International Journal of Research in Engineering and Applied Sciences(IJREAS)
Available online at http://euroasiapub.org/journals.php
Vol. 7 Issue 6 , June-2017, pp. 15~27
ISSN (O): 2249-3905, ISSN(P): 2349-6525 | Impact Factor: 7.196 | Thomson Reuters Researcher ID: L-5236-2015

A MODULARIZED APPROACH FOR REAL TIME VEHICULAR SURVEILLANCE WITH A CUSTOM HOG BASED LPR SYSTEM

Vivek Joy[1],

[1]M Tech in Network Engineering, Rajagiri School of Engineering & Technology, Rajagiri Valley, Kakkanad, Kochi, Kerala

Binu A[2],

[2]Assistant Professor, Department of Information Technology, Rajagiri School of Engineering & Technology, Rajagiri Valley, Kakkanad, Kochi, Kerala

Kuttyamma A J[3]

[3]Professor Department of Information Technology, Rajagiri School of Engineering & Technology, Rajagiri Valley, Kakkanad, Kochi, Kerala

ABSTRACT

This Vehicle detection, tracking and license plate recognition are complex computer vision problems that has many use cases in parking systems, vehicle surveillance, traffic monitoring, speed estimation of vehicles etc... A vehicle entering the targeted compound or area is detected, tracked and counted which helps in knowing the number of vehicles entering the compound. The vehicle license plate number is an unique identifier for the vehicle. The vehicle license plate number is detected and recognized in order to know the details of the vehicle and distinguish it from other vehicles. If any anomaly is detected like the vehicle going on the wrong track this would be detected using tracking algorithms and the license plate number will be sent to the respected authorities. The vehicle detection is done using background subtraction algorithm. The count of the vehicle can be amounted by using centroid estimation and tracking. A tracked vehicles license plate number is then detected using a customized HOG based SVM detector and it is localized and the license plate number is read using an OCR algorithm. If any anomaly is detected like the vehicle going the wrong track an alert will be sent with the license plate number to the respected authority.

Keywords—Vehicle detection, Background Subtraction, HOG, SVM, License Plate, LPR

Introduction

Vehicle surveillance is very crucial in areas and compounds where a large number of vehicles enter and leave every day. Monitoring vehicles in these areas is not an easy task for humans as the number of vehicles passing each time will be high. An automated surveillance system is required in such cases which can be achieved with the help of computer vision techniques. Computer vision is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images or videos. It seeks to automate tasks that the human visual system can do.

Vehicle surveillance is a complex computer vision task which can be divided into small modularized tasks and can be achieved with common image processing and machine learning techniques. The main tasks in Vehicle surveillance is

• Vehicle Detection - The first main task is to detect a vehicle in a video. Each vehicle should be detected separately and its frames should be grabbed for future purposes.

• Vehicle Tracking - The vehicle after detection should be tracked. Tracking is important as it helps in eliminating redundancy of detecting the same vehicle again. Tracking is also done to detect any anomaly in the movement of the vehicle. For example if the vehicle is going in the wrong track then it is an anomaly and this can be reported to a central authority or the vehicle can be notified somehow to correct its track.

• License Plate Detection and Recognition - If the surveillance system plans on keeping a history of the vehicles and also if it needs to notify the vehicle owner if any anomaly occurs or if any violation occurs then it is important to detect and recognize the license plate of the vehicle. The License plate is the only unique identification for the vehicle. It is what distinguishes every vehicle from each other. License plate detection and recognition is a complex task as it requires localization of the license plate area, detection of text region and recognition of the characters in the license plate.

In the paper, we focus on detecting the vehicle, tracking the vehicle throughout the video and detecting and recognizing the license plate of the vehicle, in real time. Real time detection, tracking and license plate recognition is achieved using a module based approach. Each of these tasks is achieved separately using different image processing and machine learning techniques. All the modules are perfectly synced to get the real time speed. OpenCV[9] computer vision library was used to achieve all the tasks. This library is open source and contains all the image processing and computer vision based algorithms and techniques. The library has support for different programming languages and here python programming language was used to perform the operations. The rest of paper is organized as follows. Section 2 describes the different mainly used techniques in computer vision in the area of video surveillance for detection, tracking of the object and text recognition. Previous work in this area is given in section 3. Section 4 presents the proposed method to detect vehicles, track the vehicles, detect and localize their license plates and recognize the license plate numbers. Section 5 provides the experimental set up and results. Section 6 concludes the paper followed by reference.

