Road Sign Detection and Tracking from Complex Background

By
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.

Abstract

Road sign detection and tracking is one of the major tasks in a visual driver-assistance system. This paper describes an approach to detecting and tracking road signs appearing in complex traffic scenes. In the detecting phase, two neural networks are developed to extract color and shape features of traffic signs, respectively, from scenes images. A process characterized by fuzzy-set discipline is then invoked to search for road signs in the images on the basis of the extracted features. In the tracking phase, Kalman filter is employed to predict the positions and sizes of the road signs located earlier. The experimental results show that the proposed method can perform well in both detecting and tracking road signs present in complex scenes and in various weather and illumination conditions.

Keyword: Road sign detection and tracking, HSI color model, Neural networks, Fuzzy integration, Kalman filter

I. Introduction

Road signs are usually designed using particular colors, such as red, green, blue, and orange, and simple geometric shapes, including circle (e.g. railroad advance warning), rectangle (e.g. regulatory signs or guide signs), octagon (e.g. STOP signs), and equilateral triangle (e.g. YIELD signs). Thus, color and shape characterizations of road signs provide useful a priori information for automatic road sign detection [2, 3, 4, 5, 6]. While road signs are made up particular colors and simple shapes, these do does not reduce the difficulty of road sign detection and tracking. For example, on account of weather and lighting conditions, outdoor illumination typically varies from time to time. Moreover, the paint on signs also deteriorates with time. To overcome these difficulties, we developed a visual system for detection and tracking of road signs.

II. The Detection and Tracking System

The outline of our road sign detection and tracking system is shown in Figure 1. First, one image, $S^{(t)}$, of an input video sequence S is fed into the road sign detection and tracking system, as Figure 2(a) shows. Second, the color image is split into H (hue), S (saturation), and I (intensity) channels [1, 7], and only the hue values of some particular colors is calculated to form image $H^{(t)}$ [Figure 2(b)] for extracting the color feature. The color feature is defined as the centers of some special color regions. Third,

an edge detection method is applied to acquire the gradient values to construct the edge map, $E^{(t)}$ [Figure 2 (c)], for extracting the shape feature. The shape feature is defined as the centers of some fixed shape regions. Color and shape features are detected by neural network 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. In other words, these two neural networks have the same structure and the sketch of the neural network is shown in Figure 3.

In both of the neural networks, if (R, G, B) indicates the red, green, and blue color values of one pixel in the color image and input to neuron q (k, l) on the input layer. Then, the output x of neuron q on the input layer in color feature detection neural network is defined as follows:

$$x = \frac{360 - |h - h_0|}{360}, \text{ where}$$

$$h = \begin{cases} \frac{180}{\pi}\cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}}\right\} & \text{if } (G-B) \ge 0 \\ -\frac{180}{\pi}\cos^{-1}\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}}\right\} & \text{if } (G-B) < 0 \end{cases}$$

and h_0 indicates the hue value of the standard "specific" color. The weight $w_{kl,ij}$ between neuron q (k, l) on the input layer and neuron p (i, j) on the output layer is defined as follows:

$$w_{kl,ij} = \begin{cases} \frac{1}{2\pi r} & \text{if } T_1 \le r \le T_2 \\ \frac{-1}{2\pi r} & \text{if } 0 < r \le T_1 \\ 0 & \text{otherwise} \end{cases}$$

where $r = \sqrt{(i-k)^2 + (j-l)^2}$; T_1, T_2 are the inner and outer radii, respectively, of the ring region in the road signs.

On the other hand, the output x of neuron q (k, l) on the input layer in shape feature detection neural network is defined as follows:

$$x = \sqrt{(I_{kl} - I_{k-1\,l})^2 + (I_{kl} - I_{k\,l-1})^2},$$

where $I = (R+G+B)/3$. The weight $w_{kl,ij}$ between neuron q (k, l) on the input layer and neuron p (i, j) on the output layer is defined as follows:

$$w_{kl,ij} = \begin{cases} \frac{-1}{2\pi r} & \text{if } T_1 < r < T_2 \\ \frac{1}{2\pi r} & \text{if } T_1 - C \le r \le T_1 \text{ or } T_2 \le r \le T_2 + C \\ 0 & \text{otherwise} \end{cases}$$

here C is a constant depends on the method of edge detection.

