Perception (Computer Vision) I

Basic problems

- Detection
- Segmentation
- Labeling
- Pose estimation

1. General formulation

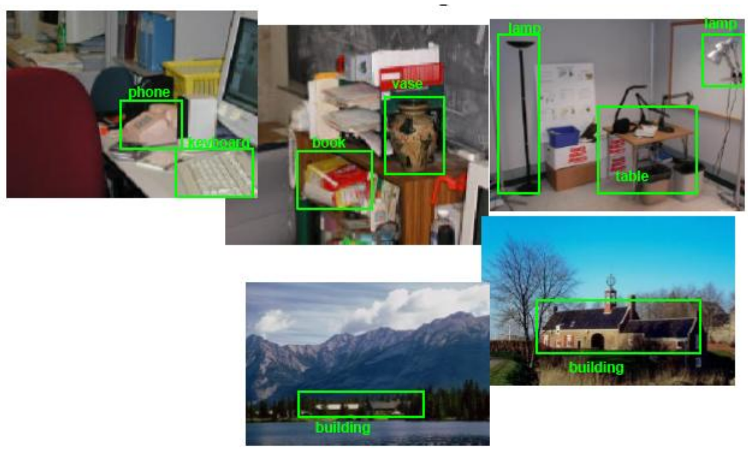
1b. Popular image representations

2. Connection with earlier discussion on factor graphs

3. Other approaches

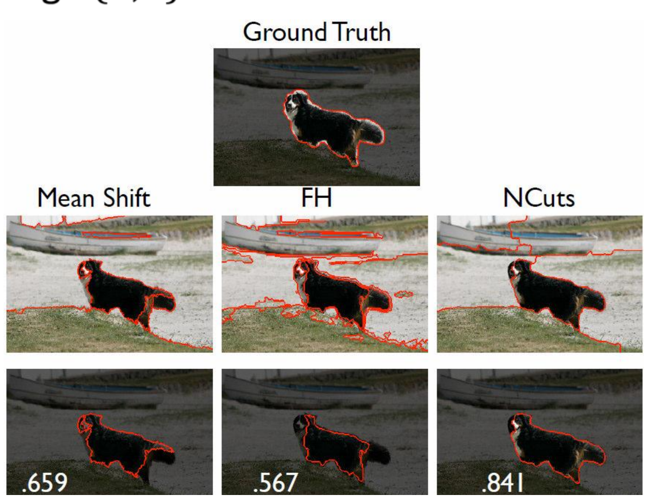
Basic problem

- Detection
- $f: \mathcal{X} \rightarrow \mathcal{Y}$ $\mathcal{Y}$ = all possible positions (and scales) of object



Basic problem

- Foreground/background segmentation
- $f: \mathcal{X} \rightarrow \mathcal{Y}$ $\mathcal{Y}$ = all possible 0/1 labelings of image $\{0,1\}^n$



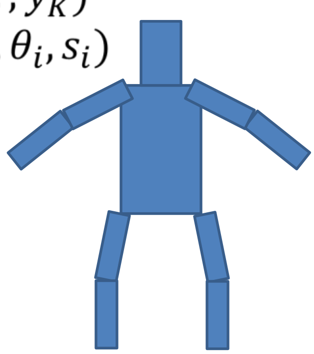
Basic problem

- Labeling
- $f: \mathcal{X} \rightarrow \mathcal{Y}$ $\mathcal{Y}$ = all possible labelings of image $\{1, \ldots, L\}^n$

Basic problem

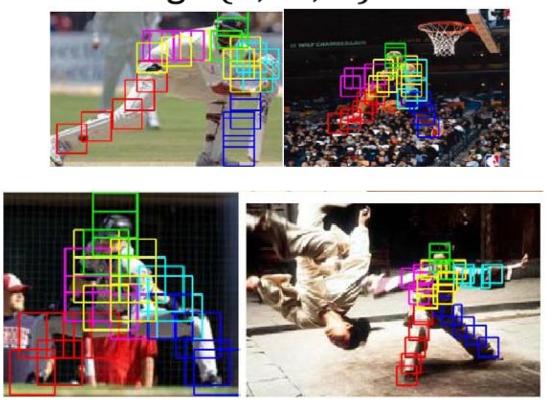
- Pose estimation
- $f: \mathcal{X} \rightarrow \mathcal{Y}$ $\mathcal{Y}$ = all possible poses (u, v, θ, s) of image $\{1, \ldots, P\}^K$

$$y = (y_1, \ldots, y_K)$$
$$y_i = (u_i, v_i, \theta_i, s_i)$$



Basic problem

- Pose estimation
- $f: \mathcal{X} \rightarrow \mathcal{Y}$ $\mathcal{Y}$ = all possible poses (u, v, θ, s) of image $\{1, \dots, P\}^K$