Computer Vision Techniques For Video Surveillance

Video to computer is actually a collection of frames. Video Surveillance using computer vision techniques is actually applying image processing and machine learning techniques to a collection of frames. These techniques consider a group of frames that constitute a video instead of single frames. The different techniques used in Video Surveillance are • Background Subtraction - Background subtraction is done to distinguish moving objects from non-moving objects in a video. In video surveillance background subtraction algorithm helps to detect the moving objects in the video and subtracts the non moving or static objects in the video. Background subtraction mainly uses a technique called history, it uses few frames in the video to calculate the static parts of the video and these static parts are subtracted from the video in the next frame onwards which results in the video having only moving objects.

• Contour Detection - It is a technique used to find the coordinates of the objects of our interest by analysing shapes in an image. It is used for object detection and localization purposes. Bounding boxes are drawn around the objects of our interest after detecting contours in the experimental setup to show the objects of interest in the video. Usually a rectangle is drawn around the objects of interest. The objects of interest from the background subtracted video is detected and localized using the contour detection technique in Video Surveillance.

• Centroid Estimation - The detected Object of interest needs to be tracked for different purposes like taking count of the object, tracking the correct movement of the objects and also it helps in eliminating duplicate detection of the same object. Tracking is done by calculating the centroid of the detected object by using the contour points. The centroid of the rectangle points after contour detection is found out and it is tracked in the subsequent frames. The object count is usually taken when the objects centroid passes through a virtual line in the video.

• Object Detection using Classifiers - If the object of interest has a well defined shape and structure with edges and similar color properties then an object detector can be trained to detect the object more accurately and robustly. It also is faster than other methods. There are different types of classifiers for detecting objects. Haar Cascade classifiers, lbp classifiers, HOG feature based SVM classifiers etc. The HOG feature based SVM classifier is considered the best among these classifiers and it has great accuracy when it comes to detecting objects with well defined shapes. Usually an object detector is trained for identifying some objects in Video Surveillance.

• Text Recognition - Text recognition is mainly used in Video Surveillance areas like people monitoring, traffic monitoring, Parking systems for recognizing license plates, recognizing text from ID cards of persons etc. Text recognition is done using the technique called optical character recognition. It is done by training characters in a language written in different forms using different machine learning techniques. Tesseract OCR is an open source library used for text recognition purposes and it is the one used in most experimental video surveillance systems.

Related Works

In[2], a method for counting and tracking vehicles using blob analysis technique is proposed. Background Subtraction was done to separate moving images and non moving objects in the video. The background subtraction method used was based on an histogram statistical model and bayes decision rule classification. The background subtraction output was foreground moving objects which are represented using white pixels called blobs. Contour detection method was used to group together the blobs. These blobs are called candidate blobs. The candidate blobs are further filtered by removing the small blobs. These part was termed as blob detection in the paper. Blob analysis was the next step used, which is the detection of candidate blobs in the subsequent frames. This is achieved using K Means clustering. The K Means clustering finds the centre of the candidate blob. These Blobs was tracked by calculating the Euclidean distances between the centroids of consecutive frames. The object with minimum euclidean distance is considered to be the same vehicle so this helps in tracking the vehicle. The vehicle counting is done by placing a base line.

When the vehicle passes the base line its image is captured.

In[1] a HOG based object detector for detecting humans is proposed. The method is based on evaluating well normalized local histograms of image gradient orientations in a dense grid. The basic idea discussed here was that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. In practice this is implemented by dividing the image window into small spatial regions (cells), for each cell accumulating a local 1-D histogram of gradient directions or edge orientations over the pixels of the cell. The combined histogram entries form the representation. For better invariance to illumination, shadowing, etc., it is also useful to contrast-normalize the local responses before using them. This can be done by accumulating a measure of local histogram energy over some what larger spatial regions (blocks) and using the results to normalize all of the cells in the block. This normalized descriptor blocks is termed as Histogram of Oriented Gradient (HOG) descriptors. Tilting the detection window with a dense (in fact, overlapping) grid of HOG descriptors and using the combined feature vector in a conventional SVM based window classifier gives the human detection classifier.

