DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF COPENHAGEN

Object Detection and Recognition II

Kim Steenstrup Pedersen

Plan for today

- Dense feature template approaches:
  - Histogram of Oriented Gradients (HoG)
  - Deformable parts models (DPM)

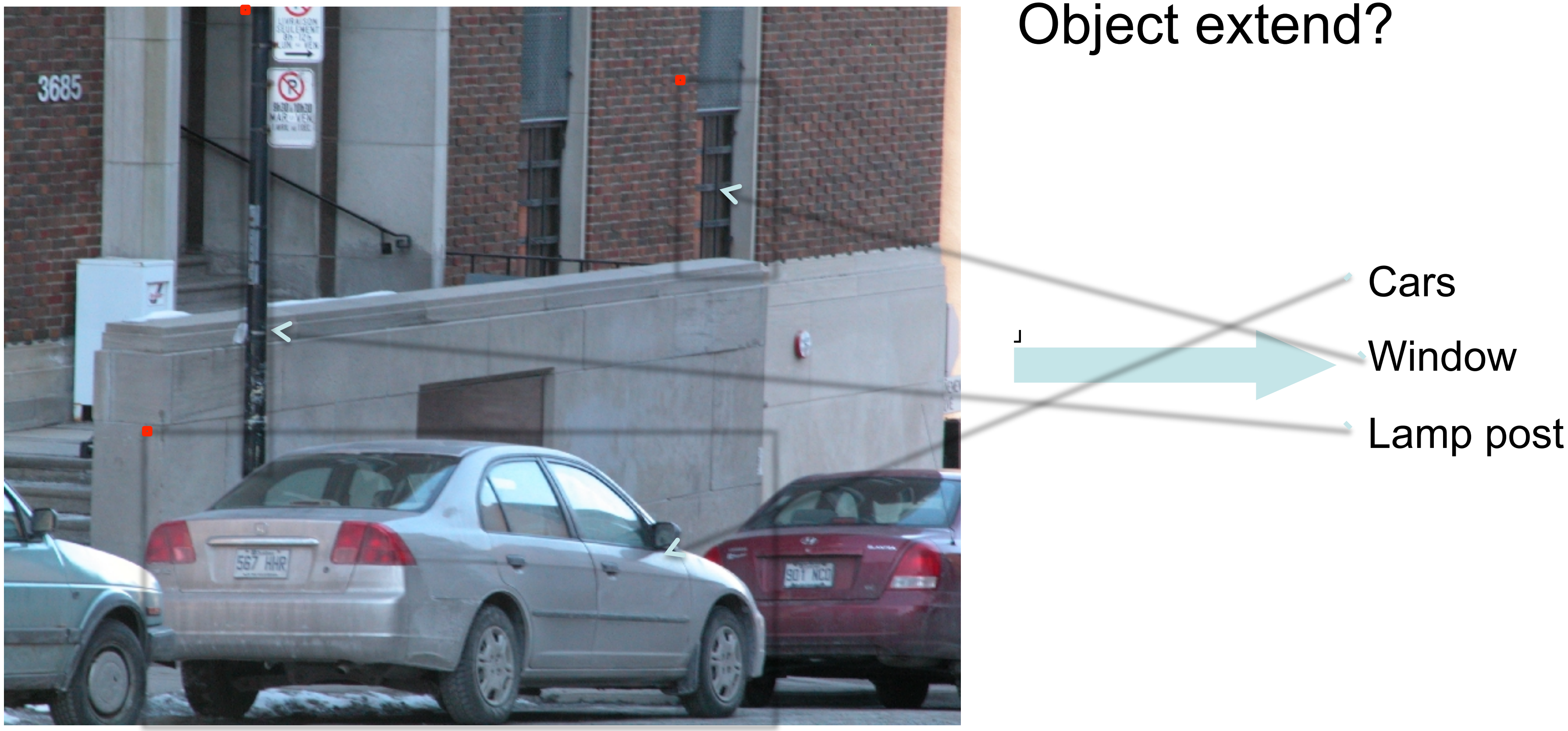
- Exercise class: Work on Assignment 2

Object detection and recognition

What is in this image?

Where?

Object extend?



Dense feature template approaches

Dense feature template approaches

Training:

- Extract features on object examples
- Train one or more classifiers to discriminate between object categories and background category

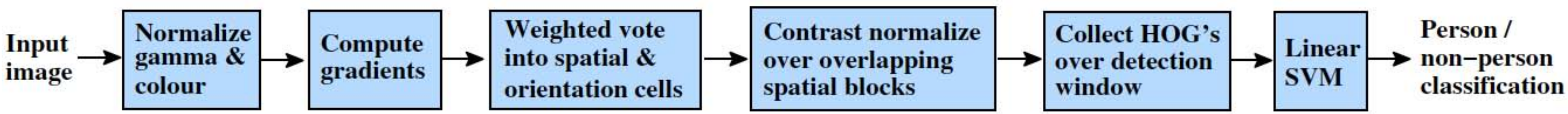
Query image:

- Apply the classifier to a detection window
- Slide the detection window across the image at all positions and scales
- Reduce nearby matches to avoid double detection (non-maxima suppression)

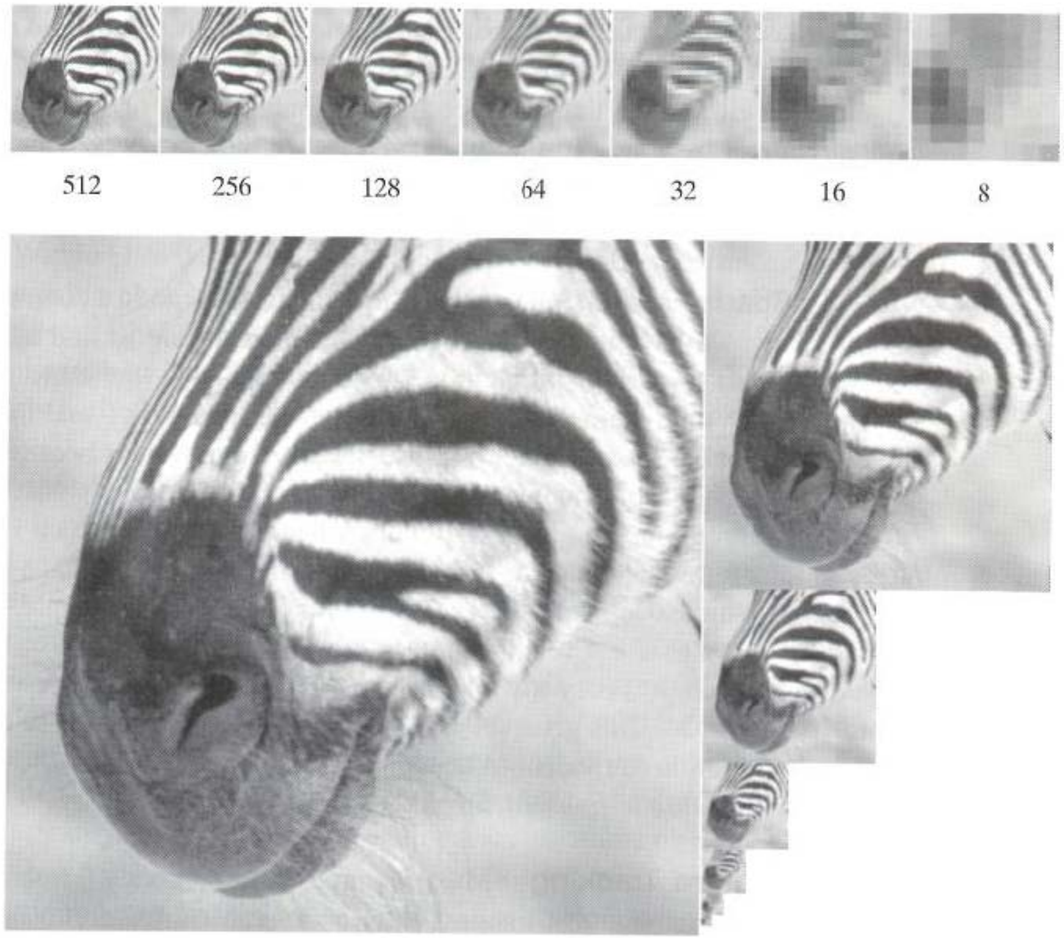
Example: Detecting people / pedestrians

Object detection using Histograms of Oriented Gradients (HoG) features (Dalal & Triggs'05)

