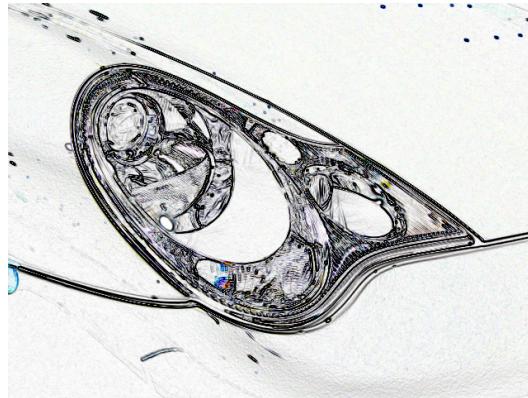


CSE578: Computer Vision

Spring 2016:

Object Detection and Recognition: HoG + SVM: Pedestrian Detection



Anoop M. Namboodiri

Center for Visual Information Technology

IIIT Hyderabad, INDIA

[Slides Generously Borrowed from Various Sources]



Challenges

- Wide variety of articulated poses
- Variable appearance/clothing
- Complex backgrounds
- Unconstrained illumination
- Occlusions
- Different Scales

Discriminative vs. Generative Models

- Generative:
 - + possibly interpretable
 - + models the object class/can draw samples
 - - model variability unimportant to classification task
 - - often hard to build good model with few parameters
- Discriminative:
 - + appealing when infeasible to model data itself
 - + currently often excel in practice
 - - often can not provide uncertainty in predictions
 - - non-interpretable

Global vs. Part-Based

- Global people detectors vs. part-based detectors
- Global approaches:
 - A single feature description for the complete person
- Part-Based Approaches:
 - Individual feature descriptors for body parts / local parts

Advantages and Disadvantages

- Part-Based
 - May be better able to deal with moving body parts
 - May be able to handle occlusion, overlaps
 - Requires more complex reasoning
- Global approaches
 - Typically simple, i.e. we train a discriminative classifier on top of the feature descriptions
 - Work well for small resolutions
 - Typically does detection via classification, i.e. uses a binary classifier

Gradient Histograms

- Have become extremely popular and successful in the vision community
- Avoid hard decisions compared to edge based features
- Examples:
 - SIFT (Scale-Invariant Image Transform)
 - GLOH (Gradient Location and Orientation Histogram)
 - HOG (Histogram of Oriented Gradients)

Computing Gradients

- One sided: $f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$
- Two sided: $f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x-h)}{2h}$

- Filter masks in x-direction

- One sided:

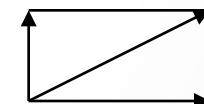
-1	1
----	---

- Two sided:

-1	0	1
----	---	---

- Gradient:

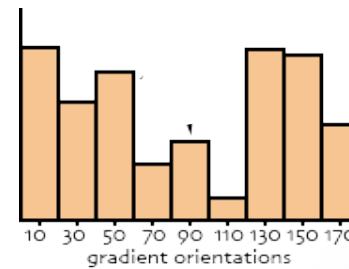
- Magnitude: $s = \sqrt{s_x^2 + s_y^2}$



- Orientation: $\theta = \arctan\left(\frac{s_y}{s_x}\right)$

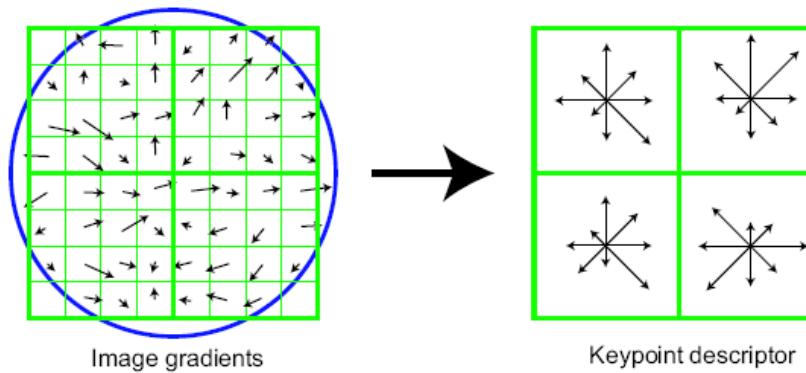
Histograms

- Gradient histograms measure the orientations and strengths of image gradients within an image region



Example: SIFT descriptor

- The most popular gradient-based descriptor
- Typically used in combination with an interest point detector



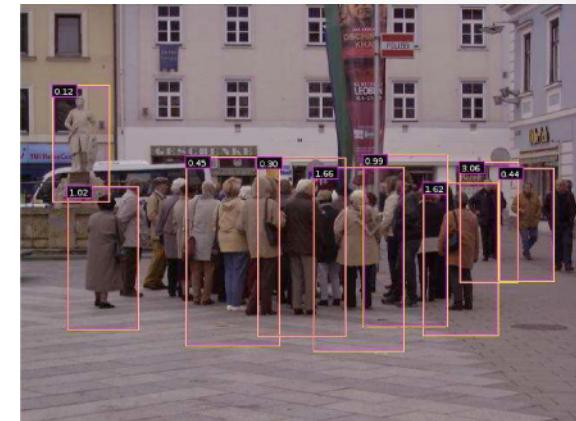
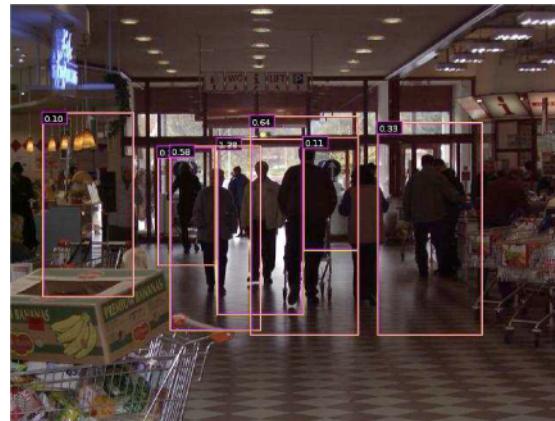
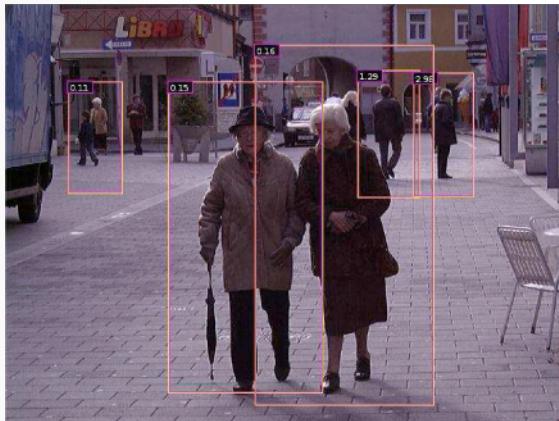
- Region rescaled to a grid of 16x16 pixels
- 4x4 regions = 16 histograms (concatenated)
- Histograms: 8 orientation bins, gradients weighted by gradient magnitude
- Final descriptor has 128 dimensions and is normalized to compensate for illumination differences

