Histogram of oriented gradients

The **histogram of oriented gradients (HOG)** is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

Robert K. McConnell of Wayland Research Inc. first described the concepts behind HOG without using the term HOG in a patent application in 1986.[1] In 1994 the concepts were used by Mitsubishi Electric Research Laboratories.[2] However, usage only became widespread in 2005 when Navneet Dalal and Bill Triggs, researchers for the French National Institute for Research in Computer Science and Automation (INRIA), presented their supplementary work on HOG descriptors at the Conference on Computer Vision and Pattern Recognition (CVPR). In this work they focused on pedestrian detection in static images, although since then they expanded their tests to include human detection in videos, as well as to a variety of common animals and vehicles in static imagery.

Contents

Theory

The essential thought behind the histogram of oriented gradients descriptor is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The image is divided into small connected regions called cells, and for the pixels within each cell, a histogram of gradient directions is compiled. The descriptor is the concatenation of these histograms. For improved accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block. This normalization results in better invariance to changes in illumination and shadowing.

The HOG descriptor has a few key advantages over other descriptors. Since it operates on local cells, it is invariant to geometric and photometric transformations, except for object orientation. Such changes would only appear in larger spatial regions. Moreover, as Dalal and Triggs discovered, coarse spatial sampling, fine orientation sampling, and strong local photometric normalization permits the individual body movement of pedestrians to be ignored so long as they maintain a roughly upright position. The HOG descriptor is thus particularly suited for human detection in images.[3]

Algorithm implementation

Gradient computation

The first step of calculation in many feature detectors in image pre-processing is to ensure normalized color and gamma values. As Dalal and Triggs point out, however, this step can omitted in HOG descriptor computation, as the ensuing descriptor normalization essentially achieves the same result. Image pre-processing thus provides little impact on performance. Instead, the first step of calculation is the computation of the gradient values. The most common method is to apply the 1-D centered, point discrete derivative mask in one or both of the horizontal and vertical directions. Specifically, this method requires filtering the color or intensity data of the image with the following filter kernels:

$$[-1, 0, 1] \text{ and } [-1, 0, 1]^\top.$$

Dalal and Triggs tested other, more complex masks, such as the 3x3 Sobel mask or diagonal masks, but these masks generally performed more poorly in detecting humans in images. They also experimented with Gaussian smoothing before applying the derivative mask, but similarly found that omission of any smoothing performed better in practice.[4]

Orientation binning

The second step of calculation is creating the cell histograms. Each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values found in the gradient computation. The cells themselves can either be rectangular or radial in shape, and the histogram channels are evenly spread over 0 to 180 degrees or 0 to 360 degrees, depending on whether the gradient is "unsigned" or "signed". Dalal and Triggs found that unsigned gradients used in conjunction with 9 histogram channels performed best in their human detection experiments. As for the vote weight, pixel contribution can either be the gradient magnitude itself, or some function of the magnitude. In tests, the gradient magnitude itself generally produces the best results. Other options for the vote weight could include the square root or square of the gradient magnitude, or some clipped version of the magnitude.[5]

Descriptor blocks

To account for changes in illumination and contrast, the gradient strengths must be locally normalized, which requires grouping the cells together into larger, spatially connected blocks. The HOG descriptor is then the concatenated vector of the components of the normalized cell histograms from all of the block regions. These blocks typically overlap, meaning that each cell contributes more than once to the final descriptor. Two main block geometries exist: rectangular R-HOG blocks and circular C-HOG blocks. R-HOG blocks are generally square grids,

represented by three parameters: the number of cells per block, the number of pixels per cell, and the number of channels per cell histogram. In the Dalal and Triggs human detection experiment, the optimal parameters were found to be four 8x8 pixels cells per block (16x16 pixels per block) with 9 histogram channels. Moreover, they found that some minor improvement in performance could be gained by applying a Gaussian spatial window within each block before tabulating histogram votes in order to weight pixels around the edge of the blocks less. The R-HOG blocks appear quite similar to the scale-invariant feature transform (SIFT) descriptors; however, despite their similar formation, R-HOG blocks are computed in dense grids at some single scale without orientation alignment, whereas SIFT descriptors are usually computed at sparse, scale-invariant key image points and are rotated to align orientation. In addition, the R-HOG blocks are used in conjunction to encode spatial form information, while SIFT descriptors are used singly.

