A System to Detect Complex Motion of Nearby Vehicles on Freeways

C. Y. Fang, C. P. Chen, C. S. Fuh, and S. W. Chen

Abstract—**We propose an improved computational model motivated by human cognitive processes for detecting changes of driving environments. The model, called dynamic visual model (DVM), consists of three major components: sensory, perceptual, and conceptual components. The proposed model is used as the underlying framework in which a system for detecting complex motions of nearby moving vehicles on a freeway is developed.**

Index Terms—**Motion Detection, Cognitive Process, Neural Networks, Fuzzy Integral.**

I. INTRODUCTION

Our objective in this study is to detect the motion of dynamic obstacles on the expressway to warn drivers so as to avoid possible traffic accidents. We first focus on detecting the motion of nearby moving vehicles. Some complex situations including more than one motion occurring synchronously or asynchronously are considers here. In the future, this system will be extended to detect other static and dynamic obstacles.

Most of the complex motions of the vehicles while driving on a freeway are combined by two kinds of motions: lane change (right lane change and left lane change) and speed change (speed-up and slow-down). For example, to overtake a vehicle ahead, drivers may first speed up, change their lane from right to left, pass the vehicle, and then change back to the original lane. In practice, the vehicle motions are permuted sequences of these motions.

Some permuted sequences of the motions of nearby vehicles are dangerous and some are not. For example, "a vehicle ahead slows down" is a dangerous behavior, but "a vehicle ahead accelerates" is not. Before detecting dangerous vehicle motions, Nine dangerous motions of only one nearby vehicle are collected by the above process. These motions include (a) Vehicle ahead slows down. (b) Right front vehicle changes lane to the left. (c) Left front vehicle changes lane to the right. (d) Vehicle ahead changes lane to the right. (e) Vehicle ahead changes lane to the left. (f) Left front vehicle slows down. (g) Right front vehicle slows down. (h) Vehicle on the left accelerates. (i) Vehicle on the right accelerates. The closer to our vehicle another vehicle moves on the freeway, the more complex and dangerous the driving environment will be.

Some motions of our vehicle itself are also dangerous, including lane change and speed change. Speed change of our vehicle can be treated as relative to other vehicles, which has been analyzed in the above discussion. However, a lane change by our vehicle should be further considered. Four conditions are of concern: beginning and end of move to left lane for our vehicle, and beginning and end of move to right lane for our vehicle.

If there are many nearby vehicles on the freeway, the above dangerous motions could occur simultaneously because vehicles move independently of each other. For example, "vehicle ahead slows down" and "vehicle on the right passes our vehicle" may occur at the same time. In this study, we propose a method to analyze and detect dangerous simultaneous motions of nearby vehicles.

II. THE IMPROVED DVM IMPLEMENTATION

A. The Implement Problems of our original DVM

We have been developed a DVM model to detect the change detection of driving environmental changes [1] and the road signs [2] successfully. And we try to develop nearby vehicle motion detection system based on our DVM model. However, it is very difficult to recognize various complex motions. Thus we proposed an improved DVM model here.

B. The Improved DVM Model

Figure 1 depicts the improved proposed DVM, which captures several aspects of the human visual process [3]. The proposed model is comprised of three major components, the sensory, perceptual, and conceptual analyzers of the human visual system.

The spatiotemporal information is first extracted by the method presented in [1] and is input to the STA neural module to form the attention maps. These attention maps are then divided into several overlapping blocks for extracting the local categorical features. These features are then input to the corresponding CART neural model of the blocks for motion classification. The outputs of a CART neural model in successive maps are integrated by a modified fuzzy integral technique for producing more confidence classification results. Such a technique has been modified to be able to integrate the temporal information. The output of the modified technique is not a number, but a set of the highly possible motions. Finally, since all the CARTs may output their classification results, and

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).

C. P. Chen and S. W. Chen are with Institute 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.

the resulting output by a CART neural model may be more than one, all the classification results should be collected to provide correct motion detection results for drivers.

The details of STA neural module, CART neural module can be accessed to our previous works [1, 2]. Figure 2 shows an example of "a nearby vehicle ahead slows down," including the input frames and their corresponding attention maps of STA neural model. The attention maps of nine dangerous motions of one nearby vehicle are shown in Fig. 3, and the attention maps of the motions of a moving vehicle are shown in Fig. 4.

