Person Detection

P J Narayanan

CS5765. Computer Vision. Spring 2013

CVIT, IIIT, Hyderabad

Person Detection

The Person Detection problem:

- Tell if a given image contains a human being (in upright position)
- If yes, draw a box around the persons to indicate the location

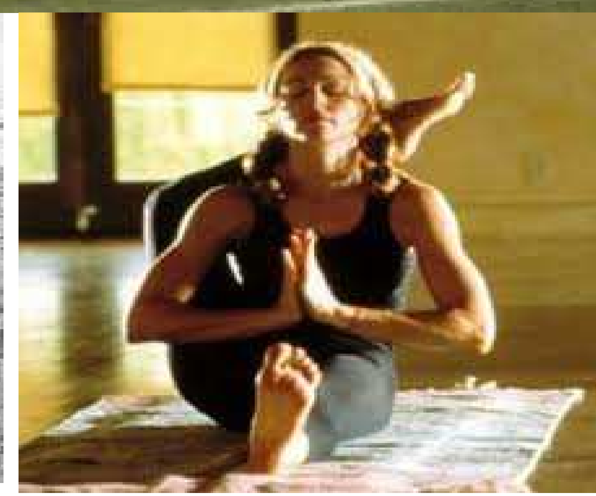
Why is this hard?

- Variation in viewpoints, scale
- Occlusion
- Variation in appearance
- Intra-class variance: Perhaps more than faces due to clothing

Different Cases





Why Person Detection?

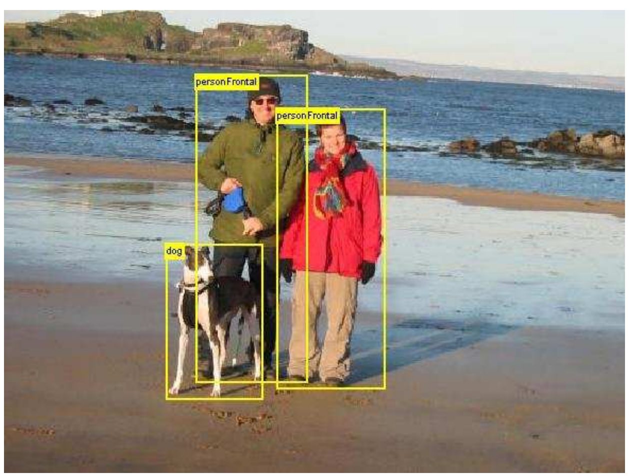
- Surveillance, Human-Computer Interaction, etc.
- A huge percentage of pictures that humans take contain humans!

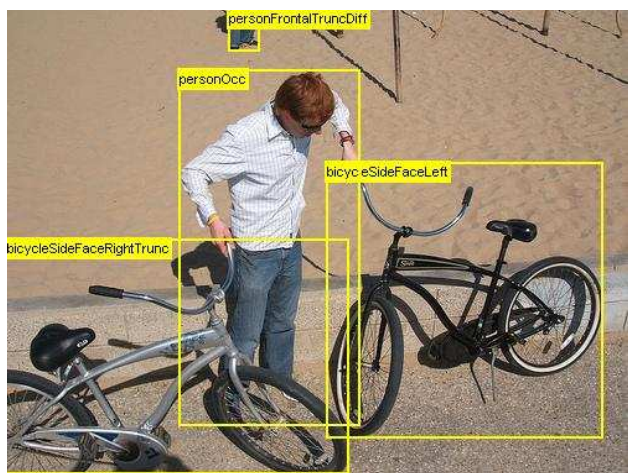
Some relevant prior efforts:

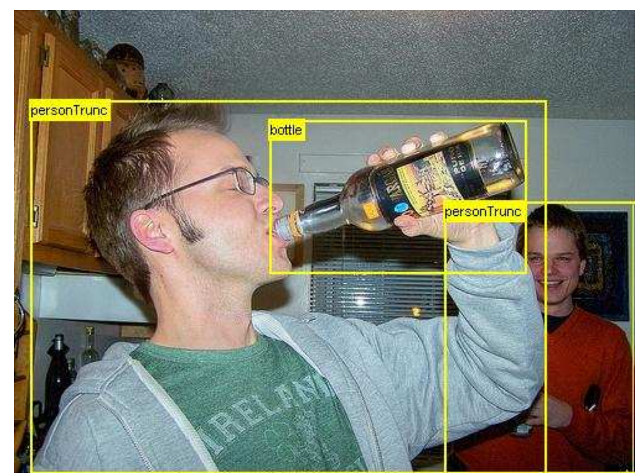
- **Gavrila/Philmin 1999:** Edges, distance transform, and template matching.
- **Papageorgiou/Poggio 2000:** Haar-features with SVM.
- **Viola and Jones 2003:** Extension of face detection
- **Dalal/Triggs 2005:** HOG features and SVM
- **Avidan et al. 2006:** Integral Histograms and cascade of classifiers for real-time person detection

Object Detection

- Recognizing object categories and marking its boundary
- Basic detector + sliding window for locations and scales
- Feature Extraction + Classification
  - Histograms, Haar, etc.
  - SVM, AdaBoost, Neural Networks, etc.



