Paper No. 01-2580

Vision-Based Stop Sign Detection and Recognition System for Intelligent Vehicles

Henry X. Liu and Bin Ran

The traffic sign detection and recognition system is an essential module of the driver warning and assistance system. A vision-based stop sign detection and recognition system is presented here. This system has two main modules: detection and recognition. In the detection module, the color thresholding in hue, saturation, and value color space is used to segment the image. The features of the traffic sign are investigated and used to detect potential objects. For the recognition module, one neural network is trained to perform the classification and another one is trained to perform the validation. Joint use of classification and validation networks can reduce the rate of false positives. The reliability demonstrated by the proposed algorithm suggests that this system could be a part of an integrated driver warning and assistance system based on computer vision technologies.

Driving-environment understanding is of great importance for the development of vision-based driver warning and assistance systems. The goal of driving-environment understanding is to detect and recognize important objects, such as other vehicles, lane markers, and traffic signs. Traffic signs provide drivers with valuable information about the driving environment. A system capable of detecting and recognizing traffic signs is an essential module of the driver warning and assistance system. It could be used as assistance for drivers to warn them about the presence of a specific sign, such as a stop sign, or of some risky situation such as driving at a speed higher than the speed limit.

The difficulties of traffic sign detection and recognition come from several aspects. First, the system should perform in real time. This capability requires efficient algorithms and high-performance hardware. In addition, the complex outside conditions should be considered, including lighting, shadows, occlusion, and different types of weather (sunny, rainy, foggy, etc.). Finally, there are a variety of traffic signs, which are characterized by their colors, shapes, and pictographic symbols.

In recent years, many studies on the detection and recognition of traffic signs have been conducted. As a part of the European research project PROMETHEUS, a vision system for traffic sign recognition was developed in the Daimler-Benz Research Center and integrated in the Daimler-Benz autonomous vehicle (*1*, pp. 635–638; *2*, pp. 624–631; *3*, pp. 12–17; *4*, pp. 165–170; *5*, pp. 213–218). The software architecture of the system integrated three hierarchical levels of data processing. At each level, specific tasks were isolated. The lowest level included specialists for color, shape, and pictogram analysis, which performed the iconic-to-symbolic data transformation. At the highest level, administration processes organized data flow as a double bottom-up and top-down mechanism to dynamically interpret the image sequence. Although the system was very fast (200-ms cycle time), the hardware was complex and expensive. It consisted of four power PCs and four transputers.

Blancard's algorithm (*6*, pp. 162–175) recognized the signs by their color and form. For the color classification, a passband filter for the chosen color (red) was attached to a black-and-white charged-coupled device (CCD) camera. Then the Sobel filter was applied, and the edges were found in the image by using the Freeman code. Some features were calculated from the resulting contours: perimeter, length and compactness, and so forth. These features were the input to the neural-network-type restricted coulomb energy for classification. The detection was at a fixed distance from the vehicle and could only tell the sign type, not the exact sign.

Escalera et al. (*7*) used color thresholding for image segmentation and shape analysis in order to detect signs. For classification, they used a neural network. Their system cannot be used to detect stop and yield signs. Piccioli et al. (*8*, pp. 278–283) used black-and-white images. After the extraction of edges, the shape analysis was used to look for the circular and triangular signs, and the classification was conducted through cross-correlation with a database.

In this paper, the proposed algorithm has two main modules: detection and recognition. In the detection module, color thresholding in hue, saturation, and value (HSV) color space is used to segment the image. The features of traffic signs are investigated and used to detect potential objects. Once the potential objects are found, they are clipped from the background, normalized to a specified size, and sent to the recognition module. Basically, the recognition module consists of a classification neural network and a validation neural network. The input of the classification neural network is the intensity image of the potential object, and the output is which sign it is. An overview of the proposed system is presented next, followed by presentation of the traffic sign detection and recognition modules. Representative experimental results and conclusions are offered last.

