Road Sign Detection from Complex Backgrounds

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

†Department of Information and Computer Education
National Taiwan Normal University
Taipei, Taiwan, R. O. C.
Email: violet@ice.ntnu.edu.tw
‡Department of Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan, R. O. C.

Abstract

This paper describes an approach to detecting road signs from complex backgrounds. The input data are color image sequences acquired using a single camcorder mounted on a moving vehicle. Two neural networks are developed to extract color and shape features, respectively, from image sequences. A process characterized by fuzzy discipline is then introduced to determine road sign candidates based on the extracted color and shape features. Experimental results have manifested the applicability of the proposed method. Incremental strategies applied to constantly incoming video data may improve the performance of the method.

Keyword: Road sign detection; HSI color model; Neural networks; Fuzzy integration

1. Introduction

Automatic recognition of road signs can be of use in such applications as driver assistant systems, maintenance of traffic facilities, route planning and management, as well as administration of transportation securities and efficacy all toward the ultimate goal of intelligent transportation systems (ITS) [10].

Two major steps constitute road sign recognition: (1) localization and (2) recognition. This study concerns the localization of road signs from complex scenes. Road signs can be broadly categorized in terms of functionality into three classes: specification, warning, and prohibition. Each class is characterized by a few particular colors, e.g., blue and green (specification), yellow and orange (warning), and red (prohibition). While colors provide crucial evidence for road sign detection, the colors of signs generally degrade with time. Moreover, the chromatic properties of colors could be altered by shading, shadowing, highlighting, and different weather, air, as well as lighting conditions. Road sign detection has to rely on other supports in addition to the color one. In this study, structural evidence of road signs is incorporated. The scenes under our consideration may be taken from freeways, expressways, highways, boulevards, streets, and country roads, which involve considerably distinct backgrounds and road signs. Reliable detection of road signs from various scenes could be rather challenging.

Many techniques [1, 3, 4, 5, 6, 7] have been developed to detect road signs from scenes. Most of them worked on single input images [3, 4] and some on grey level images. In this paper, video color image sequences are employed because they provide rich information for the purpose. Lalonde and Li [3] reported a color indexing approach to identifying road signs. There was a prebuilt database in their system. In the database, road signs were described in terms of color histograms. Identifying a road sign extracted from an image was carried out simply by comparing its color histogram with those in the database. Their method did not take into account the structural evidence of road signs. Piccioli et al. [7] incorporated both color and edge information to recognize road signs. They claimed that their technique could readily be applied to temporal image sequences and would improve the performance.

In this paper, a structural evidence of road signs is presented, referred to as circular shape features. The methods for detecting color and shape features are addressed in the following section. Besides, a process characterized by fuzzy discipline is introduced for integrating information derived from the continually incoming video sequences. The experimental results with real images are demonstrated in Section 3, followed by concluding remarks and future works in Section 4.

2. Road Sign Detection System

In general, road signs convey traffic information through their shapes, colors, and contents. To attract human attention, road signs are usually designed using particular colors and shapes. Thus, color and shape are two important features of road signs, which are useful to be a priori information for automatic road sign detection.

The outline of our road sign detection system is shown in Figure 1. One image $S^{(t)}$ of an input video sequence, *S*, is fed into the road sign detection system. First, the detection system splits the color image channels to calculate the hue image, $S_h^{(t)}$, for extracting the color feature. The color feature is defined as the centers of some special color regions. Second, an edge detection method is applied to acquire the edge image, $S_e^{(t)}$, for extracting the shape feature. The shape feature is defined as the centers of some fixed shape regions. Third, the color and shape features are integrated to locate the centers of road signs, and this is the result of road sign detection. Finally, the car speed is predicted to update the parameters of our feature extraction procedures, and it provides the immediate information for later processing successive image $S^{(t+1)}$.

Two neural networks are developed to detect color and shape features, respectively. Both of the neural networks are constructed by two layers of neurons, one is input layer, the other is output layer. The synapses between input layer and output layer are full connection. Figure 2 shows the sketch of these two neural networks, and they will be explained in detail later.

