Group 20

Video Surveillance Project

CS771A: Machine Learning Techniques

OVERVIEW

We are given a dataset consisting of the day time surveillance videos from the security cameras of IIT Kanpur from the Main Entrance Gate of the Institute and we are supposed to use different Machine Learning Techniques to achieve the objectives and report the results.

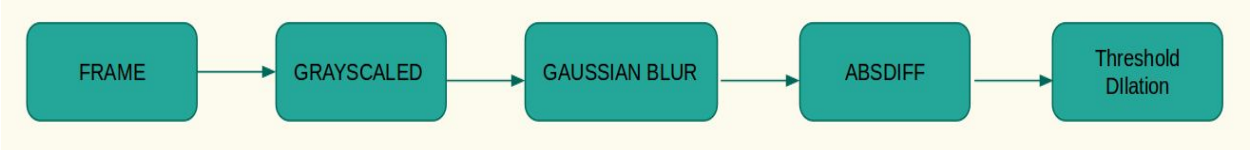
OBJECTIVES

1. Separate foreground frames from background frames in real time with high accuracy.
2. Classify different objects in the videos - pedestrians, vehicles, etc..

Background Separation

Approach

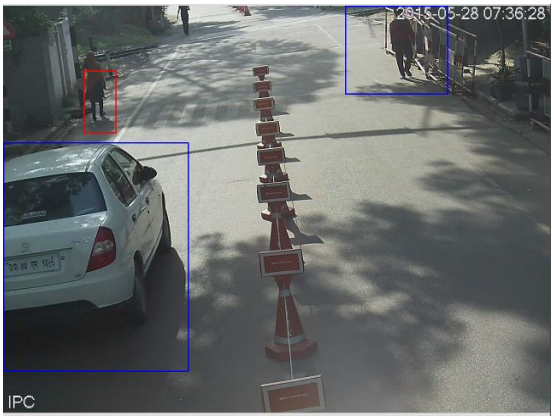
For real time background separation for every video we save the first frame and for the consecutive frames we convert it into grayscale then apply gaussian blur to it so as to smoothen the image and then we take absolute difference of it from the first frame. On the resulting image we apply threshold dilation so as to fill in the gaps and increase the brightness of the detected area.



One of the problem with this approach is that if first frame of the input video has some objects then those objects are detected as background and result in bad results.

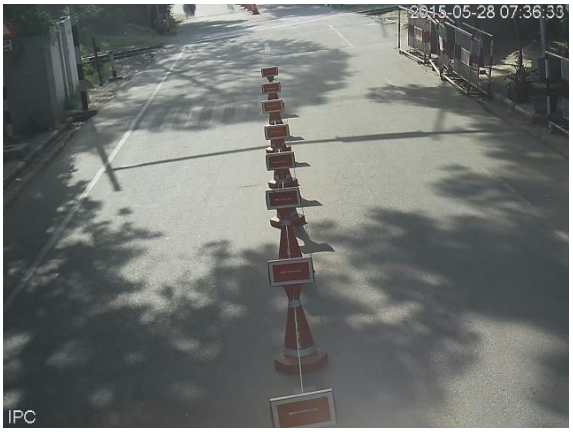
To overcome this problem we tried different approach in which we take temporal median of every pixel to generate background and then we take the absolute difference of the frames from this generated background.

Results



Frame from datasample1.mov

Background Subtracted frame



Background generated using temporal median

First frame

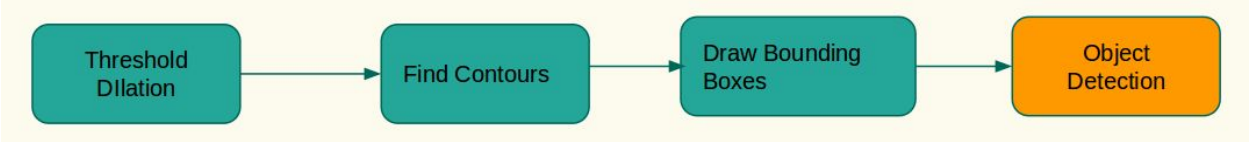
No objects from the first frame

Objects from the first frame create hindrance

Object Detection

Approach

For object detection we take the background separated frame after the threshold dilation of the frame and find contours in this frame and draw the bounding boxes around these contours.



