PEDESTRIAN DETECTION USING BACKGROUND SUBTRACTION AND HISTOGRAM OF ORIENTED GRADIENTS

A Report Submitted

in Partial Fulfillment of the Requirements

for the Degree of

Bachelor of Technology

in

Computer Science & Engineering

by

Group NO : 23
M.V.Harshitha(20134117)
Kiran Parwani(20134167)
Abhishek Bhardwaj(20134157)
J.Nikhil Chandra(20134160)
B.Yashwanth(20134150)

to the

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY
ALLAHABAD

April, 2017

UNDERTAKING

We declare that the work presented in this report titled "*PEDESTRIAN DETECTION USING BACKGROUND SUBTRACTION AND HISTOGRAM OF ORIENTED GRADIENTS*", submitted to the Computer Science and Engineering Department, Motilal Nehru National Institute of Technology, Allahabad, for the award of the ***Bachelor of Technology*** degree in ***Computer Science & Engineering***, is our original work. We have not plagiarized or submitted the same work for the award of any other degree. In case this undertaking is found incorrect, We accept that our degree may be unconditionally withdrawn.

April, 2017
Allahabad

(M.V.Harshitha
Kiran Parwani
Abhishek Bhardwaj
J.Nikhil Chandra
B.Yashwanth)

CERTIFICATE

Certified that the work contained in the report titled "*PEDESTRIAN DETECTION USING BACKGROUND SUBTRACTION AND HISTOGRAM OF ORIENTED GRADIENTS*", by M.V.Harshitha, Kiran Parwani, Abhishek Bhardwaj, J.Nikhil Chandra, B.Yashwanth has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

(Dr.Rajitha.B)
Computer Science and Engineering Dept.
M.N.N.I.T, Allahabad

April, 2017

Preface

This project focuses on Human Detection. Detecting human beings accurately in a visual surveillance system is crucial for diverse application areas including abnormal event detection, human gait characterization, congestion analysis, person identification, gender classification and fall detection for elderly people. The first step of the detection process is to detect an object which is in motion.
Object detection could be performed using background subtraction, optical flow and spatial-temporal filtering techniques. Once detected, a moving object could be classified as a human being using shape-based, texture-based or motion-based features.

Acknowledgement

We would like to take this opportunity to express our deep sense of gratitude to all those who helped us directly or indirectly in our thesis work.First we would like to thank our supervisor Dr. Rajitha.B, for being great mentor we could ever have. Her advice, constant encouragement and constructive criticism were source innovative ideas and cause behind the successful completion of this dissertation.
We wish to express our sincere gratitude to Prof. Rajeev Tripathi,Director, MNNIT Allahabad and Prof. Neeraj Tyagi, Head, Computer Science and Engineering Department, for providing us with all facilities required for completion of this work.
We are also greatly indebted to our friends who helped us with their suggestions and ideas at critical time.

Contents

Chapter 1

Introduction

Finding people in images has attracted much attention in recent years for practical applications such as visual surveillance, vehicle auxiliary driving and image understanding. It is a next step after the development of successful face detection algorithms. The detection of humans has become an own research field.
However, unlike other object detection, human detection has some of its own characteristics. Humans usually have many different appearances in pose and style, and the background of the images or videos is often cluttered and has on general describable structure. Hence, human detection in image or videos is a challenging task for the variable appearance and various poses, which can influence the algorithm of choice. The articulated pose, style and color of clothes, illumination conditions in outdoor scene are the major aspects that affect the detection results.

1.1 Overview

Pedestrian Detection process can be divided into three stages [3]:

1. Region of Interest Selection i.e Motion Detection
2. Selection of effective features
3. Classification of Pedestrians

1.2 Background Modelling and Motion Detection

Motion detection is critical to many automated visual applications. A high degree of sensitivity and robustness is often desired from detection mechanisms. Motion can be detected by measuring change in speed or vector of an object.
The simplest way of accomplishing detection is through building a representation of the scene background and comparing each new frame with this representation. This procedure is known as background subtraction. Some of the OpenCv functions for background subtraction purposes are listed as :

- BackgroundSubtractorMOG
- BackgroundSubtractorMOG2
- BackgroundSubtractorGMG

The above mentioned functions are based on Gaussian Mixture-based Background/ Foreground Segmentation, statistical background image estimation and per-pixel Bayesian segmentation. The general idea behind some of these algorithms is to represent each pixel of a scene using a probability density function (PDF).
But, these techniques are bounded by limitations such as explicitly handling dynamic changes of the background e.g., gradual or sudden (as in moving clouds); motion changes including camera oscillations and high frequency background objects (tree branches, sea waves, etc.) and changes in the background geometry.
To overcome these problems we can use clustering approach. The fundamental problem of cluster background subtraction involves a decision whether a cluster of pixels belongs to the background or foreground object based on the descriptors like colour difference histograms and cluster densities.[2]

1.3 Human Detection

The problem of human detection is to automatically locate people in an image or video sequence and has been actively researched in the past decade. Existing surveys decompose a human detection method into two components:

- Features
- Classifiers

In general, the process of detecting human objects from images/videos can be performed in the following sequential steps: extracting candidate regions that are potentially covered by human objects, describing the extracted regions, classifying/verifying the regions as human or non-human, and post-processing (e.g. merging the positive regions or adjusting the size of those regions).[4]
A robust feature set is required that allows the human form to be discriminated cleanly, even in cluttered backgrounds under difficult illumination. These features can be computed from low-level information such as edge, texture, colour, or motion. Some edge based features are:

1. Pixel level edge-based features: They refer to the edge-based features computed at individual pixels.The location of each edge pixel can be encoded by its spatial distance to the nearest edge pixel on a template modelling the human shape. The templates can be as simple as parallel edge segments,rectangular contours, or small curves and segments called edgelets.

2. Region level edge-based features: In contrast to computing edge-based features at pixel level, edge features obtained from local image regions have also been explored. Compared with the pixel level features, the region level features are to a certain extent adaptive to the local deformation of the human shape. A number of ways have been proposed to compute the region level edge-based features like Histogram Of Oriented Gradients (HOG).

