ABHIYANTRIKI
An International Journal of Engineering & Technology
(AIJET)
Vol. 1, No. 1 (November, 2014) http://www.aijet.in/ eISSN: 2394-627X

Review on Automatic Text and Sign Detection from Traffic Panels using Sliding Window Approach

Monali R. Patil*
M. Tech. Student
Deptt. of Computer Science & Engineering
G. H. R. A. E. T., Nagpur University, India

Amit Pimpalkar
Assistant Professor
Deptt. of Computer Science & Engineering
G. H. R. A. E. T., Nagpur University, India

Abstract

Traffic panel detection and recognition is use to support road maintenance and to help drivers. It is use to detect traffic panels and recognize the information present on street-level images. The images of traffic panel are taken by high resolution digital cameras or smart phones. It recognizes text and symbol accurately. To recognize text, system extracts local descriptors after applying green and white color segmentation. Then, images are classified using Naïve Bayes and represented as a "bag of visual words". In images if a traffic panel has been detected then Text detection and recognition method is applied on it to automatically store the information contain on the panels. We propose the system which uses spatial extension to BOVW such as sliding window, branch and bound. To recognize text exactly, we compute the prior probabilities of all the words using unigram language model .The language model completely based on a dynamic dictionary. Various algorithm use which is based on SIFT descriptors to recognize single characters and also on HMMs to recognize whole words.

Keywords*: Bag of Visual Words (BOVW), HMMs, SIFT Descriptors.*

Author for correspondence monalipatil963@gmail.com

1. Introduction

Traffic panels provide drivers information about the route by means of different traffic signs and text strings. Various shape of panel is used to show different traffic information such as rectangular panel which is commonly use to show direction of any route and distance. Triangular panel is used to show turning direction, Circular panel use to show only speed limit. There are many panels on the roads that can be confused with traffic panels, like advertisement panels.

Various public organizations and private companies have started to record street-level images. The most well-known service is Street View provided by Google. Text reading from images is a challenging problem. This is important because there is a huge variety of text appearance because of different writing styles, sizes, textures, fonts, colors, and layouts, also the presence of geometrical distortions, partial occlusions, and different shooting angles that may cause deformed text. Images suffer from blurring if it is taken from moving vehicle and its affect the segmentation step.

Segmentation method extracts objects i.e. traffic signs from the background using color information. Object localization is an important task to understand the images such as to separate text and sign from the background, or to find the spatial relations between objects in an image. For this functionality branch and bound is used.

The aim of this work is to detect traffic panels and to recognize the information and sign inside them using text detection and recognition method, for this Multi-frame integration is use to improve efficiency of system. For detecting possible rectangular planes color segmentation and shape classification is done on image. A dictionary is created which contain the words that the system is able to recognize, dataset consist panel that have a blue or a white background. An Optical Character Recognition (OCR) system has been taken as a starting point to identify the misspelled words if the detected traffic panel image is blurred. OCR takes each character as input, and then computes its feature vector and using a kNN approach the object is classified into a class.

Fig. 1: (a) Turning direction with speed (b) Stop sign (c) Speed limit sign

2. Review of Literature

Various type of the information present on traffic panels. Automatic visual classification of the information contained on road panels has not been much research. A. Reina, et al. [2] worked to detect candidates to be traffic panels by using image segmentation for blue and white colors region on image using HSI space. These candidates were classified according to their shape by method correlating the radial signature of their FFT (Fast Fourier Transform) with the pattern corresponding to rectangular shape. A homography between the original plane and the reoriented one, was carried out for an image transformation to correct the angular deviation of the panel in an image. Finally, SVM classifiers applied on the image to classify every symbol and character in grey-scale image. A priori information that can be known from the panels did not take into account because the kind of information present on the panel depends on the situation over the panel itself. H. Gómez-Moreno, et al. [3] presented the segmentation methods that can be categorized into edge detection, color-space thresholding and achromatic/chromatic decomposition. The segmentation algorithm consist four stages: 1) Segmentation 2) Detection 3) Recognition 4) Tracking. For color segmentation SVMs was use to provide better result but it need some improvements when it applied to achromatic colors. Best methods were Ohtaor RGB Normalized that was normalized with respect to illumination and no need of improvement when it used with Hue Saturation Intensity (HSI) spaces.

