Detection and Recognition of Road Traffic Signs - A Survey

Sumi K. M.

PG Student

Department of Computer Science & Engineering

Federal Institute Of Science And Technology

Angamaly, Ernakulam, India

Arun Kumar M. N., PhD

Associate Professor

Department of Computer Science & Engineering

Federal Institute Of Science And Technology

Angamaly, Ernakulam, India

ABSTRACT

This paper presents a survey on the detection and recognition of traffic signs which has a number of important application areas that include advanced driver assistance systems, road surveying and autonomous vehicles. This has been thoroughly studied for a long time. But still it remains a challenging problem in computer vision due to the different types and the huge variability of informations present in them. This problem can be divided into two stages; first stage will be detection of region that cantains traffic sign candidates and second will be character recognition. For the detection and recognition of text and sign from traffic boards, appropriate methods must be applied to obtain accuracy. This method can be achieved by doing a survey on different methods used to detect and recognize text and signs from traffic boards. In this paper, various techniques for the detection and recognition of traffic signs are explained. A comparitive study is performed on these techniques and their performance is explained.

Keywords

Detection, Recognition, Maximally stable extremal region (MSER), Optical Character Recognition(OCR)

1. INTRODUCTION

Road signs are used to regulate traffic, warn drivers, and provide useful information to help make driving safe and convenient. In order to make clear what kind of informations are being provided by any particular sign, different design styles are used for the signs. These differences are based on shapes and colors. While the design and use of road signs has been established long, they do not provide a perfect solution to make driving a safe one, partly because due to the fact that drivers are human only. For example, most of the accidents have been caused by the failure of a driver to notice a stop sign, either due to a lack of paying attention at a critical moment or due to adverse conditions that impede visibility. The drivers, at night are easily distracted or blinded by the headlights of oncoming vehicles. In bad weather situations(e.g., rain, fog, and snow) traffic signs are less likely than normal to attract a drivers attention. All these situations make driving more difficult and so, can lead to more and more traffic accidents. In most of the tasks that require visual ability, the computer vision systems are now increasingly being used to aid or replace human beings.

Traffic Sign Recognition technology is a process by which a vehicle is able to recognise the traffic signs along the road such as "hump ahead" or "narrow bridge". This is a part of the features collectively called ADAS. These are systems to help the driver in the driving process. They automate/adapt/enhance vehicle systems for safety and better driving. Safety features are efficiently designed to avoid traffic collisions and accidents by offering methods that alert the driver to potential traffic problems, or to avoid traffic collisions by adopting safeguards and taking control over the vehicle. Adaptive features may automate lighting, provide control, automate braking mechanism, incorporate GPS or traffic warnings, connect to various smartphones, alert driver to other vehicles or dangers, keep the driver in the correct lane, or show what features are in blind spots.[source:wikipedia]

Automatic traffic sign detection and recognition enhances safety by informing the driver about speed limits or possible dangers like icy roads, massive road works, or zebra crossings. The traffic signs can be divided into several categories, based on their color and shape, e.g., red-rimmed circular prohibition signs, triangular warning signs, and blue information signs. The simplified pictoric representation make them easily perceivable and comprehensible. Traffic symbols have several characteristic features that may be used for their detection and identification purposes. They are designed in specific colors and shapes, with the text or symbol in high contrast to the background. Because traffic signs are generally oriented upright and facing the camera, the amount of rotational and geometric distortion is limited.

Text contained on road signs carries many useful information required for driving; it describes the traffic situation currently existing, defines the right way, provide warnings about potential risks, and permits or prohibits road access. Automatic text detection on road signs can help a driver to be aware of the traffic situation and surrounding environments by watching and highlighting signs which are ahead and/or have been passed. Some systems can also read out text on road signs with a synthesized voice. Thus the cognitive load on drivers can be reduced by such a multimedia system and can enhance safety while driving. Furthermore, this can be integrated together with some navigation systems like global positioning system (GPS) based systems.