Circular HOG blocks (C-HOG) can be found in two variants: those with a single, central cell and those with an angularly divided central cell. In addition, these C-HOG blocks can be described with four parameters: the number of angular and radial bins, the radius of the center bin, and the expansion factor for the radius of additional radial bins. Dalal and Triggs found that the two main variants provided equal performance, and that two radial bins with four angular bins, a center radius of 4 pixels, and an expansion factor of 2 provided the best performance in their experimentation(to achieve a good performance, at last use this configure). Also, Gaussian weighting provided no benefit when used in conjunction with the C-HOG blocks. C-HOG blocks appear similar to shape context descriptors, but differ strongly in that C-HOG blocks contain cells with several orientation channels, while shape contexts only make use of a single edge presence count in their formulation.[6]

Block normalization

Dalal and Triggs explored four different methods for block normalization. Let v be the non-normalized vector containing all histograms in a given block, $\|v\|_k$ be its k-norm for $k = 1, 2$ and e be some small constant (the exact value, hopefully, is unimportant). Then the normalization factor can be one of the following:

L2-norm: $f = \dfrac{v}{\sqrt{\|v\|_2^2 + e^2}}$

L2-hys: L2-norm followed by clipping (limiting the maximum values of v to 0.2) and renormalizing, as in[7]

L1-norm: $f = \dfrac{v}{(\|v\|_1 + e)}$

L1-sqrt: $f = \sqrt{\dfrac{v}{(\|v\|_1 + e)}}$

In addition, the scheme L2-hys can be computed by first taking the L2-norm, clipping the result, and then renormalizing. In their experiments, Dalal and Triggs found the L2-hys, L2-norm, and L1-sqrt schemes provide similar performance, while the L1-norm provides slightly less reliable performance; however, all four methods showed very significant improvement over the non-normalized data.[8]

Object recognition

HOG descriptors may be used for object recognition by providing them as features to a machine learning algorithm. Dalal and Triggs used HOG descriptors as features in a support vector machine (SVM) [9]; however, HOG descriptors are not tied to a specific machine learning algorithm.

Performance

In their original human detection experiment, Dalal and Triggs compared their R-HOG and C-HOG descriptor blocks against generalized Haar wavelets, PCA-SIFT descriptors, and shape context descriptors. Generalized Haar wavelets are oriented Haar wavelets, and were used in 2001 by Mohan, Papageorgiou, and Poggio in their own object detection experiments. PCA-SIFT descriptors are similar to SIFT descriptors, but differ in that principal component analysis is applied to the normalized gradient patches. PCA-SIFT descriptors were first used in 2004 by Ke and Sukthankar and were claimed to outperform regular SIFT descriptors. Finally, shape contexts use circular bins, similar to those used in C-HOG blocks, but only tabulate votes on the basis of edge presence, making no distinction with regards to orientation. Shape contexts were originally used in 2001 by Belongie, Malik, and Puzicha.

The testing commenced on two different data sets. The Massachusetts Institute of Technology (MIT) pedestrian database contains 509 training images and 200 test images of pedestrians on city streets. The set only contains images featuring the front or back of human figures and contains little variety in human pose. The set is well-known and has been used in a variety of human detection experiments, such as those conducted by Papageorgiou and Poggio in 2000. The MIT database is currently available for research at http://cbcl.mit.edu/cbcl/software-datasets/PedestrianData.html. The second set was developed by Dalal and Triggs exclusively for their human detection experiment due to the fact that the HOG descriptors performed near-perfectly on the MIT set. Their set, known as INRIA, contains 1805 images of humans taken from personal photographs. The set contains images of humans in a wide variety of poses and includes difficult backgrounds, such as crowd scenes, thus rendering it more complex than the MIT set. The INRIA database is currently available for research at http://lear.inrialpes.fr/data.

The above site has an image showing examples from the INRIA human detection database.