Histograms of Oriented Gradients

- Gradient-based feature descriptor developed for people detection
 - Authors: Dalal & Triggs (INRIA Grenoble, F)
- Global descriptor for the complete body
- Very high-dimensional
 - Typically ~4000 dimensions

HOG

Very promising results on challenging data sets

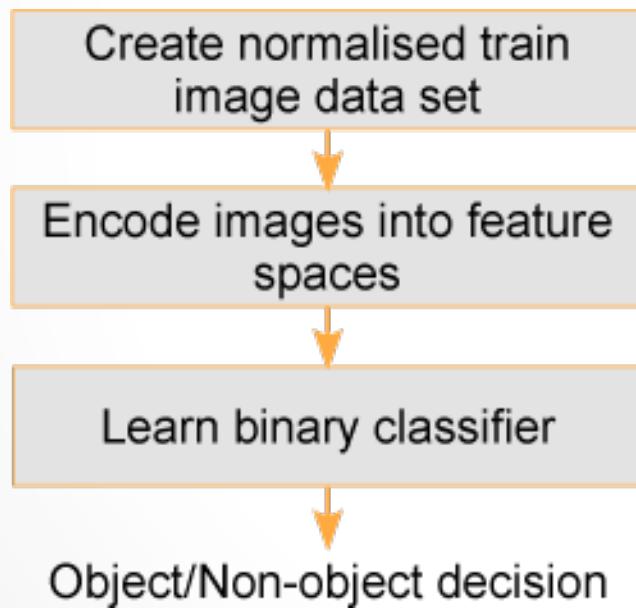


Phases

1. Learning Phase
2. Detection Phase

Detector: Learning Phase

1. Learning



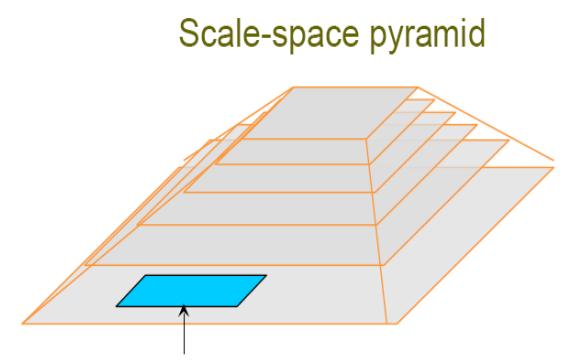
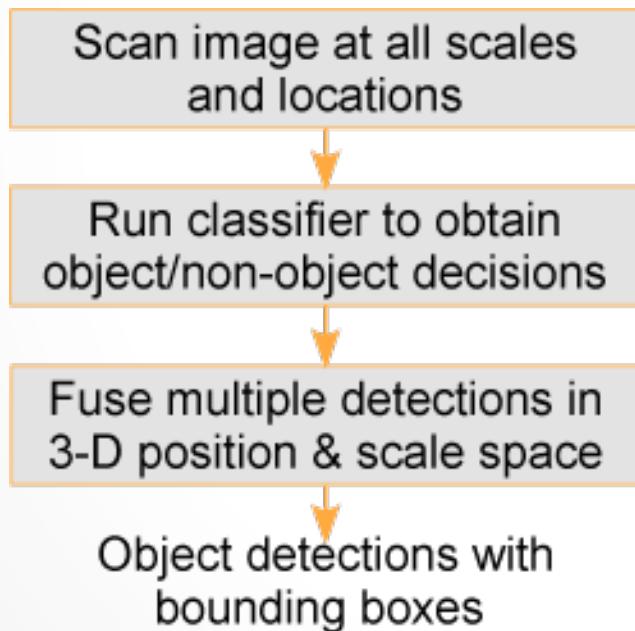
Set of cropped images containing pedestrians in normal environment