Fourth, the color and shape features are integrated to locate the centers of road signs. The color and shape features extracted by neural networks, the locations of the centers of road signs from the viewpoints of color and shape characterizations respectively, are not exactly the same, thus the fuzzy approach is utilized for accomplishing the integration step.

The fuzzy degree of pixel (i, j) in the input image belonging to road signs is then defined as follows:

$$\mu(i,j) = \frac{w_c \times y^c_{k'l'} + w_s \times y^s_{m'n'}}{D + \varepsilon},$$

where $y^c_{k'l'} = \max_{(k,l)\in N_{ij}} (y^c_{kl})$; $y^s_{m'n'} = \max_{(m,n)\in N_{ij}} (y^s_{mn})$; $D = \sqrt{(k'-m')^2 + (l'-n')^2}$; N_{ij} indicates the set of neurons which are neighbors of the neuron (i, j); w_c and w_s are the weights of color and shape features, respectively, and ε is a positive constant to avoid division by zero.

Now, the location function of road signs, $L(i, j)$, can be defined:

$$L(i,j) = \begin{cases} 1 & \text{if } \mu(i,j) > \mu(i_1,j_1) \text{ for all} \\ & i - R \le i_1 \le i + R,\ j - R \le j_1 \le j + R \\ & \text{and } \mu(i,j) > T_3 \\ 0 & \text{otherwise} \end{cases}$$

here R is the initial size of road sign, and T_3 is a threshold to detect local maximum of membership values. The integration results of Figures 2 (b) and (c) are shown in Figures 2 (d) and (e).

After the integration step, a simple verification process is utilized for verifying the authenticity of the road sign candidates, and the result is shown in Figure 2 (f). Finally, the

parameters, including the positions and radii, of the road signs are predicted by Kalman filter technique [8] to update the parameters of the feature extraction procedures, and these parameters of road signs provide the immediate information for later processing successive image $S^{(t+1)}$. The operation of the Kalman filter [9] is shown in Figure 4, and we define the state vector s, the measurement vector z, the transform matrices H and A as follows:

$$s = \begin{bmatrix} x \\ y \\ 1 \\ \overline{r} \\ 1 \end{bmatrix}, \quad z = \begin{bmatrix} x \\ y \\ 1 \\ \overline{r} \\ 1 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$A = \begin{bmatrix} \frac{1}{1-\frac{v\Delta t}{Rf}r} & 0 & 0 & 0 \\ 0 & \frac{1}{1-\frac{v\Delta t}{Rf}r} & 0 & 0 \\ 0 & 0 & 1 & \frac{-1}{1-\frac{v\Delta t}{Rf}r} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where f is the camera constant; v means car velocity; R is the radius of road sign; r indicates the projection radius of road sign; x and y indicate the projection x-axis and y-axis distance between the road sign and camera lens, respectively; and Δt indicates the time between time t and $t+1$.

III. Experimental Results

In the experiment, we got some video sequences from the camcorder mounted on a vehicle, and converted these analog signals into digital image sequences by computer software (Adobe Premlere 5.1). The size of each image is 320×240 pixels and the time between two successive images is 0.2 second. Figure 5 shows the experimental results with sequence S_1, the first fifteen frames of this sequence are shown in (a) to (o).

In the beginning, when the first one image of the sequence S_1 fed in, our system detected the red circular road sign candidates whose radius is eight pixels by extracting the color and shape features. And then, Kalman filter is utilized to estimate the parameters of road signs with the speed of the vehicle for the next image. Thus, if the next image feeds in, we can only search the neighborhood of the estimated centers, instead of the whole image, of the road signs to track their trajectories.