W. Wu, et al. [4] proposed a method to detect text present traffic panels from video. In this, two rules were use: 1) Apply a divide-and-conquer strategy to divide the single task into two subtasks, 2) Detect text from each video frame by combining (2D) two-dimensional image features with the (3-D) three-dimensional geometric structure information of objects which was extracted from video sequence. This rule gave 88.9% text detection rate and 9.2% a false hit rate. To extract the region of same color k-means algorithm was used and traffic panel candidates were detected by searching for flat regions perpendicular to the camera axis. A multi-scale text detection algorithm was performed on each candidate traffic panel area. Gaussian mixture models and geometry alignment analysis was used to combines adaptive searching, edge

detection and color analysis. For recognition all detected text lines are extracted, but how the recognition is carried out was not mentioned. A. González, et al. [5] described an approach to the VISUAL Inspection of Signs and panel ("VISUALISE") which was an automatic inspection system. It was based on light retro reflection principle. It used an active infrared illuminator when it comes into contact with the Signs and then panels are reflected. It was capable of classifying a sign or panel into a certain class of material by comparing the luminance measurements with a model obtained through a prior calibration process. The inspection process can be divided into offline process and online process. In Online process, necessary devices integrated in vehicle with all software applications to recording the input sequences corresponding to actively illuminated roads by the onboard infrared illumination system. In offline process, an image processing device based on a personal computer processed the recorded sequences after processing result contain the retro reflection and the contrast values of every sign and panel. Ashish Emmanuel et al. [6] presented an Optimal Text Recognition and Translation System for Smart phone. Any language can be converted into any other language. Genetic algorithms were used to process the image to maximize the quality for text conversion from image. The size of processed image is reduced and the image segmentation helps in reducing it further thus letting to use multiple asynchronous attempts for translating text.

J. Yebes, et al. [7] introduced approach complemented the functionality of a traffic signposting inspection system based on computer vision, which was able to collect data related to the maintenance state of traffic signs and panels automatically which was based on SIFT descriptors to recognize single characters, numbers and symbols and HMMs used to recognize whole word. The text extraction algorithm was divided into three main steps. The first one was the segmentation, which is based on a canny edge detector, Second a geometrical transformation was applied to the image so that the inclination of the panel, due to distortion, is corrected. At last the location of the different elements on the panel is found out. A wrong extraction of the foreground objects was avoided by separating horizontal edges and vertical edges of the image. The advantage of this system was it only works with geometrical features and does not use a method based on colour. So it easily adapted to other traffic signposting regulations. T. Gevers and C.M. Snoek [8] suggested the properties of color descriptors that never changes using a naming and classifying of invariance with respect to photometric transformations. Image domain and video domain were used to assessed distinctiveness of color descriptors .this paper presented three color descriptors First, color descriptor based on histograms second color moments and color moment invariants and third was color descriptors based on SIFT. The addition of color descriptors over SIFT improves category recognition. For automated recognition of individual object and scene categories, color descriptor with an appropriate level of invariance was selected. Christoph H. Lampert [9] proposed Efficient Sub window Search (ESS), a method for object localization. Branch and bound scheme was use that allows maximization of quality functions of large class over all possible sub images. The branch-and-bound search quickly identifies an interested region and to find the exact maximum it focuses on its computation in this interested region. Every image was divided into a grid of regions and represents each grid cell by a separate histogram. The proposed method hierarchically split the parameter space into disjoint subsets. For each of the subsets, desired parts of the parameter space are evaluated first and if their upper bound shows that they cannot contain the maximum then that parameter space of larger parts does not have to be examined. Further, extensions of ESS have object localization tasks require the detection of number of object location in an image. Bram Alefs et al. [10] presented a system

for detection road sign from panel based on edge orientation histograms. This system can be able to detect 85% of the objects from width of 12 pixels and 95% for objects of width of 24 pixels at a low false alarm rate. Edge orientation histograms can be calculated efficiently using integral images and they are static for small variations in rotation and position. It indicates the use of local indicators in same way as weak classifiers for AdaBoost. The application of this system consists of three steps; image acquisition, image processing and CAN interface. Third step CAN interface manages the communication with the vehicle network according to proprietary transport protocol. This system performs at 7.5 frames per second on a Laptop PC.

3. Proposed Work

The propose system is Automatic Aggregation of Text and Sign on a traffic panels Using spatial extensions to BOVW (Bag of Visual Word). This information is use to assist the driver and supporting road maintenance. We will create a Database which contains images of traffic panel these panel detected by using image segmentation for blue and white background colors. Segmentation is use to find out the location of text and signs based on their color and thus reduce search space. Color-based segmentation is use which is based on a thresholding of the input image in some color space. Spatial extension such as sliding window approach is use to find out the location of text and sign from image. It based one valuating a quality function over many rectangular sub regions of the image and taking its maximum as the object's location local optimization methods can be applied instead of global ones, by first identifying promising regions in the image and then using a discrete gradient ascent procedure to refine the detection. Text and sign are considered as a feature which is extract from panel by using the Harris Laplace salient point detector. Fast Filter Method is use to remove irrelevant feature. After removing irrelevant feature image is reconstructed and represented as BOVW. We use spatial Extension to BOVW such as branch and bound. Most of the time extracted text is blurred or some character are erased due to environmental maladies so OCR will use to identify the misspelled words or symbols on traffic panels. Unigram probabilistic language model is use to recognize word and we computes the prior probabilities of all the words to recognize text exactly. We use multi-frame integration at each single frame of the recognized information. This refers to estimation and noise reduction technique. There are some panels with uniform white or green backgrounds in the image with letters inside them, so system will consider them as traffic panels generating false positives. Panels are validating using sensitivity and specification defines as:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (1)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (2)$$

