Department of Computer Science & Engineering

University of Moratuwa

ViDasa

Vision Based Pedestrian Crossing Controller

Literature Review

Project Supervisor

Dr. Chathura de Silva

Project Members

Amila Fernando	110159E
Dulitha Kularathne	110305B
Hiran Kulasekara	110306E
Thilina Madumal	110339G

Date of Submission: 18th of September 2015

Table of Contents

List of Figures iii

Acronyms iv

1 Introduction 1

1.1 Overview 1

1.2 Problem Specification 2

1.3 Our Solution 4

1.3.1 Methodology 5

2 Literature Review 7

2.1 Pedestrian Detection (Human body detection in road environments) 7

2.1.1 Motion Based Detection 8

2.1.2 Direct Detection 11

2.1.2.1 Holistic Detection 11

2.1.2.2 Patch Based Detection 14

2.1.2.3 Part Based Detection 16

2.2 Vehicle Detection, Classification and Tracking 20

2.2.1 Motion Based Vehicle Detection and Segmentation Approaches 20

2.2.1.1 Background Subtraction Method 21

2.2.1.2 Feature Based Method 22

2.2.1.2.1 Detection of Vehicles using Haar-like Features 22

2.2.1.3 Frame Differencing and Motion Based Method 24

2.2.2 Camera Calibration Approaches 24

2.2.3 Vehicle Tracking Approaches 25

2.2.3.1 Region-Based Tracking Methods 25

2.2.3.2 Contour Tracking Methods 26

2.2.3.3 3D Model-Based Tracking Methods 27

2.2.3.4 Feature-Based Tracking Methods 28

2.2.3.5 Color and Pattern-Based Methods 29

2.3 Algorithms 30

2.3.1 Kalman Filter 30

2.3.2 AdaBoost Algorithm 30

Reference v

List of Figures

Figure 1: Methodology Block Diagram ... 6
Figure 2: Codebook Generation ... 15
Figure 3: Definition of body parts. ... 16
Figure 4: Modeling a human head as an ellipse. ... 17
Figure 5: Detecting a head using ellipse concept ... 18
Figure 6: Detecting human heads and shoulders from an overhead ... 19
Figure 7: First 5 Haar-like features used in Viola and Jones method ... 23
Figure 8: Cascade Classifier Architecture ... 23
Figure 9: Haar-like Features used in Vehicle Detection ... 23
Figure 10: Camera Setting ... 25
Figure 11: Average optical flow vectors (red arrows). ... 27
Figure 12: System overview ... 28
Figure 13: Pseudocode for AdaBoost ... 30

Acronyms

SVM	Support Vector Machine
BS	Background Subtraction
AP	Affinity Propagation
SIFT	Scale Invariant Feature Transform
CCTV	Closed-Circuit Television
HOG	Histogram of Oriented Gradients

1 Introduction

1.1 Overview

Traffic congestion in urban areas has become a key problem in many countries including Sri-Lanka. Due to this problem huge number of people suffer in their day-to-day life, desperately wasting their valuable time on roads trying to get to their destination. It also affects Industries, corporate sector, and government sector dropping down productivity which directly result in degradation of economic figures in a country. When it comes to developing countries like Sri-Lanka the effect of this particular issue becomes worse. Unlike developed countries, spending a lot of money and resources on traffic congestion reduction is not very feasible and desirable by third world countries. In spite of the difficulties third world countries should somehow take measures to reduce traffic congestion in urban areas if they are to thrive for development.

There are two key solutions for the high traffic congestion which should go hand in hand. One is to construct new roads, renovate already existing roads, and developing alternative transportation methodologies. The second option is to provide efficient traffic controlling. Obviously the first option costs a lot of money and resources, so the developing countries like Sri-Lanka find it difficult to spend a lot in the second option. In that circumstances our Vision Based Pedestrian Crossing Controller which is comparably low cost (in installation, maintenance, and upgrading) but effective, and has quite number of extended usages, comes in handy.

The aim of the project is to develop a software system that analyzes the real-time video footage taken by a 360 view fish-eye lensed Closed-Circuit Television (CCTV) camera mounted at the pedestrian crossing and generate parametric signals to be sent to the traffic light controller. The decision should be based on number of Pedestrians waiting to cross the road on both sides and vehicle movements in terms of average speed of arrival and type of vehicle (i.e. motorcycle, large vehicles, small vehicles etc.). In addition Vehicles violating the red light condition & other traffic rules such as overtaking on pedestrian crossing will be detected and identified. Furthermore all the statistical data derived from the real-time video footage will be sent to a central system.

Extended usages of this system provide solutions for the early mentioned problem of traffic congestion in urban areas. Statistical data that is sent to a central system by each Vision Based Pedestrian Crossing Controller can be used to provide real-time traffic condition updates on different roads and areas for interested parties. Also the statistical data can be used to come-up with traffic models, planning day-to-day traffic handling, and recognizing traffic patterns in specific areas that would be helpful in planning new road constructions, renovating roads, and developing other alternative transportation methods.

In the remaining two subsection of the Introduction, the report presents the 'Problem Specification' which describes the problem with references and 'Our Solution' under which our approach for solving the specified problem is described respectively.

1.2 Problem Specification

As number of vehicles increases, the requirement for better vehicle traffic controlling methods arises. From the early days manual traffic controlling has been a good solution for some extent but with the limitation of manual traffic controlling there had been a trend on automatic traffic controlling. Although automatic traffic controlling methodologies have been experimented and implemented for a long time still manual traffic controlling is playing a major role in traffic controlling on roads. Prevailing automatic traffic controlling systems can be categorized into two parts,

1. Fixed Time Control
2. Dynamic Time Control

Typical Dynamic Time Control use following methodologies for vehicle and pedestrian detection,

I. In-pavement detectors (sensors are buried under the ground)
II. Non-intrusive detectors (over road detectors such as cameras, electromagnetic waves and acoustic sensors)
III. Non-motorized user detection (manually press button or give command vocally to audio detectors)

These methodologies have various drawbacks when we consider them individually. High installation and maintenance cost in In-pavement detectors, user-unfriendly and inefficient behavior in Non-motorized detectors, and difficulty in maintaining and installing Non-intrusive detectors except for cameras, are some of the drawbacks in above listed detectors.

Fixed Time Controllers don't address the issue that we described at the beginning because it doesn't provide adaptive traffic controlling which is an essential requirement for reducing traffic congestion on roads. Therefore we could identify the necessity of an adaptive traffic controlling system. Vehicle and pedestrian detection is a key component of an adaptive traffic controlling system. While achieving that goal easy installation, maintenance, and upgrading should be preserved. Thus we can conclude that the vision based traffic controlling systems apparently the best desirable solution for an adaptive traffic controlling system.

As we refer the literature on already developed adaptive traffic controlling systems we can identify several key systems developed by various parties. Those systems have their own advantages and disadvantages. When it comes to developing countries like Sri-Lanka, the compliance of those systems are quite questionable. The very reason is the high cost in installation and maintenance. When it comes to low cost solutions, accuracy and reliability figures are low. Let's have a look at on some of those major adaptive traffic controlling systems available.

1. Meadowlands Adaptive Signal System for Traffic Reduction(MASSTR) [2]
    a. Uses cameras, controlling software and wireless and fiber optics communication
    b. Has high installation cost
    c. Timing is determined by the flow of traffic
2. InSync adaptive traffic control system [3]
    a. Works with existing traffic control cabinets and controllers
    b. Has two hardware components, IP video cameras and a processor
    c. Low cost compared to MASSTR
3. Split Cycle Offset Optimization Technique (SCOOT) [4]
    a. Uses other vehicle detectors other than CCTV cameras
    b. Give less priority for pedestrians and bicyclist
    c. Use online computer
    d. Wireless communication
    e. Expensive to install and maintain
4. Sydney Coordinated Adaptive Traffic System (SCATS) [5]
    a. Use detectors such as inductive loops in the road pavement for vehicle and push button to pedestrians
    b. Inputs can be not accurate as system uses electronic sensors
    c. Highly expensive to install and maintain

In most of the available systems their main concern is the traffic flow and give less priority for the pedestrians. When generating signals to be sent to the signal lights most of the systems just go through very simple logic without much care for optimizing the signal decision. Also to reduce the traffic congestion effectively traffic controlling systems need to consider not only the traffic flow but also the types of vehicles. For an example it takes significant amount of time for a long-vehicle to accelerate than a car or van. So if there are long vehicles better let them pass. But in the meanwhile it is not practical to keep the pedestrians wait for a long time as well because then the pedestrians will tend to violate red light conditions that can result even in catastrophic scenarios worsening the traffic congestion.

Considering all the above facts and with the guidance and initial idea of Dr. Chathura de Silva, we came up with the project idea to develop a Vision Based Pedestrian Crossing Controller that can be extended further into a complete Vision Based Adaptive Traffic Controlling System with little more extra effort . In the following section we describe our solution for the above problem specification in an abstract level.

1.3 Our Solution

Proposed Vision Based Pedestrian Crossing Controller is a software system that will basically detect the pedestrians waiting to cross the road and analyze the traffic flow and the types of vehicles and come-up with optimum traffic light signals to be sent to the traffic light controller. This will happen continuously on real-time ensuring the efficient and adaptive traffic and pedestrian controlling.

One 360-CCTV camera with a fish-eye lens will be mounted at the pedestrian crossing as required and the decoded video footages coming from the CCTV camera will be analyzed and the following information will be extracted on real-time.

1. No of Pedestrians waiting to cross the road on both sides and pedestrian movements on the crossing.
2. Vehicle movements in terms of average speed of arrival and the types of vehicles (i.e. motorcycle, large vehicles, small vehicles etc.)
3. Vehicles and pedestrians violating the red light condition & other traffic rules such as overtaking on pedestrian crossing.(optional)

Then based on the observed parameters a small AI which is specifically designed to generate controlling signals will generate controlling signals to be sent the traffic light controller. In the meantime statistical data will be sent to a central server for further use and for persisting them.

