Traffic Sign Detection using Computer Vision Techniques

Spring 2018, EE 5805: Directed Study

Advisor: Prof Dr. Michael C. Roggeman

Sarala Ravindra

Abstract

With great accomplishments in fields like autonomous driving, self-driving technologies the use of image data to detect and recognize traffic signs has been a topic of great interest for the computer vision community.The traffic sign detection along with traffic sign recognition is a part of driver assistance systems which require machines to detect road signs to understand them.High performing traffic sign detection and recognition systems are considered beneficial in assisting human driving especially in conditions of bad weather, sun glare, human negligence. I intend to study the various approaches in detecting traffic signs in real-time and report my understanding of it.

1 Introduction

Road traffic signs are informative and/or instructional signs put to use alongside the roads to guide the road users - cyclists, motorists, pedestrian.There are diverse conventions used in the deployment of traffic sign boards.The region, its history, convention treaty etcetera play roles in the type of traffic signs used.For instance the most of the European countries follow the famous 1968 Vienna Convention on Road Signs and Signals again in different flavoring in spite of the uniformity and the standardization, United States have route based road signs again not so uniform throughout the region, most Asian countries seems to be little influenced by the European approach but their signs regulations are different from one another. Due to much uniformity and availability of datasets the European traffic signs are used in my study. In the broader sense, the traffic signs can be categorized based on the type of information they convey. A sign could fall under one of these groups:Priority- signs of high importance such as Stop and yield signs, Regulatory, signs conveying regulations such as speed limits, Obligatory- signs the road users are obliged to follow for smooth traffic and less social disturbance such right only and school zone, Warning- signs which indicate traffic warnings such as road construction,slippery roads, Informative- other information such as route, dead end. Figure 1 depicts this categorization.

The techniques used in traffic sign identification aims to segment the region of traffic sign either by using template-based methods, feature-based methods, color information or some combination of these.It is important to choose an algorithm which is computationally less expensive but provides expected results to be used real-time traffic sign detection systems. Usually, the detection process is followed by the recognition algorithm or sometimes it is done in a single algorithm as in some of the neural networks based approaches.Histogram of Oriented Gradients (HOG) features, Haar

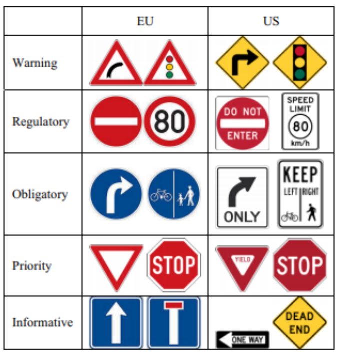

Figure 1: Different signs in EU and the US region [2]

cascades, Gaussian mixture models, Convolutional Neural Networks(CNN) are some of the famous approaches employed in traffic sign detection and recognition systems.

2 Detection methods

This section details the approaches I studied for the detection of traffic signs in real-time traffic images.

2.1 Color based detection

Algorithms which employ color based detection use the fact that traffic signs usually make use of distinctive colors for easily being spot by the road users.Though most of the image acquisition devices by default acquire images in the RGB format, RGB is not widely used in color based segmentation due to implicit distribution of the intensity amongst the three channels.Figure 2 shows a good example of RGB segmentation results but conditions such as varying light conditions, shadows etc degrade the output based on RGB color thresholding which is shown in Figure 3.Instead color-spaces with explicit intensity channel such as YCrCb, HSV are used for better color-based segmentation results.The comparison in Figure 3 makes this evident.HSV is more suitable to define colors than the RGB color model.Other approach is to use ratios of R,G,B values and comparing with the predetermined constants obtained by inspecting a number of traffic signs in images as the [5].The author works in RGB to reduce the computational load due to RGB to HSV color conversion which is mostly used.

Apart from these a very interesting approach I came to implement was based on the modification of HLS(Hue Lightness Saturation) called the Improved HLS (IHLS) color segmentation approach.This model provides independence between chromatic and achromatic components of the image.Steps describing this algorithm [1] are as follows

1. Convert RGB image to IHLS image using the equations

$$\begin{aligned} H &= \theta \qquad\qquad if \quad B \leq G \\ &= 360 - \theta \qquad otherwise \end{aligned}$$

$$\theta = arccos\left(\frac{R - (G + B)/2}{(R^2 + G^2 + B^2 - R * G - R * B - G * B)^{\frac{1}{2}}}\right)$$

$$S = max(R, G, B) - min(R, G, B)$$
$$L = 0.212R + 0.715G + 0.072B$$

2. Calculate the global mean,Mean of the luminance image

$$Mean = \sum \frac{All \quad the \quad luminance \quad image \quad pixel \quad intensities}{m * n}$$

   where (m,n) = Size of the luminance image

3. Obtain normalized mean Nmean by equation

$$NMean = Mean/256$$

4. Calculate required Euclidean distance threshold th, given by

$$th = exp(-NMean)$$

5. Specify the HLS values of the reference color as in step 1

6. Calculate Euclidean distance between the reference HLS vector and the individual pixel HLS vectors

$$d = \left((S_2cos(H_2) - S_1cos(H_1))^2 + (S_2sin(H_2) - S_1sin(H_1)^2)\right)^{\frac{1}{2}}$$

   where subscripts denote the vector set.

