CS771: Machine Learning: Tools, Techniques, Applications
Project Report

Group 36
Deepak Kumar (13229)
Shivam Chaturvedi (13657)
Divyanshu Shende (13264)
Gowtham Sai (13463)

April 15, 2016

1 Problem Statement

The aim of the project is to classify objects into pedestrian, 2/3/4 wheelers etc from the given CCTV footage. The problem also includes first separating foreground frames from the background frames to get images of objects for label prediction in real time.

2 Our Approach

We broke down the problem into 3 steps. Each of them is explained below.

2.1 Feature Extraction

The first task was to vectorize the data given which means crating a *feature vector*. So features had to extracted from image data on which we could train the classifier. This was done by using algorithms like **SIFT**(Scale-Invariant Feature Transform) and **SURF**(Speed-Up Robust Features). These algorithms allowed us to extract key-points from image data. We then formed a K-means clustering(with 200 clusters) of these key-points over the entire collection of images. This allowed us to create 200-dimensional feature vector for each image by noting the number of key-points in the image that lie in a particular cluster.

2.2 Classification

The next step was to build a classifier. This was done using labelled image data obtained through crowd-sourcing. The feature vectors were obtained using the procedure described above. The number and type of classes depended on the type of classification : *coarse* and *fine*. In coarse classification, the classes were *Person, 2-wheeler, 3-wheeler, 4-wheeler*. In fine classification, the classes were *Person, Bicycle, Motorbike, Car, Autorickshaw, Rickshaw*.

For each of the feature extraction algorithms and for each type of classification, we used two different classifiers: *Support Vector Machines* and *Random Forest Classifiers*. These were trained on vectorized and labelled data.

2.3 Foreground Extraction and Video Labelling

This was one of the key steps. After training the classifier, we needed to use it in real time. This required extracting regions of interest(ROI) in the video. This was done by splitting the video into frames and then extracting ROI's in the frame. The *BackgroundSubtractorMOG2()* method of OpenCV enabled us to extract the foreground which we split into ROI using contours. The co-ordinates of each ROI in the frame were obtained and the corresponding image was cropped and sent to the classifier. The classifier then assigned a label to the image which was used for labelling the ROI in the original video.

3 Libraries and Tools Used

The following libraries were used in the project.

1. **Scikit Learn** - For Random Forest Classifier, SVM and K-Means
2. **OpenCV Python** - For feature extraction, image vectorization, foreground extraction.

4 Results

The results for *coarse classification* are mentioned below. Note that we first split all videos into frames and then partitioned the obtained images into labelling and testing data. This explains the high accuracies obtained below.

4.1 SIFT Features and SVM classifier

The table below shows kernel type versus accuracy for different values of c (inverse of margin in SVM) using SIFT features.

Kernel	**c = 0.1**	**c = 1**	**c = 10**
Linear	84.25	85.25	81.41
Polynomial	66.0	81.41	88.08
RBF	78.83	92.58	97.33

Best Kernel - RBF
Best value of $c = 10$

4.2 SURF Features and SVM classifier

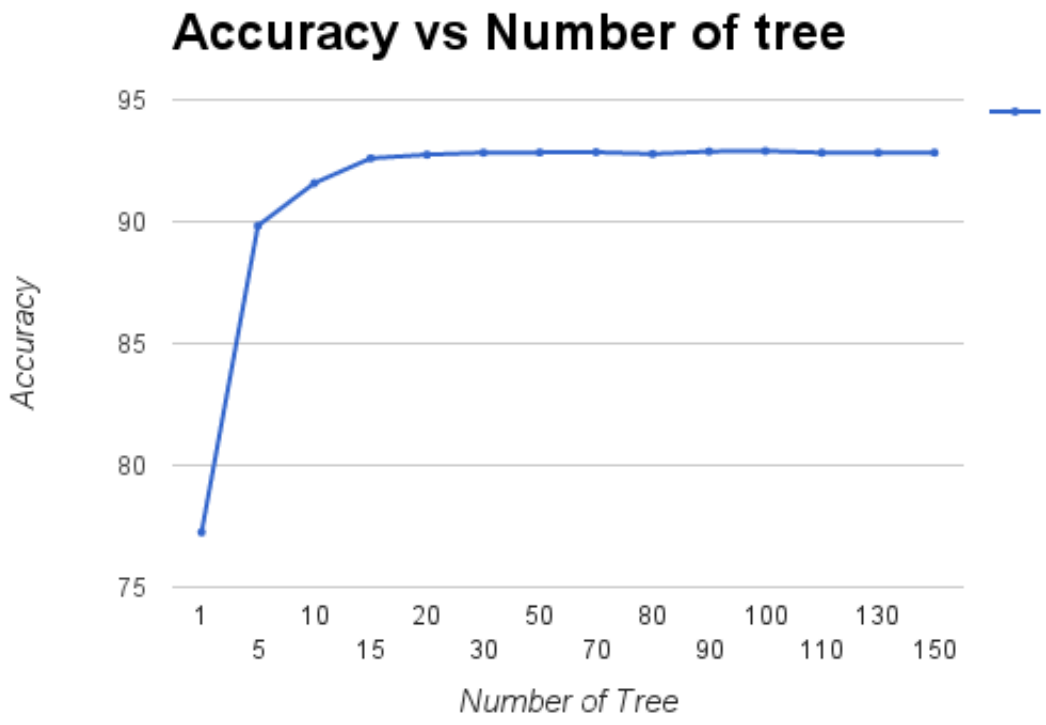
The table below shows kernel type versus accuracy for different values of c (inverse of margin in SVM) using SURF features.