Identified vehicles' details violating the red-light condition & overtaking on pedestrian crossing will also be sent to the central system with the timestamps.

When the green-light is on for the traffic and red-light for the pedestrians the software system will analyze the number of pedestrians waiting to cross the road, the traffic flow, and the types of vehicles in the traffic flow and make decisions based on those parameters. While the red-light is on for traffic and green-light is on for pedestrians the system will analyze the pedestrian movements on the crossing and take decisions based on those parameters. In both cases the system will simultaneously analyze the video footage for vehicles and pedestrians violating the red-light conditions. Additionally it will check for overtaking of vehicles on the crossing.

1.3.1 Methodology

There will be four main components of the Vision Based Pedestrian Crossing Controller software system.

I. Vehicle detection and classification package
II. Pedestrian detection package
III. Traffic rule violation detection package (optional)
IV. Decision making and controlling package

Vehicle Detection and Classification Package will pass the detected average speed of the traffic flow and the vehicle types to the Decision Making and Controlling Package. With parallel to that Pedestrian Detection Package will pass either the number of pedestrians waiting to cross the road or the pedestrian movements on the crossing. Simultaneously Traffic Rule Violation Detection Package will pass identified vehicles' details that violate traffic rules and the timestamps indicating the time of the occurrence of the violation.

Decision Making and Controlling Package will generate the optimal adaptive signals to be sent to the traffic light controller. If there are any information regarding traffic rule violation the controlling package will send them to a central server. Statistical data received by the controlling package will also be persisted in the same central system. The following diagram shows a graphical representation of our solution.

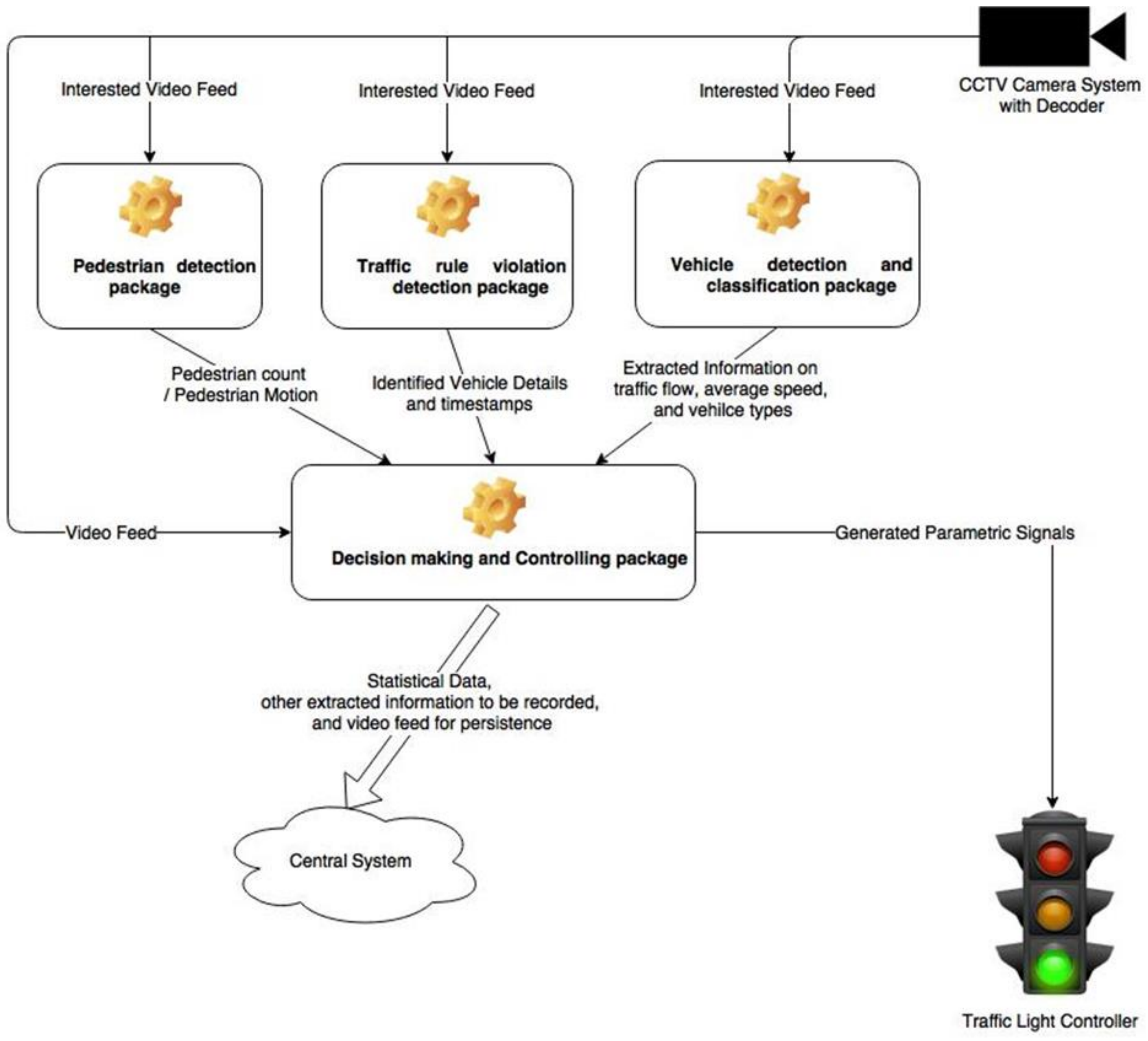

CCTV Camera System with Decoder
Interested Video Feed
Interested Video Feed
Interested Video Feed
Pedestrian detection package
Traffic rule violation detection package
Vehicle detection and classification package
Pedestrian count / Pedestrian Motion
Identified Vehicle Details and timestamps
Extracted Information on traffic flow, average speed, and vehilce types
Video Feed
Decision making and Controlling package
Generated Parametric Signals
Statistical Data, other extracted information to be recorded, and video feed for persistence
Central System
Traffic Light Controller

Figure 1: Methodology Block Diagram

Since we are interested in delivering a software system for Vision Based Pedestrian Crossing Controller the key research areas for this project are;

1. Vision based pedestrian detection
2. Vision based pedestrian tracking
3. Vision based vehicle detection
4. Vision based vehicle Tracking

Thus we have surveyed and analyzed the already available literature on these categories. In the next chapter we have presented the literature we have analyzed under above categories.

2 Literature Review

2.1 Pedestrian Detection (Human body detection in road environments)

The literature for detecting human bodies can be divided into two main categories. They are methods that uses a motion based pre-processing technique to get some identifiable contours that can be matched with human contours and direct detection methods that are most of the times, based on image segmentation methodologies that can detect humans directly from the images based on various feature extraction methodologies like shape, colour, or even sometimes motion in order to detect humans.

The common methodology used to detect image features using the relative motion of objects is background subtraction. Usually in background subtraction methods, the main focus is on extracting the foreground objects from the background of an image and detect or classify the relevant objects like humans, animals, vehicles into categories. So the methods are based on the motion of the image sequence. They cannot detect nonmoving objects in an image sequence. The relative motion of detectable objects in an image sequence is a main requirement in order to use a method like background subtraction.

Direct techniques operate on features extracted from image or video patches and classify them as human or non-human. In these methods various techniques are used to classify a given input as human or not, based on the features detected from images. These features include shape in the form of contours or other descriptors, skin color detection, motion, or combinations of these. These techniques can further be categorized into part based, patch based, and holistic methodologies based on the detection mechanism used.

In our project scope, the requirements are as follows.

1. Pedestrians count should give a positive value if the pedestrians are present. (the count needs not to be highly accurate in densely crowded environments)
2. If no pedestrian is present, it should not indicate as pedestrians being present. (no false positives in zero pedestrian occasions)
3. Road environment can be anything and the background structure may vary depending on the targeted place.
4. Typical Sri Lankan weather environment is expected (Rainy/Sunny/Dusky/Night)
5. Pedestrians may have different clothing styles but have an average shape model.

So considering all the above requirements, following sub-topics on pedestrian detection methodologies are evaluated by going through relevant literature sources.

2.1.1 Motion Based Detection

In this method, the essential feature is to detect the moving objects. Often the detections are based on blobs and etc., which causes problems like presence of shadows and difficulty in distinguishing and separating people.

Since the background subtraction method is a motion based methodology, pedestrian detection at a pedestrian controller, is hindered by some constraints. The pedestrians are normally getting gathered around a particular area and they tend to wait in an approximately still posture until the lights show the signal to cross the road. In that case the motion is very difficult to detect. Even though the small motions are detected in a small scale, it is barely sufficient to map to a certain shape or do any comparisons like in systems suggested by Maojun [27].

The system should detect pedestrians in an open area where all kinds of disturbances can possibly occur. For example lighting and various factors can change in a typical road environment. Toyama et al. [19] describes the process and complications of background subtraction well. Those complications can be summarized as below as they are relevant to this project.

1. Moved Objects: Objects that are supposed to be considered part of the background might change its position.
2. Time of Day: Depending on the time of the day (day/night/dusk/rainy & etc.) lighting conditions may vary.
3. Light Switch: Illumination of the objects in the background scene can change due to various reasons.
4. Waving trees: This is a common scenario in the backgrounds of the road.
5. Motionless person: A foreground object that becomes motionless cannot be distinguished from background objects (This is an important scenario in this project).
6. Shadows: Depending on the time of the day, the shadows can vary and that might give some concatenated blob structures in the filtered model.

Stauffer and Grimson [20] present a solution for formulating a mixture model. Lee [21] claims that this model has become standard for the mixture model approach. However the solution does not completely address all the complications. So there are still no universal solutions to the complex problem of background subtraction.

