Critical Motion Detection of Nearby Moving Vehicles in a Vision-Based Driver-Assistance System

Shen Cherng, Chiung-Yao Fang, Chia-Pei Chen, and Sei-Wang Chen, *Senior Member, IEEE*

Abstract—**Driving always involves risk. Various means have been proposed to reduce the risk. Critical motion detection of nearby moving vehicles is one of the important means of preventing accidents. In this paper, a computational model, which is referred to as the dynamic visual model (DVM), is proposed to detect critical motions of nearby vehicles while driving on a highway. The DVM is motivated by the human visual system and consists of three analyzers: 1) sensory analyzers, 2) perceptual analyzers, and 3) conceptual analyzers. In addition, a memory, which is called the episodic memory, is incorporated, through which a number of features of the system, including hierarchical processing, configurability, adaptive response, and selective attention, are realized. A series of experimental results with both single and multiple critical motions are demonstrated and show the feasibility of the proposed system.**

Index Terms—**Assembly of adaptive-resonance-theory (ART) neural networks, driver-assistance system (DAS), dynamic visual model (DVM), fuzzy integral, spatiotemporal attention (STA) neural network.**

I. INTRODUCTION

DRIVING a car always involves risk. Although modern technology cannot eliminate risk, it can, at least, reduce risk. Various means toward the goal of driving risk reduction have been reported. They can be categorized into two classes: 1) passive and 2) active. Passive approaches (e.g., seat belts and airbags) are intended to reduce the degree of injury in case of an accident, whereas active methods attempt to prevent accidents in advance. Driver-assistance systems (DASs) [1], [6], [30], [31], [33], [39] are one kind of active system that brings potentially hazardous conditions to the driver's attention as soon as possible.

Various functions constitute a DAS, of which automatic obstacle avoidance [2], [4], [7], [9], [10], [18], [19], [32], [40] is one of the most important functions. Obstacles can broadly be divided into two categories: 1) static and 2) dynamic.

Static obstacles are fixed with respect to the ground, such as refuge islands, guard rails, traffic signals, and stopped vehicles. Dynamic obstacles are movable, such as animals, pedestrians, and moving vehicles. However, all objects are moving while the observer moves. We refer to the motions that result from the observer's movement as induced motions. Static obstacles have only induced motions, whereas dynamic obstacles have both induced and self motions [40], [42]. Induced motions are predictable, whereas self motions are not. Dealing with dynamic obstacles is typically more difficult than dealing with static ones. In this paper, moving vehicles, which are a kind of dynamic object, are considered.

A. Critical Motions

Regardless of the diverse behaviors of a moving vehicle, its motion can always be decomposed into a sequence of simple motions. Every simple motion is characterized by a change in either speed or direction. Speed changes include acceleration and deceleration. Direction changes include left and right drifts. Various motion behaviors can be generated by concatenating simple motions. A generated motion behavior may be safe or perilous. However, an individual safe motion may become hazardous when extraneous traffic conditions are considered. We refer to the motions taking external traffic conditions into account as conditioned motions. Conditioned risky motions form what we call critical motions in this paper.

Considering one nearby moving vehicle, it may cause several potential critical situations.

1) The right (left) front vehicle slows down.
2) The right (left) front vehicle changes lanes to the lane of the observer vehicle with lower speed than the observer vehicle.
3) The vehicle in front slows down.
4) The vehicle in front changes lanes to the right (left) with lower speed than the observer vehicle.
5) The right (left) rear vehicle passes the observer vehicle.

The aforementioned situations all result in decreasing the distance between the nearby and the observer vehicles. In general, several vehicles may surround the observer vehicle, and multiple critical situations can occur simultaneously.

B. Sensor Overview

Many sensors are available for detecting vehicles [3], [8], [21]–[24], [35], [38], [44], such as laser radar, passive far infrared, reflected light detector, microwave radar, millimeter-wave radar, acoustic array, and ultrasonic detector. Each sensor has its own advantages and disadvantages [20]. In this paper,

S. Cherng is with the Department of Computer Science and Information Engineering, Chengshiu University, Kaohsiung 833, Taiwan (e-mail: cherng@msu.edu).

C.-Y. Fang and S.-W. Chen are with the Laboratory for Computer Vision and Image Processing, Department of Computer Science and Information Engineering, National Taiwan Normal University, Taipei 116, Taiwan (e-mail: violet@csie.ntnu.edu.tw; schen@csie.ntnu.edu.tw).

C.-P. Chen is with the Chunghwa Telecom Laboratories, Taoyuan 32617, Taiwan (e-mail: peggychen@cht.com.tw).

a video camera is utilized. A video camera that possesses a large field of view can collect much information for a variety of applications, such as wide-area monitoring [13], unusual traffic event detection [12], [29], road sign recognition [11], traffic sign/signal identification [14], and road/driving condition analysis [43]. However, sophisticated algorithms are generally necessary to accomplish these tasks.

C. Proposed System

One computational model, referred to as the dynamic visual model (DVM), is proposed in this paper to detect the critical motion of nearby moving vehicles. Unlike static visual models that worked on still images, the DVM performs visual analysis based on video sequences. Video sequences are different from still images in several respects. Still images are taken at independent time instants. Individual images are unrelated to one another and can separately be processed. Video images are acquired at (usually rapid) sequential instants in time, causing successive images to temporally be correlated. Each video sequence records a certain period of time in an environment. To interpret environmental variations during the period, the video sequence has to be dealt with collectively. Furthermore, still images contain only spatial information, and they are suitable for structural analysis of static scenes, such as forms of objects and their spatial relationships. Video sequences, which include both spatial and temporal information, are feasibly employed to study structures and motions of dynamic scenes.

Video sequences are, in a sense, analogous to the streams of signals produced by the photoreceptors in the retina. The manner of signal processing in the human visual system may provide useful clues for video processing. In the next section, we review the psychophysical evidences that characterize the proposed DVM. Its architecture and workflow are then addressed in Section III. The system for detecting critical motions of nearby vehicles based on the DVM is presented in Section IV. Section V demonstrates experimental results, followed by the concluding remarks and future work given in Section VI.

II. Psychophysical Eviences

The proposed DVM grasps several aspects of the human visual system, including hierarchical processing, neural configurability, adaptive response, and selective attention.

A. Hierarchical Processing

According to psychophysicists [25], [28], there are hierarchical organs that exist along the path from the retina to the areas of the brain concerned with association. Each organ performs a distinct level of information analysis, such as sensory, perceptual, syntactic, and semantic analysis. Although the organs play different roles in a visual activity, they possess similar anatomical structures [15], [36]. Every organ is composed of a large number of interconnected neurons, which are organized into one or more layers. The proposed DVM consists of three analyzers, which are referred to as the sensory, perceptual, and conceptual analyzers. They are arranged in a hierarchy. Two neural modules—a spatiotemporal attention (STA) neural net-

work and an assembly of adaptive resonance theory II (ART2) neural networks—are introduced to realize the perceptual and conceptual analyzers, respectively.

B. Neural Configurability

Although the brain possesses an enormous capacity to process information, it consists of a finite number of neurons. Three characteristics regarding the nervous system have often been mentioned to illustrate the capacity of the brain: 1) connectionism, 2) parallelism, and 3) configurability. The first two characteristics are well known, but we are interested in the third characteristic. Two attributes—adaptive organizing and multitasking—feature the configurability of the nervous system. Considering a mental task, it is first decomposed by the brain into a set of independent processes that can separately be carried out. The processes are then allocated to cognitive units. The units, which may originally be unrelated to one another, are organized and cooperate for the mental task. We call this attribute the adaptive organization of the nervous system. Furthermore, a cognitive unit need not be available when it is asked to participate in a task. A cognitive unit can join in several distinct tasks at a time. We call this the multitasking of the nervous system.

To incorporate the feature of configurability in the proposed DVM, a memory, which is called the episodic memory, is introduced. This memory provides not only storage but also representations for preserving and retrieving long-term, short-term, and transient information. In addition to configurability, several functions of the DVM, such as adaptive response and selective attention, are also realized through the episodic memory.

C. Adaptive Response

We all have the experience of getting out of the way of a quickly moving object before identifying what it is. This illustrates that we can perceive motion earlier than form and meaning. However, motions are ubiquitous throughout the retina. How can we attend to the motions that originate from the objects of our interest?

According to vision physiologists [15], [36], the eye contains two types of ganglion cells—parvocellular and magnocellular cells—and their associated pathways. Parvocellular pathways are responsible for structural analysis, whereas magnocellular pathways are devoted to motion analysis. Structure and motion analyses are performed in parallel in the respective pathways. However, the two analyses are not independent, because the two pathways interact with each other at a number of neural layers. It is the exchange of information that the motions associated with the objects of interest are attended. This characteristic has been referred to as the adaptive response of the visual system and is of importance to our application. It suggests that we could perceive vehicle motions before vehicle identification.

