

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
 - - Hard to build good models with a few parameters

Discriminative vs. Generative Models

- **Generative:**
 - + Possibly interpretable
 - + Models the object class/can draw samples
 - - Model variability unimportant to classification task
 - - Hard to build good models with a few parameters
- **Discriminative:**
 - + Appealing when infeasible to model data itself
 - + Often excels in practice for classification
 - - May 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
 - Better able to deal with moving body parts
 - Better 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

Advantages and Disadvantages

- **Part-Based**
 - Better able to deal with moving body parts
 - Better 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

- Extremely and successful in the vision
- Avoids hard decisions vs. edge based features
- Examples:
 - SIFT (Scale-Invariant Image Transform)
 - GLOH (Gradient Location and Orientation Histogram)
 - HOG (Histogram of Oriented Gradients)

Computing Gradients

- Derivatives

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

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

Computing Gradients

- Derivatives

- 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
----	---	---

Computing Gradients

- Derivatives

- 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:

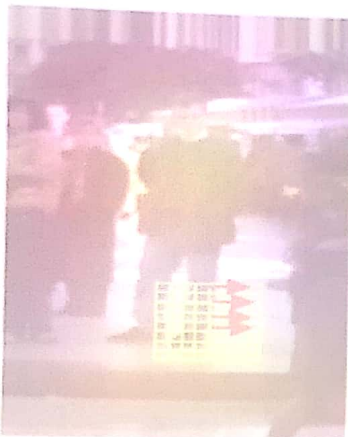
- Magnitude: $s = \sqrt{s_x^2 + s_y^2}$

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



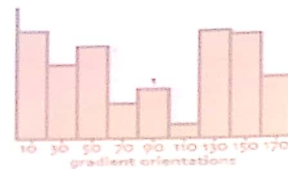
Histograms

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



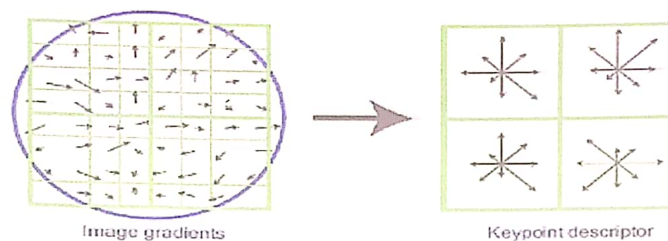
Histograms

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



Recap: SIFT descriptor

- Note that SIFT was a 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 (HOG)

- Histogram of Oriented Gradients for Human Detection
 - Navneet Dalal & Bill Triggs (INRIA Rhône-Alps), CVPR 2005
- Global descriptor for the complete body
- Very high-dimensional: ~ 4000 dimensions
- Significant improvement over SoTA