As for the results, the C-HOG and R-HOG block descriptors perform comparably, with the C-HOG descriptors maintaining a slight advantage in the detection miss rate at fixed false positive rates across both data sets. On the MIT set, the C-HOG and R-HOG descriptors produced a detection miss rate of essentially zero at a 10^{-4} false positive rate. On the INRIA set, the C-HOG and R-HOG descriptors produced a detection miss rate of roughly 0.1 at a 10^{-4} false positive rate. The generalized Haar wavelets represent the next highest performing approach: they produced roughly a 0.01 miss rate at a 10^{-4} false positive rate on the MIT set, and roughly a 0.3 miss rate on the INRIA set. The PCA-SIFT descriptors and shape context descriptors both performed fairly poorly on both data sets. Both methods produced a miss rate of 0.1 at a 10^{-4} false positive rate on the MIT set and nearly a miss rate of 0.5 at a 10^{-4} false positive rate on the INRIA set.

Further development

As part of the Pascal Visual Object Classes 2006 Workshop, Dalal and Triggs presented results on applying histogram of oriented gradients descriptors to image objects other than humans, such as cars, buses, and bicycles, as well as common animals such as dogs, cats, and cows. They included with their results the optimal parameters for block formulation and normalization in each case. The image in the below reference shows some of their detection examples for motorbikes.[10]

As part of the 2006 European Conference on Computer Vision (ECCV), Dalal and Triggs teamed up with Cordelia Schmid to apply HOG detectors to the problem of human detection in films and videos. They combined HOG descriptors on individual video frames with their newly introduced internal motion histograms (IMH) on pairs of subsequent video frames. These internal motion histograms use the gradient magnitudes from optical flow fields obtained from two consecutive frames. These gradient magnitudes are then used in the same manner as those produced from static image data within the HOG descriptor approach. When testing on two large datasets taken from several movies, the combined HOG-IMH method yielded a miss rate of approximately 0.1 at a 10^{-4} false positive rate.[11]

At the Intelligent Vehicles Symposium in 2006, F. Suard, A. Rakotomamonjy, and A. Bensrhair introduced a complete system for pedestrian detection based on HOG descriptors. Their system operates using two infrared cameras. Since human beings appear brighter than their surroundings on infrared images, the system first locates positions of interest within the larger view field where humans could possibly be located. Then support vector machine classifiers operate on the HOG descriptors taken from these smaller positions of interest to formulate a decision regarding the presence of a pedestrian. Once pedestrians are located within the view field, the actual position of the pedestrian is estimated using stereo vision.[12]

At the IEEE Conference on Computer Vision and Pattern Recognition in 2006, Qiang Zhu, Shai Avidan, Mei-Chen Yeh, and Kwang-Ting Cheng presented an algorithm to significantly speed up human detection using HOG descriptor methods. Their method uses HOG descriptors in combination with the cascading classifiers algorithm normally applied with great success to face detection. Also, rather than relying on blocks of uniform size, they introduce blocks that vary in size, location, and aspect ratio. In order to isolate the blocks best suited for human detection, they applied the AdaBoost algorithm to select those blocks to be included in the cascade. In their experimentation, their algorithm achieved comparable performance to the original Dalal and Triggs algorithm, but operated at speeds up to 70 times faster. In 2006, the Mitsubishi Electric Research Laboratories applied for the U.S. Patent of this algorithm under application number 20070237387.[13]

At the IEEE International Conference on Image Processing in 2010, Rui Hu, Mark Banard, and John Collomosse extended the HOG descriptor for use in sketch based image retrieval (SBIR). A dense orientation field was extrapolated from dominant responses in the Canny edge detector under a Laplacian smoothness constraint, and HOG computed over this field. The resulting gradient field HOG (GF-HOG) descriptor captured local spatial structure in sketches or image edge maps. This enabled the descriptor to be used within a content-based image retrieval system searchable by free-hand sketched shapes.[14] The GF-HOG adaptation was shown to outperform existing gradient histogram descriptors such as SIFT, SURF, and HOG by around 15 percent at the task of SBIR.[15]

In 2010, Martin Krückhans introduced an enhancement of the HOG descriptor for 3D pointclouds.[16] Instead of image gradients he used distances between points (pixels) and planes, so called residuals, to characterize a local region in a pointcloud. His histogram of oriented residuals descriptor (HOR) was successfully used in object detection tasks of 3d pointclouds.[17]