D. Attention Models

Many attention models [16], [28] have been reported to realize the mechanism of adaptive response. They can be categorized into two classes: 1) voluntary and 2) involuntary.
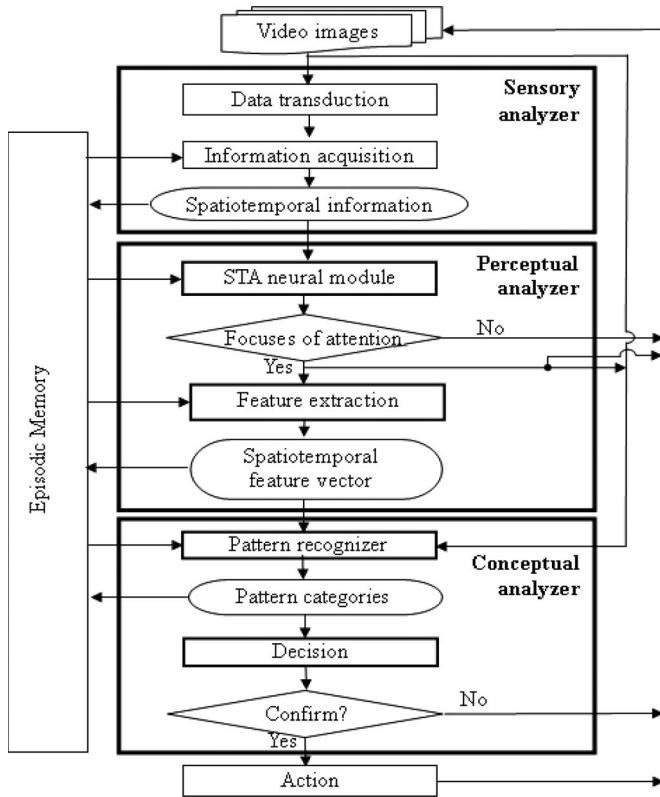
Fig. 1.   Proposed DVM.

The voluntary class forces attention to anticipated targets, whereas the involuntary class results from sudden or unexpected events. In this paper, the motions associated with vehicles are of concern; thus, the voluntary model is considered. Two principles of selectivity, i.e., filtering and amplification, characterize the voluntary model. Filtering selectivity gradually attenuates unwanted events until they are filtered out from the center of attention. Amplification selectivity progressively intensifies selected events until they seize the focus of attention. Both mechanisms of selectivity are implemented in the proposed DVM.

Where does attentive selection occur in the visual system? In [28], Broadbent argued that attentive selection appears only in the sensory layer, where selected information is passed on for further processing. Norman argued that attentive selection occurs in both the sensory and semantic layers. The sensory layer contributes bottom-up stimuli, and the semantic layer provides top-down expectation. As a consequence, some neurons that are (are not) activated by the bottom-up stimuli may not (may) be stimulated due to the top-down stimuli. Treisman further extended Norman's model by allowing attentive selection to take place in all neural layers. In this paper, Norman's idea is adopted, with which attentive selection occurs only in the first two layers of the DVM.

III. DVM

Fig. 1 shows the proposed DVM, which is composed of three analyzers that are referred to as the sensory, perceptual, and conceptual analyzers. Each analyzer constitutes a number of stages. The sensory analyzer consists of two major stages:

1) data transduction and 2) information acquisition. The perceptual analyzer consists of an STA pattern detection stage and an STA feature-extraction stage. The conceptual analyzer is composed of the stages of pattern recognition and decision. The data flows among the stages of the analyzers are primarily through a memory called the episodic memory.

A. Workflow of the DVM

Considering an input video sequence, the sensory analyzer of the DVM first reduces (e.g., subsampling and downsampling), converts (e.g., wavelet and Fourier transforms) and represents (e.g., hierarchical and distributed representations) the video sequence to increase the efficiency and effectiveness of subsequent processing. We call this stage the data transduction stage. The transduced video sequence is then forwarded to the information acquisition stage, in which the spatiotemporal information of target objects is extracted from the video sequence. The *a priori* knowledge of target objects is provided by the episodic memory, which is prestored. Instead of details, syntactic (e.g., color and texture) or semantic (e.g., function and behavior) hints of the objects are preserved in the memory.

The spatiotemporal information that was extracted in the sensory analyzer serves as a stimulus to the STA neural network (see Section III-B) in the perceptual analyzer. The STA neural network mimics the adaptive response of the human visual system. Two mechanisms, i.e., filtering selectivity and amplification selectivity, are implemented in the neural network. The filtering selectivity attenuates the neural activations innervated by unwanted stimuli (e.g., noise and extraneous information) based on their natures of randomness. Amplification selectivity intensifies the activations excited by the stimuli associated with target objects. If such stimuli continue, a focus of attention will eventually be established in the STA neural network.

A focus of attention, hereafter referred to as an attention pattern, preserves both the spatial and temporal information of an event. The DVM recognizes an event based on its attention pattern. Once an attention pattern emerges in the STA neural network, the feature-extraction stage is initiated, in which a mixed top-down and bottom-up process is evoked to extract local features from the attention pattern. The extracted features are arranged into a 1-D vector, called a spatiotemporal feature vector, which is to be fed to a pattern recognizer embedded in the conceptual analyzer.

The pattern recognizer consists of an assembly of ART2 neural networks [5]. Each ART2 determines the class of the attention pattern based on a feature vector extracted from a particular portion of the pattern. The decisions of individual ART2s are combined through a fuzzy integral process (see Section III-C) to arrive at the final decision. If this decision is validated, an adequate action is taken, and the system process repeats.

B. STA Neural Network

The STA neural network shown in Fig. 2 is configured as a two-layer network: one layer for input and another layer for output. The output layer is also referred to as the attention layer.
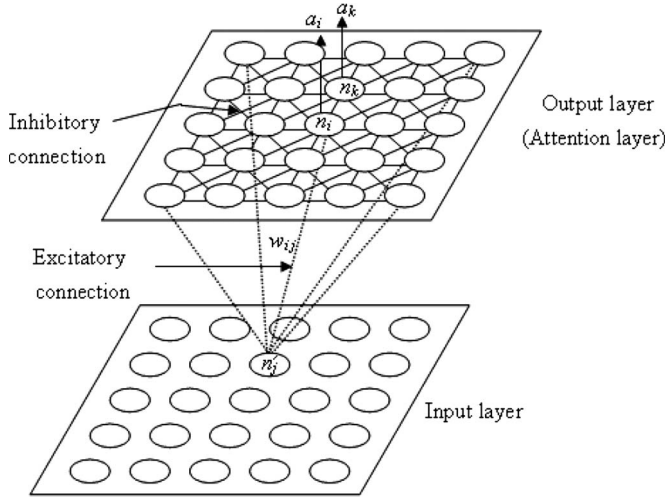
Fig. 2. STA neural network.

Fig. 3. Activation of an attention neuron in response to a stimulus.

Focuses of attention, if any, are formed in this layer. The neurons in the attention layer are arranged into an interconnected 2-D array. There is no synaptic link among input neurons, although they are also organized into a 2-D array. Both the input and output layers have the same size as the input images and are fully connected to each other.

Let w_{ij} denote the weight of the link between input neuron n_j' and attention neuron n_i and be characterized by a Gaussian, i.e., $w_{ij} = G(\boldsymbol{r}_{ij})$, where $\boldsymbol{r}_{ij}$ is the position vector of attention neuron n_j, which corresponds to the input neuron n_j' with respect to the attention neuron n_i. The weight vector for attention neuron n_i is denoted $\boldsymbol{w}_i = (w_{i1}, w_{i2}, \ldots, w_{im})$, where m is the number of input neurons. The input to attention neuron n_i due to input stimulus $\boldsymbol{x}$ is $\boldsymbol{I}_i^v = \boldsymbol{w}_i \cdot \boldsymbol{x} = \sum_{j=1}^m w_{ij} x_j$.

The lateral interaction among attention neurons is characterized by a Laplacian of Gaussian $\nabla^2 G(\boldsymbol{r})$, where $\boldsymbol{r}$ is a position vector that originates in the center of the function. The input to neuron n_i due to lateral interaction is defined as $I_i^l = \sum_{k \in N_i, k \neq i} [u_{ik} \cdot \nabla^2 G(\boldsymbol{r}_k - \boldsymbol{r}_i) \cdot a_k]$, where N_i represents the neighbors of neuron n_i and a_k is the activation of neuron n_k. Let the dynamic behavior of a_k be governed by

$$\dot{a}_i = A(-pa_i + qB(net_i)) \tag{1}$$

where p and q are positive constants, p specifies the decay rate of a_i, and q weights the net input $net_i = I_i^v + I_i^l - \Gamma$, in which Γ is a threshold to restrict the effects of noise. Functions $A(\cdot)$ and $B(\cdot)$ are

$$A(x) = \begin{cases} x, & \text{if } x > 0 \\ dx, & \text{if } x \leq 0 \end{cases}, \quad B(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases} \tag{2}$$

where $0 < d < 1$.