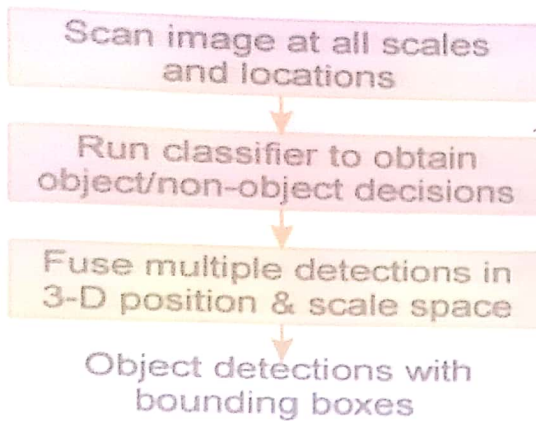
Detector: Learning Phase

1. Learning

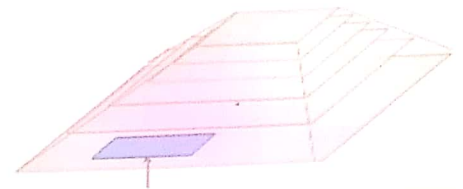


Detector: Detection Phase

2. Detection



Scale-space pyramid



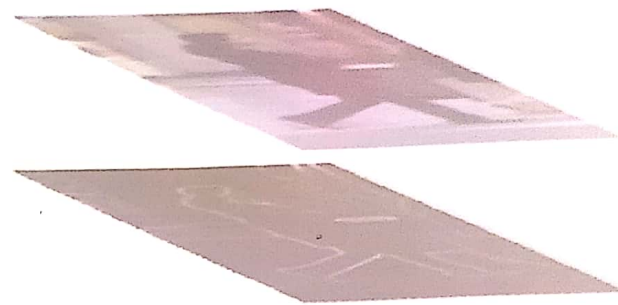
Sliding window over each scale

Simple SVM prediction



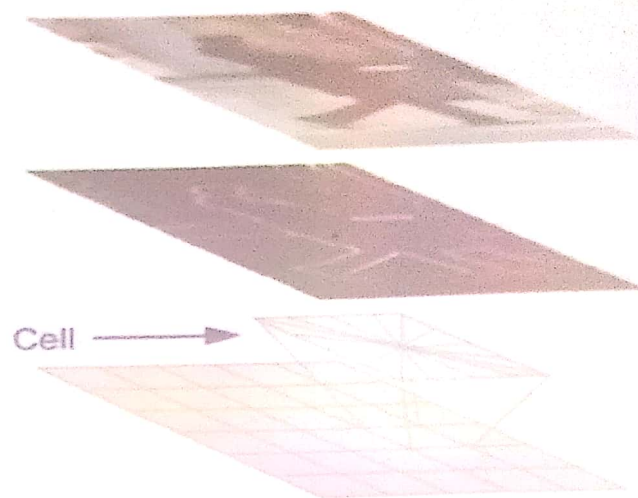
Descriptor

1. Compute gradients on an image region of 64×128 pixels



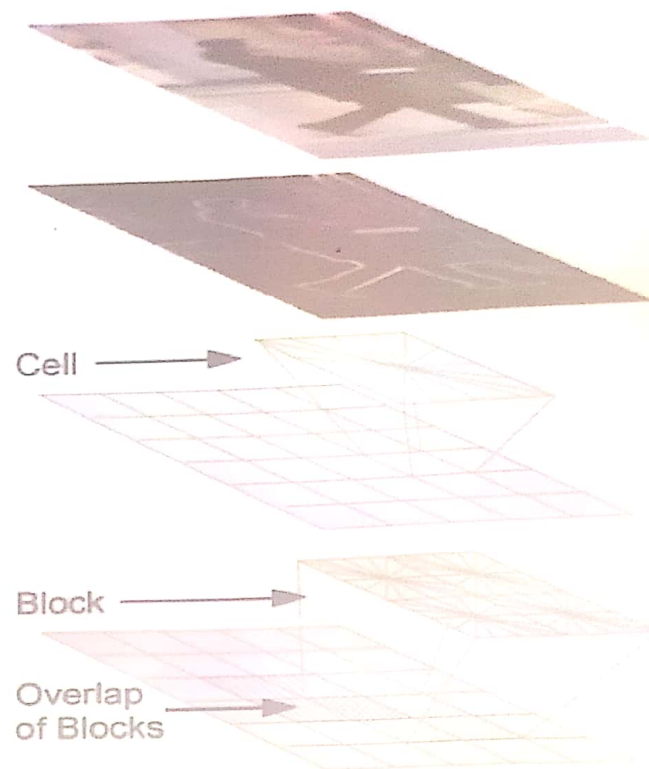
Descriptor

1. Compute gradients on an image region of 64x128 pixels
2. Compute histograms on 'cells' of typically 8x8 pixels (i.e. 8x16 cells)



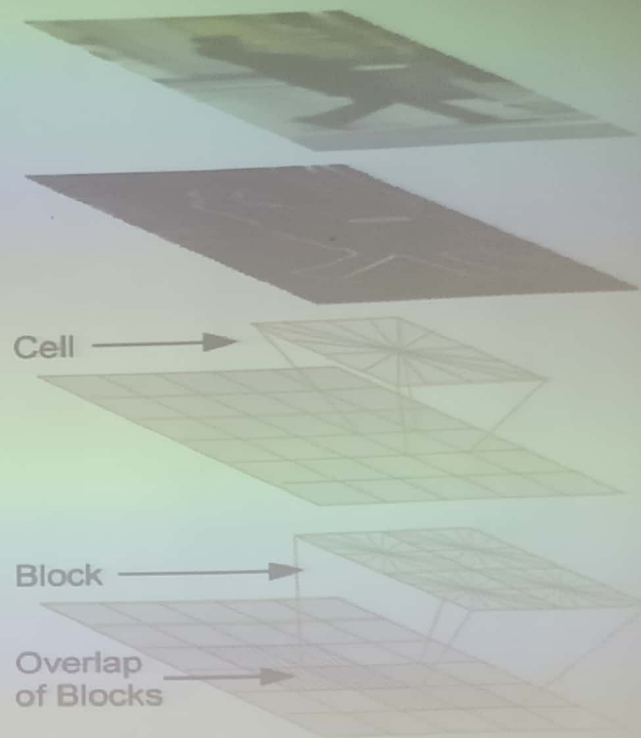
Descriptor

1. Compute gradients on an image region of 64x128 pixels
2. Compute histograms on 'cells' of typically 8x8 pixels (i.e. 8x16 cells)
3. Normalize histograms within overlapping blocks of cells (typically 2x2 cells, i.e. 7x15 blocks)



