2017 2nd IEEE International Conference On Recent Trends in Electronics Information & Communication Technology (RTEICT), May 19-20, 2017, India

Automatic Traffic Sign Detection and Recognition in Video Sequences

Swathi M
Dept. of ECE,
Siddaganga Institute of Technology, Tumakuru,
mswathi.sm@gmail.com

K.V. Suresh
Dept. of ECE,
Siddaganga Institute of Technology, Tumakuru,
sureshkvsit@sit.ac.in

Abstract— Traffic sign detection and recognition plays a prominent role in advanced driver assistance systems (ADAS). Traffic signs provide information about the traffic rules, road conditions, route directions and assist the drivers for better and safe driving. This article proposes an algorithm for traffic sign detection and recognition in video frames. The detection is performed in the hue-saturation-value (HSV) color space using both color and shape features and the recognition is performed using the multilayer perceptron (MLP) trained with histogram of oriented gradients (HOG) features. The proposed method implemented on Microsoft Visual Studio-2010 using OPENCV 2.4.8.0 image processing library. Experimental results reveal that a recognition rate of 97.14% and 95.71% is achieved on real-time videos.

***Keywords*— ADAS, Histogram of oriented gradients, Hue-saturation-intensity, Multilayer perceptron, Traffic signs**

I. INTRODUCTION

Nowadays, we have seen an exponential rise in the number of vehicles on roads. Due to this, a huge amount of accidents are happening all over the world. According to ‘World Road Statistics’- a report published by the International Road Federation (IRF), India has recorded second highest road accident deaths in 2015 [1]. These accidents are mainly because of the driver inability to process all the visual information that is available while driving. In order to avoid these accidents advanced driver assistance systems (ADAS) are developed to automate the vehicle systems for safety and better driving. The main applications of ADAS are in blind spot detection, emergency brake assist, traffic sign recognition, lane departure warning etc. As an application of ADAS, traffic sign recognition system plays an important role in assisting the drivers along the routes.

The traffic authorities have placed the traffic signs along the roads to warn about the dangerous road conditions ahead and to provide necessary information to the driver. Traffic signs used in India are similar to those used in United States. Traffic signs are categorized into 3 main types namely, Mandatory, Cautionary and Informatory signs. Mandatory signs provide the information about the rules and regulations to be followed by the driver. Cautionary signs warn the driver about the hazardous road conditions ahead. Informatory signs assist the drivers along the routes by providing the information about the route directions. Sometimes, heavy traffic, weather conditions or miss attention of drivers causes a chance of missing a sign and it might lead to accidents. It is necessary to detect and recognise these traffic signs automatically and alert the driver about the situation.

Developing a traffic sign detection and recognition system in real-time is a more challenging task because of the various factors that affect the performance of the system. Some of the challenges involved in this area are listed here.

- The detection of traffic signs is affected with the varying lighting conditions due to time of the day, weather conditions etc.
- The paint present on the signs may fade with the time
- The traffic signs may be damaged, tilted or rotated due to storms or winds resulting in geometric distortions
- Variation in scale as the vehicle approaches the traffic sign
- Other similar objects present in the scene under consideration
- Occlusion by the trees, street lamps, buildings, pedestrians etc.
- Characteristics and viewing angle of the image acquisition system
- Detection may suffer due to vehicle vibration and motion blur
- Shadows due to trees and other objects

The main objective of this work is to develop an algorithm that can be used to detect and recognize the traffic signs in the real-time videos. In this work, only mandatory and cautionary signs with red background are considered for the implementation. The video segmentation is done in the HSV color space and the shape constraints are applied onto the segmented video frame to detect the signs. Then the MLP is trained with the HOG features to recognize the traffic signs.

The remainder of this paper is organized as follows: Section II provides a brief literature review of the past work done in the area of traffic sign detection and recognition, the proposed method for the traffic sign detection and recognition is discussed in Section III. Section IV provides the experimental results and finally Section V concludes the paper.