2.1 Detection Using Color

Although road signs are mainly made by particular colors, such as red, green, blue, and orange, this fact does not reduce the difficulty of road sign detection. On account of the weather conditions, the outdoor illumination varies, and can not be under control. Moreover, the paint on signs also deteriorates with time. Thus, we should look for a suitable color model for road sign detection.

Many color models are provided to process color image [2, 8], including RGB (Red, Green, Blue) model, normalized RGB model, HSI (Hue, Saturation, Intensity) model, TekHVCT model, XYZ tristimulus coordinates, CIE (the International Commission on Illumination) (Lab) space, CIE (Luv) space, YIQ and YUV color system, and so on. Different color models can be utilized for different applications. Since the HSI color model is defined based on human color perception [2], and the designing point of road signs is focused on capturing human attention, HSI model may be suitable for the application of road sign detection.

In HSI color model, the values of saturation channel and intensity channel should inevitably be under the influence of light (sunny or shady) and shadow. Nevertheless, hue channel is well known to be invariant to lightness [2, 8]. In consideration of the uncertainty of the weather and the natural and artificial damages of road signs, we think hue channel is the best choice among color features in road sign detection.

2.2 The Neural Network for Color Feature Detection

Since the color model used by the camcorder is RGB model, we first transfrom RGB model to HSI model. In fact, only the hue channel is needed. Let $(R, G, B)_{kl}$ indicate the red, green, and blue color values of the pixel (k, l) in the color image and input to neuron *q* on the input layer. If, for example, the neural network would be used to extract the "red" color, and h^c is defined as follows:

$$h^c(k,l)=\begin{cases}\dfrac{180}{\pi}\cos^{-1}\left\{\dfrac{\frac{1}{2}[(R_{kl}-G_{kl})+(R_{kl}-B_{kl})]}{[(R_{kl}-G_{kl})^2+(R_{kl}-B_{kl})(G_{kl}-B_{kl})]^{1/2}}\right\} \\ \qquad \textbf{if } (R_{kl}\text{-}G_{kl})>0, (R_{kl}-B_{kl})>0, \textbf{and } (G_{kl}-B_{kl})\ge 0 \\ -\dfrac{180}{\pi}\cos^{-1}\left\{\dfrac{\frac{1}{2}[(R_{kl}-G_{kl})+(R_{kl}-B_{kl})]}{[(R_{kl}-G_{kl})^2+(R_{kl}-B_{kl})(G_{kl}-B_{kl})]^{1/2}}\right\} \\ \qquad \textbf{if } (R_{kl}\text{-}G_{kl})>0, (R_{kl}-B_{kl})>0, \textbf{and } (G_{kl}-B_{kl})<0 \\ 0 \qquad\qquad \textbf{otherwise}\end{cases} \tag{1}$$

Then, the output of neuron *q* should be defined by

$$x_{kl}^c=\begin{cases}180-h^c(k,l)+R_o & \textbf{if } h^c(k,l)\ge R_0 \\ 180+h^c(k,l)-R_o & \textbf{if } h^c(k,l)<R_0\end{cases}$$

(2)

where R_0 is the hue value of the standard "red" color.

First, function h^c calculates hue values when the red value is larger than green and blue values in the RGB model. This is the loosest constraint about the definition of "red". Second, from equation (2), neuron *q* works like a similarity function, the more similar between the color of an image pixel and the standard "red" color, the larger output value of the neuron *q*.