Yang and Ramanan. Articulated pose estimation with flexible mixtures-of-parts. CVPR. 2011.
V. Ferrari, M. Marin-Jimenez, A. Zisserman: "Progressive Search Space Reduction for Human Pose Estimation", CVPR 2008

Basic problems

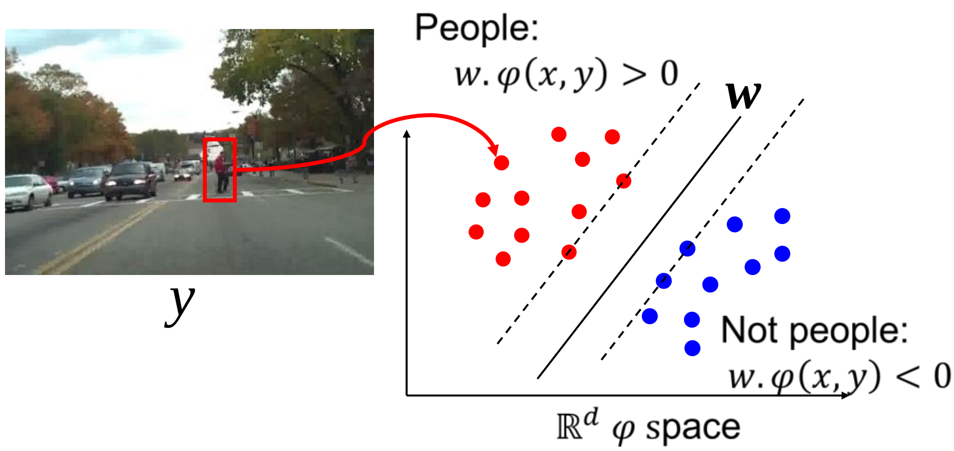
- Detection
- Segmentation
- Labeling
- Pose estimation
- Possible formalization (structured prediction):

$$f: \mathcal{X} \rightarrow \mathcal{Y}$$
$$g: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$$

$$f(x) = y^* = argmax_{y \in \mathcal{Y}}\ g(x, y)$$

Example: Linear classifier for detection

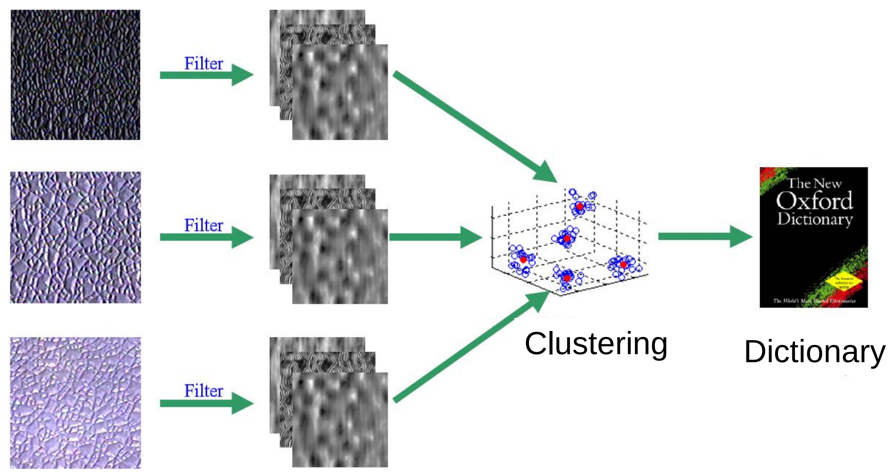
$$g(x, y) = w.\varphi(x, y)$$



- Slight digression: A couple of popular features $\varphi(x, y)$ for vision problems

