Histogram of Oriented Gradients

Summary written by Vedant Jain
February 2, 2021

Summary: This paper looks into a new method for selecting features based on using the magnitude and phase of orientation based gradients on small patches of the image. They showed that the performance of their approach as far better than other attempts.

Related work: The authors looked at other attempts at extracting pedestrian features. Specifically they looked at methods using wavelets as functions that could be used in convolution to extract features. [1] They also looked at work using adaboost to train based on a chain of more complex measurements. Finally, they looked at other methods that used Gaussian filters and combinations of orientation based gradients. They noted that this method was more complex then their own yet resulted in poorer outcomes.

Approach: Their approach essentially hopes to harness local gradients in an image without necessarily knowing their exact location in 2D space. Practically to perform this calculation they first normalize the gamma and colour of the image. Then compute local 1-D gradients which get weighted into spatial and orientation cells. These are then normalized for the small histogram blocks. This isn't the individual patches but actually includes a larger group of the blocks. Finally each individual HOG is collected into an array which becomes the main features that then get applied to the SVM model. Key features to note in their work is that they have overlapping patches that create overlapping HOG descriptors. Their approach essentially involved them selecting 1239 of the images to train which included pedestrians and 1218 person free images to to be without pedestrians. Then then trained the SVM model based on the initial 12180 detectors. The 12180 comes from breaking up an image into smaller patches. They then retrained their model with some hard examples, ie features that looked like pedestrians and were actually classified falsely. They plotted their results on a Detection Error Tradeoff Curve very similar to a ROC.

Datasets, Experiments and Results: The authors used the established MIT pedestrian database. This database contains 509 training images and 200 test images. They found that their method performed too well on that criteria so they also created their own data set of 2478 images. Their algorithm greatly outperformed that of wavelets and application of PCA-SIFT methods, which were other leading methods.

Strengths: This paper does a good job of describing their method and other competing methods that have been good at feature extraction. What I find really helpful is that they go into depth of how changing certain parameters of their method changes the quality of their results. For example they mentioned how increasing the gradient scale results in poorer performance. They also describe how the correct number of orientation bins at around 9 bins from 0-180 degrees resulted in some of the best results. Finally they went into different methods of applying normalization to the patches. They did this in a very clever method by normalizing in relation to an individual block but having th blocks overlap so that a specific measurement would be normalized according to two different blocks, thereby allowing a more standard normalizing procedure.

Weaknesses: One weakness in their method is that this method has only been explicitly tested in pedestrians. It could be that the method runs into much more problems when it has to deal with high contrast patterns which could significantly challenge a method such as image gradient descent. The authors should have gone deeper into the fundamentals of their method as that was abundantly clear at the introduction of the paper and required some background researching to be up to speed with all the pertinent methods.

Reflections: The author describes a new method to extract features from images of pedestrians. They describe how successful their method is as well as the hyper parameters which describe different ways which can affect the usefulness of the features they extract. They go deep into explaining the draw backs and strengths of some of these changes. Finally, they determine how bluring an image initially, wide orientation binning, fine spatial binning, and low quality normalization can all cause bad performance. They emphasize that these considerations must be taken when developing new feature extracting algorithms.

References

[1] C. P. A. Mohan and T. Poggio. Example-based object detection in images by components. *PAMI*, 23(4):349–361, 2001.