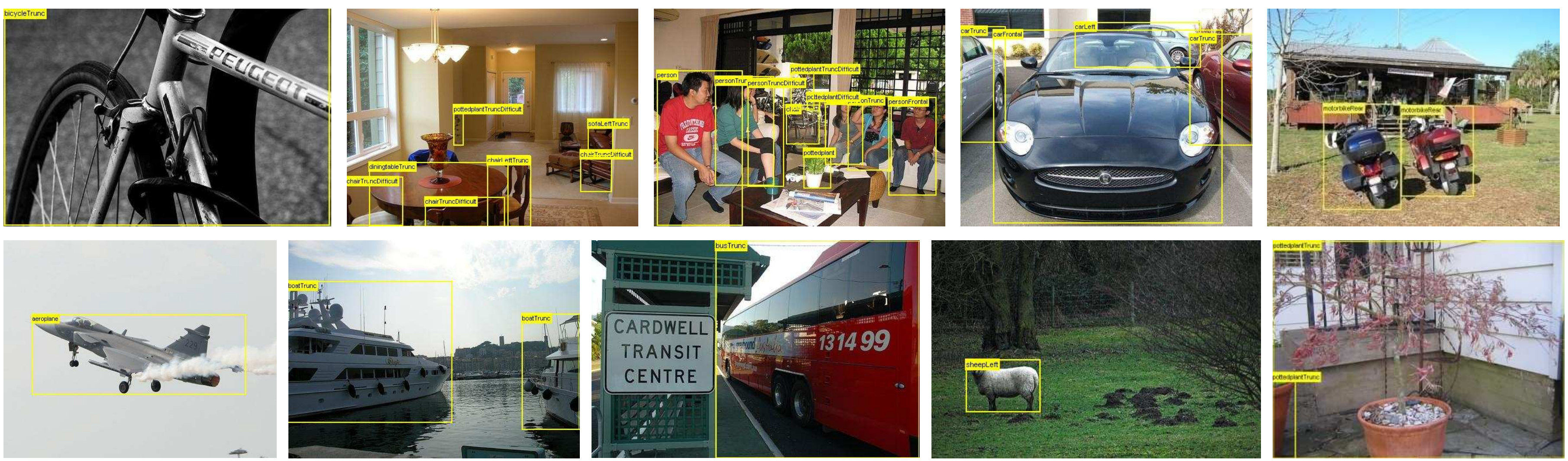




Material from PASCAL page/presentation

PASCAL VOC Challenge

- 20 object classes: aeroplane, bicycle, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, train, TV
- Downloaded from Flickr, not filtered for quality
- Complex scenes with size, occlusion, lighting, etc.
- Detection and annotation as per norms of all objects