Each government has imposed a sets of rules and regulations to ensure an efficient traffic system. Everyone, specially the driver must

obey these rules and regulations for a secure travel. Some of these laws are represented visually as traffic boards that contain different signs and texts. Different categories of traffic signs are there which we see along the roads. An efficient driver must notice each of the road signs and have to act accordingly. Otherwise dangerous things can happen. There are chances a driver may not notice the road signs in front of his vehicle due to lack of care or human perception errors. Since the percentage of road accidents is increasing day by day, it is desirable to have an automatic road sign detection and recognition system to assist the driver for ensuring a safe travel.[15]

Several methods have been proposed to detect and recognize the traffic signs present on the roads. This paper presents a survey on these various techniques.

2. DETECTION AND RECOGNITION TECHNIQUES

Much researches have been focussed on the detection and recognition of text in natural scenes. Approaches to this kind of problem can be broadly classified into two groups: region based methods and connected component (CC)-based methods. Local features, such as texture are used to locate text regions in region-based text detection. Whereas in CC-based methods, text characters are segmented individually by using information like intensity variation, edges and color contrast. They mainly consist of three phases: the first stage detect CCs within the image, the second stage based on their features, eliminate unlikely CCs and the final stage group the remaining CCs into words or lines.[10]

Maximally Stable Extremal Region (MSER)

The algorithm includes observing the candidates for traffic signs among the outlined scene search regions. MSERs are defined to be regions that approximately maintain their shape regardless several image threshold levels. This region detector is robust to lighting and contrast variations and detects high-contrast regions, that make it suitable for the detection of traffic signs.[10]

Hue, Saturation, Value color space(HSV)

HSV (Hue, Saturation and Value) - specifies a type of color space. It has three components: hue, saturation and value. Hue gives the color. Saturation represents the grey range in the color space. The brightness of the color which varies in accordance with color saturation is represented by the value channel. [source:http://www.tech-faq.com/hsv.html]

Traffic text sign candidates are also detected using HSV thresholding. Each frame is transformed into the HSV color space first, then a threshold is applied to hue and saturation channels both. The value channel is not considered to make the system remain invariant to changes in brightness.[10]

Hue, Saturation, Intensity color space(HSI)

In HSI color model, each color is represented by three components: hue(H), saturation(S), intensity(I). The Hue component specifies the color itself. The Saturation component signals how much the color is polluted with white color. The Intensity range is between [0,1].[source:http://www.blackice.com/colorspaceHSI.htm]

HSI model is suitable for road-sign detection since it is based on human color perception and the color used for road signs is selected to capture human attention. Light (sun or shade) and shadows influence the values of saturation and intensity. On the other hand,

hue is largely invariant to such changes in daylight. For this reason, hue was chosen for the color features in the road-sign detection system.[2]

Support Vector Machines(SVM)

SVM is a supervised learning method which defines a hyperplane to separate data into different classes. The data points which define hyperplane's maximum margin are regarded as "support vectors". Although primarily a binary classifier, by training many one-against-one binary SVMs, multiclass classification can also be achieved by SVM. SVM classification is fast, highly accurate and efficient, and less prone to overfitting compared to many other classification methods. It is also possible to very quickly train an SVM classifier[7].

Optical Character Recognition(OCR)

This is a process of electronically extracting text from images. It involves recognition of printed or written text characters by a computer. Character-by-character photo scanning of the text is followed by the analysis of the scanned-in image. Then character image is translated into character codes, like ASCII, commonly used in data processing. OCR is a technology that enable different types of documents converging, like scanned paper documents, PDF files or image files which are captured by a digital camera into searchable and editable data.[source:wikipedia]

Histogram of Oriented Gradient(HOG)

HOG is a widely used shape descriptor used for traffic sign detection due to its superior performance. In order to extract HOG with color information, the original HOG calculates gradients for each color channel and takes the gradient with the largest norm. In order to extract HOG descriptors, an image is divided into blocks, which are formed by overlapping cells. Each cell consists of non-overlapping pixels and for each cell a local 1-D histogram of edges orientation is derived. Each pixel contributes to the formulation of the histogram by the magnitude and also the orientation of its gradient. Orientation angles are measured into bins and a vote is assigned to the suitable bin to which the orientation value belongs for each pixel and this vote is assigned by the magnitude. The local histograms are accumulated over blocks in order to attain illumination invariance and are then concatenated to create the descriptor.[12]

3. LITERATURE SURVEY

Many approaches have been proposed to deal with traffic sign detection and recognition. Most of these systems typically involve two tasks: finding the locations and sizes of traffic signs in natural scene images (traffic sign detection) and recognizing the detected traffic signs to interpret its meaning (traffic sign recognition). Being designed with regular shapes and conspicuous colors, traffic signs attract human driver's attention so as to be easily captured by human drivers.