II. LITERATURE REVIEW

This section provides a brief review of the past work in the area of traffic sign detection and recognition. Since traffic signs are designed with specific colours and shapes, researchers have used the color and shape features to detect the signs. The authors in [2] used RGB color space for the traffic sign detection using color distance method. Segmentation in the RGB color space reduces the computational complexity however the method is highly sensitive to illumination changes. To overcome the illumination problem other color spaces which separate the chromatic and achromatic components are used. The authors in [3] have used the hue-saturation-intensity (HSI) color space to detect the traffic signs. The second method to overcome illumination issues is by using the shape as a feature. In [4], Garvilla used a distance transform approach to detect the signs. Since it is a template matching approach, it results in high computational cost. The authors in [5] used Hough transform to detect the circular and triangular signs. A radial symmetry transform is used in [6] to detect the traffic signs. The radial symmetry transform detects only the regular polygons and hence is not suitable for the signs with some geometrical distortions. Use of color and shape features separately increases the interferences. A combination of both these features would yield better results. The authors in [7] performed the color segmentation in HSV color space and the detection is performed by using Hough transform.

The recognition can be done by feature matching and machine learning approaches. The authors in [8] performed a comparative analysis of the three feature matching approaches such as scale invariant feature transform (SIFT), speeded up robust features (SURF) and binary robust invariant scalable keypoints (BRISK) and showed that, SIFT provides the better recognition accuracy. Popular machine learning algorithms used in the recognition stage are support vector machines (SVM) and neural networks. The authors in [3] used one to one multilayer perceptron neural networks trained with the resilient back propagation algorithm. A traffic sign recognition system based on the deep convolutional neural network is proposed in [9]. Even though, it can provide good recognition rate, it requires a high computation complexity. The authors in [4] used SVM trained with the HOG features for the traffic sign classification. HOG is invariant to scale variations. In [10], traffic signs are detected using Hu moment invariants and neural networks. Since Hu moments are invariant to rotation, scale and translation it provides high recognition accuracy.

In this section, some of the commonly used techniques for traffic sign detection and recognition are discussed. Even though ample amount of research has been carried out in this area, none of the approaches solved all the problems faced by the system. The proposed system mainly addresses the real-time issues like illumination variations, shadow from other objects, scale variations and the detection of geometrically distorted signs.

III. PROPOSED METHODOLOGY

A. Overview of the Proposed System

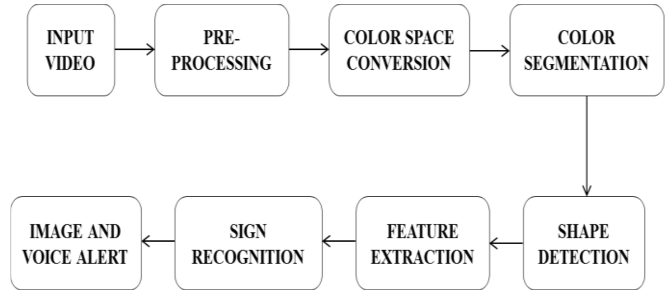

Figure 1. Block diagram of proposed traffic sign detection and recognition system

The block diagram of proposed traffic sign detection and recognition system is shown in Figure 1. The system has two main stages: Traffic sign detection and recognition stages. The detection stage involves the localization of the traffic signs in the video frame and the recognition stage identifies the type of the detected traffic signs.

The input to the system is a video containing the traffic signs. Each frame is extracted from the video and is resized to 500 x 500 pixels. The Region of Interest (ROI) is selected such that, the probability of occurrence of traffic signs in the region is more. Since the traffic signs are designed with specific colours and shapes, these features are used for the detection of the traffic signs. The ROI is converted to HSV color space and shape detection is performed using the Douglass-Peucker algorithm to detect the traffic signs in the segmented video frame. The detected signs are resized, cropped and saved. Then, the HOG features extracted from the detected signs are used to train the multilayer perceptron neural network using backpropagation learning to recognize the type of the traffic signs. Finally, an image and voice alert is given to alert the driver about the traffic signs ahead.

B. Traffis Sign Detection

Traffic sign detection involves identifying the position of traffic signs in video frames. The literature lists several works that combine both color and shape features to reduce the interferences. It includes three main stages: color conversion from RGB to HSV color space, color segmentation and shape detection.