Fig. 3 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 about $1/pd$ to decay. $d < 1$; thus, the decay time is longer than the rise time.
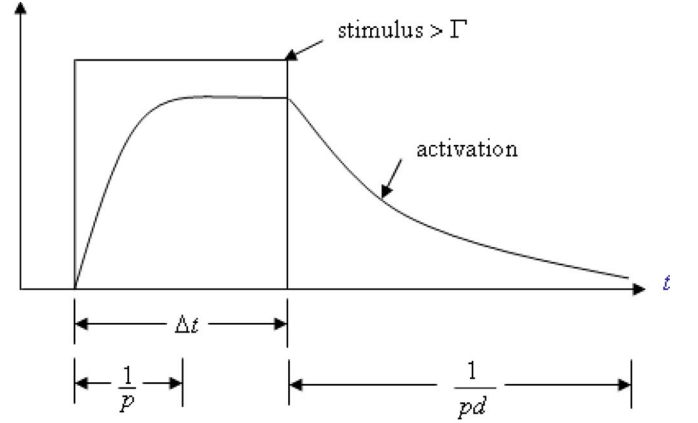
C. Decision by Fuzzy Integral

Fuzzy integrals [41], [45] have been generalized from the Lebesque or Riemann integral. In this study, the Sugeno fuzzy integral [37], which was extended from the Lebesque integral, is considered. Let $f : S \to [0, 1]$ be a function defined on a finite set S and $g : P(S) \to [0, 1]$ be a set function defined over the power set of S. Function $g(\cdot)$, referred to as a fuzzy measure function, satisfies the axioms of boundary conditions, monotonicity, and continuity. Sugeno further imposed on $g(\cdot)$ an additional property: $\forall A, B \subset S, A \cap B = \phi$. We have

$$g(A \cup B) = g(A) + g(B) + \lambda g(A) g(B), \quad \lambda \leq 1. \tag{3}$$

The fuzzy integral of $f(\cdot)$ with respect to $g(\cdot)$ is defined as

$$e = \int_S f(s) \cdot g = \sup_{\alpha \in [0,1]} \{\alpha \wedge g(A_\alpha)\} \tag{4}$$

where $\wedge$ represents the fuzzy intersection, and $A_\alpha = \{s \in S | f(s) \geq \alpha\}$.

Fuzzy integral provides an elegant nonlinear numeric approach that is suitable for integrating multiple sources of information to arrive at a value that indicates the degree of support for a particular hypothesis. Suppose that we have hypotheses, $H = \{h_i, i = 1, \ldots, n\}$, from which a decision d will be made. Let e_{h_i} be the integral value evaluated for h_i. We determined the final decision by $d = \arg \max_{h_i \in H} e_{h_i}$.

Considering any hypothesis $h \in H$, let S be the set collecting all the information sources at hand. Function $f(\cdot)$ that receives an information source s returns a value $f(s)$ that reveals the level of support of s to the hypothesis h. The degrees of worth of information sources may be different; thus, function $g(\cdot)$ takes as input a subset of information sources and gives a value that reflects the degree of worth of the set of sources relative to the other sources. Let $d(s) = g(\{s\})$. Function $d(\cdot)$ is referred to as the density function of $g(\cdot)$.

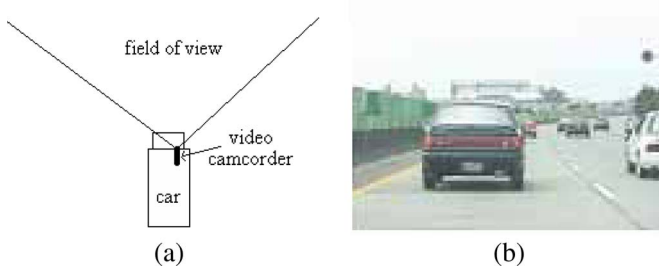In general, densities $d(s)$, $s \in S$ are readily estimated.

Fig. 4.   Video camera is mounted in the front of the vehicle.

For any subset $A = \{s_i, i = 1, \ldots, m\}$ of S, the fuzzy measure of A can be computed recursively from (3), i.e.,

$$g(A) = \sum_{i=1}^{m} d(s_i) + \lambda \sum_{i=1}^{m-1} \sum_{j=i=1}^{m} d(s_i)d(s_j)$$

$$+ \cdots + \lambda^{m-1} d(s_1), \ldots$$

$$d(s_m) = \left(\prod_{s_i \in A} (1 + \lambda d(s_i)) - 1 \right) / \lambda. \tag{5}$$

$g(S) = 1$; thus, the value of λ can be determined by solving $\lambda + 1 = \prod_{s_i \in S} (1 + \lambda d(s_i))$.

In (4), there are $2^{|S|}$ subsets of S needed to perform fuzzy integral. Let $S' = \{s'_1, s'_2, \ldots, s'_{|S|}\}$ be the sorted version of S such that $f(s'_1) \geq f(s'_2) \geq \cdots \geq f(s'_{|S|})$. We can rewrite (4) as

$$e = \int_S f(s) \cdot g = \sup_{\alpha \in [0,1]} \{\alpha \wedge g(A_\alpha)\} = \bigvee_{1 \leq i \leq |S|} [f(s'_i) \wedge g(S'_i)] \tag{6}$$

where $\vee$ specifies the fuzzy union, and $S'_i = \{s'_1, s'_2, \ldots, s'_i\}$. This equation reduces the number of subsets required to perform the fuzzy integral from $2^{|S|}$ [by (4)] to $|S|$.

IV. CRITICAL MOTION DETECTION

In this section, the system based on the proposed DVM for detecting critical motions of nearby moving vehicles while driving on a highway is addressed. The system consists of four major components: 1) sensory, 2) perceptual, 3) conceptual, and 4) episodic memory.

A. Spatiotemporal Information Extraction

The input data to the system are video sequences acquired by a video camera mounted in the front of the observer vehicle [see Fig. 4(a)]. One example image taken by the video camera is shown in Fig. 4(b). In the data transduction stage of the sensory component, we subsample images from the input video sequence. For each sampled image, we further reduce its size by downsampling. Downsampling also eliminates the interlacing effect of the video camera.

The transduced video sequence is forwarded to the information acquisition stage of the sensory analyzer, where the spatiotemporal information of scenes is extracted. Let S denote the sequence of transduced images and I^t be the tth image in S.

Let R^t, G^t, and B^t represent the three color components of I^t, i.e., $I^t = (R^t, G^t, B^t)$. We define the MAX image $I^t_{\max} = (R^t_{\max}, G^t_{\max}, B^t_{\max})$ and MIN image $I^t_{\min} = (R^t_{\min}, G^t_{\min}, B^t_{\min})$ at time t as follows:

$$R^t_{\max}(i,j) = \max\{R^t(i,j), R^{t-1}_{\max}(i,j)\}$$

$$R^t_{\min}(i,j) = \min\{R^t(i,j), R^{t-1}_{\min}(i,j)\}$$

$$G^t_{\max}(i,j) = \max\{G^t(i,j), G^{t-1}_{\max}(i,j)\}$$

$$G^t_{\min}(i,j) = \min\{G^t(i,j), G^{t-1}_{\min}(i,j)\}$$

$$B^t_{\max}(i,j) = \max\{B^t(i,j), B^{t-1}_{\max}(i,j)\}$$

$$B^t_{\min}(i,j) = \min\{B^t(i,j), B^{t-1}_{\min}(i,j)\}$$

where (i, j) denotes any pixel location. Initially, we let $I^0_{\max} = I^0_{\min} = I^0$. Image $I^t_{\max}$ ($I^t_{\min}$) preserves the maximum (minimum) color values of the input video sequence up to time t and gets much brighter (darker) with time.

Let M^t and N^t represent the intensity counterparts of $I^t_{\max}$ and $I^t_{\min}$, respectively, where $M^t(i,j) = ((R^t_{\max}(i,j) + G^t_{\max}(i,j) + B^t_{\max}(i,j))/3$ and $N^t(i,j) = ((R^t_{\min}(i,j) + G^t_{\min}(i,j) + B^t_{\min}(i,j))/3$. We compute the spatial difference image (DIF image for short) D^t at time t by $D^t(i,j) = M^t(i,j) - N^t(i,j)$, which accumulates the positions of moving objects up to time t. Next, we compute the spatiotemporal derivative (DER) images $D^{t'}$ by $D^{t'}(i,j) = |D^t(i,j) - D^{t-1}(i,j)|$, which records object motions at time t. Note here that we detect object motions from the DIF images instead of the video images, because DIF images get stronger responses than the video image. Moreover, the influence of image instability that originates in vehicle vibration is also reduced.

B. Attention Patterns

The DER images serve as the input stimuli to the STA neural network in the perceptual component. Attention will be grabbed by the STA neural network only when the neural activations over the attention layer, i.e., the attention pattern, of the neural network become sufficiently prominent (both large in size and strong in strength). The neural network then keeps its attention on the pattern until it reaches maximum, called the principal attention pattern. Different motion behaviors of vehicles generate different principal attention patterns. Fig. 5 shows nine principal attention patterns originating from nine distinct critical motions.

