Environmental Change Detection System Regarding Roads

Chiung-Yao Fang(方瓊瑤)‡†, Sei-Wang Chen(陳世旺)†, and Chiou-Shann Fuh(傅楸善)‡

†Department of Information and Computer Education
National Taiwan Normal University
Taipei, Taiwan, R. O. C.
Tel:(886)2-23622841 Fax:(886)2-3512772
E-mail: violet@ice.ntnu.edu.tw

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

(NSC 89-2218-E-003-001)

ABSTRACT

The change detection of driving circumstances is an important matter for the driver assistance system. While the computational ability of computers has been exceedingly excellent, the change detection ability of computers cannot match that of human beings. Human detection can be attributed to both parallel processing and distributed representation. In this paper, we proposed a general computational framework for modeling the parallel processing and information representation of human nervous systems. Since artificial neural networks supply ways to simulate human brains, they are utilized in this study to implement our computational framework. For example, the attention map of human beings is simulated by a temporal self-organizing feature map (TSOM), and the categorical perception is modeled by a configurable adaptive resonance theory II (CART2). In this paper, the computational framework is utilized to develop a system for detecting environmental changes. The experimental results show that our method can work well.

Index Terms--- change detection, parallel processing, distributed information representation, attention map, perception, SOM neural network, ART neural network.

I. INTRODUCTION

To understand the driving environments of vehicles immediately and remind the driver paying attention to the change of driving environments to avoid traffic accidents is one of the main objectives of vision-based driver assistance system. For understanding the complex driving environments, the vision-based driver assistance system should combined many detection subsystems, including road detection, road sign detection, and obstacle detection subsystems [3, 4]. These subsystems are all important thus it is difficult to decide which one should operate at the moment. Moreover, to keep on operating these subsystems consumes more system sources and time than to invoke them only when necessary. For example, it would be better if the road sign and obstacle detection subsystems operate only when these objects have appeared in view. Thus, it is essential to develop a subsystem to decide if any other detection subsystem should operate now.

Since the change of driving environments is extremely complicated, in this paper, we concentrate on the change detection of road conditions in expressway. Our change detection subsystem could detect some environmental changes in expressway, such as lane-change, expressway-entry, expressway-exit, tunnel-entry, tunnel-exit and viaduct-ahead conditions.

Although the computational ability of computers has been exceedingly excellent, the detection ability of computers cannot match that of human. The efficiency of human detection can be attributed to the parallel processing ability and the distributed information representation of neural networks in brain. Distributed information representation indicates that human beings memorize their experiences on the connections between the neurons, not on the neurons themselves. Therefore, the memory capacity of human can be almost unlimited [1]. On the other hand, the property of parallel processing constitution allows human to associate the distributed information for an instant even if the memory database is extremely large. Under these special properties, the neurons in neural networks can simultaneously process their own received stimuli and propagate the outputs to others. This means information, scattered every-where in the cortex, can be interchanged immediately and in parallel through the connections between neurons. By way of the exchanged process, high-level information, including abstract concepts, can be constructed and supplied to our brain to analyze and understand the various complex situations especially those cannot

work out with the sequential process. Therefore, we designed a simulated framework to model the parallel detection process and the information representation of human brain.

II. COMPUTATIONAL FRAMEWORK

The flowchart of our general computational framework of environmental change detection system, developed based on the recognition process of human, is shown in Figure 1. One video sequence regarded as a continuous stimulus is fed into our system. First, the signal noises are removed and the image size is reduced in the preprocessing stage. Second, similar to the analyzers in human brain [1], our system extracts the spatial and temporal information from the stimulus and outputs the activation of cognitive units. If the activation of cognitive units is too low, nothing can strongly attract our attention, then our system waits for the following stimulus. Otherwise, the system outputs the attentional focus since the cognitive units have been already highly excited. This stage can be implemented by temporal self-organizing feature map (TSOM).

Human beings first take notice of the attractive objects, such as moving objects or bright-color objects, and then recognize them. For example, when we are driving, the road signs become closer to us from afar. We may first be attracted by their colors and shapes, and then recognize their meaning, such as no right turn signs, no left turn signs, and so on. In our system, the recognition process, called categorical perception, is simulated by configurable adaptive resonance theory (CART).

Similar to the structure of human analyzers, the output pattern from TSOM neural network can be regarded as an input supraliminal pattern of another subsystem. On the other hand, one subliminal expectation is associated from the long-term memory (LTM), which memorizes all the learned experiences of our system. However, instead of searching all experiences in the LTM, our system only looks through the suitable part in it. Where the subliminal expectation should be looked for in the LTM depends on the previous experiences of our system and the special characteristics of the input pattern. After the subliminal expectation has been found, it is compared with the supraliminal pattern. If they are exactly equal, then perception is undoubtedly successful. However, in most successful cases, the two patterns are only similar enough to each other. Our system will properly adapt the subliminal expectation in the LTM guided by the supraliminal pattern, and the adapting step is called supervised learning. If the two patterns are a mismatch, it means our system is in a situation that has never been seen before, the system should learn the new experience through unsupervised learning.