C. Color Space Conversion

Since RGB color space is highly sensitive to illumination variations, HSV color space is chosen for the implementation. A color is represented by hue, saturation and value components in HSV color space. Hue represents the pure color, saturation represents the amount of white light added to the pure color and value gives the brightness information. As HSV color space separates the chromatic and achromatic components, it is invariant to various illumination changes. The ROI from the original video frame is converted to HSV color space. The color conversion from RGB to HSV color space is done as in (1) [11].

$$H = \begin{cases} \dfrac{60(G-B)}{(V-\min(R,G,B))} & if \quad V = R \\ 120 + \dfrac{60(B-R)}{(V-\min(R,G,B))} & if \quad V = G \\ 240 + \dfrac{60(R-G)}{(V-\min(R,G,B))} & if \quad V = B \end{cases}$$

$$S = \begin{cases} \dfrac{V-\min(R,G,B)}{V} & if \quad V \neq 0 \\ 0 & otherwise \end{cases}$$

$$V = \max(R,G,B) \tag{1}$$

1) COLOR SEGMENTATION

Segmentation is a crucial step in the traffic sign detection. It removes the unwanted background part and hence it reduces the search area in the video frames. In this work, traffic signs with red rim are considered for implementation. The segmentation is carried out by thresholding the hue and saturation values as in (2) since they represent the chromatic information in the HSV color space. In OPENCV, Hue ranges from 0 to 180, Saturation ranges from 0 to 255 and Value from 0 to 255 [11].

$$I = \begin{cases} 255 & if \quad 0 \geq Hue \leq 10 \quad and \quad 0 \geq Sat \leq 255 \\ 255 & if \quad 160 \geq Hue \leq 179 \quad and \quad 0 \geq Sat \leq 255 \\ 0 & otherwise \end{cases} \tag{2}$$

2) SHAPE DETECTION

The contours of the segmented frame are detected. The computational cost increases if all the contours are considered for processing and hence only the contours which have the area above and below the upper and lower thresholds are retained. The shape detection is carried out by a contour approximation technique called Douglass-Peucker algorithm (DP) [12]. For any given curve composed of line segments, DP algorithm finds a similar curve with less number of points. The simplified curve consists of subset of points defined in the original curve. The algorithm works as follows: Any two arbitrary points A and B are selected from the contour as the first and last point respectively. Point A and B are connected through a line segment and the farthest point C on the curve to line AB is determined. The distance *(d)* between the point C and the line is calculated. If the distance is less than the threshold, then the point is ignored and AB will be the final output. Otherwise, farthest point search is repeated for the line AC and BC. This process continues until all the points in the contour are searched. Based on the number of lines in the final curve, the contour is approximated to a polygon. The triangular approximation using DP algorithm is shown in Figure 2. The triangular and octagonal shapes are approximated if total number of lines in the final curve is 3 and 8 respectively. For approximating the circular shape, from the experimentation we have found that the total number of lines in the final curve ranges from 9 to 12. As the shape detection is performed based on the number of lines in the final curve, this method is suitable for the detection of traffic signs with some geometric distortions. The various stages of the traffic sign detection in a video frame is shown in Figure 3.

D. RECOGNITION

This stage involves identifying the type of the detected traffic signs in the video frames. The detected traffic sign is identified by training the HOG features with multilayer perceptron using back propagation learning algorithm. The traffic sign recognition involves three steps: Pre-processing, Feature extraction and Recognition.