Let $r^c(i,j,k,l)=\sqrt{(i-k)^2+(j-l)^2}$, then the weight between neuron *q* on the input layer and neuron *p* on the output layer could be defined as follows:

$$w_{kl,ij}^c=\begin{cases}A/r^c(i,j,k,l) & \textbf{if } i\ne k\textbf{ , } j\ne l, \\ & \textbf{and } T_1\le r^c\le T_2 \\ 0 & \textbf{otherwise}\end{cases} \tag{3}$$

here *A* is a constant, (i, j) could be regarded as the center of a road sign; T_1 and T_2 are the inner and outer radii, respectively, of the ring region we are interested in and they will be adapted to extract different sizes of road signs. Thus, for every neuron p on the output layer, the net input to the neuron p could be calculated by

$$net_{ij}^{c} = \sum_{k}\sum_{l} x_{kl}^{c} w_{kl,ij}^{c} \tag{4}$$

where the indices, k and l, run over all connections from input layer to the neuron p. If $f_{ij}^{c}()$ denotes the transfer function of neuron p, then the output of the neuron p could be written as follows:

$$y_{ij}^{c} = f_{ij}^{c}(net_{ij}^{c}) = net_{ij}^{c} \tag{5}$$

here the transfer function of neuron p only outputs the same input value. The output value of neuron p indicates the possibility whether there is any "red" ring whose center is at (i, j).

2.3 Detection Using Shape

The design of road sign shapes are only simple geometric shapes, including circle (e.g. railroad advance warning or civil defense evacuation route), rectangle (e.g. regulatory signs or guide signs), octagon (e.g. STOP signs), equilateral triangle (e.g. YIELD signs), and so on. Since shape feature does not vary in all weathers and light conditions, it is a useful and robust characteristic of road signs.

Shape information especially helps to remove the unsuitable color feature points, such as the pixels belonging to a "red" building or a "red" car, that is, a large "red" region but not a road sign. Although sometimes road signs may be partly occluded or damaged, the shape information is still very reliable in road sign detection.

2.4 The Neural Network for Shape Feature Detection

Edges are the fundamental elements forming shapes. Thus, the neurons on input layer work as an edge detection function. Let $(R, G, B)_{kl}$ indicate the red, green, and blue color values of the pixel (k, l) in the color image and input to neuron q on the input layer, and x_{kl}^{s} denotes the output of neuron q, then

$$x_{kl}^{s} = \sqrt{(I_{kl} - I_{k-1\,l})^2 + (I_{kl} - I_{k\,l-1})^2}\ . \tag{6}$$

Here $I_{kl} = (R+G+B)_{kl}/3$, and equation (6) could be replaced by any other edge detection equations, such as Roberts, Sobel, Laplacian, and so on.

Let $r^{s}(i,j,k,l) = \sqrt{(i-k)^2 + (j-l)^2}$. If, for example, the neural network would be used to extract the circular shape, then the weight between neuron q on the input layer and neuron p on the output layer could be defined by

$$w_{kl,ij}^{s} = \begin{cases} B/r^{s}(i,j,k,l) & \textbf{if } r > T_3, \\ & \textbf{and } r \bmod C < C/2 \\ -B/r^{s}(i,j,k,l) & \textbf{if } r > T_3, \\ & \textbf{and } r \bmod C \geq C/2 \\ 0 & \textbf{otherwise} \end{cases} \tag{7}$$

here T_3, B, and C are constants, and C depends on the method of edge detection we choose. Thus, for every neuron p on the output layer, the net input to the neuron p is given by

$$net_{ij}^{s} = \sum_{k}\sum_{l} x_{kl}^{s} w_{kl,ij}^{s} \tag{8}$$

where the indices, k and l, run over all connections from input layer to the neuron p. If $f_{ij}^{s}()$ denotes the transfer function of neuron p, then the output of the neuron p is

$$y_{ij}^{s} = f_{ij}^{s}(net_{ij}^{s}) = \begin{cases} 1 & \textbf{if } |net_{ij}^{s}| > T_4 \\ 0 & \textbf{otherwise} \end{cases} \tag{9}$$

here T_4 is a threshold, and $f_{ij}^{s}()$ is a step function. If the output of $f_{ij}^{s}()$ is 1, then we believe that there is at least a circle whose center is at (i, j).

2.5 Fuzzy Feature Integration

The shape and color features extracted by the foregoing neural networks should be integrated to locate the centers of road signs. Here we provide a fuzzy approach to accomplish this job.

