CS771A (Machine Learning: Tools, Techniques and Applications)
course project report

Foreground-background classification and ROI detection in surveillance videos

Group 10
Dheeraj Mekala (13405)
Keerti Anand (13344)
Aakash Ghosh (12003)
Saurabh Kataria (12807637)

Supervisor:
Prof. Harish Karnick

April 15, 2016

Foreground-background classification and ROI detection in surveillance videos

Dheeraj Mekala (13405), Keerti Anand (13344), Aakash Ghosh (12003), Saurabh Kataria (12807637)

Abstract

With the ever growing demand for storage and careful observation from surveillance footage, it has become necessary to devise methods that can efficiently perform surveillance tasks in a fully automated fashion. In this project report, we describe two such tasks that we have dealt with. First, classification of a video frame as foreground or background, and secondly, detecting region of interest (ROI)/ foreground object in a foreground frame. For the former problem, we have experimented with four methods (frame differencing, GMG, weighted moving averages, and adaptive selective background learning) from BGS library, two from OpenCV library (methods of gradient-I, and II), and one standalone packaged algorithm called ViBe algorithm. Evaluation has been performed with frame decision accuracy measure. For the second problem, we have manually labelled the ROI regions for 400 frames, and used it as a testing set. The ROI regions detected from above mentioned algorithm are bounded by rectangles, and compared with ground truth (self labeled set) and overlap accuracy is evaluated. We conclude by finally mentioning the challenges we faced and the possible extensions in future.

Keywords: Machine learning, background subtraction, region of interest

1. Introduction

In this Project, we aim to study frames from the security footage obtained from the surveillance cameras at the Institute Gate and classify them as foreground or background. We do this using many algorithms, which are mentioned in the Methodology section. These algorithms try to label each pixel in a frame as Foregound(White) and Background(Black). If there are a considerable number or foreground pixels in a frame, we label it as Foreground otherwise we call it Background. We also calculate the accuracy metric of foreground detection and in the second part, bound the foreground pixels in a particular frame with the best-fit box and calculate its overlapping with the corresponding annotated image. Hence, we wish to find out how do different FG-BG seperation algorithms perform on the given dataset, and how efficiently are they able to detect the ROI correctly.

2. Motivation

Foreground-Background Sepeartion finds use in many areas of Computer Vision - Video Synopsis, Region of Interest (ROI) classification, and object tracking - to name a few. Consider how a large dataset of video footage could be trimmed down by discarding the unimportant background frames (which are of no interest) and keeping only the foreground frames. Hence, we save on tape storage. Also, we can keep track of number plates of vehicles entering the campus and also find out the people who are entering the campus using facial recogonition.

3. Accuracy measure

There are 2 accuracy measures used in our project.

3.1. Frame Decision Accuracy

We defined the frame with non-static objects as Foreground frame and the frame with no non-static objects as Background frame. The accuracy measure for this classification of frames is called Frame Decision Accuracy. Small movements of non-static objects such as tree leaves were also detected, due to which the frame may be classified as a foreground frame which we don't want to. We changed this definition of foreground frame as: "The frame with at least one non-static object bounded by a rectangle with area more than a threshold value is called Foreground frame."
Mathematically:

$$Accuracy = \frac{\text{Number of correctly classified frames}}{\text{Total number of frames}} \tag{1}$$

3.2. Bounding box overlap accuracy

The bounding box overlapping was calculated by the following method: First, we bound the foreground pixels in the output frame(A) using the best-fitting rectangles. Then, we take a look at the rectangle in the corresponding annotated frame(B) (Ground Truth), we then calculate the overlapping area and take its ratio with the area of the rectangle of the ground truth. Since, there can be many rectangles, we choose that rectangle in the output frame, which has the best overlap.