Global descriptor rather than local features

Using linear SVM

Detector: Detection Phase

2. Detection

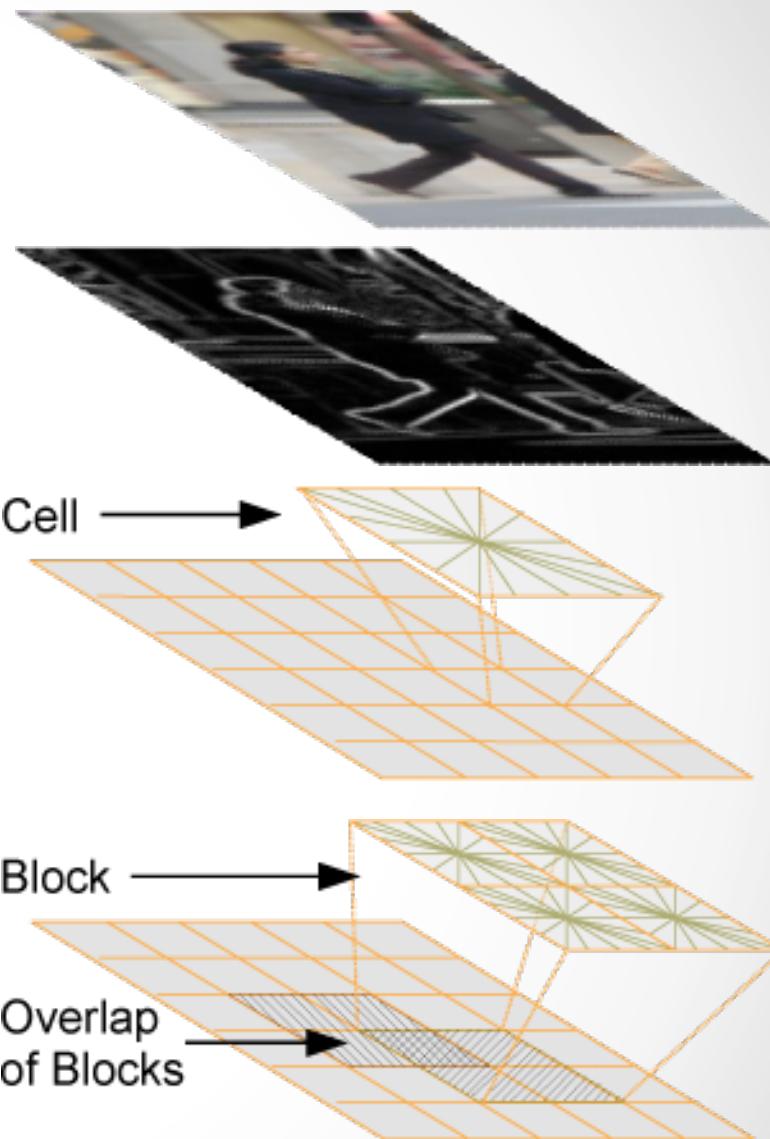


Sliding window over each scale

Simple SVM prediction

Descriptor

1. Compute gradients on an image region of 64×128 pixels
2. Compute histograms on ‘cells’ of typically 8×8 pixels (i.e. 8×16 cells)
3. Normalize histograms within overlapping blocks of cells (typically 2×2 cells, i.e. 7×15 blocks)
4. Concatenate histograms



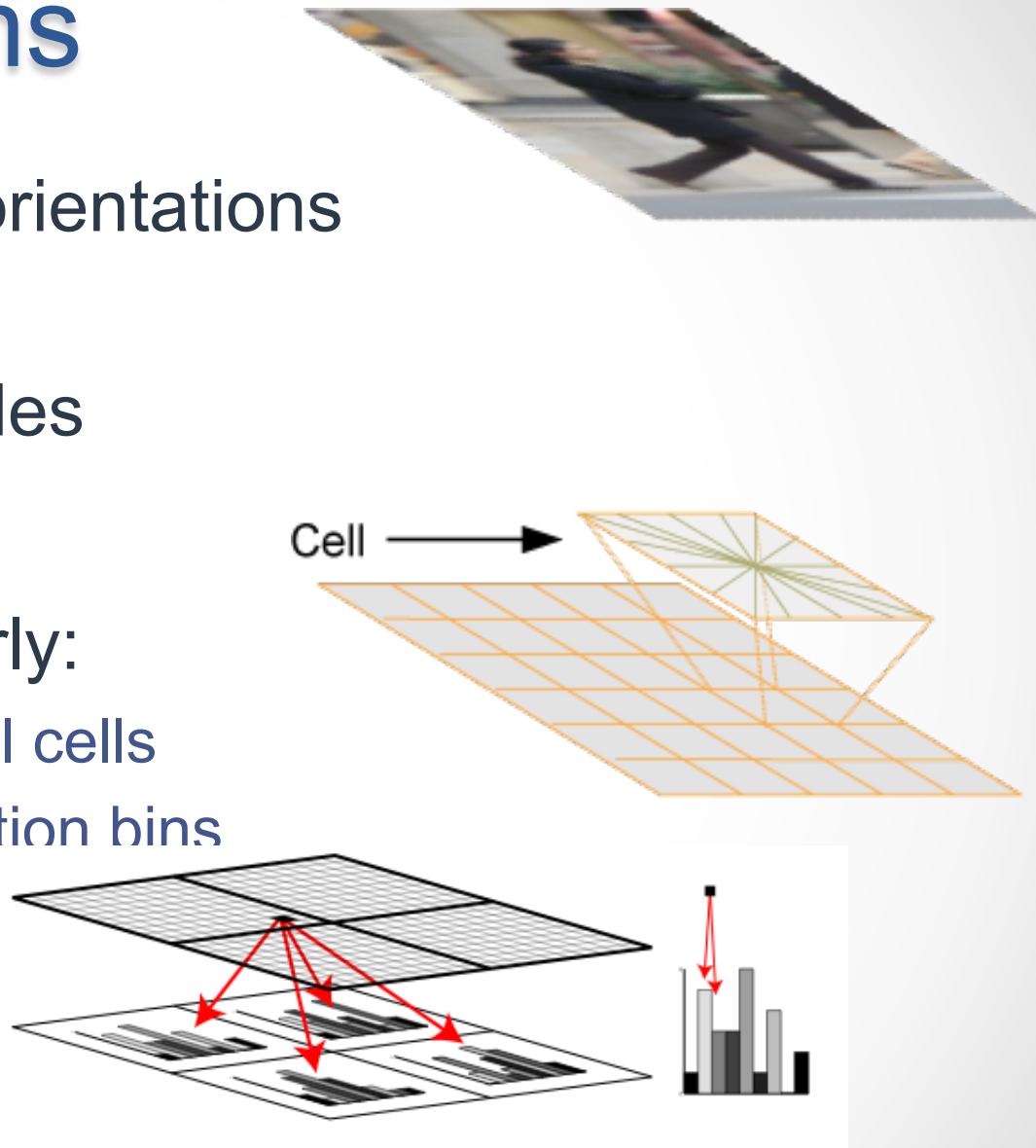
Gradients

- Convolution with $[-1 \ 0 \ 1]$ filters
- No smoothing
- Compute gradient magnitude+direction
- Per pixel: color channel with greatest magnitude -> final gradient