If a road sign is large and clear enough, we could get accurate verification result, because the larger the road sign is, the more complete information it can give us. On the other hand, verifying the road signs too late will wastes too much time to track the impossible candidates. Thus, we decide to verify a road sign when its radius is larger than ten pixels.

Our system can tolerate some tracking mistakes. If a road sign candidate misses in one image, then we replaced its detected position and size by estimated position and size, predicted by Kalman filter in previous images, until it has been detected again. Nevertheless, if a road sign candidate is continuously lost in five successive images, then we give up to estimate it.

From Figure 5 (a) (R_{cir}=8, T_{cir}=1, U=0.5), we knew that there are three road sign candidates detected by our system, but two of them are partially occluding in next image (Figure 5 (b)). This situation causes the tracking mistake. The real road sign can not be correctly detected indeed. However, the mistake was recovered later (Figure 5 (c)). Figure 5 (d) shows that the road sign is correctly detected even partially occluding in some situations. We verified the road sign candidates from the fifth image (Figure 5 (e)) of the sequence. However, verification does not have any effect on this image because there is only one correct candidate in it. Compared (d) with (e) in Figure 6, we can observe the verification effect obviously.

Figure 6 illustrates the experimental results with complex background. Here we frame the red road signs with red boxes. These examples show that our road sign detection method also can be suitable to extract the road signs in various complex backgrounds.

III. Conclusions

In summary, this paper describes an approach to detecting and tracking 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. The technique of Kalman filter is used to predict the positions and radii of the road signs. Experimental results have manifested the applicability of the proposed method.

Reference

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

[2] 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*)

[3] 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*)

[4] 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.

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

[6] 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.

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

[8] Taygeta Scientific Incorporated, "A Kalman Filter C++ Class," *http://www.taygeta.com/ skip.html/*, 1999.

[9] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," *http://www.cs.unc.edu/~welch/ kalmanIntro.html*, 1999.