$$Accuracy = \frac{S_{overlap}}{S_B} \times 100 \tag{2}$$

4. Description of dataset and methods used

From the IITK surveillance video data that was made available to us, we have used 'dec21h1330.dat' (duration = 00:02:05) for implementing/ evaluating the performance for both the tasks. The video has three channels, a high resolution of 2048 × 1536, and frame rate of 18 per second. Using the video codecs of ffmpeg module of FFmpeg library, the video was preprocessed from .dat format to .mp4 format for further analysis. Using ffpmeg again, we have extracted individual frames from our video. For the first task of foreground-background frame classification, we have used the first 2255 frames and labeled them by 0 (background frame) and 1 (foreground frame) manually. Now, using the methods (described later in this section), frames are being classified as 0 or 1. Final accuracy is calculated by the frame mismatch ratio. For the second task, we have hand labeled the first 400 frames of the video for testing set for creating the ground truth. Using the MATLAB' s "trainingimagelabeler" API, we have approximately marked every foreground object/ ROI by a rectangular bounding box. Using the same methods as before, we adaptively detect ROI in those 400 images. Using the contour library of OpenCV, we have fitted the detected ROIs with bounding boxes so that they become eligible to be compared with the ground truth. Finally, we use the bounding box overlap metric to evaluate the performance of ROI detection.

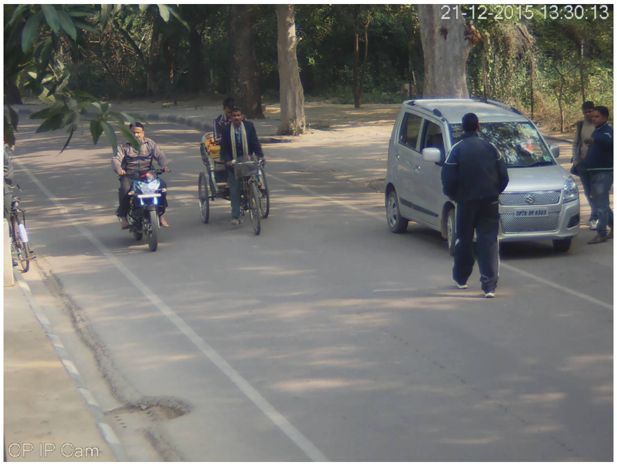

Figure 1: Sample Image from the dataset

4.1. Frame Differencing

Frame differencing is one of the most rudimentary techniques of background subtraction used. In this method, we check the difference in pixel intensities between two video frames. A change in intensity of the pixels implies some change in the image, and one of the key reasons for the change is movement.

$$|frame_i - frame_{i-1}| > Threshold \tag{3}$$

However, this method is susceptible to change in lighting conditions, camera auto-focus, brightness correction and other issues. Hence, in practice, a threshold value is used to distinguish real movement from noise.

Another possible disadvantage is that since it uses only a single previous frame, frame differencing may not be able to identify the interior pixels of a large, uniformly-colored moving object.

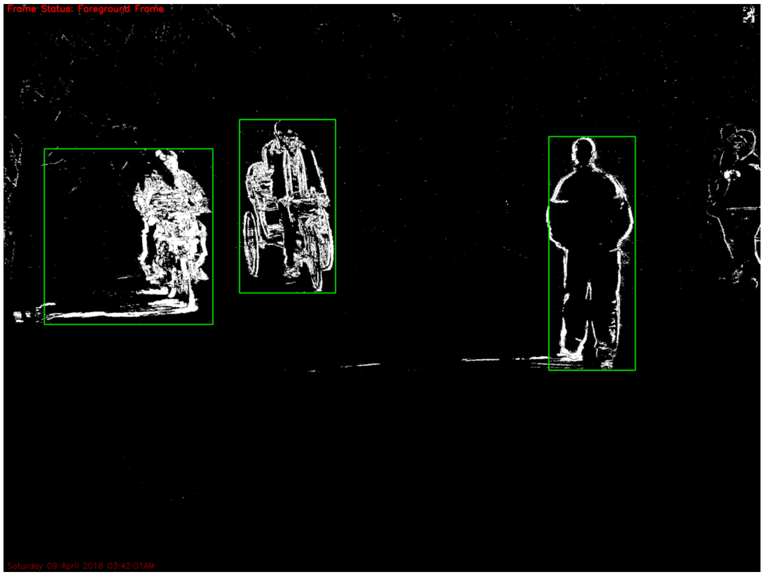
Frame Status: Foreground Frame
Saturday 09 April 2016 03:42:01AM

Figure 2: Frame Differencing

Classification Accuracy	94.46%
ROI Detection Accuracy	88.04%

4.2. GMG (Global Minimum with a Guarantee) Algorithm

This algorithm uses statistical background image estimation along with the method of per-pixel Bayesian segmentation. The first few frames, usually 120, are used for background modelling. It employs probabilistic foreground segmentation algorithm that identifies possible foreground objects using Bayesian inference. The estimates are adaptive in the sense that newer observations are given more weight than old observations to consider variable illumination. Several morphological filtering operations like closing and opening are done to remove unwanted noise.