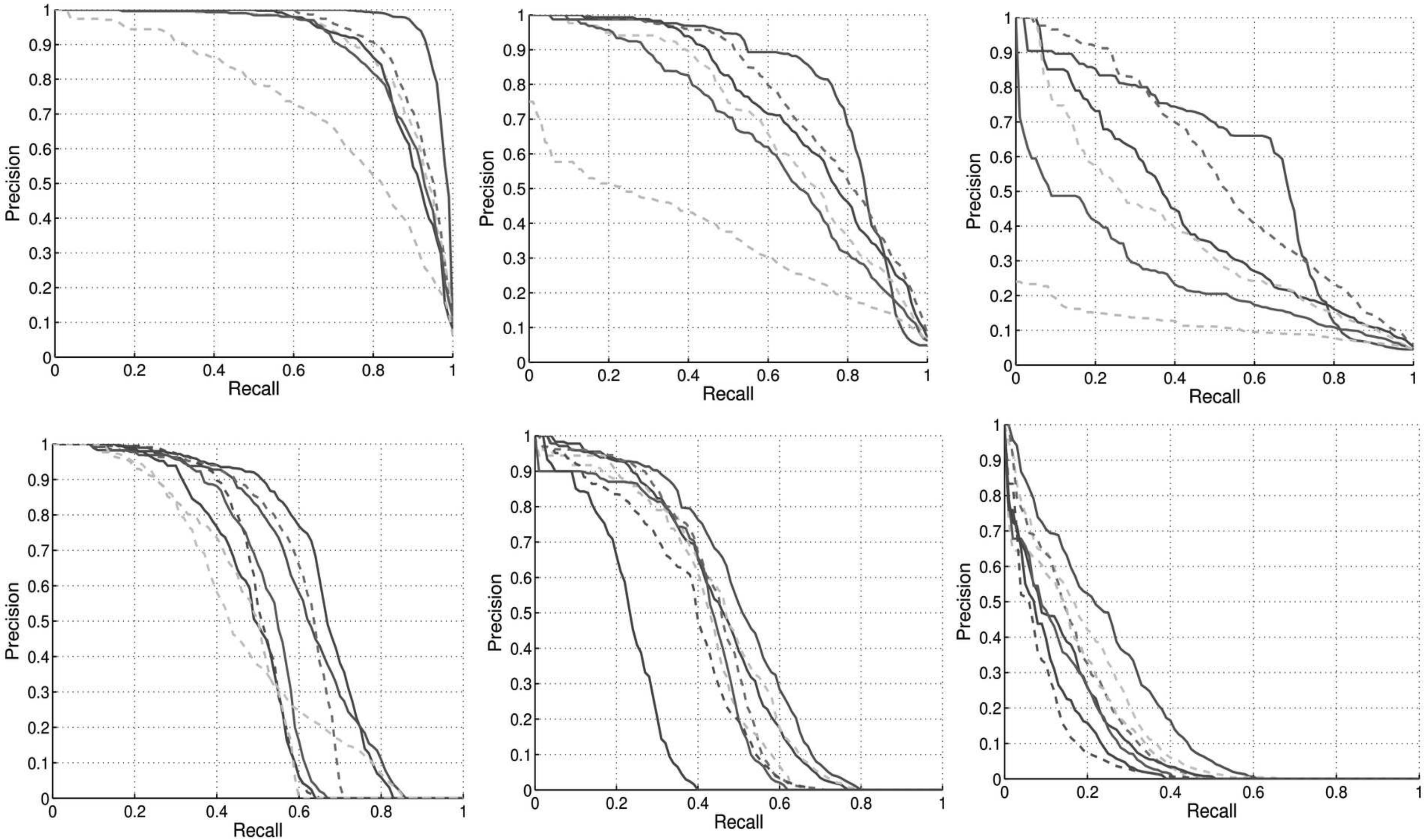
PASCAL 2011: Results

aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/ monitor
86.5	58.3	59.7	67.4	33.2	74.2	64.0	65.5	58.5	44.8	53.5	57.0	60.7	70.8	84.6	39.4	55.4	50.5	80.7	63.1
85.0	57.0	57.7	65.9	30.7	75.0	62.4	64.4	56.9	42.2	50.9	55.3	59.1	69.1	84.2	39.3	52.3	46.7	78.9	61.8
61.5	11.9	12.4	29.7	8.7	30.6	18.4	23.6	21.6	5.8	14.8	18.5	7.1	12.3	47.7	7.2	15.0	9.8	18.8	19.2
65.1	23.8	17.3	36.0	12.6	40.5	31.1	35.4	27.2	10.4	20.8	31.3	13.6	29.5	54.9	10.7	19.1	19.2	42.1	30.8
84.2	52.0	54.5	63.2	25.3	71.2	58.0	61.1	50.2	33.3	44.3	49.7	57.9	65.1	79.9	20.9	47.4	43.0	77.7	56.7
88.3	56.2	59.3	68.6	33.2	76.6	62.2	64.5	55.3	42.6	55.1	56.2	61.9	70.0	82.5	37.3	56.4	48.3	79.6	64.7
90.0	66.2	63.3	70.9	47.0	80.9	73.9	63.9	61.1	52.7	57.9	56.9	69.6	73.8	88.4	46.3	65.3	54.2	81.3	72.7
92.8	74.8	69.6	76.1	47.3	83.5	76.4	76.9	59.8	54.5	63.5	67.0	75.1	78.8	90.4	43.1	63.1	60.4	85.6	71.1
92.7	74.5	69.4	75.4	45.7	83.4	76.5	76.6	59.6	54.5	63.4	67.4	74.8	78.6	90.3	43.0	63.1	58.6	85.2	71.3
55.6	25.5	31.0	36.5	15.8	41.4	40.0	40.6	30.0	17.8	21.1	34.0	27.0	31.0	57.9	11.9	20.7	22.6	48.4	35.7
10.5	9.1	10.7	6.0	6.5	7.2	13.3	12.2	11.5	9.5	5.6	16.7	8.6	6.6	38.9	5.3	15.0	5.0	8.3	5.4
94.5	82.6	79.4	80.7	57.8	87.8	85.5	83.9	66.6	74.2	69.4	75.2	83.0	88.1	93.5	56.2	75.5	64.1	90.0	76.6
82.9	69.4	45.4	60.1	46.0	80.0	75.1	59.9	54.9	50.7	43.3	49.9	63.4	72.2	88.1	36.1	57.1	37.7	75.2	58.5
83.8	69.8	47.8	60.5	45.4	80.5	74.6	60.4	54.0	51.3	45.3	51.5	64.5	72.6	87.7	35.9	57.7	39.8	75.8	62.7
95.5	81.1	79.4	82.5	58.2	87.7	84.1	83.1	68.5	72.8	68.5	76.4	83.3	87.5	92.8	56.5	77.7	67.0	91.2	77.5
94.3	78.5	76.4	80.0	57.0	86.3	82.1	81.5	65.6	74.7	66.5	73.4	81.9	85.3	91.9	53.2	73.9	65.1	89.5	76.0
85.6	66.5	51.9	60.3	45.4	76.8	70.3	65.1	56.4	34.3	49.6	52.4	63.1	71.5	86.8	26.1	56.9	47.9	75.5	65.6
83.2	52.5	49.3	59.6	26.0	73.5	58.2	64.4	52.1	36.6	44.9	52.1	57.8	63.8	78.1	19.1	52.8	44.1	72.0	57.4
90.1	74.1	66.5	76.0	57.0	85.6	81.2	74.5	63.5	62.7	64.5	66.6	76.5	81.2	90.8	58.7	69.3	66.3	84.7	77.2

aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/ monitor
37.1	42.6	2.0	0.0	16.0	43.8	38.6	17.0	10.3	7.7	2.4	1.5	34.3	41.1	38.4	1.5	14.7	5.3	35.4	27.1
-	-	-	13.3	26.4	-	41.5	-	-	-	12.2	-	-	41.6	-	8.3	31.4	-	-	-
40.4	47.8	-	11.4	23.7	48.9	40.9	23.5	11.9	25.5	-	10.9	42.0	38.6	40.7	7.5	30.4	-	38.4	34.8
42.5	43.7	5.4	4.8	18.1	28.6	36.6	24.2	12.6	20.5	4.4	17.5	15.2	38.2	7.9	1.7	23.2	7.1	41.0	25.7
41.1	51.7	13.7	11.9	27.3	52.1	41.7	32.9	17.6	27.3	18.5	23.1	45.2	48.6	41.9	11.6	32.4	27.5	44.2	38.3
43.8	51.7	13.7	12.7	27.3	51.5	43.7	32.9	18.3	27.3	18.5	23.1	45.2	48.6	42.9	11.6	32.4	27.5	47.0	39.3
55.0	58.1	22.5	18.8	33.9	57.6	54.5	42.6	20.2	40.3	29.3	37.1	54.6	58.3	51.6	14.7	44.8	32.1	51.7	41.0
51.4	52.9	20.1	15.7	26.9	53.0	45.6	37.6	15.2	36.0	25.1	32.6	50.4	55.8	36.8	12.3	37.6	30.5	48.1	41.0
56.3	55.9	23.4	20.3	27.2	56.6	48.1	53.8	23.2	32.9	33.3	39.2	53.0	56.9	43.6	14.3	37.9	39.4	52.6	43.7
56.0	53.3	19.2	17.2	25.8	53.1	45.4	44.5	20.1	32.1	28.1	37.2	52.3	56.6	43.3	12.1	34.3	37.6	51.8	45.2
53.2	53.9	13.1	13.5	30.5	55.5	51.2	31.7	14.5	29.0	16.0	22.1	43.1	50.3	46.3	8.8	33.0	22.9	45.8	38.2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	49.2	-	-	-	-	-
56.9	43.4	16.6	15.8	18.0	52.3	38.3	48.9	12.2	29.7	32.8	36.7	45.7	54.4	30.4	16.2	37.2	34.7	45.9	44.2

PASCAL 2012: 3 Categories

Aeroplane, TV Monitor, Potted Plant



Dalal-Triggs HoG-Based Person Detector

Navneet Dalal and Bill Triggs

CVPR 2005

DT Person Detector

- A highly successful research by Dalal and Triggs. Appeared in CVPR 2005.
- Designed for pedestrians or people in upright positions, usually standing.
- Feature: **Histogram of Oriented Gradients (HoG)**
  Classifier: **Support Vector Machines (SVM)**
- An important landmark in the area of person detection. Still a competitive method.
- Made real-time using Integral HoG and a cascade of rejectors by Avidan et al. in CVPR 2006.
- Code available from the authors; widely used.

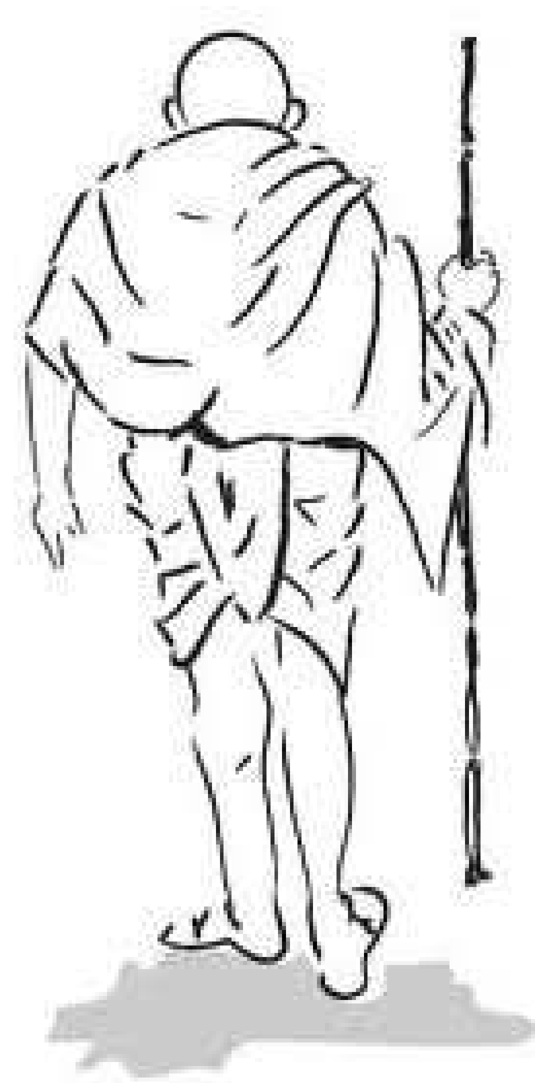
Appearances Are Deceptive!

- People are roughly similar, but not their appearances due to diverse dressing patterns.
- Appearance-based methods cannot distinguish persons from non-persons easily.
- How about *variations* in appearance instead?
- How much information do derivatives contain?

A lot!!