SYSTEM OVERVIEW

The input data for this vision-based system consists of the color image sequence taken from a moving vehicle. A single CCD camera is mounted inside the vehicle behind the windshield along the central line. It takes the images of the environment in front of the

H. X. Liu, California PATH ATMS Center, Institute of Transportation Studies, 523 Social Science Tower, University of California–Irvine, Irvine, CA 92697. B. Ran, Department of Civil and Environmental Engineering, University of Wisconsin at Madison, 2256 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706.

vehicle, including the road, vehicles on the road, traffic signs on the roadside, and sometimes incident objects on the road. The on-board computer with image-capturing card captures the images in real time (up to 30 frames/s) and saves them in the computer memory. Then the traffic sign detection and recognition system takes images from the memory and starts processing.

In this paper, the hardware system consists of a Sony CCD TR400 video camera, 166-MHz Pentium PC, and an AV Master image capture board. The recorded data taken from the driving environment are played back and processed.

As shown in Figure 1, the system has two main modules: detection and recognition. During the detection process, each color image taken with the camera is first segmented into regions according to the color values of each pixel. The HSV color space is used in the segmentation process. A median filter is applied to the segmented binary image to remove noise. In order to identify potential traffic signs, the features of the objects are investigated. Three criteria are used to distinguish potential objects from the background: object size, aspect ratio, and symmetrical level of objects. In the recognition model, both classification and validation networks are two-layer, feedforward, backpropagation neural networks. There are 900 inputs, 6 hidden neurons, and 1 output neuron in the classification network and 900 inputs, 100 hidden neurons, and 900 output neurons in the validation network. Although only the stop sign is investigated in this paper, the detection and recognition algorithms can be easily extended to identify other kinds of traffic signs.

STOP SIGN DETECTION

Four types of traffic signs are shown in the traffic code: warning, prohibition, obligation, and information. Different signs have different colors, shapes, and sizes. The meaning of the shapes and colors for U.S. road symbol signs are described in the Manual on Uniform Traffic Control Devices (*9*). For the stop sign, the color is red and the shape is octagonal.

The image taken from a moving vehicle is represented by three primary monochrome colors, namely, red, green, and blue. Figure 2*a* is an example of the input image. In the detection model, the first step is to convert the red-green-blue (RGB) color space to the HSV color space, and the image will be segmented according to the HSV value of each pixel. Objects in the segmented image are labeled according to each pixel's 8-connectivity. Potential stop signs are identified using the features of the object. The object is marked using a bounding box, and the equalized intensity images of the objects are sent to the recognition model. Figure 2*b*–*f* shows the results of the sample image.

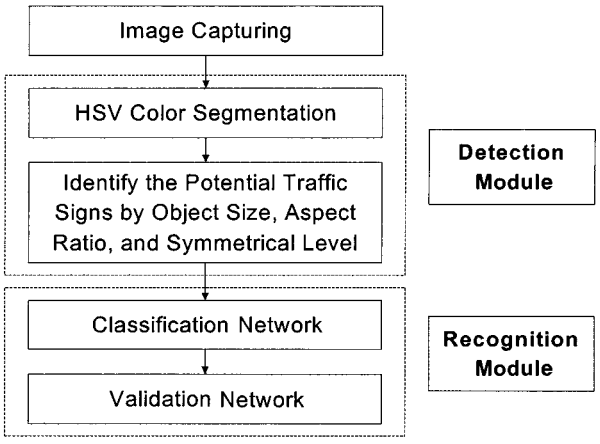


FIGURE 1 Proposed traffic sign detection and recognition system overview.

Image Segmentation

The color segmentation is based on the color space model RGB-to-HSV conversion, followed by the proper thresholds on the hue and saturation bands to extract the red, blue, and green colors of the signs. This approach provides a hue component decorrelated from the intensity and allows outdoor light variations.

The HSV color space is often used for picking colors (e.g., colors of paints or inks) from a color wheel or palette because it corresponds better to how people experience color than the RGB color space does. As hue varies from 0 to 1.0, the corresponding colors vary from red through yellow, green, cyan, blue, and magenta, and back to red. As saturation varies from 0 to 1.0, the corresponding colors change from unsaturated (shades of gray) to fully saturated (no white component). As value or brightness varies from 0 to 1.0, the corresponding colors become increasingly brighter (Figure 3).