- Bags of words
- Histograms of gradients

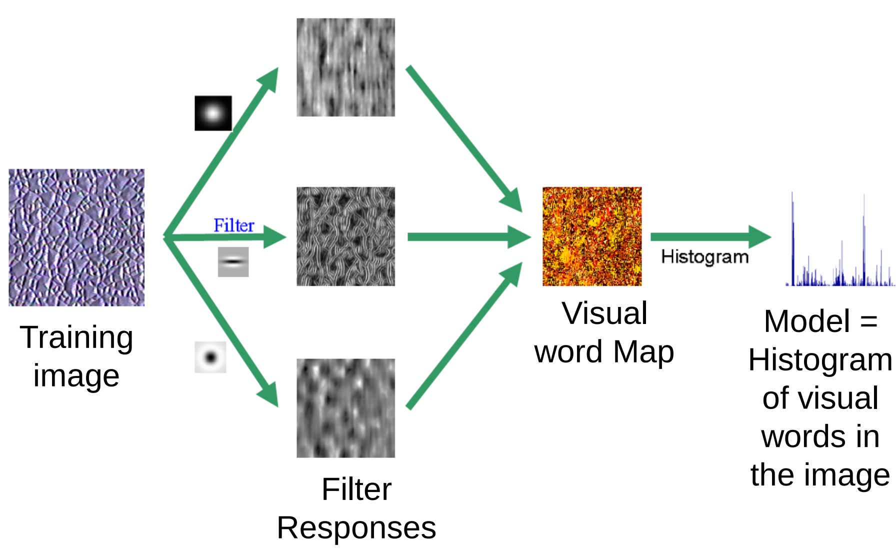
Example feature $\varphi(x, y)$: Bag of words (BoW)



Multiple training images

Local features at each pixel: E.g., filter responses over a bank of filters

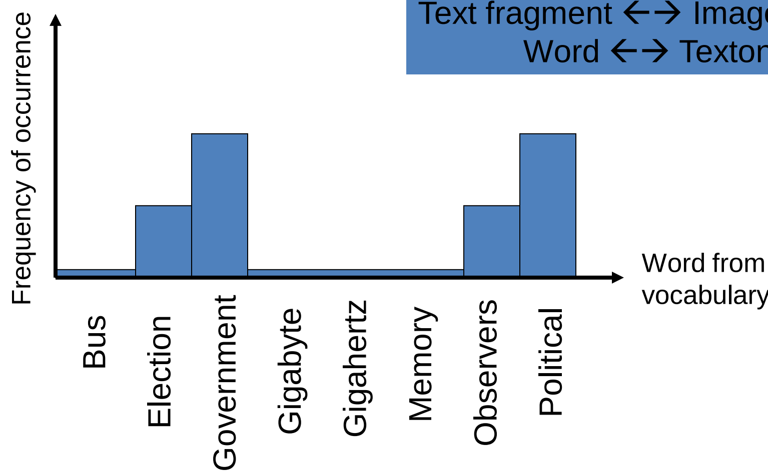
Modeling Texton Distributions



Analogy with Text Analysis

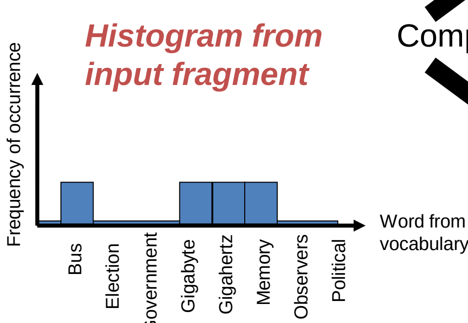
Political observers say that the government of Zorgia does not control the political situation. The government will not hold elections …

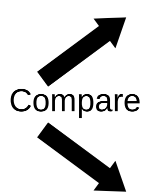
Analogy:
Text fragment ←→ Image region
Word ←→ Texton

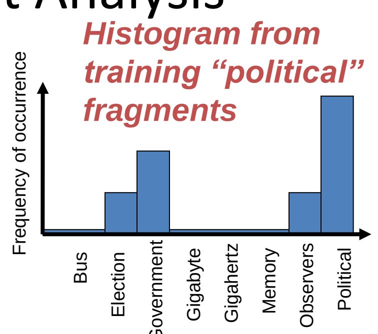


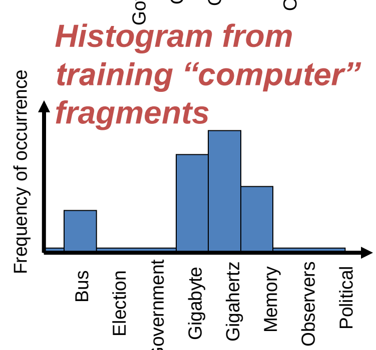
Analogy with Text Analysis

The ZH-20 unit is a 200Gigahertz processor with 2Gigabyte memory. Its strength is its bus and high-speed memory……

