**Chapter 1**

**Introduction**

In today’s world, image processing concepts such as foreground extraction and key frame extraction are used extensively to reduce computational burden on post processing of the video. Post processing of the video can mean to extract features of objects of interest, for example face recognition or perhaps, for surveillance in cameras or perhaps, for analysis in detection of traffic violations, to name a few. In any case, one problem that many image processing techniques face today is variable frame rates and large amounts of background noise that is created during recording of the video being analyzed. This variable frame rate introduces lag in the video being analyzed and in turn results in the insertion of blank frames during frame extraction. Also, depending on the quality of the recording significant amount of noise can be present in the background causing loss of accuracy and in turn compromises the efficacy of the analysis. Thus, the variable frame rate and noise hinders meaningful frame extraction by resulting in frames that have no discernible foreground movement. Detection of motion of objects to apply to traffic applications that accomplish tasks such as violation detection, depend greatly on the quality of the video being processed. Therefore, an efficient pre-processing algorithm is imperative in order to ensure best results. In addition to this the method being used for detecting such a violation needs to benefit from the pre-processing stage or take full advantage of the improved quality obtained from pre-processing.

Much of the past and on-going research done in this field aims at resolving these issues in order to improve accuracy of results. There have already been various methods proposed to deal with filtering out unwanted frames that simply burden the computer with additional processing load whilst carrying no information of use during the analysis. The method that has become most popular is known as the Gaussian Mixture Model [1]. This method models each background pixel with a mixture of K Gaussian distributions, the value of K being three to five. The time that a pixel stays in the scene is determined by the weights of the mixtures in the distribution. The most probable background colours are the ones that stay longer as determined by the weights. In this method the number of Gaussian distributions ‘K’ has to be statistically determined through pre-processing of the video being analyzed. Therefore, the method suffers the disadvantage of the time overhead required for this statistical analysis in order to obtain efficiency in results. Few more methods have been proposed since, notably the improved Gaussian Mixture method and the Bayesian Gaussian Mixture model to overcome the shortcomings of the first Gaussian Mixture Model method. The advantage in the first of the two is the adaptive choosing of the appropriate number of Gaussian distributions for modelling each pixel [2][3]. The second method incorporates Bayesian segmentation to target the specific issue pertaining to video analysis in variable lighting conditions [4].

However, all of these methods suffer from seemingly un-avoidable background noise. Also, they fail to address the issue of lag in frame rates of recording video thus, resulting in introduction of blank frames that, along with the foreground frames are unnecessarily processed during the analysis phase. Background noise can be decreased by frame averaging and blurring but this technique also decreases the quality of pixels forming the scene. In our method we have addressed concerns of variable frame rates and noise by making use of a clustering algorithm called ‘K-means Clustering’ to colour quantize the image using two clusters to obtain a clear de-noised image. This clustering is performed on the images obtained after background subtraction and thus frames that introduced as a result of the lag in frame rate are rendered completely black. We then make use of a mean squared error similarity checker algorithm and filter out the blank frames and compiling a video from frames that contain significant foreground movement thus, greatly reducing the computational burden during the violation detection phase. Furthermore methods that have been proposed to detect violation such as colour based tracking, mean-shift object tracking and Cam shift object tracking are dependent on local histogram distributions of pixels in the image [5][6]. Therefore, they inadvertently suffer the disadvantage of requiring a high quality video with perfectly preserved histogram distributions and no noise in order to track objects accurately. Our method implements a very straight forward and computationally highly economical method of detection violation, particularly signal jumps. It only requires basic geometric calculation and interfacing with a camera.