A statistical study containing different traffic signs was performed for the purpose of bounding the subspace of the HSV color space. The red color of stop signs was mapped into the HSV color space for a large number of stop signs with different backgrounds and ambient light conditions. This study indicated that stop signs are contained in the subspace spanned by $h < 0.05$ and $h > 0.95$, $s > 0.5$, and $v > 0.01$. These values were then used as discriminant functions to segment possible stop signs. A binary image is formed by only those points that fall into this subspace. Figure 2*c* is an example of the segmented image for stop signs.

Criteria for Choosing Potential Object

In order to detect potential stop signs, several features of the stop sign in the image are investigated. The features used here are object width, aspect ratio, and symmetrical level. Figure 2*d* is an example image after application of these criteria.

1. Object width (number of horizontal pixels of the object): Once the camera is calibrated in advance, the focal length of the camera is fixed. From the pinhole camera model (*10*) and the perspective equation, it is known that the width of the object in the image corresponds to the real distance between the object and the subject vehicle. In this case, 20-pixel width in the image equals about 150 ft (45 m). If the width of the object is smaller than 20 pixels, the distance will be longer than 150 ft. A traffic sign that is more than 150 ft cannot be detected because of the resolution of the camera. On the other hand, if the width of the object is greater than 20 pixels, the distance will be shorter than 150 ft. This object will be considered as a potential traffic sign.
2. Aspect ratio: Although the width of the object is greater than 20 pixels, sometimes the aspect ratio of the selected object is not in the range of a traffic sign (for example, the aspect ratio of a stop sign is about 1.0) or it may not be a real object. If the aspect ratio of the object is not in the range, it will not be selected as the potential object.

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

FIGURE 2 Example of traffic sign detection: (*a*) original color image (240 * 320), (*b*) gray scale image, (*c*) segmented image, (*d*) gray scale image after histogram equalization with potential objects being detected (drawn by bounding box), (*e*) potential Object 1 being normalized and ready to be sent to recognition module, (*f*) potential Object 2 being normalized and ready to be sent to recognition module.

3. Symmetrical level: This feature is used because the stop sign has a symmetrical shape. In order to calculate the symmetrical level of the object, the vertical symmetry axis of the object should be found. Then the symmetrical level of the object will be calculated by

$$S = \frac{D}{W}$$

where

S = symmetrical level of object (range: 0 ~ 1).
D = distance between symmetry axis by calculation and real symmetry axis, and
W = width of object.

According to the principles of the Hough transform (*10*), a "voting" scheme is used to detect the vertical symmetry axis of the object. Each vertical axis within the object area becomes a candidate. Each pair of pixels in one row within this area is forced to vote for their axis. Then, among all the axis candidates, the axis that receives the maximal number of votes will be the symmetry axis of this object.

Histogram Equalization

The image histogram equalization needs to be done before the potential objects are cropped from the gray scale image. The goal is to eliminate the influence of different light conditions, which may cause the intensity histogram of the image to be concentrated in different ranges. Histogram equalization is used to spread out the intensity and make the image easier to analyze. Figure 2*d* is the gray scale image after histogram equalization. Figure 2*e* and 2*f* show the potential object ready to be sent to the recognition model for further classification.

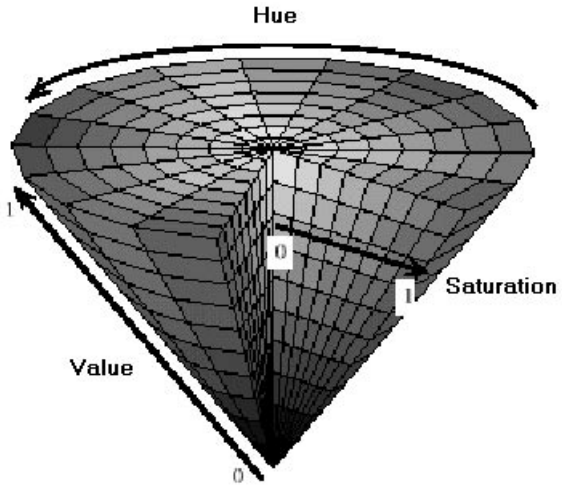


FIGURE 3 HSV color space.