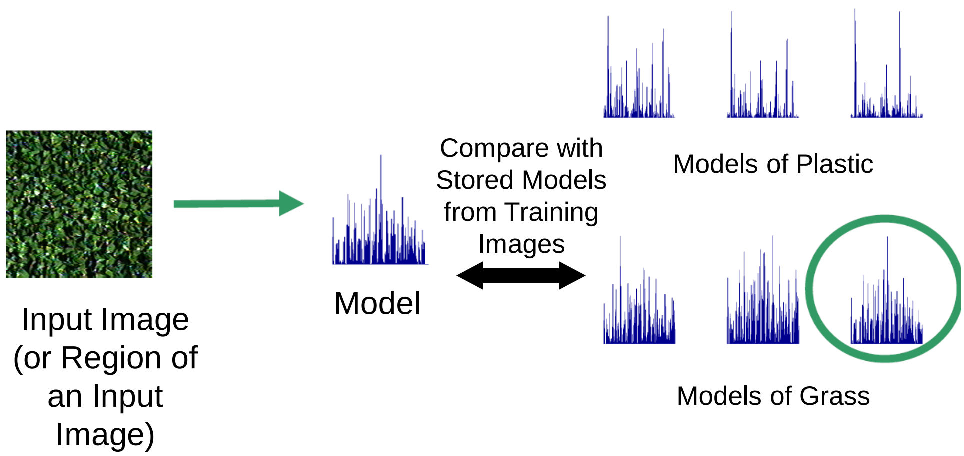




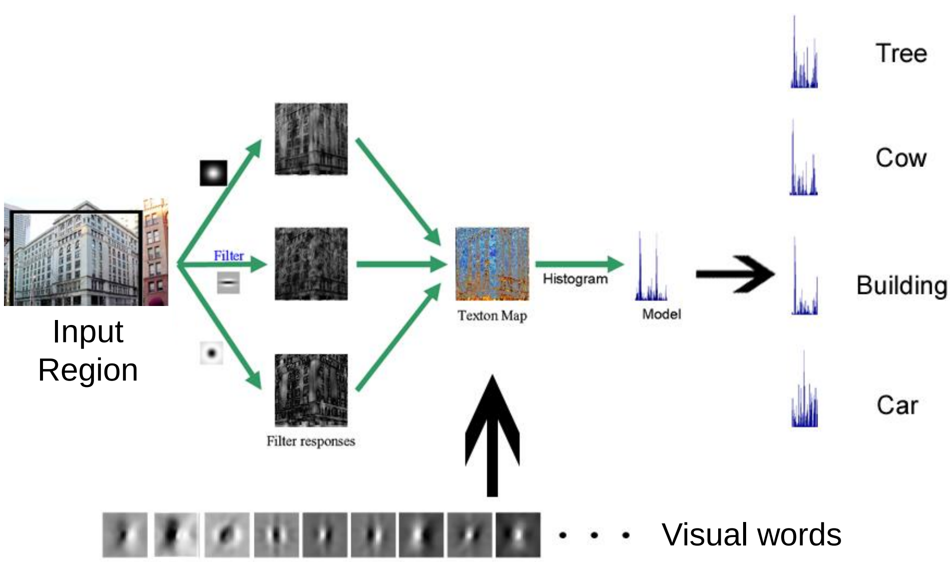




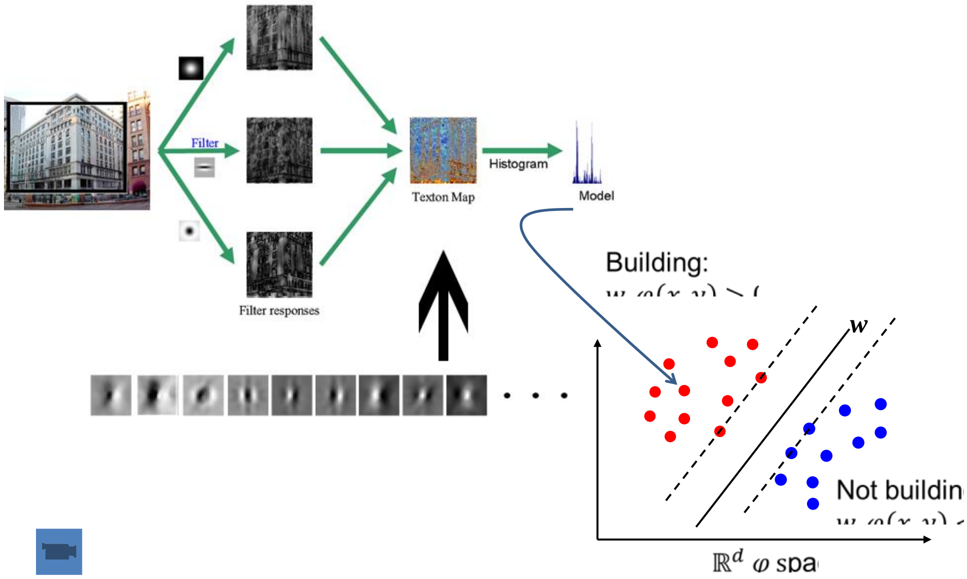
Classification: Nearest-neighbor



Example Classification (NN)