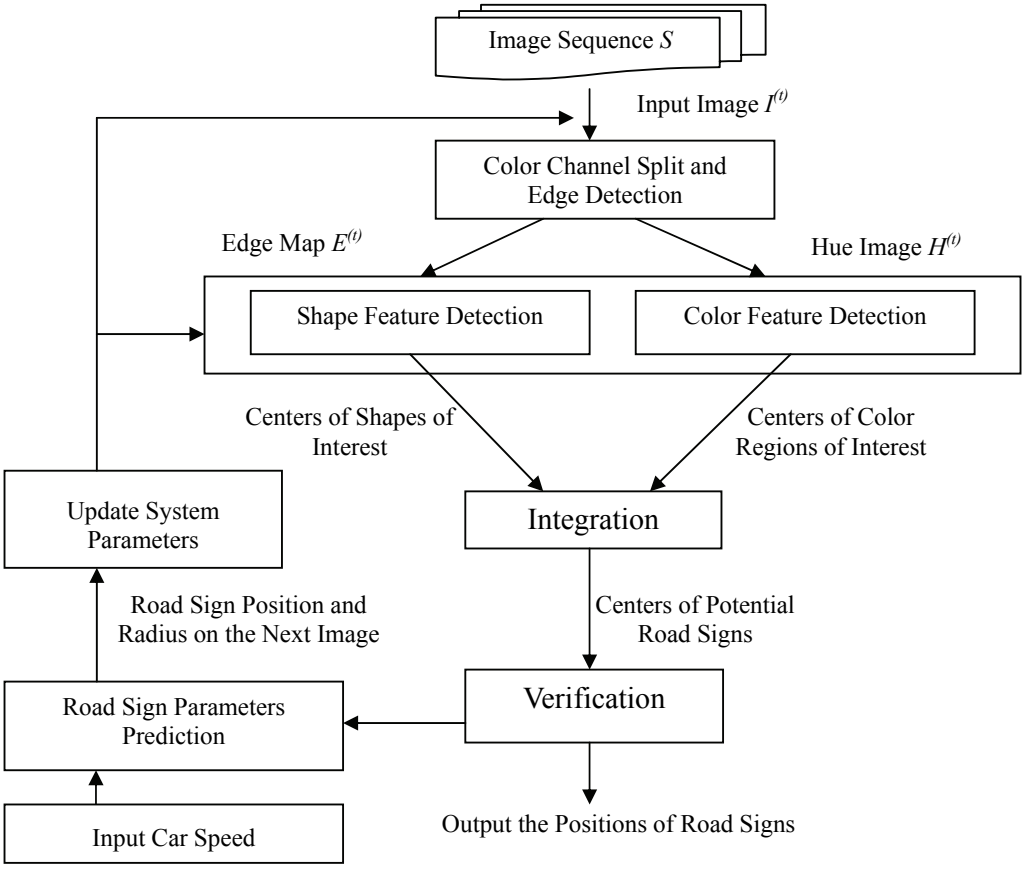

Figure 1: Outline of the road sign detection and tracking system.

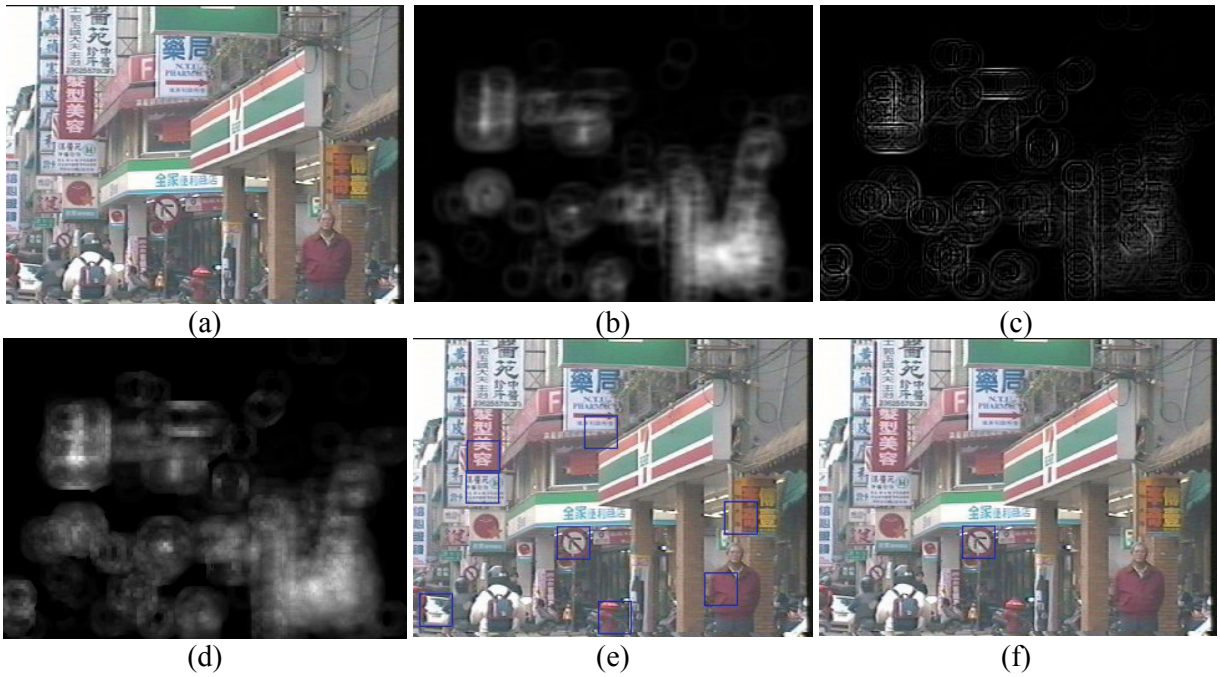

Figure 2. An example of road sign detection process. (a) An original input image. (b) The color feature map of the input image. (c) The shape feature map of the input image. (d) The integration map of color and shape features. (e) The result after integration step. (f) The result after the verification step.

