication Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan, R.O.C. From 1991 to 1994, he was an Associate Professor of the Department of Information and Computer Education at the National Taiwan Normal University, Taipei, Taiwan, R.O.C. From 1995 to 2001, he became a Full Professor of the same department. He is currently a Professor of the Graduate Institute of Computer Science and Information Engineering at the same university. His areas of research interest include neural networks, fuzzy systems, pattern recognition, image processing, and computer vision.



Chiou-Shann Fuh received the B.S. degree in computer science and information engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1983, the M.S. degree in computer science from the Pennsylvania State University, University Park, in 1987, and the Ph.D. degree in computer science from Harvard University, Cambridge, MA, in 1992.

From 1992 to 1993, he was with AT\&T Bell Laboratories, Murray Hill, NJ, and engaged in performance monitoring of switching networks. From 1993 to 2000, he was an Associate Professor in Department of Computer Science and Information Engineering, National Taiwan University, and was then promoted to a Full Professor. His current research interests include digital image processing, computer vision, pattern recognition, and mathematical morphology.