We will be discussing more features in further chapters(i.e.Literature Survey).
Since the classification of candidate regions can be considered as a binary classification problem, discriminative methods are generally used. Support Vector Machine (SVMs) are often used to classify the human and non-human descriptors by maximising the margin between these two classes.

1.4 Problem Statement

The main problem statement was efficient human detection and tracking in a surveillance video. The first step would be proper background Modelling in order to identify changes and keep track of motion.
Problems encountered with the Background during motion detection :

- Dynamic texture scenes like waving leaves, rippling water.
- Illumination variations.
- Colour based differences and noise.
- Shadow and Ghost effects.

In order to rectify the above problems Non parametric background model is to be constructed. Fuzzy cmeans clustering can provide solution for this as it analyses the contribution to each cluster for every pixel.
After the foreground detection an apt feature is required for human detection. From the machine vision perspective, it is hard to distinguish an object as a human due to its large number of possible appearances. Moreover, the human motion is not always periodic, but a combination of features could be useful in identifying humans.
For this purpose, HOGs have several benefits over other descriptor methods as they work on localized parts of the image and hence are capable of addressing occlusion problems.

1.5 Organisation of the report

The report is organised as follows :

> Chapter 2 - This chapter discusses about various pre-existing approaches of background subtraction. Background subtraction is a major technique applied for foreground segmentation to analyze motion. It also includes feature selection that covers shape, edge and motion based approaches and design of the classifier architecture.

Chapter 3 - The chapter gives detail about all the hardware and software requirements like the libraries and packages needed. It includes list of all python-Opencv libraries and supporting softwares.

Chapter 4 - This chapter explains the implemented approach and algorithm. It gives the basic workflow of code and outlines various steps. The algorithms of Fuzzy cmeans and HoG calculation are also explained stepwise.

Chapter 5 - This chapter gives the information of datasets used and analysis of obtained experimental results. It includes comparison of resulting clustered images with the ground truth.

Chapter 6- It concludes the whole report.

Chapter 2

Background and Literature Review

2.1 Determination of Region of Interest

Background Modelling and Subtraction is used to detect changes in frames which determine the region of Interest. The different background subtraction techniques published in the literature can be classified depending on the features and procedures used to construct the background model.

- The pixel-based approach models observe scenes as a set of independent pixel processes.
- The region-based approach builds background models by taking advantages of inter-pixel relations, demonstrating impressive results in handling non-stationary background.

Various Approaches :

2.1.1 Mixture Of Gaussians

Mixture of Gaussians is a widely used approach for background modeling to detect moving objects from static cameras. Stauffer and Grimson introduced an adaptive

Gaussian mixture model, which is sensitive to the changes in dynamic scenes derived from illumination changes, extraneous events, etc. Rather than modelling the values of all the pixels of an image as one particular type of distribution, they modelled the values of each pixel as a mixture of Gaussians. Over time, new pixel values update the mixture of Gaussian (MoG) using an online K-means approximation. For example in case of vehicle surveillance Gaussians are manually labeled in a heuristic manner as follows: the darkest component is labeled as shadow; in the remaining two components, the one with the largest variance is labeled as vehicle and the other one as road. This remains fixed for all the process giving lack of adaptation to changes over time. For the foreground detection, each pixel is compared with each Gaussian and is classified according to it corresponding Gaussian. The maintenance is made using an incremental EM algorithm for real time consideration. Stauffer and Grimson generalized this idea by modeling the recent history of the color features of each pixel mixture of K Gaussians.

2.1.2 Non-parametric background model

Kim and Kim proposed a non-parametric method, which was found effective for background subtraction in dynamic texture scenes (e.g. waving leaves, spouting fountain and rippling water). The model can handle the situations where the background of scene is cluttered and not completely static and can have small motions. This model estimates the probability of observing pixel intensity values based on a sample of intensity values for each pixel. The implementation of model runs for both grey level and color imagery. This approach achieves low false alarm rates with very sensitive detection. In this approach kernel-based function is employed to represent the colour distribution of each background pixel. Although the processing time was high in comparison with the adaptive Gaussian mixture model, the false positive rate of detection is signicantly low at high true positive rates.

2.1.3 Temporal Differencing

The temporal differencing approach involves three important modules: block alarm module, background modelling module and object extraction module. The block alarm module efficiently checked each block for the presence of either a moving object or background information. This was accomplished using temporal differencing pixels of the Laplacian distribution model and allowed the subsequent background modelling module to process only those blocks that were found to contain background pixels. Next, the background modelling module is employed in order to generate a high-quality adaptive background model using a unique two-stage training procedure and a mechanism for recognizing changes in illumination. As the final step of their process, the proposed object extraction module computes the binary object detection mask by applying suitable threshold values. This is accomplished using a proposed threshold training procedure.

2.1.4 Hierarchical background model

The hierarchical models first detect the regions containing foreground and then locate the foreground only in these regions, thus avoid detection failure in other regions and reduce the time and cost. By incorporating the pixel-based segmentation into the coarse-level results, a two-stage hierarchical approach is introduced. It first segments the background images into several regions by the mean-shift algorithm. Then, a hierarchical model, which consists of the region models and pixel models, is created. The region model is one kind of approximate Gaussian mixture model extracted from the histogram of a specific region. The pixel model is based on the co-occurrence of image variations described by HOG of pixels in each region. Benefiting from the background segmentation, the region models and pixel models corresponding to different regions can be set to different parameters. The pixel descriptors are calculated only from neighbouring pixels belonging to the same object.