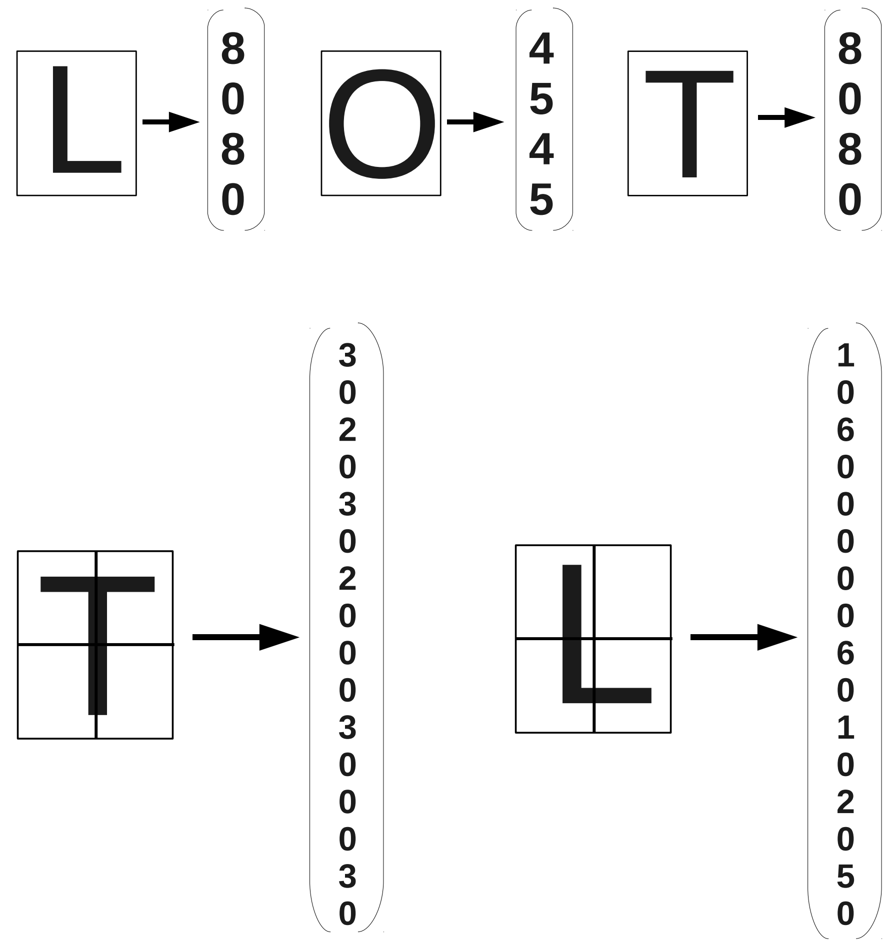
Gradient-Based Representations

- 99% of the data thrown away. Information still remains!
- Information is in boundaries or derivatives and not in regions!
- Gradient: Apply derivative operators in X and Y directions
- Gradient Magnitude: $\sqrt{I_x^2 + I_y^2}$
- Gradient Direction: $\arctan\left(\frac{I_y}{I_x}\right)$
- Sobel, Prewitt, etc. Simple: $[-1 \ \ 0 \ \ 1]$, $[-1 \ \ 0 \ \ 1]^{\mathbf{T}}$



Gradient Histograms

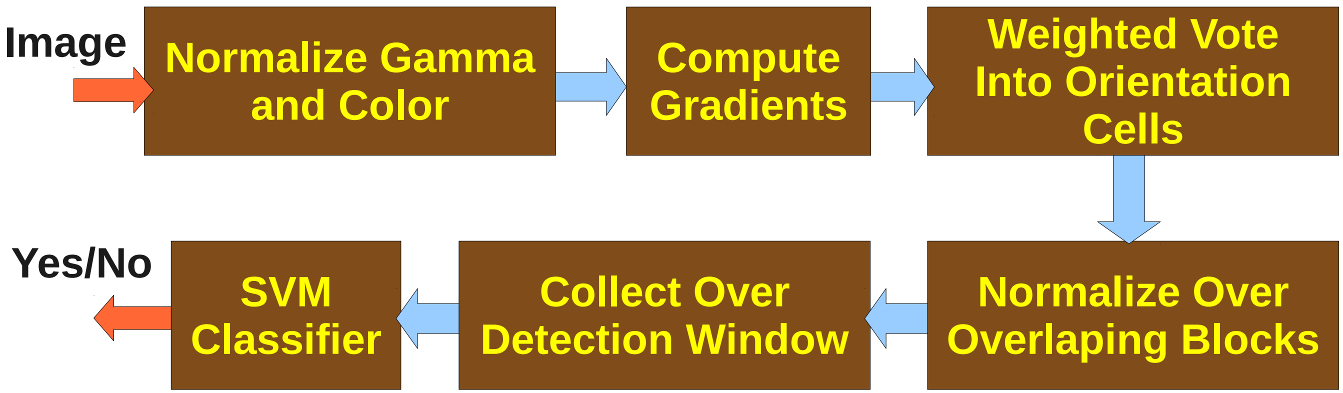
- Count the number of pixles with gradient directions 0, 45, 90, and 135 degrees in the window
- A global property of the window.
- Not discriminative enough?
- Divide image into regions; count gradients in each. Concatenate to form a single feature vector.



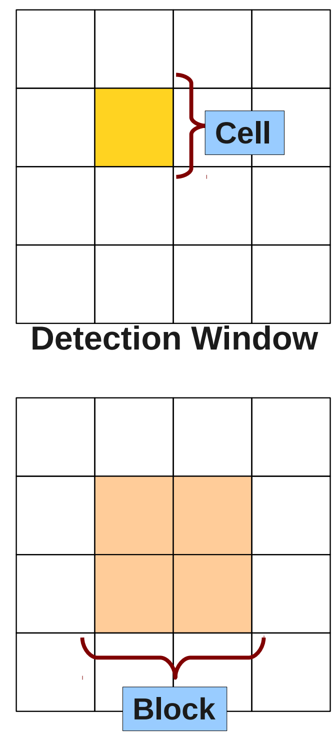
Gradient-Weighted Histograms

- Go beyond counting. Use gradient magnitude as the "vote" to the bin determined by the gradient direction.
- Bilinear interpolation of weights to spatial and orientation bins
- Counting is not powerful enough.
- Orientations can be in many directions, "signed" and "unsigned", etc.
- Spatial cell sizes can vary