Here we introduce the following improvements, including attention map partition, categorical feature extraction, temporal integral process, and the collection of classification results.

1) Attention Map Partition

In order to use many CART neural networks to recognize the complex motions we should first partition the attention map. In our experiments, we divide the attention maps into five partially overlapping blocks. Each block corresponds to a window, shown in Fig. 5 (a). The boundaries of these five windows are colored cyan (b1), yellow (b2), blue (b3), green (b4), and red (b5), as shown by the example of Fig. 5 (b).

This design of the partition is based on the needs of our application. The block covered by b_1 contains the main region of our vehicle's lane ahead. The driver should be especially vigilant to dangerous motions of nearby vehicles occurring in this window, for example when the vehicle ahead slows down. The blocks covered by b_2 and b_3 contain the lanes on the left and right, respectively, relative to our vehicle. The motions (e.g., the vehicle on the right accelerates) occurring in these windows are potentially dangerous to our vehicle. Finally, b_4 and b_5 cover regions to which drivers may pay less attention since motions occurring in these windows would have only a slight effect on our traffic safety.

Every CART neural network has the job to recognize the motions occurring within its window. Each CART neural network can recognize only one motion shown in the input attention map. However, the simultaneous motions can be recognized individually and successfully if they occur in different windows. Our system can detect at most five motions in an attention map at the same time.

If the attention maps are divided into non-overlapping spatial windows, then some motions will not be detected successfully. Such behaviors possibly occur at the boundary of a window, rather than inside a window. If the windows overlap and their sizes are large enough so that each motion is covered with at least one window, then the boundary problem will not happen. Therefore, overlapping block segmentation is necessary, and the window size and degree of overlap depend on the nature of an application.

2) Categorical Feature Extraction

Based on these partitioned attention maps, categorical features are extracted for classification. We think that skewness may be a suitable feature. First, each partitioned attention map is divided into ten equal horizontal regions, as shown in Fig. 6 (a). For each region (whose width is M), the mean intensity values for each columnare calculated and denoted by $\mathbf{c}_i = \{c_{i1}, c_{i2}, ..., c_{iM}\}$, $i = 1, 2, ..., 10$.

Let $\bar{\mathbf{c}}_i$ be the mean horizontal position of the intensity means, $c_{i1}, c_{i2}, ..., c_{iM}$, in the ith region, and m_{i2}, m_{i3} be the normalized second and third moments, respectively. The skewness of intensity is

$$g_{i1} = \frac{m_{i3}}{m_{i2}\sqrt{m_{i2}}}, \text{ where } \bar{\mathbf{c}}_i = \frac{\sum_{j=1}^{M} j c_{ij}}{\sum_{j=1}^{M} c_{ij}},\ m_{i2} = \frac{\sum_{j=1}^{M}\left(jc_{ij} - \bar{\mathbf{c}}_i\right)^2}{\sum_{j=1}^{M} c_{ij}},$$

$$\text{and } m_{i3} = \frac{\sum_{j=1}^{M}\left(jc_{ij} - \bar{\mathbf{c}}_i\right)^3}{\sum_{j=1}^{M} c_{ij}}.$$

Figure 6 (b) shows the ten skewness features of the partitioned attention map shown in Fig. 6 (a) for each horizontal division. Using the same method, we can calculate the vertical skewness features from the same partitioned attention map.

The above 20 features of divided window b_i are arranged as a feature vector, $(f'^1_i, f'^2_i, \cdots, f'^m_i)$ (where m = 20), to be input to CART, neural network. The values of these features will fall into the interval $(-\infty, \infty)$. Since our CART neural networks accept only positive real numbers, these values should be mapped into interval [0, 1] by

$$f_i^m = \left(\frac{f'^m_i}{T} + 1\right)\Big/2, \text{ where } T = \sum_{m=1}^{20} f'^m_i .$$

These features, $(f_i^1, f_i^2, \cdots, f_i^m)$, are then fed into their corresponding CART neural networks (CART$_i$) and, using the method to be introduced in next section, the decision whether a complex motion has been detected will be made even if more than one nearby vehicle moves relative to ours.