Cell histograms

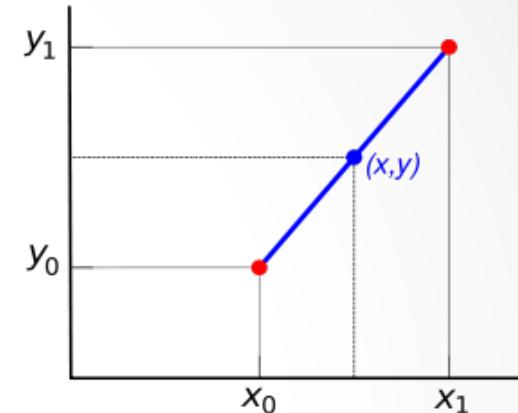
- 9 bins for gradient orientations (0-180 degrees)
- Filled with magnitudes
- Interpolated trilinearly:
 - Bilinearly into spatial cells
 - Linearly into orientation bins



Linear and Bilinear Interpol. for Subsampling

Linear:

$$y = y_0 + (x - x_0) \frac{y_1 - y_0}{x_1 - x_0}$$

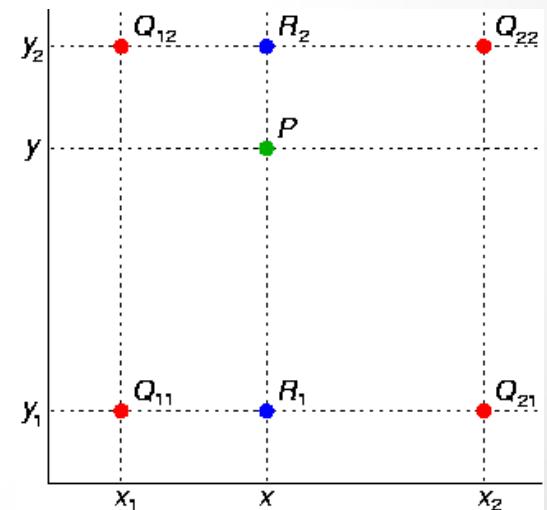


Bilinear:

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \quad \text{where } R_1 = (x, y_1),$$

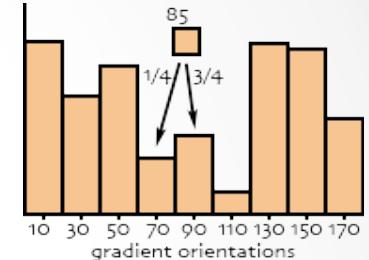
$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \quad \text{where } R_2 = (x, y_2).$$

$$f(P) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2).$$

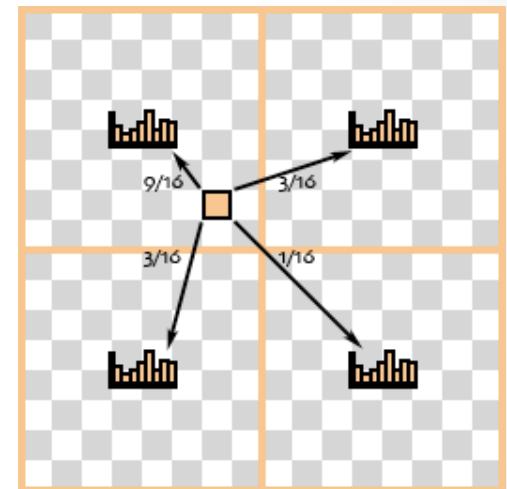


Histogram Interpolation Example

- $\theta=85$ degrees
- Distance to bin centers
 - Bin 70 -> 15 degrees
 - Bin 90 -> 5 degrees
- Ratios: $5/20=1/4$, $15/20=3/4$

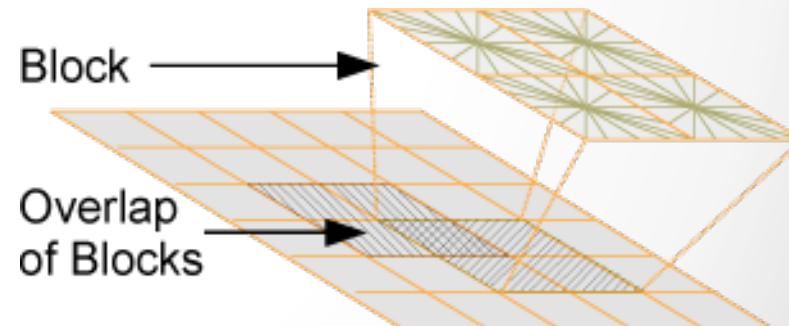
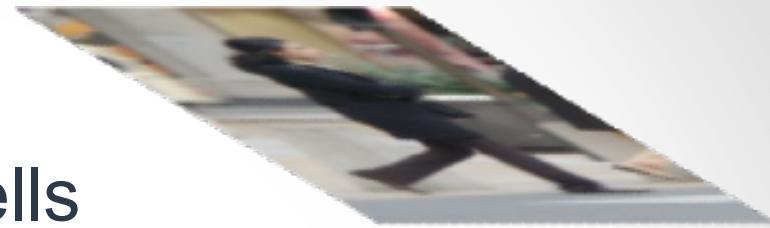