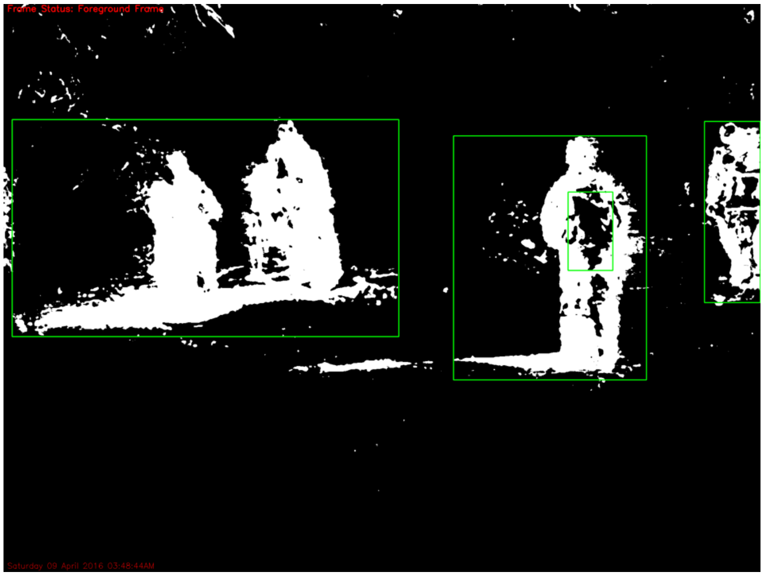
Frame Status: Foreground Frame
Saturday 09 April 2016 03:48:44AM

Figure 3: GMG Algorithm

Classification Accuracy	91.70%
ROI Detection Accuracy	94.54%

4.3. Weighted Moving Averages

In this algorithm, the background model at each pixel location is based on each pixel's recent history. A weighted average is used to model the algorithm, where recent frames have higher weight, and hence more importance than previous frames. The background model is thus computed as a chronological average from the pixel's history. In this method, no spatial correlation is used between different (neighbouring) pixel locations.

The average is usually taken to be a Gaussian average, and a Gaussian probability density function is fitted on a fixed number of the most recent frames.

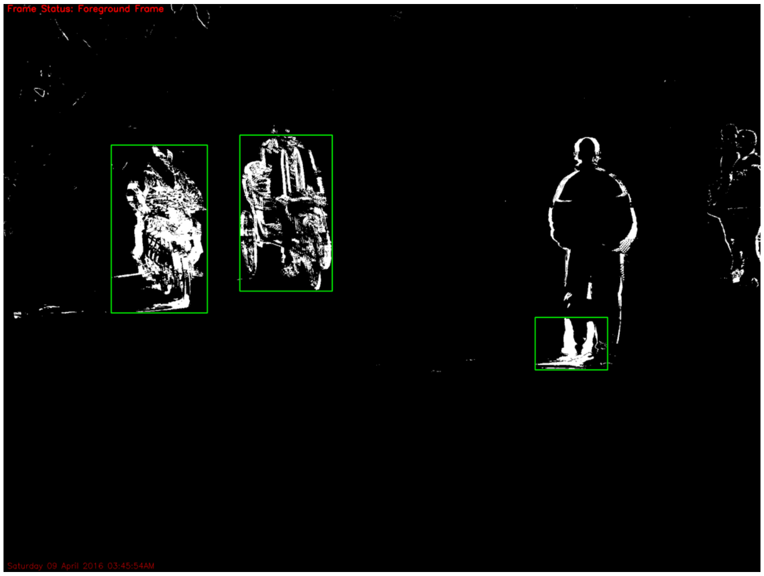
Frame Status: Foreground Frame
Saturday 09 April 2016 03:45:54AM

Figure 4: Weighted Moving Averages

Classification Accuracy	87.09%
ROI Detection Accuracy	87.42%

4.4. Adaptive Selective Background Learning

This method was implemented for experimental purposes. Specifically, we wished to see the effect of adaptive background subtraction (ghost elimination), and shadow removal. As can be observed from the output, the algorithm is adaptive. Ghost removal is quite effective here. Also, the shadows are not affecting the bounding box result that we desired.