The background objects can change with the time. So a static learning methodology of the background is not suitable for a dynamic environment like a road environment. As suggested by Baisheng[28] the background must be statically learned initially in order to detect foreground objects. Using this approach to train the model, the training effort can be reduced and detection cost can be reduced as well. When a person enters the area of interest, the reference background scene has to be matched with the current frame and anything that moves are segmented easily as foreground objects. This way, computing overhead can be reduced. But in a road environment, this is hardly applicable as the environment can change with the time. Another reason against this approach is that it is difficult to record the particular environment under a certain time where there are no moving objects and only the background objects are present. Another approach can be used to detect the number of people coming into the area of interest and going out of that area and take the difference as the count to be waiting. The pedestrians that are moving along the zebra crossing can be detected using an overhead camera as proposed by Alessandra Fascioli, Rean Isabella Fedriga and Stefano Ghidoni [22].

Since our proposed system is performance critical, systems like Maojun, Z et al [27] cannot be adopted to our project. Our system should entirely run on a simple hardware. So a high performing hardware is not desirable and not cost effective either.

Another issue is that the camera angle needs to be carefully decided. Detection can be occluded by the vehicles coming in between. That totally depends on the angle and the positioning of the CCTV camera. If a background subtraction method is to be used, the best approach is to position the camera in a way such that effect of other irrelevant moving objects are minimal. If vehicles are in an area of disturbance, then the background subtraction method consumes more computations to deliver results at the same time causing a reduction in the true positive rate and increasing false positives. Since the capturing is done by a wide angle camera [23], the interested areas can be decoded and analyzed. Anyhow it is better to avoid the noise from the vehicle movement.

Following are some of the brief reviews in relevant literature in the area of human detection based on background subtraction methodologies.

In order to detect people in complex scenes, backgrounds, Maojun, Z et al [27] proposes a method which creates a panoramic image from the background before a person enters. Whenever someone enters the region of interest, the process of background subtraction starts. Both the captured image and the constructed background image are compared to detect foreground objects (human bodies) using the background subtraction algorithm which is based on logarithmic intensities. In order to run this on a personal computer in real time, experiments show that it has to have a high performance capability. [27]

In Baisheng, C. et al. [28], it proposes a background model initiation and maintenance algorithm for video surveillance. It uses a frequency ratio mechanism to model the background. First the initial background is statically learned using the frequency of the pixel intensity values during training period. The frequency ratios of the intensity values for each pixel at the same position in the frames are calculated; the intensity values with the biggest ratios are incorporated to model the background scene. Then secondly, background maintenance model is proposed to adapt to the changes in the scene. The scene changes are due to the environmental changes and the moving objects transforming into still objects in the scene. Examples like the changes in the lighting condition, sun being blocked by the clouds and etc. can be taken as possible environmental changes. There also can be extraneous events like a moving person decides to wait for some time or may be a person parking a car and etc. So then after a three stage process is carried out in order to detect the objects in the foreground. They are thresholding, noise clearing and shadow removal. Their algorithm is proved to perform in the real time conditions and it's also proven to be robust for such conditions.

Some systems also have been developed for anomalies detection and in that context, human body detection has become a crucial area that has been focused on that domain. In M et al. [13] a video surveillance system for outdoor environments is proposed and human detection is performed using background subtraction methodology. He proposes a video surveillance system for outdoor environments, in this case the context of a park, which can highlight objects of interest and recognize behavior like theft and violence. The focus is on detecting the human figures in the outdoor scenes using a static CCTV camera. It is an example based learning technique to detect people in the dynamic scenes. The classification is purely based on shapes of people and not on the image content. An adaptive background subtraction approach has been used for detecting the objects of interest. The shapes of the contours are represented by geometrical information extracted from images in horizontal or vertical directions. Then patterns are defined and feature vectors are classified using a neural network of three layers. The specialty of this approach is that person figures are detected individually as well as in crowded environments. It is also capable of detecting humans within a dynamic scene where moving objects are not only the human figures but also other types of figures are involved. But when the blob retrieved by background subtraction which actually belongs to a human shape, is somewhat connected with another blob of a non-human moving object like a moving car, then the blob is not properly classified as a human shape. The reason is that the blob is highly modified and it is far more different than a human shaped blob. In that case the binary shape is considered modified and they were not detected. Since such scenarios are even sometimes hard for human eyes to detect, those scenarios can be neglected. This method can be concluded as a fast, reliable, and robust method. Another important point is that it can be easily adapted to be able to detect other moving objects as well. This method has a potential of serving in the area of anomalies detection largely.

Another real time human detection system was proposed by Wren et al. [14]. The background model uses a Gaussian distribution in the YUV colour space at each pixel and it is continually updated. In this approach also blobs are used to model human figures and also the spatial and colour components are used. Kalman filter is used to estimate the spatial parameter as they are constant with the dynamically changing blob. The probability of a pixel being part of the background scene or the blob, is evaluated next. Each pixel is then assigned to either to be in the background or the blob after doing some morphological filtering. After this, statically those models for backgrounds and person figures are updated and person blobs are initialized using a contour detection methodology in which the head, hands and feet are located. In that operation, skin colour feature is also used to initialize hand and face. This is more focused towards finding single human figures and also makes several domain specific assumptions. This is also targeted towards the real time detection scenarios.

2.1.2 Direct Detection

Direct detection varies from the background subtraction methods as in these methods the feature extraction of a particular model is performed instead of trying to detect foreground objects. Various features can be used to detect human figures. We can categorize direct detection approaches as holistic detection, patch detection, part based detection.

2.1.2.1 Holistic Detection

In this approach, the system scans a whole frame captured by a video footage in order to detect an object of interest. The detector fires the image features if the image features inside a particular local search window meets a certain criteria like features of a human shape. The methodologies are of two types. One type uses global features like edge template. Some use local features like HOG descriptors. The performance can get affected easily with various occlusions and clutter in the background.

There are some major drawbacks in holistic detectors like HOG detector. As G.Thomas Prathiba and Y.R.Packia Dhas [24] have described in their approach, the main drawbacks are as follows.

"We have shown that using locally normalized histogram of gradient orientations features similar to SIFT descriptors in a dense overlapping grid gives very good results for person detection, reducing false positive rates by more than an order of magnitude relative to the best Haar wavelet based detector. Histograms of oriented Gradients may achieve more accurate counting and detection results when the crowd is small. Disadvantages of this process are high time consumption, shows only results for a small crowd with few occlusions, and it also required high resolution images."

Since the significant reductions of the false positive rate, it seems very suitable for our application of pedestrian detection. In the road environment, the system should not give false positives. I.e. if no pedestrian is present at the scene, it should not detect any other object as a pedestrian. So in that case the system should have a very low rate of false positives. But note that the crowd size should be small in order to achieve good results. In our case the crowd size can be dense in a small area because in a peak time, there can be many pedestrians gathering around a corner to cross the road. So it is a main requirement to be able to capture the pedestrians in a densely crowded scene. Another disadvantage is that it requires a significant time to compute. Since this is a real time environment, the detections should be made in a very fast pace which stands as a main objective in our way. Occlusions also can occur and the above mentioned system shows low performance in the presence of occlusions.

The system suggested by Papageorgiou, C. and T. Poggio.Trainable [15] which is based on an edge template pattern matching, they use a dictionary of Haar-like wavelets. Through that their system is able to identify important characteristics of human body class by ignoring unnecessary noise that can be seen in the pixel level representations. Using a trained SVM by a large number of training samples both representing human and non-human examples, the system detect human figures. Their core detection system implements a brute force approach where it is not required to capture anything from the relative motion of the objects like in background subtraction methodologies. It does not make any assumptions on the scene structure or the number of people in the scene, or not even the movement of the camera. They directly apply this approach to video frame sequences and it helps ignoring the dynamic information significantly. They use two-rectangular based Haar-like features that can detect the edge features and as a deviation of normal feature calculation, they use an over complete (an overlapped version of the rectangles) in order to calculate the Haar-like features. Then the template feature vector which is of 1326 dimensions, are classified using a SVM. The detection is done using a detection window of size 128x64, shifted all over the image. In order to detect the pedestrians in a multi-scale, they do resizing of the images and runs the detection window over each and every resized image. Since this is a brute force search mechanism, the system can be very slow.

A main advantage of the system suggested by Papageorgiou, C. and T. Poggio.Trainable [15] is that it does not require any pre-training of the background just as many of other holistic detector. Since it uses a simple brute force approach it needs not to make any assumptions on the structure or number of people in the scene, or the movement of the camera. As in an open environment camera might shake a little with the wind and in this approach that can be ignored. One disadvantage of this approach is, the need of a large dataset. In Sri Lankan environment, people wear diverse set of clothing and due to that the training sample size can get even larger because this considers some edge features using 2 rectangular Haar-like features for formation of feature vector. Use of a brute force searching approach always will hinder the performance. So as for our targeted environment, it is not very promising as a real time pedestrian detector.

The HOG features are utilized by the system suggested by N. Dalal, B. Triggs [16] which uses the local features. The HOG descriptor that they use has a few key advantages over other descriptors as it operates on local cells. It is invariant to geometric and photometric transformations. But for object orientations, it gives a variant feature descriptor. As Dalal and Triggs suggests, by using various sampling methods, the movement of the body of pedestrians can be ignored if the pedestrians are assumed to maintain an upright standing position which almost in every case can be considered to be true. The next step of calculation is creating the cell histograms. An orientation based histogram is formed with each pixel in each cell giving a weighted vote as contribution towards it. Those weights are summed into a histogram channel based on the values found in the gradient computation. The orientation of the channel can vary from 0-360 degrees or 0-180 degrees depending on the values that are considered; whether signed or unsigned gradient values. They suggest use of unsigned gradients in conjunction with 9 histogram channels as that particular set of parameters gives the best performance in the human detection tasks based on their experiments. As for the gradient calculation, even though it is possible to consider some function of the pixel gradient, they suggest using the pure gradient value itself as it gives good results. As the next step the descriptors are generated using those cell histograms after performing normalizations. The block based approach is used for this task and a block is defined as a concatenation of individual cells over a rectangular region (Circular block descriptors are also tested in their experiments). The parameters that they suggest are, 8x8 cell blocks of 16x16 pixel cells with 9 histogram channels. After performing block normalization, the values are trained using a support vector machine (SVM). Using supervised learning methodology, the human and non-human figures can be identified.