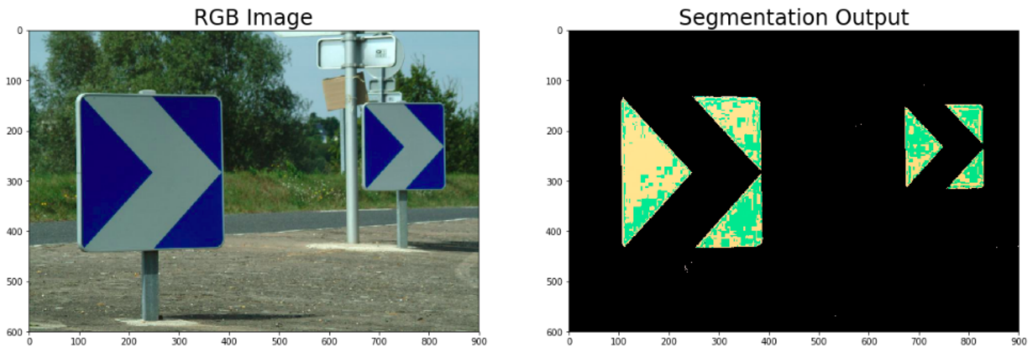

Figure 2: Segmentation results with RGB color thresholding.

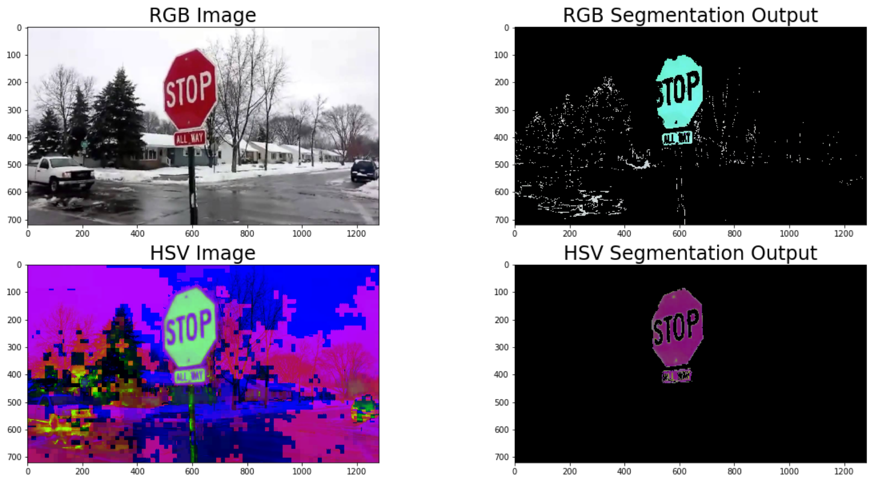

Figure 3: RGB and HSV Segmentation Result Comparison.

7. Any pixel whose Euclidean distance is lesser than or equal to the threshold defined in step 4 is considered to be the object pixel while the rest of the image pixels are considered to be non-object pixels

This approach is more robust to lighting condition and background variation than other color based segmentation methods.The results of the same is shown in Figure 4.Though colors are important feature in segmenting sign region from the images other objects in the image with similar color or similar colored image background are some of the common instances where even a robust color based segmentation algorithm fails to detect the expected region of interest.It simply provides multiple images patches or major portion of the image as the region of interest making the segmentation based detection step less useful.

. . .

