Smart Driver Assistant

Abha Tewari, Sahil Khan, Aditya Krishnan, Tanmay Rauth, Jyoti Singh
Dept. of Computer Engineering, VESIT,
Chembur, Mumbai -400 074.
Email: skhan28297@gmail.com

***Abstract* - This paper proposes a smart driver assistant system that will help drivers avoid accidents during lane departures by providing prompt and quick markings of road lanes. This paper also proposes a novel system for the automatic detection and recognition of traffic signs. It detects the blobs using MSER i.e Maximally Stable Extremal Regions which provides similar results under different lighting conditions. Recognition is based on a cascade of Convolutional Neural Networks (CNN) that were trained using histogram of oriented gradient (HOG) features.**

Keywords - Maximally Stable Extremal Region, Convoluted Neural Networks, Histogram of Oriented Gradient

I. Introduction

With the advent of smartphones, Android has become the dominating mobile OS functioning on over 1.2 million devices worldwide[1]. Android provides an efficient SDK which when used with Android Studio IDE can help create applications quickly and easily. World Health Organisation estimated about 1.35 million death all around the globe due to road traffic which can be approximated to about 1 death in 25 seconds in the year 2016. A majority of these accidents occur due to lack of attentiveness while departing lanes / lane splitting. The purpose of this paper is to create a smart driver assistant which will help the driver make rational decisions based on the real-time environment. The system marks lanes in front of the car with image processed highlighting and colored tracking. Along with this, the system will ensure that the driver never misses a traffic sign. The system constantly grabs each and every traffic symbol along the path and makes the information available to the driver through voice assistant. The following sections provides sufficient background and insight into our objective to develop the Smart Driver Assistant

II. Related Work

To get rid of the perspective of the image , one can observe that the lanes appear to converge at the horizon line and these lanes are now vertical and parallel.The paper Real time Detection of Pavement Markers in Streets[2] gives an insight to find the pavement markers. The main assumption here is that the lanes are parallel or almost parallel to the camera. This focuses the attention on only a subregion of the input image reducing the run time considerably.

In Detection and Recognition of Road Traffic Signs[3] the Detection is performed using a novel application of Maximally Stable Extremal Regions (MSERs) and Recognition is performed with histogram of oriented gradient (HOG) features, which are classified using a linear support vector machine (SVM).

The paper Traffic Road Sign Detection and Recognition for Automotive Vehicles[4] gives an intuition on the algorithms that are used to detect and recognise the traffic signs. According to this papers the original image is resized and exposed to a red colour threshold. The object is detected using Breadth-First Search (BFS) and the possible signboard region is cut. A sobel operator for precise edge detection is applied and then the sign in the original RGB image is extracted to real signboard.

The paper Lane Detection Techniques[5] describes various limitations in the existing lane detection models and gives a basic understanding of the solutions that can be used to eliminate these limitations .The process of lane detection starts by taking the image of the road with the help of a camera. The coloured image is converted into its grayscale image to minimise the objective of processing the image. Then the basic filtering techniques are used to remove the noise presented in the grayscale image as it can affect the efficiency of detection. Then these filtered images are sent to the canny edge detectors which uses automatic thresholding techniques for edge detection. After the process of edge detection this intermediate processed image is fed to the line detectors which provides us the right and left boundary segments[6] the image is then superimposed over the original image to get final image.

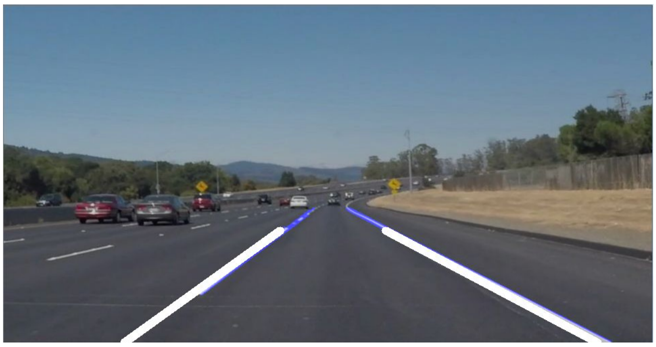

Fig 1. Prototype of lane detection system

The paper Lane detection and tracking using B-snake[7] gives solution to the problem of identifying the lane when the pavement markers are not clear or absent. In real world there are generally two methods taken into consideration for lane detection , they are feature-based technique and model based technique. The feature-based technique localizes the lanes in the road images by combining the low-level features, such as painted lines or lane edges, etc.