- Using the sliding detection window approach $(64 \times 128 \text{ pixels})$.
- Normalization of image prior to detection:
  - Use all RGB channels
  - Optional: Apply power law gamma correction to each channel $I_{\text{out}} = cI_{\text{in}}^{\gamma}$
- Compute gradient image pyramid:

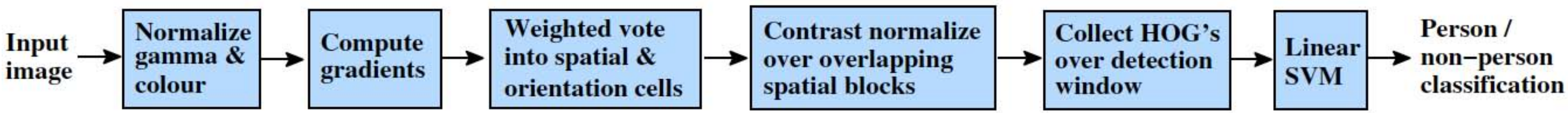


Recall: The Gaussian pyramid



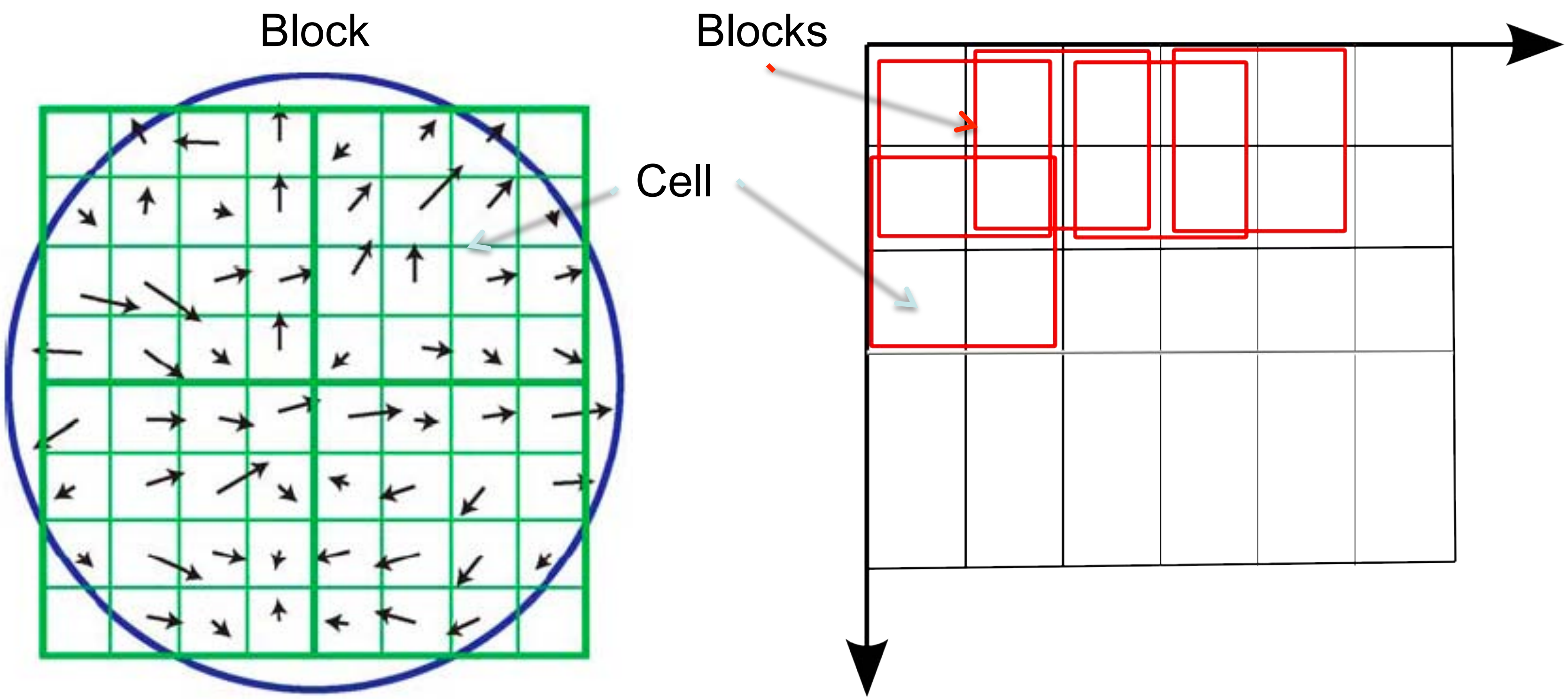
Object detection using Histograms of Oriented Gradients (HoG) features (Dalal & Triggs'05)

- Using the sliding detection window approach $(64 \times 128 \text{ pixels})$.
- Normalization of image prior to detection:
  - Use all RGB channels
  - Optional: Apply power law gamma correction to each channel $I_{\text{out}} = cI_{\text{in}}^{\gamma}$
- Compute gradient image pyramid:
  - For each color channel, compute intensity gradients
  - For each pixel, pick the gradient from the color channel with largest gradient magnitude (simple color gradient)
  - Detail: Dalal & Triggs do not pre-smooth the image and use the derivative approximation filter $[-1,0,1]$



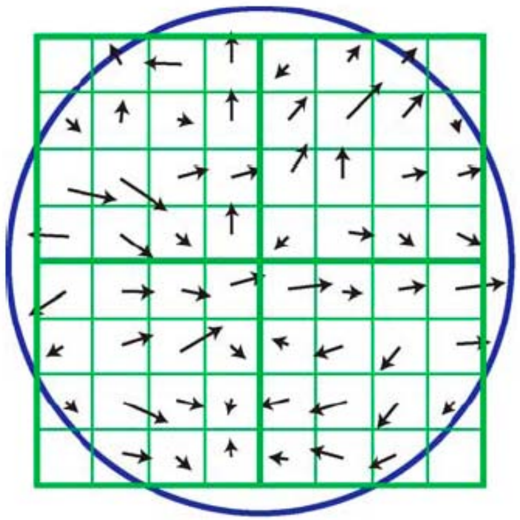
Histograms of Oriented Gradients (HoG) feature
(Applied to detection window)

- Divide the detection window into 8 x 8 pixels non-overlapping cells.
- Divide the detection window into 16 x 16 pixels overlapping (8 pixel stride) blocks covering 2 x 2 cells.



Histograms of Oriented Gradients (HoG) feature
(Applied to detection window)

- In each cell:
  - Compute a 9-bin gradient orientation histogram for the range $0°$ to $180°$ (different from SIFTs range of $0°$ to $360°$)
  - Use gradient magnitude weighting and linear interpolation in neighbor bins (just as in SIFT).
  - Use a Gaussian window function centered on the block to weigh gradients magnitude contributions to histograms as a function of distance to block center (just as in SIFT).