2.2 Feature Selection

The feature selection is a key problem in the process of human detection. The selected features must embody the objective characteristics, proper features can improve classification accuracy, while the improper features may lead to misjudge. Global features like body shape or silhouettes can be used to express the difference between human and other objects. However, these features are not flexible since human can have many kinds of poses and shapes. So its hard to model human for a trained classifier.[7]

2.2.1 Shape and Edge Based

Histogram Of Oriented Gradients

Dalal and Triggs [1] present a human detection algorithm with excellent detect precision. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy. The essential thought behind the histogram of oriented gradients descriptor is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The image is divided into small connected regions called cells, and for the pixels within each cell, a histogram of gradient directions is compiled. The descriptor is the concatenation of these histograms. For improved accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block. This normalization results in better invariance to changes in illumination and shadowing. The HOG descriptor has a few key advantages over other descriptors. Since it operates on local cells, it is invariant to geometric and photometric transformations, except for object orientation. Such

changes would only appear in larger spatial regions. Moreover, as Dalal and Triggs discovered, coarse spatial sampling, fine orientation sampling, and strong local photometric normalization permits the individual body movement of pedestrians to be ignored so long as they maintain a roughly upright position. The HOG descriptor is thus particularly suited for human detection in images.

2.2.2 Appearance Based

Haar features

Appearance features are mainly to capture the colour and texture and they are also extracted in local image regions. Haar feature is another commonly used appearance feature, was proposed by Paul Viola and Michael Jones [8] in their paper, "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images.

Haar features are used to detect the presence of feature in given image. Each features result in a single value which is calculated by subtracting the sum of pixels under white rectangle from the sum of pixels under black rectangle. To compute the rectangle features rapidly integral image concept is used. It need only four values at the corners of the rectangle for the calculation of sum of all pixels inside any given rectangle. Voila Jones algorithm uses a 24 x 24 window as the base window size to start evaluating these features in any given image.

If we consider all the possible parameters of the haar features like position, type and scale then we have to calculate the 160,000 features in this window but this is practically impossible. The solution of this problem is to use the adaboost algorithm. Adaboost is a machine learning algorithm which helps us to find the best features among the 160,000. These features are the weak classifiers. Adaboost construct a strong classifier as a linear combinaion of these weak classifiers. In Haar cascade, an image will be a human face if it passes all the stages. If it is not passed any one of the stage it means the image is not a human face.

Local Binary Pattern

Local binary pattern (LBP) which was proposed in 1990 has been found to be a powerful feature for texture classification. This phenomenon can also be used to describe the appearance of the human body. The basic idea is to summarize the local structure in an image by comparing each pixel with its neighbourhood. Similarly to HOG, an image region is encoded by the histogram of LBPs computed at all pixels in that region . The computed LBP feature vector can be processed using the Support vector machine or some other machine-learning algorithm to classify images. Such classifiers can be used for face recognition or texture analysis. LBP is well-known for its robustness against illumination changes, discriminative power, and computational simplicity. Many variants and extensions of LBP have also been developed.

Local Ternary Pattern(LTP) is an extension of LBP. Unlike LBP, it does not threshold the pixels into 0 and 1, rather it uses a threshold constant to threshold pixels into three values. It allows 3-valued quantisation of local intensity difference. Compared with the LBP, the LTP achieved better detection performance. LBP and LTP were generalised to the so-called local intensity distribution (LID) descriptor in which more quantisation levels were adopted and the neighbouring pixels in an LBP/LTP pattern were assumed to be independent.

2.2.3 Motion Based

The availability of motion information can be exploited in human descriptors. Motion can be used to discriminate one object from another if the motion patterns are different and thus, plays important role in object description. This is especially true for non-rigid objects such as pedestrians which often perform cyclic movements. To encode the motion of human objects, temporal features are defined from the temporal difference or optical flows.

Histogram Of Flows

For the use of optical flows, histogram of flows (HOF)[6] was proposed. The HOF was computed in a similar manner with the HOG. The HOF can be used to describe the boundary motion as well as internal motion (i.e. the motion of internal regions of

the human body). The HOG was extracted on the flow images to encode the motion of pedestrians. It would be useful to confirm the presence of a human object if the body parts or joints can be tracked through motion. In particular, spatio-temporal patterns modelling the motion of joints were represented by trajectories and learned from the training data. Given a hypothesis of being a human, the trajectories were computed based on tracking feature points using the dense optical flow method. The hypothesis was then validated by finding a subset of those trajectories that best matched the spatio-temporal patterns learned off-line.

2.3 Design of Classifier Architecture

After features are extracted from each image, some classifiers for supervised learning such as neural network, support vector machine are then used to classify objects based on sample data. Discriminative classification techniques aim at determining an optimal decision boundary between pattern classes in a feature space.

Nueral Network

Neural network is an effective tool for image classification and recognition. In the process of pedestrian detection, multi-layer neural networks have been utilized in conjunction with adaptive local feature in the hidden network layer. Pedestrian detection is a popular research topic due to its paramount importance for a number of applications, especially in the fields of automotive, surveillance and robotics. Despite the significant improvements, pedestrian detection is still an open challenge that calls for more and more accurate algorithms. In the last few years, deep learning and in particular Convolutional Neural Networks emerged as the state of the art in terms of accuracy for a number of computer vision tasks such as image classification, object detection and segmentation, often outperforming the previous gold standards by a large margin.

Support Vector Machine

Support vector machine (SVM)[5] have evolved as a powerful tool to solve pattern classification problems. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that

assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. The cascade architecture is tuned to detect almost all pedestrians while rejecting no-pedestrians as early as possible.

Chapter 3

Requirement Analysis

3.1 Hardware Requirement

The proposed technique is implemented on a 64-bit Ubuntu 14.04 platform with AMD A6-5200 APU with Radeon(TM) HD Graphics x 4 and 4 GB RAM.

3.2 Software Requirement

The Description of various Software and Module requirements are as follows -

1. Software Tools Required -
    - opencv
    - PyQt4 packages for GUI purposes
2. Opencv and its Modules Required -
    - numpy
    - matplotlib
    - scipy
    - cv2
    - skfuzzy

3. Python Packages Required -

   - GCC 4.4x or Higher
   - CMake 2.6 or Higher
   - Git and pkgconfig
   - GTK+2.x or Higher including headers (libgtk2.0dev)
   - python2.6 and opencv3.0
   - ffmpeg or libav development packages
   - pip 9.0.0 package
   - Optional libtbbdev, libdc13942.x, libjpegdev, libpngdev etc . . .