Kernel	**c = 0.1**	**c = 1**	**c = 10**
Linear	89.91	89.33	86.75
Polynomial	82.16	86.50	90.33
RBF	59.25	94.33	97.83

Best Kernel - RBF
Best value of $c = 10$

4.3 SIFT Features using Random Forest Classifier

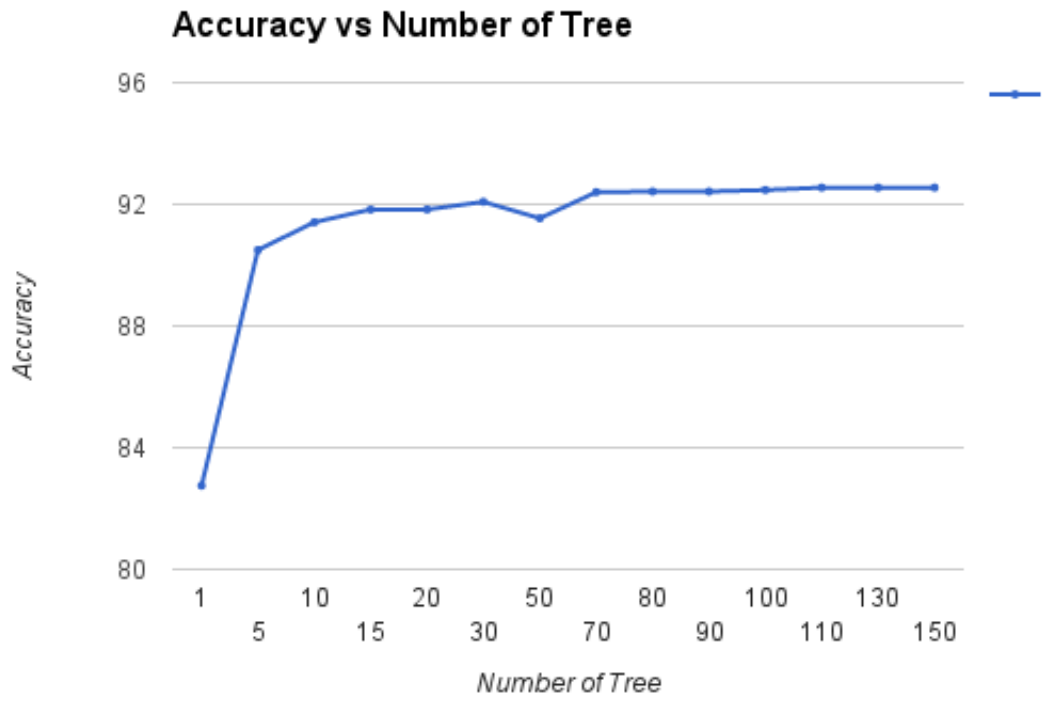
The Random Forest Classifier was used while working with SIFT features. The chart below shows accuracy as a function of the number of trees in the forest.



The accuracy increases on increasing the number of trees.

4.4 SURF Features using Random Forest Classifier

The Random Forest Classifier was used while working with SURF features. The chart below shows accuracy as a function of the number of trees in the forest.



The accuracy increases on increasing the number of trees.

4.5 Conclusion

Both SIFT and SURF features are close to each other in comparison. Also, SVM and Random Forest, both performed well with both types of features. The results of both feature extraction algorithms and both classifiers were almost the same.

5 Challenges Faced

1. Incorrectly labelled data.
2. Shadows in images. This is because the area has lots of trees.
3. Occluded objects made classification difficult.

4. For bicycles and motorcycles, data labelling was done by cropping person with the vehicle. But background foreground separator draw contours around entire bicycle + person. Since our classifier was trained only for bicycle, it was confusing object with person and bicycle.

6 Possible Improvisations

1. Our classifier sometimes predicts different labels on different times for the same object. This could be prevented by using the fact that images the images fed to the classifier are in fact of the same object since it is the same video. This would improve the quality of labelling.

2. Other features can be used like HOG, ORB and Caffe.