In a real situation, several vehicles may surround the observer vehicle, and multiple critical motions may simultaneously occur. Unlike single critical motions, multiple ones have no particular principal attention patterns. This is because attention patterns of individual critical motions rarely reach their maximums at the same time. Furthermore, multiple critical motions may have their attention patterns overlapping or be connected to each other. We deal with multiple critical motions by identifying individual motions one by one. After completing
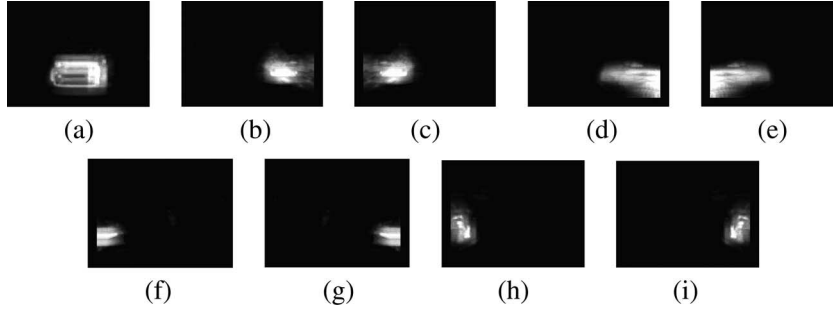
Fig. 5. Principal attention patterns that originate in distinct critical motions of a nearby vehicle. (a) Vehicle ahead slows down. (b) Right front vehicle changes lanes to the left. (c) Left front vehicle changes lanes to the right. (d) Vehicle ahead changes lanes to the right. (e) Vehicle ahead changes lanes to the left. (f) Left front vehicle slows down. (g) Right front vehicle slows down. (h) Left rear vehicle passes observer. (i) Right rear vehicle passes observer.
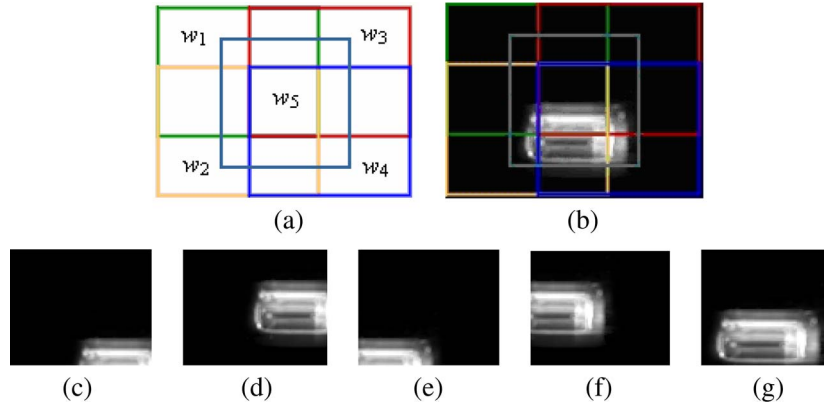

Fig. 6. (a) Five overlapping windows, denoted by w_1, w_2, w_3, w_4, and w_5. (b) Attention pattern divided by the five windows. (c) Portions of the attention pattern that correspond to the five windows.

the processing of the last motion, the system initializes the attention layer of STA neural network and the MAX, MIN, DIF, and DER sequences. In addition, note that our system performs initialization every fixed time interval T if no critical motion is detected within this period. We define T as $T = c \cdot \max_i t_i$, where c is a positive integer larger than 1, and t_i is the time needed to complete the critical motion of type i.

C. Spatiotemporal Feature Vector

The principal attention patterns that results from different critical motions may be different in shape, location, distribution of neural activations, or combinations of these. In this section, we discuss how we can extract these attributes from an attention pattern.

As shown in Fig. 5, every critical motion generates a principal attention pattern located at a particular position. Determining the location of a principal attention pattern is affected by camera vibration, uneven road surface, and road slope, resulting in only an approximated location of the pattern. To overcome this problem, we divide the attention layer of the STA neural network into five overlapping windows [see Fig. 6(a)], denoted by w_1, w_2, w_3, w_4, and w_5. Fig. 6(b) shows a principal attention pattern superimposed by the five overlapping windows, and Fig. 6(c)–(g) show the windowed attention patterns. We determine the location of a principal attention pattern as the window, called the position window, which covers the largest portion of the pattern, called the dominant portion of the pattern. In the
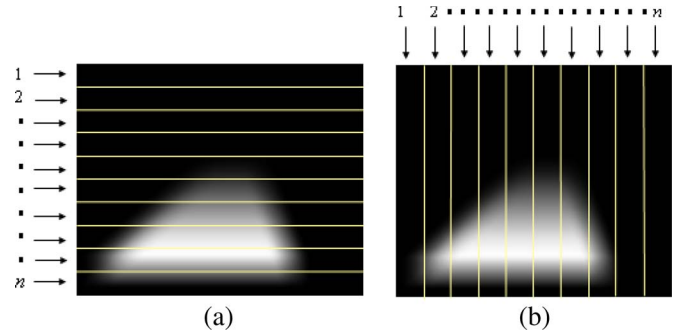

Fig. 7. Example to illustrate local feature extraction from a window.

example in Fig. 6, since window w_5 covers almost the entire pattern, this window is determined to be the position of the attention pattern.

Consider multiple critical motions, whose attention patterns may overlap one another. The issue of pattern overlapping is, in a sense, similar to the issue of shape occlusion encountered in applications of object recognition. Local features have commonly been employed to deal with shape occlusion. In the following, we define local features of attention patterns and discuss how we can extract them from an attention pattern.

Note that the overlapping windows separate an attention pattern into parts and that every window can, at any time, contain at most one dominant pattern. As shown in Fig. 5, although principal attention patterns of different critical motion may have the same position window, in practice, such motion
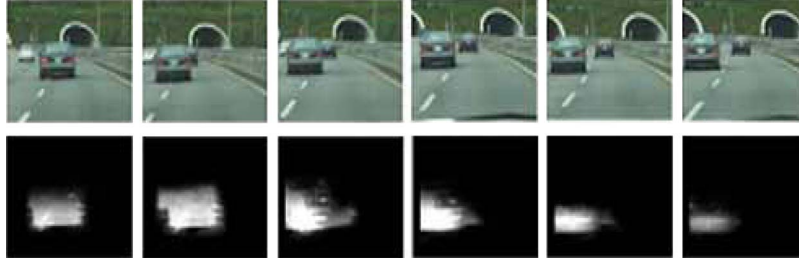
Fig. 8.   Vehicle directly in front changes lanes to the left.

will not occur at the same time; otherwise, accidents will occur. As a consequence, we can always extract local features of individual attention patterns from some windows.

Considering a window, first, we divide the window into $n(= 10$ in experiments) equal horizontal zones [see Fig. 7(a)]. For each zone, we average the neural activations of each column. Let $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_m$ be the average activations of the columns, where m is the number of columns. Next, we compute the mean column by $\bar{c} = \sum_{i=1}^{n} i \cdot \bar{a}_i / \sum_{i=1}^{n} \bar{a}_i$, with which we calculate the second M_2 and third M_3 moments of the distribution of average neural activations, $M_2 = \sum_{i=1}^{n}(i \cdot \bar{a}_i - \bar{c})^2 / \sum_{i=1}^{n} \bar{a}_i$ and $M_3 = \sum_{i=1}^{n}(i \cdot \bar{a}_i - \bar{c})^3 / \sum_{i=1}^{n} \bar{a}_i$. The skewness s of the distribution of the average neural activations is defined as $s = M_3/(M_2\sqrt{M_2})$ [17], [27], which expresses the degree of nonsymmetry of the distribution. The skewness of a distribution can be positive or negative, indicating a right or left long tail of the distribution.

For each window, $2n$ skewness values (i.e., n from horizontal zones and another n from vertical zones) are calculated. The $2n$ skewness values form the feature vector of the window. Five feature vectors are extracted from the five overlapping windows. Each feature vector is then fed into an ART2 neural network in the conceptual component of the system.

D. Pattern Recognition

There are five ART2 neural networks that constitute the pattern recognizer. Each neural network receives the feature vectors that were extracted from a particular overlapping window. Consider a neural network, in which each input feature vector matches against a set of patterns that have been prestored in the neural network through an offline training process. Each stored pattern corresponds to a specific type of critical motion. If the input feature vector matches a stored pattern well, the associated type of critical motion is regarded as that generating the feature vector. If no pattern is matched, a "do not care" signal is returned by the neural network.