Chapter 4

Proposed Work

4.1 Alogrithms Implemented

4.1.1 Fuzzy CMeans Clustering

Clustering techniques are mostly unsupervised methods that can be used to organize data into groups based on similarities among the individual data items. Hard clustering methods are based on classical set theory, and require that an object either does or does not belong to a cluster. Hard clustering means partitioning the data into a specified number of mutually exclusive subsets.
Fuzzy clustering methods, however allow the objects to belong to several clusters simultaneously, with different degrees of membership. In many situations, fuzzy clustering is more natural than hard clustering. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership. It is based on minimization of the following objective function in the equation(1) :

Objective func: $$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m ||x_i - c_j||^2 \qquad (1)$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the i^{th} of d-dimensional measured data, c_j is the dimension center of the cluster, and $|| * ||$ is any norm expressing the similarity between any

measured data and the center.
Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} given by equation(2) and the cluster centers c_j in equation(3):

Membership func:
$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \frac{||x_i - c_j||}{||xi - c_k||}^{\frac{2}{m-1}}} \quad (2)$$

Cluster j:
$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m} \quad (3)$$

4.1.2 Histogram Of Oriented Gradients

HOG stands for Histograms of Oriented Gradients. HOG is a type of feature descriptor. The intent of a feature descriptor is to generalize the object in such a way that the same object (in this case a person) produces as close as possible to the same feature descriptor when viewed under different conditions. This makes the classification task easier.
The HOG person detector uses a sliding detection window which is moved around the image. At each position of the detector window, a HOG descriptor is computed for the detection window. This descriptor is then shown to the trained SVM, which classifies it as either person or not a person.
The procedure of computing HOG feature is composed of the following steps[9]:

1. **Normalize gamma and Colour of input image.**

2. **Compute the Gradients**: The magnitude of gradient is given by equation(4) and orientation by equation(5)

   Magnitude: $|G| = \sqrt{I_x^2 + I_y^2}$ (4)
   Orientation: $\theta = \arctan\frac{I_y}{I_x}$ (5)

3. **Orientation Binning** : 8x8 pixel size cells are computed with 9 orientation bins for $[0^0, 180^0]$ interval. For each pixels orientation, the corresponding

orientation bin is found and the orientations magnitude $|G|$ is voted to this bin.

4. **Descriptor Blocks** : To normalize the cells orientation histograms, they should be grouped into blocks. From the two main block geometries, the implementation uses R-HOG geometry. Each R-HOG block has 2x2 cells and adjacent R-HOGs are overlapping each other for a magnitude of half-size of a block.

5. **Block Normalization** : L2-Norm normalization is implemented using norm(vec) method given by equation(6):

   norm(vec): $$f = \frac{v}{\sqrt{||v||_2^2 + e^2}} \qquad (6)$$

6. **Detector Window**:The detector window size is 64x128 pixels. This result in 8x16 cells and 7x15 R-HOG blocks, since blocks are overlapping. Each R-HOG block has 2x2 cells, which also has 1x9 histogram vector each. So the overall size of R-HOG descriptor of a window is 7x15x2x2x9.

7. **Classifier**: When a test image is given to the system, two half-sizes are generated, resulting in three images with scales 1, 1/2 and 1/4 of the original image. For each of these generated images, a window of size 64x128 is scrolled across the entire image with 32 pixel horizontal and 64 pixel vertical step sizes. And for each window that is cropped, a classifier is run.

4.2 Proposed Approach : Algorithm

4.2.1 Overview Of Steps

Our proposed algorithm comprises the following steps as stated in figure 1 :