In[4] license plate recognition using template matching techniques is proposed. Template matching is done by first securing the images of the characters that are usually seen in the license plates and is stored in a database. The license plates that are extracted from an image is subjected to image segmentation to make the text visible clearly. The characters are then extracted using contour method and these characters are matched using template matching method with the templates in the database and the recognition is successfully done.

In[3] Tesseract OCR engine was used to do character recognition. They proposed a way for recognizing text and using it in a mobile application with speech to text capabilities for helping blind people. The method they processed consists of thresholding of the image to differentiate between text background and foreground as usually text background and foreground have different contrast. Connected component analysis is done to find blobs of each character. The characters are cropped and send to the tesseract OCR engine for recognition. This method improves the recognition results.

Proposed Method

In the proposed method Vehicle Surveillance is done by detecting the vehicle, Tracking the vehicle and detecting and recognizing the license plate of the vehicle. Vehicle Detection is done using the background subtraction method, vehicle tracking is done by contour and centroid estimation methods, license plate is detected and localized using a custom classifier for detecting license plates and recognition of license plate number is recognized by training an Optical Character Recognition classifier using Tesseract OCR library. The Proposed system design is shown in the figure below:

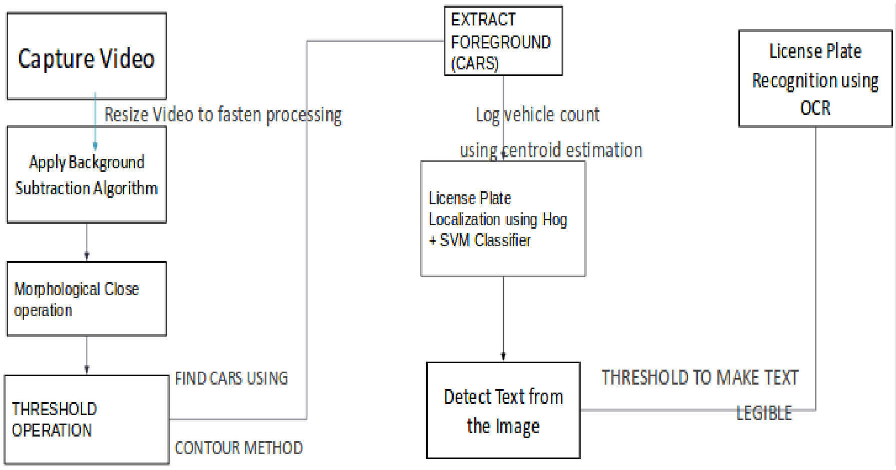

Fig. 1: System Design

The Video is first captured from the surveillance camera frame by frame. The captured frames are of high resolution processing of which requires huge resources, to fasten the processing the video frames are resized to a quarter of its length. The latter operations are done on this resized video frame and the final results are merged to the original video frame. The next step is background subtraction - It is the process of extracting moving foreground from static background. It is easier if there is already an image of the static background to begin with, but this is not the case in most scenarios. The solution is a technique called history. The background subtraction algorithm takes a few frames in the video as history frames. The algorithm understands the static parts of the video from the history and subtracts these parts to extract the foreground. Another major hurdle is the presence of shadows. Since shadow is also moving, simple subtraction will mark that also as foreground which complicates things. So taking all these into consideration the background subtraction algorithm used here is called Background Subtractor MOG.It is a Gaussian Mixture-based Background/Foreground Segmentation Algorithm. It uses a method to model each background pixel by a mixture of K Gaussian distributions (K = 3 to 5). The weights of the mixture represent the time proportions that those colours stay in the scene. The probable background colours are the ones which stay longer and more static. The shadows all have the same properties in a video. The algorithm manages to identify shadows and mark it as gray in the subsequent frames. each pixel is characterized by its intensity in the RGB color space. Then, the probability of observing the current pixel value is considered given by the following equation