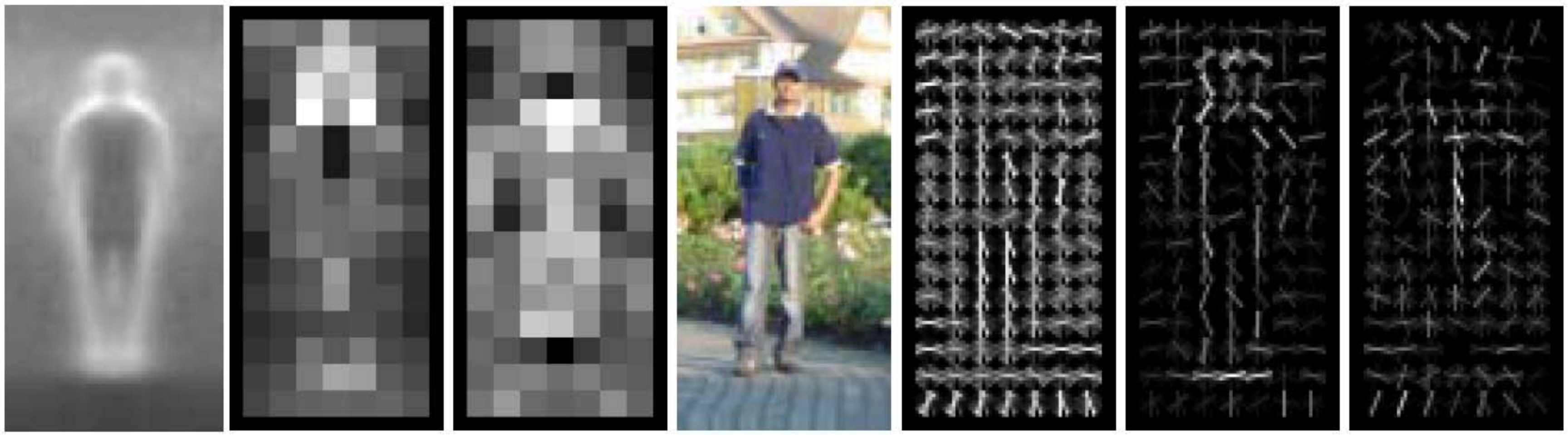


Histograms of Oriented Gradients (HoG) feature
(Applied to a detection window)

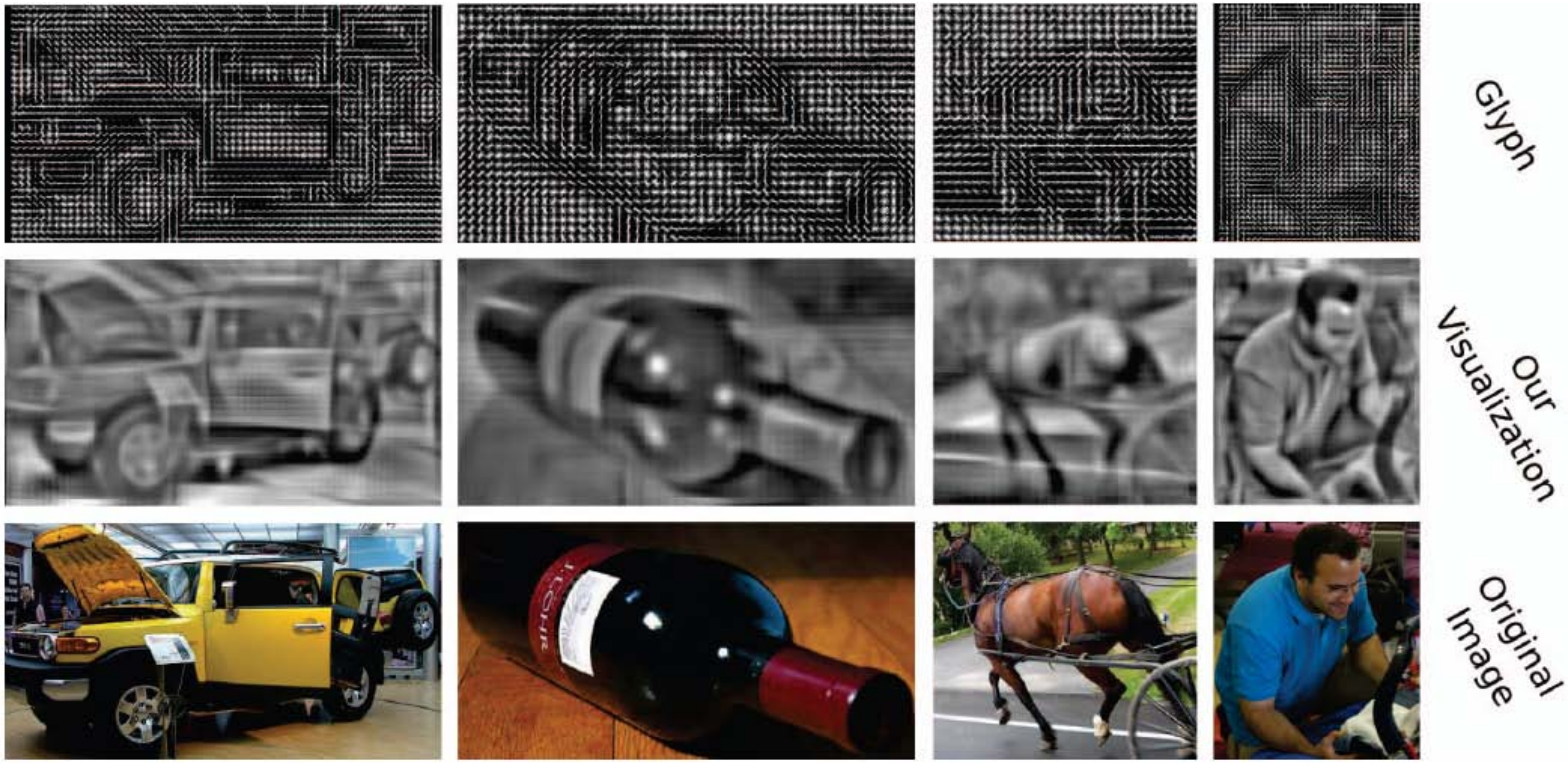
- For each block:
  - Concatenate cell histogram vectors to a feature vector $\mathbf{F}$
  - Normalize feature vector:
    - Euclidean: $\mathbf{F} = \mathbf{F} \Big/ \sqrt{\|\mathbf{F}\|^2 + \varepsilon^2}$
    - Peak clipping followed by renormalization (just as in SIFT)

- For detection window:
  - Concatenate block feature vectors to form a joint feature vector for whole of detection window
  - Dimensionality for 64 x 128 = 8192 pixels detection window: 9 bins x (7 x 15) blocks = 945 dimensions
  - Apply a classifier to the joint feature vector – the detection window (Dalal & Triggs uses a linear Support Vector Machine (SVM))

HoG features visualized



More human-friendly visualization of HoG



Vondrick et al 2013

Dense feature template approaches

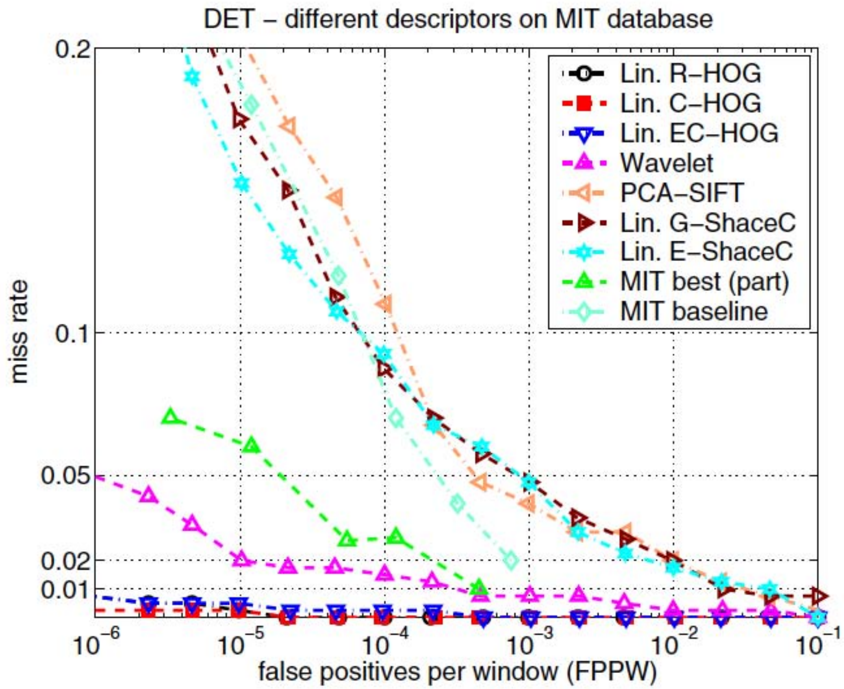
Training:

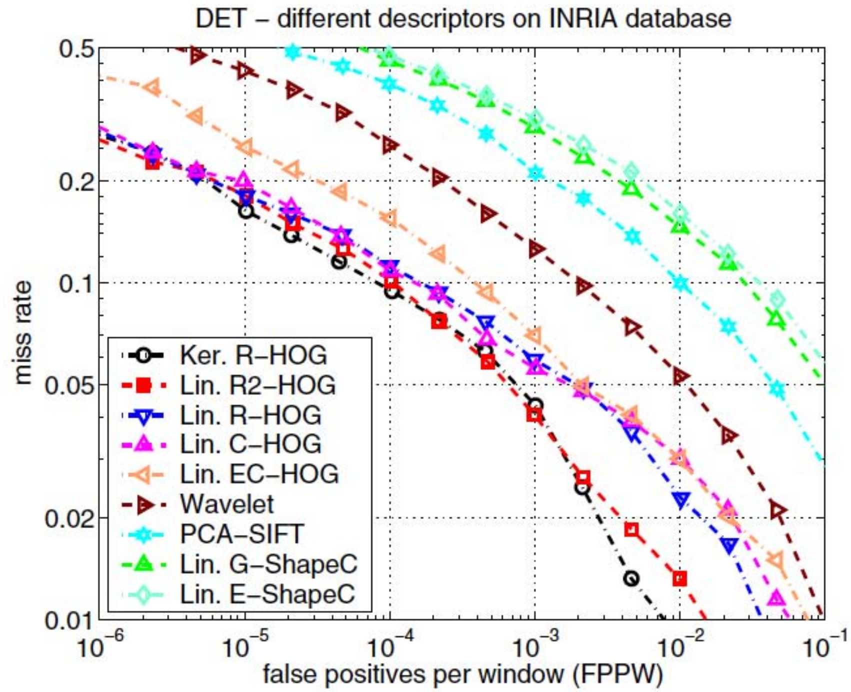
- Extract features on localized object examples
- Train a classifier to discriminate between object categories and background category

Query image:

- Apply the classifier to a detection window
- Slide the detection window across the image at all positions and scales
- Reduce nearby matches to avoid double detection (non-maxima suppression)

HoG results on detecting humans





Confusion matrix revisited

- Receiver operating characteristic (ROC):
  - TPR = TP / P (Recall)
  - FPR = FP / N
- Miss rate:
  - 1-recall = FN / P
- False positives per window tested:
  - FPPW = FP / (P+N)

		actual value		
		p	***n***	**total**
prediction outcome	***p'***	True Positive	False Positive	P'
	n'	False Negative	True Negative	N'
	total	P	N	

From Wikipedia 2012

Problems

This approach:

- Not really robust to occlusions – e.g. body parts being partially hidden, or intra-class variation.

- Not robust to rotations, e.g. a rotation of the object in the image plane.

Part based models

- Objects consist of parts which can help resolve partial occlusion.

- Being able to detect parts can provide cues about the presence of an object (feedback loop).

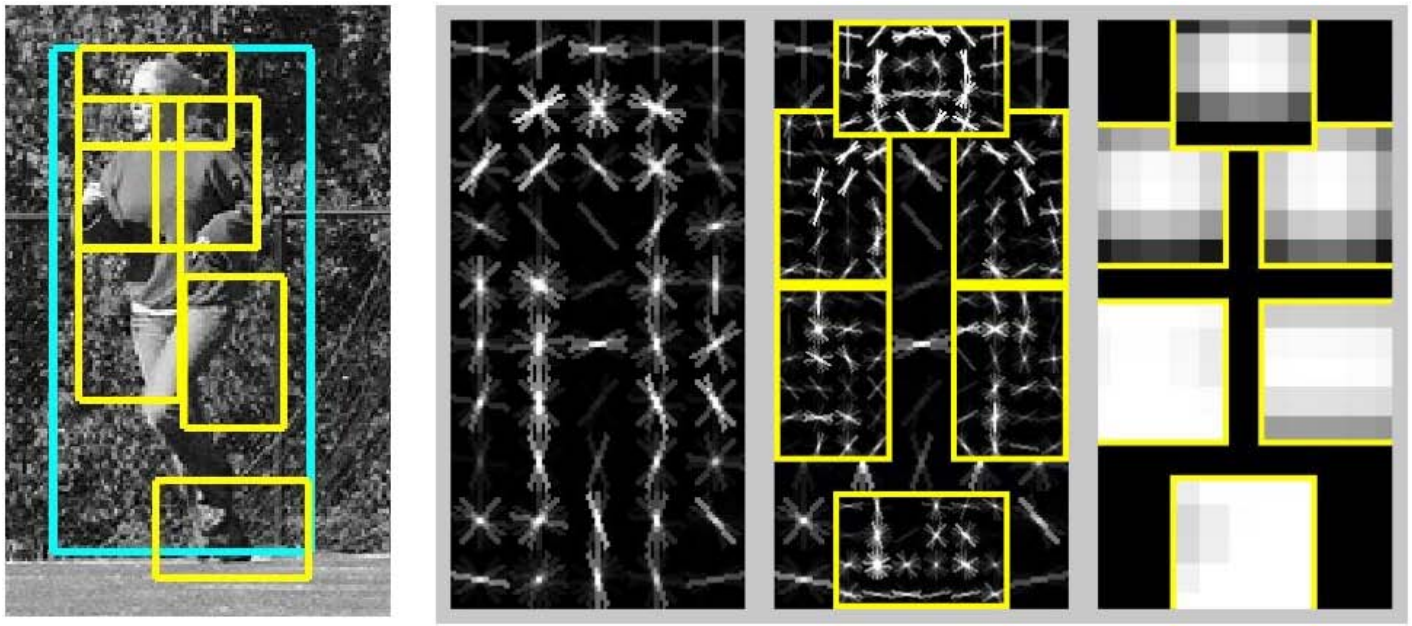
- Parts allow us to generalize models to handle large intra-class variation.

Deformable Parts Model (DPM)