**1.1 Literature Survey**

Many researchers have continued to innovate, contribute and increase performance of the image processing systems, since their inception. Innovation was revolutionized during the world wars when image processing and object tracking were indispensible tools with regard to air and naval warfare. Concepts that went into this endeavor have been improved upon with the advent of UAV’s [7]. In more recent times, because of ever increasing traffic and population the need for efficient traffic management systems became a necessity. Image processing has found much application in this field. Many tracking algorithms, background subtraction methods and de-noising techniques have been proposed and implemented to help aid with the pre-processing of traffic scene recordings, an unavoidable step. Although, it is still not perfect, significant progress has been achieved with regard to improving quality of footage, removing noise and frame rate lags, removing shadows and illumination changes, etc. We aim to further this research and apply to detecting violations of the nature of signal jumps in a computationally economic, yet accurate way.

**1.1.1 Issues**

Employing image processing techniques to achieve improved quality of traffic scene recordings to facilitate violation detection posses a daunting, yet very interesting challenge. These challenges include dealing with background noise, variable frame rates, illumination changes and shadow removal. Also, the need for removal of certain parts of the video that can add no value to the analysis of the scene is vital with regard to decreasing computational burden. The following explains these issues in brief.

* **De-noising**: Noise refers to minute disturbances in the pixel shading of an image. The recording on a camera cannot exactly simulate how the scene looks in reality. There is always present background noise occurring because of the internal operation of a camera. The optics involved cannot eliminate all the noise. In still pictures, this is a simpler problem. However, in highly dynamic video recordings such as traffic scenes, de-noising the footage is a very challenging task.
* **Foreground extraction**: Foreground extraction refers to the separating of moving objects in the scene from background objects, including non-static background objects. For example, a tree slightly displaced by the motion of a wind current in the scene is still considered a background object. Defining what can be categorized as a foreground object in a dynamic scene is a delicate problem and is also highly application specific.
* **Variable frame rate**: Videos when recorded have a feature associated with them called frame rate or “frame rates per second (fps)”. The video can be recorded at 10fps or 25fps, whatever is the case, and the algorithm used for processing of the video must work in both cases. Also, the lag due to shifting frames introduces many unwanted frames that become evident during extraction and can bring down the computational efficiency of the task being carried out.

**1.1.2 Existing systems**

The research regarding de-noising, background subtraction for variable frame rate videos has long been a trending topic of interest among the image processing research community. As it is, various systems already exist to tackle the above mentioned problems, which are ubiquitous to all domains of further application. The following points introduce these systems.

* **K-SVD de-noising algorithm:** This method makes use of sparse and redundant representations over trained dictionaries, which represents the content of the image. Therefore, it uses Bayesian treatment of images to obtain de-noised images [8].
* **Adaptive wavelet thresholding:** It is also developed in a Bayesian framework and makes use of the generalized the Gaussian distribution. The image is divided into wavelet sub bands , and the thresholding process is adaptive to each sub band to accomplish de-noising of the image [9].
* **Global de-noising:** In this method a “Global Filter” is applied in which each pixel in the image is estimated from all other pixels in the image statistically, as opposed to applying adaptive filters to patches of the image [10].
* **Gaussian Mixture Model:** This method models each background pixel with a mixture of K Gaussian distributions, the value of K being three to five. The time that a pixel stays in the scene is determined by the weights of the mixtures in the distribution [1].
* **Improved Gaussian Mixture Model:** In this method the K distributions used for modelling is appropriately determined for each pixel in the image. Thus, it is more adaptive than the Gaussian Mixture Model [2][3].
* **GMM adaptive to variable lighting conditions:** This method incorporates per pixel Bayesian segmentation into the Gaussian Mixture Model in order to account for videos recorded in variable lighting conditions [4].
* **Adaptive variable frame rate coding:** This method adjusts the frame-rate of the video dynamically and adaptively, making use of information from already existing video encoders [11].
* **Intermittent motion coding:** This method involves disabling of motion coding during periods of inactivity in the video. Thus it records only parts of the video were active foreground movement is involved for further processing [12].

All of the methods explained above incur considerable overhead with regard to time or CPU usage. The Gaussian Mixture Model based methods cannot efficiently deal with variable frame rates in videos. The de-noising algorithms apply blurring techniques to the image which alter the quality of the image in question. The variable frame rate coding techniques make use of video encoder information, the compilation of which involves CPU overhead. Also, recording only during periods of activity means that the definition of activity in the scene has to be pre-determined in advance, and done so using extensive statistical analysis.