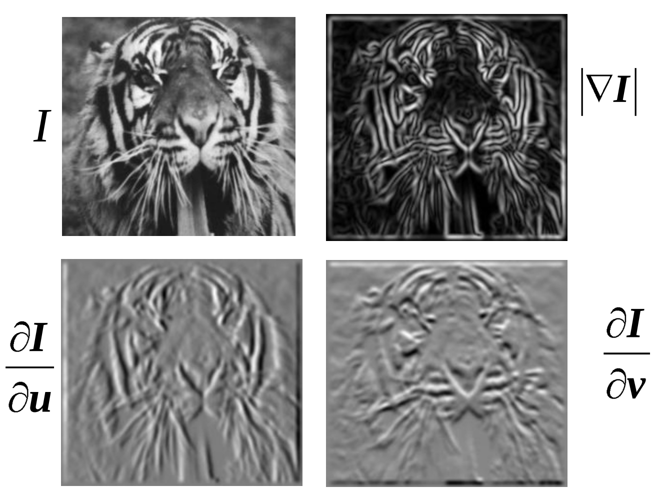


Classifier on BoW

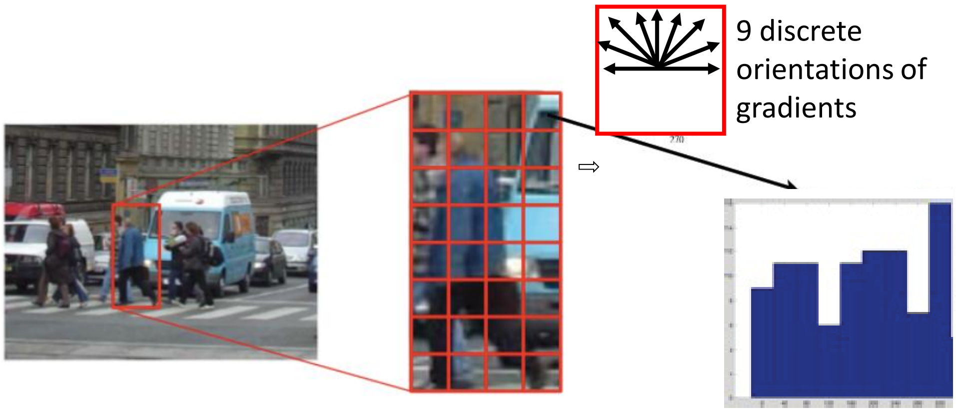


Example feature $\varphi(x, y)$: Histogram of gradients

- Gradients in images capture the local variation of intensity = vector of derivatives of image in coordinates u and v
- Captures local “shape” and local “texture”