III. PREPROCESSING STAGE

The input data of our environmental change detection system are color video sequences recorded by a camcorder mounted on a moving vehicle. Under the influence of camcorder and vehicle jolt, the input images become highly unstable and the difficulty of condition change detection is greatly increased. Thus, we design a preprocessing step to surmount this difficulty.

The flowchart of the preprocessing stage is shown in Figure 2. One image $I(t)$ of an input video sequence, I, is fed into the road condition change detection system. First, the image $I'(t)$ is sub-sampled from image $I(t)$ to reduce the time complexity (320X240 → 160X120 pixels). Second, the low-intensity image $L(t)$ and the high-intensity image $H(t)$ are accumulated separately. The low-intensity image preserves the minimum values of corresponding pixels between images $I'(t)$ and $L(t\text{-}1)$, i.e. $L(t)=\min_{\text{pixel-wide}}(I'(t),L(t-1))$; $L(0)=I'(0)$. On the other hand, the high-intensity image preserves the maximum values of corresponding pixels between images $I'(t)$ and $H(t\text{-}1)$, i.e. $H(t)=\max_{\text{pixel-wide}}(I'(t),H(t-1))$; $H(0)=I'(0)$. Third, after subtracting the pixel values in image $L(t)$ from those on the corresponding positions in image $H(t)$, we get the difference image $D(t)$, i.e. $D(t)=H(t)-L(t)$. Fourth, the second difference image, $S(t)$, represents the absolute difference between images $D(t\text{-}1)$ and $D(t)$, i.e. $S(t)=|D(t)-D(t\text{-}1)|$. Finally, image $S(t)$ is then fed into TSOM neural network as the input stimulus to simulate the attention map. If the stimulus is strong enough to attract system attention (e.g. lane-change, tunnel-entry, and tunnel-exit), then the focus of attention is output to the CART subsystem. Moreover, the images $L(t)$, $H(t)$, and $D(t)$ are all reset for the next change detection. Otherwise, image $D(t\text{-}1)$ is replaced by $D(t)$, i.e. $D(t) \leftarrow D(t\text{-}1)$, and the system waits for the next image.

Figure 4 illustrates an experimental result of the preprocessing stage in our change detection subsystem. In Figure 4, column (a) shows ten images of a video sequence input to our system. The first image is on the top of this column and the following input images are arranged downward in temporal order. Columns (b) and (c) show the high-intensity and low-intensity image sequences respectively. To maintain these two image sequences is able to avoid the affect of camcorder and car jolt. Afterwards, the difference image sequence D calculated from H and L is in column (d). Finally, column (e) indicates the second difference image sequence S. The second difference images represent the change amount of driving environment from a time t to t+1. If there is no driving environment change in the short time period, then the second difference image $S(t)$ should be dark. In this example, image $S(t)$ can detect tunnel-entry, but is robust to jolt because lane markers will fall in side lane marker cluster during a jolt, not during tunnel-entry.

IV. TEMPORAL SELF-ORGNIZING FEATURE MAP

In our environmental change detection system, the input video sequences not only carry the spatial information in each image, but also hide the temporal information between the successive images. Thus, we create a temporal self-organizing feature map (TSOM), which can accept continuous stimuli and whose feature map can describe both spatial and temporal topological relations. Since TSOM is utilized to model the attention map in our brain, the feature map of SO layer can be called the attention map.

The TSOM neural network, as Figure 3 shows, is structured as a two-layer network: one input layer and one output layer. The output layer is more often referred to as an SO layer. Neurons on this layer are arranged into a 2D array in which neurons are interconnected to one another. These connections are called within-layer connections. There are no synaptic links among input neurons; they are, however, fully connected to the SO neurons. These connections are called between-layer connections. Between-layer connections are always excitatory, while within-layer connections are almost always inhibitory.

Suppose that the input layer of the neural network consists of m neurons and the output layer comprises n neurons. Let w_{ij} denote the strength of the link between output neuron i and input neuron j. The strength vector of output neuron i is written as $\mathbf{w}_i = (w_{i1}, w_{i2},\ldots,w_{im})$. The input to output neuron i due to innervation $\mathbf{x}$ at time t is defined by

$$I_i^v(t)=\sum_{j=1}^{m} w_{ij}\cdot x_j(t)$$

Let u_{ik} be the synaptic strength of the connection between neurons i and k with the respective positions $\mathbf{r}_i$ and $\mathbf{r}_k$. The input to output neuron i due to lateral interaction can be formulated as,

$$I_i^l(t)=\sum_{k\in N,k\neq i}\left[u_{ik}\cdot M(\mathbf{r}_k-\mathbf{r}_i)a_k(t-1)\right]$$