$$P(X_i) = \sum_{i=1}^{K} \omega_{i,t}\eta(X_i, \mu_{i,t}, \epsilon_{i,t}) \quad (1)$$

where K is the number of distributions ω is a weight associated to the i th gaussian at time t with mean μi,t and standard deviation €i,t η is a gaussian probability density function given by equation:

$$\eta(X_i, \mu, \epsilon) = \frac{1}{2\pi^{n/2}|\epsilon|^{1/2}} e^{\frac{1}{2}(X_i-\mu)\epsilon^{-1}(X_i-\mu)} \quad (2)$$

each pixel is characterized by a mixture of K Gaussians.Once the background model is defined, the different parameters of the mixture of Gaussians must be initialized. The parameters of the MOG's model are the number of Gaussians K,the weight ωi,t associated to the i th gaussian at time t, the mean μi,t and the covariance matrix €i,t

$$\epsilon_{i,t} = \sigma^2_{i,t} I \quad (3)$$

The background subtraction operation gives the moving foreground frames but a captured video contains noises and the backgrounds subtracted frames will contain these noises as small white pixels. As these are small pixels it can be removed using a morphological open operation. Morphological transformations are some simple operations based on the image shape. It is normally performed on binary images. It needs two inputs, one is our original image, second one is called structuring element or kernel which decides the nature of operation. Two basic morphological operators are Erosion and Dilation. Erosion is erodes away the boundaries of foreground object which are white pixels. The kernel slides through the image ,a pixel in the original image (either 1 or 0) will be considered 1 only if all the pixels under the kernel is 1, otherwise it is eroded (made to zero). All the pixels near boundary will be discarded depending upon the size of kernel. So the thickness or size of the foreground object decreases or simply white region decreases in the image. It is useful for removing small white noises. Dilation is just the opposite of erosion. Here, a pixel element is 1 if at least one pixel under the kernel is 1. So it increases the white region in the image or size of foreground object increases. Normally, in cases like noise removal, erosion is followed by dilation. Because, erosion removes white noises, but it also shrinks our object. So we dilate it. Since noise is gone, they won't come back, but our object area increases. It is also useful in joining broken parts of an object. The frames obtained after background subtraction is subjected to morphological opening which is just another name of erosion followed by dilation. It is useful in removing small noises. Erosion erode away the small white pixels in the image(the noises) and this may result in erosion of some of the boundaries of the objects of our interest (vehicle), The dilation operation after will enhance the boundaries of the objects(vehicles). Morphological opening can be represented using the following equation

$$A \circ B = (A \ominus B) \oplus B \quad (4)$$

where A is the image and B is the kernel, denote erosion and $\oplus$ denote dilation. The noises are now removed and we have clear foreground images, but the problem of shadows still exists. The background subtraction has marked the shadows with white pixels of lower thresholds(gray color). So a method called Simple thresholding is done to eliminate the shadows in the frames. If pixel value is greater than a threshold value, it is assigned one value (white), else it is assigned another value (black). The threshold value is set to a value greater than gray color(threshold value = 127). The result is the complete elimination of the shadows from the frames. The resulting frames now will have some significant moving white pixels but all these moving white pixels may not be our object of interest(vehicle). We want to detect the pixels or points that consist of our object of interest (or Region of Interest(ROI)) to do further processing on the frames. This requires some shape finding and detection techniques. One such technique that is commonly used in computer vision is called contours. Contours can be explained simply as a curve joining all the continuous points (along the boundary), having same color or intensity. The contours are a useful tool for shape analysis and object detection and recognition. The contours in the shape of the vehicle size are filtered out from the frames. In experimental setup the contours are usually drawn on images to show that we have found the object of our interest. A bounding box in the form of rectangle is drawn along the contour points of the object. These concludes the detection part of the paper. After vehicle detection the next step is to count and track the number of vehicles entering the area or compound. The vehicle detection is done on each frame and we have to identify the same vehicle and distinguish other vehicles in subsequent frames to count and track the vehicles. This can be done by a method called centroid estimation. The centroid of the contour points of the vehicle is found out in each frame and its distance between consecutive frames is calculated. If the distance is less than a threshold value then it belongs to the same vehicle. The vehicle count is taken when the centroid passes a virtual line drawn in the frames to determine the count of the vehicles. When the centroid of the vehicle passes the virtual line the car count is incremented by a factor of 1. Centroid of the contour points is calculated using the technique called moments. Image moments help you to calculate some features like center of mass of the object, area of the object, centroid etc... Centroid is given by the equations:

$$C_x = \frac{M_{10}}{M_{00}} \tag{5}$$

$$C_y = \frac{M_{01}}{M_{00}} \tag{6}$$

The same method is also used to track the direction of the vehicle movement. It is used to find out whether the vehicle is entering and leaving through the correct entry and exit points. If the vehicle goes in the wrong direction an alert is generated and send to the respected authorities. Two virtual lines are drawn. If the vehicle crosses the first line and then crosses the second in the entry scenario then it is travelling in right direction, for the exit case scenario opposite of this method is used to check the direction. The virtual line for counting the vehicle is drawn at a position such that the vehicle can be captured with the license plate being visible for the plate detection and recognition purposes. The captured image is sent to a custom classifier creat ed for identifying license plates. The classifier detects the license plate and the area containing the license plate is localized and send for further processing. The classifier that is being used is trained based on Histogram of Gradients features. The training of the classifier is a quite complex task but the result obtained is quite accurate. The training of the classifier is done as shown in the Fig 2.

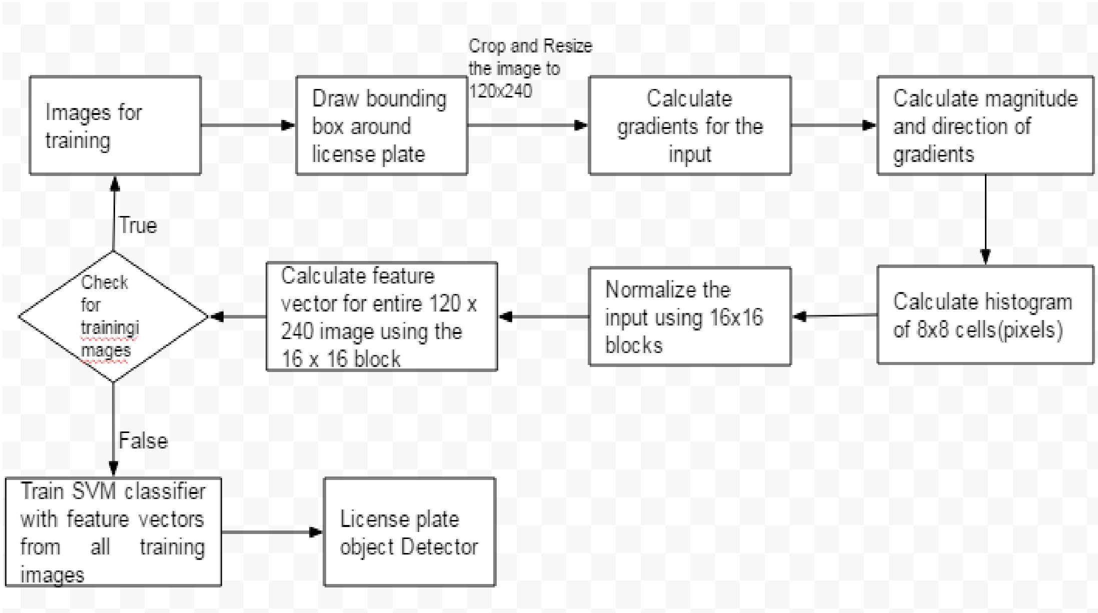

Fig. 2: Training License plate object detector

The images for training are the images of cars with license plate numbers clearly visible. After selecting the training images the images are one by one prepared and its hog features are calculated. First a bounding box is drawn near the object of interest(license plate) and the image is cropped to that region as a 120x240 image patch. The next step is calculating gradients of the image patch, gradients are the derivatives in the x and y direction of the image. The magnitude and direction of the gradients are calculated using the equations