E. Fuzzy Decision

Each ART2 neural network in the pattern recognizer receives a sequence of feature vectors extracted from the corresponding overlapping window and returns a series of classification decisions of the attention pattern under consideration. The series of decisions is integrated using a fuzzy integral process to arrive at the final decision.
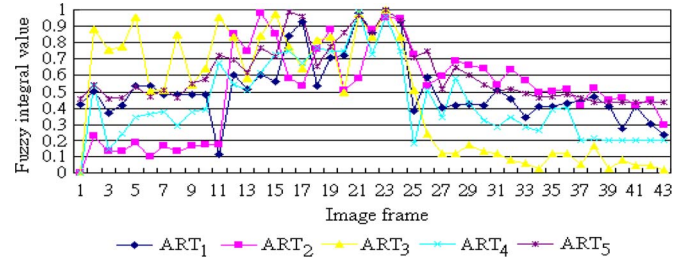

Fig. 9.   Fuzzy integral values calculated by the five ART2 neural networks for the 43 video images used in the second experiment.

Let $s_1, s_2, \ldots, s_t$ be the sequence of decisions up to time t made by a neural network. The set $S = \{s_1, s_2, \ldots, s_t\}$ then forms what we call the collection of information sources. Let $P = \{\boldsymbol{p}_0, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_T\}$ be the set of principal attention patterns of distinct critical motions, in which $\boldsymbol{p}_0$ represents a null pattern. Set P here serves as a hypothesis set. For each hypothesis $\boldsymbol{p}_i \in P$, we define a support function $f_i : S \to [0, 1]$ for the hypothesis as

$$f_i(s_j) = e^{-\alpha \| p'_j - p_i \|} \tag{7}$$

in which α is a positive constant, and $\boldsymbol{p}'_j \in P$ is the attention pattern decided by decision s_j. The aforementioned equation states that the more similar the decided pattern $\boldsymbol{p}'_j$ to hypothesis pattern $\boldsymbol{p}_i$, the higher the degree of support of decision s_j to hypothesis $\boldsymbol{p}_i$.

Next, we define the fuzzy density function $d_i : S \to [0, 1]$ as

$$d_i(s_j) = \min\left\{ \frac{n'_j}{n_i}, \frac{n_i}{n'_j} \right\} \tag{8}$$

where n_i and n'_j are the numbers of nonzero pixels of patterns $\boldsymbol{p}_i$ and $\boldsymbol{p}'_j$, respectively. Function $\min\{\cdot, \cdot\}$ forces value $d_i(s_j)$ to fall between [0, 1]. The aforementioned equation states that the closer the numbers (n_i and n'_j) of nonzero pixels of patterns $\boldsymbol{p}_i$ and $\boldsymbol{p}'_j$, the higher the degree of worth of decision s_j. Having defined the densities of individual decisions, we are ready to calculate the degree of worth $g_i(S_k)$ of any subset $S_k \subseteq S$ of decisions by (5).

Let $S' = \{s'_1, s'_2, \ldots, s'_{|S|}\}$ be the sorted version of S such that $f_i(s'_1) \geq f_i(s'_2) \geq \cdots \geq f_i(s'_{|S|})$. Substituting $f_i(s'_j)$ and $g_i(S_j)$, $(1 \leq j \leq |S|)$ into (6), we can calculate the fuzzy integral e_i of hypothesis $\boldsymbol{p}_i$. The aforementioned process repeats for all the hypotheses in P. Let $e_{\max} = \max_{\boldsymbol{p}_i \in P} e_i$. We collect the hypotheses with $e_{\max} - e_i < \tau$ (τ is a threshold) in a set called the decision set. Let D_1, D_2, D_3, D_4, and D_5 be the decision
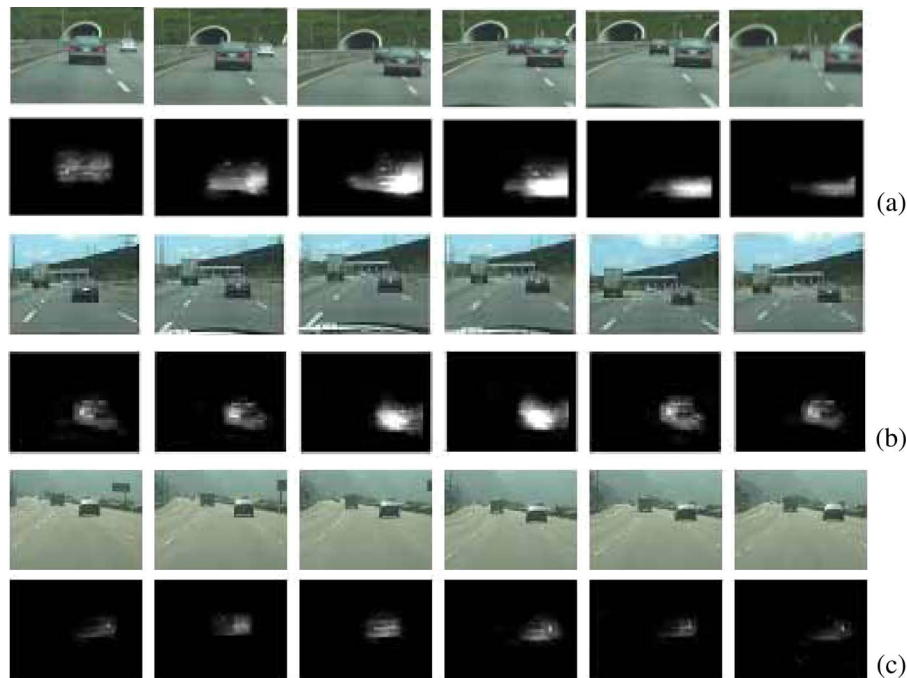
Fig. 10. Vehicles at different distances ahead change lanes to the right.

sets of the five ART2 neural networks. We determine the final decision set D as $D = \{\boldsymbol{p}_i \in P | \exists D_j, \boldsymbol{p}_i \in D_j \wedge e_i > \tau'\}$, where τ' is a threshold. The critical motions that correspond to the attention patterns in D are reported by the system. If D is an empty set, no critical motion is detected. If D contains more than one pattern, multiple critical motions are detected.

V. EXPERIMENTAL RESULTS

The input data for our system were acquired using a progressive video camera (SONY TRV-900) mounted in the front windshield of a vehicle while driving on an expressway at a speed of around 90 km/hr. The input video sequences at a rate of 30 frames/s were downsampled to a rate of 6 frames/s, which is the processing speed of our current system. The critical motions that were considered in this paper take more than 3 s to complete; thus, we further reduce the size (640×480 pixels) of each image to 160×120 pixels to increase the processing rate.

Fig. 8 shows the first experimental example, in which a vehicle ahead changes lanes to the left. A video sequence of 8 s (240 video images in total) was used in this experiment. Of these, 43 images were processed, and only six of them are displayed in the first row of the figure. The corresponding sequence of attention patterns generated by the vehicle in the STA neural network is shown in the second row. An attention pattern is examined by the pattern recognizer only if the pattern has grown sufficiently prominent. Such a pattern is first decomposed into five subpatterns according to the predefined overlapping windows. Next, each subpattern is recognized by an ART2 neural network in the pattern recognizer. The results of individual neural networks are then integrated to arrive at a consistent decision. Instead of depending only on an attention pattern, our system determines the motion behavior of a nearby

vehicle by aggregating a series of decisions made from a sequence of attention patterns.

Fig. 9 shows the (maximum) fuzzy integral values calculated by the five ART2 neural networks for the 43 input images in the aforementioned example. We divide the figure into three vertical parts at images 12 and 24. In the first part (from images 1 to 12), the variance of the fuzzy integral values that the ART2 neural networks returned is larger than both the second (from images 13 to 24) and third (from the degree of confidence in a decision, a large variance of images 24 to 43) parts. Since a fuzzy integral value reflects the degree of confidence in a decision, a large variance of fuzzy integral values indicates a high level of disagreement among the decisions of the ART2 neural networks. Although both the second and third parts of the figure have relatively small variances of fuzzy integral values, our system has decided in favor of the decision made by the second part, because its average (0.78) fuzzy integral value is much larger than that (0.42) of the third part. Our system locates the video segment to make a decision based on the variances and averages of fuzzy integral values of the video segments.