Histogram of Gradients: HoG

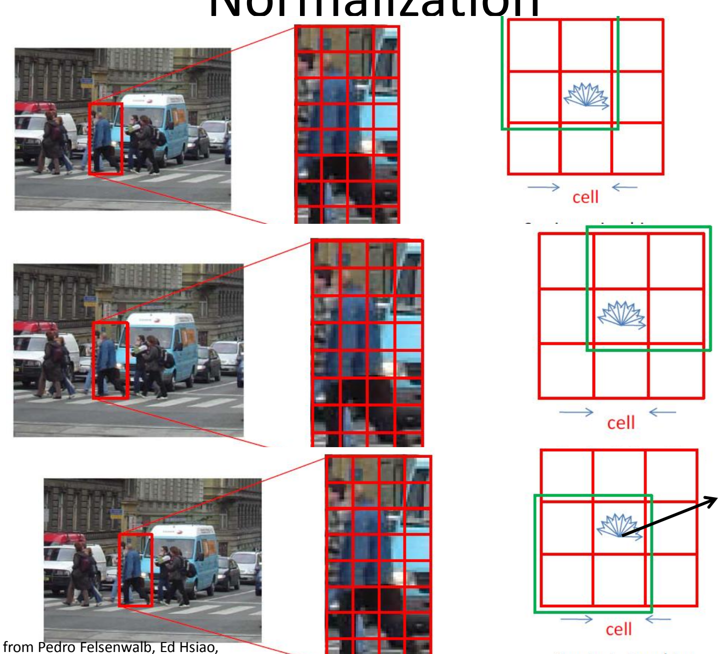


8x4 cells

Histogram of gradient magnitude over 9 discretized orientations

Slide adapted from Pedro Felsenwalb, Ed Hsiao, Varun Ramakrishnan

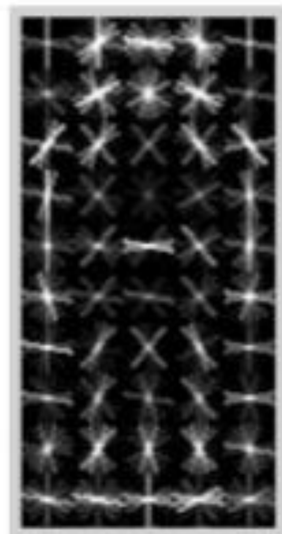
Normalization



Slide adapted from Pedro Felsenwalb, Ed Hsiao, Varun Ramakrishnan

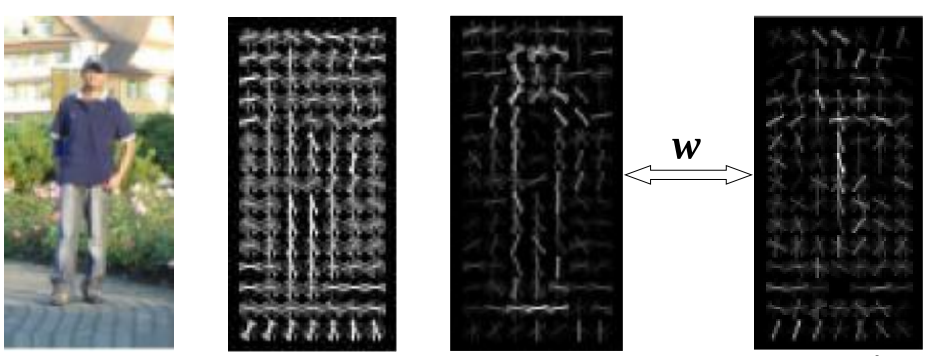
Feature vector

- $\varphi(x, y) \in \mathbb{R}^d \; d \approx 1000$



Classification

$\boldsymbol{w}.\varphi > 0$



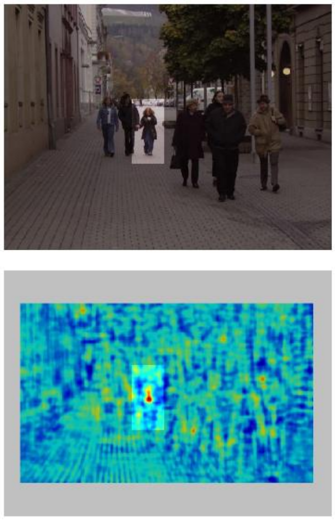
Example detection window

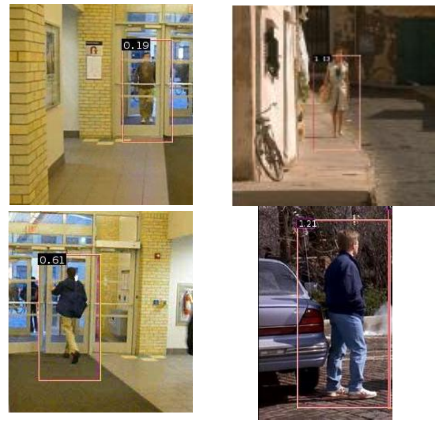
HOG descriptor

Positive components: Peole class