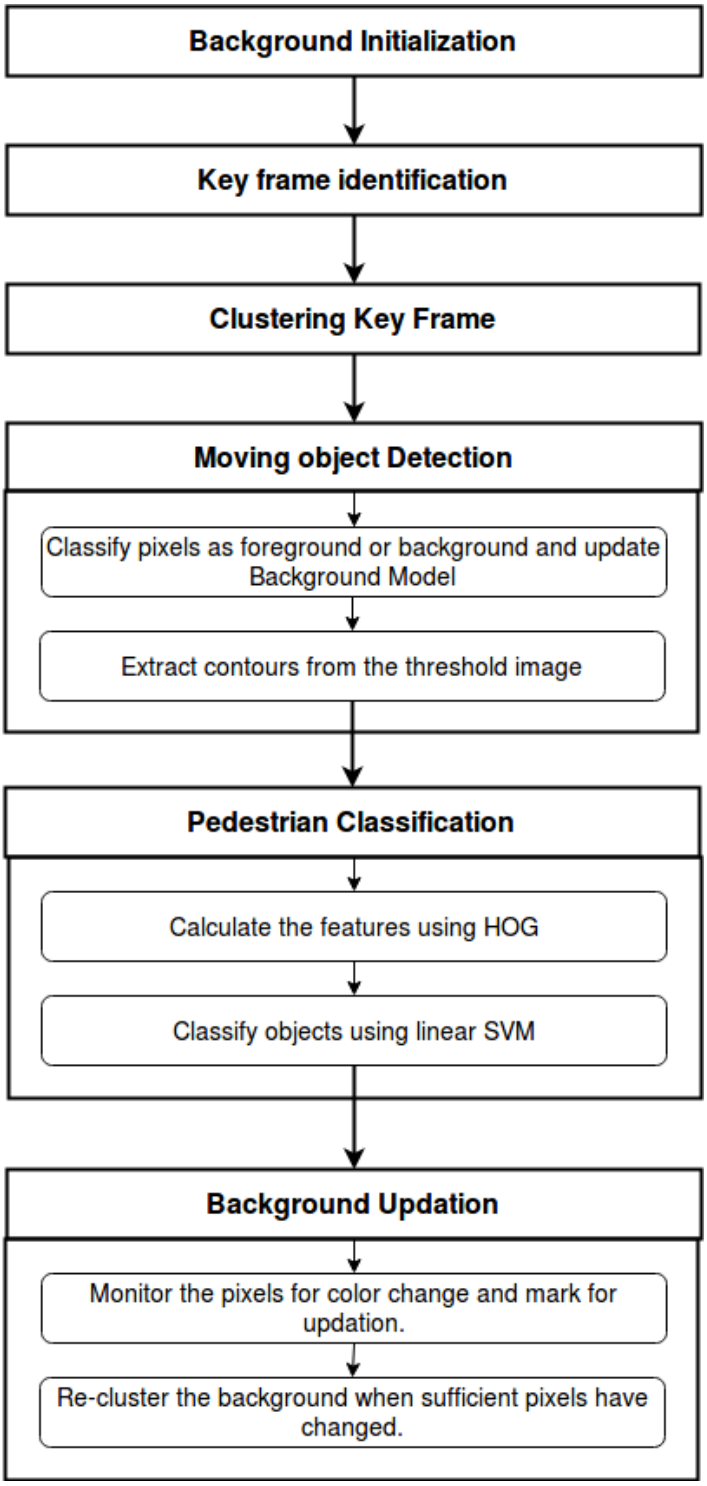

Figure 1: Flow of Algorithm

- **Background Initialization** : Compute the cluster centers and initial membership matrix of background frame by using fuzzy cmeans clustering given by algorithm 1.

- **Selection Of Key Frames** : Key Frames are selected on basis of absolute difference between the current and previous frames taken in grayscale.

- **Clustering Key Frame** : For every key frame obtained from above step is passed to Fuzzy_predict function that returns final membership matrix using the initial centers.

- **Foreground Detection and Background Updation** : Every pixel is classified as either foreground (0) or background (255) depending on the similarity measure. For every Background pixel, the Reference membership matrix u_i and background image are updated using learning rate α.This step of process is described by algorithm 2.

- **Contour Extraction** : The above generated Back and White image is passed to contour extraction function that detects the changed area.

- **Computing HOG and Classification** : The Changed area is cropped and HOG feature is calculated for this selected area given by algorithm 4. Classification whether it a person or not is done using the Linear SVM Classifier based on computed HOG feature.

- **Pixel Colour Monitoring** : The pixels in contours which have been rejected by HOG are monitored for their colour change. If the new colour appears continuously over bg_colour_freq frames then the pixel is marked for updation.

- **Background Re-clustering** : When the number of pixels marked for updation exceeds $bg\_cluster\_thresh$ then the updated background image is reclustered using Fuzzy cmeans algorithm.The last two steps are described in algorithm 3.

4.2.2 Algorithm

Algorithm 1 Fuzzy CMeans Clustering Algorithm

1: **Input Data** : X = $x_1, x_2.....xn$ *where i = 1,2,...,N*
2: **Output Data** : U membership matrix.
3: Set the values of c, q, ξ
4: Initialize $U = [u_{ij}]$ matrix, $U^{(0)}$
5: **for** each $k \in maxiter$ **do**
6: **for** each $j \in c$ **do**
7: $c_j \leftarrow \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m}$
8: **end for**
9: **for** each $i \in N$ **do**
10: **for** each $j \in c$ **do**
11: $u_{ij} \leftarrow \frac{1}{\sum_{k=1}^{C} \frac{||x_i - c_j||}{||xi - c_k||}^{\frac{2}{m-1}}}$
12: **end for**
13: **end for**
14:
15: **if** $||U^{(k+1)} - U^{(k)}|| > \xi$ **then**
16: return to step 4
17: **end if**
18: **end for**

Algorithm 2 Foreground Detection and Background Maintenance

Input Data : U, U^b membership matrices of current frame and background
bg_frame - 3D array (RGB) of the current background
$frame$ - 3D array (RGB) of the current frame
α is the learning rate
Output Data : Updated bg_frame and difference image B.
for each $i \in N$ **do**
    Similarity ρ : $\rho(u_i^b, u_i) \leftarrow \sum_{j=0}^{c-1} min(u_j^b, u_j)$
    if $\rho(u_i^b, u_i) > th$ **then**
        $B[i] \leftarrow 1$
        $u_i^b \leftarrow (1-\alpha) * u_i^b + \alpha * u_i$
        $x \leftarrow i/num\_cols$
        $y \leftarrow i\%num\_cols$
        for each $j \in (0, 3)$ **do**
            $bg\_frame[x][y][j] \leftarrow (1-\alpha) * bg\_frame[x][y][j] + \alpha * frame[x][y][j]$
        end for
    else
        $B[i] \leftarrow 0$
    end if
end for

Algorithm 3 Background Updation

Input Data : $frame$ is a 3D array (RGB) of the current frame
bg_frame - 3D array (RGB) of the current background
$temp\_bg$ - 3D array which stores the recent colour value of each pixel
$freq\_bg$ - 2D array that stores number of previous frames for which $temp\_bg$ appeared
bg_color_freq - threshold for $freq\_bg$
$mark\_up$ - 2D binary array that marks pixels rejected by HOG present in the contours
$pixels$ - number of pixels marked for updation
$bg\_cluster\_thresh$ - threshold on $pixels$ for re-clustering of the updated bg_frame
Output Data : Updated bg_frame
for each $x \in num\_rows$ **do**
    for each $y \in num\_cols$ **do**
        if $mark\_up(x, y) = 0$ **then**
            if $freq\_bg(x, y) = 0$ **then**
                for each $k \in (0, 3)$ **do**
                    $temp\_bg(x, y, k) \leftarrow frame(x, y, k)$
                end for
            else
                if $temp\_bg(x, y, k) - frame(x, y, k) < color\_thresh(RGB)$ **then**
                    for each $m \in (0, 3)$ **do**
                        $temp\_bg(x, y, m) \leftarrow avg(frame(x, y, m), temp(x, y, m))$
                        $freq\_bg(x, y) \leftarrow freq\_bg(x, y) + 1$
                    end for
                    if $freq\_bg(x, y) = bg\_color\_freq$ **then**
                        for each $a \in (0, 3)$ **do**
                            $bg\_frame(x, y, a) \leftarrow temp\_bg(x, y, a)$
                            $pixels \leftarrow pixels + 1$
                            $freq\_bg(x, y) \leftarrow 0$
                        end for
                    end if
                end if
            end if
        end if
    end for
end for
if $pixels > bg\_cluster\_thresh$ **then**
    $fuzzy(bg\_frame)$
    $pixels \leftarrow 0$
end if