3) The Temporal Integral Process for a CART

The fuzzy integral [4] is a suitable technique for integrating the classification results. It is a nonlinear numerical approach for integrating multiple sources of uncertain information or evidence to arrive at a value which expresses the degree of confidence in a particular hypothesis or decision. To apply this technique to our application, the confidence function and the fuzzy measure function are of concern.

Since the input to a conceptual component is a sequence of attention maps, each CART neural network will output a successive string of classification results. Suppose the attention map is first segmented into partial overlapping blocks through attention map partition, where each block corresponds to a window. The categorical features extracted from window i form a feature vector $(f_i^1, f_i^2, \cdots, f_i^m)$. This vector is then fed

0-7803-8125-4/03/$17.00 © 2003 IEEE

into its corresponding CART neural network, CART_i, for classification. The outputs of the CART neural networks are shown in Fig. 7. Here we suppose that in CART_i the focus of attention map is formed at time $t-r_i+1$, and the time is regarded as the beginning of the motion. Without loss of generality, we discuss only the fuzzy integral process of CART_i; the integral process of the other CARTs is similar.

The output string of labels of CART_i from time $t-r_i$ to t can be represented by $s_i^{t-r_i+1}s_i^{t-r_i+2}...s_i^t$, where r_i is the frame number. Let $L=\{l^0,l^1,...,l^{r_i}\}$ denote the set of all labels recognized by CART_i, including a null label l^0, and let each label l^k in L have its

corresponding stored pattern p^k. The labels output from the CART_i neural network all belong to this set, so that $s_i^{t-r_i+j}\in L$, j = 1, 2, ..., r_i. Here set L is the same as the sources of information in the fuzzy integral technique.

Let $h(\cdot)$ and $g(\cdot)$ be a confidence function and a fuzzy measure function, respectively. Function $h: L \rightarrow [0,1]$ is defined on set L and function $g: 2^L \rightarrow [0,1]$ is defined over the power set of L.

Both the definitions of confidence function and fuzzy measure function depend on their application. In our system, the confidence function is based on $s_i^{t-r_i+j}$, the label output by CART_i at time $t-r_i-j$, measured by other outputs is defined as

$$h^{s_i^{t-r_i+j}}(s_i^{t-r_i+j})=\begin{cases} w(j,k) & \text{if } l^j=l^k \\ \dfrac{w'(j,k)}{\left\|p^j-p^k\right\|} & \text{otherwise,} \end{cases}$$ where

$$w(j,k)=\frac{1}{1+e^{-\alpha|r_i-j||r_i-k|}} \text{ and } w'(j,k)=\frac{1}{1+e^{-\beta|r_i-j||r_i-k|}}.$$

Distance $\left\|p^j-p^k\right\|$ is the city block (L_1) distance between patterns p^j and p^k, and $l^j, l^k \in L$, where j, k = 1, 2,..., r_i. Symbols α and β are positive parameters.

Let the number of non-zero pixels of one stored pattern be $\#p$, and the number of such pixels falling in the union of windows i or j be $\#p_{(i,j)}$, then the measure function can be calculated by

$$g(S)=\frac{\#p_S}{\#p}, \text{ where } S\in 2^L.$$

Finally, the fuzzy integral for $s_i^{t-r_i+j}$ is defined by

$$e^{s_i^{t-r_i+j}}=\int h^{s_i^{t-r_i+j}}(s)\circ g=\sup_{\alpha\in[0,1]}\{\alpha\wedge g(F_\alpha)\},$$

where $\wedge$ indicates the fuzzy intersection characterized by a t-norm and $F_\alpha=\{s\mid h^{s_i^{t-r_i+j}}(s)\geq\alpha\}$.

The intermediate decision set of individual CART_i includes those labels whose integral is maximum or very close to the maximum. We define a set $\mathbf{d}_i^t$ which includes the possible category candidates selected by CART_i.

$$\mathbf{d}_i^t=\{l_i^s \mid \text{the corresponding label of } \arg(e^s \mid \left|e^s-e^{\max}\right|\leq\Gamma)\}$$

where $e^{\max}=\max_{j=1,...,r_i}\{e^{s_i^{t-r_i+j}}\}$ and Γ is a distance threshold.

Since the CART can usually classify the input attention maps correctly, most of the time there is only one candidate in the set. However, sometimes the set contains more than one candidate, which indicates that the CART gets too little information to clearly classify the map. In this case, the results of other CARTs should be consulted to make a final classification. Therefore, the system needs to collect the classification results of the CARTs.