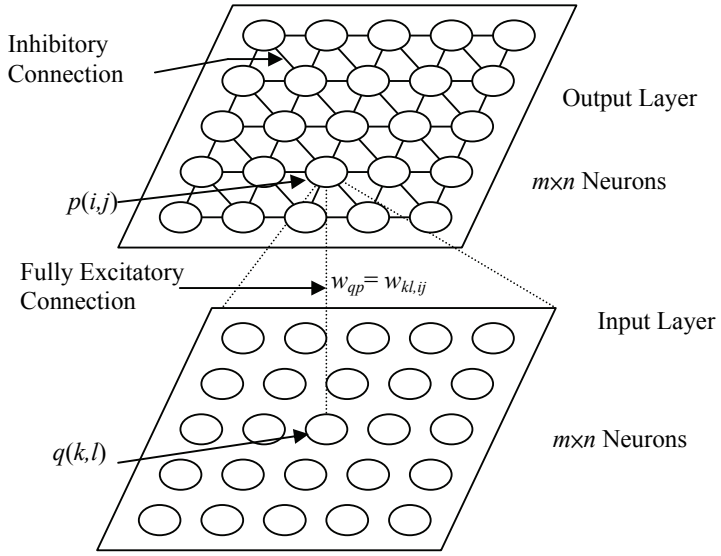

Figure 3: Sketch of the neural network for color (shape) feature detection, where (*i, j*) is the position of neuron *p* on the output layer, (*k, l*) is the position of neuron *q* on the input layer, and $w_{kl,ij}$ is the weight between neuron *q* on the input layer and neuron *p* on the output layer.

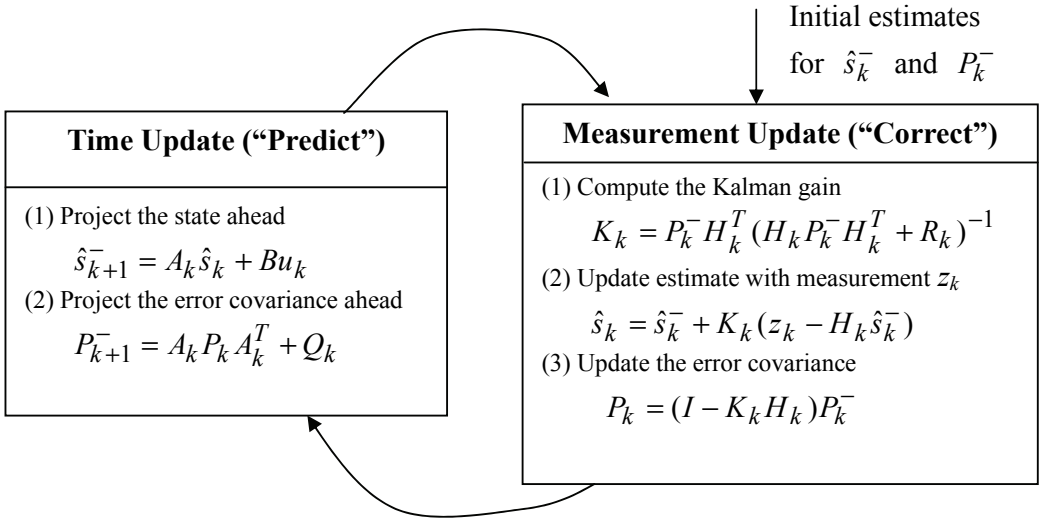

Figure 4: A complete picture of the operation of the Kalman filter [9], where s_k means the state vector; $\hat{s}_k^-$ is the a priori state estimate at step *k*; $\hat{s}_k$ is the a posteriori state estimate; P_k indicates the a posteriori estimate error covariance; P_k^- is the a priori estimate error covariance; z_k is the measurement vector; H_k is the transform matrix between z_k and s_k; A_k is the transform matrix between s_k and s_{k+1}; K_k means the Kalman gain; B is the weight matrix of the control input; u_k is the control input; and R_k and Q_k indicate the variances of the measurement noise and the process noise, respectively.