Algorithm 4 HOG Computation

Input Data : I ∈ $x_1, x2_2.....x_n$ where i ∈ 1,2,...,N ,D_X ,D_Y,k ∈ [0,8]
Output Data : Feature vector f
for $i \in N$ **do**
    $I_X^i \leftarrow I * D_X$
    $I_Y^i \leftarrow I * D_Y$
    $|G| \leftarrow \sqrt{I_x^2 + I_y^2}$
    $\theta \leftarrow arctan\frac{I_y}{I_x}$
end for
for $i \in N$ **do**
    in every 8X8 cell find orientation O in $[0^0, 180^0]$ interval
    k = O % 20
    $G_i \in k^{th}$bin
end for
Normalize orienation histograms into B blocks.
for $j \in B$ **do**
    norm(vec) $f_j = \frac{v}{\sqrt{||v||_2^2 + e^2}}$
end for

Chapter 5

Result Analysis

5.1 Dataset Collection

Datasets used for testing purposes :

1. PETS Dataset
2. Manual Dataset : Videos of Group Members.

5.2 Output Analysis

Result Obtained on one of video of manual dataset.
Output Figures at every stage :

- Figure 2 : When the first frame is grabbed then it is passed on to the fuzzy cmeans algorithm which gives us the clustered image. We can set number of clusters, error and iterations for the calculation of membership matrix.
- Figure 3 : When a key frame gets identified then it is passed to fuzzy_predict function which gives the clustered image based on the cluster centres computed for the background frame.
- Figure 4 : After clustering the key frame, it is pixel wise compared with background clustered frame and pixels are classified as either foreground(i.e.

white) or background pixel (i.e. black). This image obtained from background subtraction helps to detect the motion.

- Figure 5 : It white box gives the cropped image where motion has been detected after the extraction of contours from the background subtracted image. Basically we are extracting the moving objects in the foreground.

- Figure 6 : After obtaining the cropped image we compute HOG for the image and pass it to linear svm for classifying whether it is a pedestrian or not. This figure represents the detected pedestrian.