For every pixel (i_1,j_1) in the input image, its fuzzy degree belonging to road signs, the membership function, μ, is defined as follows:

$$\mu(i_1, j_1) = y_{i_1 j_1}^{s} \sum_{i_2}\sum_{j_2} y_{i_2 j_2}^{c} \times f^{m}(i_1, j_1, i_2, j_2) \tag{10}$$

$$f^{m}(i_1, j_1, i_2, j_2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(i_1-i_2)^2+(j_1-j_2)^2}{2\sigma^2}} \tag{11}$$

where σ is a constant; $y_{i_1 j_1}^{s}$ is the output of the neuron (i_1, j_1) on the output layer for the neural network of shape feature point detection; and $y_{i_2 j_2}^{c}$ is the output of the neuron (i_2, j_2) on the output layer for the neural network of color feature point detection.

Since there may be more than one road sign in an image, we select the local maximum in sub-regions to locate the road signs. Let the initial size of road signs be *SIZE*; the locations of road signs be defined by

$$Location(i, j) = \begin{cases} 1 & \textbf{if } \mu(i, j) = \max\limits_{\substack{i-SIZE/2 \leq i_1 \leq i+SIZE/2 \\ j-SIZE/2 \leq j_1 \leq j+SIZE/2}} \mu(i_1, j_1) \\ 0 & \textbf{otherwise} \end{cases} \tag{12}$$

If the output of *Location*(i, j) is 1, then we think it has high possibility that there exists a road sign whose center is at (i, j).

3. Experimental Results

At present, the system is first experimented to detect the red circular road signs, and it can be extended to detect other road signs later. Figures 5 and 6 show the two input video sequences, S_1 and S_2, respectively. Sequence S_1 is taken in the high way, and S_2 is taken on the street. In these video sequences, the size of each image is 320×240 pixels and the time between two successive images is 0.2 second.

Figure 3 shows the experimental results of one frame, the fifth frame of the video sequence S_1. First, the input image is shown in Figure 3(a). There is a speed limit sign in the image. Second, the output results of the neural network for color feature detection is shown in Figure 3(b) and the white pixels indicate the color feature points of the input image. In this step, we defined $R_0 = 0$, $B = 1$, and the initial values of T_1 and T_2 are 5 and 15, respectively. Third, the output results of the neural network for shape feature detection are shown in Figure 3(c), and the white pixels indicate the shape feature points of the input image. Here we defined $A = 100$, $C = 6$, $T_3 = 5$, and $T_4 = 40$. Finally, Figure 3(d) is the integrated result of the road sign detection system (where $\sigma = 2$ and $SIZE = 20$). The system uses one white rectangle to indicate one location of road signs. In this example, our system detects the speed limit sign correctly.

Figure 4 shows another example provided by video sequence S_2 with $R_0 = 0$, $T_1 = T_3 = 10$, $T_2 = 22$, $T_4 = 40$, $A = 100$, $B = 1$, $C = 6$, $\sigma = 2$ and $SIZE = 20$. Although there is actually only one no right turn sign in the traffic scene, our system detects more than one. We believe the false alarm will be refined by the following images if the car speed could be predicted precisely.

Now, we can make an observation on the above examples. The result of the example shown in Figure 3 is mostly affected by color feature. The main cause is that the number of shape feature points is larger than that of color feature points. On the other hand, since the number of color feature points is larger than that of shape feature points, the result of the example shown in Figure 4 is mainly affected by shape feature. In short, color and shape features may have different weights in different examples.

Finally, Figures 5 and 6 illustrate more experimental results of the different frames in the video sequences S_1 and S_2, respectively. In these examples, we applied simple linear method to predict the car speed, but we did not use a priori knowledge to enhance the correctness of road sign detection.

4. Conclusions and Future Work

Our paper describes a method for detecting road signs from complex background. Input a color sequence images acquired by a single camcorder, two neural networks are developed to extract the color and shape features, respectively. A fuzzy approach to integrating color and shape features is introduced for selecting the candidates of the centers of road signs. Experimental results show that our system works well.