As N. Dalal, B. Triggs [16] the system is invariant to geometric and photometric transformations which is a good feature for the current situation. A drawback of this system is that it is variant to the object orientations. That may not be a problem in our project's context. If we can assume that the pedestrians are standing facing towards the road, then it provides a stable environment for this detector to perform well. Another advantage is that the movement invariant behavior of the detector assuming that pedestrians are standing in the upright position. It can be identified as a fair assumption as everyone is on their feet and no one would change the standing posture during the time waiting for the signal lights.

2.1.2.2 Patch Based Detection

This method is a combination of detection and segmentation. Different local features are focused and each match over a local hypothesis casts one vote for the global hypothesis.

System suggested in Leibe et al. [17] is based on a visual vocabulary of small object parts and aggregates evidence from local image descriptors. The training is performed by extracting image patches of a predetermined scale. The patches are rescaled and grouped under an agglomerative clustering scheme where the resulting clusters form a compact representation of the local object structure. So the cluster centers are stored as the codebook entries. The learning is on the spatial occurrence distribution of each codebook entry on the object category. Due to the variety of pedestrian shapes and appearances observed in images, the aggregation of local information alone is often not discriminative enough. They therefore propose algorithms, which combines local with global information. In that system, they tested their results on a very diverse dataset where all kinds of occlusions can be found. Other than the detection, the system is also capable of getting an exact count of the number of pedestrians. Furthermore they also have gone to an extent where they could even tell the precise locations of each and every pedestrian. The specialty of the system was that it could detect a pedestrian figure when the figure is extremely occluded. Maybe sometimes only a small part of the body is displayed. Since an exact count is taken, the results on each and every hypothesis is taken into account while ignoring the additional detections on the same hypothesis as false positives. The learning is focused on learning a codebook of local appearance. When detecting, the extracted local features are matched against a set of codebook entries and if the results are positive the relevant hypothesis is cast one vote over the main hypothesis. The advantage of this approach is only a small number of training images are required. For the training data they recorded 44 sequences of 35 different people walking parallel to the camera image plane in front of two different backgrounds.

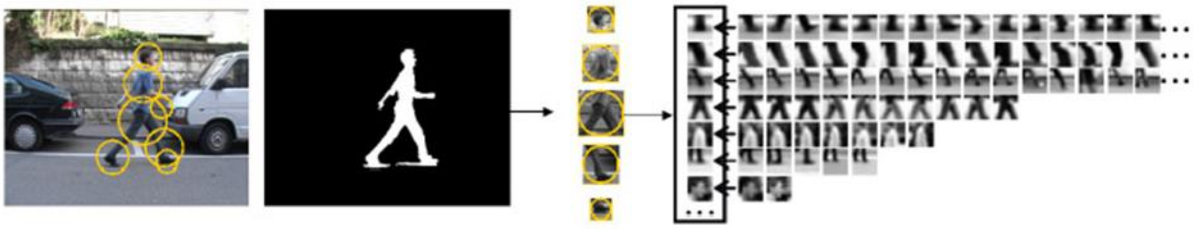


Figure 2: Codebook Generation

In the system from Leibe [17], they highlight that the system is tested with samples with high amount of occlusions present in the image set. So the system can be identified as a robust one. Another main advantage is that it gives the precise count and the precise location of the pedestrian. Precise location detection is not a requirement in our context but the precise count is elegant even though it is not crucial as long as it does not give false positives when pedestrians are not present. As mentioned above, the ability to detect pedestrians in very occluded situations like in a dense crowd, which is same as our context of detection, is an outstanding feature. In a crowded environment sometimes only a small part of the body may be displayed and if the system can detect that, the precision of the count is very reliable. On the other hand the size of the training dataset is very small. Since it does not require any background structure pre-configuration, it stands as one of the most suitable methods among the various other methods mentioned.

As a whole, patch based systems seem to be more suitable than other mechanisms as they consider the local features rather than global features. Since the local features themselves do not show a good precision, the global features are also combined and the combination performs well in terms of accuracy and robustness.

2.1.2.3 Part Based Detection

Basically in a crowd, detecting full human bodies can be a problem as the body gets occluded by other human figures and might miss some of the figures and also there is a possibility of a high false positive rate. So detecting parts like heads [25] would be a solution. The part based approaches focus on detecting a single or multiple parts of the human body specifically to verify the presence of a human figure. So the pedestrians are modeled as a collection of parts. Though this approach seems very promising, part detection itself can be a difficult task.

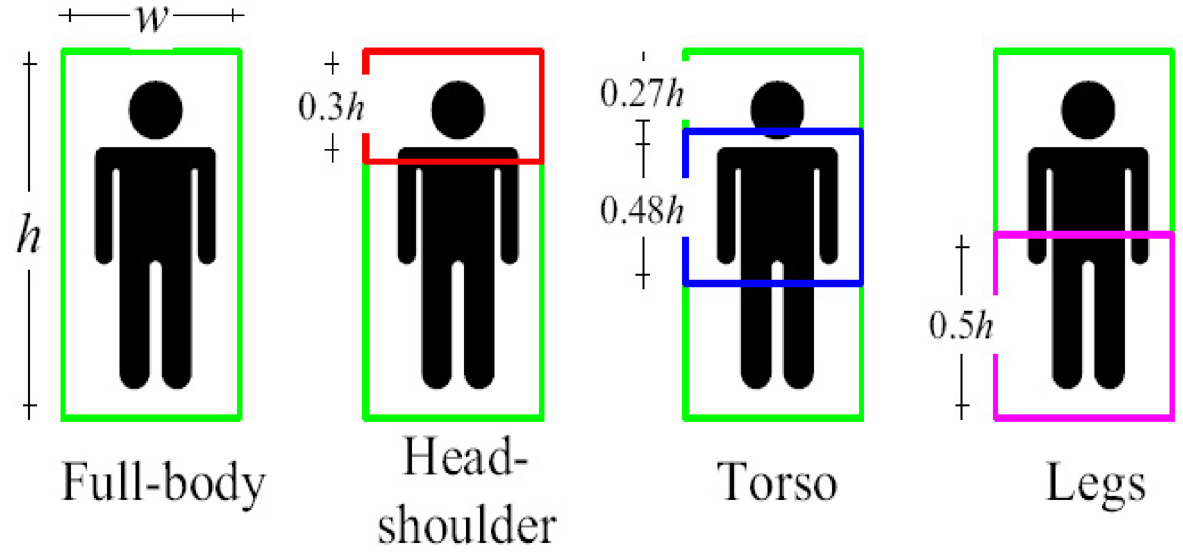


Figure 3: Definition of body parts.

Wu and Nevatia [18] present a people detection system crafted towards a surveillance scenario. It learns multiple part detectors for full body, head-shoulder, torso and legs. The feature used for detection is called edgelet which is a short line or curve segment, which encodes both the edge intensity and orientation. The shape of a person is modeled as an ellipse. This improves occlusion as the ellipse is tighter than a commonly used bounding box. The authors have used front and back views of pedestrians to model the figures. Then they show that the edgelet features perform better than Haar-wavelet features. The results show that for crowded scenes, the performance of full body and leg detectors decreases significantly where lower body is occluded. One advantage is that the detection rate decreases slowly with the degree of occlusion.

In our context, the pedestrians only need to be modeled from one side. As a convention the same camera view is supposed to use in every location to be implemented. Since the system by, Wu and Nevatia [18] focuses on the edgelet concept it is capable of detecting an occluded figure in a crowd. As a highlight, the considerations on all three aspects, edges, intensity and orientation can be noted as important approaches taken towards the reduction of false positive rate. The precision is acceptable as it gives better performance than Haar-like features. But a main disadvantage is that in a scene where pedestrians are standing in front of the camera, the crowd can be very dense such that lower body parts of pedestrians who are in the middle of the crowd can be occluded. So the system [18] performs poor in such scenarios. But on the other hand, the rate of reduction of the detection precision is low compared with the rate of increase of the degree of occlusion which results in an acceptable level of detection. In the targeted context, if the crowd is very much dense, the detection needs not to be very precise as a threshold value can be defined for large number of count values as either way crowd being larger than a particular count value would impose the same level of urgency as same as a crowd with a count at the threshold value. So the time allowed for the vehicles on the road would be fixed for such situations where a large crowd is waiting to cross the road.

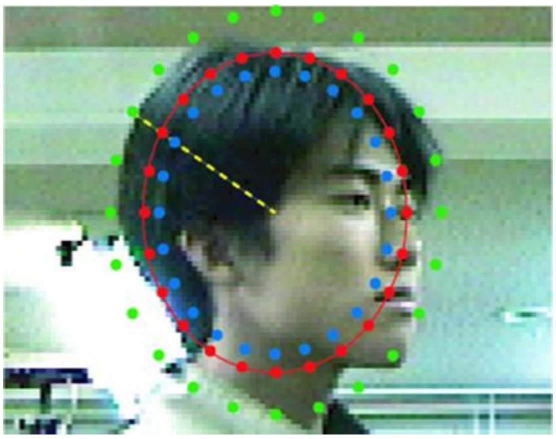


Figure 4: Modeling a human head as an ellipse.

