The papers focus on using simple HOG for human detection. And it outperforms other more complicated methods. And idea is to use histogram of local 1-d gradient representation of small regions and combine them. The method chooses 1-d centred [-1, 0 ,1] mask because it works empirically well then others masks. Using color information doesn’t change the result much. Each pixel in a small region called cell will vote for an edge orientation. Thus, the cell will be represented by an orientation gradient. Mentioned in paper, the singed gradient is somehow ignored because human is more discriminated than other objects. Then a number of cells form a lager region called block.

Normalization decreases the variability of the object. The key of HOG is to use overlapped blocks. It increases the performance by avoiding accidentally assigning a wrong cell to the final components. It is a very good and simple way of normalization. Vertical and vertical+horizontal block pairs work significantly better. It indicates that the most common human posture is standing in the pictures. Also contrast normalization help to decrease the false positive rate.

Overall, the method is simple and straightforward. But the most parameters of this methods depends on the empirically experiments. The size of cell, block and detector window is dependent on the image size. It is highly data-driven method. However, it points out that smoothing can decrease the performance as the method relies on the edge of human body. For feature based method, one thing is to know what the most important feature that contributing to object detection.