However, many difficulties are there to identify the traffic signs by computer algorithms due to color degradation, illumination changes, cluttered background, motion blur, partial occlusion or shadowing, etc. In most of these approaches, MSER and HSV color thresholding techniques are used for the detection of traffic signs and HOG feature extraction method is used for the traffic sign recognition purpose. For the recognition of text based traffic signs, OCR method is used.

Dilip *et al*. [1] analyzed color information contained on the images for the detection of traffic signs. This had the capability of detection and recognition of traffic signs even with bad visual artifacts those originate from different conditions. They used the Speeded up Robust Features (SURF) algorithm - an outstanding feature detector and descriptor which is rotation and illumination invariant for detection purpose. After detecting the shape of the traffic signs, optical character recognition (OCR) method is used to recognize the character present in it. A technique based on Maximally Stable Extremal Regions (MSER) region and canny edge detector was also supervised for character recognition in traffic sign detection process. They provided an adaptive and standalone system which works without any kind of previous training. They also reported an average success rate of 92.18%, the result being independent of the number of traffic signs contained in the image.

Chiung-Yao Fang *et al*. [2] used two neural networks to extract the traffic sign's color and shape features from the input scenes images. For extracting color information they believed that the HSI model is suitable for road-sign detection, since it is based on human color perception and the color used for road signs is selected to capture human attention. Hue was chosen for the color features in the road-sign detection system. The neurons on the input layer of the neural network for shape feature detection acted as an edge detector. Based on the extracted features, traffic signs were then located in the images. This process was primarily conceptualized in terms of fuzzy-set discipline. They assumed that the speed of the vehicle and the physical sizes of road signs were known. Then the traffic signs located in the previous phase were tracked through image sequences using a Kalman filter. Experimental results demonstrated that their method performs well in both detecting and tracking of road signs that were present in complex scenes and in various weather and illumination conditions.

Wu *et al*. [3] found candidate regions, by using a combination of Shi and Tomasi features[13], Gaussian mixture models, and also geometric analysis. Discriminative feature points are detected in every video frame using the Shi and Tomasi algorithm and partitioned them into clusters. Then they selected the candidate road sign regions that correspond to feature point's clusters using a vertical plane criterion. Text on candidate road sign areas are detected and tracked. All detected text lines are extracted for recognition, given a satisfactory resolution in the text extraction and recognition step. The recognition was done by integrating an optical character recognition (OCR) system. The authors reported a detection rate of 88.9% and a false detection rate of 9.2%, based on a data set of 22 video sequences, each around 30 s long.

Reina *et al*. [4] detected every potential panel on the input captured image. They segmented regions of interest based on color information, by applying a threshold to the chrominance and luminance channels in the Lab color space. The candidate blobs are extracted from the input image by using HSI color space thresholding and computed shape classification. Each blob's shape was classified by applying FFT to the signature of blobs. They found the rectangular regions by making a comparison of the fast Fourier transform (FFT) signature of each blob to the FFT signature of a rectangular shaped reference. The four points that represent the corners of the rectangular region were next found by taking the peaks of the FFT signature. A homography transformation matrix is then applied to reorient the located and properly separated potential panels in an attempt to vertically align text characters. Finally every potential characters are extracted from the panel using OCR method.

Bascon *et al*. [5] proposed a Support Vector Machine(SVM) based automatic system for road-sign detection and recognition. This system covered all existing Spanish traffic-sign shapes including circular, rectangular, triangular, and octagonal signs. Their recognition system was based on the generalization properties of SVMs. HSI color space thresholding was used for extracting candidate blobs from the input image. Then these blobs were classified using linear SVMs which was later recognized using SVMs with Gaussian kernels. The authors reported that the system is invariant to rotation, translation, scaling, and even to partial occlusions.