- Distance to bin centers
 - Left: 2, Right: 6
 - Top: 2, Bottom: 6
- Ratio Left-Right: $6/8$, $2/8$
- Ratio Top-Bottom: $6/8$, $2/8$
- Ratios:
 - $6/8 * 6/8 = 36/64 = 9/16$
 - $6/8 * 2/8 = 12/64 = 3/16$
 - $2/8 * 6/8 = 12/64 = 3/16$
 - $2/8 * 2/8 = 4/64 = 1/16$



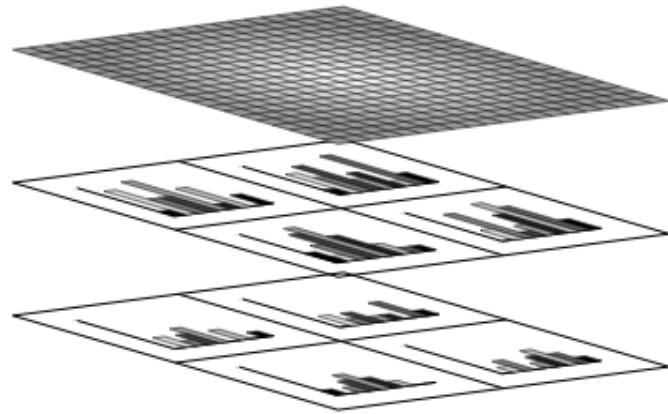
Blocks

- Overlapping blocks of 2x2 cells
- Cell histograms are concatenated and then normalized
 - Note that each cell several occurrences with different normalization in final descriptor
- Normalization
 - Different norms possible (L2, L2hys etc.)
 - We add a normalization epsilon to avoid division by zero



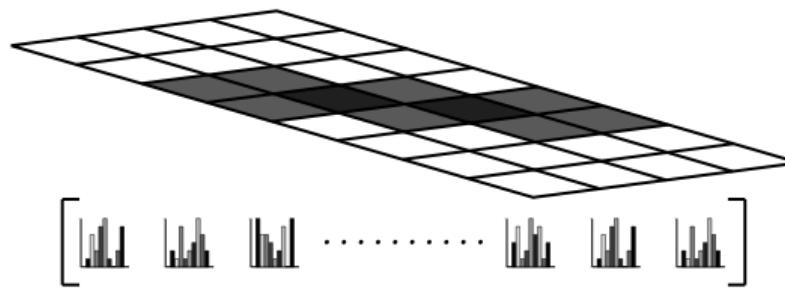
Blocks

- Gradient magnitudes are weighted according to a Gaussian spatial window
- Distant gradients contribute less to the histogram

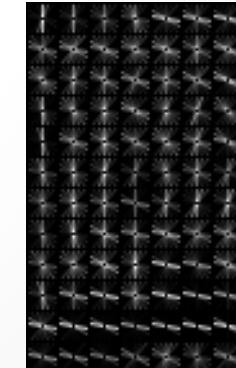
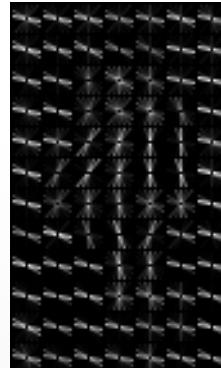
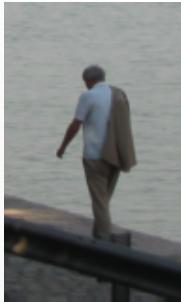


Final Descriptor

- Concatenation of Blocks



- Visualization:



Engineering

- Developing a feature descriptor requires a lot of engineering
 - Testing of parameters (e.g. size of cells, blocks, number of cells in a block, size of overlap)
 - Normalization schemes (e.g. L1, L2-Norms etc., gamma correction, pixel intensity normalization)
- An extensive evaluation of different choices was performed, when the descriptor was proposed
- It is not only the idea, but also the engineering effort