4) Collection of Classification Results

Referring to Fig. 7, suppose that the intermediate decision set of CART_i includes m_i elements. Then $\mathbf{d}_i^t=\{l_i^1,l_i^2,...,l_i^{m_i}\}$, here l_i^j is the output label of CART_i in the intermediate decision set, for i = 1, 2, ..., n, and j = 1, 2, ..., m_i. Then the final decision set is

$$\mathbf{D}^t=\{l \mid l=\arg(f(l'))\ \text{where } f(l')>\Gamma',\ \text{for all } l'\in L\},$$

where $f(l')=\bigvee_{\substack{i=1,2,...,n;\\ j=1,2,...,m_i}}(\delta(l',l_i^j)e_i^j)$, $l_i^j\in\mathbf{d}_i^t$, and e_i^j is the corresponding integral value of l_i^j. Function

$$\delta(l',l'')=\begin{cases}1 & \text{if } l'=l'' \\ 0 & \text{otherwise}\end{cases},$$

and Γ' is a threshold to filter the weaker candidates.

Set $\mathbf{D}^t$ includes all the motions of nearby vehicles detected by the CARTs at time t. The number of labels in set $\mathbf{D}^t$ indicates the number of nearby vehicle motions. If more than one label is contained in set $\mathbf{D}^t$, then a complex motion occurs at time t.

III. Experimental Results and Discussions

The input data to our system was acquired using a video camcorder mounted in the front windshield of a vehicle while driving on the expressways. In our experiments, each video sequence was down-sampled to a rate of 5 frames per second before being sent to the system. Furthermore, each 640 x 480 pixel input image was reduced to 160 x 120 pixels by sub-sampling. We downsample input video sequences for the purpose of reducing the processing load on the computer. Likewise, we subsample video images for reducing the processing time.

A. Experimental Results

The CART neural networks should be well trained before being tested. The training attention maps of CARTs are shown in Figs. 3 and 4. In our experiments from each category of motion we select only one representative map for training. These maps, selected from real sequences, are partitioned and then the categorical features of each block are extracted to train its corresponding CART neural network.

Fig. 8 shows the first experimental example. The input attention maps, shown in Fig. 8 (a), are created by a simple motion sequence: a vehicle ahead changes lane to the left. Fig. 8 (b) shows the integral values according to the outputs of different CARTs. The motion of the vehicle in Fig. 8 (a) corresponds to Table 1, row (e). As can be seen in the table, CART has a significantly higher value for g (0.86), indicating a higher likelihood for "vehicle ahead changes lane to the left." This then is the decision made by our system.

Fig. 9 shows another example of the sequence---a vehicle on the right lane changes lane to the left. In this case, since the motion patterns of this kind of motions cover a large region of the attention map. Both $CART_3$ and $CART_5$ are important, because their measure function values are 0.8633 and 0.5398 (shown in Table 1,motion case (b)). However, since the vehicle gradually moves into the region of CART1, its integral value (see Fig. 9 (b) frames 8 and 9) also increases. All the outputs of the CARTs are integrated by our system for determining the final result. We can observe that while even camera vibration and vehicle motion make the integral values of CARTs in some frames decrease, these incorrect temporal results do not greatly affect the final results of our system. Our system is then robust with respect to vibration and motion.

Twelve simple motion sequences (328 frames in total) have been tested in the experiments. Most of the results are correct (97.9%), and only seven frames are incorrect. This is enough information for our system to decide which motion occurred. Actually, our system only outputs a result for each input sequence, as shown Fig. 10. Notice that we select only one training map in each motion category, making total number of training maps thirteen. If we select more training maps, the accuracy rate will increase.

Collecting real image sequences including the complex motions is difficult, so we simulate them by overlapping two or more simple motions in sequences. The overlapped sequences are generated as shown in Fig. 11. The first and second rows show a simple motion sequence, and the third row shows their overlapped results. The complex motion sequence is "a left front vehicle changes lane to the right," and "a vehicle on the right accelerates." Using this method, we can overlap two simple motion sequences with any time difference and generate numerous reasonable complex motions.