A human head (overall head with the facial components) detection methodology is suggested by Akihiro Sugimoto, Mitsuhiro Kimura and Takashi Matsuyama [25]. Their method is to detect the human head as ellipses in the scene using a camera targeted towards the crowd from a certain distance using a predetermined angle (depression angle) and then run various algorithms for detected ellipse contours in order to verify whether the detected ellipse regions are human heads. The detection is based on the change in the gradient magnitude of intensities along the ellipse perimeters.

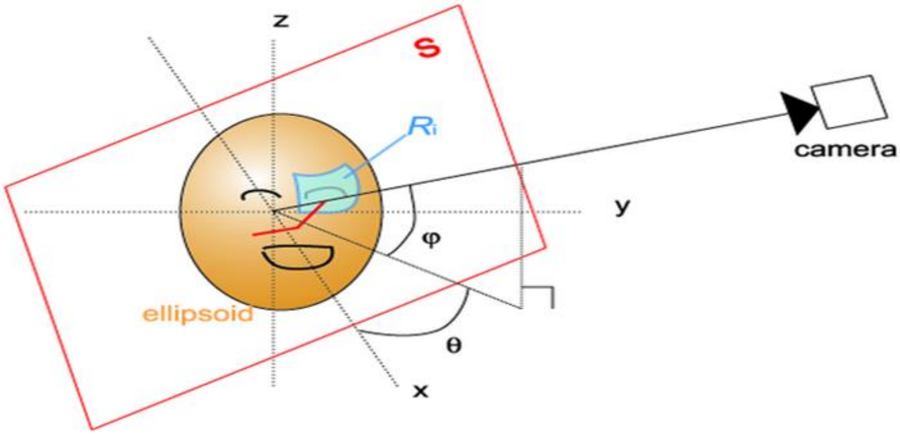


Figure 5: Detecting a head using ellipse concept

They [25] have used a database of 300 images, each person having each and every orientation of the face. People were ranging uniformly from age 15 to 65. They recorded the recognition accuracy varying the depression angle of the camera against all expected face orientations. They also have run the process for all the image in the database (9600 images). For low depression angles the results showed a good accuracy rate even for face orientations with large angles like 330 degrees (75-85%). But when the depression angle got increased the accuracy have fallen down drastically around 15-30%. The reason for relatively low accuracy for high orientation angles is because the lack of exposure of facial components. Only one cheek is shown for like 270 degree angles and the detection has become unstable for them. Also that accuracy becomes higher as the angle of depression of the camera becomes smaller. The small angle of depression of the camera means that the face is captured from the horizontal direction of the face and that the facial components clearly appear in the image. So for a good accuracy the facial components had to be clearly visible. This model is limited in aspects for angle changes. But in case of a pedestrian detection scenario for our context, pedestrians will most of the time stand facing the camera and the camera can be placed on the other side of the road which would give a considerable distance between the man and the camera so that it is easy to get near horizontal trajectories for angle of depression. For this method it is required to fix the camera at a distance as in our case on the other side of the road focusing the pedestrians standing on the opposite side. As a cost reduction approach, we use a single fish eye lens camera for detection in which case the camera is placed over the top with a high depression angle, for which this system performs poorly as with the difficulty of capturing the facial components.

The strengths of the system [25] can be stated as follows.

1. Reduction of the false positives by selecting ellipses.
2. Reduces the computational power required due to pre-selection of ellipses.
3. Good accuracy for horizontal angles.(suitable for pedestrian detection at crossings)
4. Use of Gabor Wavelet filter which is robust against illumination changes.
5. Ability to derive facial component using natural component configuration.
6. Can detect different orientations with good accuracy.

So the system [25] is suitable for our scenario in every aspect except for the issue with the high camera angle.

As in part based systems, since they focus on particular body parts of the human body, we have an opportunity to select the most suitable body element as for the scenario. In our context detecting head can be a very elegant approach because all the pedestrian heads are being exposed and it has a small chance of being occluded and those occlusions can be ignored considering the requirements of the project. But a problem might arise as when people use hats, caps or on a rainy day when people largely use umbrellas as in such situations head area is occluded partially or entirely.

Head and shoulder areas also can be used for detection as suggested in “Human Detection using HOG Features of Head and Shoulder Based on Depth Map” [26]. But the same issues of occlusions might occur as the camera is placed over the head.

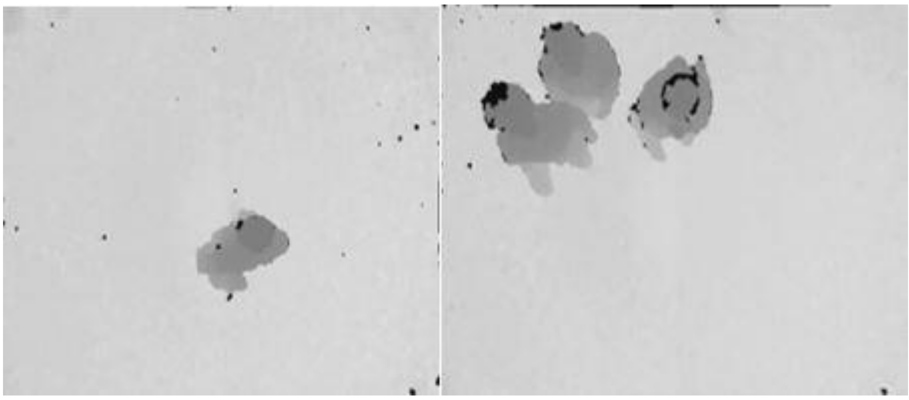


Figure 6: Detecting human heads and shoulders from an overhead

So in part based detectors the detection process is efficient but placing the camera to capture the image with minimum occlusions and with maximum desirable parts is a major challenge.

2.2 Vehicle Detection, Classification and Tracking

When developing an adaptive vision based pedestrian crossing controlling system, vehicle detection system is a major part. So for a vehicle detection algorithm which is crucial and performance critical, we have to make sure that it meets optimal level of efficiency, accuracy, and robustness. In our research on vehicle detection and classification using surveillance cameras (a sub part of the vision based pedestrian crossing controlling system) we figured it out that many researchers have investigated on implementing intelligent traffic surveillance which can be used in applications like intelligent transportation systems, transportation planning, traffic engineering and so on. These kind of applications extract information by analyzing surveillance video streams using image processing techniques. Some of these information are vehicle count, vehicle trajectory, vehicle tracking, vehicle flow, vehicle classification, traffic density, vehicle velocity, traffic lane changes, license plate recognition, and etc. Throughout the past two decades researchers have succeeded in applying image processing techniques such as vehicle detection, classification, and tracking in real world applications such as:

- Automated toll levy systems
- Classify vehicles on highways
- Automated traffic lane violations
- Automated license plate detectors

In our research on vehicle detection and classification, we found out three step methodology to implement vehicle detection and classification system. They are as follows:

1. Motion Vehicle Detection and Segmentation Approaches
2. Camera Calibration Approaches
3. Vehicle Tracking Approaches

Later in this chapter we discuss these three approaches in details. Though these kind of approaches have invented and improved by researchers, other factors such as background obstacle, weather conditions, and etc. affect the performance of these approaches.

2.2.1 Motion Based Vehicle Detection and Segmentation Approaches

Due to dynamic nature of the traffic surveillance, analyzing traffic surveillance, detecting, and segmenting moving vehicle still a tough task. One of the best approaches taken by researchers to analysis traffic surveillance is extracting moving section of the video stream. Analyzing two images of video stream which is taken at different intervals is the most popular method of computer vision. This method is widely used in areas like video surveillance, medical diagnosis and treatment, underwater sensing and so on. These kind of moving object detection techniques are useful in vehicle detection as traffic is moving most of the time. Another problem arises in vehicle detection and classification is moving of traffic becomes slow and vehicles overlap on each other due to heavy traffic conditions. These problems can be reduced by using feature based approaches which uses some unique features of a vehicle to detect it.

In our research on vehicle detection and classification, we figured out three methods which can be used in traffic surveillance analysis.

1. Background Subtraction Methods.
2. Feature Based Methods.
3. Frame Differencing and Motion Based methods.

2.2.1.1 Background Subtraction Method

Methods which are involved in background subtraction are widely used in computer vision. The basic concept of background subtraction is extracting moving object or blob (foreground of input image) from the static image (background of the input image). In practical situations we cannot always give static image of the background as we cannot define the initial stage of a dynamic environment. In these situations background image can be generated by analyzing series of images from the video stream.

An upgrade to background subtraction method is use of statistical and parametric based techniques to extract foreground from the background. One of these techniques that has been utilized by researchers is Gaussian probability distribution model which assign value for each pixel according to distribution. For pixel in point (x, y) in image, the algorithm analyses the same point (x, y) for series of images to build up the Gaussian probability model. Then this method uses knowledge of the probability model to categorize the same pixel at (x, y) point in new image as foreground or background with the use of following equation.

$$I(x,y) - Mean(x,y) < C * Std(x,y)$$

I = intensity of pixel (x, y), Mean(x, y) = Mean intensity of pixel (x, y), C = constant, Std (x, y) = Standard deviation of intensity of pixel (x, y). By using statistical techniques in background subtraction can beneficial in order to detect moving objects in some limited dynamic environments (background need not to be completely stationary). In our project, detecting vehicles to be done in dynamic environment such as dynamic weather conditions and different times of the day in which light conditions can change. So in non-controlled environments like traffic surveillance, using this statistical approaches can be more beneficial. [31][32]

2.2.1.2 Feature Based Method

Analyzing traffic surveillance become somewhat harder problem due to dynamically changing environment. When implementing an adaptive traffic controlling system, detecting vehicles in a heavy traffic condition is critical. In situations like vehicle overlapping and heavy traffic conditions, methods like background subtraction can give incorrect results. So the researchers try to move towards feature based methods which will analyze sub-features like the edges and corners of vehicles. These methods segment moving objects from background by collecting and analyzing the set of these features from the movement visible in the subsequent frames. Feature based methods show significant improvement in analyzing computationally difficult views and efficiency compared to background subtraction methods. [33]

A research group [34] proposed a new feature based method for vehicle detection in low resolution aerial imagery. This approach uses scale–invariant feature transform (SIFT) to extract key points in the images. A trained SVM is used to predict whether the set of SIFT key points belongs to vehicle structure or not in the image. The SIFT is a graphical operations like translation, rotation, scaling, and enlightenment which is simply pretentious by a noise and small distortion. Finally, Affinity Propagation (AP) algorithm used to separate clustering of SIFT key points in the images which is already classified by SVM. Since our vision based pedestrian crossing controlling system is performance critical system, the delay incurred by noise and distortion can be more crucial.

