HOG BASED VEHICLE DETECTION USING CLASSICAL SUPERVISED LEARNING TECHNIQUES

Jenish Raj Bajracharya, Himanshu Thakur

December 14, 2021

ABSTRACT

With the advent of self-driving cars, accurate detection of on-road vehicles is a prime challenge in computer vision. This paper proposes a vehicle detection system on a static image frame. The proposed method employs Histogram of Oriented Gradients(HOG) for feature extraction and linear SVM, Naive Bayes, Decision Trees for classification. We study the influence of various color spaces on the performance of the classifier along with an robust analysis on the effects of the HOG parameters.

Keywords HOG, SVM, Naive Bayes, Decision Trees, Vehicle Detection

Code Repository: GitHub

1 INTRODUCTION

One of the crucial components of a driverless car is its ability to detect obstacles on roads such as pedestrians, lanes, other vehicles, and objects. As a result, the technology can be extensively used to avoid fatal accidents and maintain safety. But the accuracy and speed of vehicle detection are crucial for practical implementation.

The HOG(Histogram of Oriented Gradients) based method, along with an SVM(Support Vector Machine) or Naive Bayes classifier is a popular approach. HOG makes use of gradient information present in each pixel cell. Since a vehicle has explicit edges in x and y direction, the HOG can be a good feature extractor. Similarly, SVM and Naive Bayes are widely used classifiers.

Thus, we develop a pipeline for detecting vehicles using HOG and a classifier. Our aim with this project is to study the applicability of HOG as a potential for vehicle detection. We experiment with the HOG parameters and the color spaces to find the optimal feature extractor and use it to train our classifier.

2 LITERATURE SURVEY

In 2005, Dalal and Triggs [1] described the HOG-based method for human detection. In their breakthrough paper, they experimented with various parameters for optimal feature extraction. The values they found closely resemble the values we found optimal for our project. Then in 2010, Creusen, Wijnhoven, Herbschleb, *et al*, [2] experimented with different color spaces for traffic sign detection. The paper concluded that the choice of color space significantly influences the performance of the HOG. The optimal color choice depends on the type of object being detected. For example, the HSV and RGB color spaces are less suitable for traffic sign detection and detection of objects with high color variation. In their experiment, they also concluded that the saturation and intensity of a single color channel are irrelevant because the performance along one channel is almost identical to the performance in the entire color space. Furthermore, they found that YCrCb and LAB color space provides the best performance.

Lately, deep learning techniques have gained mainstream. Ding *et al.* [3] presented an approach by extracting vehicle semantic key-points and using a CNN(Convolutional Neural Network). Later, Satar and Dirik [4] proposed combining SSD(Single Shot Multibox Detector) model with a CNN. Similarly, Alfasly et al. [5]proposed a two-fold framework using variational feature learning and LSTM (Long Short Term Memory) to learn the relationship of vehicles from different points of view. Recently, Fast R-CNNs and YOLO introduced by Redmon *et al.* [6] are used extensively for object detection and are considered state of the art methods.

3 Methodology

3.1 ABOUT THE DATASET

The labeled images of vehicles and non-vehicles come from the GIT vehicle image data database [7] and the KITTI vision benchmark suite [8]. The labeled images contain 8968 non-vehicles and 8792 vehicles. All the training images have the same dimension (64×64), so cleaning is not required. The vehicle images were taken under different lighting conditions varying the orientation. On the other hand, the non-vehicle images consist of roads, trees, traffic signs, pavements, and asphalts.

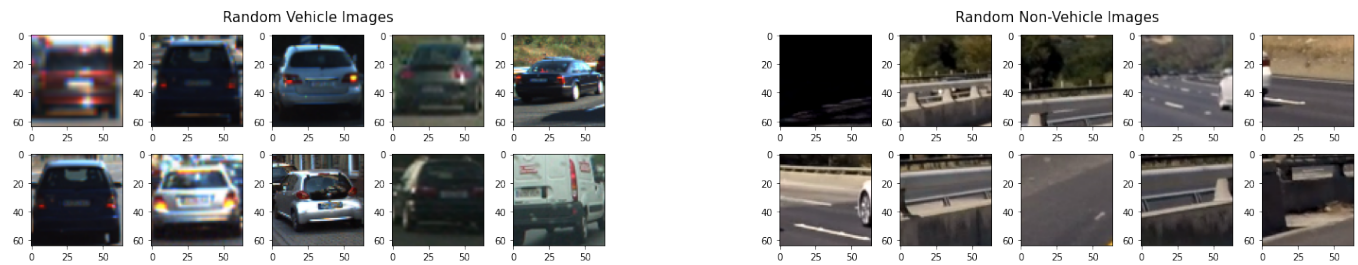


Figure 1: Random Vehicle and Non-vehicle Images

3.2 HOG and Feature Extraction

After the image preprocessing stage, features for vehicles and non-vehicles are extracted using HOG descriptors. In HOG, pixels from an image are grouped into small cells and for each cell, the gradient direction is computed and grouped into a specified number of oriented bins, and the sum is calculated. Stronger gradients contribute more weights to their respective bins thus, noise is reduced. Performing this for all the cells gives a structural representation of the image. We used skimage's HOG function and experimented with its parameters and the color space to extract the features.