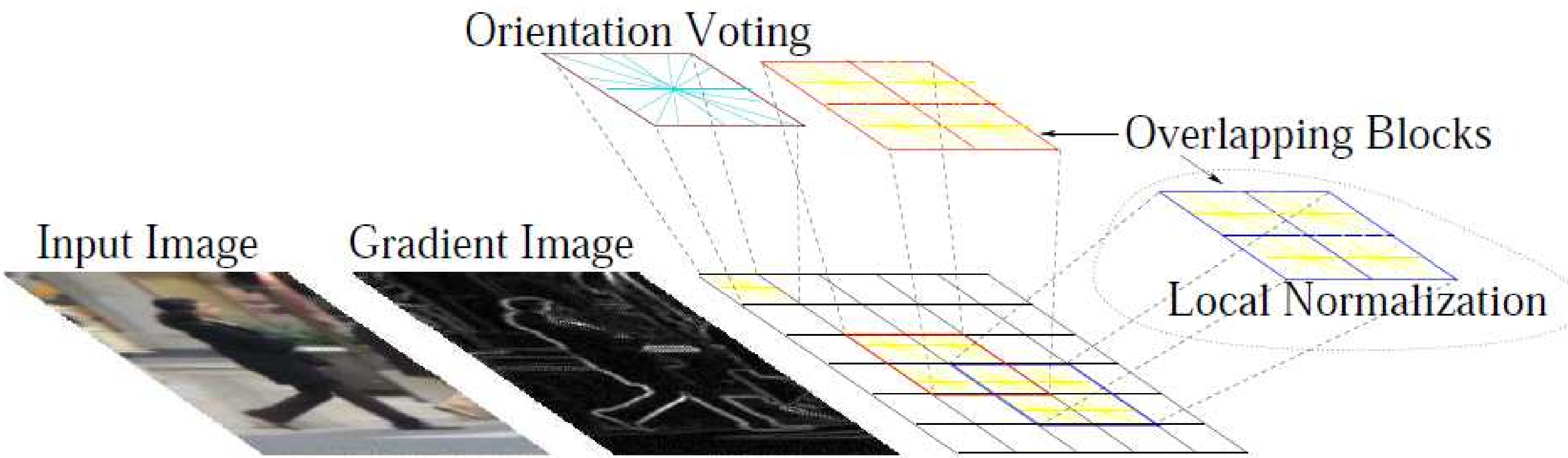
Overall Idea



- Divide window into cells, group cells into blocks
- Compute of gradients in each cell; compute histogram of gradients
- Normalize the gradient vector
- Classify using Support Vector Machines

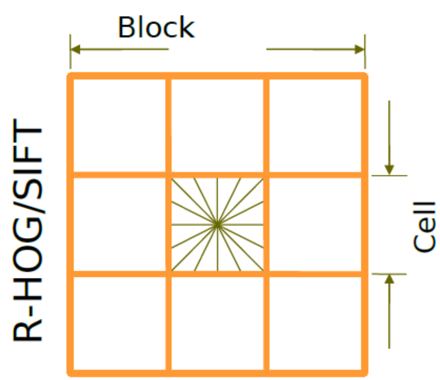


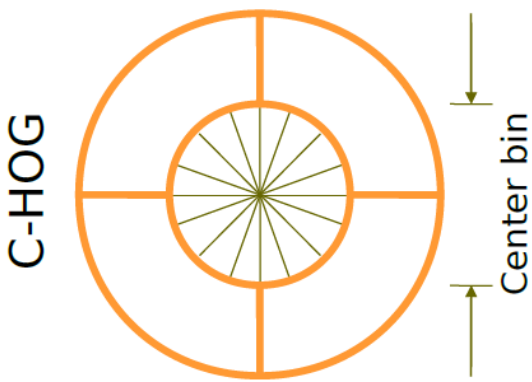
Overall Idea



Details

- 64×128 detection window, 8×8 cell, 2×2 block
- 9-bin orientation histograms
- $4 \times 9 = 36$ entries per block
- 7×15 blocks per detection window
- HOG descriptor: $36 \times 105 = 3780$ dimensional vector!
- R-HOG: Rectangular HOG and C-HOG: Circular HOG.

