Pedestrian detection
VIC Assignment 2

Elias Aouad
CentraleSupélec
elias.aouad@student.ecp.fr

1. Main Idea

In this assignment, we are asked to perform pedestrian detection in a video. In other words, for each frame of the video, output bounding boxes that correspond to the region of the frame that contain a pedestrian. To this extent, HoG features will be used with an SVM on top to identify the regions that will most probably contain a pedestrian.

Basically, the idea was to use the already implemented HoG feature extraction method on Open-CV and use it with the default SVM for people detection. The first pipeline was as follows : for each frame of the video,

1. Use the multi-scale detection method with Open-CV's default parameters to output the eligible regions for pedestrian location
2. Save each region as a bounding box with the corresponding frame ID and bounding box ID

However, this baseline had many disadvantages :

- Too slow : took over 20 minutes to run it
- Outputs too many boxes in the same region, and many boxes were referring to the same pedestrian
- Outputs bigger boxes than necessary
- Was not very efficient : got a shy score of **3%** on the provided video.

Hence, it seemed like the default people detector of Open-CV was having issues dealing with the size/scales of our image and computation time.

2. Methodology

To overcome those issues, we must help the detector focus on the most interesting regions. The pipeline is as follows.

First, a background subtractor was applied. It is already implemented in Open-CV (MOG2), and it removes all unimportant parts of the image, focusing only on people.

Second, morphological transformations were applied on the image. The aim would be to get a binary image with values of 1 where we have people, and 0 everywhere else. However, the background subtractor outputs a noisy image. Hence, a closing transformation was performed to fill the erased parts of a person, and second, an opening transformation was performed to clear all undesired random white points. After that, dilation was performed to strengthen the whiteness of already white regions. By doing so, we have better chances of ending up with very white regions where we have pedestrians on the image.

Once we have a clean binary image, all we want is to find the contours of the image. A contour is a set of points that delimit a white region shape of a binary image. In fact, there is an already implemented method on Open-CV for doing so, and it works best with binary or thresholded grayscale images. We just have to use this function to output the contours of our preprocessed image.

Next, this contour was transformed into a bounding box for the desired region. Once we have that, the image was cropped into the selected region and the original multi-scale detection method was applied (with the HoG features and the default SVM parameters for person detection). Before that, we needed to make sure that our cropped image satisfied minimal dimension requirements : area higher than 100, x-shape and y-shape higher than 115. Otherwise the method would fail and the kernel would crash. To avoid complications with crashing kernels, the image was cropped into a bigger portion in order to meet the requirements.

Once we had our detections, they were saved into a huge list. Last, to avoid overlapping boxes, non-max suppression was performed with a threshold of 0.8.

3. Results

With the provided pipeline, a score of **26.4%** was achieved on the provided video. This shows that the default person detector can improve drastically its performance if it is provided with the right region of focus.