The Fastest Pedestrian Detector in the West

Piotr Dollár¹  Serge Belongie²  Pietro Perona¹
¹ Electrical Engineering, California Institute of Technology
{pdollar,perona}@caltech.edu
² Computer Science and Engineering, University of California, San Diego
sjb@cs.ucsd.edu

Approximating Multiscale Gradients

Selected Features

- Most features not scale invariant
- Upsampling: no information created, can predict feature response
- Downsampling: information lost, but loss consistent, can compensate!

Fast Accurate Multiscale Detection

Approach:
- Approximate features at nearby scales
- Coarsely sampled image pyramid (octave step)
- Classifier pyramid within each octave

Benefits:
- Speed of classifier pyramid
- Accuracy of image pyramid
- General applicability

Method Parameters

- Fastest Pedestrian Detector in the West (FPDW)
- ACCURATE: within 1%-2% of top performing algorithm
- ROBUST: consistent performance across datasets/scenarios
- FAST: 10-100x faster than competing methods

Detection Performance

- www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/

Approximating Multiscale Features

- Distribution of image statistics in natural images invariant to scale [Ruderman & Bialek 94]
- Statistics of image independent of scene area of single pixel

Integral Channel Features [BMVC09]

- Generate and compute features efficiently using integral images over multiple registered image channels (followed by feature selection via boosting)
- Integrate heterogeneous information, few parameters, state of the art performance, fast to compute, accurate spatial localization

Integral Channel Features

Exponential Scaling Law:

- Let \( f_L(x) \) denote the channel sum computer over \( I \) after downsampling by \( 2^L \)
- Expectation of \( f_L(x_0)/(a_2) \) should depend only on \( x_1-x_2 \)
- Thus expect that: \( E(f_L(x_0)/(a_2)) = E(f_L(x_0)/(a_1)) \rightarrow n \equiv 0 \)
- If so, then following relation must hold (derivation in paper):

\[
\frac{f_L(x)}{f_L(0)} \approx \frac{f_L(0)}{f_L(0)} e^{-2Ax}
\]

- Generally applicable (should hold for most feature types)

Feature Comp. = Feature Computation

Accurate  Fast

- Dense Image Pyramid  x  km/s/n4
- Classifier Pyramid  x  kn
- Hybrid Approach  x  4kn/3

k = cost per pixel, n = # of pixels, m = # of scales / octave