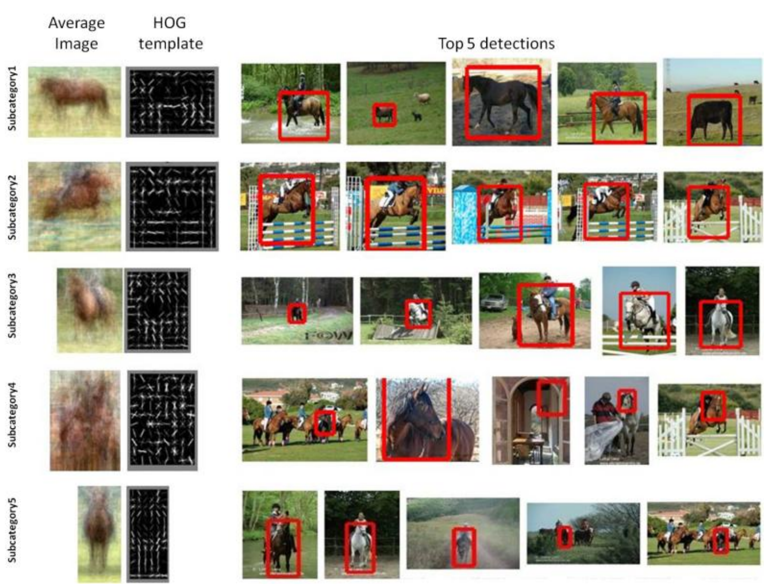
Negative components: Non-people class

N. Dalal and B. Triggs . *Histograms of Oriented Gradients for Human Detection.* CVPR, 2005



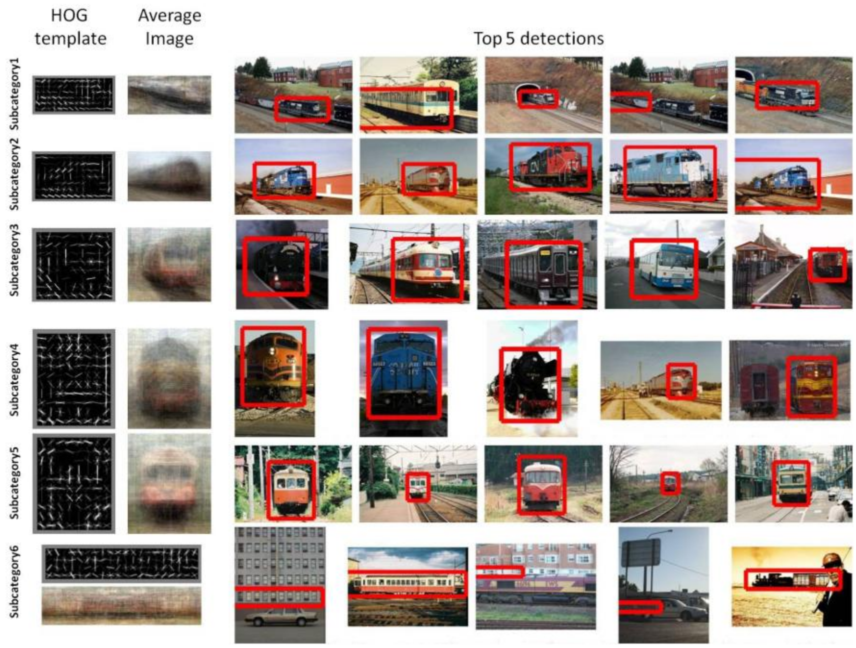


Other objects



Courtesy Santosh Divvalla 2012

Other objects

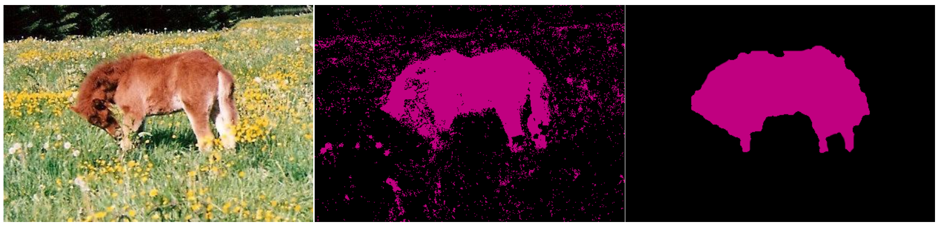


Example: Segmentation

$$g(x, y) = \sum_{i=1}^{n} g_i(x, y_i) + \sum_{i,j \in \mathcal{N}(i)} g_{i,j}(y_i, y_j)$$
$$f(x) = y^* = argmax_{y \in \{0,1\}^n}\, g(x, y)$$

$g_{i,j}(y_i, y_j)$ large if $y_i = y_j$
$g_i(x, y_i)$ large if data x at i agrees with model y_i
Example: Linear model $g_i(x, y_i) = w_{y_i}^{\cdot} . \varphi_i(x, y_i)$