where the lateral interaction $M(.)$ is approximated by the Laplacian of Gaussian $\nabla^2 g(\mathbf{r})$; N is the set of output neurons and a_k is the activation of neuron k and a_k is the activation of neuron k. Now, we obtain the net input to neuron i on SO layer as

$$net_i(t)=a_i(t-1)+A(-b_s a_i(t-1)+c_s\left[I_i^v(t)+I_i^l(t)-\Gamma\right]^+)$$

where b_s and c_s are positive constants; Γ is a threshold to avoid the noise effect; and the functions $[\cdot]^+$ and $A(\cdot)$ are defined as follows:

$$[u]^+ = \begin{cases} u & \text{if } u > 0 \\ 0 & \text{if } u \le 0 \end{cases},$$

$$A(u) = \begin{cases} u & \text{if } u > 0 \\ d_s u & \text{if } u \le 0 \end{cases},$$

where $1 > d_s > 0$. The transfer function of SO neurons is generally simulated using a sigmoid function. Thus the real activation of neuron i is

$$a_i = \psi(net_i)$$

In Figure 4, column (f) shows the experimental results of the TSOM neural network, where b_s = 0.5; c_s = 1; d_s = 0.5; and image size is 160X120. In the experiment, the focus of attention is highly concentrated in the attention map, even the camcorder and vehicle jolted along and one of the road borders is a dash line.

V. ADAPTIVE RESONANCE THEORY

As we mentioned before, the memory capacity of our brain is almost unlimited because the information is scatteringly stored on the connections among neurons, not on the neurons themselves. Thus, the various neural networks can share the same neurons with others even if they worked for different mental processes. If necessary, our brain could immediately collect a set of neurons to construct a suitable neural network for a specific mental process. After the mental process has been accomplishment, the neural network is then destroyed and the neurons in the neural network can be reused by another neural network for another mental process. Neural networks having this property are called configurable neural networks. In this section, we discuss the configurable adaptive resonance theory (CART) neural network. Moreover, we select ART2 from ART family to classify the attention map because of its unsupervised processing ability and acceptance of floating patterns.

Figure 5 sketches the ART2 architecture [2] consisting of two main modules: the attentional and the orienting modules. The attentional module is further divided into two fields: an input representation field, F_1, and a category representation field F_2. There are top-down and bottom-up full connections between these two fields. Prototypes of patterns are to be reserved on these connections in terms of their synaptic weights. The F_1 field consists of six layers, **w**, **x**, **v**, **u, p**, and **q**. Bottom-up input patterns and top-down predicted prototypes will be matched in this field. Field F_2 has only one layer, denoted **y**. This layer can be realized by any gated dipole field network and serve as a competitive layer. The orienting module consists of two components: one layer, specified **r**, and one signal generator, denoted *S*. Layer **r** is connected to layers **p** and **u** in the F_1 field which aggregates the activities of **p** and **u** and transmits the result to the signal generator *S*. It then decides based on the result whether or not to emit a reset signal to the layer **y** in field F_2. If a signal is emitted, the currently activated neuron on **y** is prohibited and the entire process is repeated;

otherwise, the neuron either modifies the prototype reserved by the neuron, or learns the input pattern as a new prototype, or rejects the input pattern because the memory has been full. In addition to the aforementioned layered structures, there are three gain control units in the F_1 field, and two gain control units in the orienting module. The gain control units, implemented by on-center off-surround networks, play the role of normalizing the activity patterns of neurons on layers.

In summary, the computational process of CART2 is as follows [2]:

(1) Initialize the fully-connected weights between fields F_1 and F_2.
(a) Initialize the top-down weights to zero.
(b) Initialize the bottom-up weights by

$$z_{ji}(0) \le \frac{1}{(1-d_a)\sqrt{M}}$$

where $0 < d_a < 1$ and M is the number of neurons on each layer in field F_1.

(2) Set all layer and sublayer outputs to zero vectors, and the cycle counter to one.

(3) Input a pattern **i** to the **w** layer, and propagate to the **x, v,** and **u** layer by equations as follows:

$$w_i = I_i + a_a u_i, \quad x_i = \frac{w_i}{e + \|\mathbf{w}\|},$$

$$v_i = f(x_i) + b_a f(q_i), \quad u_i = \frac{v_i}{e + \|\mathbf{v}\|},$$

where a_a and b_a are positive constants, and e is a small value preventing neural activities from becoming infinite when no signal is present on the layers. Function f conducts a contrast enhancement to the input pattern defined by

$$f(x) = \begin{cases} 0 & 0 \le x \le \theta \\ x & x > \theta \end{cases}$$

where θ is a positive constant less than one.

