Histograms of Oriented Gradients

Edgar A. Margffoy-Tuay
Universidad de los Andes
201412566
ea.margffoy10@uniandes.edu.co

Abstract

Histograms of Oriented Gradients (HOG) are image features that describe and capture local orientation information on an image. Reminiscent of SIFT features, HOG binarizes the gradient responses on a local window cell, which allow to describe local objects based on their gradient filter responses, based on this features, it is possible to distnguish different objects and eventually, detect them. The objective of the present report consists on designing and testing a multiscale object detector based on HOG features, similar to the baseline approach proposed by Dalal and Triggs adapted to face detection.

1. Introduction

HOG features were proposed by Dalal and Triggs on [1] to characterize and generalize pedestrian outlines and contour gradient orientations suitable to detect them on a street scene, initially, HOG descriptors were only employed on object categories that presented a similar geometry, point of view and aspect ratio, which favored the generalization of a single HOG activation template that could be applied onto a test image via convolution. Due to the similar geometry, aspect ratio and orientation present on pedestrian pictures, HOG representation was suitable to solve this binary classification problem, as Dalal and Triggs demostrated.

However, if simple HOG features were to be applied to solve a complex multiclass object that presents several of the initial HOG limitations such as viewpoint translation, geometric obfuscation and intra-class instance differences, the expected classification accuracy rate should be low, due to the introduction of several false positive detections due to the similarity between the HOG representations of different objects that should be represented differently.

To solve those limitations, several HOG improvements were proposed, such as Deformable Part Models (DPM) [2], on which not only a single local object HOG representation is used but also the relation with nearby HOG features. This approach allows to represent an object as a graph of HOG features, giving more information about geometry, location and viewpoint of an image. Another approach consists on training a single SVM classifier per each HOG object instance, this approach, denomined Exemplar SVMs [4], allows not only to classify each object instance accurately, but it also allows to obtain geometrical information suitable to more advanced tasks, such as 3D reconstruction and pose estimation, however, this method is too expensive due to the formulation and minimization of a single classification model per each example present on the dataset, which means that this approach is not scalable, however it is highly parallel, due to the independence of each exemplar classification model.

To explore the limitations of a plain HOG detector on a complex detection task, such as face detection, a simple Dalal-Triggs inspired framework is proposed to solve this task over the Wider Face detection dataset [6], a modern state-of-the-art face detection benchmark.

2. Materials and Methods

HOG feature calculation is similar to SIFT [3] representation, the main difference between both descriptors is related to gradient histogram normalization across a segmentation of a window on a set of blocks, this procedure allows to decorrelate gradient histograms on different blocks members of a single HOG window, which allows to introduce more redundancy onto the model, increasing the accuracy of the final classification model.

To detect multiple object instances on each image, a normalized sliding window of size 136×100 approach was used, this decision was taken taking in account all the possible aspect ratios of the bounding boxes present on the dataset, some of them are elogated, and other have a similar proportion. This window was also applied over different scales, to ensure that smaller background or near faces were also subject to examination. Finally, the detection pipeline was implemented without any addition nor any improvement such as DPMs or Exemplar SVMs. To calculate HOG histograms across each of the dataset images, a cell of size 8 was defined.

All the implementation was based upon Andrea's Vedaldi vlfeat and matconvnet libraries [5], as shown on the Oxford's VGG tutorial on object detection[1].

About the Datasets

- **Wider Face**: This dataset contains 32203 images, which in turn contains 393703 annotated faces (Bounding boxes) grouped on 60 event categories. All detection evaluations are based on the Intersection-over-Union (IoU) metric.

3. Results

[1] Available at: `http://www.robots.ox.ac.uk/~vgg/practicals/category-detection/`

Caltech-101						
# Train	# Test	# Words	Spatial-X	Spatial-Y	C	Accuracy (%)
20	20	600	(2, 4)	(2, 4)	10	70.85%
50	20	1000	(2, 4)	(2, 4)	10	**75.79**%
50	20	600	(2, 4)	(2, 4)	30	75.37%
50	20	600	(2, 4)	(2, 4)	5	74.94%

Table 1: Caltech-101: Accuracy test results of PHOW subject to different hyperparameter configurations

ImageNet-200						
# Train	# Test	# Words	Spatial-X	Spatial-Y	C	Accuracy (%)
50	100	1000	(2, 5)	(2, 5)	10	24.47%
50	100	600	(2, 4)	(2, 4)	10	23.24%
70	100	1200	(1, 5)	(1, 5)	10	**27.17**%

Table 2: ImageNet-200: Accuracy test results of PHOW subject to different hyperparameter configurations

References

[1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1, June 2005.

[2] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 32(9):1627–1645, 2010.

[3] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2*, ICCV '99, pages 1150–, Washington, DC, USA, 1999. IEEE Computer Society.

[4] T. Malisiewicz, A. Gupta, and A. A. Efros. Ensemble of exemplar-svms for object detection and beyond. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 89–96. IEEE, 2011.

[5] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. `http://www.vlfeat.org/`, 2008.

[6] S. Yang, P. Luo, C.-C. Loy, and X. Tang. Wider face: A face detection benchmark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5525–5533, 2016.