2.2.1.2.1 Detection of Vehicles using Haar-like Features

The method which uses Haar-like feature to detect objects was introduced by Viola and Jones in their research to detect pedestrians using pattern of motion and appearance. The process requires representative data sets to be used for training and validation including positive (presence of objects to detect) and negative (absence of objects to detect) image samples. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums[35][36]. The general idea of Viola and Jones is to describe an object as a cascade of simple feature classifiers organized into several stages.

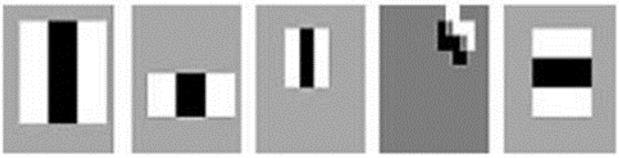


Figure 7: First 5 Haar-like features used in Viola and Jones method

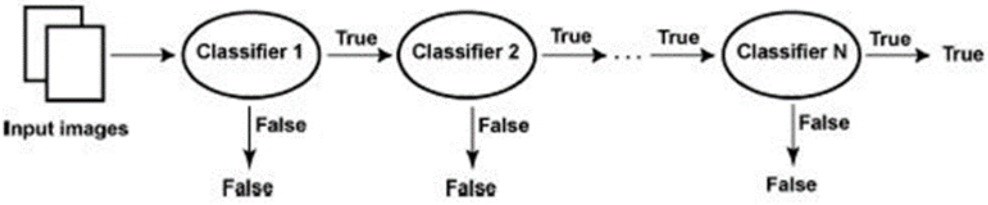


Figure 8: Cascade Classifier Architecture

Due to the performance, both speed and accuracy shown in pedestrian detection researchers tried to use cascade classifier for vehicle detection [37]. They have trained the cascade with extra Haar-like features so that the accuracy become high.

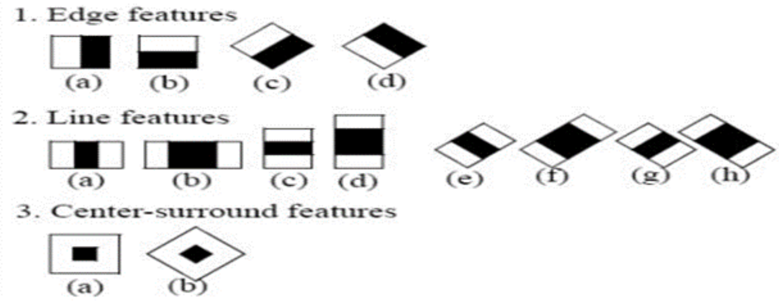


Figure 9: Haar-like Features used in Vehicle Detection

When designing a traffic light controlling system the speed of detecting vehicles is crucial. So using a process with steps to detect vehicles in such a system will increase the performance of the system.

2.2.1.3 Frame Differencing and Motion Based Method

Frame differencing is a similar process to background subtraction which is the process of subtracting two subsequent frames in an image sequence to segment the foreground object (moving object) from the background frame. But the problem in background subtraction is, it needs a method to construct the background properly. Otherwise background subtraction cannot detect foreground accurately. The frame differencing method give best solution to different illumination levels which occurred throughout the day. But frame differencing method does not work well if the time interval between the frames being subtracted is too large [38] [39]. Most of the vision based traffic controlling systems unable to detect object due to change in the lightning condition. But in these methods the difference between lightning conditions of two consecutive frames is negligible.

2.2.2 Camera Calibration Approaches

Camera calibration is an essential step in a vision-based vehicle detection system. Measuring speed and size of a vehicle cannot be done without proper camera calibration. In our pedestrian crossing system we have to measure vehicle speeds in order to give traffic signals. Otherwise vehicle will not be able to stop at pedestrian crossing if vehicle is coming at high speed. And other important factor is detecting vehicle size. It allow system to take decisions correctly according to the size of vehicle because large vehicles result in more traffic congestion than small vehicles. Camera calibration involves in mapping of points in the real world onto the image plane.

A group of researchers who have worked on automatic traffic monitoring systems suggested to use 2D spatio-temporal images [1]. This system uses a TV camera to keep track of vehicles for the highway traffic within two slice windows for each traffic lane. The purpose of this system was classifying the passing vehicles by using these 3D measurements (height, width and length) in addition to counting the vehicles and assessing their speeds. This system has showed efficient and robust performance in different lightning conditions.

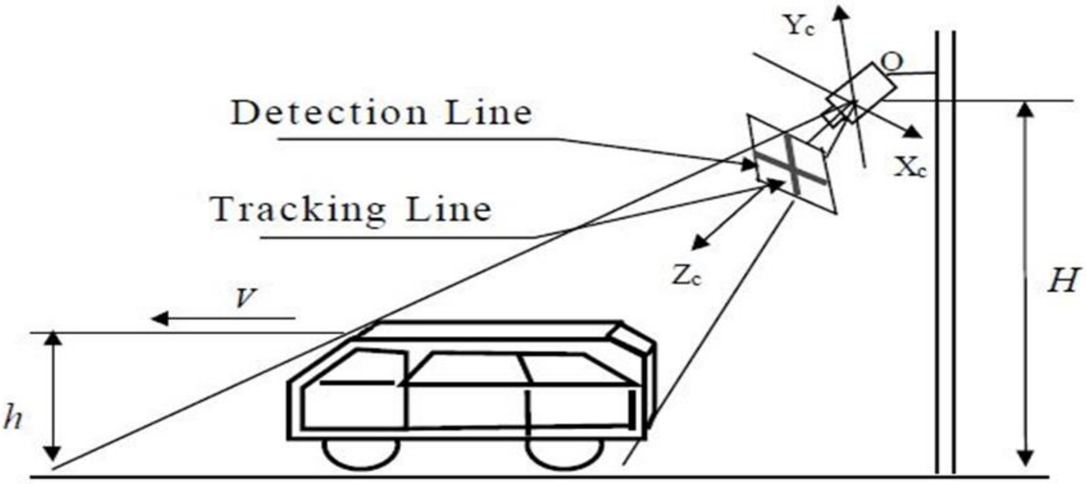


Figure 10: Camera Setting

2.2.3 Vehicle Tracking Approaches

Tracking moving vehicles in video streams has been an active area of research in computer vision. The object tracking using video processing is an important in visual-based surveillance systems and represents a challenging task. To track the physical appearance of moving objects such as the vehicles and identify it in dynamic scene, it has to locate the position, estimate the motion of these blobs and follow these movements between two consecutive frames in video scene. Several vehicle tracking methods have been illustrated and proposed by several researchers for different issues. These methods are:

1. Region-Based Tracking Methods
2. Contour Tracking Methods
3. 3D Model-Based Tracking Methods
4. Feature-Based Tracking Methods
5. Color and Pattern-Based Methods

2.2.3.1 Region-Based Tracking Methods

In these methods, the regions of the moving objects (blobs) are tracked and used for tracking the vehicles. These regions are segmented from the subtracting process between the input frame image and prior stored background image.

S. Gupte, et al [6] introduced a model based vehicle detection, tracking, and classification which works efficiently under most weather conditions. Their vehicle model is based on the assumption that the scene has a flat ground. A vehicle is modeled as a rectangular patch whose dimensions depend on its location in the image. The dimensions are equal to the projection of the vehicle at the corresponding location in the scene. A vehicle consists of one or more regions, and a region might be owned by zero or more vehicles. The region tracking stage produces a conflict-free association graph that describes the relations between regions from the previous frame and regions from the current frame. The vehicle tracking stage updates the location, velocity, and dimensions of each vehicle based on this association graph. The location and dimensions of a vehicle are computed as the bounding box of all its constituent blobs. The velocity is computed as the weighted average of the velocities of its constituent blobs. The vehicle's velocity is used to predict its location in the next frame. A region can be in one of five possible states. The vehicle tracker performs different actions depending on the state of each region that is owned by a vehicle.

When the tested results taken into consideration, the system could detect and track 90% of the vehicles correctly and classify 70% vehicles correctly from detected 90%.Errors in detection were most frequently due to occlusions and/or poor segmentation. The two vehicles will continue to persist as a single vehicle if the relative motion between them is small. So this system will not be useful in Sri Lankan traffic conditions with large number of vehicle occlusions. Another thing to note is that the images we used are grayscale. Since their segmentation approach is intensity based, vehicles whose intensity is similar to the road surface are sometimes missed, or detected as fragments that are too small to be reliably separated from noise. So that is another difficulty in using this system in our project.