Our system predicts the car speed by linear method now, but in fact the velocity variation of the car is nonlinear. If the system could include a precise technique of car speed prediction, such as Kalman filter [9], then its functionality will be improved. Furthermore, if we could accurately track motion trajectories of the road signs by incorporating the information of the image sequence, we can detect the road sign more robustly. These are our future works.

Acknowledgments

This work is supported in part by the National Science Council, Taiwan, Republic of China, under contract NSC 88-2213-E-003-003.

References

[1] A. Guiducci, "3D Road Reconstruction from a Single View," *Computer Vision and Image Understanding*, Vol. 70, pp. 212-226, 1998.

[2] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, Massachusetts, 1993.

[3] M. Lalonde and Y. Li, "Detection of Road Signs using Color Indexing," Technical Report CRIM-IT-95/12-49, Centre de Recherche Informatique de Montreal, 1995. (*http://www.crim.ca/sbc/english/cime/publications.html*)

[4] M. Lalonde and Y. Li, "Road Signs Recognition," Technical Report CRIM-IT-95/09-35, Centre de Recherche Informatique de Montreal, 1995. (*http://www.crim.ca/sbc/english /cime/publications.html*)

[5] W. Li, X. Jiang, and Y. Wang, "Road Recognition for Vision Navigation of an Autonomous Vehicle by Fuzzy Reasoning," *Fuzzy Sets and Systems*, Vol. 93, pp. 275-280, 1998.

[6] P. Parodi and G. Piccioli, "A Feature Based Recognition Scheme for Traffic Scenes," *Proceedings of Intelligent Vehicles Symposium*, pp. 229-234, Detroit, 1995.

[7] G. Piccioli, E. D. Micheli, P. Parodi, and M. Campani, "A Robust Method for Road Sign Detection and Recognition," *Image and Vision Computing*, Vol. 14, pp. 209-223, 1996.

[8] W. Skarbek and A. Koschan, "Colour Image Segmentation--A Survey," Thchnischer Bericht 94-32, Technical University of Berlin, 1994.

[9] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," *http://www.cs.unc.edu/~welch/kalmanIntro.html*, 1999.
[10] Intelligent Transportation Systems, *http://www.iot.gov.tw/its/,* 1999.