Fig. 10 shows three examples of a vehicle in front changing lanes to the right. In each example, the input video sequence is shown in the first row, and the corresponding sequence of attention patterns is displayed in the second row. Comparing the sequences of attention patterns of these three examples, it is clear that the attention patterns in the sequence of the first example are all more prominent than those of the second. This is because the preceding vehicle in the first example is closer to the observer vehicle than that in the second example. Our system successfully recognized the behaviors of the preceding vehicles in these two examples. However, our system did not give any response to the input sequence of the third example, because the preceding vehicle is too far to generate sufficiently prominent attention patterns. More experimental results of distinct
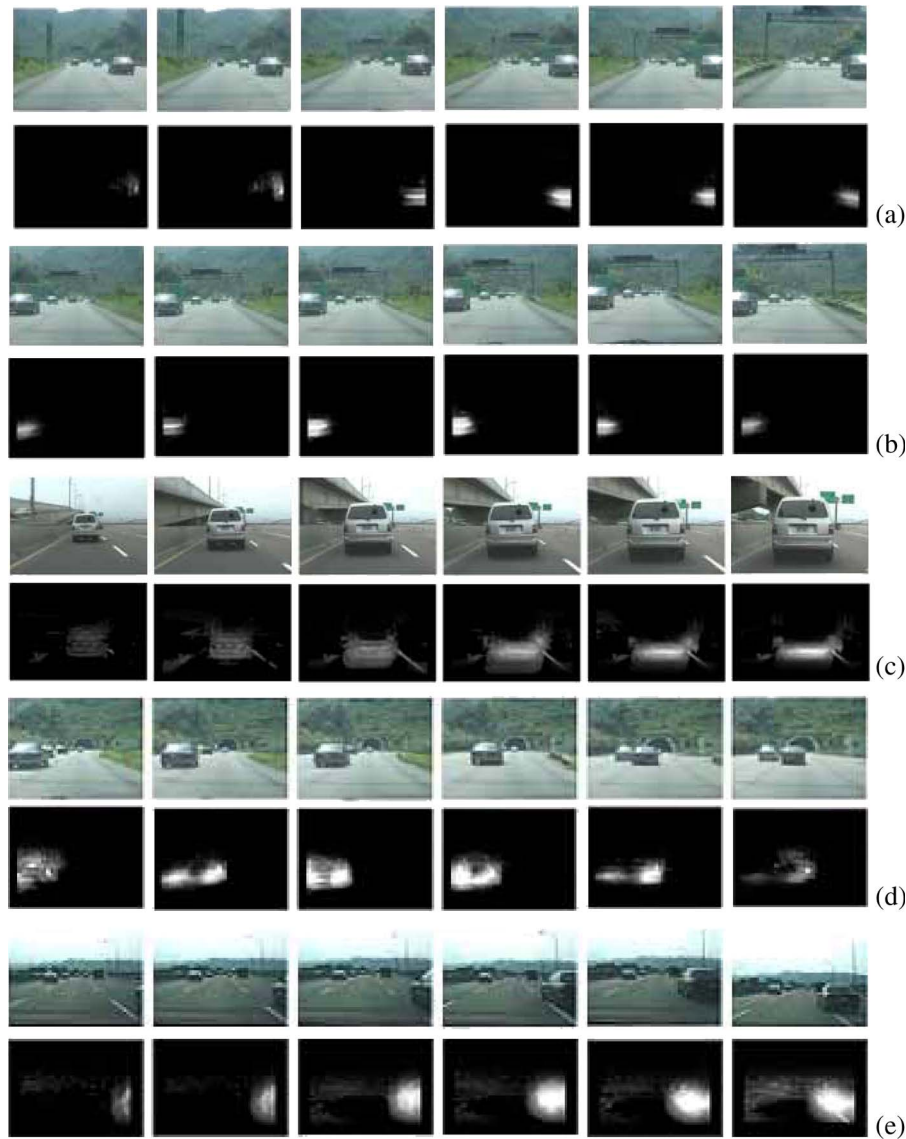
Fig. 11. Experimental results of different critical motions. (a) Right front vehicle slows down. (b) Left front vehicle slows down. (c) Vehicle directly ahead slows down. (d) Left front vehicle changes lanes to the right. (e) Vehicle passing on the right.

critical motion is shown in Fig. 11: 1) a right front vehicle slowing down, 2) a left front vehicle slowing down, 3) a vehicle directly in front slowing down, (d) a left front vehicle changing lanes to the right, and (e) a vehicle passing on the right.

Fig. 12 demonstrates the performances of our system under different illumination conditions, including an input video sequence taken in a tunnel [see Fig. 12(a)] and one acquired at night [see Fig. 12(b)]. Here, we neglected the video sequences that involve rapid changes of illumination (e.g., entering or exiting a shadow cast by a passing cloud or a large building, entering or exiting a tunnel, or direct expose of the camera to sunlight as the vehicle turns), because the time interval of a change in illumination is typically short. A noted attention pattern is hardly generated in the STA neural network within a short period. However, if the attention pattern of a critical motion is nearly matured, an illumination change will ruin the pattern and lead to a missing detection of the critical motion.

So far, we have considered only the cases of single critical motions. In the following, three experimental results of multiple

critical motions are demonstrated. The first example is shown in Fig. 13, where a vehicle in front of the observer vehicle changes lanes to the left, whereas a vehicle in the left lane slows down. The system detected only the motion behavior of the latter vehicle. The front vehicle is distant; thus, the associated attention pattern is small and weak, and the system did not respond to this pattern. Fig. 14 shows the second example, in which two preceding vehicles successively change lanes; the first one changes lanes to the left and the second to the right. The attention patterns generated by the two vehicles are different in both time and location; thus, the system easily distinguished between the two patterns and successfully recognized the motion behaviors of the two vehicles.

The final example of multiple critical motions is shown in Fig. 15. Two vehicles pass the observer vehicle on its right; the second vehicle then changes lanes to the left. The input video sequence is a clip of a long video sequence; thus, our system detected only the second passing vehicle at the beginning of the input video sequence. The system later sensed the second
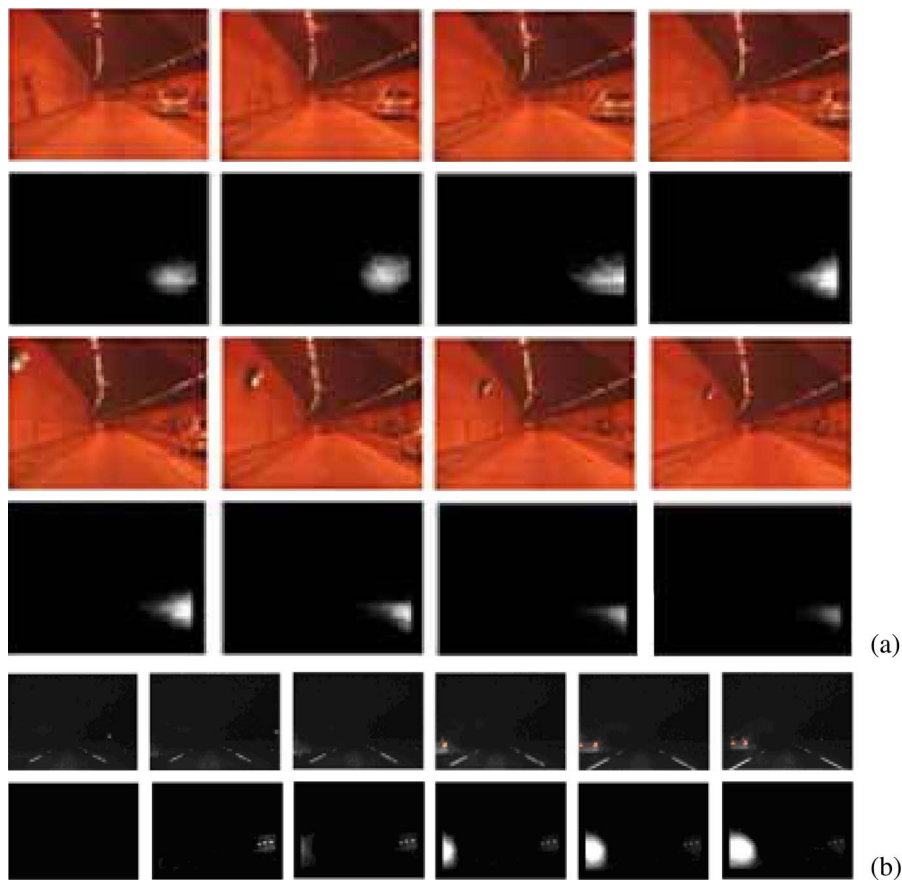
Fig. 12. Experimental results with the input video sequences taken (a) in a tunnel and (b) at night.
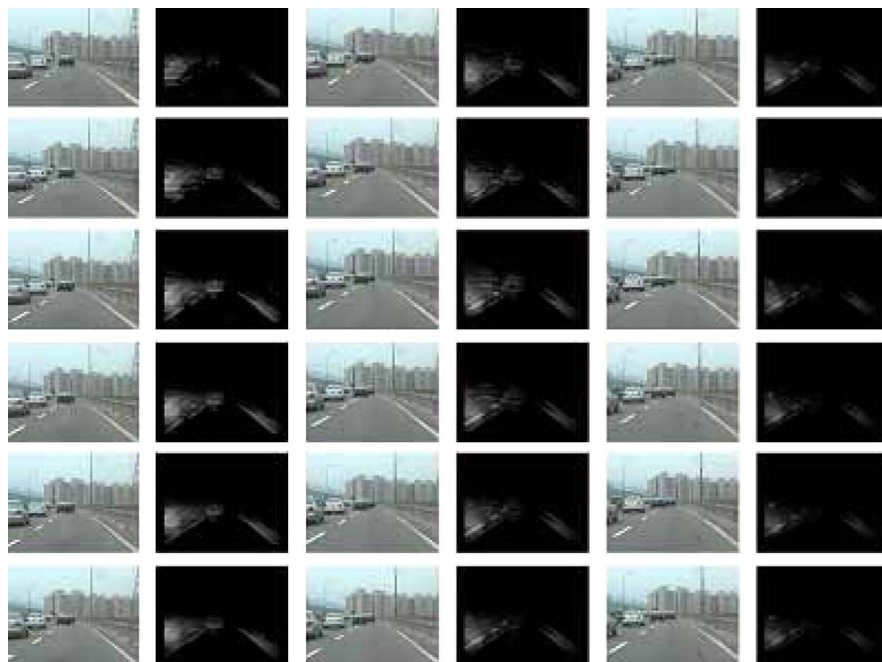

