Histograms of Oriented Gradients for Human Detection

Authors: Navneet Dalal and Bill triggs
Journal: Computer Vision
Keywords: Human Recognition, SVM, HOG, SIFT, MIT, wavelets
Year: 2005

Report:

1. **Abstract:**

"We study the question of feature sets for robust visual object recognition, adopting linear SVM based human detection as a test case. After reviewing existing edge and gradient based descriptors, we show experimentally that grids of Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for human detection. We study the influence of each stage of the computation on performance, concluding that fine-scale gradients, fine orientation binning, relatively coarse spatial binning, and high-quality local contrast normalization in overlapping descriptor blocks are all important for good results. The new approach gives near-perfect separation on the original MIT pedestrian database, so we introduce a more challenging dataset containing over 1800 annotated human images with a large range of pose variations and backgrounds."

Introduction

- The paper's motivation is human recognition in different poses by a new approach of combining HOG and linear SVM to get the better results.
- The authors also focused on examining the each phase from computation on performance to the relatively ocarse spatial binning, which they hoped that result in the increase in the performance.
- The MIT data set of pedestrians are taken as the training data set and computed the frequency results on them.
- In addition they have created a new dataset with different poses and backgrounds of humans as their approach results in the near-perfect solution with the existing database.
- They have observed the better results compared to other approaches like kernel SVM, PCA-SIFT etc., with the defined limitations.

Previous works

- All the previous works have been significantly concentrated on the object detection, where human detection had been one of the least observed fields.
- All the works done on human detection has also done on parts based methods by identifying faces, hands, legs and profiles etc.,
- On the other hand, the authors concentrated on the single window human detection, which they affirm that had not been done before.

Content

    - The content mainly deals with the object recognition in which svm process is used as an approach.
- **The main idea of the procedure is that the object shape can be characterized by the local gradient's distribution and the edge directions even without knowing the equivalent edge position and gradients.**
- **Initially the image is splitted into small windows in which one dimensional histograms of gradient and edge directions are derived and combined to result in a normalized vector.**
- **They normalized the whole image vectors and named the descriptor blocks which had been normalized as "Histograms of Oriented Gradient" descriptors.**
- **According to the procedure as shown in the image, initially the input image is normalized by gamma and color which is followed by computing the gradients and edge directions.**
- **Then, the weights are given to the orientation cells and spatial cells. That leads to the some of the overlapping spatial blocks.**
- **Then they normalized the overlapped spatial blocks and collected HOGs from the detection windows.**
- **On the HOGs collected, they are implementing the linear support vector machine which classifies into the person or non-person batches.**
- **As an experiment they have taken around 1200 images for training as positive examples. From around 12000 patches, they have taken the one tenth of it for training as negative set.**
- **They initially trained the networks with negative set and searched for false positives.**
- **Then they have retrained with the 12000 patch samples and with the hard examples to generate the detector.**
- **For the SVM training, which is the final phase of training, they had taken into a corresponding amount of ram and it improved the individual detectors significantly.**

Results

The author, despite of concentrating on one or two results, he derived a good set of results of which all come through the process and some advices are given to increase the efficiency of the results.

- **For comparative results, graphs are drawn between DET(Detection Error Tradeoff) and FPPW, for different procedures, in which R-HOG and C-HOG had shown the better results.**
- **The graphs are given the error rate for false positives and false negatives, which had shown a significant results in the HOG based kernels with low error rates.**
- **It is stated that replacing linear SVM with Gaussian Kernel increased the optimistic results by 3%.**
- **It is advised that smoothing should not be done before HOG as that is resulting in bad results.**
- **Strong local contrast normalizations are essential for good results of the HOG were also pointed out.**

- **At the end of the experimentation, the authors even noticed the decrease of false positive rates.**

Discussion and Thoughts

- **Very good pointers are given the introduction itself regarding topics and the content in each passage, which helps the user to navigate easily through the paper.**
- **Datasets are made free and their references are given in the paper.**
- **In addition to the details, an explanation about R-HOG and C-Hog are also given briefly.**
- **The paper has solved many results in comparative had given many alternatives to increase the precision of the results.**
- **Only pedestrian cases are taken as the datasets. Other poses of the humans like sitting, driving, etc., are not taken into account.**
- **The procedures of HOG, SIFT and SVM are not given , which suggest that reader should have prior knowledge regarding that. However their advantages and applications are illustrated**

Conclusion

- **Finally, it can be deduced that it is a paper that is mainly concentrated on Computer Vision with the applications of SVM and HOG.**
- **The authors also stated that HOG will best the PCA-SIFT, shape context and wavelet.**
- **It also gave very near perfection to the MIT database, where as haar-like wavelets outperformed the test.**