Where,

TP: True Positive TN: True Negative
FP: False Positive FN: False Negative

The sensitivity shown in equation (1) relates to the system's ability to find out positive samples and the specificity shown in equation (2) relates to the system's ability to find out negative samples, if system detects a panel and that panel is traffic panel then it is true positive. If detected panel is advertisement panel not a traffic panel then it is false positive. If panel is present in frame or image but it is not detected by system then it is True negative.

In figure 2, database contains images of traffic panel. Images contain noises so, Fast Filter Method is used to remove irrelevant feature. After removing irrelevant feature, image is reconstructed and represented as Bag of Visual Word. Panels are green in color and text on it is in black color, so color segmentation is performed. For that Color descriptor, Scale Invariant Feature Transform (SIFT) and Transformed Color Histogram (TCH) are used to detect green and white traffic panel. To extract the feature, Harris-Laplace Salient Point Detector will use. Panels contain text and sign, to classify these object Naïve Bayes is use. Text reading Algorithm is use to recognize the text and OCR will use to correct the misspelled word. For clustering we will use k-means algorithm which cluster the feature discrete number of visual words.

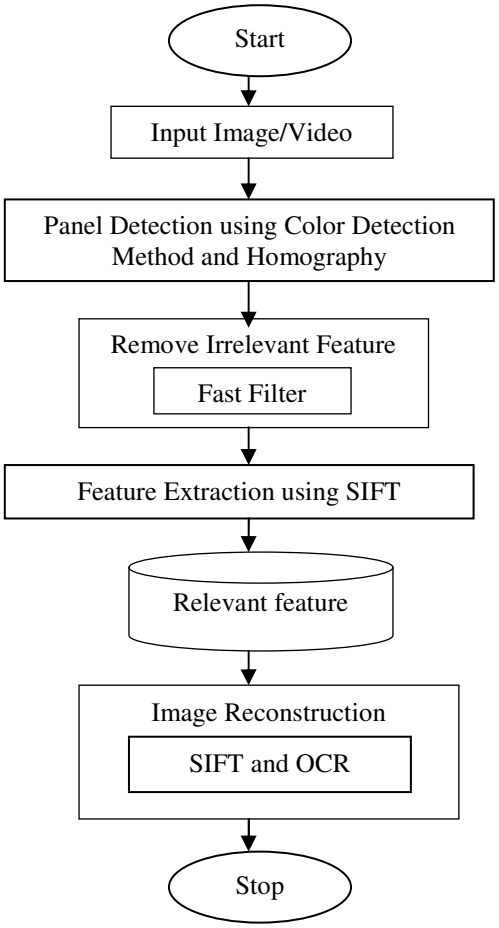

Fig. 2: Data flow diagram of proposed system

Image Acquisition

The first stage of system is image acquisition. Images of traffic panels are taken by live camera or mobile camera. Live camera image will not be in a perfect lighting and background. Image is pre-processed to enhance the image. The color image is converted to a gray scale image and image smoothing, noise removal are performed.

Fast Filter Method

Image taken by camera or mobile contain irrelevant feature that feature are remove by fast filter method. This method has two parts, first how to decide whether a feature is relevant to the class or not. To decide relevant feature method, use a user-defends threshold value SU (Symmetrical Uncertainty) values for feature. Second, how to decide whether such a relevant feature is redundant or not when considering it with other relevant features. For this SU values captured F-correlations in order to decide whether a relevant feature is redundant or not. Redundancy is because if the level of correlation between two features in S0 is high. To avoid redundancy, one of them may be removed from S0. For a feature Fi in S0, the value of SU_i quantizes the extent to which Fi is correlated to the class C. We will obtain quantized estimations about the extent to which Fi is correlated to the rest relevant features in S0. For measuring correlations between features Symmetry is a desire property symmetrical uncertainty defined as follows:

$$SU(X, Y) = 2 \left[\frac{IG(X|Y)}{H(X)+H(Y)} \right] \quad (3)$$

Where, IG is Information Gain. It is the amount by which the entropy of *X* decreases reflects additional information about *X* provided by Y and it is given by

$$IG(X|Y) = H(X) - H(X|Y) \quad (4)$$

Image Reconstruction

Image reconstruction creates 2D or 3D images from incomplete or scattered data. Image sharpening is use to make an image a readable, usable and useful. But there is problem with image reconstruction if mage contain noise and unwanted data that can interrupt the clarity of an image, so to remove noise or to correct the data OCR is use. In OCR, SIFT descriptor use to extract alphabetic characters, numbers and symbols. A SIFT key point described by a geometric frame of four parameters, the key point center coordinates x and y, its scale and its orientation angle. The fourth parameter is orientation angle because some characters are symmetrical, like 6 and 9 or b. From a set of reference images, SIFT key points of objects are first extracted and the corresponding descriptor vectors are stored in a database. These images consists all upper case and lowercase letters, all the numbers since 0 to 9. From new image character or symbol is recognized by comparing each feature vector from the new image to the database and finding candidate matching features based on Euclidean distance of their feature vectors.