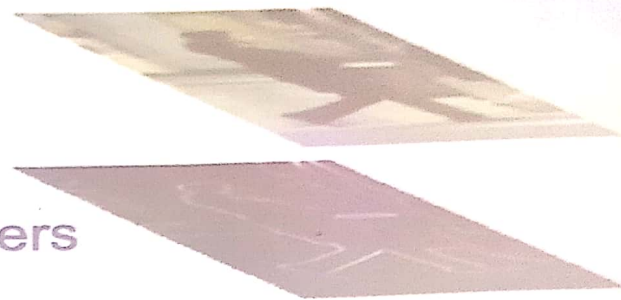
Descriptor

1. Compute gradients on an image region of 64x128 pixels
2. Compute histograms on 'cells' of typically 8x8 pixels (i.e. 8x16 cells)
3. Normalize histograms within overlapping blocks of cells (typically 2x2 cells, i.e. 7x15 blocks)



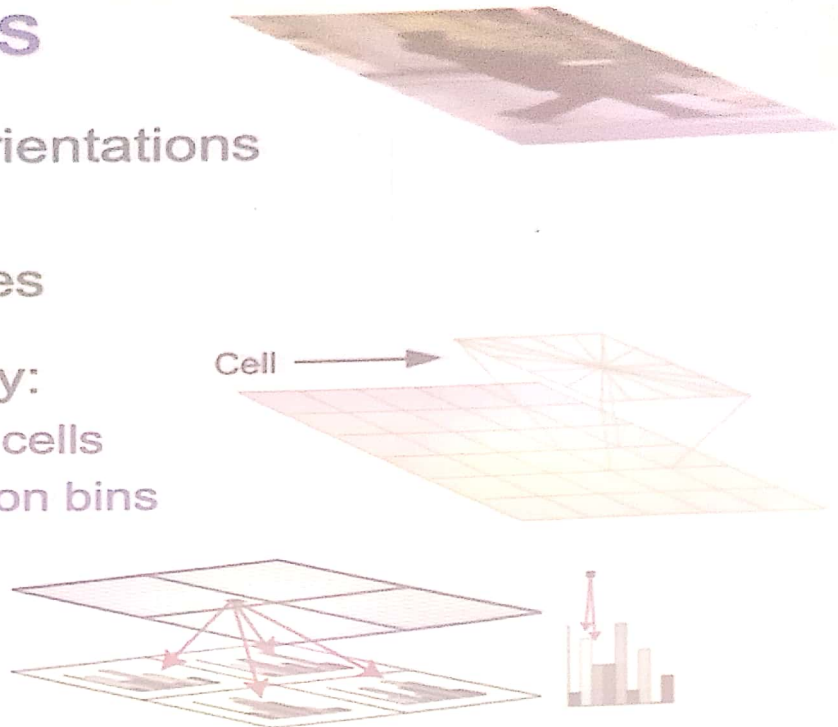
Gradients

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



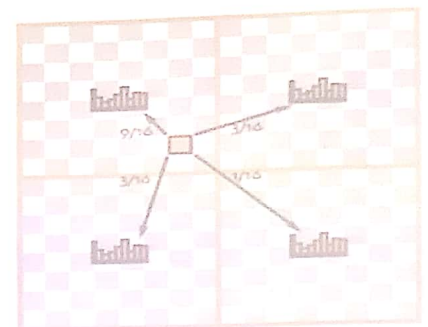
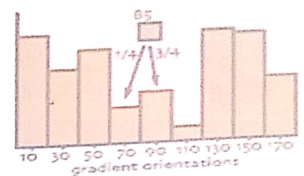
Cell histograms

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



Histogram Interpolation Example

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

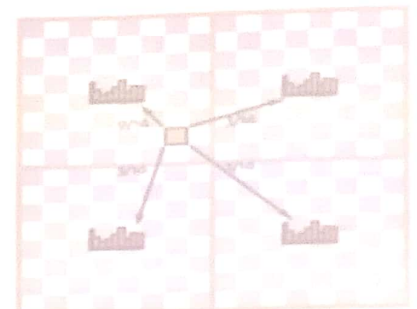


Histogram Interpolation Example

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

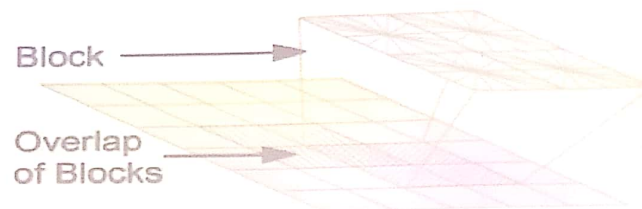


- Distance to bin centers
 - Left: 2, Right: 6
 - Top: 2, Bottom: 6
- Ratio Left-Right: $6/8$, $2/8$
- Ratio Top-Bottom: $6/8$, $2/8$
- Ratios:
 - $6/8 * 6/8 = 36/64 = 9/16$
 - $6/8 * 2/8 = 12/64 = 3/16$
 - $2/8 * 6/8 = 12/64 = 3/16$
 - $2/8 * 2/8 = 4/64 = 1/16$