Fig. 13. Vehicle in front of the observer vehicle changes lanes to the left, whereas a vehicle in the left lane slows down. The system detected only the motion behavior of the latter vehicle.

passing vehicle changing lanes to the left. Immediately afterward, the system indicated a vehicle passing on the right. However, this vehicle is actually the vehicle that had passed earlier. It emerged again after the second passing vehicle changed lanes to the left. The emerged vehicle generated an attention pattern. Although the pattern was weak, it overlapped that of the
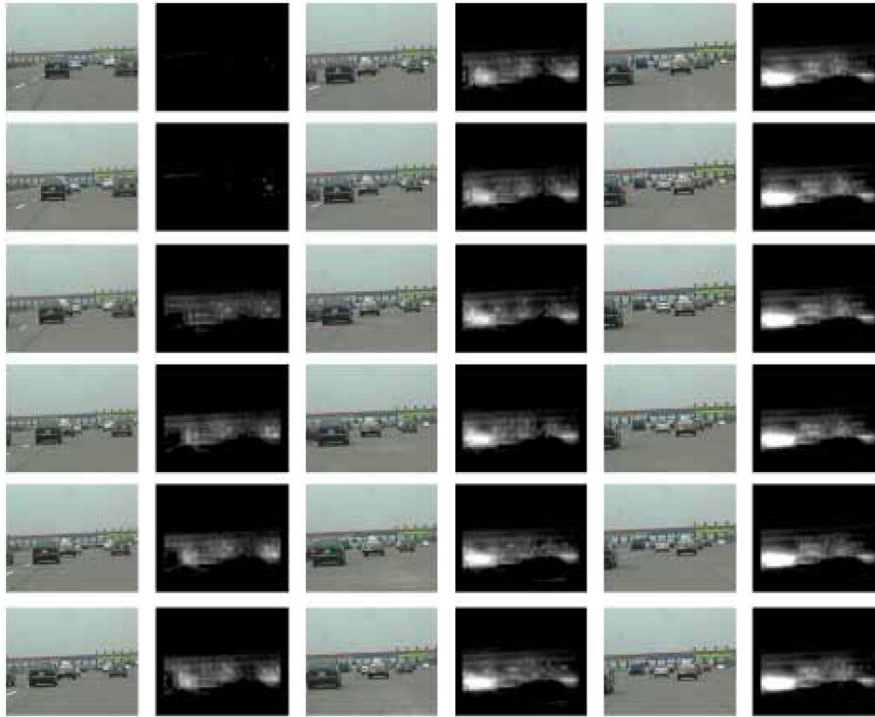
Fig. 14. Two preceding vehicles change lanes successively. The first vehicle changes lanes to the left, and the second vehicle changes lanes to the right.
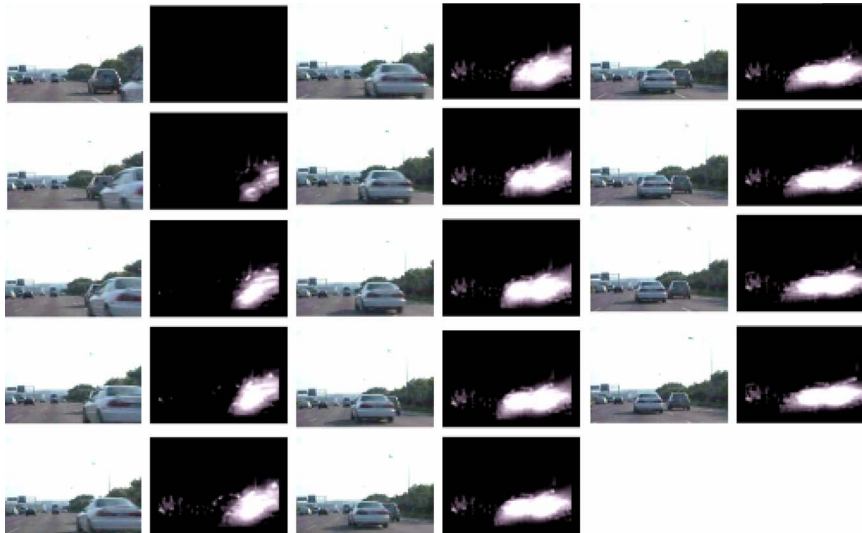

Fig. 15. Two vehicles successively pass the observer vehicle from its right. The second vehicle changes lanes to the left. Our system signaled two correct warnings and one false alarm in this input video sequence.

second passing vehicle. The resulting attention pattern tricked the system out of a right passing vehicle. In this example, the motion behaviors of the second passing vehicle were correctly recognized, whereas the first passing vehicle resulted in a false alarm.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have developed a system of detecting critical motions of nearby moving vehicles while driving on a highway. The computational framework of the system was characterized by a computational model called the DVM, which performs visual analysis based on video sequences. Video sequences have been known to suffer from the inherent drawbacks of instability, blurring, and color distortion. The DVM possesses a number of intriguing features that enable the system to get around the drawbacks of video sequences.

The abilities of coping with diverse weather (e.g., sunny, foggy, cloudy, rainy, and snowing) and illumination (e.g., dawn, sunrise, sunshine, dusk, and nighttime) conditions are also essentially due to the DVM. Different weather and illumination conditions lead to different intensity values of the input video images. The DVM detects vehicle motions based on the DIF image that is the difference between the MIN and MAX images. The effect of global illumination is eliminated during image differencing.

The proposed system recognizes critical motions of nearby vehicles by matching the attention patterns generated in the STA neural network against the principal attention patterns prestored in the ART2s through an offline supervised learning process. Although the ART2s can perform online unsupervised learning, their plasticity to significant patterns and stability to irrelevant patterns heavily rely on the quality of the attention patterns provided by the STA neural network. Increasing either the learning ability of the ART2 neural network, the detecting power of the STA neural network, or both would improve the performance of our system.

Our current system has performed well on single critical motions but was sometimes confused with multiple critical motions. Better strategies for resolving the pattern overlapping issue would improve the performance of multiple critical motion detection. Although the current system cannot distinguish between false and actual alarm situations, it is feasible to rate alarm cases according to potential dangers [12]. Currently, only template patterns of small vehicles were considered. To deal with large vehicles (e.g., buses, trucks, and towers), the associated template patterns should be included. However, this will inevitably worsen the pattern-overlapping issue.

References

[1] 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

[2] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Trans. Image Process.*, vol. 7, no. 1, pp. 62–81, Jan. 1998.

[3] M. Bertozzi, A. Broggi, A. Fascioli, and S. Niclele, "Stereo vision-based vehicle detection," in *Proc. IEEE Symp. Intell. Veh.*, Detroit, MI, 2000, pp. 39–44.

[4] M. Bertozzi, A. Fascioli, and A. Broggi, "Performance analysis of a low-cost solution to vision-based obstacle detection," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Tokyo, Japan, 1999, pp. 350–355.

[5] G. A. Carpenter and S. Grossberg, *Pattern Recognition by Self-Organizing Neural Networks*. Cambridge, MA: MIT Press, 1991, pp. 397–423.

[6] H. Cheng, N. Zheng, X. Zhang, J. Qin, and H. van de Wetering, "Interactive road situation analysis for driver assistance and safety warning systems: Framework and algorithms," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 1, pp. 157–167, Mar. 2007.

[7] S. Denasi and G. Quaglia, "Obstacle detection using a deformable model of vehicles," in *Proc. IEEE Symp. Intell. Veh.*, Tokyo, Japan, 2001, pp. 145–150.

[8] M. P. Dubuisson, S. Lakshmanan, and A. K. Jain, "Vehicle segmentation and classification using deformable templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 3, pp. 293–308, Mar. 1996.

[9] A. Eidehall, J. Pohl, F. Gustafsson, and J. Ekmark, "Toward autonomous collision avoidance by steering," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 1, pp. 84–94, Mar. 2007.

[10] S. Ernst, C. Stiller, J. Goldbeck, and C. Roessig, "Camera calibration for lane and obstacle detection," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, 1999, pp. 356–361.