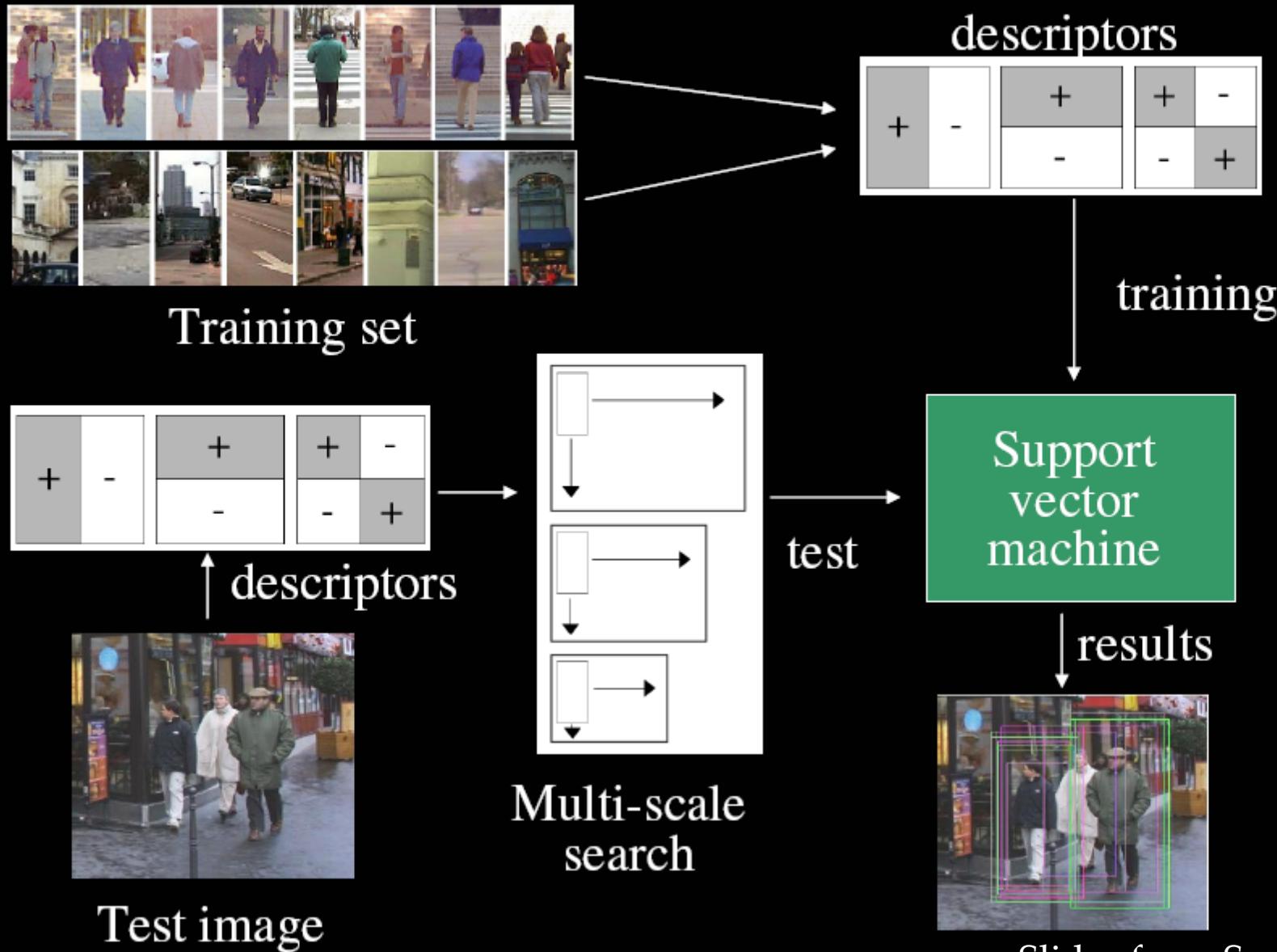
Training Set

- More than 2000 positive & 2000 negative training images (96x160px)
- Carefully aligned and resized
- Wide variety of backgrounds



Support Vector Machine Detector

(Papagergiu & Poggio, 1998)



Dynamic Pedestrian Detection

Viola, Jones and Snow, ICCV 2003



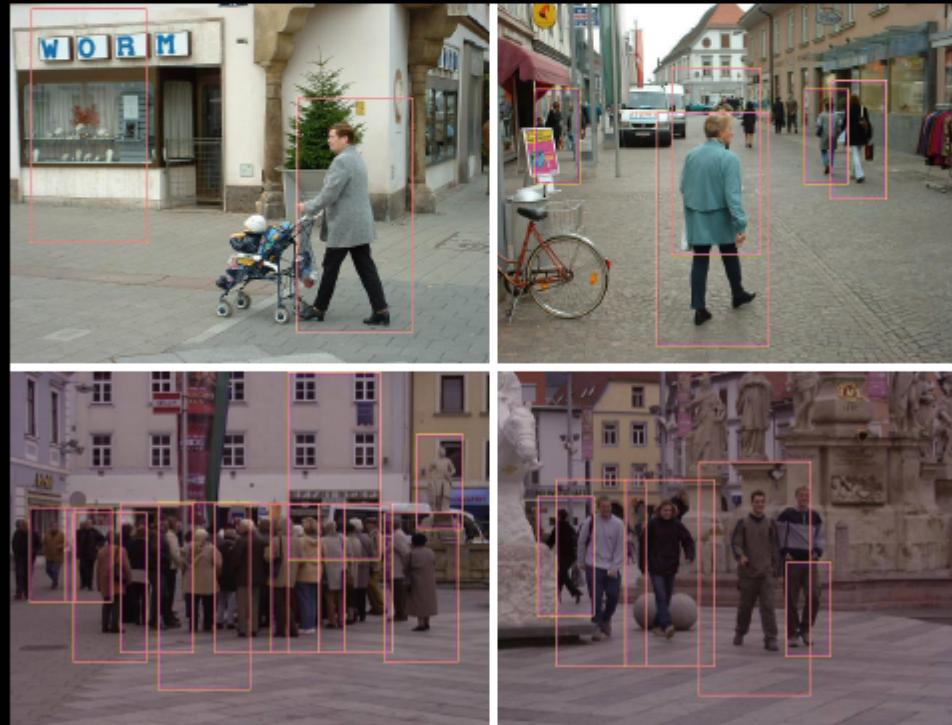
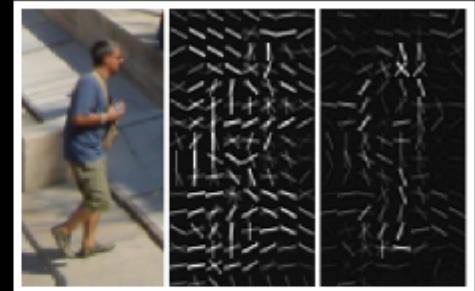
- Train using AdaBoost, about 45,000 possible features
- Efficient and reliable for distant detections (20x15), 4fps

2d Global Detector

Dalal and Triggs, CVPR 2005

- 3-D Histogram of Oriented Gradients (HOG) as descriptors
- Linear SVM for runtime efficiency
- Tolerates different poses, clothing, lighting and background
- Currently works for fully visible upright persons