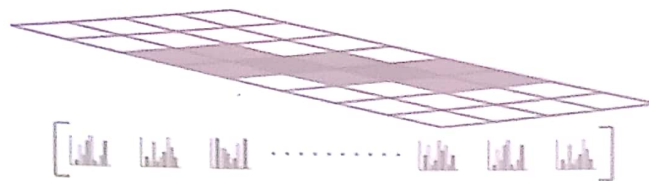
Blocks

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



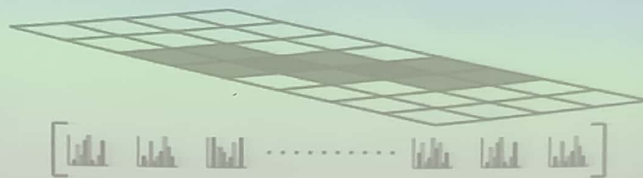
Final Descriptor

- Concatenation of Blocks



Final Descriptor

- Concatenation of Blocks



- Visualization:

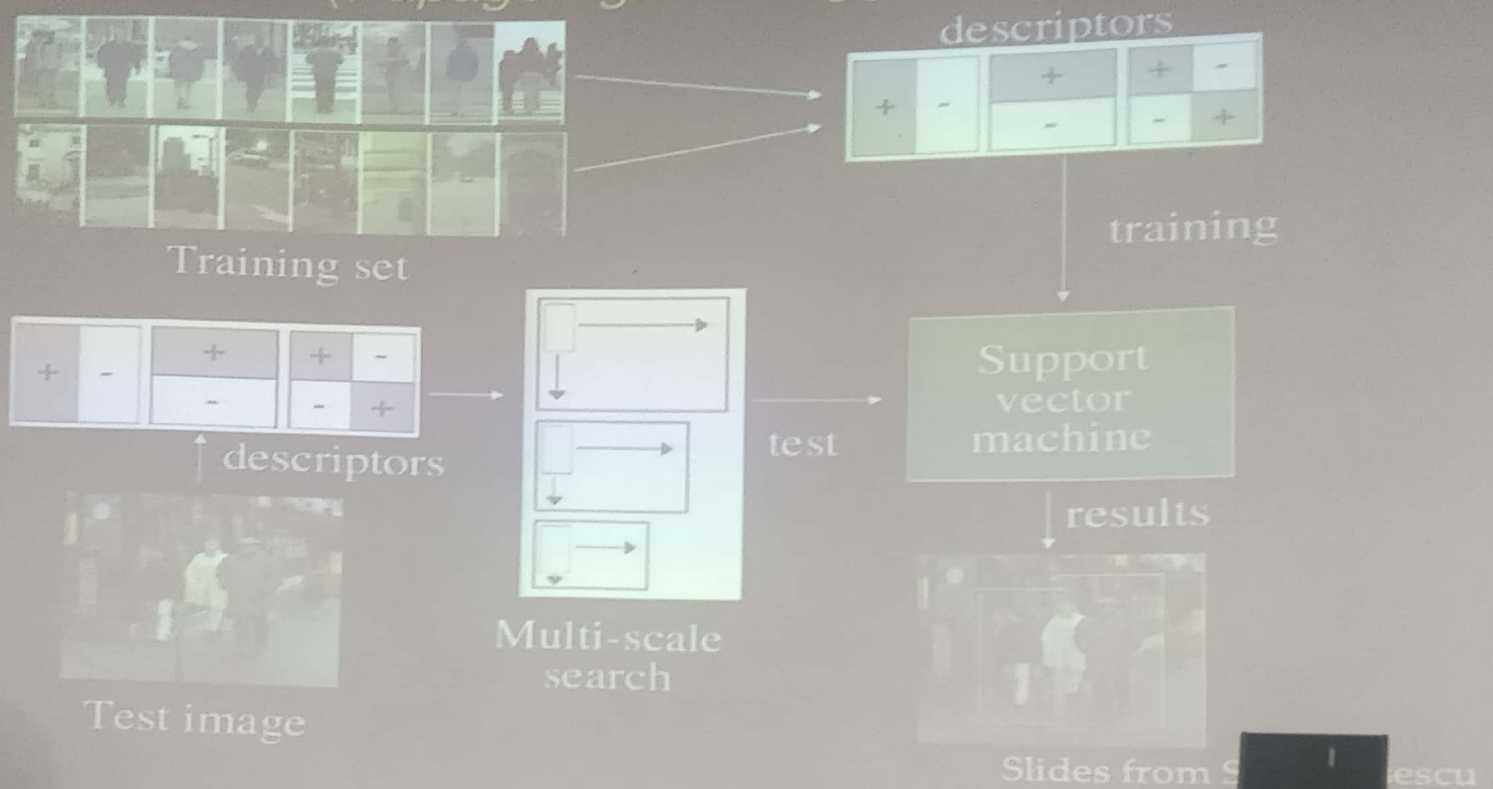


Engineering