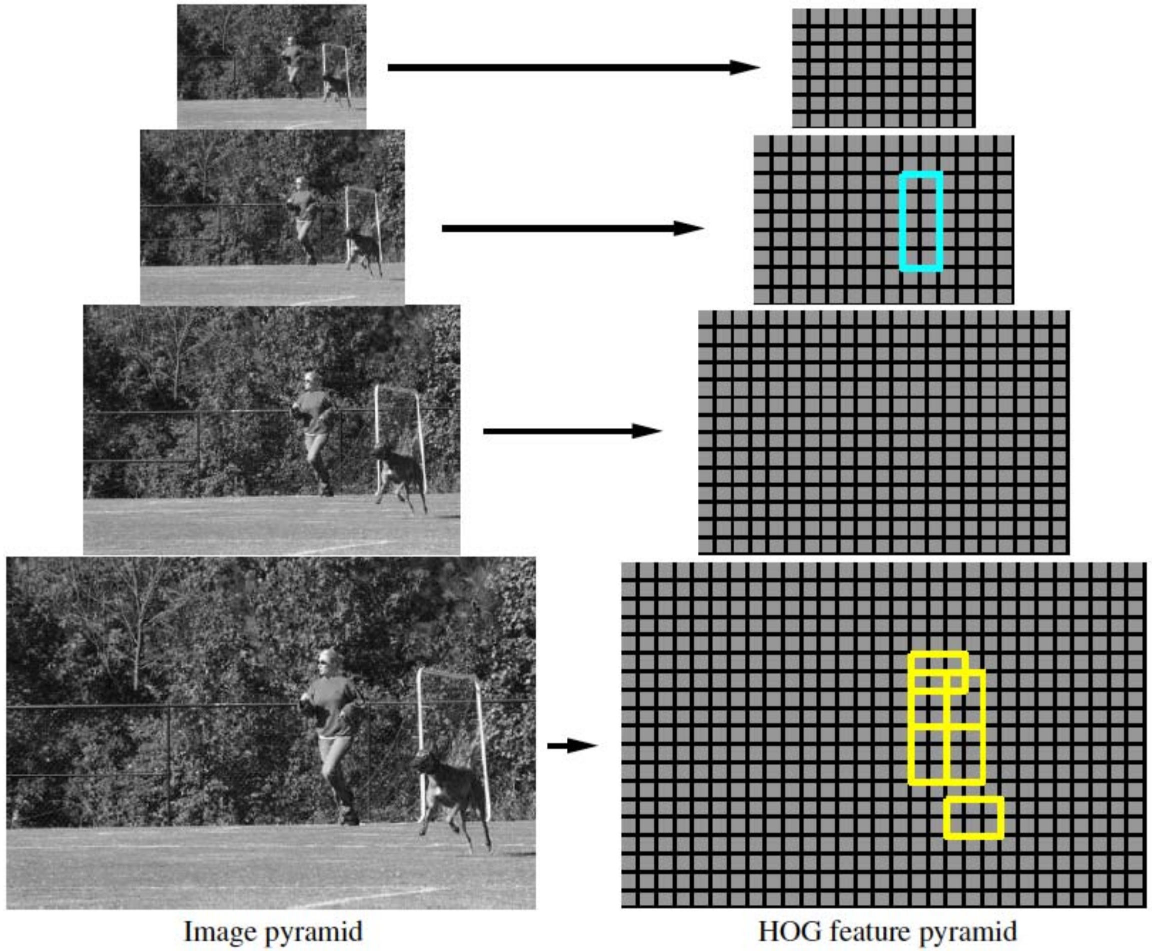
by Felzenszwalb et al CVPR 2008 & IJCV 2010

Deformable parts model consists of:

- A root filter (HoG)
- Parts models (HoG at higher resolution)
- Spatial parts placement model within root filter (a quadratic cost function per part)



Multi-scale detection in DPM

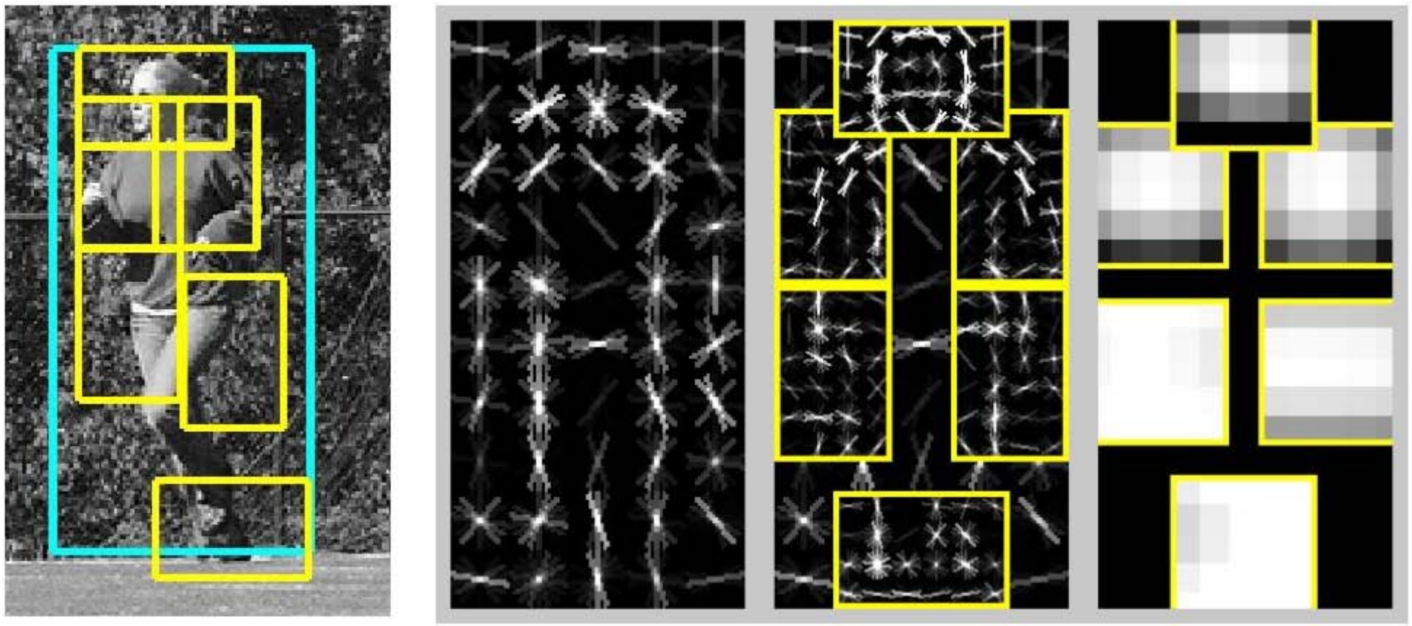


Deformable Parts Model (DPM)
by Felzenszwalb et al CVPR 2008 & IJCV 2010

- Detection by filtering HoG pyramids

$$\text{Response}(x,y,l) = \sum_{x',y'} F(x',y') \cdot \text{HoG}(x+x', y+y', l)$$

- Apply filter across the HoG pyramid and compute Score(x,y,l) = Root + Parts – Deformation cost



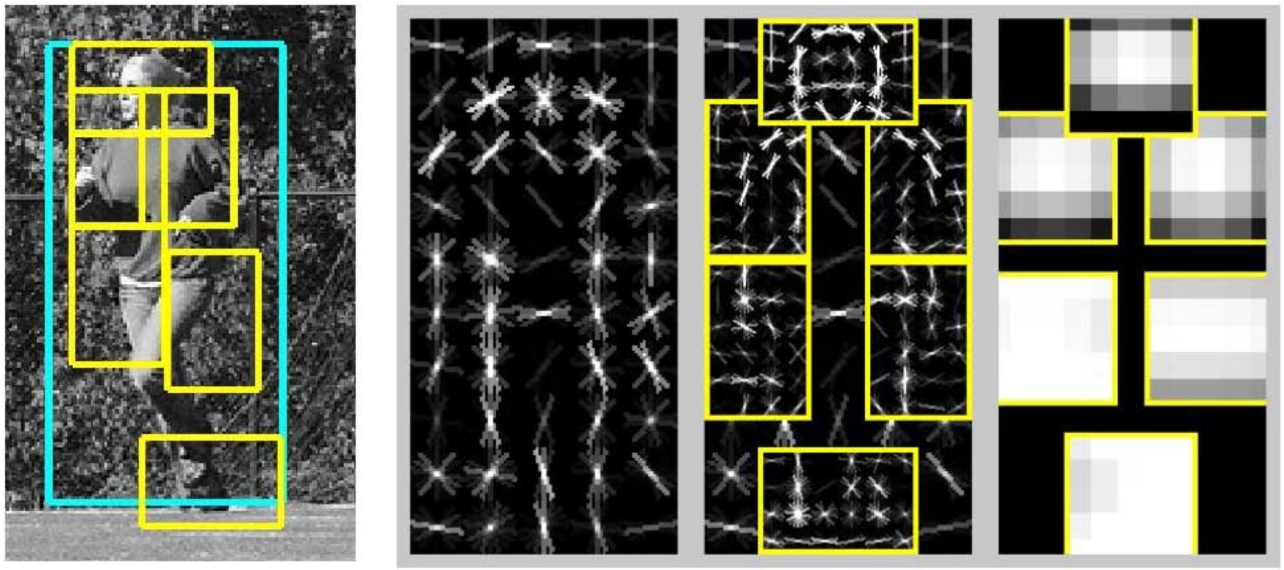
Deformable Parts Model (DPM)
by Felzenszwalb et al CVPR 2008 & IJCV 2010

- Filters and parts displacement parameters are learnt with a SVM classifier using latent variable SVM classifier