Error Measure

For Object Detection, accuracy is defined as the total area of intersection of training bounding box and testing bounding box divided by the total average area of training and testing bounding boxes.

Results

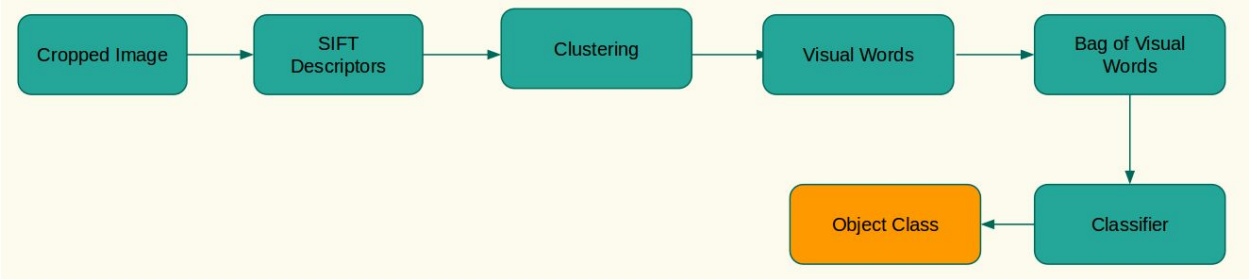
Background Separation with **Temporal Median** as a background, we get an accuracy of **68.2%**

Background Separation with **First Frame** as a background, we get an accuracy of **48.2%**

Object Classification

Approach

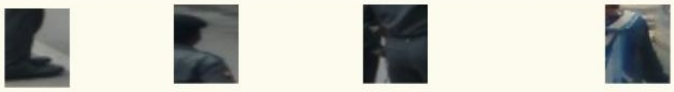
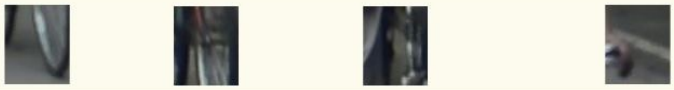
For training a classifier we crop the image of the labelled objects and find their SIFT descriptors. The SIFT descriptors thus obtained are clustered using K-Means to get a vocabulary of Visual words. These Visual words are then used to get a fixed size feature vector of each object and then a classifier is trained using these feature vectors and labels.



The trained classifier is then used to classify objects from the video in each frame as and when the objects are detected.

Visual Words

The following visual words were obtained for different classes of objects.

Person	
Motorcycle	
Car	
Bicycle	
Rickshaw	

Results

Adaboost Classifier with maximum 100 estimators

Classification	Accuracy
Person vs Rest	68.8%
Motorcycle vs Rest	78.0%
Car vs Rest	61.0%
Bicycle vs Rest	56.4%
Person vs 2-wheeler vs 4-wheeler	50.0%

Adaboost Classifier with maximum 500 estimators

Classification	Accuracy
Person vs Rest	71.9%
Motorcycle vs Rest	78.2%
Car vs Rest	57.4%
Bicycle vs Rest	53.0%
Person vs 2-wheeler vs 4-wheeler	46.8%

Random Forest Classifier

Classification	Accuracy
Person vs Rest	64.0%
Motorcycle vs Rest	74.4%
Car vs Rest	60.0%
Bicycle vs Rest	50.3%
Person vs 2-wheeler vs 4-wheeler	45.4%

Support Vector Classification using RBF kernel

Classification	Accuracy
Person vs Rest	73.9%
Motorcycle vs Rest	78.4%
Car vs Rest	43.3%
Bicycle vs Rest	49.3%
Person vs 2-wheeler vs 4-wheeler	38.6%

Support Vector Classification using Sigmoid kernel

Classification	Accuracy
Person vs Rest	73.9%
Motorcycle vs Rest	78.4%
Car vs Rest	43.3%
Bicycle vs Rest	49.3%
Person vs 2-wheeler vs 4-wheeler	38.6%

References

1). Scikit-learn

2). Open-CV

3). PyImage Object Detection

4). SIFT