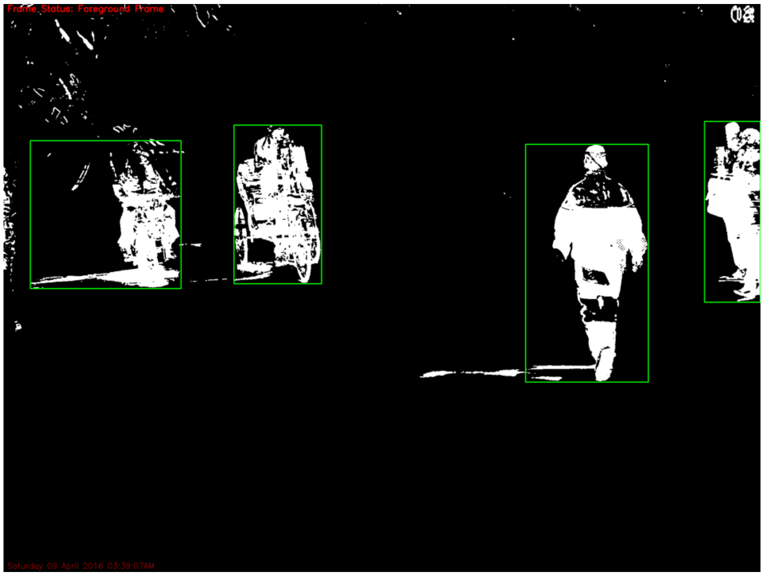
Frame Status: Foreground Frame
Saturday 09 April 2016 03:39:07AM

Figure 5: Adaptive Selective Background Learning

Classification Accuracy	81.02%
ROI Detection Accuracy	91.18%

4.5. Method of Gradients - 1

It is a Gaussian Mixture-based Background/Foreground Segmentation Algorithm. It uses a method to model each background pixel by a mixture of K Gaussian distributions (K = 3 to 5). The weights of the mixture represent the time proportions that those colours stay in the scene. The probable background colours are the ones which stay longer and more static.

This method can cope with multimodal background distributions.Each pixel modeled with a mixture of Gaussians and thus it is flexible to handle variations in the background.

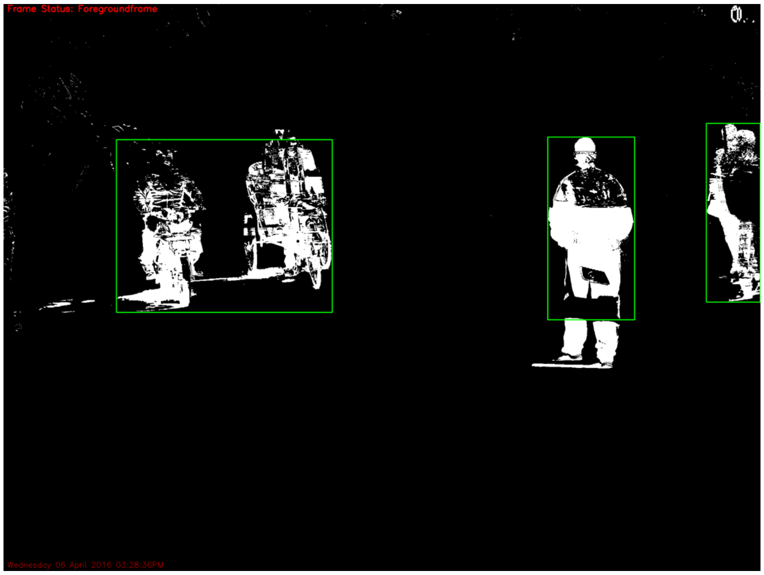
Frame Status: Foregroundframe
Wednesday 06 April 2016 03:28:36PM

Figure 6: Method of Gradients - 1

Classification Accuracy	97.44%
ROI Detection Accuracy	84.09%

4.6. Method of Gradients - 2

It is also a Gaussian Mixture-based Background/Foreground Segmentation Algorithm. A key difference from the previous algorithm is that it selects the appropriate number of gaussian distribution for each pixel. It provides better adaptibility to varying scenes due illumination changes etc.