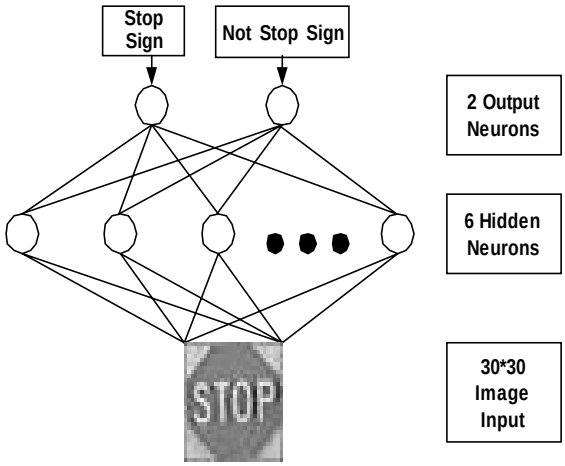


FIGURE 4 Neural network structure.

STOP SIGN RECOGNITION

Although a potential object satisfies the three criteria above, it may be a real stop sign or it may be something else. Sometimes a patch of red wall will be selected as one of the potential objects. The recognition model is used to distinguish the stop sign from other potential objects.

The classification takes place through a neural network trained to recognize stop signs. The neural network is a two-layer, feedforward, backpropagation training network. The input for the neural network is a 30- by 30-pixel gray scale image, and the output is whether the potential object is a stop sign. Once the stop sign is recognized, the validation network performs input reconstruction and decides whether the input for a stop sign can be validated. The joint use of classification and validation networks can reduce the rate of false positives.

CLASSIFICATION NEURAL NETWORK

Network Structure

The classification neural network is shown in Figure 4. The neural network has 30-by-30 images as its input pattern and two output neurons in its output layer to identify the object. Because it is hard to obtain some important features from the nature of the object and give them to the neural network as the input patterns, the raw image input is given to the network and the network is allowed learn from experience what features are important. The object region of a gray scale image is taken as the input. Because the image size of the object region is different from time to time, it is normalized to 30 by 30 pixels and taken as the input pattern. The two output neurons represent whether the potential object is a stop sign. If the first output neuron is active, this object is a stop sign; otherwise it is not. Since noisy input vectors may result in the network's not creating perfect 1 and 0, the output of the neural network passes through the competitive function and the bigger one will be active. The result of this postprocessing is the output that is actually used. The transfer function of the neural network is a two-layer sigmoid function. The sigmoid transfer function was picked because its output range (0 to 1) is perfect for learning to output boolean values.

Parameter Setting

The hidden (first) layer has six neurons. The choice of this number is based on experiments. Tenfold cross-validation is used here. The relationship between the average error rate and the number of hidden units is shown in Figure 5. The number of hidden neurons was

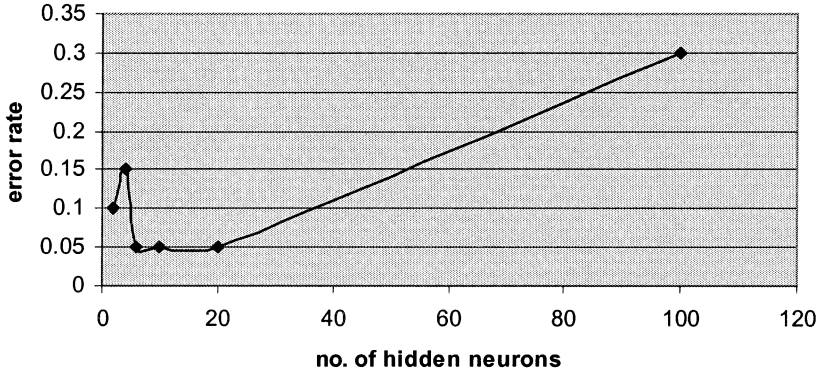


FIGURE 5 Error rate versus number of hidden neurons.