- 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

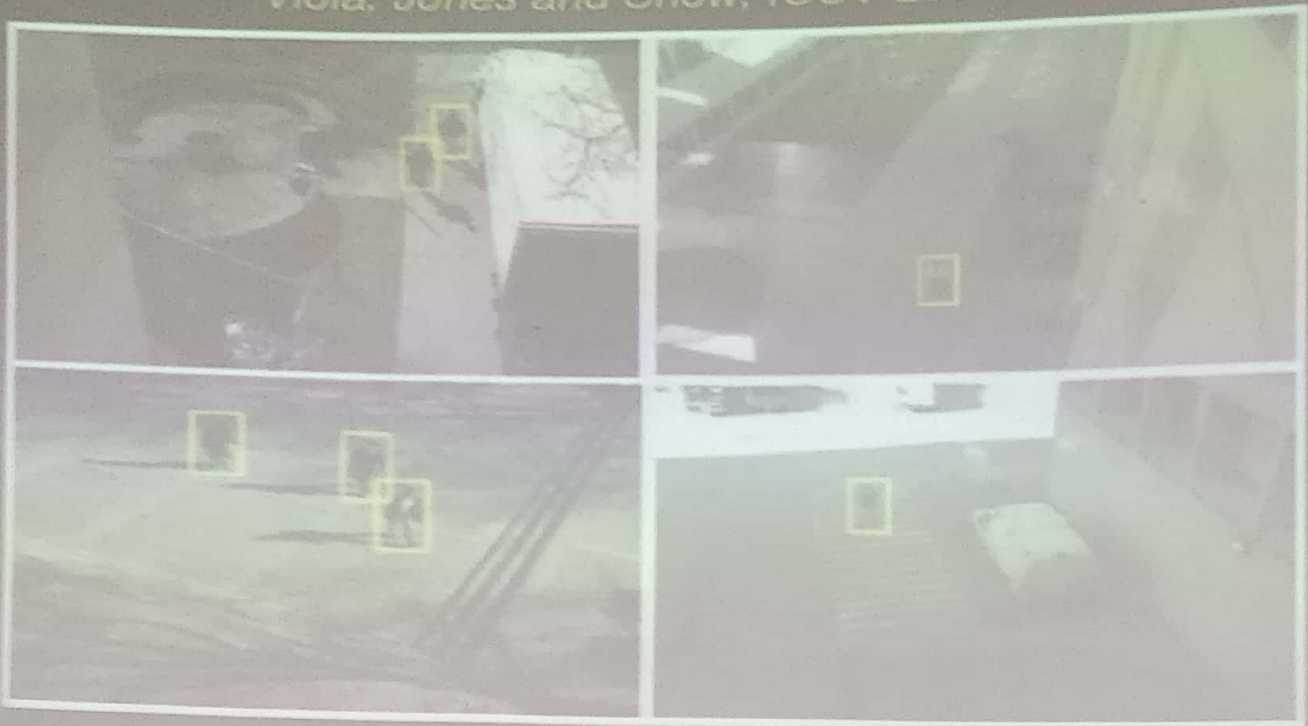
Support Vector Machine Detector

(Papagerogiu & 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

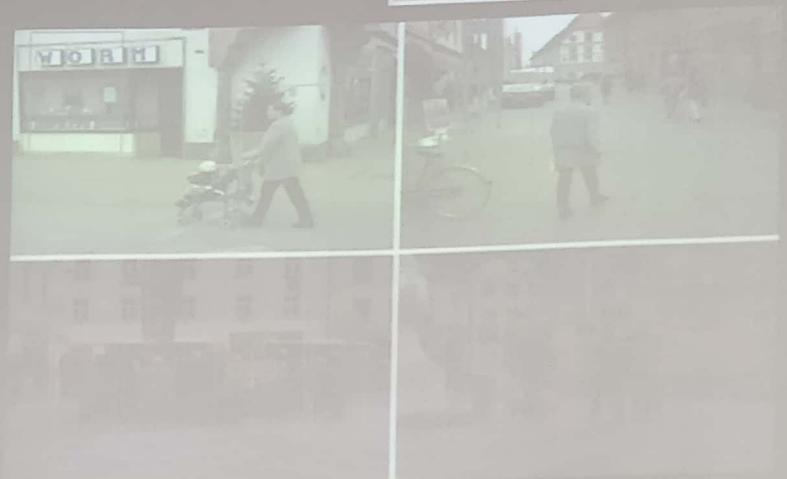
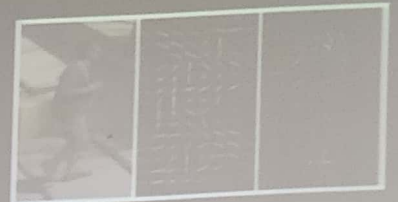
Slides from S. Ionescu