2.2 Shape based detection

Similar to color based detection, the shape based detections use the characteristic shape of the sign boards to detect their presence in the image scene.The regular shape based detections follow a simple approach of looking for objects of some shape.Thus edge detection methods such as canny, are employed as the initial steps.Then the template of the sign boards are run across the edge detection output image to check if any similar shaped objects are present.This method must employ multi-scale templates (such as a template pyramid) to be able to detect signs of different sizes.The majority of the traffic sign boards are either circular, triangular, rectangular or octagonal in shape.Inability to incorporate different shapes, multiple non-sign objects with similar

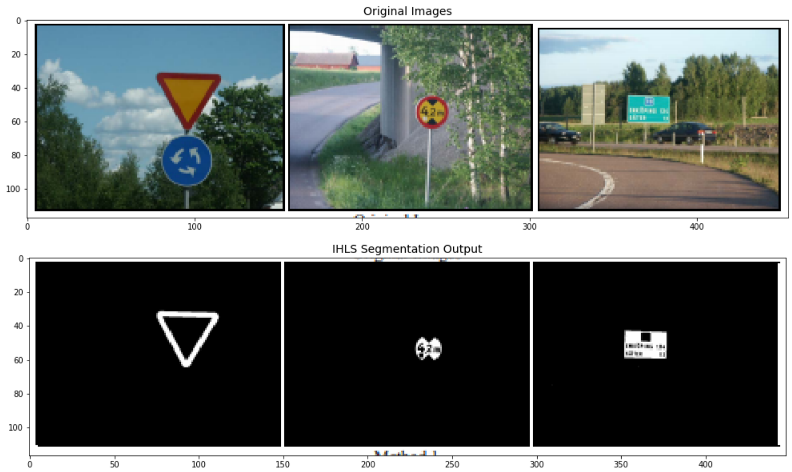

Figure 4: IHLS Color Segmentation Results.

appearance, poor edge detection due to noise etcetera makes the shape detection method alone too risky in applications such as traffic sign detection in real-time.Thus the shape based methods are usually used as an after step once the other detection steps are employed either to assure the detection of sign board or identify the sign board detected by different method in the previous step.One example of such implementation is summarized below [2].Note this algorithm only looks for signs which the author considers to be of more significance -Warning, Regulatory, Obligatory, Priority.The informative sign boards such as the ones giving direction information etcetera are ignored.Only Red, Blue, yellow signs are considered.

1. Convert image from RGB to HSV color-space
2. Apply HSV color range method to obtain region of interest based on color (Define separate color ranges for separate colors)
3. Obtain individual masks of each of the required reference colors
4. Make a gray scale copy of the masked input image
5. Obtain Canny edge image
6. For each of the masked image obtain corresponding edge images and perform sign board shape recognition on each of them
    - Detect circles applying Fast Radial Symmetry (FRS) method.
    - Detect triangles and rectangles by means of Harris corner method.
    - Differentiate a triangle and a rectangle by the relative location of the corner points.
7. Use the color information and shape information to put the possible sign board detected into a class or more safely use a traffic sign recognition algorithm.

The FRS method outputs an image which peaks at the center of the circle while the Harris corner points the corners of the 2-D shapes captured on the image.The major disadvantage of shape based techniques is that the object of interest might go unnoticed when the image is captured in angles /orientations where the object shape is not recognizable.So suitable image orientation correction methods must be employed in such cases.For example when an ellipse is detected in the region of interest (determined by color segmentation) the axes length and their positions could be used to correct the image projection.The FRS output to detect circle is shown in Figure 5 .

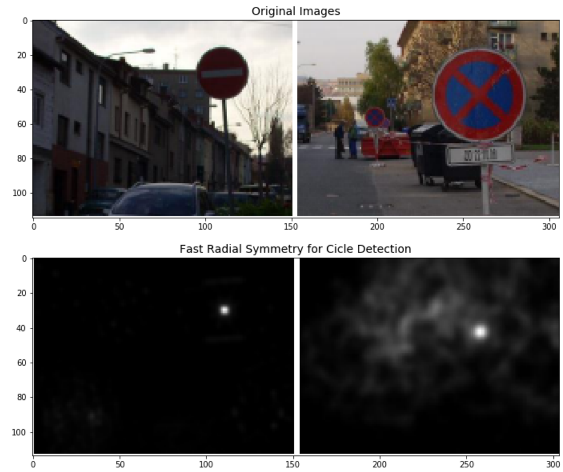

Figure 5: Shape based detection by FRS method for circular sign.

2.3 Other feature based detection

Other than color and shape many other features such as SURF (Speeded-Up Robust Features) descriptors, HOG descriptors, BRISK, ORB (Oriented FAST and Rotated BRIEF) etcetera are used in the detection of traffic sign board.Descriptors are found on the image region in the region of interest and is usually classified using a machine learning technique ,brute force method of matching, probability projections.