1) Pre-Processing

To perform the recognition, the detected traffic signs are cropped and resized to 24 x 20 pixels. Since the bounding box used is of rectangular shape, the cropped traffic sign will have

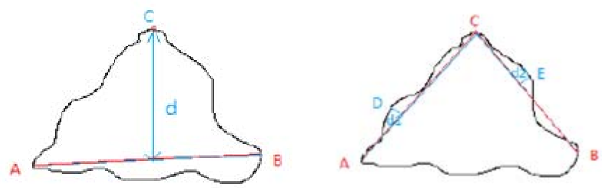

Figure 2. Triangular approximation using Douglass-Peucker algorithm, Courtesy: [12]

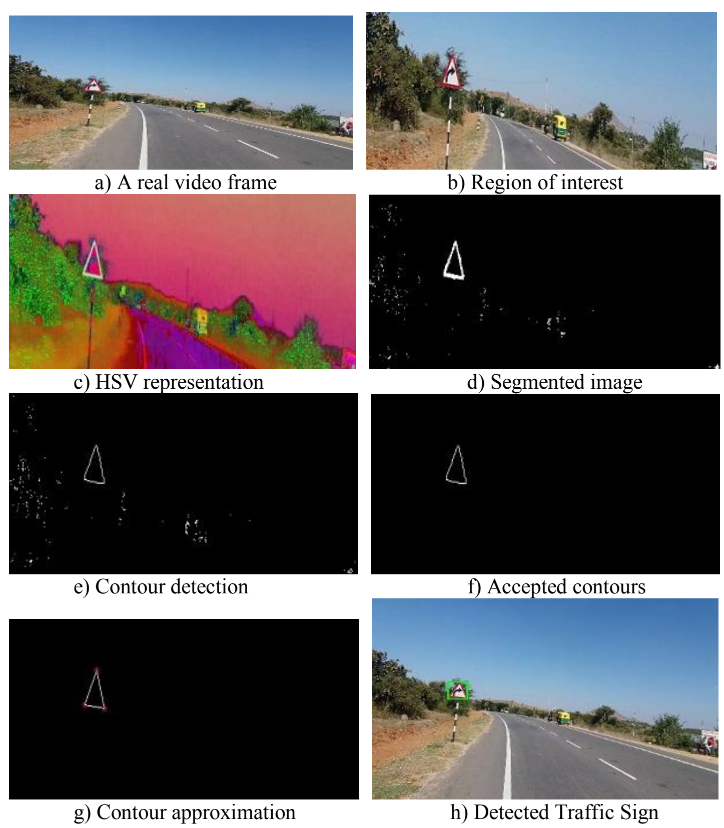

Figure 3. Various stages of the traffic sign detection in a video frame

some background which may yield undesirable results during recognition. In order to remove the background, hole filling operation is done inside the contours. Finally, AND operation is performed on the hole filled and the detected traffic signs to remove the background interferences. Hole filling or region filling is a morphological operation based on dilation followed by complementation and intersection. The main aim of hole filling operation is to fill the holes surrounded by a 8 connected boundaries of foreground pixels.

Let *A* be the original image and *B* be the structuring element,

- Compute the complement of the image A i.e. A^c by,

$$A^c = 1 - A \tag{3}$$

- Generate an image X_0 of all zeros with the same size as that of image *A*. Select any pixel within the 8 connected boundary and set it to1
- Perform hole filling operation by iterative application of dilation operation on *X* with the structuring element *B* until $X_k = X_{k-1}$ as in (4)

$$X_k = \left(X_{k-1} \oplus B\right) \cap A^c \tag{4}$$

- The image X_k contains all the filled holes and *A* contains the boundaries. Perform union of X_k and *A* to form an image *Y* containing all the filled holes and the boundaries

$$Y = X_k \cup A \tag{5}$$

FEATURE EXTRACTION

In this paper, HOG features are used to describe the traffic signs and they are trained using MLP for the recognition. HOG was first described in 1986. It is extensively used in the area of the object detection, especially in the human detection [13]. HOG uses the local object appearance and shape within an image by the distribution of the intensity gradients and edge directions to describe an object. An algorithm for HOG descriptor is as follows:

- The detected traffic sign image is resized to 24x20 pixels
- Gradient Computation: Apply a centred one dimensional derivative mask is applied on to the image in both horizontal and vertical directions with the kernels shown in (6)

$$D_x = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \; and \; D_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \tag{6}$$

For any image *I*, the gradient along the *x* and *y* directions is given by $I_x = I * D_x$ and $I_y = I * D_y$. Then the magnitude (*G*) and the angle (Θ) are calculated from I_x and I_y as in (7)

$$G = \sqrt{I_x + I_y}\,, \qquad \Theta = \arctan\frac{I_y}{I_x} \tag{7}$$

Orientation binning: In this step, for each cell the histograms are computed. The gradient directions are quantized into 9 bins. Based on the values in the gradient computation, each pixel in the cell cast a vote for an orientation based histogram channel. For the unsigned gradients the histogram channels ranges from 0 to 180 and for the signed gradients it ranges from 0 to 360.