The overlapped sequence is different from the simple motion sequences themselves because sometimes they interfere with each other by overlapping. Thus complex motion sequences are more difficult to classify. Eighteen complex motion sequences (1074 frames in total) have been generated and tested in our experiment. Most of the results are correct (93.3%), and only 36 frames are incorrect. We will collect more real image sequences to test our system in the future.

B. Discussions

Figure 12 shows an example of "a nearby vehicle ahead changes lane to the right." Since in this case our vehicle keeps a safety distance from that vehicle, its motion is not dangerous. We can observe that the attention map is weakly stimulated. This kind of motion will not be detected by our system.

There are many other dynamic objects, including pedestrians, bicycles, and motorcycles, that should be detected. We think the simple motions of these objects may be easily detected by training additional maps in the CART neural networks. However, since these kinds of objects may move in groups, the complex motions of these objects seems very difficult to be correctly classified by this system.

Dividing the attention map into five regular overlapping blocks is not a prefect solution for vehicle motion detection because the motions of vehicles may occur anywhere on the road. A better method is to divide the map based on contours of focuses of attention. Each region of attention focus indicates a motion. If each motion region can be exactly detected and extracted from the attention map, then the classification performance of CARTs will increase. We are working on this topic using by a level set technique.

IV. Conclusions and Future Work

In this study, we presented a method to recognize the complex motions of nearby moving vehicles. First, we found the dangerous motions which may affect traffic safety. The corresponding attention maps of these motions are created and partitioned into five overlapping blocks. The skewness features of each partitioned attention map are then computed and stored in the long term memory of their corresponding CART neural networks.

When an input image sequence is fed into this system, our system extracts spatial and temporal information from the stimulus and outputs the activation. If the activation is too low, then nothing can strongly attract our attention, and our system waits for the next stimulus. Otherwise, the system outputs the focus of attention map, which is divided into five blocks to extract skewness features. These features are then input to their corresponding CART neural networks for classification. Using the temporal fuzzy integral, individual classifications are made by each CART*i*, and these results are collected to get the final result.

Acknowledgment

This work was supported by the National Science Council, Republic of China, under Contract NSC 90-2213-E-003-002. The authors gratefully acknowledge the assistance of Prof. Robert R. Bailey of National Taiwan Normal University for his many helpful suggestions in writing this paper and for editing the English.

References

[1] C. Y. Fang, S. W. Chen, and C. S. Fuh, "Automatic Change Detection of Driving Environments in a Vision-Based Driver Assistance System," *IEEE Transations on Neural Networks*, accepted at Oct. 28, 2002.

0-7803-8125-4/03/$17.00 © 2003 IEEE

[2] C. Y. Fang, S. W. Chen, and C. S. Fuh, "Road Sign Detection and Tracking," *IEEE Transations on Vehicular Technology*, accepted at Jan. 06, 2003.

[3] C. Martindale, *Cognitive Psychology---A Neural-Network Approach*, Brooks/Cole, Pacific Grove, California, pp. 95-116, 1991.

[4] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic Theory and Applications*, Prentice-Hall, 1995.

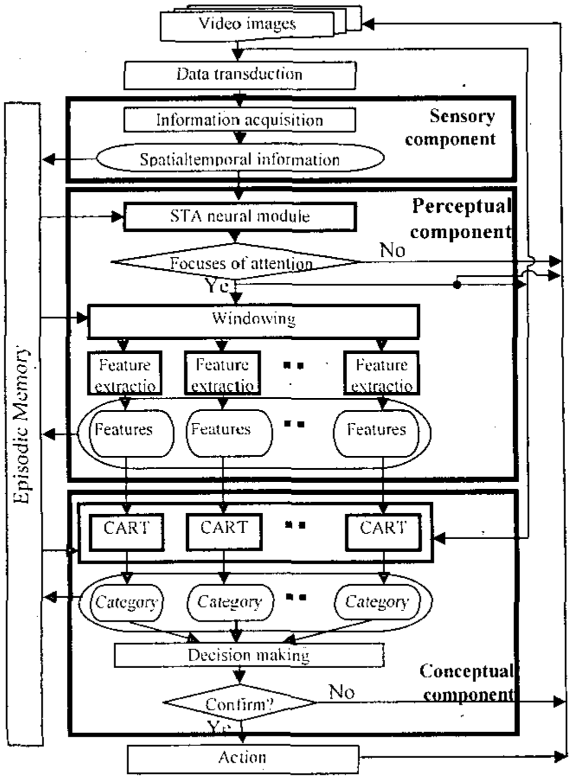


Fig. 1. The proposed improved DVM.