lane segments that are detected by traditional image segmentation. Accordingly, this technique requires the studied road having well-painted lines or strong lane edges, otherwise it will fail[8]. On the other hand, the model-based technique just uses a few parameters to represent the lanes. Assuming that lane shape can be represented by line or a curved line like hyperbole , the process can be done by taking into considerations these model parameters. Using these techniques, the system gives satisfactory results even in varying conditions of noises , shadows and illumination that are there in the images that are captured by the camera.This system is applicable to both marked and unmarked lane.

III. Smart Driver Assistant System

The system actively tracks the road lane and identifies traffic symbols. For this, the in-built camera of the smartphone will be used. This camera will constantly capture a video sequence at a suitable frame rate which will then be used to identify and mark lanes and traffic signs.

Beginning with lane detection and tracking, the general method of lane detection is to first take an image of road with the help of a camera in the vehicle. Conversion of image to grayscale image is important to reduce the processing time. Noise present in the image affects the accuracy of the system hence it becomes important to eliminate these noise components from the image. For this purpose lateral filters like bilateral filters, trilateral filters and gabor filters are used. To obtain the edges in an image using automatic thresholding, the edge detector makes use of a canny filter. To obtain the boundaries on both side of the image, the resultant image from the above step is further segmented using the line detectors. The lane boundary scan mainly consists of Hough transform. It uses the information of the edge detected image returned by the Hough transform to perform the scan. The scan results in a an array of points towards the left and right side of the line. The final step in this process is to properly fit a curved/straight line according to the need to get the resultant image.

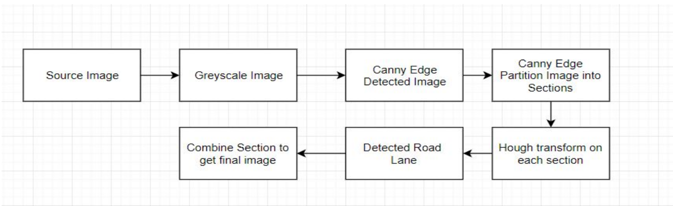

Figure 2 : Lane detection basic flow diagram

The sign which is placed at the side of roads to impart information to road users is known as road signs or traffic signs. There are four types of traffic signs that are shown in the traffic code: a)warning b)prohibition c)obligation and d)informative[9]. These traffic signs are given different geometrical shapes based on features like form and colour. The warning signs are triangular in shape with the vertex at the top. These signs consists of a white or blue background and are bordered in red. If the signs are located in an area where there's public work taking place, then the signs are given a yellow background. If the traffic sign has to portray a sense of obligation or information, then the signs are given a circular shape with blue background[10]. In order to detect the location of the traffic sign in an image, two main properties are to be considered, the colour of the traffic sign as well as the shape of the traffic sign.