Though RGB ratio based color thresholding, IHLS methods prove to perform satisfactorily in many environments they only rely on single feature (color).Thus for robust detection methods techniques which incorporate every useful feature has to be applied.The Convolutional Neural Network(CNN) is an example of such a technique.In the general, neural network requires descent amount of training data to perform well in detection/recognition of traffic sign boards. The CNNs such as Fast Region based CNN (RCNN) outperform the traditional sign detection methods.High testing accuracy of RCNN makes it a sought out method in object detection.The only major drawback is that this method is a region based proposal method meaning the number of region as proposed as region of interest for the recognition algorithm to work with.Though it is better in terms of speed compared to the sliding window approach, this selective search approach still makes room for unnecessary computation. This is overcome by a technique called You Only Look Once(YOLO).The architecture of the single CNN used for both detection and classification is shown in Figure 6 used on PASCAL VOC.It is a 24 layer CNN followed by 2 fully connected layer. Uses 1*1 reduction layer and 3*3 Convolutional filters for reduction. The image be divided into S*S grid and each of the grid produce B bounding boxes (thickness of the box signifying object presence) after passed onto the Neural network.Let S = 7 and B = 2 and number of object class be 20.

Object detection with YOLO

1. Divide image into 7*7 grids

2. For each of the grid cell produce B = 2, bounding boxes and their confidence scores using the CNN model(Training explained).

   - $Confidence = P(object) * IOU(truth, prediction)$
   - Each bounding box produced has 5 parameters attached x,y,w,h (x,y-locate the center of bounding box relative to the grid cell it was produced by,w,h - width and height of the bounding box relative to the image center) and the confidence score of the box (P(Object) for that box)
   - The output is of $length = S \times S \times (B * 5 + C) = 7 * 7 * (2 * 5 + 20) = 1470$

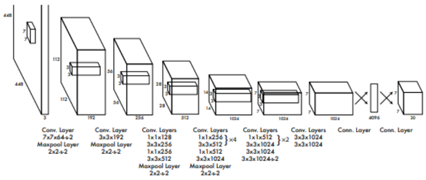

Figure 6: YOLO Architecture given in the source paper[3].

3. Use Non-Maximum Suppression to remove the redundant boxes.This is done by taking the intersection of one box with the other and retain only the box with higher confidence if the intersection of the box areas is greater than a set threshold say 0.5,else both of the boxes are retained and the process is continued till all of the boxes are checked for overlapping box areas and confidence levels.The result is only boxes with object presence probability above threshold without redundant boxes around.

4. Find class-specific confidence scores for each box given by

$$P(class_i/object) * P(object) = P(class_i) * IOU(truth, prediction)$$

   where $P(class_i/object)$ is the ith class confidence of a grid cell given the object is present in it and IOU(truth,prediction) is the Intersection Over Union of truth and prediction score.If no object in the cell the confidence of any class of object must be 0 otherwise it must be equal to the IOU(truth,predicted).

5. Output the boxes with corresponding highest class confidence score.

Training CNN model for YOLO object detection

- The CNN model is trained on ImageNet 1000-class competition dataset [4] .The training network resolution is increased from 224*224 to 448*448 to account for fine resolution required for object detection purpose.

- For the forward pass all of the layers except the last layer employ leaky ReLU(Rectified Linear Unit) activation function given by

$$\begin{aligned} \phi(x) &= x, x > 0 \\ &= 0.1x, otherwise \end{aligned}$$

  The last layer is a linear activation function.The weights of the model are randomly initialized.

- The error function is calculated by the sum of squared differences of the bounding box location in the train set actual output and the predicted output and the confidence score predictions and the actual confidence scores

The results of implementation of YOLO for traffic sign detection using a pre-trained model is shown in figure 7.The YOLO implementation is faster compared to both traditional and other neural network based methods due to its interaction with the image is only once.But its training period is long and is dependent on the model parameters, the threshold etcetera.Another disadvantage is that it requires better resolution images and/or smaller grid sizes to not miss detection of near-by objects, otherwise there are chances of them ending up in same grid and thereby having higher chances objects' presence being missed. .

.

3 Conclusion

Traffic sign detection can be tackled in number of ways.The YOLO detection method in spite of longer training period seems to be a better model due to high accuracy rate, detection and localization at the same step and high speed in the testing. According to [3] the YOLO method is 100 times and 1000 times faster than Fast-RCNN and RCNN models respectively.

Figure 7: Traffic sign detection using YOLO apprach[3].

References

[1] Hasan Fleyeh. Color detection and segmentation for road and traffic signs. In *Cybernetics and Intelligent Systems, 2004 IEEE Conference on*, volume 2, pages 809–814. IEEE, 2004.

[2] Karel Horak, Pavel Cip, and Daniel Davidek. Automatic traffic sign detection and recognition using colour segmentation and shape identification. In *MATEC Web of Conferences*, volume 68, page 17002. EDP Sciences, 2016.

[3] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[4] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[5] Michael Shneier. Road sign detection and recognition. In *Unmanned Systems Technology VIII*, volume 6230, page 623016. International Society for Optics and Photonics, 2006.