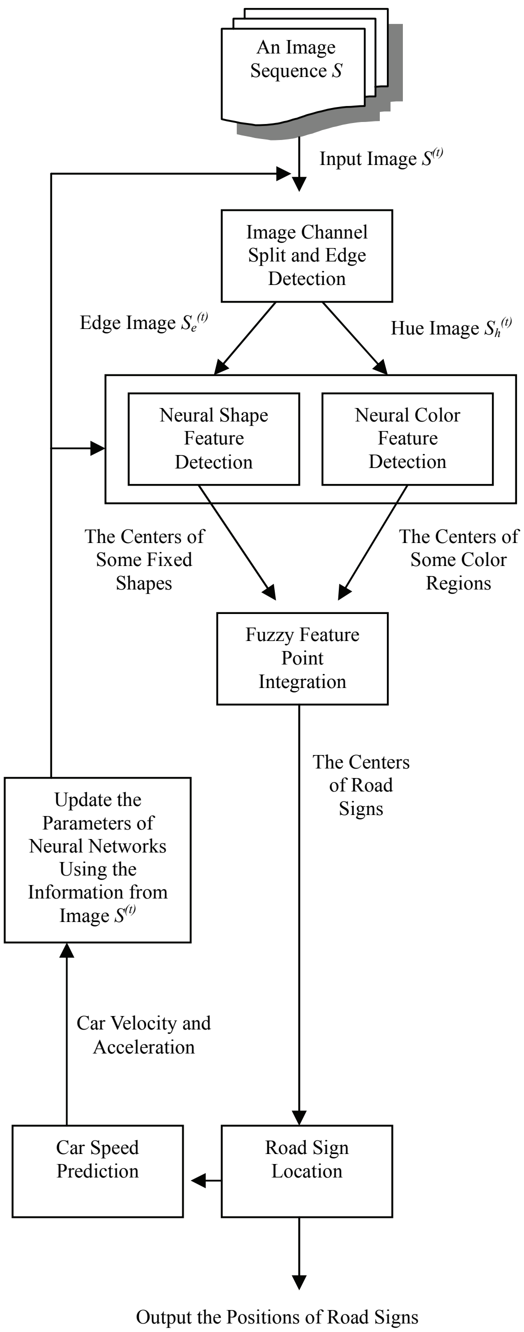


Figure 1: Outline of road sign detection.

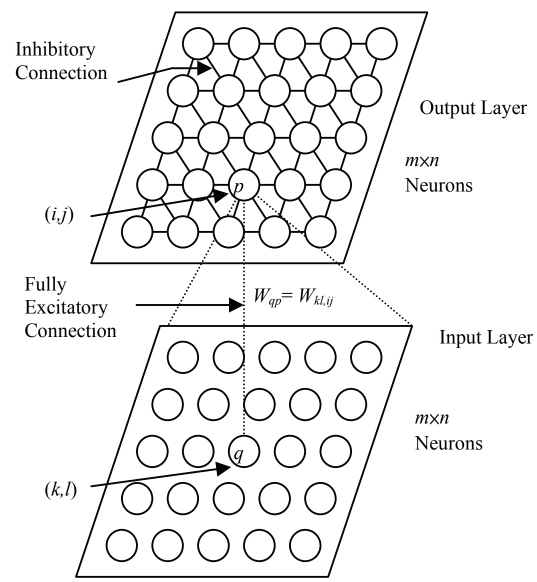
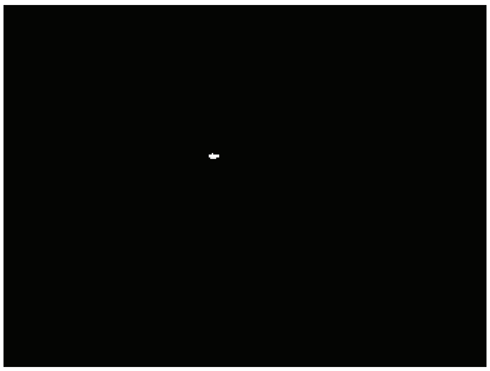
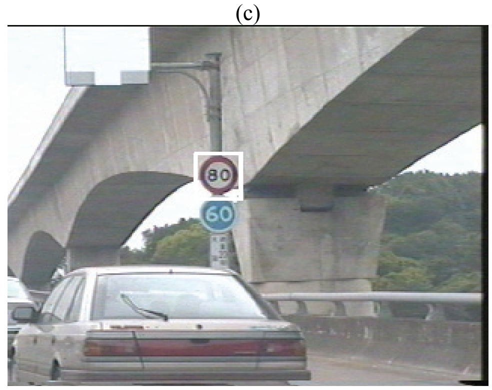


Figure 2: Sketch of the neural network for color (shape) feature detection, where (i, j) is the position of neuron p on the output layer, (k, l) is the position of neuron q on the input layer, and $W_{kl,ij}$ is the weight between neuron q on the input layer and neuron p on the output layer.

(a)

(c)



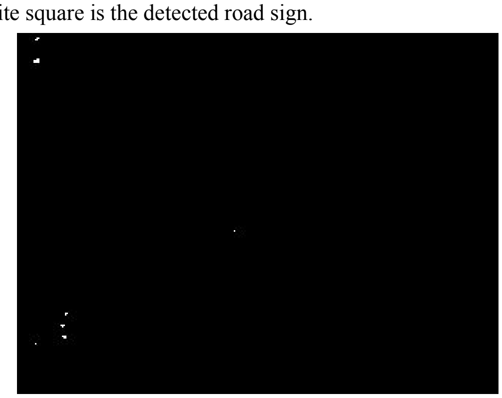


(b)

(d)

Figure 3: The experimental results of one frame in video sequence S_1, (a) the input image; (b) the output of the neural network for color feature detection; (c) the output of the neural network for shape feature detection; and (d) the integrated result of the road sign detection system. White square is the detected road sign.