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



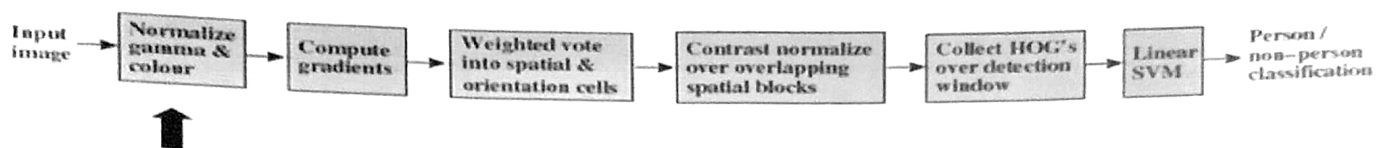
Slides from S. Ionescu

Feature Sets

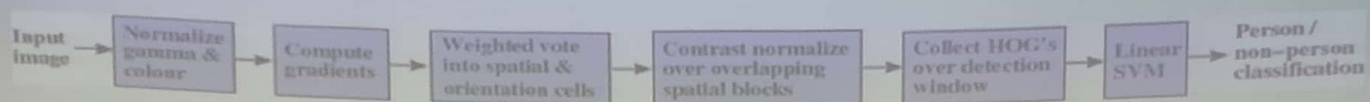
- Haar wavelets + SVM:
 - Papageorgiou & Poggio (2000)
 - Mohan et al (2001)
 - DePoortere et al (2002)
- Rectangular differential features + adaBoost:
 - Viola & Jones(2001)
- Parts based binary orientation position histogram + adaBoost:
 - Mikolajczyk et al (2004)
- Edge templates + nearest neighbor: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

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:**
 - Gavrilu & Philomen (1999)
- **Dynamic programming:**
 - Felzenszwalb & Huttenlocher (2000),
 - Lofte & Forsyth (1999)
- **Orientation histograms:**
 - C.F. Freeman et al (1996)
 - Lowe(1999)
- **Shape contexts:**
 - Belongie et al (2002)
- **PCA-SIFT:**
 - Ke and Sukthankar (2004)



- Color Space: Tested with
 - RGB
 - LAB
 - Grayscale
- Gamma Normalization and Compression
 - Square root
 - Log
 - None



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

centered

-1	1
----	---

uncentered

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

cubic-corrected

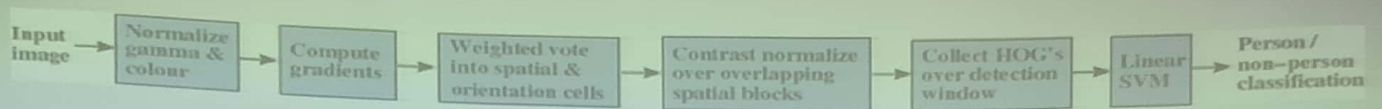


0	1
-1	0

diagonal

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

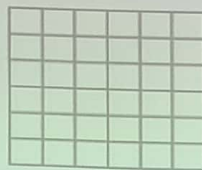
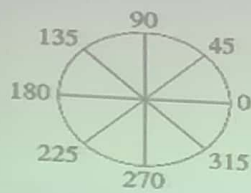
Sobel