**1.2 Motivation**

When trying to develop a system that can detect signal jumps by vehicles in a traffic scene, firstly, the video recording of the scene needs to be free of background noise. Moreover, an efficient background modelling technique is required in order to obtain an accurate foreground mask, while also ensuring that non-static background objects are not recorded in the foreground mask. Also, when footage is recorded, the frame rates per second introduce lag in the video and thus results in insertion of unwanted frames that slow down the CPU during analysis of the traffic scene for violation detection.

After all this is accomplished, already existing methods such as mean-shift and cam shift can be used effectively for object tracking. However, these techniques make use of feature extraction methods such as SURF which are highly CPU intensive tasks [13]. The widely popular Haar Cascade Vehicle Classifier can also be used to track vehicles in the scene and obtain characteristics [14]. But, the question is to detect a signal violation is can there be a more computationally efficient method? We have developed a system that makes use of simple geometric calculations to detect signal jumps in traffic scenes.

**1.3 Problem Statement**

Key frame extraction in videos with variable frame rate followed by detection of traffic violation from the newly compiled video. The nature of the traffic violation considered in this project is signal jump. The video in the problem is recorded in a real life traffic scene.

**1.4 Objectives**

As a result of the video input being a recording of a real scene, background noise and the “frames per second” at which the video has been recorded need to be accounted for before proceeding with detection of signal jump. The vehicles in the scene need to be identified and the frame pertaining to the signal jump of the vehicle concerned is saved and time stamped as per the date and time of violation.

**1.5 Scope**

The system takes a MPEG-4 or AVI format video recording of a real life traffic scene, performs background modelling for inclusion of non-static background images. Next, background subtraction is performed against the modelled background to obtain the foreground mask of the vehicles in the scene. Following this, the foreground mask images are colour quantized using clustering and a mean square error similarity checker algorithm filters out unwanted frames. The new set of frames is re-compiled into a video to being violation detection.

The input to the next system is the newly compiled video. The output is a time stamped frame capture of the signal jump of the vehicles in the scene. We have also included a simulation of post processing of number plate extraction and driver information gathering to demonstrate how this step can be carried out in future. Each stage of the processing from start to finish is decomposed into user interactive operations in order to clearly show what happens in each phase of processing.

**1.6 Methodology**

A video of a traffic scene is recorded using a standard resolution camera. The video is given as input to the system. The background in the traffic scene is modelled using the running average method [15][16]. By assuming a reasonably high regulator value, non-static background objects is successfully incorporated into the background. Background subtraction of all the frames is carried out against this background. We have used the standard frame differencing for background subtraction. Other methods can also be used to the same effect [17].

Next we cluster the foreground masks obtained using K-means Clustering with two clusters thus obtained two coloured colour quantized images of the foreground. The foreground images so obtained contain no noise and are as clear as the original frame [18]. After colour quantizing the images, the frames responsible for the lag are rendered completely blank. Hence, we make use of a simple Mean Squared Error similarity algorithm to filter out these blank frames and recompile the video [19].

From the newly compiled video we contour the image and then identify the largest contour in the image [20]. Thus, this contour represents the vehicle of interest. The contour is then approximated into the shape of a polygon whose centroid can be calculated [21]. The distance from the centorid to the edge of the frame is calculated. When the distance is less than an assumed threshold, the frame of occurrence of violation is captured and time stamped.

Next, the back view of the vehicle is captured using a electric signal to a secondary camera. Subsequently, the plates are extracted and numbers are identified using a digit classifier trained using support vector machines. The numbers are identified using histogram of oriented gradients [22]. Following this step information is gathered from the database indexed by the identified numbers are the filtered results are displayed.

**1.7 Organisation of Thesis**

**1.8 Summary**