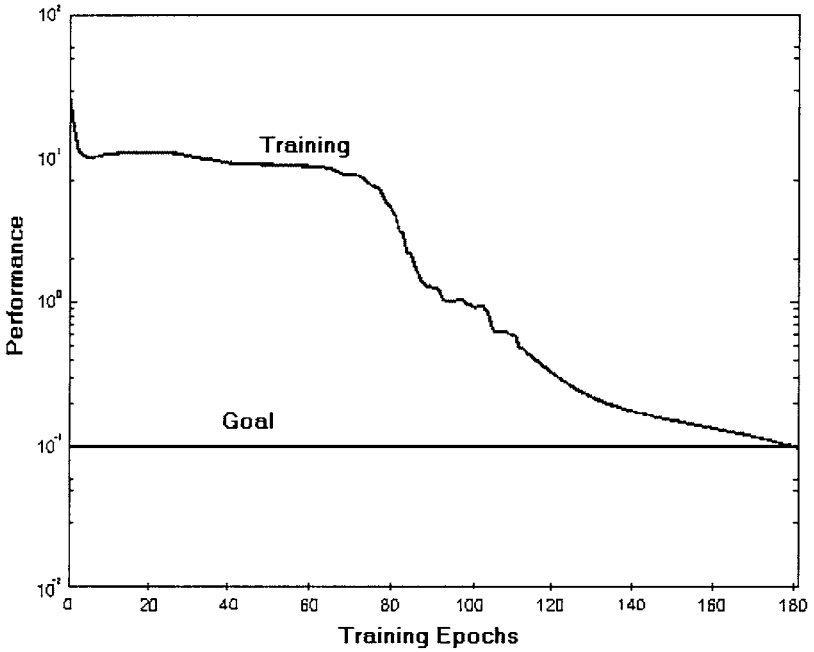


FIGURE 6 Learning curve of classification network.

set at six because it has the same error rate with 10 and 20, and obviously it is computationally cheaper than the other two values.

Network Training

In total, there are 241 examples in the data set. All of them are the cropped and normalized 30-by-30 gray scale images. These images are the potential objects identified by the detection module from the images taken in the real driving environment under different conditions. Among them, some are stop signs, but some are not. For the training set, 200 examples are randomly selected, and the other 41 examples are the test set. Figure 6 is the learning curve of the network. The mean value of the sum of squares of the network errors is used as the performance function, and the training continues until the network sum-squared error falls below 0.1.

Validation Network

In order to reduce the false positive rate, a validation network is trained using autoassociation. The purpose is to remap the output to the input in an attempt to reconstruct the input after it has been classified. Once the input has been re-created by the validation network, a measure of the closeness between the actual input and the validated replica of the input needs to be computed. A decision on whether the input can be validated will be made on the basis of the measure of closeness. If it cannot be validated, the classification will be overturned.

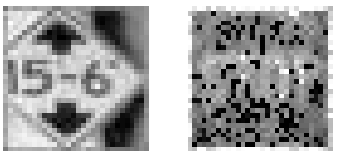
One way to ensure that the network explicitly represents specific features from the current input in its hidden representation is to force the network to re-create the input in its output (*11*). Here the validation network consists of 900 inputs, 100 hidden neurons, and 900 outputs. The desired activation pattern for the output units is identical to the input pattern. This network is trained by using the same data with the classification network. The number of hidden neurons is set to 100 because the smaller number does not guarantee the convergence.

Regarding the measure of closeness, a straight pixel-by-pixel distance metric such as Euclidean distance would be a poor measure because the input image and reconstructed image have very different means and variances (*11*). A more straightforward technique for determining the closeness of two images is to compute the correlation coefficient between them. The correlation coefficient r between image R and image T (R and T are the same size) is defined as follows:

$$r = \frac{\sum_{x,y} R(x,y)T(x,y) - \left[\sum_{x,y} R(x,y)\right]\left[\sum_{x,y} T(x,y)\right]}{\sqrt{\sum_{x,y} R(x,y)^2 - \left[\sum_{x,y} R(x,y)\right]^2}\sqrt{\sum_{x,y} T(x,y)^2 - \left[\sum_{x,y} T(x,y)\right]^2}}$$

where x, y are the spatial coordinates of the image.