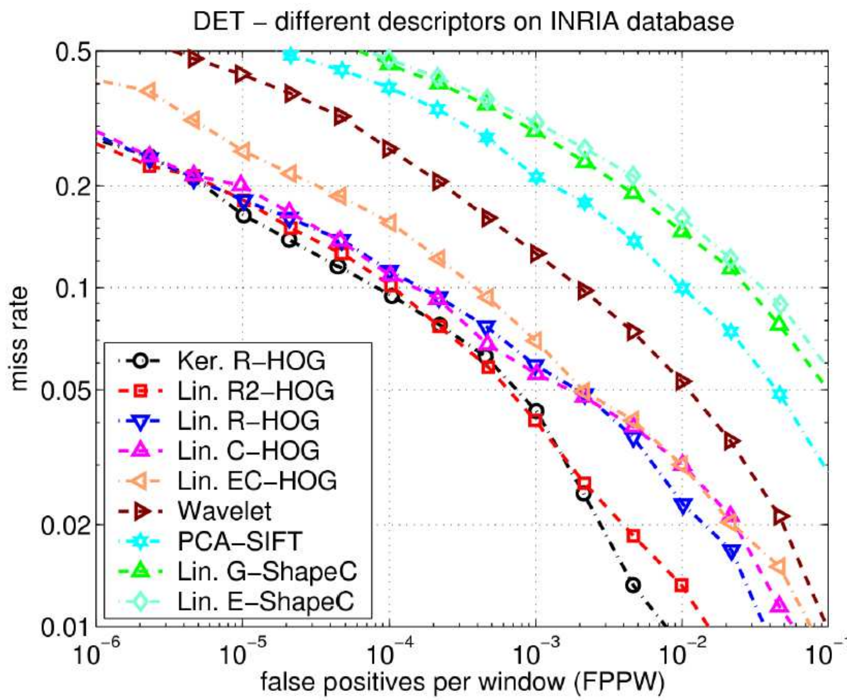


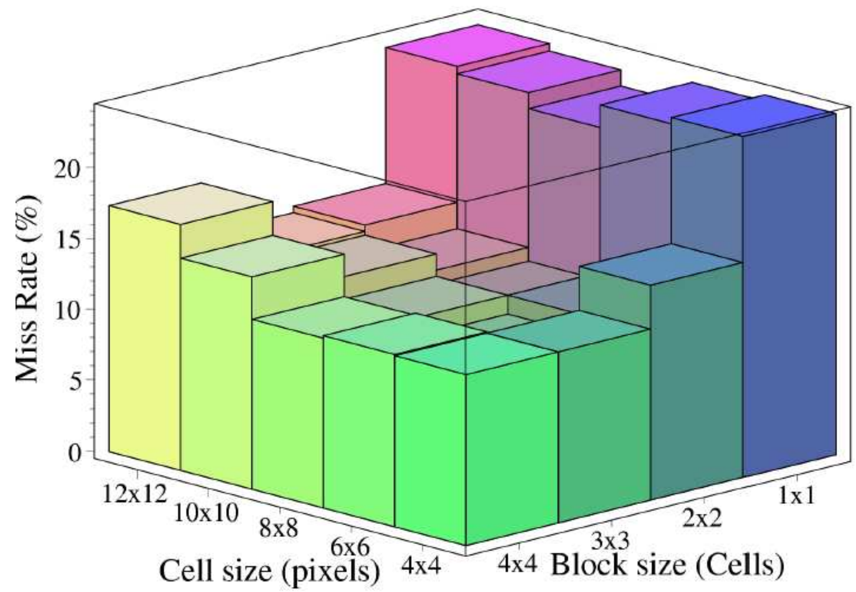


More Details

-
- Training dataset:
  - 1239 positive images with their mirror reflections
  - 12180 random patches from 1218 images with no persons as negative examples
  - Train once and run on 1218 negative images. Add false positives as hard examples for a **retraining** step. Performance up by 5% at 10^{-4} FPPW
- Testing: Created own dataset called INRIA dataset

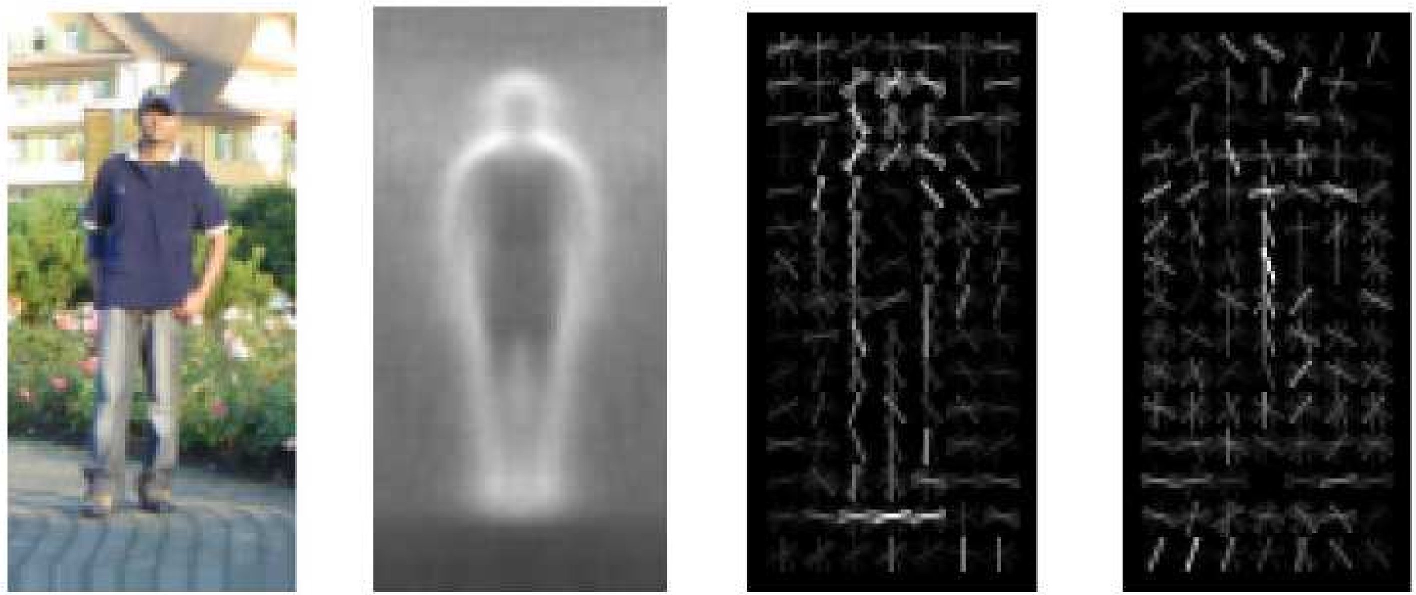
Some Results





Detection Error Tradeoff (DET): Miss rate (i.e., 1 - recall) against FPPW or False Positives Per Window.