3.3 Model Training

The input to our training pipeline is the feature vectors. The length of each feature vector is fixed because our training images are of the same size. Since this is a two-class classification problem we used three different ML models(SVM, Naive Bayes, Decision Trees). For each learning algorithm default parameters along with the feature, vectors were passed.

3.4 Vehicle Detection

The last step is to perform vehicle detection on a static image frame. For this, we used the sliding window search with varying window sizes. The window sizes decreased for the farthest portion of the image because the size of the vehicles is relatively small when they are at a distance from the camera.

4 RESULTS

4.1 HOG Parameters

In skimage's HOG function, we can specify the color space, number of orientations, pixels-per-cell, and cells-per-block for computing the HOG features across a single color channel. Since feature extraction is a crucial part of the pipeline getting the right set of parameters helps reduce false positives. To figure out the best hyper-parameters, we experimented with a couple of commonly used values. Upon experimentation, we concluded that YCrCb color space, pixels-per-cell of 10 with the orientation of 10 and block-size of 2 worked best because more features were preserved [Figure 5]. Although, most literature suggests pixels-per-cell of 8 with an orientation of 9 and block-size of 2.

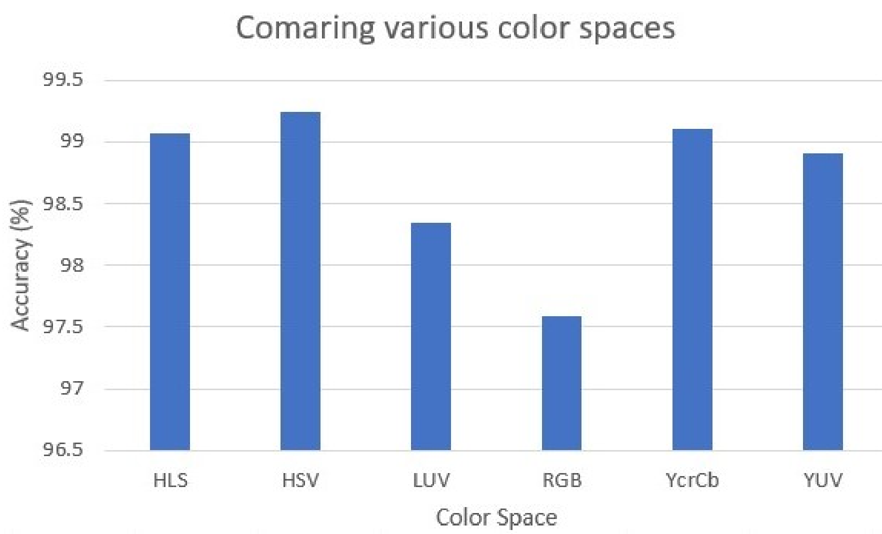


Figure 2: Accuracy for various color space

4.2 Color Space

Since color spaces play a crucial role in the extraction of features, multiple experiments were performed varying the color spaces but keeping the HOG parameters constant. Figure 2 gives a comparison of the performance for six different color spaces based on the percent of misclassified images. For the classification of vehicles and non-vehicle, we used the SVM pipeline.

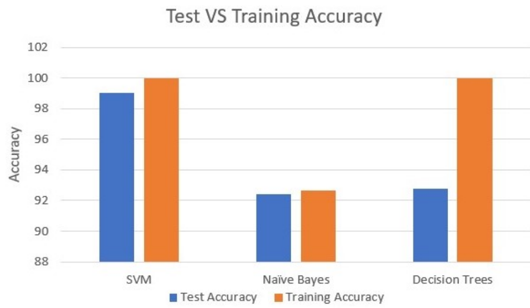
Surprisingly, HSV gives the lowest percentage of misclassified images. This is odd because the HSV color space has strongly correlated channels. Thus, for images with high color variations(vehicles), HSV and RGB may not be suitable [2]. Although HVS achieves best performance during classification, it performs poorly when tested against real images. This is probably because of our inherent bias in our dataset. The 64×64 vehicle images are zoomed in to capture the car but, real images have a lot of background color variation thus resulting in false positive. The results can be observed below.