$$M = \sqrt{g_{x^2} + g_{y^2}} \quad (7)$$

$$\theta = arctan\frac{g_y}{g_x} \quad (8)$$

For every pixels there are two values for gradients the magnitude and the direction. The next step is to calculate histogram of the gradients, for this the image patch is divided into 8x8 cells. The histogram is a collection of bins of equal length the equal length is the direction from 0 to 180 degrees. So 9 equal bins are created which are 0,20,40,60,80,100,120,140,160. The magnitude value that we calculated for the pixels are matched to the corresponding direction values of the pixel and is send the corresponding bin. At last after considering all the pixels we get histogram of gradients for the 8x8 cell. Like that we calculate for all the pixels. The next step is normalization, normalization is done so that the detector works on all lighting conditions. The image is considered as 16x16 blocks, each block contains 4 8x8 cells ie.. 4 histograms. There are 9 bins in each histogram which gives a total of 36 x 1 feature vector when concatenated. The normalization is done by finding the square root of sum of squares of the 36 features and dividing all the features by this value. The 120x240 block contains 7 blocks in the horizontal direction and 15 in vertical direction. So there are a total of 36x7x15=3780 dimensional vectors. The process is repeated for all the remaining images, when the HOG features of all the images are calculated a SVM classifier is trained with all the feature vectors. The SVM classifier after training acts as the object detector. After the localization of the license plate region the text portion in the license plate should be detected and extracted out for recognition purposes. The text portion is detected using connected component analysis. This is a method used to find out connected components in an image using a feature detector called MSER(Maximal Extremal Regions). Each character is a connected component of a particular shape. The connected components are all detetcted as a block and is cropped out from the image. The text is then thresholded so that any non text portions are eliminated from the image. The text is then fed to the Optical Character Recognition library called Tesseract. Tesseract is an open source OCR engine that is used for the purposes of OCR font training and character recognition. The license plates will have many different kinds of fonts and changes from vehicle to vehicle. To correctly recognize the characters it is important that many fonts are trained using the library before the recognition process. The Tesseract OCR library has a very good recognition rate but it does not recognize the characters if the images are not preprocessed according to certain conditions. The images should have the text as the clearly visible portion. This can be achieved using thresholding, the images should not be skewed and it should be scaled to a certain factor for good recognition accuracy.

Experimental Setup and Results

The Experimental setup was done using Surveillance footage recorded in a DVR of a CCTV camera. The computer vision library that was used for the setup is OpenCV[9] library. The OpenCV library is an open source computer vision library that supports many programming languages like C++, python, Java etc. The experiment was done using Python programming language on linux 16.04 operating system. The HOG SVM object detector for license plate was trained using the help of a library called dlib which is a C++ based library for machine learning purposes.The Optical Character Recognition library used was Tesseract 3.0 with the help of pytesseract library which acts as a wrapper for tesseract so that it can be used with python. The Video is captured frame by frame and Background subtraction is applied to it. Fig 3 shows the result of Background Subtraction.