(4) Propagate the output of **u** sublayer to **p** and **q** sublayers using equations as follows:

$$p_i = u_i + \sum_j g(y_j) z_{ij}, \quad q_i = \frac{p_i}{e + \|\mathbf{p}\|}$$

Function g is a transfer function of the neurons on layer **y** given by,

$$g(y_J) = \begin{cases} d_a & T_J = \max_k \{T_k\} \\ 0 & \text{otherwise} \end{cases}$$

where J indicates the winner on layer **y** and d_a is a constant between 0 and 1. Input T_k is the net input from layer **p** to the kth neuron of layer **y**:

$$T_k = \sum_i p_i z_{ik}$$

where z_{ik} is the bottom-up weight from the ith neuron of layer **p** to the kth neuron of layer **y**.

(5) Repeat steps (3) and (4) until the values of sublayers on the field F_1 is stable.

(6) Calculate the output of the **r** layer using by

$$r_i = \frac{u_i + c_a p_i}{e + \|\mathbf{u}\| + \|c_a \mathbf{p}\|}$$

where c_a is a constant subject to the constraint that $(c_a d_a / (1 - d_a)) \le 1$.

(7) The orienting subsystem decided whether to output the reset signal.
(a) If $\rho / (e + \|\mathbf{r}\|) > 1$, then reset the winner on field F_2, and set the cycle counter to one, go to step (3).
(b) If $\rho / (e + \|\mathbf{r}\|) \le 1$, and cycle counter is one, then increment the cycle counter, and go to step (8).
(c) If $\rho / (e + \|\mathbf{r}\|) \le 1$, and cycle counter is greater than one, then go to step (11).

(8) Propagate the output of the **p** sublayer to field F_2. Calculate the net inputs of neurons in field F_2.

(9) According to function g, only the winner on field F_2 has nonzero output.

(10) Repeat steps (4) through (7).

(11) Modify bottom-up weights and the top-down weights between fields F_1 and F_2:

$$z_{ij} = z_{ji} = \frac{u_i}{1 - d_a}$$

(12) Remove the input vector. Restore all inactive F_2 neurons. Return to step (1) with a new input pattern.

In our application, the input images are the sub-sampled attention maps. The attention maps are sub-sampled to avoid the noise effect and to reduce the time complexity (160X120 → 80X60 pixels). Before classification stage, the training stage should be finished to memorize the learned patterns into the LTM. Once a supraliminal pattern, an attention map, input ART2, the subliminal expectation should be looked for in the LTM and compared with the supraliminal pattern. If they are similar enough, then the input pattern is successfully classified and learned into our system by supervised learning. On the other hand, if none of subliminal expectations is similar to the input pattern, the system should learn the input pattern to be a new class through unsupervised learning.

Figure 6 shows the experimental result of the CART neural network. There are 20 attention maps input to the CART neural network, and these attention maps are classified into ten classes. The maps in columns (a), (b), and (c), all happen when vehicles change to left lane, and the maps in columns

(f) and (g) both happen when vehicles change to right lane. Column (d) indicates the map class that vehicles enter the tunnel, and the maps in column (e) show the vehicles exit the tunnel. Besides, column (h) shows the expressway-entry class and column (i) shows the expressway-exit one. Finally, column (j) shows the viaduct-ahead case. This classification result is correct, where $a_a = 0.1$; $b_a = 1$; $c_a = 0.1$; $d_a = 0.9$; $e = 0.0000001$; $\rho = 0.99$; $\theta = 0.0001$.

VI. CONCLUSION

This paper described a method to detect the road condition change, including lane-change, expressway-entry, expressway-exit, tunnel-entry, tunnel-exit and viaduct-ahead conditions. The input data of our environmental change detection system is color video sequences recorded by a camcorder mounted on a moving vehicle. Thus first, we design a pre- processing stage to surmount this difficulty from the influence of camcorder and vehicle jolt. Second, a TSOM neural network is created to model the attention map in our brain for detecting the environmental changes. Finally, CART2 neural network is utilized to classify the environmental changes. We will try to detect more changes of driving environments in the future.

REFERENCES

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

[2] J. A. Freeman and D. M. Skapura, *Neural Network s---Algorithms, Applications, and Programming Techniques*, Addison-Wesley, Reading, Massachu- setts, 1992.

[3] G. Salgian and D. H. Ballard, "Visual Routines for Autonomous Driving," *Proceedings of the 6th International Conference on Pattern Recognition*, Bombay, India, pp. 876-882, 1998.

[4] S. Tsugawa, "Vision-Based Vehicles in Japan: the Machine Vision Systems and Driving Control Systems," *Proceedings of IEEE International Symposium on Industrial Electronics*, Budapest, pp. 278-285, 1993.

