Vehicle Recognition in Dashcam Videos

Alexandre Robin
CentraleSupélec

Abstract

The aim of this project is to perform vehicle detection in dashcam videos with traditional computer vision methods, that is without using neural networks.

1. Introduction

This project is about detecting vehicles in dashcam videos. In order to do so we splitted the project in two halves. The first part will focus on a classification task that aims to discriminate vehicles and non-vehicles. To perform such a task we need a dataset. We will see that two techniques are available: using an external dataset or extracting a dataset from the dashcam videos. Once a dataset for classification is ready to use, we extract features from it and proceed to classfication. With this first model trained we can then move on to the second part of the project that focuses on finding vehicles inside the dashcam videos.

2. Classification

2.1. Features

The main challenge of the project is to perform feature extraction with traditional methods not using neural networks. We investigated several methods in order to extract features from the images.

Spatial binning The spatial binning technique consists in grouping pixels spatially - by dividind the image into a grid of cells, and computing some statistics (such as the average or sum) to aggregate the pixel values within each cell and create a lower-resolution representation of the original image. This technique allows to creates small dimension features retaining important information about the overall structure of the image.

Color histogram The color histogram technique consists in representing the distribution of colors of an image with a histogram to provide a quantitative measure on how the colors are distributed across chanels (Red Green Blue in this project). The key concept is to capture the frequency of occurrence of different color values or intensities present in the image. To construct this histogram, we divide the image into discrete bins terms of color values. The histogram is computed by counting the number of pixels that fall into each bin for each color channel. This technique allows to cath the overall color composition of an image, making it less sensitive to lightning conditions. We don't use theses featues when working on gray images.

Histogram of Oriented Gradients The Histogram of Oriented Gradients technique can be broken down to several steps. The first step consists in computing an image's gradients (magnitude and orientation). Once the gradients are available, we split the image into cells in which we will compute an histogram of oriented gradients: the orientation space is split in several bins to which magnitudes are proportionnaly added depending on their orientation proximity. For example, supposing we consider 9 bins over the unsigned space of orientations (going from 0 to 180 degrees) a pixel with gradient magnitude 10 and gradient orientation 110 will contribute with 5 unit to the 100 degree bin and 5 to the 120 bin because it is halfway between. Finally once histograms are computed for every cell, cells are gathered in blocks and normalized. The features are obtained by concatenating the histograms of each cell. The resultant feature vector provides structural information - shape and edge - of the image while being robust to variations in scale, orientation, and lighting conditions thanks to normalization.

2.2. Model

The classification task is performed with a Linear Support Vector Classifier that has been fine-tuned with cross-validation. Two different models have been tried, one for each dataset we considered.

2.3. Dataset

Two dataset have been considered. The first one is an external dataset of 64x64 vehicle and non-vechicle images. The second one has been extracted from the dashcam videos. For each box containing a vehicle in the training set, we have searched for a box of same size not containing a vechicle in the image, between a line considered to be the bottom of the sky and a line considered to be the top of the dashboard. When not possible or too complicated we searched for a 64x64 image not containing a vehicle. Each image of vehicle and non-vehicle has then been resized to 64x64 pixels.

3. Recognition

Once the classifier trained we implemented sliding windows of several sizes in the dashcam videos' images and then proceeded to 64x64 pixel resizing and classification in each window. In order to build final boxes, the positive overlapping windows were aggregated to build a wider window covering the whole vehicle.

4. Results

The results were not very good (around 0.39). An improvement could have been to perform a Principal Component Analysis on the feature space to reduce dimensionality.