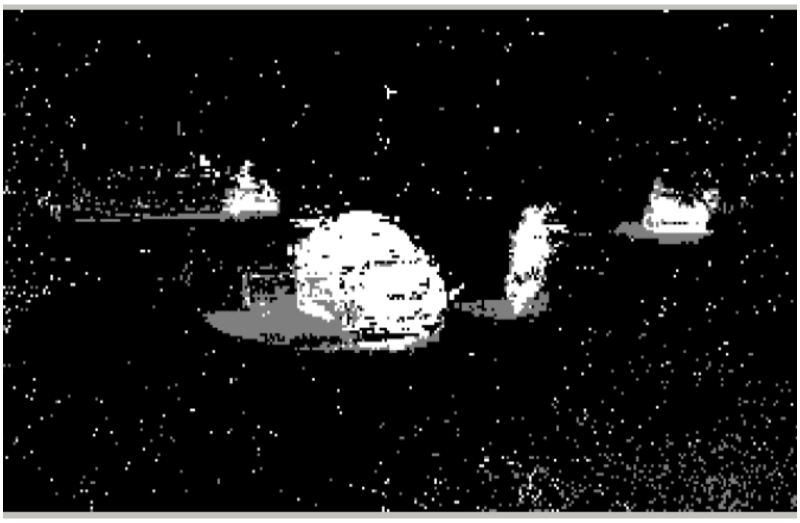

Fig 3: Background Subtraction result

The background subtraction result contains noises and shadow. The shadow is the gray marking in the sides of the vehicles. The noises are the small grainy white pixels in the image. Morphological opening and thresholding operations are done to eliminate the noises and shadows. The results are depicted by Fig 4

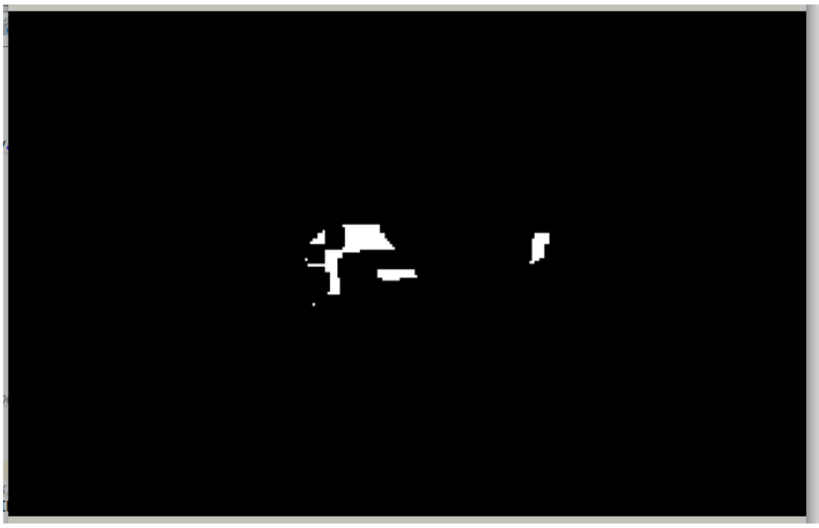

Fig 4: Morphological and Thresholding result

The vehicle is detected by contour technique and it is tracked using the centroid estimation method. Here the vehicle that is being focused on is car. The car count is taken when the vehicle passes a fixed point. The result is shown in Fig 5

Fig 5: Contour detection and centroid estimation result

The image of the vehicle is captured when it crosses a fixed point and it is sent to the license plate HOG SVM detector. Fig 6 shows the detection result of the HOG SVM license plate detector.

Fig. 6: HOG SVM license plate detection result

The license plate after detection is cropped and the text detection and thresholding operations are applied to process the image for OCR library.The cropped plate is shown in Fig 7.

Fig 7. Cropped license plate after detection

The text detection and processing result is shown in Fig 8

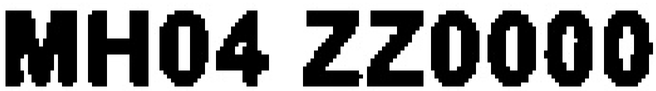

Fig 8: text detection result

The processed text is then send for recognition to the tesseract OCR library and the recognized text result is shown in Fig 9