Importance weight responses



Feature Sets

- Haar wavelets + SVM:
 - Papageorgiou & Poggio (2000)
 - Mohan et al (2001)
 - DePoortere et al (2002)
- Rectangular differential features + adaBoost:
 - Viola & Jones(2001)
- Parts based binary orientation position histogram + adaBoost:
 - Mikolajczk et al (2004)
- Edge templates + nearest neighbor:
 - Gavrila & Philomen (1999)
- Dynamic programming:
 - Felzenszwalb & Huttenlocher (2000),
 - Loffe & Forsyth (1999)
- Orientation histograms:
 - C.F. Freeman et al (1996)
 - Lowe(1999)
- Shape contexts:
 - Belongie et al (2002)
- PCA-SIFT:
 - Ke and Sukthankar (2004)

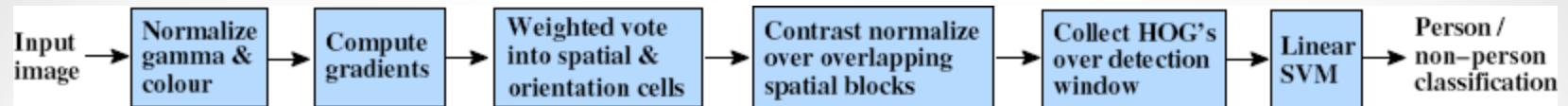


Method Summary





- Tested with
 - RGB
 - LAB
 - Grayscale
- Gamma Normalization and Compression
 - Square root
 - Log



-1	0	1
----	---	---

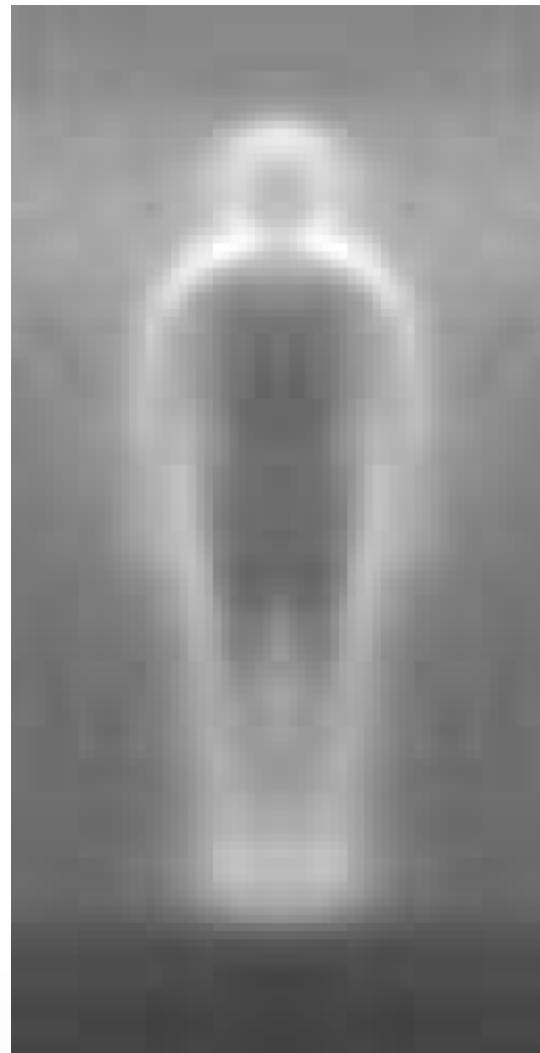
centered

-1	1
----	---

uncentered

1	-8	0	8	-1
---	----	---	---	----

cubic-corrected

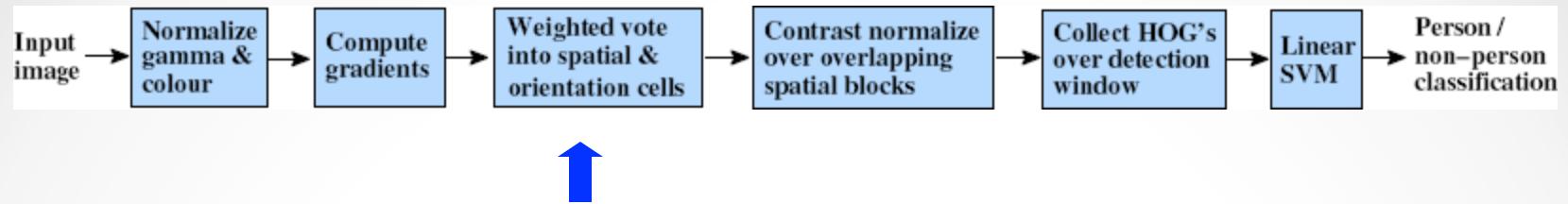


0	1
-1	0

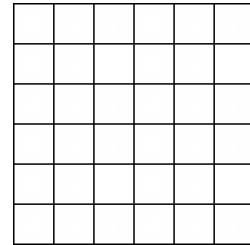
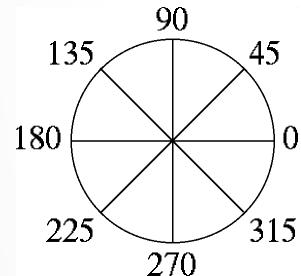
diagonal