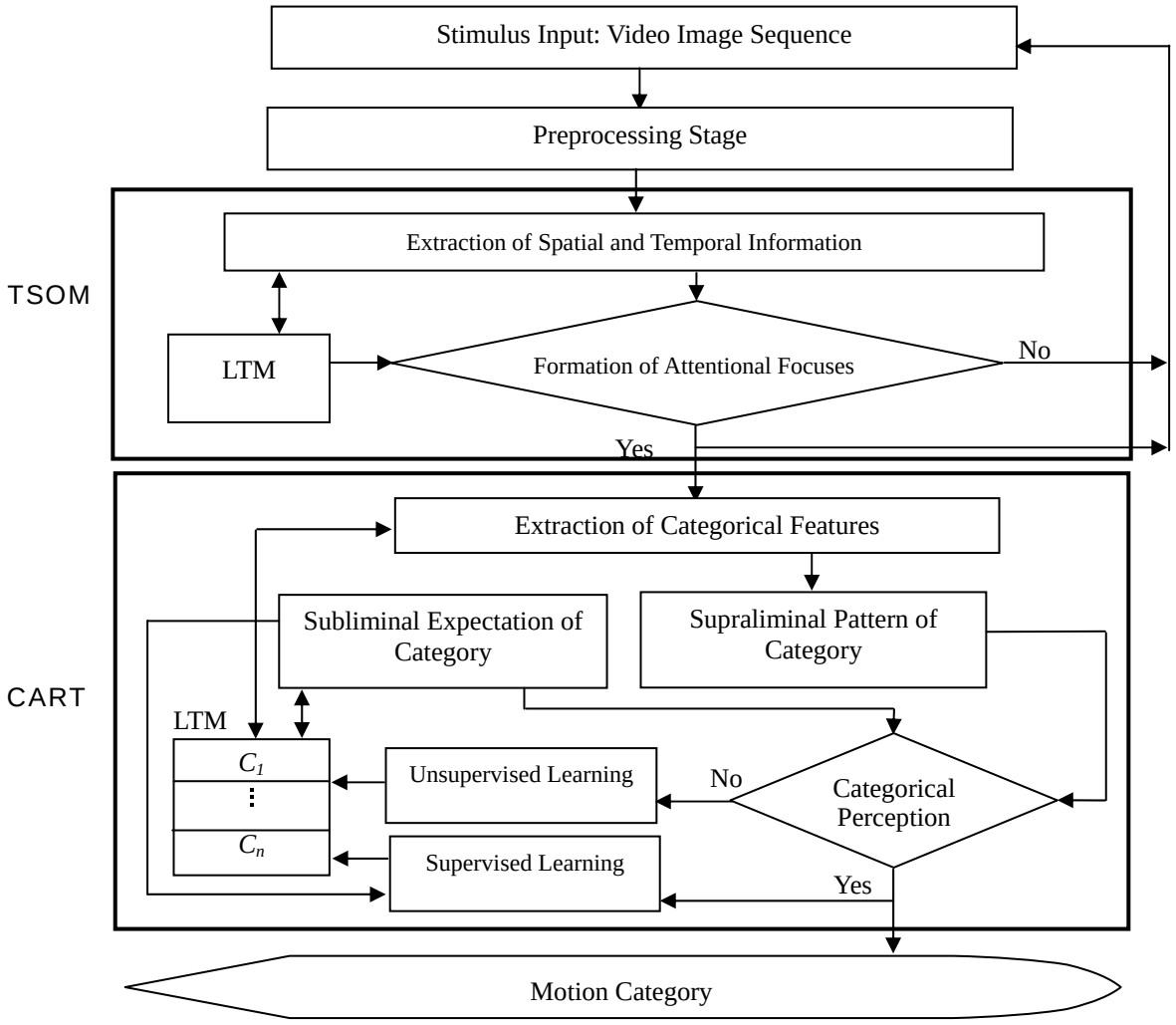


Figure 1: The structure of environmental change detection system.