2.2.3.2 Contour Tracking Methods

Contour Tracking Methods depend on contours (the boundaries of vehicle) of vehicles in vehicle tracking process. A real-time vehicles tracking and classification technique on highway have offered by A. Ambardekar, et al [7]. This approach employs optical movement and un-calibrated camera parameter knowledge to detect a vehicle pose in the 3D world. The proposed approach uses two new techniques: color contour based matching and gradient based matching, and it showed promising results when it tested for tracking, foreground object detection, vehicle recognition and vehicle speed assessment methods. In this approach, Optical flow algorithm estimates the motion of each pixel between two image frames. They use optical flow to estimate how different blobs are moving. Assuming that the vehicles move in the forward direction, optical flow gives a good estimate on how vehicles are oriented (see figure 1).

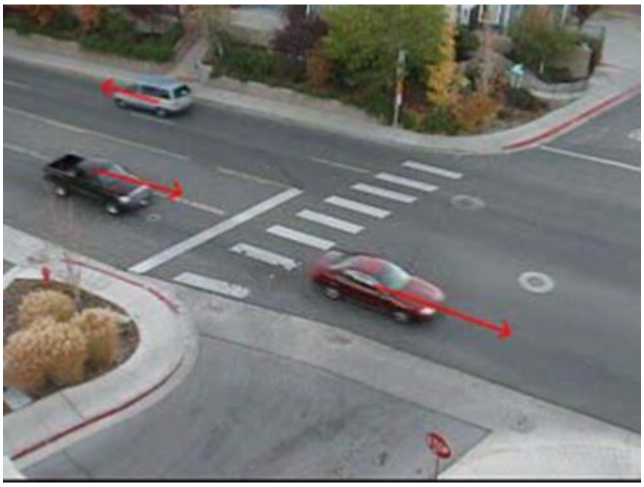


Figure 11: Average optical flow vectors (red arrows).

This information is used by the reconstruction module to obtain a 3D model of the vehicle with the correct orientation. Then detect the edges using the canny edge detector in the regions where the objects were found by the foreground object detection module. These edges are used in the vehicle detection and classification stage. The main advantage of this tracking algorithm is its speed. In terms of traffic parameter collection, they keep a record of how each track was classified in each frame, the number of active tracks (vehicles) present at any time, velocity of each vehicle at current time, and average velocity of each vehicle during the entire time when it was visible in the camera's field of view. The velocity of the vehicle can be calculated by using the tracks' location information.

2.2.3.3 3D Model-Based Tracking Methods

K. ZuWhan and J. Malik [8], suggested a new 3D model-based vehicle detection and depiction framework which is based on a probabilistic boundary feature grouping, which is used for vehicle detection and tracking process. In this method, first apply a stabilization algorithm to all images to reduce impact of moving the unstable camera. For each image, we manually assign several "static" (background) areas for the stabilization. Then, for each frame, we find corners from these areas, find their matches in the previous frame, estimate camera transformation (affine), and generate new images using the transformation matrix. Then apply the vehicle detection and description algorithm. This proposed method needs three cameras and uses three frames of three cameras to detect track and classify vehicles.

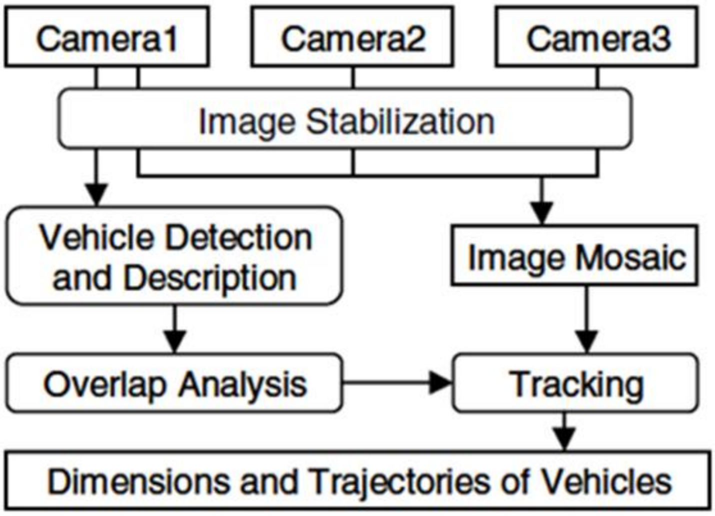


Figure 12: System overview

The tracking is performed on the mosaic of all three images. Then manually calibrated all three cameras and rectify images using the calibration parameters and attach them to generate the mosaic image. Brightness and contrast levels of all three cameras are different from each other. Therefore, algorithm had to adjust the brightness and contrast levels by examining those of overlapping areas: then estimate the mean and standard deviation of the intensity pixels, and (linear) transform all the intensity levels of the images so that the brightness and contrast levels of all three images be the same. The tracking is performed based on the zero-mean cross correlation matching.

If the Overall design is taken into consideration, this implementation is difficult to be applied in to our project because of its low performance in tracking moving vehicles. Another reason is this implementation needs three cameras instead of a single camera. So it is not cost effective as well.

2.2.3.4 Feature-Based Tracking Methods

An iterative and distinguishable framework based on edge points and modified SIFT descriptors as features uses in similarity process, these features represents a large region of set of features forms a strong depiction for object classes. The proposed framework by M. Xiaoxu and W. E. L. Grimson [9] showed a good performance for vehicle classification in surveillance videos despite of significant challenges such as limited image size and quality and large intra-class dissimilarities. Their feature extraction method includes four steps: Extract edge points, attach a descriptor to each edge point, Segment edge points into point groups and form features from edge point segments. This system still has higher error rate than the expected. So this does not useful in our project.

An automatic unique visual-based expressway surveillance approach for segmenting and tracking vehicles introduced by N. K. Kanhere, et al [10]. This approach suffers from difficulties with existence of rigorous occlusion due to low-level floor position of camera on the roadside. In this paper, the particular vehicles are detected, segmented, and tracked in image sequence by assembling, bunching, and approximating of the 3D world coordinates of vehicle's feature points.

2.2.3.5 Color and Pattern-Based Methods

A technique to traffic supervision using color analyzing is used by H. Mao-Chi and Y. Shwu-Huey [11]. That system uses the color analyzing technique by embracing the YCrCb color space for the construction of preliminary background, segmenting foreground, vehicle location, vehicle tracking, shade elimination, and background updating algorithms that are used throughout the proposed system. Through the practical experiments, this system is proven to be working under several weather situations, and it is insensitive to lighting. The system uses an algorithm based on YCrCb color space as it has robustness to brightly light backgrounds and preciseness of shadows detection. The system was implemented on Intel Celeron 2 GHz PC. The test results shows that 90% of the vehicles were correctly detected and tracked. Most of the detecting and tracking errors occur near the start point of the detection zone due to occlusion of vehicles. But once the vehicles move towards the camera, after one or two frames, they can be correctly detected and tracked. When considering the accuracy, although vehicles have look-alike colors with the background, with the help of YCrCb color system, all vehicles were correctly detected and tracked with shadow removed. Weaknesses of this method are, this solution cannot operate well under extreme weather conditions such as rainy day, night. It is unable to overcome vehicle overlap and occlusion conditions. All in all this system is not useful in for our project.

Another color and pattern based tracking method was introduced by G. D. Sullivan, et al [12]. It is a model-based system for real-time traffic supervision using continuous visual tracking and classification of vehicles for busy multi-lane highway scenes. In this proposed work, the researchers use the orthographic approximations for matching process. This system consists of three improved main levels: (1) using 1-D patterns of shape and posture theory (2) theory tracking (3) using 2-D patterns of theory verification.

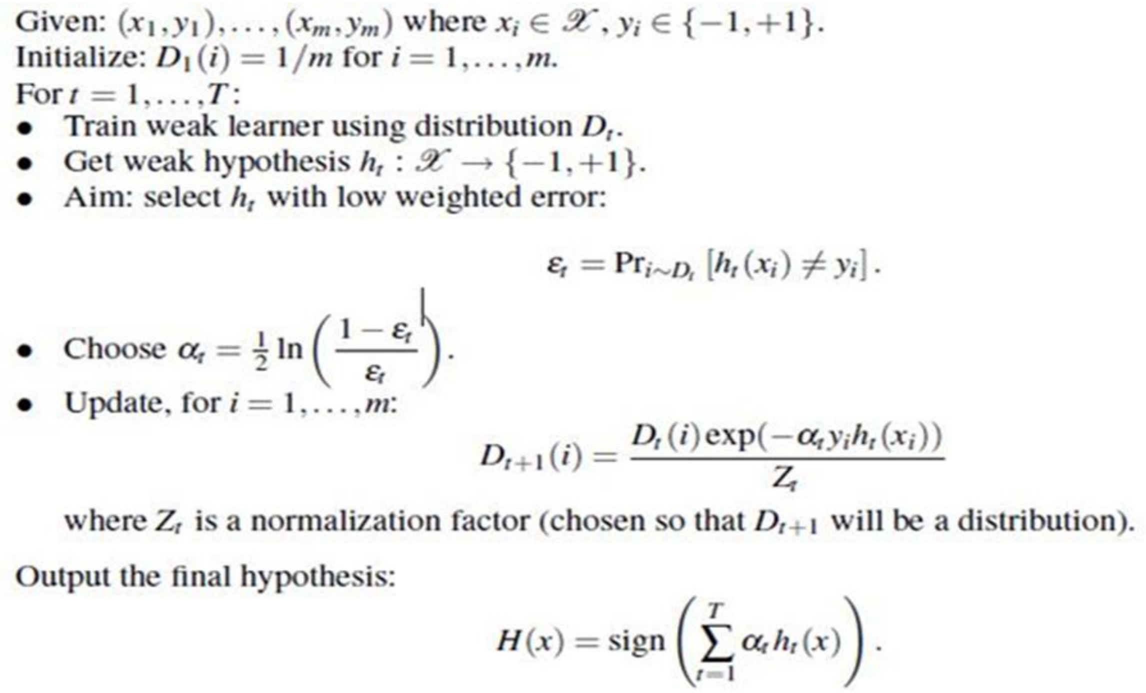
2.3 Algorithms

2.3.1 Kalman Filter

The Kalman filter has many applications in technology. The main function of Kalman Filter is to utilize measurements recorded over time which contain random variations and inaccuracies to generate values that tend closer to the measurement true values and connected values that resulted from calculations. The Kalman filter calculate associated values by predicting a value to estimate the uncertainty of that predicted value, and compute a weighted average of the predicted value and measured value. Kalman filter was successfully used for vehicle tracking by A. Salarpour et. Al. [29] and F. Dellaert and C. Thorpe [30].