Jack *et al*. [6] proposed a novel system for the real-time detection and recognition of traffic symbols. They detected candidate regions using MSER approach. Candidates are detected based on the traffic sign background colour, for these backgrounds persist within the MSER process. Traffic signs with white and red/blue background were detected separately in a different manner. To confirm a candidate region as a traffic sign, and to classify them to the exact type of sign, they performed a recognition stage. HOG features were computed for the detected candidate signs and then Random Forest was used for classification purpose. A cascaded classification approach was employed. A set of synthetically generated images was used as training dataset, which was created from template images taken from an on-line database by applying some randomised distortions. This method avoided the need for real training images and made it easy to include all possible signs. The authors reported that the system is accurate even at high vehicle speeds and achieves an overall $F_{measure}$ of 0.85.

Jack *et al*. [7] used the same methods for detection and recognition of traffic sign candidates as specified in [6]. They used MSER for detection and HOG features were computed for each of them for recognition. But unlike in [6], classification was performed using Support Vector Machines(SVM). The regions were classified using a cascade of multiclass SVMs [14]. They used SVM classifier because they believed that it can be trained very quickly and easily. The authors reported an overall accuracy of 89.2% for the classifier and achieved an $F_{measure}$ of 0.84.

Fatin Zaklouta *et al*. [8] evaluated the performance of k-d trees, random forests, and support vector machines (SVMs) for the efficient traffic-sign classification using different-sized histogram-of-oriented-gradient (HOG) feature descriptors and distance transforms (DTs). To improve the k-d tree classification, they introduced a spatial weighting scheme to prioritize the interior regions of the traffic sign. The Fisher's criterion and random forests approaches were used for the most significant feature selection to reduce the memory requirements and enhance the performance, which are particularly important factors in embedded system applications. They used the German Traffic Sign Recognition Benchmark (GTSRB) data set which contained 43 classes and more than 50000 images and the corresponding HOG 2 feature for the evaluation of the random forests and Fisher's criterion for feature-space reduction. It was found that the k-d tree accuracy is improved by up to 15% when applying a spatial weighting, which focused the similarity measure on the interior of the traffic sign. The random forests yield the highest classification accuracy of up to 97%.

Gonzalez *et al*. [9] made use of MSERs for the detection of both traffic signs and text characters. They applied a method based on color segmentation and the BOVW algorithm was applied on each frame to detect white and blue traffic panels. They extracted local descriptors at some keypoints of interest after applying blue and white color segmentation. Then, images were represented as a "bag

of visual words". These regions were then classified using both support vector machines and Naive Bayes classifiers. This method was applied to single images, with no use of temporal information, and the emphasis placed on the traffic sign's geolocalization using Global Positioning System information. The height of the text itself was used to approximate the real-world size, and therefore the distance from the camera to the traffic signs. All of the results were based on 10763 images taken from Google Street View. The authors provided individual detection and recognition rates for words, symbols, and numbers at all the short, medium, and long distances. These rates ranged between 13.09% and 90.18% for detection and between 8.51% and 87.50% for recognition.

Jack Greenhalgh *et al*. [10] used scene structure to define search regions within the image, inorder to find the traffic sign candidates. Maximally stable extremal regions (MSERs) and hue, saturation, and value color thresholding were used to locate a large number of candidates. By applying temporal and structural informations as constraints, they are then reduced. Text contained within detected candidate regions were interpreted for recognition. For that, individual text characters were detected as MSERs and were grouped into lines. Then they were interpreted using optical character recognition (OCR). Recognition accuracy is vastly improved through the temporal fusion of text results across several consecutive frames. The method was evaluated comparatively and achieves an overall $F_{measure}$ of 0.87.

Nadra *et al*. [11] used a color based segmentation method to generate traffic sign candidate regions. They selected the HSV color space as it is based on human color perception. The segmented regions are analyzed and precise location of candidates are found. To encode the detected traffic signs, they extracted the HoG features and then feature vectors are generated. These vectors are used as an input to an SVM classifier which identifies the traffic sign class. Finally, a monocular tracking method based on optical flow was performed to ensure a sequential capture of the recognized traffic sign while accelerating the execution time. They applied the Lucas-Kanade tracker as it has a high performance to find the exact match under illumination. The method affords high precision rates under different challenging conditions. The authors reported an average improvement of 2.53% in the Recall rate, 3.56% in the Precision rate, and 3.12% in the $F_{measure}$ rate.