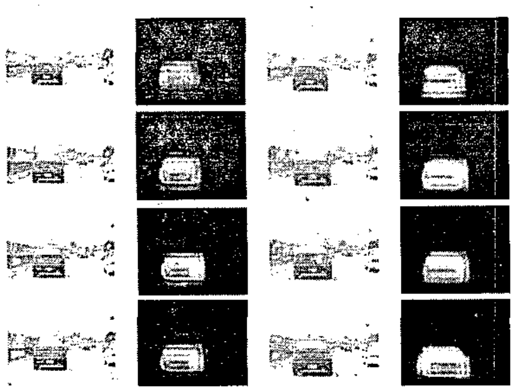


Fig. 2. A nearby vehicle ahead slows down.

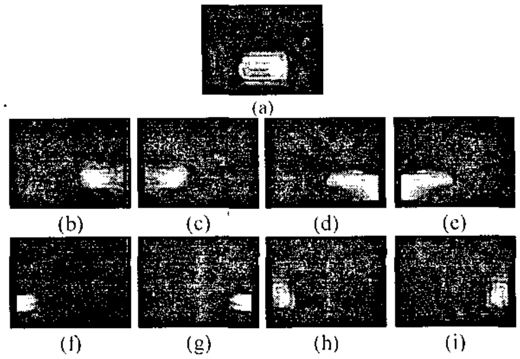


Fig. 3. Motions of a nearby moving vehicle. (a) Vehicle ahead slows down. (b) Right front vehicle changes lane to the left. (c) Left front vehicle changes lane to the right. (d) Vehicle ahead changes lane to the right. (e) Vehicle ahead changes lane to the left. (f) Left front vehicle slows down. (g) Right front vehicle slows down. (h) Vehicle on the left accelerates. (i) Vehicle on the right accelerates.

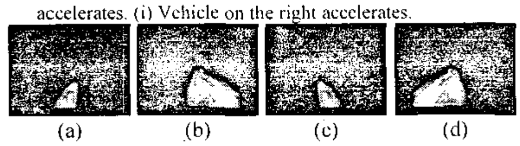


Fig. 4. Motions of a moving vehicle. (a) Beginning of move to left lane for our vehicle. (b) End of move to left lane for our vehicle. (c) Beginning of move to right lane for our vehicle. (d) End of move to right lane for our vehicle.

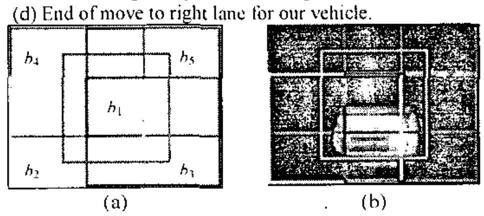


Fig. 5. The attention map partition. (a) The five divided windows whose colors are cyan (b_1), yellow (b_2), blue (b_3), green (b_4), and red (b_5). (b) An attention map divided by these five windows. (c) The five overlapping blocks.

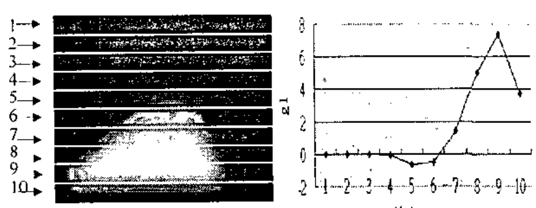


Fig. 6. An example of divisions. (a) The horizontal divisions. (b) The skewness features of the partitioned attention map, for each horizontal division.

Table 1. Some values of fuzzy measure function $g(\cdot)$.