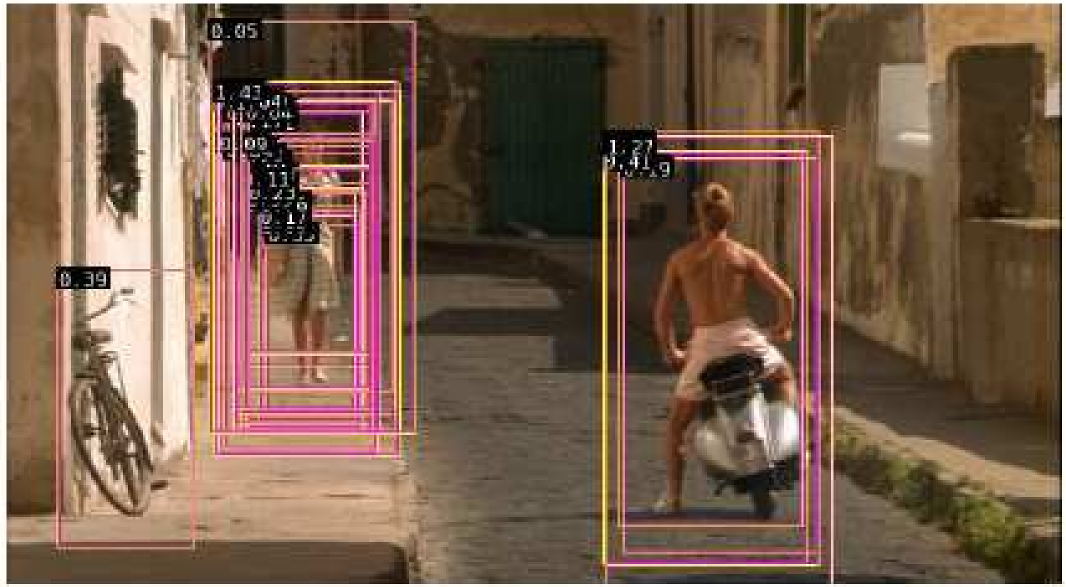
What's Important?

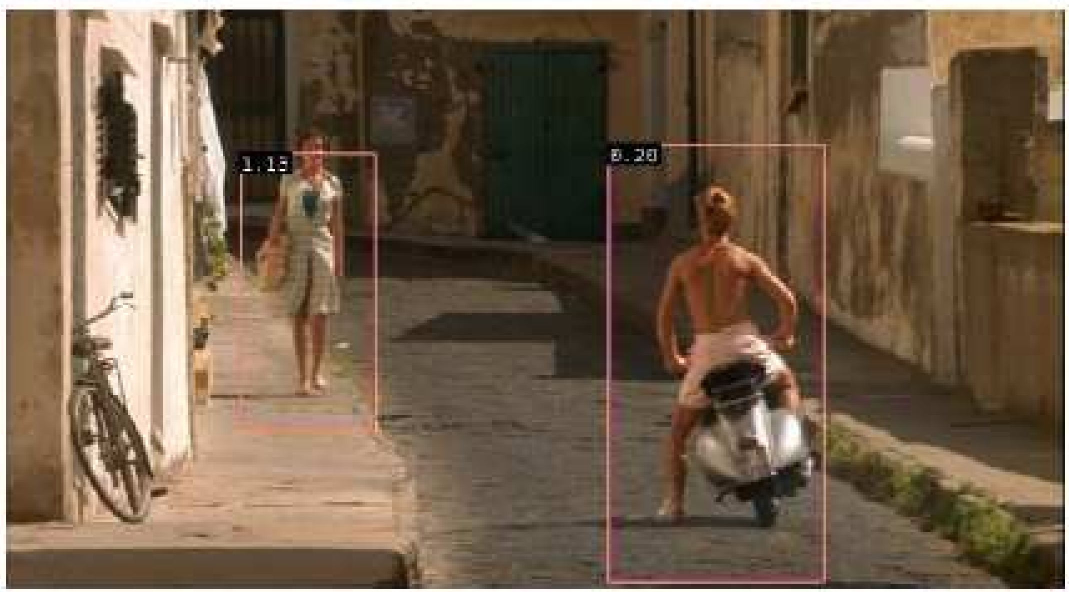


Input example | Average gradients | Weighted pos wts | Weighted neg wts

Head and neck seem to be weighted positive by SVM.
Interior is weighted negative.

Multiple Detections





Performance Study

Performance by varying many parameters by Dalal.

- Gamma and colour normalization: Not much effect. Colour helps. Use largest magnitude of all channels
- Simple $[-1 \;\; 0 \;\; 1]$ gradient with no smoothing best!
- 9 orientation bins between 0 to 180 degrees. Interpolate magnitude between neighbouring bins
- Normalizing 36-vectors per block worked well. Simple L_2 norm is fine.
- 16-pixel border for context enhances performance
- Gaussian kernel for SVM improves performance

Human detection rate of 90% with 10^{-4} false positives per window (FPPW).

D&T's Conclusions

- HOG outperforms features like wavelets
- Do not smooth images at start to remove noise!
- Instead, detect edges at fine level with gradient magnitude voting. Smooth spatially later
- Local contrast normalization is very useful

Read their CVPR05 paper. A good example of performing *exhaustive* experiments to form conclusions on a typical, training based recognition problem.

Thank You!

Many figures are from different web sources