Yi Yang *et al*. [12] proposed a color probability model that deal with color information of traffic signs inorder to enhance the specific colors of traffic signs and suppress background colors as well as to reduce the search space and detection time. For that they transformed the input color image to traffic sign probability maps. Traffic sign proposals were then extracted by finding Maximally Stable Extremal Regions (MSERs) from probability maps. Next they filtered out the false positives and classified the remaining traffic sign proposals into their super classes by using Support Vector Machine (SVM) based on a novel color HOG feature. They further classified the detected traffic signs into their specific sub classes by employing convolutional neural network. The experimental results on both German (GTSDB and GTSRB) and Chinese (CTSD) roads showed that both the detection and classification methods achieve the state-of-the- art performance together with significantly improved computational efficiency. The authors reported that the detection method is 20 times faster than the existing best method and overall classification accuracies are 98.24% and 98.77% for GTSDB and CTSD, respectively.

4. CONCLUSION

In this paper, we studied and compared different techniques used for the detection and recognition of traffic signs. Much research have been focussed on the detection and recognition of traffic signs, but only less have been conducted on text contained on the traffic signs. MSER and OCR are the most popular and efficient method used for this purpose. MSER and HSV thresholding gives more performance for the detection process. Additonal structural and temporal constraints can be used to further reduce the false positives detected. OCR provides better recognition performance. Future scope in this area includes introducing more efficient techniques for text detection and recognition from traffic signs.

5. REFERENCES

[1] Dilip Singh Solanki, Dr. Gireesh Dixit, "Traffic Sign Detection and Recognition Using Feature Based and OCR Method, "*in Proc. IJRSET* vol. 2, no. 2, pp. 32-40, 2002

[2] Chiung-Yao Fang, Sei-Wang Chen, Chiou-Shann Fuh, "Road Sign Detection and Tracking, "*IEEE Trans. Veh. Tech.*, vol. 52, no. 5, pp. 264-278, Sep. 2003.

[3] W. Wu, X. Chen, and J. Yang, "Detection of text on road signs from video,"*IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 4, pp. 378-390, Dec. 2005.

[4] A. Reina, R. Sastre, S. Arroyo, and P. Jimenez, "Adaptive traffic road sign panels text extraction,"*in Proc. WSEAS ICSPRA*, 2006, pp. 295-300.

[5] S. Maldonado-Bascon, 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, Jun. 2007.

[6] J. Greenhalgh and M. Mirmehdi, "Traffic sign recognition using MSER and random forests, "*in Proc. EUSIPCO*, Aug. 2012, pp. 1935-1939.

[7] J. Greenhalgh and M. Mirmehdi, "Real-time detection and recognition of road traffic signs, "*IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1498-1506, Dec. 2012.

[8] F. Zaklouta and B. Stanciulescu, "Real-time traffic-sign recognition using tree classifiers, "*IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1507-1514, Dec. 2012.

[9] A. Gonzalez, L. Bergasa, and J. Yebes, "Text detection and recognition on traffic panels from street-level imagery using visual appearance,"*IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 228-238, Feb. 2014.

[10] Jack Greenhalgh and Majid Mirmehdi, "Recognizing Text-Based Traffic Signs,"*Trans. Intell. Transp. Syst.*, vol. 16, no. 3, Jun. 2015.

[11] Nadra Ben Romdhane, Hazar Mliki, Mohamed Hammami, "An Improved Traffic Signs Recognition and Tracking Method for Driver Assistance System,"*in Proc. ICIS*, Jun. 2016, pp. 26-29.

[12] Yi Yang, Hengliang Luo, Huarong Xu, and Fuchao Wu, "Towards Real-Time Traffic Sign Detection and Classification,"*Trans. Intell. Transp. Syst.*, vol. 17, no. 7, Jul. 2016.

[13] J. Shi and C. Tomasi, "Good features to track,"*in Proc. CVPR*, 1994, pp. 593-600.

[14] C. Cortes and V. Vapnik, "Support vector networks,"*J. Mach. Learn.*,vol. 20, no. 3, pp. 273-297, Sep. 1995.

[15] Aparna A. Dalve , Sankirti S. Shiravale, "A Survey on Real Time Text Detection and Recognition from Traffic Panels, "*in Proc. IJRASET*, vol. 3, Dec, 2015.