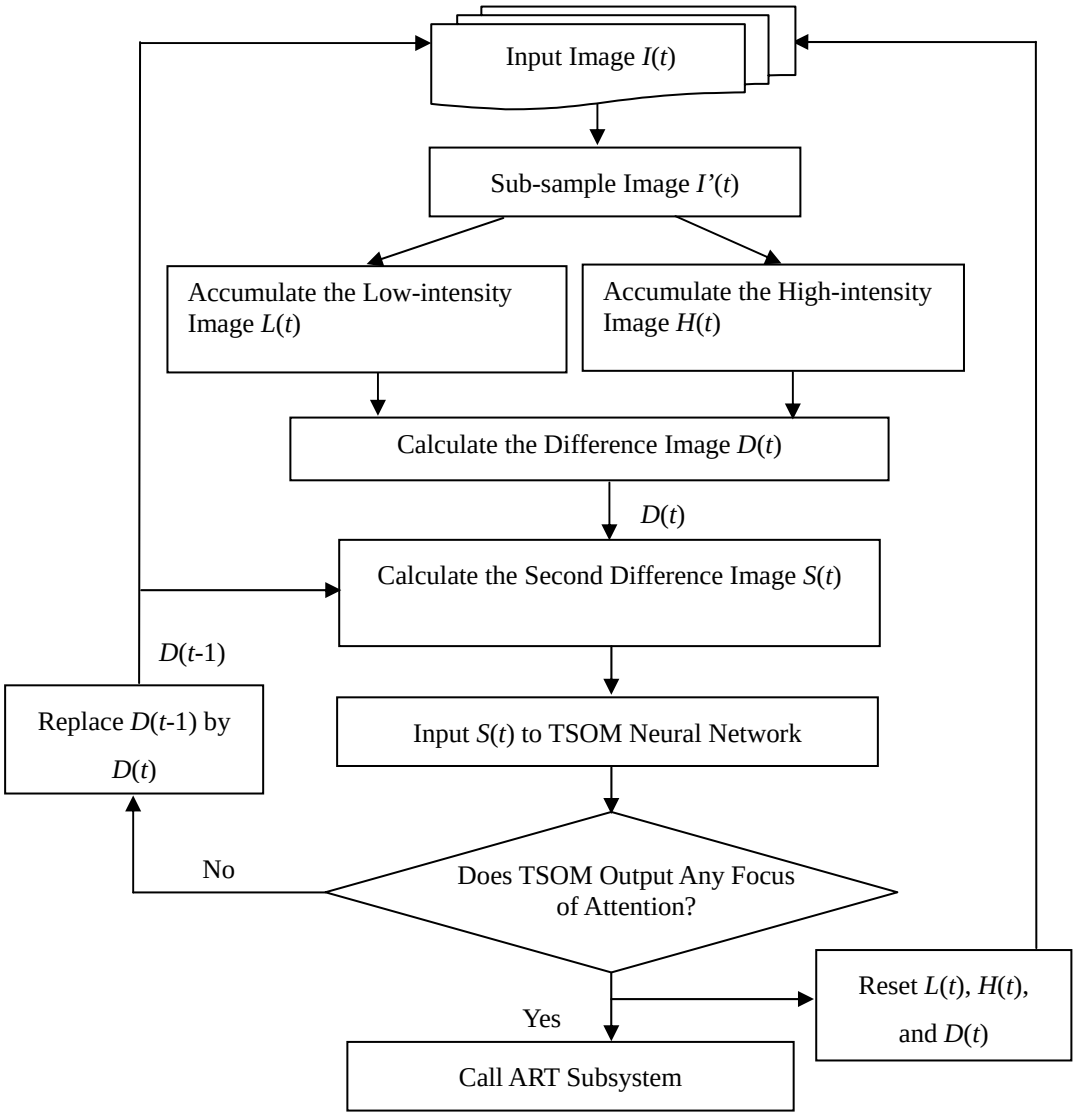


Figure 2: The flowchart of the preprocessing stage of environmental change detection system.