Color Segmentation

In image segmentation, an image is separated into different parts that correspond to something that can easily separate and view as individual objects. The segmentation process is based on various features found in the image. This might be color information that is used to create histograms, or information about the pixels that indicate edges or boundaries or texture information. Some of colors descriptors are SIFT Hue histogram, TCH. We use TCH because it is invariance to scale and shift of intensity also it is not affected by shadow, lightening effect, and viewpoints in the scene. TCH transforms RGB (Red Green Blue) into normal distributions of the color channels for the image patches pointed by the detected key points.

4. Conclusion

The images of traffic panel that is blurred because of the image stitching are correctly detected as panels by this system. Images taken at early morning or sunset suffer from illumination effect, low contrast, and shadows on the panel, these panels are correctly detected due to TCH descriptor. This work improve recognition rate since it use multi-frame integration technique. We use prior knowledge concept to reduce the false panel detection rate.

References

[1] Patil, M. & Pimpalkar A. (2014). Review on text and sign detection from traffic panel using spatial extension to BOVW. *International Journal of Enhanced Research in Science Technology & Engineering*. *3*(10), 139-144.

[2] Reina, R. Sastre, & Arroyo, S. (2006). Adaptive traffic road sign panel text extraction. *In*: Proceedings of the 5th WSEAS International Conference on Signal Processing, Robotics and Automation (pp. 295-300).

[3] Maldonado-Bascón, S., Gómez-Moreno, H., Gil-Jiménez, P., & Laurent-Arroyo, S. (2010). Goal evaluation of segmentation algorithms for traffic sign recognition. *IEEE Transactions on Intelligent Transportation System. 11*(4), 917–930.

[4] Chen, X., Yang, J., & Wu, W. (2005). Detection of text on road signs from video. *IEEE Transactions on Intelligent Transportation System. 6*(4), 378–390.

[5] González, A., Garrison, M., & Llorca, D. (2011). Automatic traffic signs and panels inspection system using computer vision. *IEEE Transactions on Intelligent Transportation System*. *12*(2), 485–499.

[6] Emmanuel, Ashish S., & Nithyanandam, S. (2014). An optimal text recognition and translation system for smart phones using genetic programming and cloud. *International Journal of Engineering Science and Innovative Technology*. *3*(2).

[7] González, A., Bergasa, L., & Yebes, J. (2010). Automatic information recognition of traffic panels using SIFT descriptors and HMMS. *In*: Proceedings of ITSC (pp. 1289–1294).

[8] Gevers, T. & Snoek, C. M. (2010). Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis Machine Intelligent. 32*(9), 1582–1596.

[9] Lampert, C., Blaschko, M., & Hoffman, T. (2009). Efficient sub-window search: a branch and bound framework for object localization. *IEEE Transactions on Pattern Analysis Machine Intelligent. 31*(12), 2129–2142.

[10] Alefs, B., Eschemann, G., Ramoser, H., & Beleznai, C. (2007). Road sign detection from edge orientation histograms. *In*: Proceedings of IEEE Intelligent Vehicles Symposium, Istanbul, Turkey.

[11] Shekhar, R. & Jawahar, C. V. (2012). Word image retrieval using Bag of Visual Words. *In*: Proceedings of 10th IAPR International Workshop on Document Analysis Systems.

[12] Foong, Oi-Mean, Sulaiman, S., & Ling, K. K. (2013). Text signage recognition in android mobile devices. *Journal of Computer Science*. *9*(12), 1793-1802.

[13] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*. *20*(3), 273–297.

[14] Lafuente-Arroyo, S., Maldonado-Bascón, S. (2006). Road sign tracking with a predictive filter solution. *In*: Proceedings of 32nd IEEE IECON (pp. 3314–3331).

[15] Maldonado-Bascón, S., & Lafuente-Arroyo, S. (2007). Road-sign detection and recognition based on support vector machines. *IEEE Transactions on Intelligent Transportation System*. *8*(2), 264–278.

[16] González, A., & Bergasa, L. M. (2013). A text reading algorithm for natural images. *Image Vis. Computer*. *31*(3), 255–274.