LAB, LUV, YCrCb, and YUV all achieved similar results. Thus, YCrCb was used a the default color space.

Figure 3: Vehicle Detection with HSV

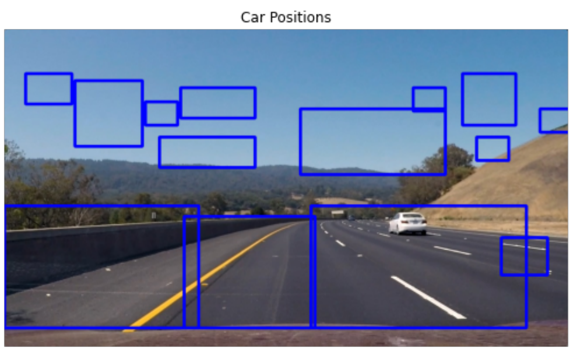
4.3 Comparing Different Models

Of all the three different models used to detect vehicles, SVM performed significantly better than the other two. As a classifier, the accuracy of SVM was very high, and similarly, when tested against real highway images, the false positives were very less 4.

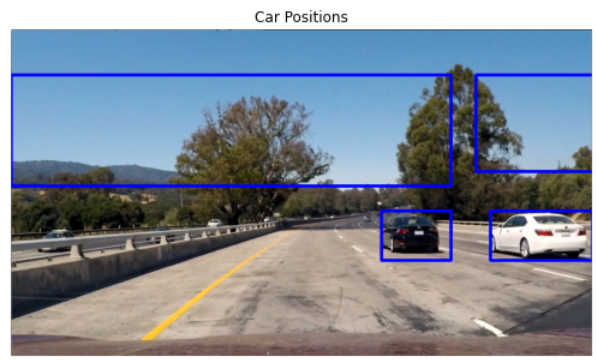


(a) Accuracy Scores

(b) SVM



(c) Naive Bayes



(d) Decision Trees

Figure 4: Comparing Different Models

4.4 Results on video

When implemented on a video footage the pipeline performed poorly. There were lots of false positive results as compared to a similar static image frame. This is probably because distortions in the video caused by the application of lossy compression. So, for artifacts that are similar to vehicles show false positive results. Gaussian smoothing can be applied on the video but still the may suffer from false positives.

5 DISCUSSIONS

The described method along with an SVM classifier performs performed reasonably well on highway images, but the performance decreases significantly when implemented on video footage. Tuning the HOG parameters can be a solution but the optimal values may differ with different videos making it computationally expensive and is slow. Rather deep learning approaches such as YOLO, CNNs, RNNs that are faster, inexpensive, and more accurate can be used.

Although real-time detection cannot be achieved via this approach and there exist far better technologies, we get a good understanding of how color spaces and image gradients affect image features.

6 CONCLUSION

In this paper, we studied the influence of color spaces on the performance of feature extraction and concluded that decorrelated colorspace such as YCrCb works best. We studied the performance of HOG subject to different parameters and concluded that the optimal values of the HOG parameter closely resemble the parameters for human detection proposed by Dalal and Triggs [1]. We then proceeded to study the performance of the various classifiers. Amongst the classifier, SVM performed the best in detecting vehicles in a static image frame but, its performance on video footage decreases significantly. Implementing a vehicle tracking system can be one future step towards improving this method.

The tracking will reduce the number of false positives as we search for positive detections that appear in consecutive frames which can provide better detection.

7 ACKNOWLEDGEMENT–

We would like to express our heartfelt gratitude to Professor Subhashis Banerjee, who guided us throughout the course and gave us a good understanding of the theories and practicalities of ML. We would also like to thank out TAs Bhavesh Neekhra, Chanda Grover, Shweta Prasad, and Yash More for their work and valuable feedback throughout the course.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

[2] Creusen IM, Wijnhoven RGJ, Herbschleb E, With, de PHN. Color exploitation in HOG-based traffic sign detection. In Proceedings of the 17th IEEE International Conference on Image Processing (ICIP 2010), 26-29 September 2010, Hong Kong, Hong Kong. Piscataway: Institute of Electrical and Electronics Engineers. 2010. p. 2669-2672 https://doi.org/10.1109/ICIP.2010.5651637

[3] Ding W, Li S, Zhang G, Lei X, Qian H. Vehicle pose and shape estimation through multiple monocular vision. 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO). 2018;

[4] Satar B, Dirik AE. Deep Learning based vehicle make-model classification. Artificial Neural Networks and Machine Learning – ICANN 2018. 2018;:544–53.