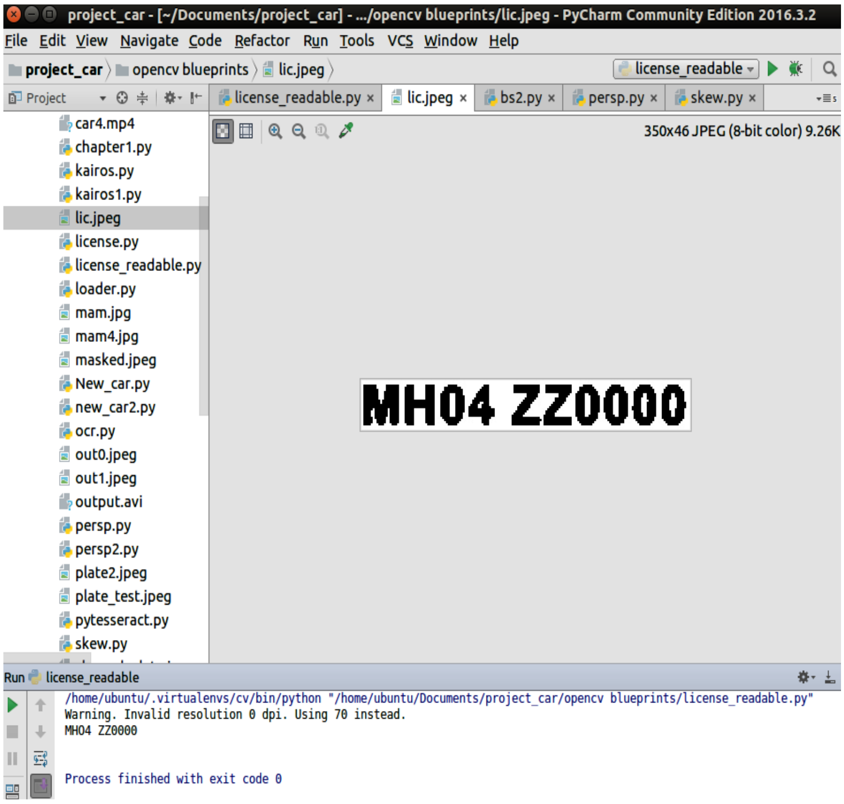

Fig. 9: Result of the recognized license plate number

Conclusion

Vehicle Detection, tracking and license plate recognition is very useful for video surveillance purposes as it automates many processes without human intervention and can be achieved in a reasonably lower expense. The vehicle detetction was done successfully using the background subtraction, morphological and thresholding operations and contour detection. The tracking and counting of each vehicles was done using the centroid estimation method and centroid tracking method. A custom HOG feature based SVM classifier was trained for detecting the license plate for accurate detection. The License plate detection and localization was successful with the HOG SVM plate detector and the detected plate's text was extracted using MSER connected component analysis. The text was the processed using thresholding method before sending to the recognnizer which was trained for popular license plate fonts. The Text was recognized by the Tesseract OCR engine with good accuracy. This technique has many applications. They can be combined with databases and data mining techniques to make useful observations. The detected license plate number can be used to grab the owner details of the vehicle from the website of the motor vehicle department. These details can be used to know the visiting times of the owner to the particular destination. Classifications can be done to identify the type of vehicles. The applications that can be thought of are limitless.

References

[1] Navneet Dalal ,Bill Triggs. Histograms of Oriented Gradients for Human Detection.Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognitio, CVPR 2005.

[2] G.Salvi .An Automated Vehicle Counting System Based on Blob Analysis for Traffic Surveillance. in PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING COMPUTER VISION AND PATTERN RECOGNITION, 1; 397-402; Image processing, computer vision, & pattern recognition; IPCV 2012

[3] Sakchai Tangwannawit, Wanida Saetang.Recognition of Lottery Digits using OCR Technology International Conference on Signal-Image Technology & Internet-Based Systems , 2016.

[4] Animesh Chandra Roy, Muhammad Kamal Hossen, Debashis Nag. License Plate Detection and Character Recognition System for Commercial Vehicles based on Morphological Approach and Template Matching. International Conference on Electrical Engineering and Information Communication Technology (ICEEICT,2016.

[5] C.E. Anagnostopoulos, I.E. Anagnostopoulos, V. Loumos, and E. Kayafas, A license plate recognition algorithm for intelligent transportation system applications,IEEE Trans. on Intelligent Transportation Systems, vol. 07, no. 03, pp. 377-382, 2006.

[6] M. Piccardi, ”Background subtraction techniques: a review”, IEEE International Conference on Systems, Man and Cybernetics, 2004.

[7] Meng Cao, Anh Vu, Barth, M., ”A Novel Omni-Directional Vision Sensing Technique for Trafc Surveillance”, IEEE Intelligent Transportation Systems Conference, 2007, pp. 678-683.

[8] Spector, A. Z. 1989. Achieving application requirements. In Distributed Systems, S. Mullende

[9] http://docs.opencv.org/3.0-beta