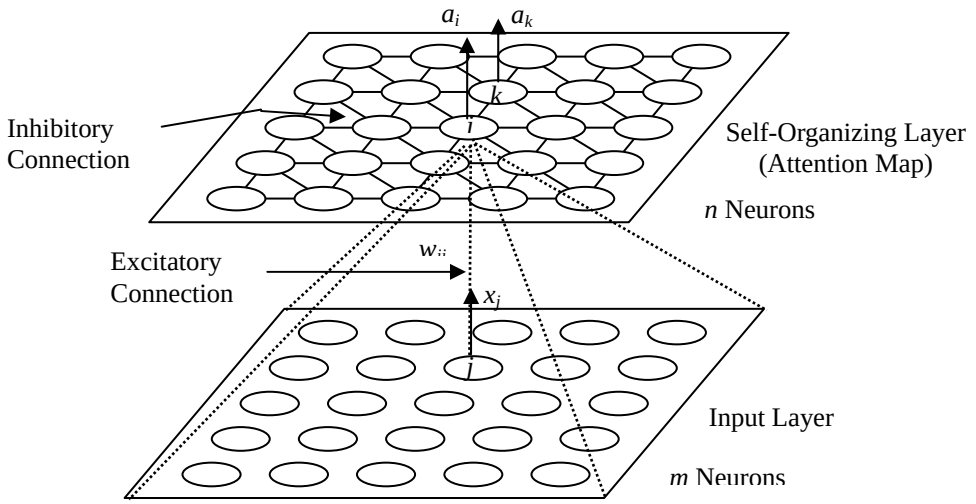
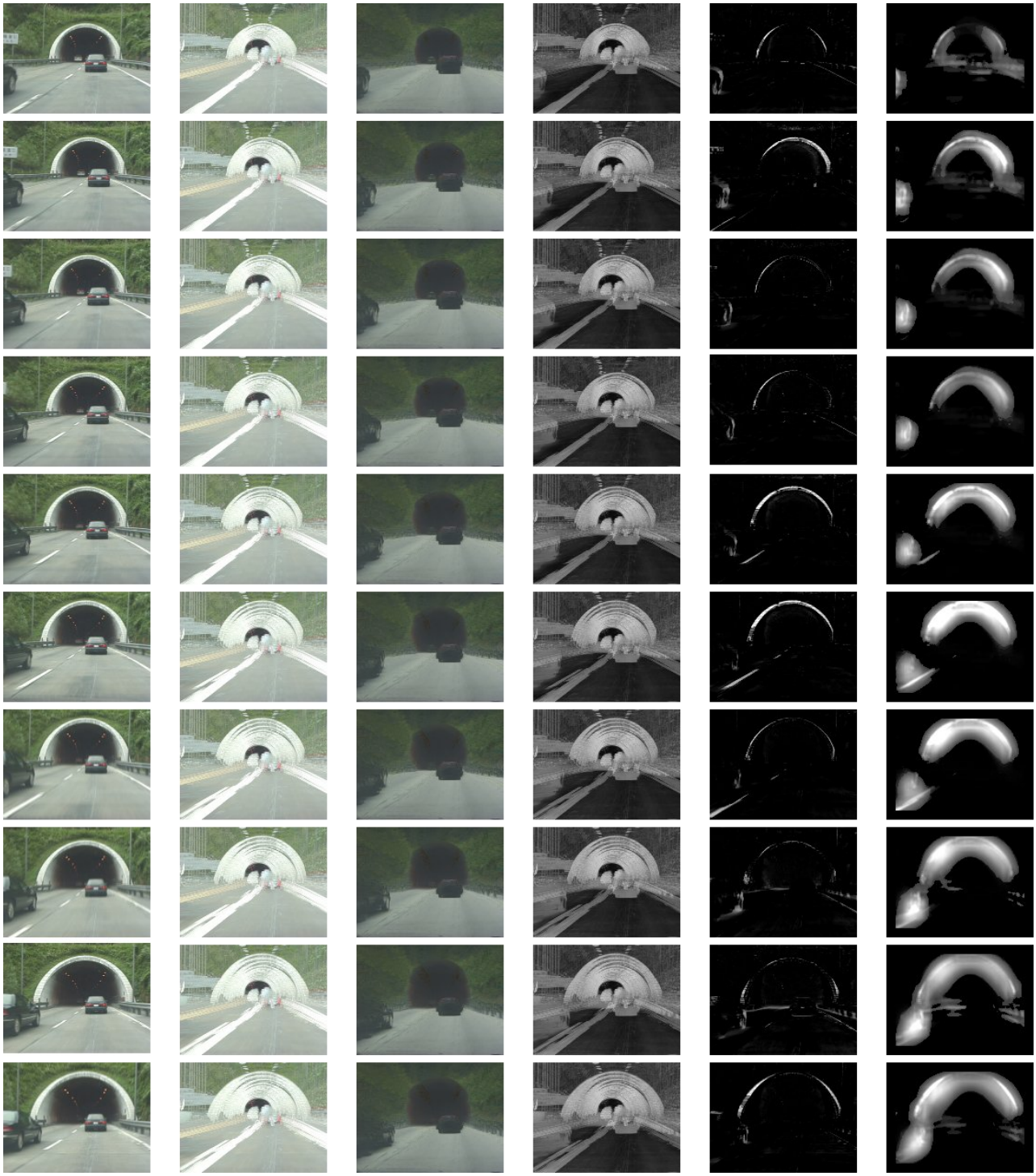


Figure 3: The TSOM neural network.



(a) (b) (c) (d) (e) (f)

Figure 4: One example illuminates the preprocessing stage of the environmental change detection system. (a) The input image sequence *I*. (b) The high-intensity image sequence *H*. (c) The low-intensity image sequence *L*. (d) The difference image sequence *D*. (e) The second difference image sequence *S*. (f) The output sequence of TSOM neural network, where b_s = 0.5; c_s = 1; d_s = 0.5, and image size is 160X120.