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

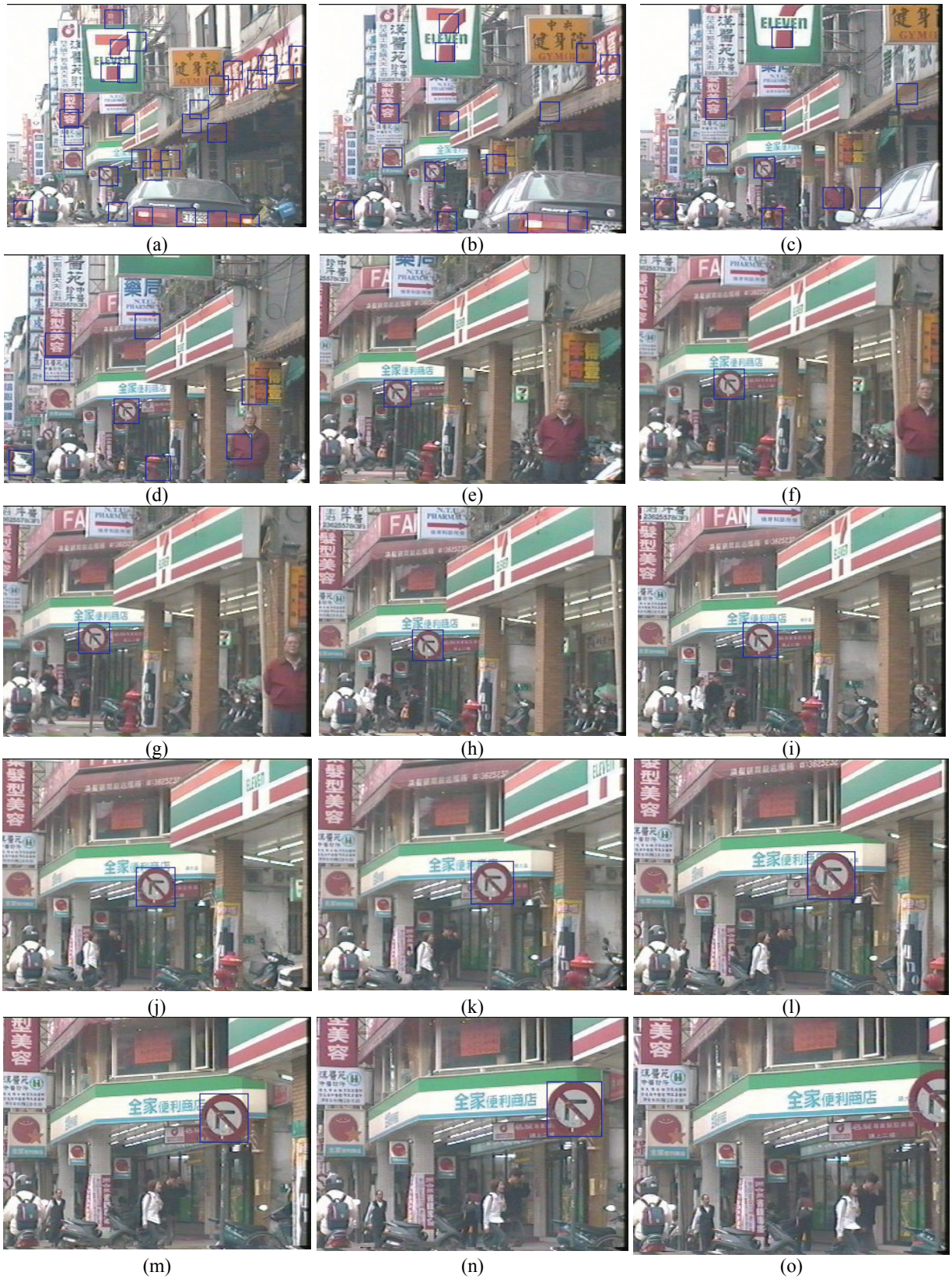

Figure 6: The experimental results with video sequence S_2. The first fifteen frames of this sequence are shown in (a) to (o). The size of each image is 320×240 pixels and the time between two successive images is 0.2 second.