It provides an option of selecting whether shadow to be detected or not. If so programmed, it detects and marks shadows, but decreases the speed. Shadows will be marked in gray color.

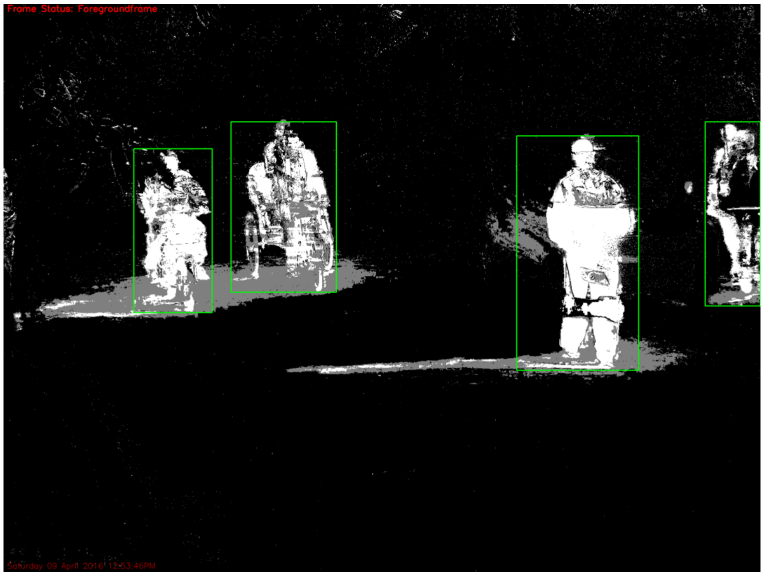
Frame Status: Foregroundframe
Saturday 09 April 2016 12:53:46PM

Figure 7: Method of Gradients - 2

Classification Accuracy	90.93%
ROI Detection Accuracy	80%

4.7. ViBE (Visual Background Extractor) Algorithm

ViBE is a stand-alone software package used for background subtraction from moving images. Several techniques are employed to provide an estimate of the temporal probability density function of a pixel, the discussion of which is beyond the scope of the project. ViBe's , in contrast, imposes the influence of a value in the polychromatic space to be limited to the local neighborhood.

In practice, ViBe does not estimate the pdf, but uses a set of previously observed sample values as a pixel model. It thus employs methods of nearest neighbour classification and its variations.

A clear advantage of ViBE is that it is able to produce spatially coherent results directly without the use of any post-processing method.

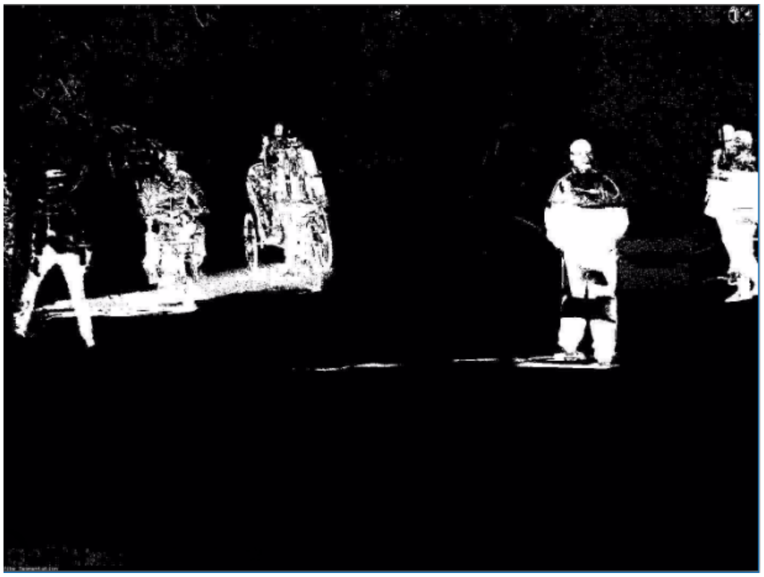

Figure 8: ViBE Algorithm

5. Challenges in Implementation

5.1. During Data Pre-processing

- The frames were not labelled properly. So, we labelled about 400 frames manually for computing bounding box overlap accuracy and about 2300 frames for computing frame decision accuracy.

- We used FFMPEG for video analysis where we faced some difficulties in converting the video from .dat to .avi format.