[5] Alfasly SA, Hu Y, Liang T, Jin X, Zhao Q, Liu B. Variational representation learning for vehicle re-identification. 2019 IEEE International Conference on Image Processing (ICIP). 2019;

[6] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016;

[7] User S. GTI [Internet]. Welcome to GTI. [cited 2021Dec14]. Available from: http://www.gti.ssr.upm.es/data/Vehicle_database.html

[8] KITTI Vision Benchmark Suite [Internet]. The Kitti Vision Benchmark Suite. [cited 2021Dec14]. Available from:http://www.cvlibs.net/datasets/kitti/

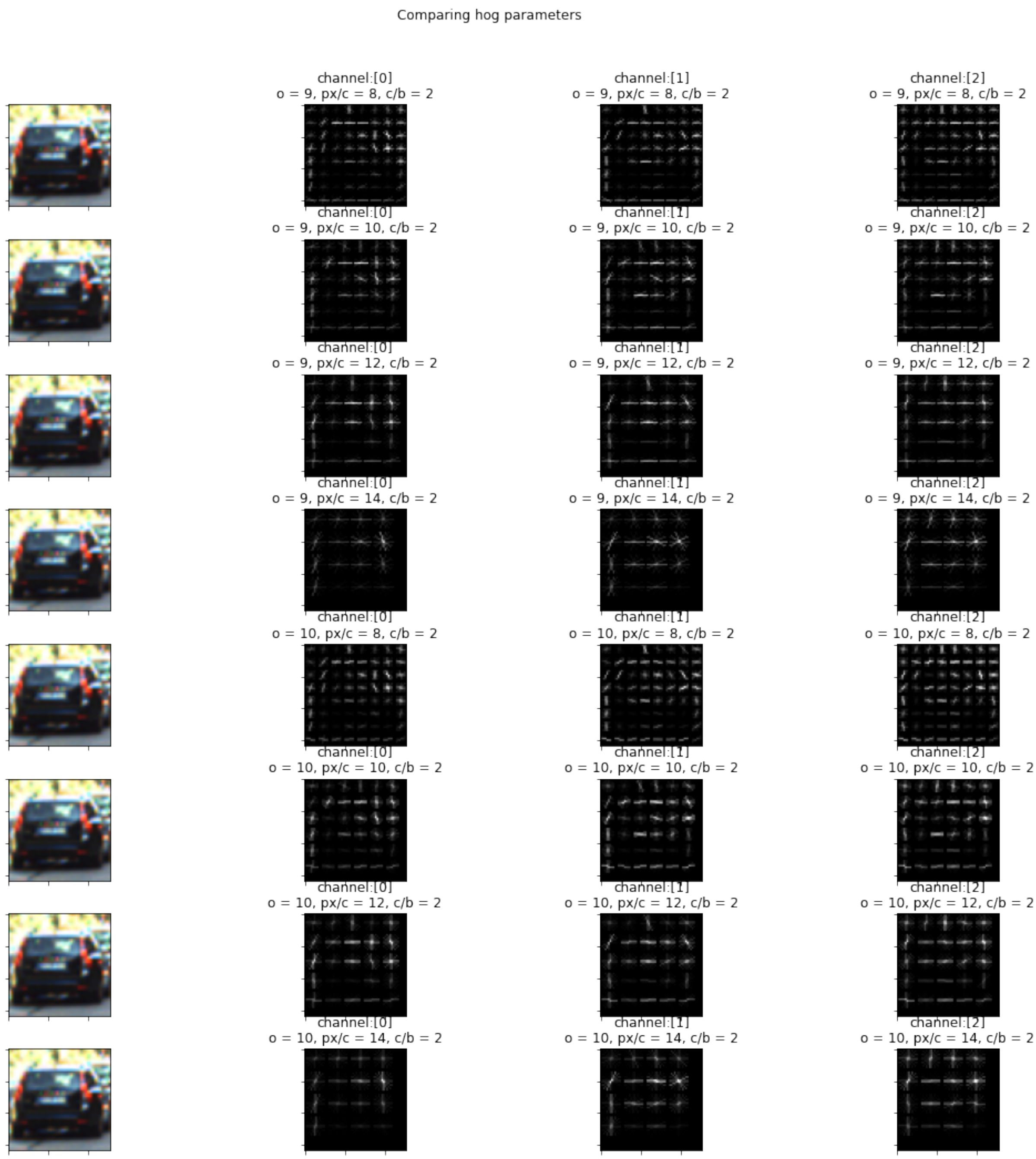


Figure 5: Comparing Different Hog Parameters