Figure 7 illustrates how the validation network is used to reduce false positives. By using the classification network, Figure 7*a* is classified as a stop sign. However, the correlation coefficient between



(a) **(b)**

FIGURE 7 Validation network for overturning decision of classification network (correlation coefficient = 0.0895): (*a*) real input, (*b*) input reconstructed by validation network.

the real input (Figure 7*a*) and the input reconstructed by the validation network (Figure 7*b*) is very low. Therefore, this classification decision is overturned.

EXPERIMENTAL RESULTS AND CONCLUSIONS

In this paper, all the algorithms were implemented by using Matlab 5.3. In the experiment, 540 images were used to detect the potential objects, and 524 images were detected correctly; that is, if there was a stop sign in the image, this stop sign was detected and ready for classification. Thus, the correct detection rate in the experiment was about 95 percent. About 26 images do contain stop signs, but the signs were not detected. The major reason for the missing classification is the occlusion; some stop signs are shadowed and occluded by a tree. The 241 training examples for the classification network and validation network came from those 485 correctly detected images (some images contained no potential signs). The 241 training examples were divided into a training set (200 examples) and a test set (41 examples). After the network training, two examples were classified to the wrong categories by the classification network. One of them is false positive, and it is prevented by the validation network (Figure 7). Thus, the error rate of the recognition model in the experiment is about 0.025.

In summary, a novel approach to the detection and recognition of stop signs has been presented. The algorithm has two main modules: detection and recognition. In the detection module, color thresholding in hue, saturation, and value (HSV) color space was used to segment the image. The features of traffic signs were investigated and used to detect potential objects. For recognition, the joint use of classification and validation networks can reduce the false positive rate. The experiment showed that the system works robustly.

Although only the stop sign was investigated in this study, the detection and recognition algorithms can be easily extended to identify other kinds of traffic signs. This further identification is one direction for future research. Moreover, no object tracking is considered in this study. The correct use of image sequence information would make the computation cheaper and faster. The third direction for this research is to improve detection accuracy by engaging more sophisticated algorithms.

REFERENCES

1. Bartneck, N., and W. Ritter. Color Segmentation with Polynomial Classification. In *Proc., 11th International Conference on Pattern Recognition*, International Association on Pattern Recognition, Vol. 2, 1992.
2. Besserer, B., S. Estable, and B. Ulmer. Multiple Knowledge Sources and Evidential Reasoning for Shape Recognition. In *Proc., 4th IEEE International Conference on Computer Vision*, Berlin, 1993.
3. Ritter, W. Traffic Sign Recognition in Color Image Sequences. In *Proc., IEEE Intelligent Vehicles Symposium*, Detroit, Mich., 1992.
4. Zheng, Y., W. Ritter, and R. Janssen. An Adaptive System for Traffic Sign Recognition. In *Proc., IEEE Intelligent Vehicles Symposium*, Paris, 1994.
5. Estable, S., et al. A Real Time Traffic Sign Recognition System. In *Proc., IEEE Intelligent Vehicles Symposium*, Paris, 1994.
6. Blancard, M. Road Sign Recognition: A Study of Vision-Based Decision Making for Road Environment Recognition. In *Vision-Based Vehicle Guidance*, Springer-Verlag, New York, 1992.
7. Escalera, A., L. Moreno, et al. Road Traffic Sign Detection and Classification. *IEEE Transactions on Industrial Electronics*, Vol. 44, 1997, pp. 848–858.
8. Piccioli, G., et al. Robust Road Sign Detection and Recognition from Image Sequences. In *Proc., IEEE Intelligent Vehicles Symposium*, Paris, 1994.
9. *Manual on Uniform Traffic Control Devices for Streets and Highways*. FHWA, U.S. Department of Transportation, 1988.
10. Trucco, E., and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall, Englewood Cliffs, N.J., 1998.
11. Pomerleau, D. *Neural Network Perception for Mobile Robot Guidance*. Kluwer Academic Publishers, Boston, Mass., 1993.

Publication of this paper sponsored by Committee on Vehicle-Highway Automation.