Motion cases \ CARTs	$CART_1$	$CART_2$	$CART_3$	$CART_4$	$CART_5$
(a) Vehicle ahead slows down.	0.776	0.659	0.725	0.290	0.316
(b) Right front vehicle changes lane to the left.	0.372	0.069	0.863	0.042	0.539
(c) Left front vehicle changes lane to the right.	0.372	0.863	0.069	0.539	0.042
(d) Vehicle ahead changes lane to the right.	0.330	0.124	0.679	0.043	0.335
(e) Vehicle ahead changes lane to the left.	0.330	0.679	0.124	0.335	0.043

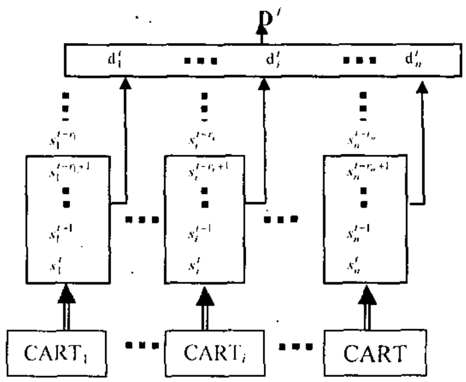
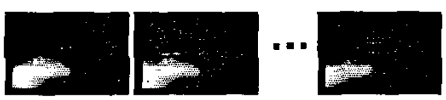
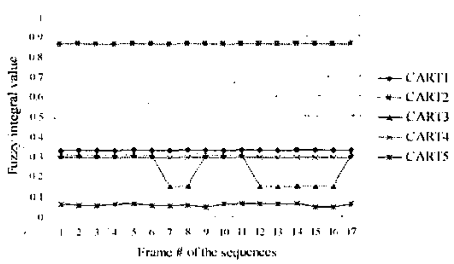


Fig. 7. The decision outputs from the CART neural networks.

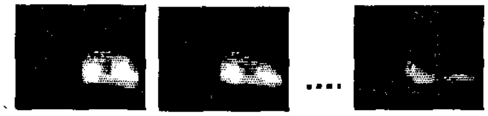


(a) Input attention maps.

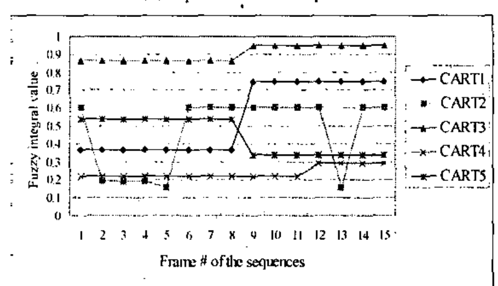


(b) Integral values according to the outputs of the CARTs.

Fig. 8. Example of "a vehicle ahead changes lane to the left."



(a) Input attention maps.



(b) Integral values according to the outputs of the CARTs.

Fig. 9. Example of "a vehicle on the right lane changes lane to the left."

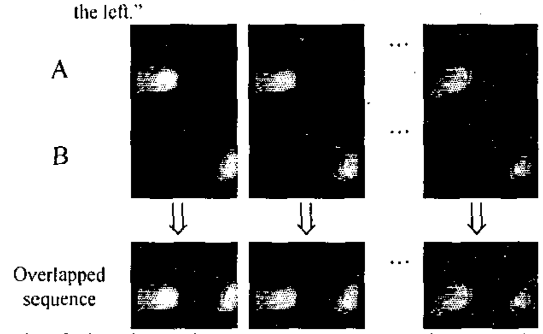


Fig. 10. Complex motion sequence overlapped by Sequences A and B.

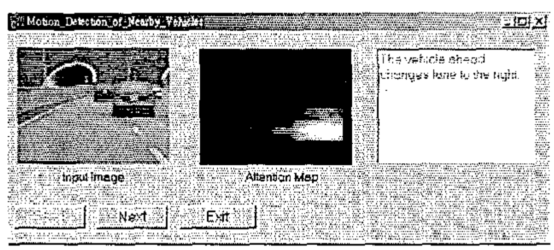


Fig. 11. Visual interface for reporting the results of nearby vehicle motion detection.

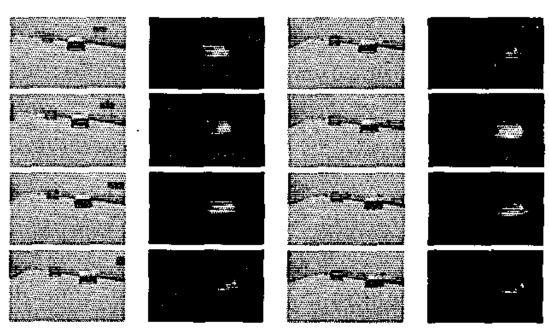


Fig. 12. A nearby vehicle ahead changes lane to the right in a safety distance.

0-7803-8125-4/03/$17.00 2003 IEEE