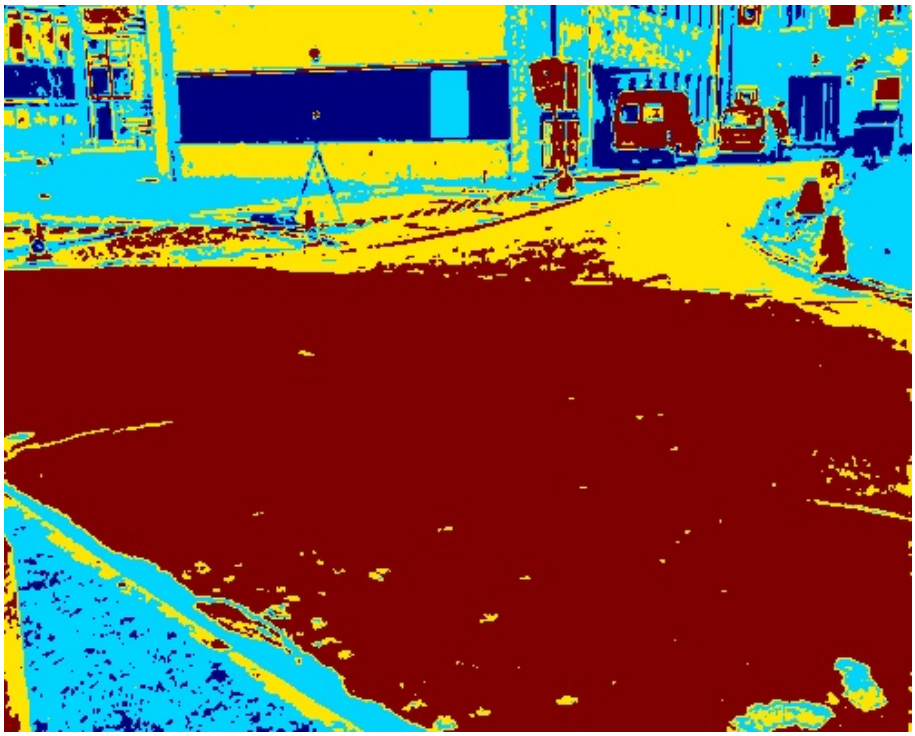

Figure 2: Clustering of Background frame

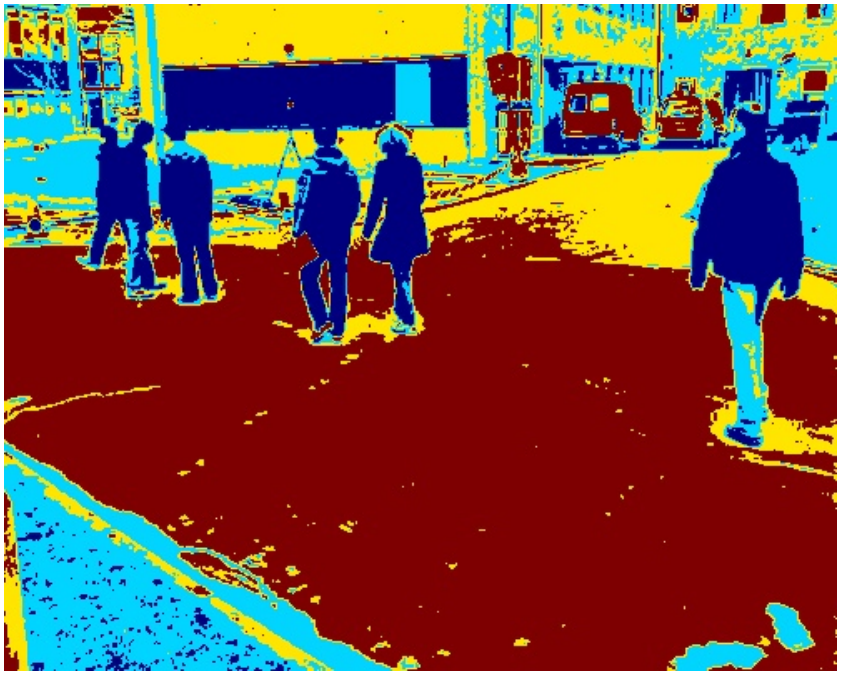

Figure 3: Clustering of Key frame

Figure 4: Background Subtraction

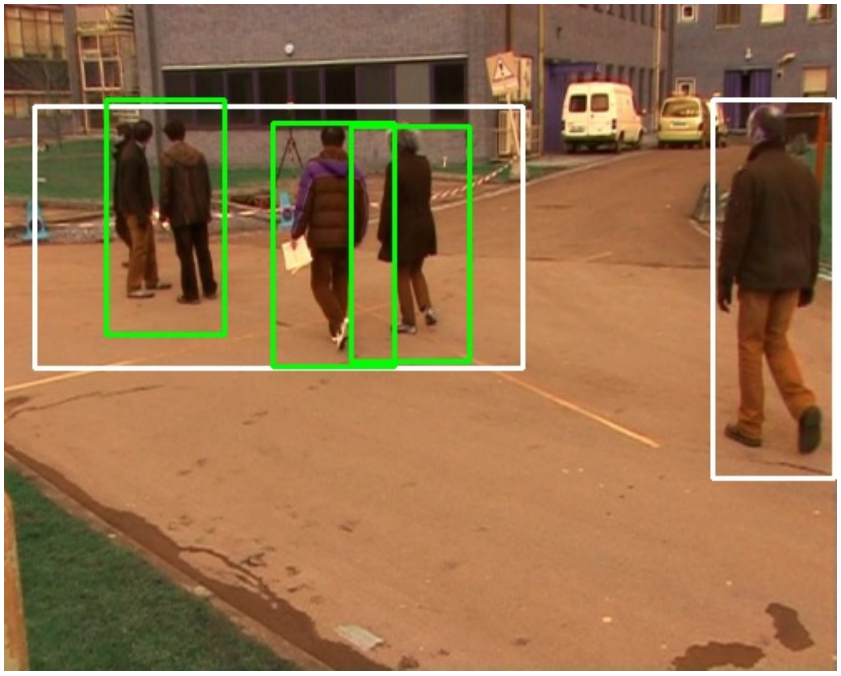

Figure 5: Motion Detection

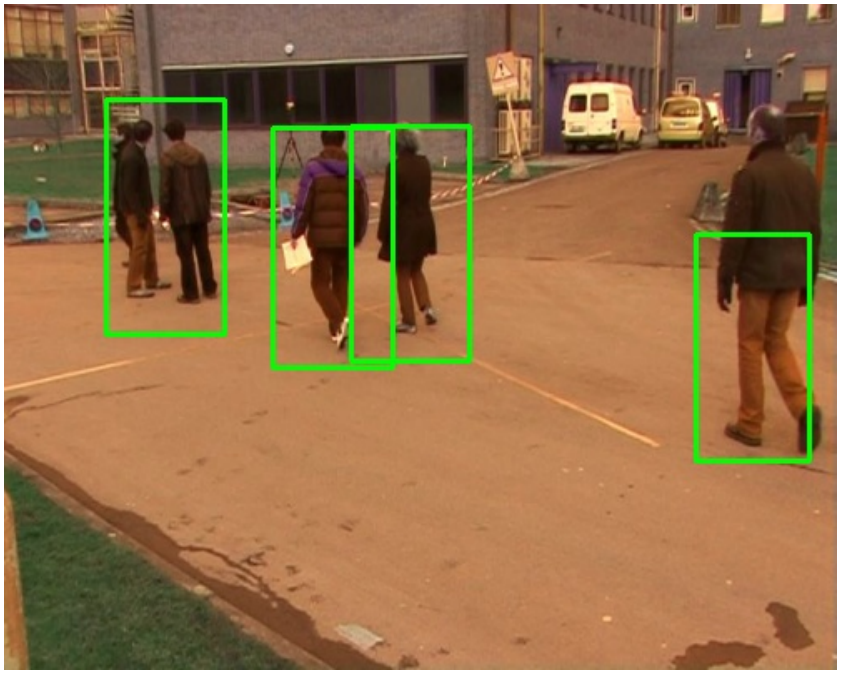

Figure 6: Pedestrian Detection

5.3 Results for PETS Dataset

In this thesis, a human detection method based on motion detection is presented. Motion detection is used to extract moving regions, which can be scanned by sliding windows. Every sliding window is regarded as an individual image region and HOG features are calculated. At last, the detected image objects can be categorized into pre-defined groups of humans and other objects by using SVM classifier.

Table 1 presents the results for motion detection in the four PETS Datasets. The total of 320 frames have been analysed and average accuracy obtained is 96.01%.

Table 2 presents the results of Human Detection in the four PETS Datasets. The total of 500 frames have been analysed and average accuracy obtained is 89.08%.

Table 3 presents the results of Human Detection by application on only HoG on each frame of the four PETS Datasets. The total of 350 frames have been analysed and average accuracy obtained is 68.61%.

Fuzzy	Datasets			
	A	B	C	D
True Positive	219	230	531	360
True Negative	0	0	0	0
False Positive	0	0	0	0
False Negative	10	0	27	26
Total No of Frames	60	50	130	80
Total No of People	229	230	558	386
Accuracy(%)	95.63	100	95.16	93.26

Table 1: Fuzzy Results

Fuzzy and HOG	Datasets			
	A	B	C	D
True Positive	229	226	972	338
True Negative	5	0	0	0
False Positive	7	0	0	0
False Negative	30	18	110	48
Total No of Frames	100	50	250	100
Total No of People	259	244	1082	386
Accuracy(%)	86.34	92.60	89.83	87.56

Table 2: Fuzzy and HOG results

HOG	Datasets			
	A	B	C	D
True Positive	227	178	469	257
True Negative	0	0	0	0
False Positive	23	39	117	14
False Negative	41	30	265	100
Total No of Frames	60	50	160	80
Total No of People	268	208	734	357
Accuracy(%)	78.00	72.06	55.11	69.27

Table 3: HOG results

Final Conclusion :

The above results clearly show that the accuracy of proposed Fuzzy and HOG approach is much better than only HOG implementation.