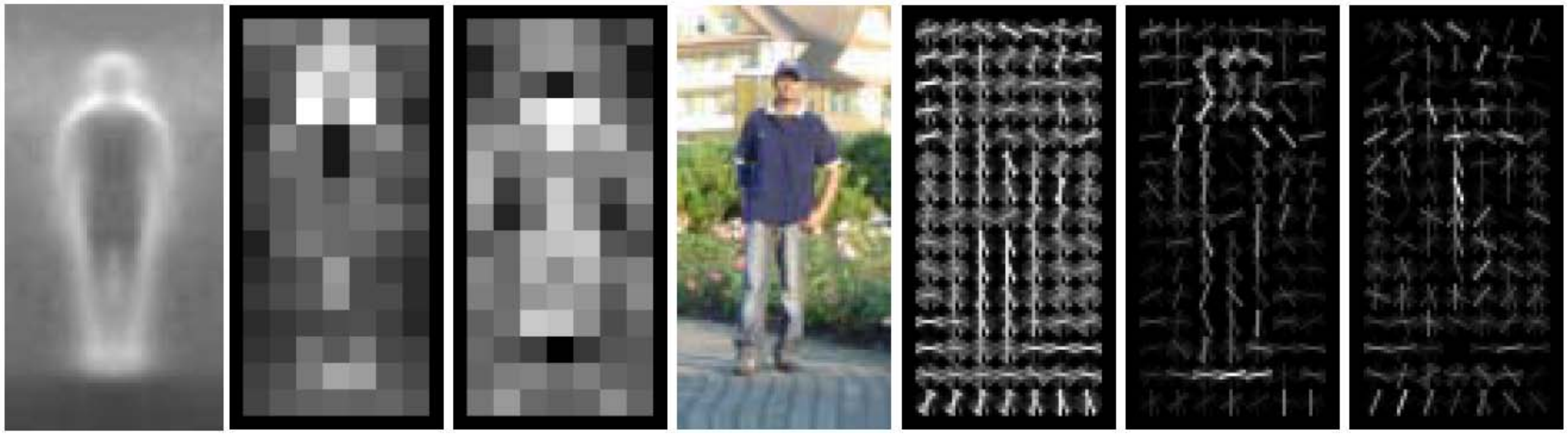
$$f_\beta(x) = \max_z \beta \cdot \Phi(x,z) \quad \text{, labels } f_\beta(x) > 0 \ \text{ or } \ f_\beta(x) \le 0$$

$$\beta = \left(F_0, F_1, \ldots, F_n, d_1, \ldots, d_n\right)$$

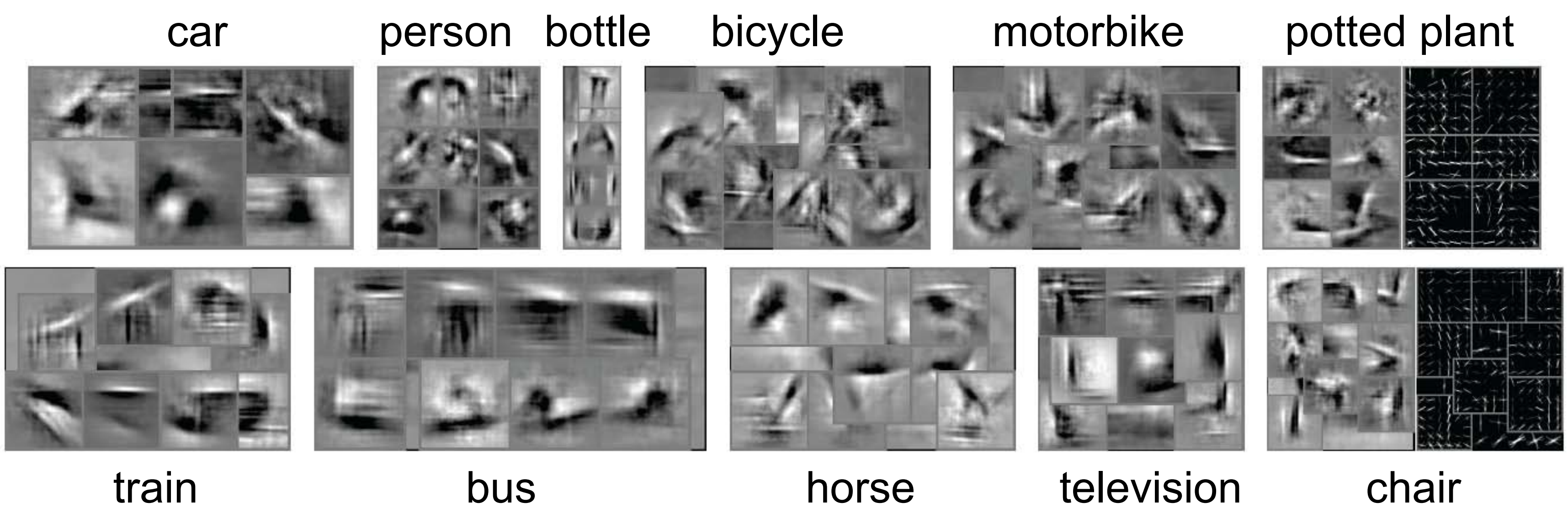
$$\Phi(x,z) = \left(\mathrm{HoG}_0(x), \mathrm{HoG}_1(x), \ldots, \mathrm{HoG}_n(x), -\phi_d(x_1), \ldots, -\phi_d(x_n)\right)$$