[11] C. Y. Fang, C. S. Fuh, S. W. Chen, and P. S. Yen, "A road sign recognition system based on dynamic visual model," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, Madison, WI, 2003, vol. 1, pp. 750–755.

[12] C. Y. Fang, H. L. Hsueh, and S. W. Chen, "Dangerous driving event analysis system by a cascaded fuzzy reasoning Petri net," in *Proc. IEEE 9th Int. Conf. Intell. Transp. Syst.*, Toronto, ON, Canada, Sep. 17–20, 2006, pp. 337–342.

[13] G. S. K. Fung, N. H. C. Yung, and G. K. H. Pang, "Vehicle shape approximation from motion for visual traffic surveillance," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Oakland, CA, 2001, pp. 610–615.

[14] D. Ghica, S. W. Lu, and X. Yuan, "Recognition of traffic signs by artificial neural network," in *Proc. IEEE Int. Conf. Neural Netw.*, Perth, Australia, 1995, vol. 3, pp. 1444–1449.

[15] R. L. Gregory, *Eye and Brain*, 3rd ed. New York: McGraw-Hill, 1978.

[16] A. Hurlbert and T. Poggio, "Visual information: Do computers need attention," *Nature*, vol. 321, no. 6071, pp. 651–652, Jun. 1986.

[17] T. Isogai, "On a measure of multivariate skewness and a test for multivariate normality," *Ann. Inst. Stat. Math.*, vol. 34, no. 1, pp. 531–541, Dec. 1982.

[18] K. Kaliyaperumal, S. Lakshmanan, and K. Kluge, "An algorithm for detecting roads and obstacles in radar images," *IEEE Trans. Veh. Technol.*, vol. 50, no. 1, pp. 170–182, Jan. 2001.

[19] T. Kato, Y. Ninomiya, and I. Kato, "An obstacle detection method by fusion of radar and motion stereo," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 3, pp. 182–188, Sep. 2002.

[20] L. A. Klein, *Sensor Technologies and Data Requirements for ITS*. Boston, MA: Artech House, 2001, pp. 306–307.

[21] D. Koller, J. Weber, and J. Malik, "Robust multiple car tracking with occlusion reasoning," Univ. California at Berkeley, Berkeley, CA, Tech. Rep. UCB/CSD-93-780, 1993.

[22] D. Koller, J. Weber, and J. Malik, "Towards real time visual based tracking in cluttered traffic scenes," in *Proc. Intell. Veh. Symp.*, Paris, France, Oct. 24–26, 1994, pp. 201–206.

[23] A. Kuehnle, "Symmetric-based vehicle location for AHS," in *Proc. SPIE Transp. Sensors Controls: Collision Avoidance. Traffic Manag., ITS*, Orlando, FL, 1998, vol. 2902, pp. 19–27.

[24] S. Kyo, T. Koga, K. Sakurai, and S. Okazaki, "A robust vehicle detection and tracking system for wet weather conditions using the IMAP-vision image processing board," in *Proc. IEEE Intell. Transp. Syst.*, Tokyo, Japan, 1999, pp. 423–428.

[25] M. W. Levine and J. M. Shefner, *Fundamentals of Sensation and Perception*, 2nd ed. Pacific Grove, CA: Brooks/Cole, 1991, ch. 14.

[26] M. Lützeler and E. D. Dickmanns, "Road recognition with MarVEye," in *Proc. IEEE Symp. Intell. Veh.*, Stuttgart, Germany, 1998, pp. 341–346.

[27] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970.

[28] C. Martindale, *Cognitive Psychology: A Neural-Network Approach*. Pacific Grove, CA: Brook/Cole, 1991.

[29] Y. Matsushita, S. Kamijo, K. Ikeuchi, and M. Sakauchi, "Image processing based incident detection at intersections," in *Proc. 4th Asian Conf. Comput. Vis.*, 2000, pp. 520–527.

[30] D. Mitrovic, "Reliable method for driving events recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 198–205, Jun. 2005.

[31] A. Polychronopoulos, M. Tsogas, A. J. Amditis, and L. Andreone, "Sensor fusion for predicting vehicles' path for collision avoidance systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 549–562, Sep. 2007.

[32] Y. Ruichek, "Multilevel- and neural-network-based stereo-matching method for real-time obstacle detection using linear cameras," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 1, pp. 54–62, Mar. 2005.

[33] 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 No. 2121.

[34] R. Sekuler and R. Blake, *Perception*. New York: Knopf, 1985.

[35] N. Shimomura, K. Fujimoto, T. Oki, and H. Muro, "Depth-based target segmentation for intelligent vehicles: Fusion of radar and binocular stereo," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 3, pp. 196–202, Sep. 2002.

[36] R. S. Snell, *Clinical Neuroanatomy for Medical Students*, 3rd ed. Boston, MA: Little, Brown, 1992.

[37] M. Sugeno, "Fuzzy measure and fuzzy integrals: A survey," in *Fuzzy Automatic and Decision Processes*, M. M. Gupta, G. N. Saridis, and B. R. Gaines, Eds. Amsterdam, The Netherlands: North-Holland, 1977, pp. 89–102.

[38] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection using evolutionary Gabor filter optimization," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 1, pp. 43–53, Jun. 2005.

[39] 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 Congr., Detroit, MI, 1999.

[40] T. Viéville, E. Clergue, and P. E. D. S. Facao, "Computation of ego-motion and structure from visual and internal sensors using the vertical cue," in *Proc. IEEE Int. Conf. Comput. Vis.*, Berlin, Germany, 1993, pp. 591–598.

[41] Z. Wang and G. J. Klir, *Fuzzy Measure Theory*. New York: Plenum, 1992.

[42] D. Willersinn and W. Enkelmann, "Robust obstacle detection and tracking by motion analysis," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, 1997, pp. 356–361.

[43] M. Yamada, K. Ueda, I. Horiba, and N. Sugie, "Discrimination of the road condition toward understanding of vehicle driving environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 2, no. 1, pp. 26–34, Mar. 2001.

[44] Y. Zhu, D. Comaniciu, M. Pellkofer, and T. Koehler, "Reliable detection of overtaking vehicles using robust information fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 4, pp. 401–414, Dec. 2006.

[45] H. J. Zimmermann, *Fuzzy Set Theory and Its Applications*, 2nd ed. Boston, MA: Kluwer, 1991.

Shen Cherng was born in Taipei, Taiwan, in 1955. He received the M.S. degree in physics from Tsinghua University, Hsinchu, Taiwan, in 1981, the M.S. degree in civil engineering from Michigan State University, East Lansing, in 1985, and the Ph.D. degree in biomedical science from Oakland University, Rochester, MI, in 1993.

From 1986 to 1989, he was a Research Assistant with the Department of Neurology, Henry Ford Hospital, Detroit, MI, working on medical physics. Then for eight years, he was a Professional Engineer with the Department of Management and Budget, Lansing, MI. In 2002, he joined the faculty of the Department of Electrical Engineering, Chengshiu University, Kaohsiung, Taiwan, where he is currently an Associate Professor with the Department of Computer Science and Information Engineering. His research interests include antenna design, radio-frequency identification, and pattern recognizing. He is the author or coauthor of more than 20 articles in scientific journals, conference proceedings, and books.

Mr. Cherng is a registered Professional Engineer in the State of Michigan.

Chiung-Yao Fang received the B.Sc. and M.Sc. degrees in information and computer education from the National Taiwan Normal University, Taipei, Taiwan, in 1992 and 1994, respectively, and the Ph.D. degree in computer science information engineering from the National Taiwan University, Taipei, in 2003.

From 2004 to 2006, she was an Associate Professor with the Department of Information and Computer Education, National Taiwan Normal University, where she is currently an Associate Professor with the Laboratory for Computer Vision and Image Processing, Department of Computer Science and Information Engineering. Her research interests include neural networks, fuzzy systems, pattern recognition, and computer vision.

Chia-Pei Chen was born in Taipei, Taiwan, on March 23, 1977. She received the B.S. degree in computer science and information engineering from Tunghai University, Taichung, Taiwan, in 1999 and the M.S. degree in computer science and information engineering from the National Taiwan Normal University, Taipei, in 2003.

From 2003 to 2007, she was a Firmware Engineer with the Phoenixtec Power Company, Taipei. Since 2007, she has been an Assistant Researcher with the Chunghwa Telecom Laboratories, Taoyuan, Taiwan. Her research interests include pattern recognition, image processing, 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 the National Central University, Taoyuan, Taiwan, in 1974 and 1976, respectively, and the M.Sc. and Ph.D. degrees in computer science and engineering from Michigan State University, East Lansing, in 1985 and 1989, respectively.

In 1990, he was a Researcher with the Advanced Technology Center, Computer and Communication Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan. He is currently a Full Professor with the Laboratory for Computer Vision and Image Processing, Department of Computer Science and Information Engineering, National Taiwan Normal University, Taipei, Taiwan. His research interests include pattern recognition, image processing, computer vision, and intelligent transportation systems.