-1	0	1
-2	0	2
-1	0	1

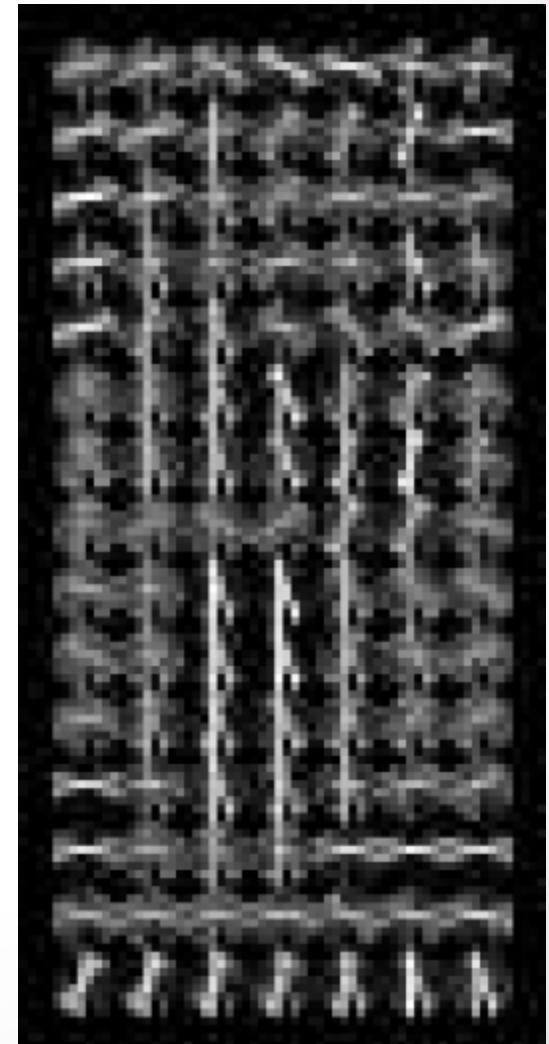
Sobel

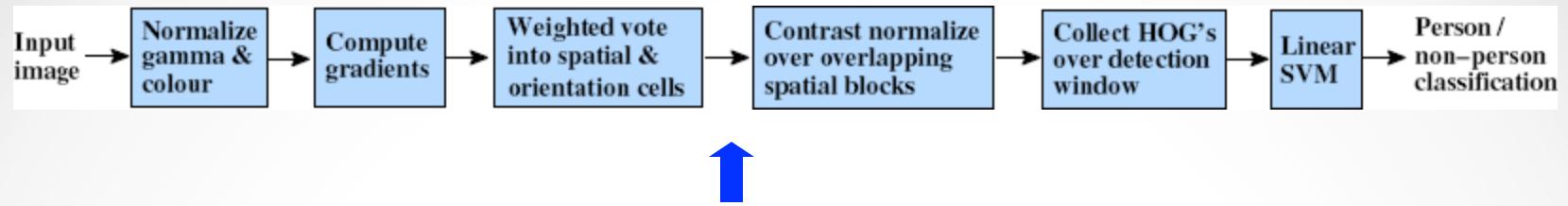


- Histogram of gradient orientations
 - Orientation
 - Position

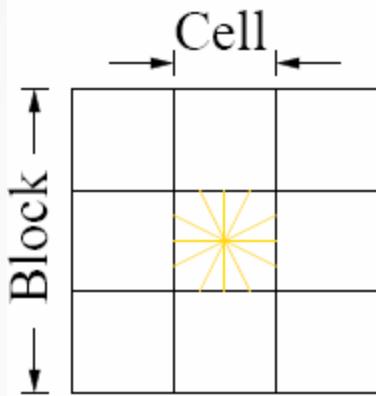


- Weighted by magnitude

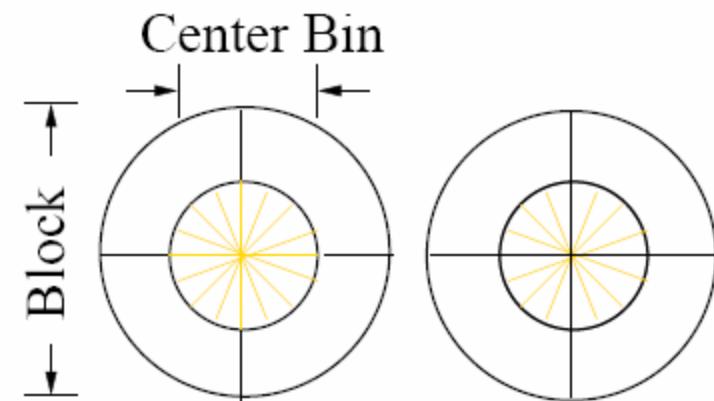




R-HOG

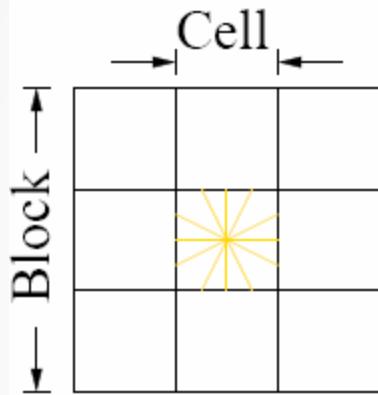


C-HOG

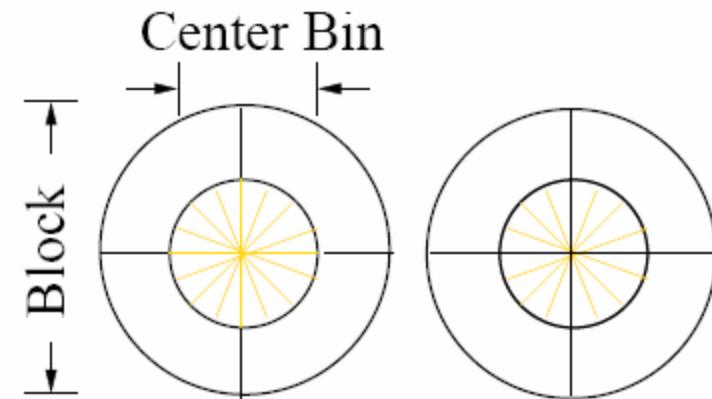




R-HOG



C-HOG



$$L1 - norm : v \longrightarrow v / (\|v\|_1 + \epsilon)$$

$$L2 - norm : v \longrightarrow v / \sqrt{\|v\|_2^2 + \epsilon^2}$$

$$L1 - sqrt : v \longrightarrow \sqrt{v / (\|v\|_1 + \epsilon)}$$

L2 - hys : L2-norm, plus clipping at .2 and renormalizing