See also

- Corner detection
- Pedestrian detection
- Feature (computer vision)
- Feature detection (computer vision)
- Feature extraction
- Interest point detection

- Object recognition
- Scale-invariant feature transform

References

1. "Method of and apparatus for pattern recognition" (http://www.google.co.uk/patents/US4567610).
2. "Orientation Histograms for Hand Gesture Recognition" (http://www.merl.com/publications/TR94-03).
3. "Histograms of Oriented Gradients for Human Detection" (http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf) (PDF). p. 2.
4. "Histograms of Oriented Gradients for Human Detection" (http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf) (PDF). p. 4.
5. "Histograms of Oriented Gradients for Human Detection" (http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf) (PDF). p. 5.
6. "Histograms of Oriented Gradients for Human Detection" (http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf) (PDF). p. 6.
7. D. G. Lowe. Distinctive image features from scale-invariant keypoints. (http://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf) IJCV, 60(2):91–110, 2004.
8. "Histograms of Oriented Gradients for Human Detection" (http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf) (PDF). p. 6.
9. "Histograms of Oriented Gradients for Human Detection" (http://lear.inrialpes.fr/people/triggs/pubs/Dalal-cvpr05.pdf) (PDF). p. 1.
10. "Object Detection using Histograms of Oriented Gradients" (http://www.pascal-network.org/challenges/VOC/voc2006/slides/dalal.pdf) (PDF).
11. "Human Detection Using Oriented Histograms of Flow and Appearance" (http://www.acemedia.org/aceMedia/files/document/wp7/2006/eccv06-inria.pdf) (PDF). (original document no longer available; similar paper (http://lear.inrialpes.fr/pubs/2006/DTS06/eccv2006.pdf))
12. "Pedestrian Detection using Infrared images and Histograms of Oriented Gradients" (http://www.ce.unipr.it/people/broggi/publications/iv2006-pd-aziz.pdf) (PDF).
13. "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients" (http://seaboy.tistory.com/attachment/cfile25.uf@116008124A989DCB0E0C23.pdf) (PDF).
14. "Gradient Field Descriptor for Sketch based Image Retrieval and Localisation" (http://personal.ee.surrey.ac.uk/Personal/J.Collomosse/pubs/Hu-ICIP-2010.pdf) (PDF).
15. "A Performance Evaluation of the Gradient Field HOG Descriptor for Sketch based Image Retrieval" (http://personal.ee.surrey.ac.uk/Personal/J.Collomosse/pubs/Hu-CVIU-2013.pdf) (PDF).
16. Krückhans, Martin. "Ein Detektor für Ornamente auf Gebäudefassaden auf Basis des "histogram-of-oriented-gradients"-Operators" (http://www.ikg.uni-bonn.de/uploads/tx_ikgpublication/krueckhans2010detektor.pdf) (PDF). (german)
17. "Semantic 3D Octree Maps based on Conditional Random Fields" (http://www.uni-koblenz.de/~agas/Documents/Lang2013S3O.pdf) (PDF).

External links

- http://www.mathworks.com/matlabcentral/fileexchange/33863 An implementation for Matlab (mex file)
- http://www.cs.cmu.edu/~yke/pcasift/ - Code for PCA-SIFT Object Detection
- http://lear.inrialpes.fr/software/ - Software Toolkit for HOG Object Detection (Research Team homepage)
- https://web.archive.org/web/20100502032344/http://www.navneetdalal.com/software - Software Toolkit for HOG Object Detection (Navneet Dalal homepage)
- http://dlib.net/imaging.html#scan_fhog_pyramid - C++ and Python Software Toolkit for HOG Object Detection

- http://pascal.inrialpes.fr/data/human/ - INRIA Human Image Dataset
- http://cbcl.mit.edu/software-datasets/PedestrianData.html - MIT Pedestrian Image Dataset

Retrieved from "https://en.wikipedia.org/w/index.php?title=Histogram_of_oriented_gradients&oldid=825987013"

This page was last edited on 16 February 2018, at 15:25.