Figure 7 represents the screenshot of the GUI.

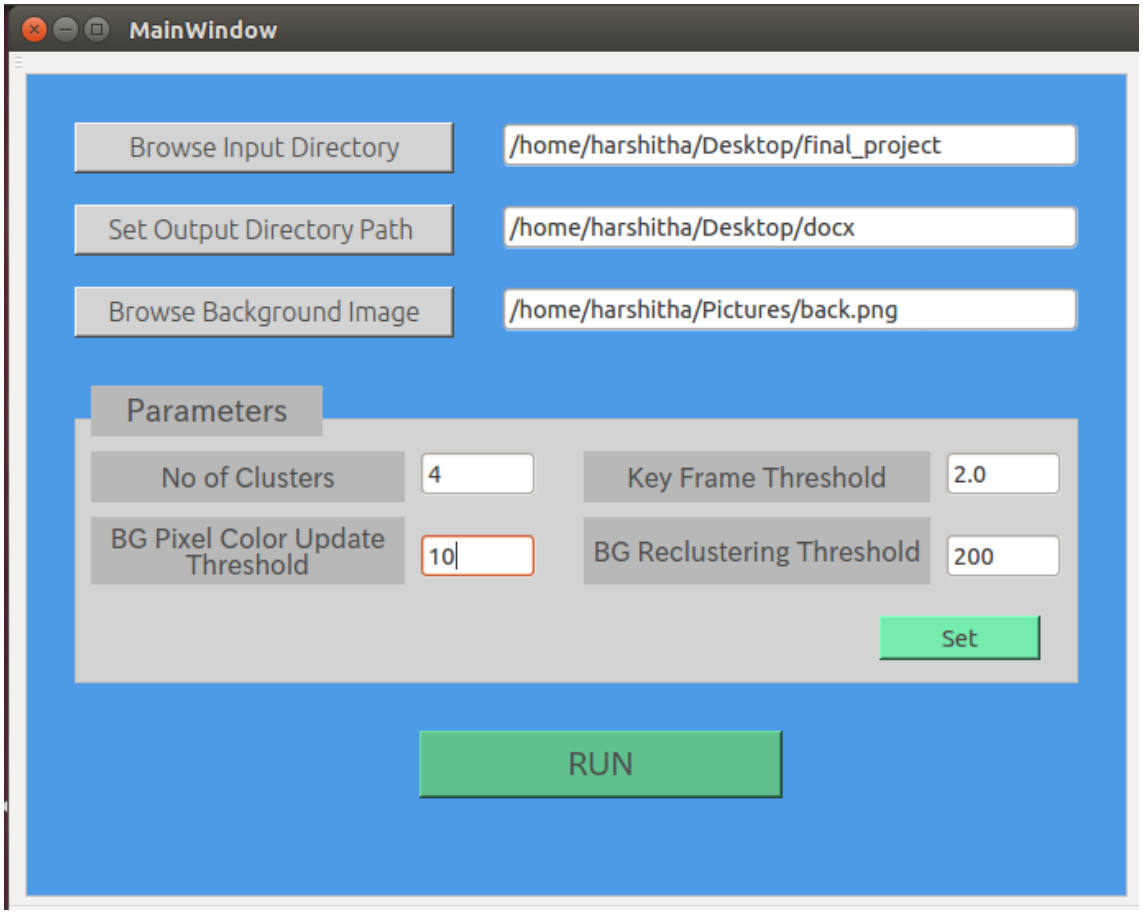

Figure 7: GUI

Chapter 6

Conclusion

6.1 Conclusion

Detecting human beings accurately in a surveillance video is one of the major topics of vision research due to its wide range of applications. It is challenging to process the image obtained from a surveillance video as it has low resolution. The detection process occurs in two steps: object detection and object classification.
We used Fuzzy cmeans clustering based background subtraction for object detection and HOG based approach for object classification. From the machine vision perspective, it is hard to distinguish an object as a human due to its large number of possible appearances. Moreover, the human motion is not always periodic, but a combination of features could be useful in identifying humans.
By employing fuzzy based clustering the false positive rate of detection was significantly low at high true positive rates. Hence we could efficiently detect changing areas. This in turn reduced the load on HOG feature computation and made the human detection system more robust. The Future Direction of work would be to track the trajectory of human detected and improvisation of efficiency and time complexity of present system.

References

[1] Dalal, N., and Triggs, B. Histograms of oriented gradients for human detection. In *CVPR* (2005), pp. 886–893.

[2] Deepak Kumar Panda, S. M. Detection of Moving Objects Using Fuzzy Color Difference Histogram Based Background Subtraction. *IEEE* (2015).

[3] Duc Thanh Nguyen, Wanqing Li, P. O. O. Human detection from images and videos. *JOURNAL OF COMPUTERS 6*, 8 (2011).

[4] Hou Beiping, Z. W. Fast Human Detection Using Motion Detection and Histogram of Oriented Gradients.

[5] Hsu, C.-W., and Lin, C.-J. A Comparison of Methods for Multi-class Support Vector Machines.

[6] Navneet Dalal, B. T., and Schmid, C. Human Detection using Oriented Histograms of Flow and Appearance. *IEEE* (2005).

[7] Subrata Chakraborty, S. M. E. H. Human detection in surveillance videos and its applications - a review. Master's thesis, 2013.

[8] Viola, P., and Jones, M. Rapid object detection using a boosted cascade of simple features. pp. 511–518.

[9] Yld, C. An implementation on histogram of oriented gradients for human detection.