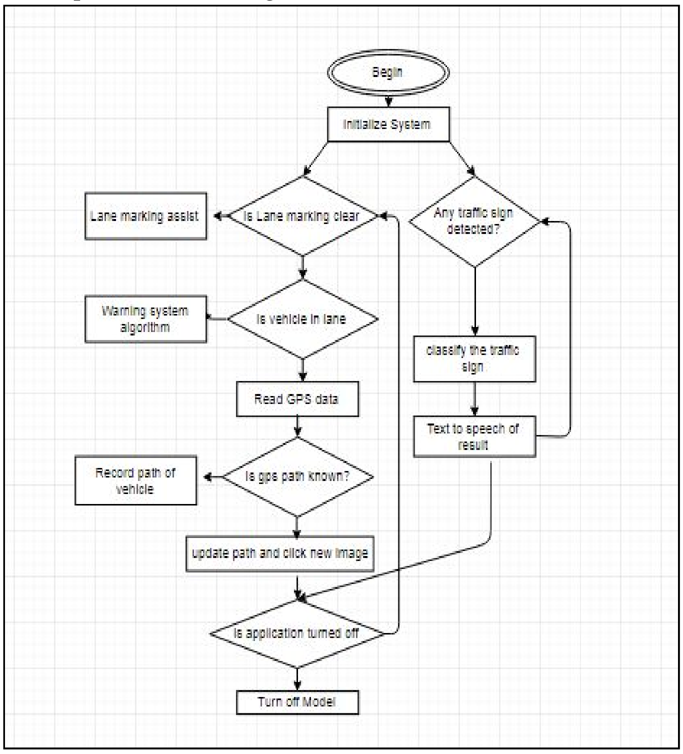

Figure 3 : Combined flow diagram of Lane detection and warning system

White background traffic signs can be detected with the help of MSERs of the grayscaled image. The video is nothing but a series of frame. Each frame of the video are binarized using a variety of threshold levels. Using this, the connected components at each thresholded level are found. MSERs are selected for the components that are connected and retain their shape at different threshold levels. This property of the MSERs ensures stability of the image irrespective of the lighting and contrast conditions which in turn increases the efficiency of the system[11]. Some other advantages of using MSERs include reducing the number of candidate regions with the help of connected components and several other features.

Figure 4 : Traffic sign detection

The features mentioned above include the geometrical height and width of the sign, aspect ratio of the image, region perimeter, area as well perimeter and area using bounding box method. The process of eliminating the unnecessary connected components i.e the components that does not fit the criteria facilitates in enhancing the speed of the process and also its accuracy. The following Table 1 depicts the parameters that are used as restrictions for the features that were determined using empirical methods[12].

Feature	**Minimum**	**Maximum**
Width	14	100
Height	14	110
MSER/bounding box perimeter	0.3	1.2
MSER/bounding box area	0.4	1

Table 1 : Properties used to sort connected components

The detection of the red and blue backgrounded traffic signs are performed in a contrasting manner as compared to the traffic signs having white backgrounds. Instead of finding out the MSERs of the image that is grayscaled, the frame is converted into normalised red/blue image from the original RGB image. The normalised red/blue of an image is calculated such that for each pixel in the original image, the value of the pixel is replaced with the maximum of the ratio of the blue channel to the sum of all the channels and the ratio of the red channel to the sum of all the channels. This ensures that the value of the pixel replacing the original pixel has higher values for red and blue components as compared to other coloured components[13]. For this normalised red/blue image , the MSERs are calculated in the same way as the traffic sign with white background.

The next step in the process is the recognition stage. The recognition stage uses the candidate region found above as a traffic sign and helps to allocate the exact type of sign to it. HOG features are used for proper characterisation of the candidate region. The HOG feature helps to represent the gradient orientation of the image. This process is repeated for every detected candidate region. For finding out the magnitude and orientation of each of the pixel, a sobel filter is used which operates by calculating the horizontal and vertical derivatives in an image[14]. The HOG application for recognition of traffic symbols are extremely convenient as the traffic symbols consists of well founded geometric shapes and varying contrast edges that encircles a spectrum of orientations. In this process, there is no need for rotational invariance since the traffic signs are positioned in a location where they are fairly visible and are mostly upright and facing the camera which limits rotational and geometrical distortions. A cascade of multilayer Convolutional Neural Networks are then used for the classification of the traffic sign.

Taking the inspiration of the biological nervous system such as the human brain , an information processing paradigm was developed which is known as the Artificial Neural Network (ANN). The unique structure of the information processing system is the primary element of this paradigm. The difference between the normal neural networks and the convolutional neural network is that the neurons in the CNNs have learnable biases and weights. The function of each neuron is to receive some input at the beginning, then calculate a dot product and at one's discretion perform it with nonlinearity. Taking raw input image at one end of the network and to classify the scores at the other end is the single differentiable function of the whole network. Despite this, CNN still has a loss function on the last layer.