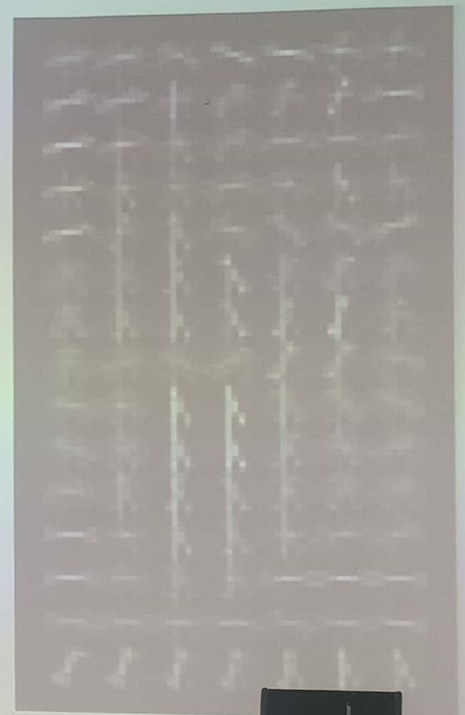
- Histogram of gradient orientations

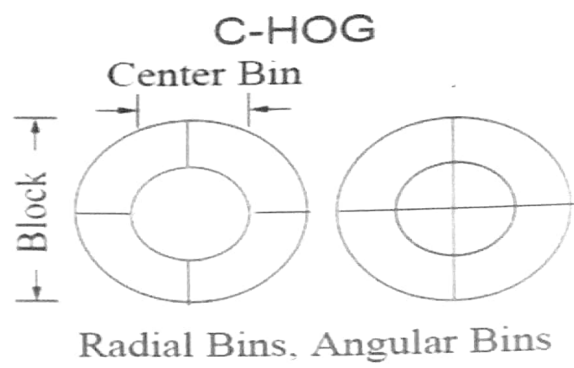
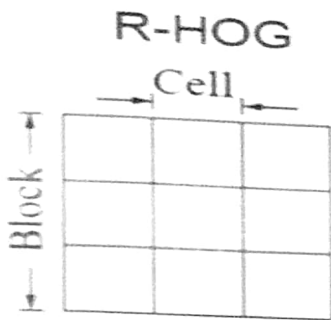
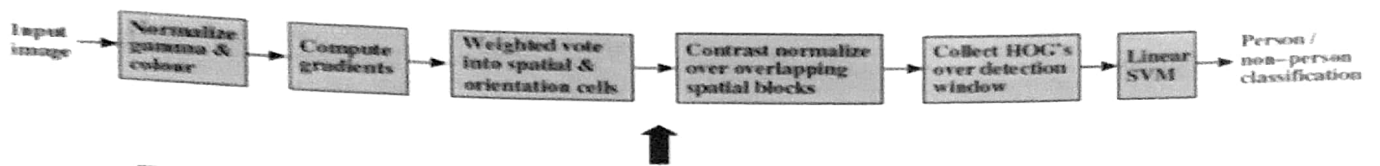
-Orientation

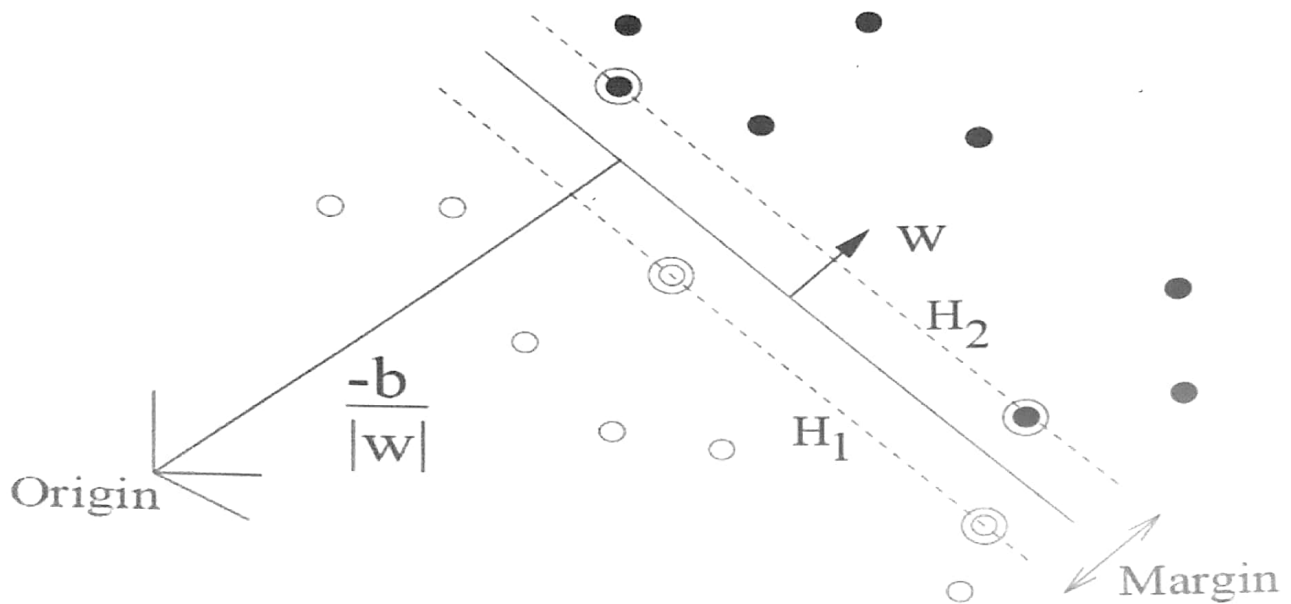
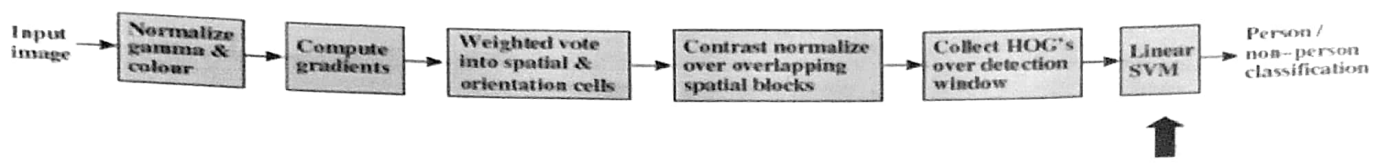
-Position