2.3.2 AdaBoost Algorithm

The AdaBoost algorithm was introduced by Freund and Schapire as the first practical boosting algorithm in machine learning area [40]. Machine learning based concept of boosting is an approach of creating a highly accurate prediction rule by combining many relatively weak and inaccurate rules.



```
Given: $(x_1, y_1), \ldots, (x_m, y_m)$ where $x_i \in \mathscr{X}$, $y_i \in \{-1, +1\}$.
Initialize: $D_1(i) = 1/m$ for $i = 1, \ldots, m$.
For $t = 1, \ldots, T$:
  • Train weak learner using distribution $D_t$.
  • Get weak hypothesis $h_t : \mathscr{X} \rightarrow \{-1, +1\}$.
  • Aim: select $h_t$ with low weighted error:
        $\varepsilon_t = \Pr_{i \sim D_t}[h_t(x_i) \neq y_i]$.
  • Choose $\alpha_t = \frac{1}{2} \ln\left(\frac{1 - \varepsilon_t}{\varepsilon_t}\right)$.
  • Update, for $i = 1, \ldots, m$:
        $D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$
    where $Z_t$ is a normalization factor (chosen so that $D_{t+1}$ will be a distribution).
Output the final hypothesis:
        $H(x) = \operatorname{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$.
```

Figure 13: Pseudocode for AdaBoost

According to the algorithm each round t = 1;:::; T, a distribution Dt is computed over the m training examples. Then the given weak learner or weak learning algorithm is applied to find a weak hypothesis ht: X à {-1, +1}, with low weighted error relative to distribution Dt. The final or combined hypothesis H computes the sign of a weighted combination of weak hypotheses. The formula as follows:

$$H(x) = \sum_{t=1}^{T} \propto_t h_t(x)$$

Combined hypothesis H will be used as the highly accurate prediction rule.

Reference

[1] Z. Zhigang, et al., "A real-time vision system for automatic traffic monitoring based on 2D spatiotemporal images," in Applications of Computer Vision, 1996. WACV '96. Proceedings 3rd IEEE Workshop on, 1996, pp. 162-167.

[2] Masstr.njmeadowlands.gov, 'Meadowlands Adaptive Signal System for Traffic Reduction (MASSTR)', 2015. [Online]. Available: http://masstr.njmeadowlands.gov/. [Accessed: 15-Jun-2015].

[3] Reggie Chandra, Chris Gregory. (2012, March) InSync Adaptive Traffic Signal Technology: Real-Time Artificial Intelligence Delivering Real-World Results. [Online]. Available: https:// rhythmtraffic.com/paper

[4] Scoot-utc.com, 'Split Cycle Offset Optimisation Technique (SCOOT)', 2015. [Online]. Available: http://www.scoot-utc.com/. [Accessed: 15- Jun- 2015].

[5] Scats.com.au, 'Sydney Coordinated Adaptive Traffic System (SCATS)', 2015. [Online]. Available: http://www.scats.com.au/. [Accessed: 15- Jun- 2015].

[6] S. Gupte, et al., "Detection and classification of vehicles," Intelligent Transportation Systems, IEEE Transactions on, vol. 3, pp. 37-47, 2002.

[7] A. Ambardekar, et al., "Efficient Vehicle Tracking and Classification for an Automated Traffic Surveillance System," in International Conference on of Signal and Image Processing, 2008, pp. 1-6.

[8] K. ZuWhan and J. Malik, "Fast vehicle detection with probabilistic feature grouping and its application to vehicle tracking," in Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, 2003, pp. 524-531 vol.1.

[9] M. Xiaoxu and W. E. L. Grimson, "Edge-based rich representation for vehicle classification," in Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, 2005, pp. 1185-1192 Vol. 2.

[10] N. K. Kanhere, et al., "Vehicle Segmentation and Tracking in the Presence of Occlusions," TRB Annual Meeting Compendium of Papers, Transportation Research Board Annual Meeting, pp. 89-97, 2006.

[11] H. Mao-Chi and Y. Shwu-Huey, "A real-time and color-based computer vision for traffic monitoring system," in Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on, 2004, pp. 2119-122 Vol.3.

[12] G. D. Sullivan, et al., "Model-based vehicle detection and classification using orthographic approximations," Image and Vision Computing, vol. 15, pp. 649-654, 1997.

[13] Leo, M. et al. Shape based people detection for visual surveillance systems. In Audio and Video Based Biometric Person Authentication. 4 th International Conference, AVBPA 2003. Proceedings Lecture Notes in Computer Science. Springer-Verlag, Berlin, Germany. 2003. Vol.2688

[14] Wren, C.R., A. Azarbayejani, T. Darrell and A.P. Pentland, 1997.Pfinder: Real-time tracking of the human body. IEEE Trans. Pattern Anal. Mach. Intel., 19: 780-785.

[15] Papageorgiou, C. and T. Poggio.Trainable Pedestrian Detection. In:Proceedings of International Conference on Image Processing (ICIP'99),Kobe, Japan, October 1999.

[16] N. Dalal, B. Triggs, “Histograms of oriented gradients for human detection”,IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005

[17] B.Leibe, E. Seemann, and B. Schiele. "Pedestrian detection in crowded scenes" IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005

[18] Bo Wu and Ram Nevatia, "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors",IEEE International Conference on Computer Vision (ICCV), 2005

[19] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In International Conference on Computer Vision. 1999. Kerkyra, Corfu, Greece.

[20] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In Computer Vision and Pattern Recognition. 1999. Fort Collins, Colorado.

[21] D.-S. Lee, Effective Gaussian Mixture Learning for Video Background Subtraction. IEEE Transactions on pattern analysis and machine intelligence, 2005. 27(5): p. 827-832.

[22] Alessandra Fascioli, ReanIsabella Fedriga, and Stefano Ghidoni,Vision-based Monitoring of Pedestrian Crossings, InProcs.14textsuperscriptthIntl.Conf.on Image Analysis and Processing, Modena, Italy, September 2007.

[23] Image processing for pedestrian detection using a high mounted wide-angle camera, Martin Lars Svante Johansson, G¨oteborg, August 2012.

[24] G.Thomas Prathiba , Y.R.Packia Dhas, Literature survey for people counting and human detection, IOSR Journal of Engineering (IOSRJEN), Jan. 2013

[25] AkihiroSugimoto, Mitsuhiro Kimura, Takashi Matsuyama:Detecting Human Heads with their orientations. Progress in Computer Vision and Image Analysis 2009: 261-282

[26] Tian, Q, Zhou, B, hua Zhao, Wei, Y. & wei Fei, 'Human Detection using HOG Features of Head and Shoulder Based on Depth Map.' 2103, JSW 8.

[27] Maojun, Z. et al. A visual method for real-time detecting persons in complex scenes. In Proceedings of the 21 st IEEE Instrumentation and Measurement Technology Conference. IEEE, Piscataway, NJ, USA. 2004. 438-40 Vol.1

[28] Baisheng, C. et al. A novel background model for real-time vehicle detection. In Yuan, B. et al (eds). 7 th International Conference on Signal Processing Proceedings IEEE. IEEE, Piscataway, NJ, USA. 2004. 1276-9 vol.2

[29] A. Salarpour, et al., “VEHICLE TRACKING USING KALMAN FILTER AND FEATURES," Signal & Image processing : An International Journal (SIPIJ) Vol.2, No.2, June 2011

[30] Frank Dellaert and Chuck Thorpe, "Robust Car Tracking using Kalman filtering and Bayesian templates," Conference on Intelligent Transportation Systems, 1997.

[31] T. E. Boult, et al., "Frame-rate omnidirectional surveillance and tracking of camouflaged and occluded targets," in Visual Surveillance, 1999. Second IEEE Workshop on, (VS'99), 1999, pp. 48-55.

[32] K. Toyama, et al., "Wallflower: principles and practice of background maintenance," in Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, 1999, pp. 255-261 vol.1.

[33] G. C. De Silva, "Automation of Traffic Flow Measurement Using Video Images," Master of Engineering, University of Moratuwa, Sri Lanka, 2001.

[34] Samir Sahli, Yueh Ouyang, Yunlong Sheng, Daniel A. lavigne “Robust vehicle detection in low-resolution aerial imagery” an Image Science group, 2010.

[35] P. Viola, M. Jones 2001. Rapid Object Detection using a Boosted Cascade of Simple Features, Conference on Computer Vision and Pattern Recognition CVPR, Hawaii, December 9-14, 2001.

[36] P. Viola, M. Jones 2003. Detecting Pedestrians Using Patterns of Motion and Appearance, International Journal of Computer Vision, USA, June 23, 2003.

[37] M. Oliveira, V. Santos, "Automatic Detection of Cars in Real Roads using Haar-like Features," in computer vision, Department of Mechanical Engineering, University of Aveiro, Portugal.

[38] B. Han, et al., "Motion-segmentation-based change detection," SPIE Defence & Security Symposium 2007, pp. 65680Q-65680Q, 2007.

[39] L. Vasu, "An effective step to real-time implementation of accident detection system using image processing," Master of Science, Oklahoma State University, USA, 2010.

[40] R. E. Schapire, "Explaining AdaBoost," Princeton University, Dept. of Computer Science, 35 Olden Street, Princeton, NJ 08540 USA.