Root classifier similar to original HoG

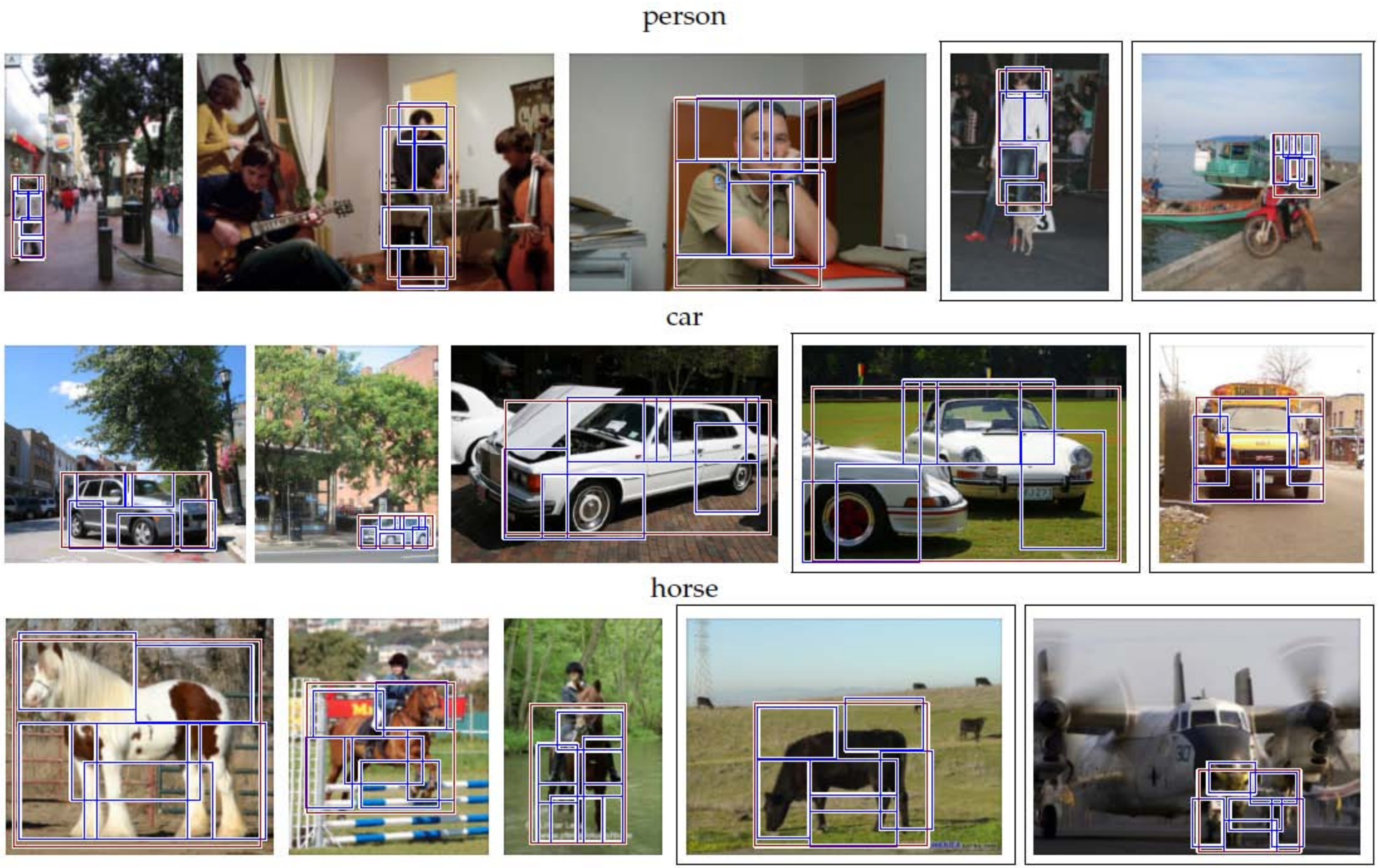


Visualization of some DPM parts features

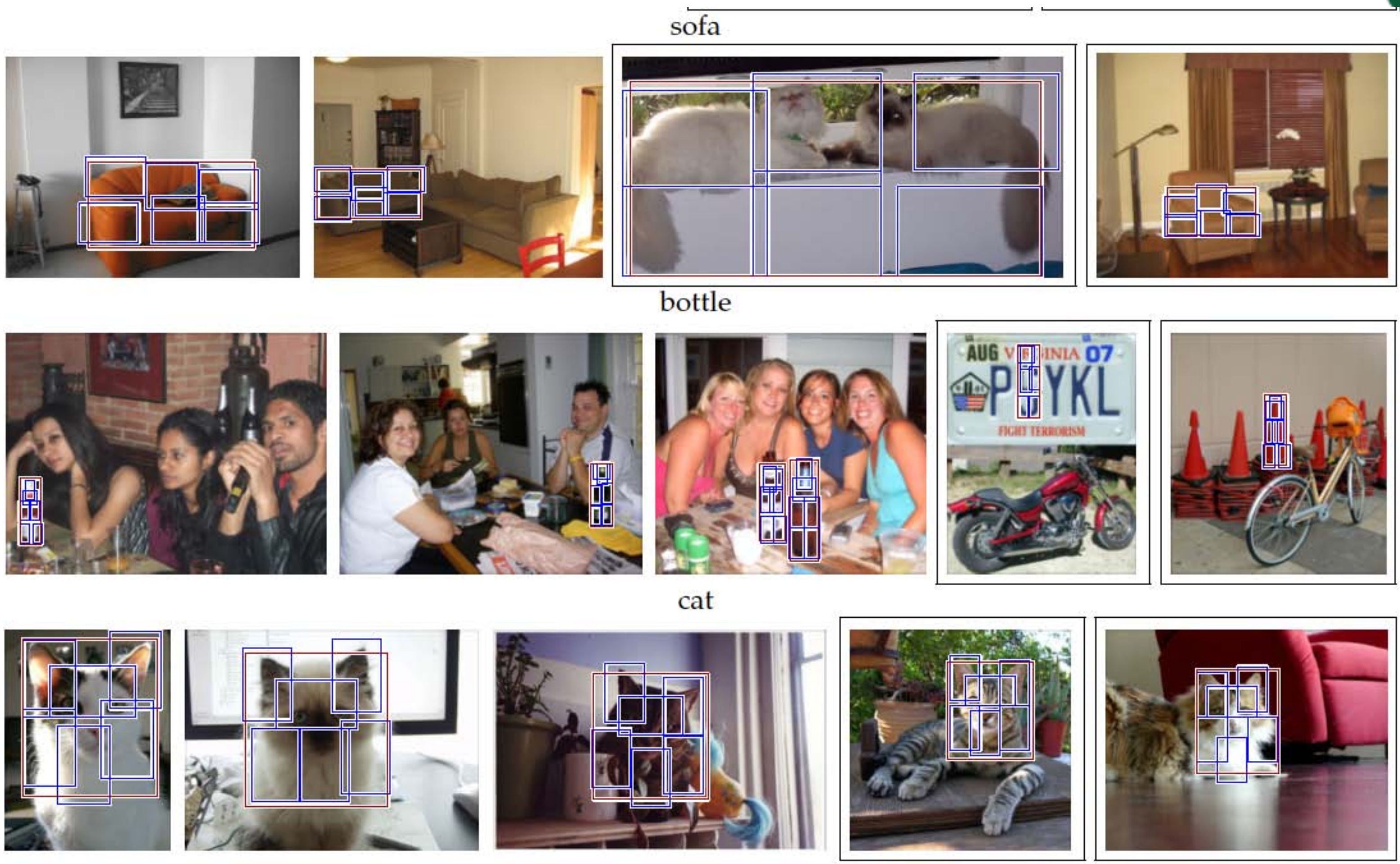


Vondrick et al 2013

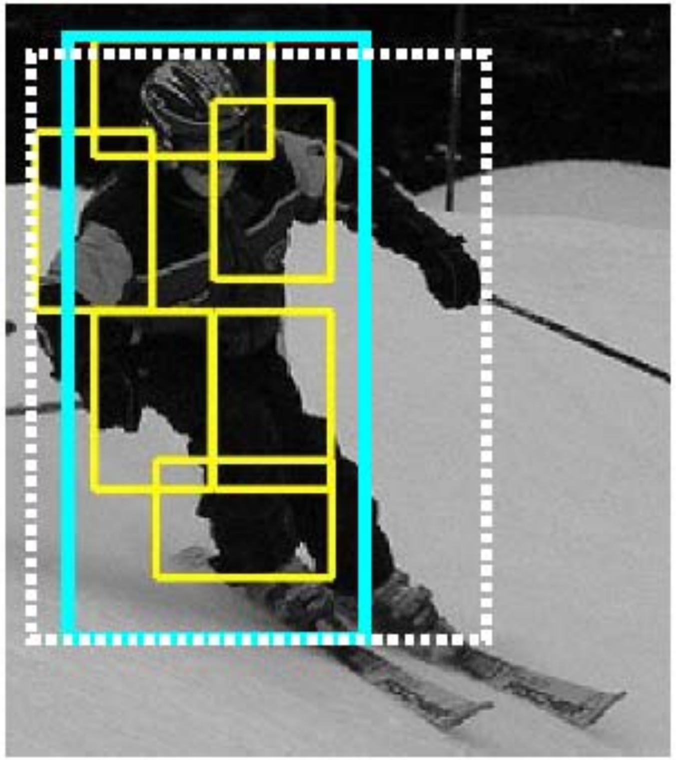
DPM results



DPM results

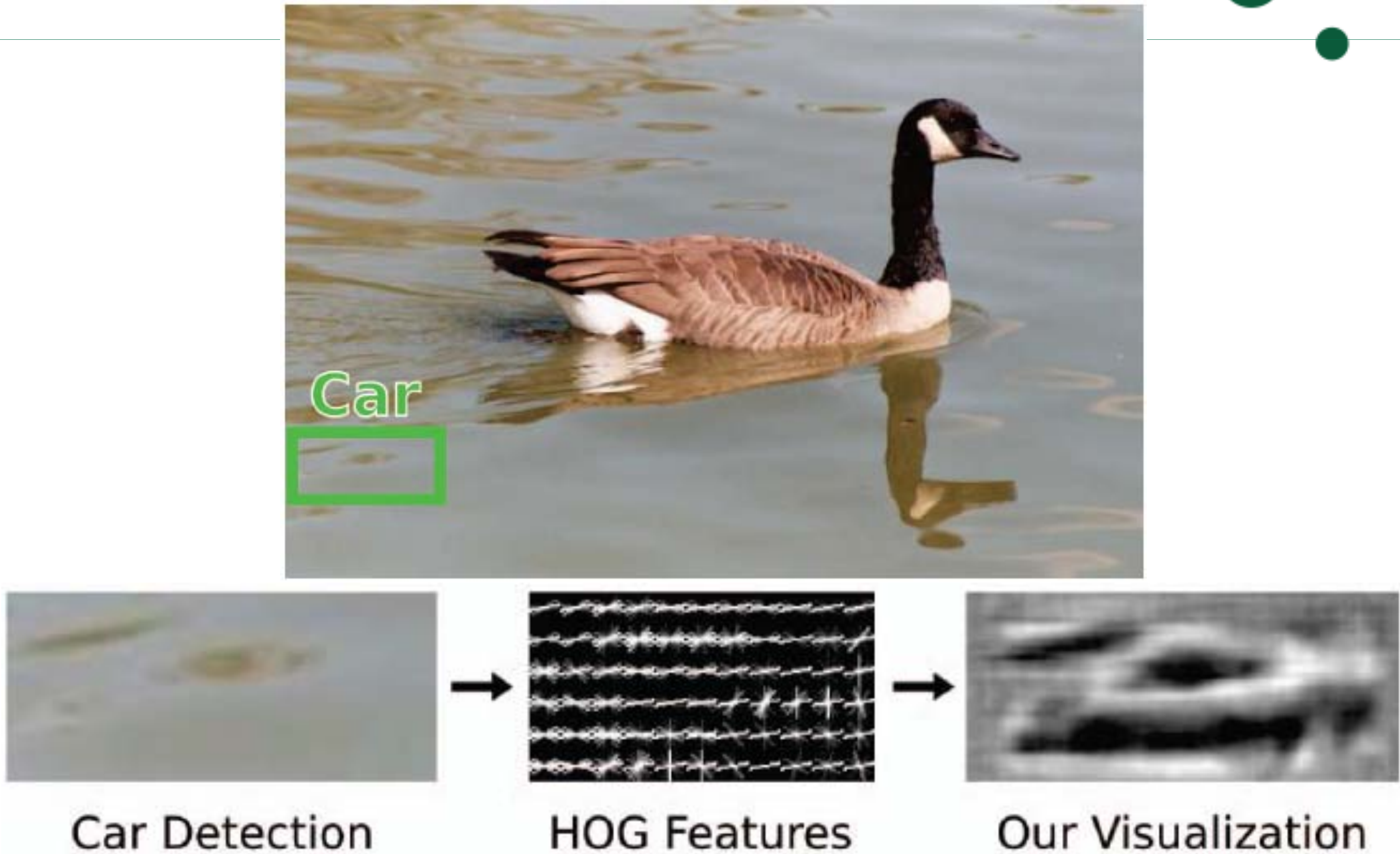


Hard negative examples for person detector



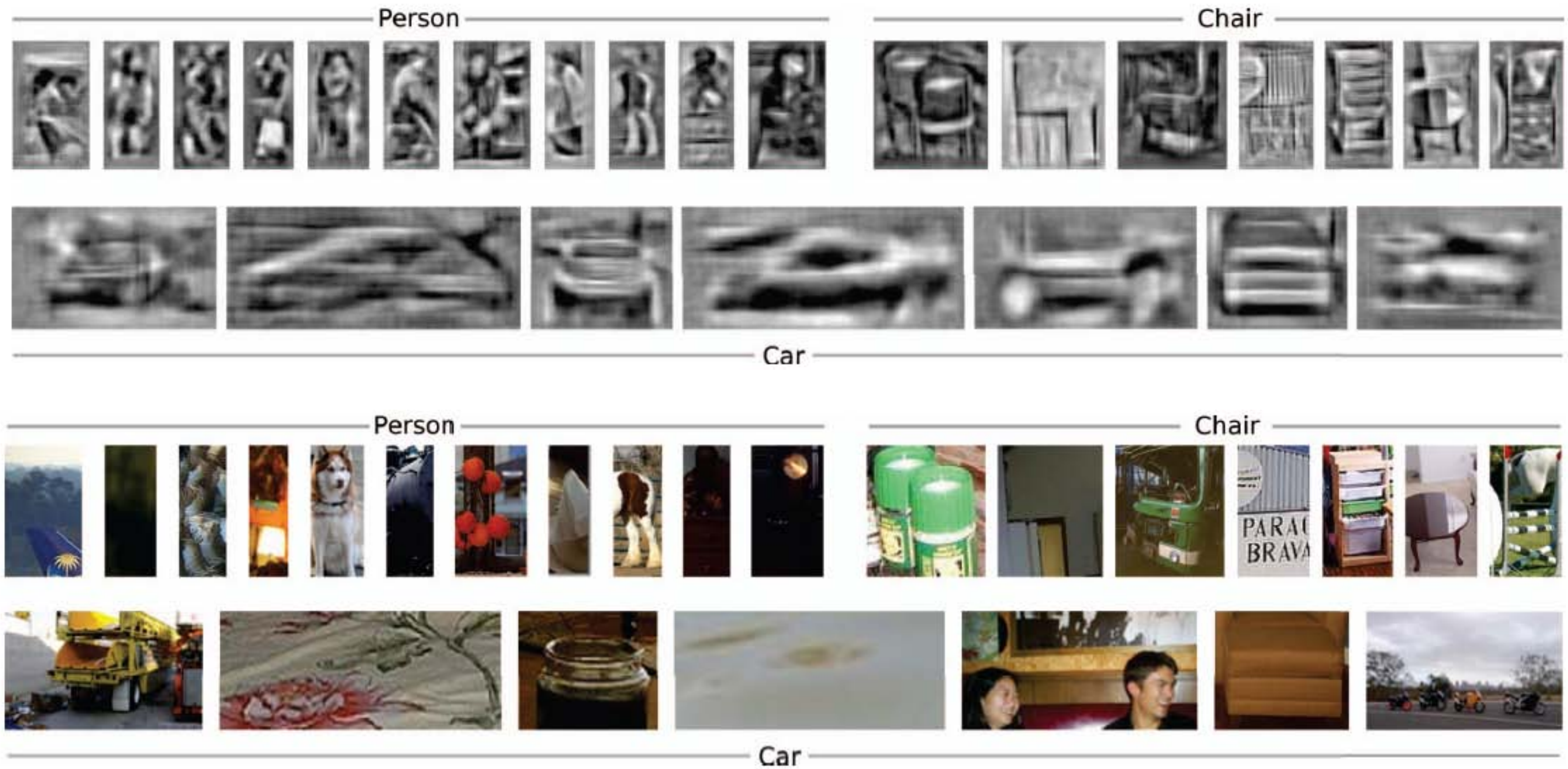
- Felzenszwalb et al propose a learning strategy to improve performance of this type of examples.

Visualization of the cause of failure



Vondrick et al 2013

Do you see the same objects as DPM?



Vondrick et al 2013

Other approaches in object recognition / detection

- Bag of visual words together with interest points
- Shape features (instead of only texture / local structure)
- Object taxonomies (e.g. using language resources such as WordNet – a database of cognitve synonyms)
- Domain specific knowledge (e.g. taxonomy of flower or bird species)

Summary

- For object detection and recognition we can use:
  - Interest point based approaches (can be extended to bag of visual words)
  - Dense feature template approaches
  - Parts models
- Dense feature template approaches:
  - Histogram of Oriented Gradients (HoG)
  - Deformable parts models (DPM)
- HoG captures first order structure locally and encodes higher order structure through cell grid. Maybe we can do better?

Literature

Reading material:

- Everingham et al IJCV 2010 (State of the art)
- Dalal & Triggs (CVPR'05) (HoG details)
- Felzenszwalb et al CVPR 2008 (DPM details)

Or

- Felzenszwalb et al IEEE T-PAMI 2010 (DPM details)

Additional material:

- Vondrick et al (ICCV 2013) (Visualizing HoG and DPM)