- There were some compatibility issues between OpenCV version-2 and version-3.

5.2. During Algorithm Runs

- **Ghost elimination**
  There is a history parameter. Let this be 'd' which keeps track of the grayscale value of a particular pixel in the previous 'd' frames. This history parameter results in formation of ghost which results in poor frame decision accuracy. We tuned this history parameter to improve the accuracy.
  We obtained 500 as the optimum history parameter.

- **Shadow elimination**
  There is a parameter which is a measure of the effect of shadow in the background separated grayscale image. We tuned this parameter to improve the accuracy. The bounding box overlap accuracy improved with this shadow parameter where as the frame decision accuracy decreased. So, we stopped tuning this parameter at an intermediate value.

- **Quality of image**
  Quality of image effects the bounding box overlap accuracy as the bounding box is constructed based on the grayscale values of the pixels. There is a parameter named threshold parameter which classifies the pixel values. This classification effects the quality of image which in turn effects the bounding box overlap accuracy. We tuned this parameter to improve the bounding box overlap accuracy.
  We obtained 127 as optimum threshold parameter.

- **Unable to save output of VIBE algorithm**
  There were resolution issues with VIBE algorithm. VIBE algorithm is computationally expensive than the other algorithms we implemented. We were able to run VIBE algorithm on the video input with low resolution where we didn't obtain appreciable accuracy.

- **Multiple rectangles bounding the same region of interest**
  There were more than one rectangle bounding the parts of region interest simultaneously. We were unable to fix the number of rectangles bounding an object of interest.

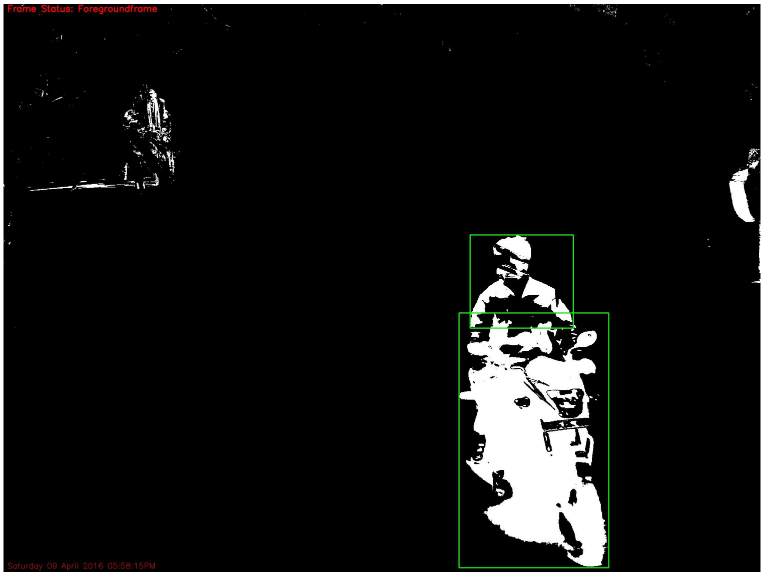
Frame Status: Foregroundframe
Saturday 09 April 2016 05:58:15PM

Figure 9: Multiple Bounding Boxes

- **Noise in the video**
  Small movements of non-static objects such as tree leaves were also detected which is not needed. We avoided the detection of these movements by tuning the threshold parameter as described above and by putting area constraint on the object of interest detected by the algorithm as described in section 4.1.

6. Possible Improvements

- **Object tracking**
  We can increase the accuracy of ROI detection by keeping track of previously detected objects. Justification of doing this is: when a ROI is detected at some location, there is a high probability of it being there in next frame too. Methods based on optical flow can be employed here.

- **Increase in performance**
  We could have improved the accuracy of certain algorithms by enhancing accuracy of both ground truth and by rigorous tuning of parameters.

- **Labelling of object of interest**
  As we have obtained the region of interest bounded by a rectangle. We could have labelled the object as 'person', 'vehicle'...

- **Face detection**
  As we have obtained a good accuracy in the region of interest detection, face detection is not too far to implement.

7. References

- http://opencv.org/
- https://www.ffmpeg.org/
- https://github.com/andrewssobral/bgslibrary
- http://www.telecom.ulg.ac.be/research/vibe/
- http://in.mathworks.com/help/vision/ug/label-images-for-classification-model-training.html