- Weighted by magnitude



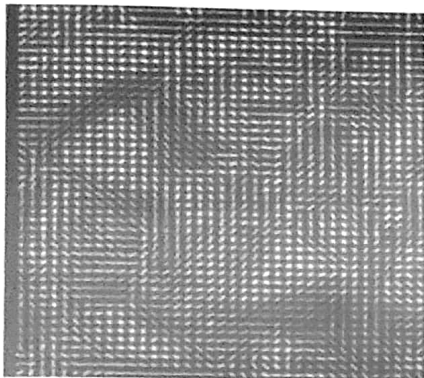




HOGgles: Visualizing Object Detection Features

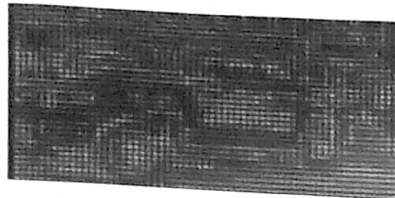
Carl Vondrick, Aditya Khosla, Hamed Pirsiavash, Tomasz Malisiewicz, and Antonio Torralba (MIT), ICCV 2013

- Sparse dictionary based encoding of images.
- Use weights from sparse HOG Basis to form image

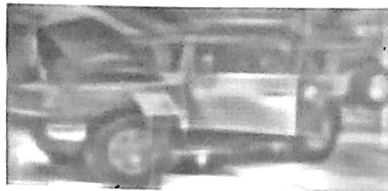


IHOG: Inverting HOG

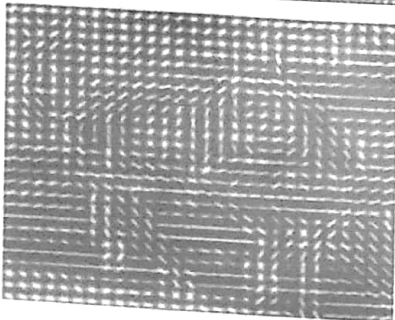
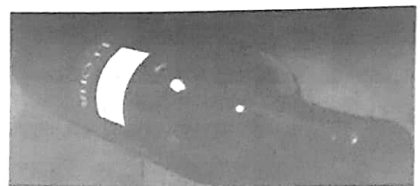
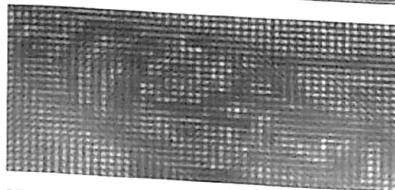
HOG



Inverse



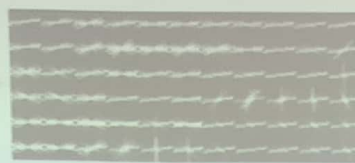
Original



What went wrong? IHOG



Car Detection



HOG Features



Our Visualization

Newer Approaches, Datasets

1. P. Felzenszwalb, D. McAllester, D. Ramanan, "A Discriminatively Trained, Multiscale, Deformable Part Model", CVPR, 2008. (*PAMI Longuet-Higgins Prize, 2018*)
 2. S. Walk, N. Majer, K. Schindler and B. Schiele, "New Features and Insights for Pedestrian Detection", CVPR 2010.
 3. Piotr Dollár, S. Belongie and P. Perona, "The Fastest Pedestrian Detector in the West", BMVC 2010
 4. P. Dollár, R. Appel and W. Kienzle, "Crosstalk Cascades for Frame-Rate Pedestrian Detection", ECCV 2012. (*faster than [3]*)
- Caltech Pedestrian Detection Benchmark
http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/
 - WIDER Face & Pedestrian Challenge – Tr2: Pedestrian Detection
<http://wider-challenge.org>