Three convolutional layers will be used for feature extraction and one fully connected layer as a classifier. Multiscale features will be used in contrast to the general feed forward CNNs which will result the output to be forwarded to subsequent layers and also the output will be branched off and it will be fed to the classifier. All the output layers that are branched off has to undergo max pooling so that the convolutions are subsampled proportionally.

IV. Tools and Technologies

The smart driver assistant being a combination of two components i.e. lane tracking and traffic symbol analysis, we would be needing two major technologies viz Image Processing and Machine Learning. Following libraries and frameworks provide a comprehensive aid in the production of the system.

- **OpenCV** (Open Source Computer Vision) is a library of programming functions mainly aimed at real-time computer vision.
- **TensorFlow** layers module provides a high-level API that makes it easy to construct a neural network. It provides methods that facilitate the creation of dense (fully connected) layers and convolutional layers, adding activation functions, and applying dropout regularization.

- **Android Studio** provides the fastest tools for building apps on every type of Android device. We will make use of android studio to make our android application of smart driver assistant.
- **Google Text to Speech** is an API in Android operating system. It powers applications to read aloud (speak) the text on the screen which support many languages. This will help to convey results of traffic sign classification to driver.

V. Datasets

German Traffic Dataset is a set of 51,839 labeled images of 43 different German traffic signs. It comes in two separate sets. A set of 39,209 images for training and another set of 12,630 images in order to test the accuracy of our trained network. Each image is a 32×32×3 array of pixel intensities, represented as [0, 255] integer values in RGB color space. Class of each image is encoded as an integer in a 0 to 42 range. Dataset is very unbalanced, and some classes are represented way better than the others. The images also differ significantly in terms of contrast and brightness, hence some kind of histogram equalization is necessary to be applied , this should noticeably improve feature extraction. For the classification process, the german traffic sign dataset has been chosen as it is the most comprehensive dataset available and has a 90% similarity to Indian road symbols. Before the main implementation takes place, the image is preprocessed which usually involves changing the scale of the pixel to [0,1]. The image is then characterised using labels where they are hot encoded and shuffled. The general preprocessing also involves a step of histogram equalisation to increase the overall contrast of the image and it proves to be useful in most of the cases. It is necessary to work with grayscale images instead of colour ones as using color images doesn't improve the results. The amount of data mentioned above is not sufficient for a model to generalise well. It is also fairly unbalanced, and some classes were represented to significantly lower extent than the others. So to extend the dataset, a variety of flips, rotations and inverts on the available images needs to be applied to extend the counterparts of the classes. This simple trick extends the dataset from almost 40,000 to nearly 65,000.

As a result of the above process the dropout drastically improved the generalization of the model. The dropout was applied only to fully connected layers, as shared weights in convolutional layers are good regularizers themselves. Lambda=0.0001 seemed to perform best in L2 Regularization. Early stopping with a patience of 100 epochs can be used to capture the last best-performing weights and roll back when model starts overfitting training data. Validation set cross entropy loss is used as an early stopping metric, intuition behind using it instead of accuracy is that it generalizes better.

VI. Discussion

The smart driver assistant system provides useful tips to the user. The lane detection and tracker will guide the driver to stay in his own designated path. This will prevent rash lane splittings as well as wandering off from the lane due to inattentive driving or drowsiness. The system will warn the driver of inattentive swaying away from the lane. Also, if the the car in front of the driver gets too close too fast then the system will also present a warning signal. In addition, the system will consist of a fail safe that will still detect lane even in the absence of white lane markers which will be based on the B-snake algorithm.

Excellent results are obtained in clear conditions, however some false positives occur due to stop lines at cross streets, at crosswalks, near passing cars. False positives are mostly found when driving on the right lane of the street with no right lane boundary, and we detect the curb as the right lane boundary , and that's the reason for the high false positive rate. However, this is not a real problem as the curb is really a lane boundary, but not painted as such and this won't affect the objectives of the algorithm to detect lane boundaries. In the current algorithm, the work is done only on the red channel which gives us better images for white and yellow lanes than converting it to grayscale. However, there is plenty to be done to further improve this algorithm. The plan is to use \the color information to classify different lane boundaries, white solid lines, double yellow lines, etc. This will also allow to remove the false positives due to curbs being detected as lanes, as well as confusing writings on the streets, which are usually in yellow and can be filtered out.