(a)



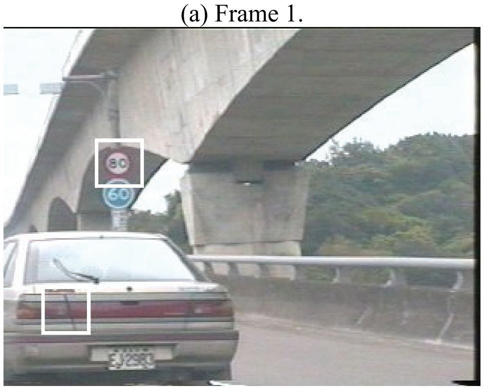
(c)

(b)

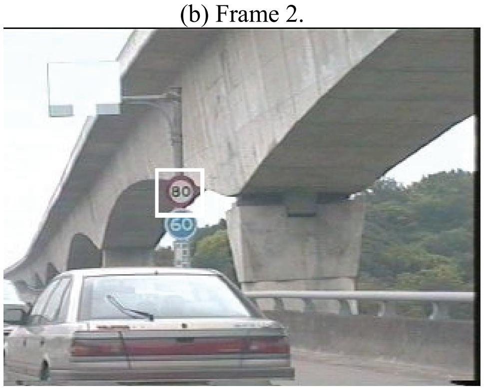
(d)

Figure 4: The experimental results of one frame in video sequence S_2, (a) the input image; (b) the output of the neural network for color feature detection; (c) the output of the neural network for shape feature detection; and (d) the integrated result of the road sign detection system. White squares are the detected road signs. There is a false alarm on the upper left corner.

(a) Frame 1.



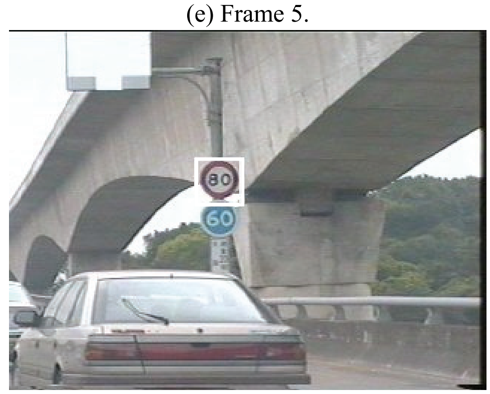
(b) Frame 2.



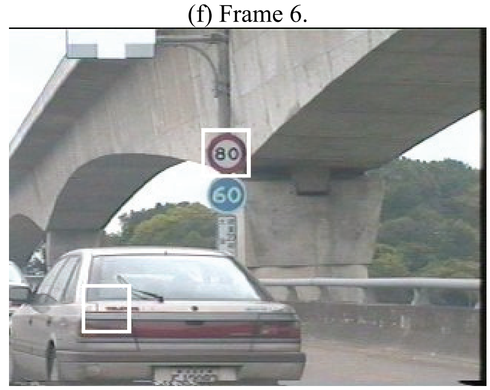
(c) Frame 3.

(d) Frame 4.

(e) Frame 5.



(f) Frame 6.



(g) Frame 7.

(h) Frame 8.

Figure 5: The experimental results with video sequence S_1. The first eight frames of this sequence are shown in (a) to (h). The size of each image is 320×240 pixels and the time between two successive images is 0.2 second.

(a) Frame 1.

(c) Frame 3.

(b) Frame 2.

(d) Frame 4.

Figure 6: The experimental results with video sequence S_2. The first four frames of this sequence are shown in (a) to (d). The size of each image is 320×240 pixels and the time between two successive images is 0.2 second.