- Descriptor blocks: Combine a group of 4 cells to form a block. Now, the length of the histogram will be 9x4=36. Since the blocks are formed without any overlapping, total number of blocks in the image will be 5x6=30 blocks. Hence the length of the feature vector becomes 5x6x4x9=1080.

2) RECOGNITION USING MLP

The recognition is carried out by training the HOG features with multilayer perceptron. A multilayer perceptron is a fully connected feed forward neural network that maps a set of input data onto a set of appropriate outputs. A multilayer perceptron architecture with 2 hidden layers is shown in Figure 4. It consists of three or more layers: input layer, hidden layers, output layer. The number of hidden layers depends on the complexity of the problem under consideration. Each node in one layer is connected to the node in the other layer with certain weight. In the multilayer perceptron, each node is a neuron with nonlinear activation function except for the input layer. The activation function used in this neural network architecture is a sigmoid function as given in (8).

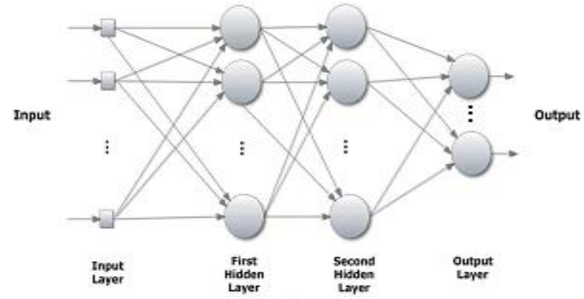

Figure 4. Multilayer perceptron architecture

$$f(x) = \beta * \frac{1 - e^{-\alpha x}}{1 + e^{-\alpha x}} \tag{8}$$

The learning is carried out by the backpropagation algorithm which learns by updating the weights after each data is processed by comparing the amount of error in the output with the expected result. In this work, triangular and circular signs are trained separately to reduce the computational complexity. An input layer with 1080 (length of the feature vector) nodes and hidden layers with 500 and 100 nodes are used for the recognition.

IV. EXPERIMENTAL RESULTS

The proposed traffic sign detection and recognition system is tested on a video captured from a vehicle moving at a speed of 30 to 50 KMPH using SAMSUNG GALAXY J5 Smartphone. The videos are collected in the rural and urban areas from 6 AM to 6 PM to capture all the variations of illuminations. In this work, the database is divided into 3 groups based on the time of the day: Morning, Afternoon and Evening. The detection rate for these three databases using both color and shape features are shown in TABLE II and the detection rate is calculated from (9). From TABLE I we can conclude that, the videos taken in the afternoon and evening provides better detection rate than the videos taken in early morning.

$$Detection\ Rate = \frac{No.\ of\ correctly\ detected\ signs}{No.\ of\ signs\ in\ the\ video} \times 100 \quad (9)$$

For the traffic sign recognition stage, 21 classes are trained using the multilayer perceptron. Out of 21 classes, 7 classes are circular signs and 13 classes are triangular signs and one miss detected class. 70 images of each class are considered and hence a total of 1470 images are trained using MLP. Then, 10 images of each class are used to test the network. A recognition accuracy achieved for both circular and triangular signs is given in the TABLE II. The recognition rate is calculated from (10). The algorithm is implemented on a computer with Intel Xenon CPU 3.20GHz processor and achieves an average frame rate of about 15-26 frames/sec. Figure 5 shows the experimental results under different scenarios.

$$Recognition\ Rate = \frac{No.\ of\ correctly\ recognized\ signs}{No.\ of\ signs\ in\ test\ dataset} \times 100 \quad (10)$$