Example: Segmentation



x $\qquad$ $g_i(x, y_i)$ $\qquad$ $+g_{i,j}(y_i, y_j)$

Example from Christoph Lampert

Example: Labeling

$$g(x, y) = \sum_{i=1}^{n} g_i(x, y_i) + \sum_{i,j \in \mathcal{N}(i)} g_{i,j}(y_i, y_j, x)$$

$$f(x) = y^* = argmax_{y \in \{1,\dots,L\}^n}\, g(x, y)$$

$g_{i,j}(y_i, y_j, x)$ large if $y_i = y_j$ when data agrees at *i* and *j*

$$g_{ij}(y_i, y_j, x) = w_{ij} . \varphi_{ij}(y_i, y_j, x)$$

$g_i(x, y_i)$ large if data x at *i* agrees with model y_i

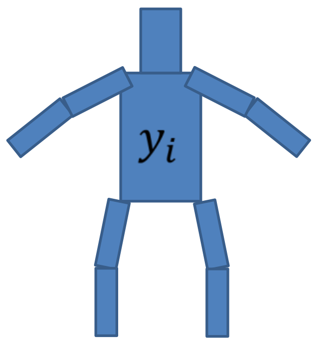
Examples: Linear, Softmax $g_i(x, y_i) = \frac{e^{w_{y_i} \cdot \varphi_i(x, y_i)}}{\sum_j e^{w_{y_j} \cdot \varphi_j(x, y_j)}}$

Example: Pose estimation

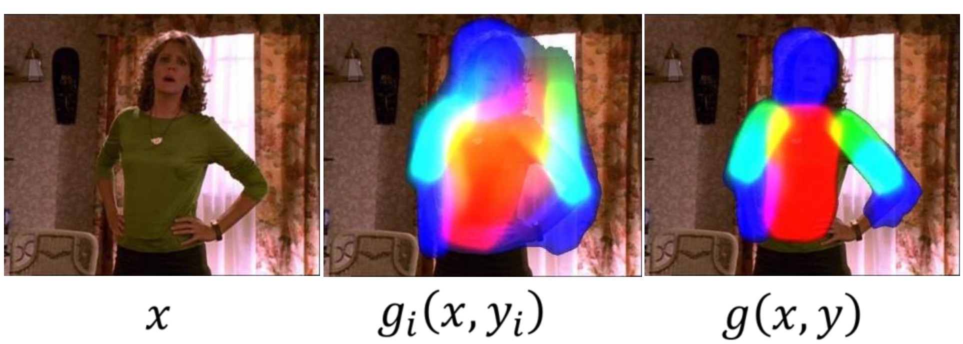
$$g(x,y) = \sum_{i=1}^{n} g_i(x, y_i) + \sum_{i,j \text{ linked}} g_{i,j}(y_i, y_j)$$

$g_i(x, y_i) = w_i . \varphi_i(x, y_i)$ φ_i = HoG at y_i

$g_{i,j}(y_i, y_j)$ = likelihood of relative positions y_i and y_j



Example: Pose estimation



V. Ferrari, M. Marin-Jimenez, A. Zisserman: "Progressive Search Space Reduction for Human Pose Estimation", CVPR 2008

Possible structured models: Factor graphs

$$g(x,y) = \sum_{i=1}^{n} g_i(x,y_i) + \sum_{i,j \text{ linked}} g_{i,j}(y_i,y_j)$$

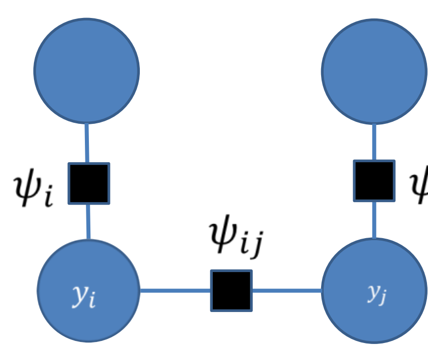
$$g(x,y)$$
$$= \log(\exp(\sum_{i=1}^{n} g_i(x,y_i) + \sum_{i,j \text{ linked}} g_{i,j}(y_i,y_j)))$$

$$\exp(g(x,y)) = \prod \psi_i(x,y_i) \prod_{i,j \text{ linked}} \psi_{ij}(y_i,y_j)$$

Factor graphs

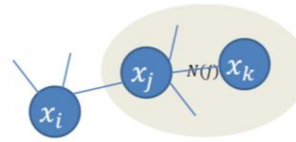
$$p(y|x) \propto \exp(g(x,y))$$

$$p(y|x) = \frac{1}{Z}\exp(g(x,y))$$
$$= \frac{1}{Z}\prod \psi_i(x,y_i) \prod_{i,j \text{ linked}} \psi_{ij}(y_i,y_j)$$



Conditional random field

Reminder of key results: Inference



- Exact algorithm on tree-structured graphs
  - Message passing
  - Max-product: compute y^*

$$\sum_{x_j} f_{ij}(x_i, x_j) \prod_{N(j)\setminus i} m_{kj}(x_j)$$

  - Sum-product: estimate marginals $p(y_i|x)$
- Approximate algorithms
  - Loopy BP
  - Sampling