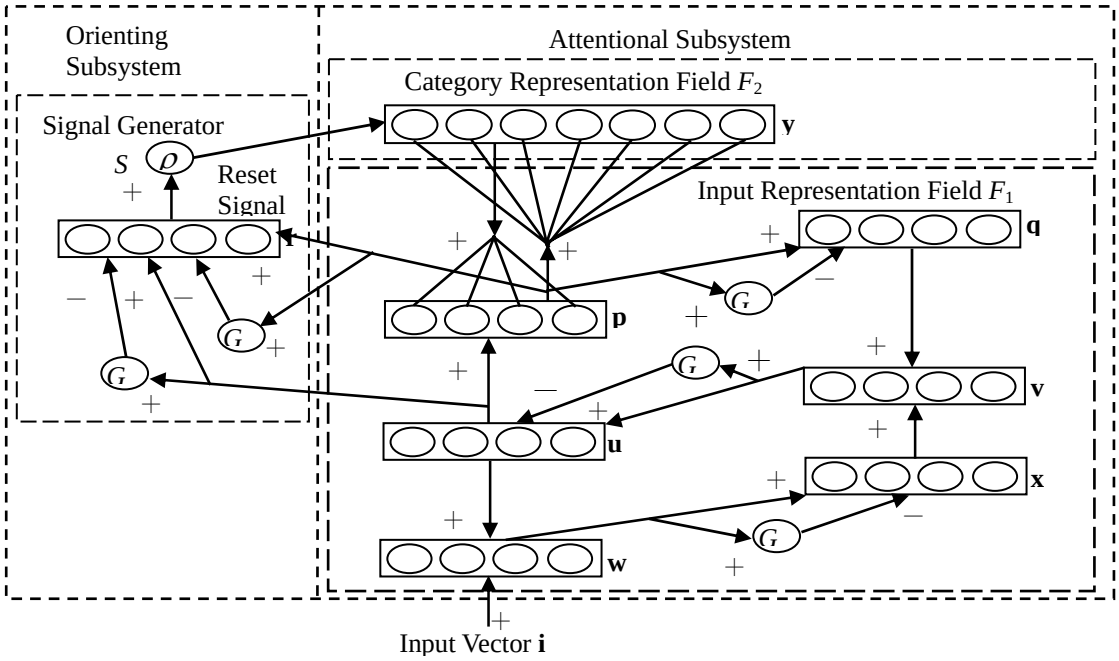


Figure 5: The ART2 neural network.

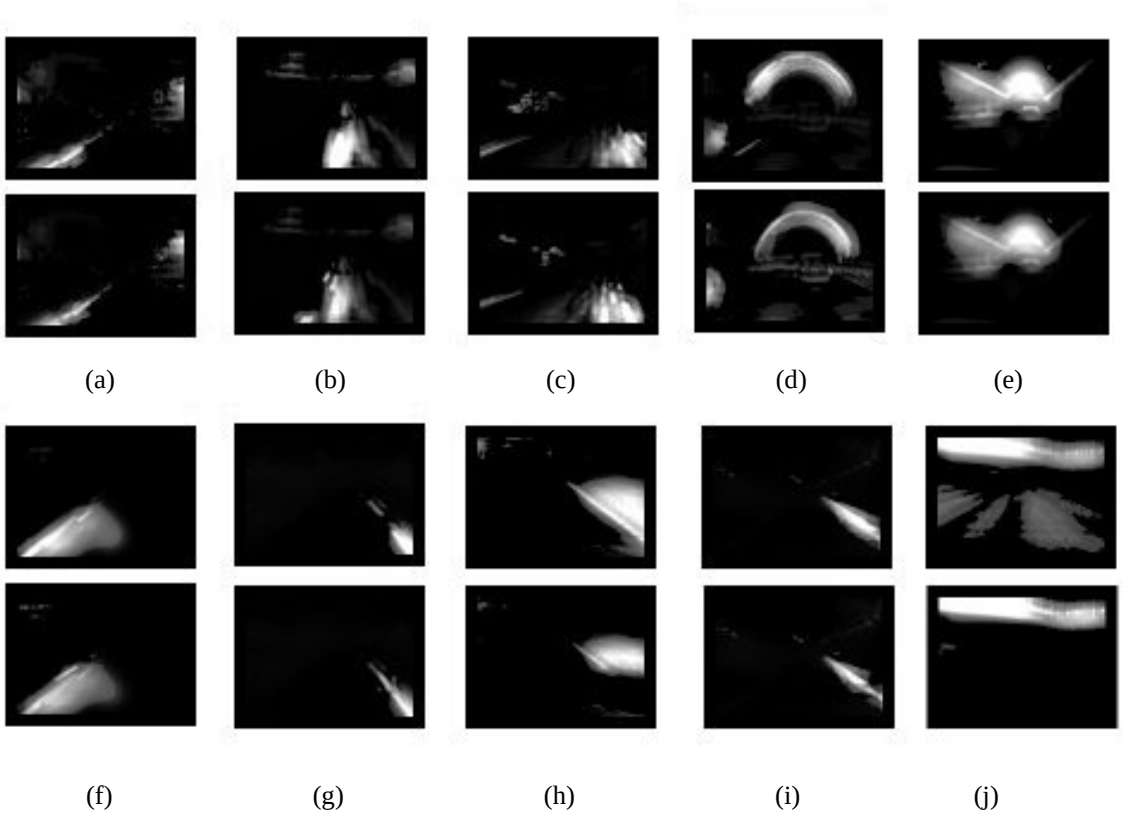


Figure 6: The experimental result of the CART neural network. There are 20 attention maps input to the CART neural network, and these maps are classified into ten classes, where a_a = 0.1; b_a = 1; c_a = 0.1; d_a = 0.9; e = 0.0000001; ρ = 0.99; θ = 0.0001, and image size is 80X60. (a) Start of change to left lane. (b) Half-way of change to left lane. (c) End of change to left lane. (d) Tunnel-entry. (e) Tunnel-exit. (f) End of change to right lane. (g) Start of change to right lane. (h) Expressway-entry. (i) Expressway-exit. (j) Viaduct-ahead.