TABLE I. DETECTION RATES FOR THREE SETS OF DATA

Database	Total no. of signs	Correctly detected signs	Detection rate
Morning	252	209	82.94%
Afternoon	210	189	90%
Evening	189	169	89.41%

TABLE II. RECOGNITION ACCURACY USING MLP

Traffic sign	Total no. of signs	Correct classification	Wrong classification	Recognition accuracy
Circular	70	68	2	97.14%
Triangular	140	134	6	95.71%

a) A video frame from morning database

b) A video frame from afternoon database

c) A video frame from evening database

d) Shadow generated by other objects

e) Sign suffered from rotation

f) Sign with geometrical distortion

Figure 5.Traffic sign recognition results under different scenarios

CONCLUSION

In this paper, an algorithm for automatic sign detection and recognition is proposed. The method uses both color and shape features to detect the traffic signs and histogram of oriented features extracted from the detected signs are used to train the multilayer perceptron using backpropagation learning algorithm. Finally, an image and voice alert is given to alert the driver about the traffic signs. The experimental results show that, the proposed traffic sign detection and recognition system can address the real-time issues like illumination variations, scale variations, vibration of the vehicle, detection of signs in the shadow environment and signs with some geometrical distortions effectively.

REFERENCES

[1] Debu, C. Road Accidents in India. Retrieved from http://www.mapsofindia.com/myindia/government/roadaccidents-in-india.

[2] F. Qin, B. Fang and H. Zhao, "Traffic Sign Segmentation and Recognition in Scene Images," 2010 Chinese Conference on Pattern Recognition (CCPR), Chongqing, pp. 1-5, 2010.

[3] Y. Y. Nguwi and A. Z. Kouzani, "Automatic Road Sign Recognition Using Neural Networks," The 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, BC, pp. 3955-3962, 2006.

[4] D. M. Gavrila, "Traffic Sign Recognition Revisited," in musterekennung 1999, Springer Berlin Heidelberg, pp. 86-93, 1999.

[5] M. A. Garcia-Garrido, M. A. Sotelo and E. Martin-Gorostiza, "Fast traffic sign detection and recognition under changing lighting conditions," 2006 IEEE Intelligent Transportation Systems Conference, Toronto, Ont., pp. 811-816, 2006.

[6] G. Loy and N. Barnes, "Fast shape-based road sign detection for a driver assistance system," 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566), pp. 70-75 vol.1, 2004.

[7] M. Kobayashi, M. Baba, K. Ohtani and L. Li, "A method for traffic sign detection and recognition based on genetic algorithm," 2015 IEEE/SICE International Symposium on System Integration (SII), Nagoya, pp. 455-460, 2015.

[8] Z. Malik and I. Siddiqi, "Detection and Recognition of Traffic Signs from Road Scene Images," 2014 12th International Conference on Frontiers of Information Technology, Islamabad, pp. 330-335, 2014.

[9] Rongqiang Qian, Bailing Zhang, Yong Yue, Zhao Wang and F. Coenen, "Robust chinese traffic sign detection and recognition with deep convolutional neural network," 2015 11th International Conference on Natural Computation (ICNC), Zhangjiajie, pp. 791-796, 2015.

[10] M. S. Hossain, M. M. Hasan, M. Ameer Ali, M. H. Kabir and A. B. M. Shawkat Ali, "Automatic detection and recognition of traffic signs," 2010 IEEE Conference on Robotics, Automation and Mechatronics, Singapore, pp. 286-291, 2010.

[11] "Miscellaneous Image Transformation- OpenCV 2.4.13.2 Documentation",http://docs.opencv.org/2.4/modules/imgproc/doc/miscellaneous_transformations.html#cvtcolor.

[12] Z. Zheng, H. Zhang, B. Wang and Z. Gao, "Robust traffic sign recognition and tracking for Advanced Driver Assistance Systems," 2012 15th International IEEE Conference on Intelligent Transport Systems, Anchorage, AK, pp. 704-709, 2012.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, pp. 886-893 vol. 1, 2005.