The traffic sign analysis information will be conveyed to the driver with the help of voice assistant and also it will be displayed on the screen of the mobile device. The driver can use voice commands to gain additional information about the traffic sign. Also, the sign analysis will be integrated with the environment of the car such as the speed. Thus, if the speed limit is detected as 60 kmph on the traffic symbol then the system will alert the driver if the car is moving at a speed much greater than the speed threshold.

VII. Conclusion

The lane detection techniques play a significant role in intelligent transport systems. In this paper various lane detection methods have been studied. However, further improvements can be done to enhance the results. In the near future, one can modify the existing Hough Transformation so that it can measure both the curved and straight roads.

In this paper, an automatic traffic sign detection and recognition system is visualized that works in real time. The region where the traffic signs are located are identified as candidate region and is calculated through the process of MSERs. MSERs are chosen as the candidate regions because these regions can be calculated irrespective of the sensitivity of the lighting conditions and the illumination in the image.

Recognition of the traffic signs are based on HOG feature which are classified using a cascade of CNNs and SVMs. The synthetic datasets are generated so as to increase the number of images in the overall datasets which will increase the accuracy of the system and the model will be trained better. This system can identify signs from the whole range of ideographic traffic symbols currently in use in Germany, which form the basis of our training data. The system retains a high accuracy at a variety of vehicle speeds.

References

[1] Car accidents. http://en.wikipedia.org/wiki/car_accident.

[2] M. Bertozzi and A. Broggi. Real-time lane and obstacle detection on the gold system. IEEE page number from 213 to 218, 19th and 20th September, 1996.

[3] M. Bertozzi, A. Broggi, G. Conte, and A. Fascioli. Obstacle and lane detection on argo. In Intelligent Transportation System, IEEE Conference on, pages 1010–1015, 9-12 Nov. 1997.

[4] David A. Forsyth and Jean Ponce. Computer Vision: A modern approach. Prentice Hall, 2002.

[5] U. Franke, D. Gavrila, S. Gorzig, F. Lindner, F. Puetzold, and C. Wohler.IEEE, 13(6):40–48, Nov.-Dec. 1998.

[6] U. Franke and I. Kutzbach. Fast stereo based object detection for stop & go traffic. In Intelligent Vehicles Symposium.

[7] P. Paclik and J. Novovicová, "Road sign classification without color information." in Proc. ASIC, Lommel, Belgium, 2000.

[8] S. Vitabile, G. Pollaccia, G. Pilato, and E. Sorbello, "Road signs recognition using a dynamic pixel aggregation technique in the HSV color space." in Proc. IAP, 2001, pp. 572–577.

[9] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," in Proc. IVS, 2005, pp. 255–260.

[10] X. Gao, L. Podladchikova, D. Shaposhnikov, K. Hong, and N. Shevtsova, "Recognition of traffic signs based on their color and shape features extracted using human vision models." J. Vis. Commun. Imag. Represent.,vol. 17, no. 4, pp. 675–685, Aug. 2006.

[11] S. Maldonado-Bascón, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras, "Road-sign detection and recognition based on support vector machines." IEEE Trans. Intell. Transp. Syst., vol. 8, no. 2, pp. 264–278, June 2007.

[12] R. Malik, J. Khurshid, and S. Ahmad, "Road sign detection and recognition using color segmentation, shape analysis and template matching." in Proc. ICMLC, Aug. 2007, vol. 6, pp. 3556–3560.

[13] M. Rahman, F. Mousumi, E. Scavino, A. Hussain, and H. Basri, "Real time road sign recognition system using artificial neural networks for Bengali textual information box." in Proc. ITSim, 2008, vol. 2, pp. 1–8.

[14] Y. Wang, M. Shi, and T. Wu, "A method of fast and robust for traffic sign recognition." in Proc. ICIG, 2009, pp. 891–895.