

Integral Channel Features

Piotr Dollár¹

pdollar@caltech.edu

Zhuowen Tu²

zhuowen.tu@loni.ucla.edu

Pietro Perona¹

perona@caltech.edu

Serge Belongie³

sjb@cs.ucsd.edu

¹ Dept. of Electrical Engineering
California Institute of Technology

² Lab of Neuro Imaging
University of CA, Los Angeles

³ Dept. of Computer Science and Eng.
University of California, San Diego

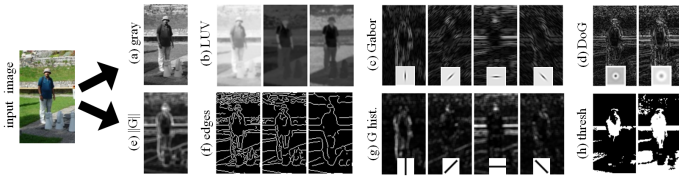


Figure 1: Multiple registered image channels are computed using various transformations of the input image; next, features such as local sums, histograms, and Haar wavelets are computed efficiently using integral images. Such features, which we refer to as *integral channel features*, naturally integrate heterogeneous sources of information, have few parameters, and result in fast, accurate detectors.

We study the performance of ‘integral channel features’ for image classification tasks, focusing in particular on pedestrian detection. The general idea behind integral channel features is that multiple registered image channels are computed using linear and non-linear transformations of the input image [6], and then features such as local sums, histograms, and Haar features and their various generalizations are efficiently computed using integral images [8]. Such features have been used in recent literature for a variety of tasks – indeed, variations appear to have been invented independently multiple times. Although integral channel features have proven effective, little effort has been devoted to analyzing or optimizing the features themselves. In this work we present a unified view of the relevant work in this area and perform a detailed experimental evaluation. We demonstrate that when designed properly, integral channel features not only outperform other features including histogram of oriented gradient (HOG), they also (1) naturally integrate heterogeneous sources of information, (2) have few parameters and are insensitive to exact parameter settings, (3) allow for more accurate spatial localization during detection, and (4) result in fast detectors when coupled with cascade classifiers.

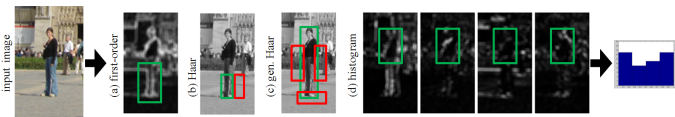


Figure 2: Examples of integral channel features: (a) A first-order feature is the sum of pixels in a rectangular region. (b) A Haar-like feature is a second-order feature approximating a local derivative [8]. (c) Generalized Haar features include more complex combinations of weighted rectangles. (d) Histograms can be computed by evaluating local sums on quantized images [7].

We show significantly improved results over previous applications of similar features to pedestrian detection [3]. In fact, full-image evaluation on the INRIA pedestrian dataset shows that learning using standard boosting coupled with our optimized integral channel features matches or outperforms all but one other method [5], including state of the art approaches obtained using HOG [2] features with more sophisticated learn-

ing techniques. On the task of accurate localization in the INRIA dataset, the proposed method outperforms state of the art by a large margin.

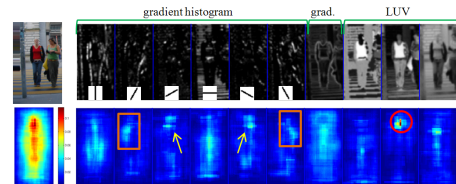


Figure 4: *Top*: Example image and computed channels. *Bottom*: Rough visualization of spatial support of trained classifier for all channels jointly (left) and separately for each channel type, obtained by averaging the rectangle masks of selected features. Peaks in different channels are highlighted.

In addition to the large gains in performance, we describe a number of optimizations that allow us to compute effective channels that take about .05-.2s per 640×480 image depending on the options selected. For 320×240 images, the channels can be computed in real time at rates of 20-80 frames per second on a standard PC. Our overall detection system has a runtime of about 2s for multiscale pedestrian detection in a 640×480 image, the fastest of all methods surveyed in [4].

Finally, we show results on the recently introduced Caltech Pedestrian Dataset [1, 4] which contains almost half a million labeled bounding boxes and annotated occlusion information. Results for 50-pixel or taller, unoccluded or partially occluded pedestrians are shown in Fig. 5. ChnFtrs significantly outperforms all other methods, achieving a detection rate of almost 60% at 1 fppi, compared to 50% for competing methods.

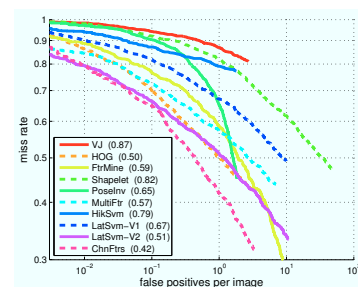


Figure 5: Results on the Caltech Pedestrian Dataset[1, 4]

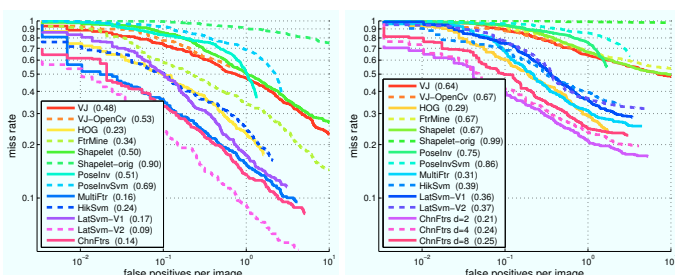


Figure 3: **Left**: INRIA Results. **Right**: Localization Accuracy.

- [1] www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/.
- [2] N. Dalal and B. Triggs. Histogram of oriented gradient for human detection. In *CVPR*, 2005.
- [3] P. Dollár, B. Babenko, S. Belongie, P. Perona, and Z. Tu. Multiple component learning for object detection. In *ECCV*, 2008.
- [4] P. Dollár, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In *CVPR*, 2009.
- [5] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model. In *CVPR*, 2008.
- [6] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, 7: 923–932, May 1990.
- [7] F. Porikli. Integral histogram: A fast way to extract histograms in cartesian spaces. In *CVPR*, 2005.
- [8] P. Viola and M. Jones. Robust real-time object detection. *IJCV*, 57 (2):137–154, 2004.