Object Recognition and Scene Understanding

student presentation

6.870

Template matching and histograms

Nicolas Pinto

Introduction

Hosts

a guy...

(who has big arms)

Antonio T...

(who knows a lot about vision)

a frog...

(who has big eyes)

and thus should know
a lot about vision...

3 papers

Lowe (1999)

Object Recognition from Local Scale-Invariant Features

David G. Lowe
Computer Science Department
University of British Columbia
Vancouver, B.C., V6T 1Z4, Canada
lowe@cs.ubc.ca

Abstract

An object recognition system has been developed that uses a new class of local image features. The features are invariant

translation, scaling, and rotation, and partially invariant to illumination changes and affine or 3D projection. Previous approaches to local feature generation lacked invariance to scale and were more sensitive to projective distortion and

Nalal and Triggs (2005)

Histograms of Oriented Gradients for Human Detection

Navneet Dalal and Bill Triggs
INRIA Rhône-Alps, 655 avenue de l'Europe, Montbonnot 38334, France
{Navneet.Dalal,Bill.Triggs}@inrialpes.fr, http://lear.inrialpes.fr

Abstract

We study the question of feature sets for robust visual object recognition, adopting linear SVM based human detection as a test case. After reviewing existing edge and gradient based descriptors, we show experimentally that grids of Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for human detec-

We briefly discuss previous work on human detection in §2, give an overview of our method §3, describe our data sets in §4 and give a detailed description and experimental evaluation of each stage of the process in §5–6. The main conclusions are summarized in §7.

2 Previous Work

Felzenszwalb et al. (2008)

A Discriminatively Trained, Multiscale, Deformable Part Model

Pedro Felzenszwalb
University of Chicago
pff@cs.uchicago.edu

David McAllester
Toyota Technological Institute at Chicago
mcallester@tti-c.org

Deva Ramanan
UC Irvine
dramanan@ics.uci.edu

Abstract

This paper describes a discriminatively trained, multiscale, deformable part model for object detection. Our system achieves a two-fold improvement in average precision over the best performance in the 2006 PASCAL person detection challenge. It also outperforms the best results in the 2007 challenge in ten out of twenty categories. The system relies heavily on deformable parts. While deformable part models have become quite popular, their value had not been

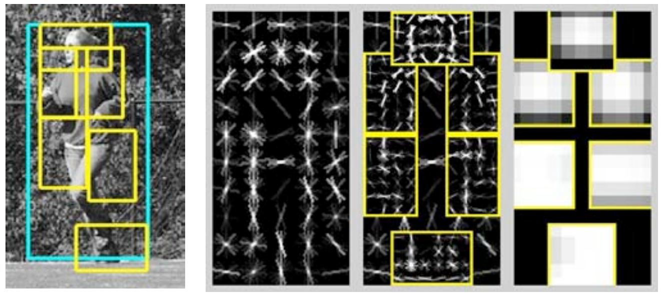
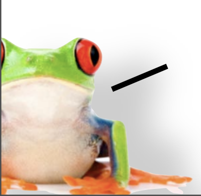


Figure 1. Example detection obtained with the person model. The



yey!!

Lowe (1999)

Object Recognition from Local Scale-Invariant Features

David G. Lowe
Computer Science Department
University of British Columbia
Vancouver, B.C., V6T 1Z4, Canada
lowe@cs.ubc.ca

Abstract

An object recognition system has been developed that uses a new class of local image features. The features are invariant

translation, scaling, and rotation, and partially invariant to illumination changes and affine or 3D projection. Previous approaches to local feature generation lacked invariance to scale and were more sensitive to projective distortion and

Nalal and Triggs (2005)

Histograms of Oriented Gradients for Human Detection

Navneet Dalal and Bill Triggs
INRIA Rhône-Alps, 655 avenue de l'Europe, Montbonnot 38334, France
{Navneet.Dalal,Bill.Triggs}@inrialpes.fr, http://lear.inrialpes.fr

Abstract

We study the question of feature sets for robust visual object recognition, adopting linear SVM based human detection as a test case. After reviewing existing edge and gradient based descriptors, we show experimentally that grids of Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for human detec-

We briefly discuss previous work on human detection in §2, give an overview of our method §3, describe our data sets in §4 and give a detailed description and experimental evaluation of each stage of the process in §5–6. The main conclusions are summarized in §7.

2 Previous Work

Felzenszwalb et al. (2008)

A Discriminatively Trained, Multiscale, Deformable Part Model

Pedro Felzenszwalb, University of Chicago, *pff@cs.uchicago.edu*
David McAllester, Toyota Technological Institute at Chicago, *mcallester@tti-c.org*
Deva Ramanan, UC Irvine, *dramanan@ics.uci.edu*

Abstract

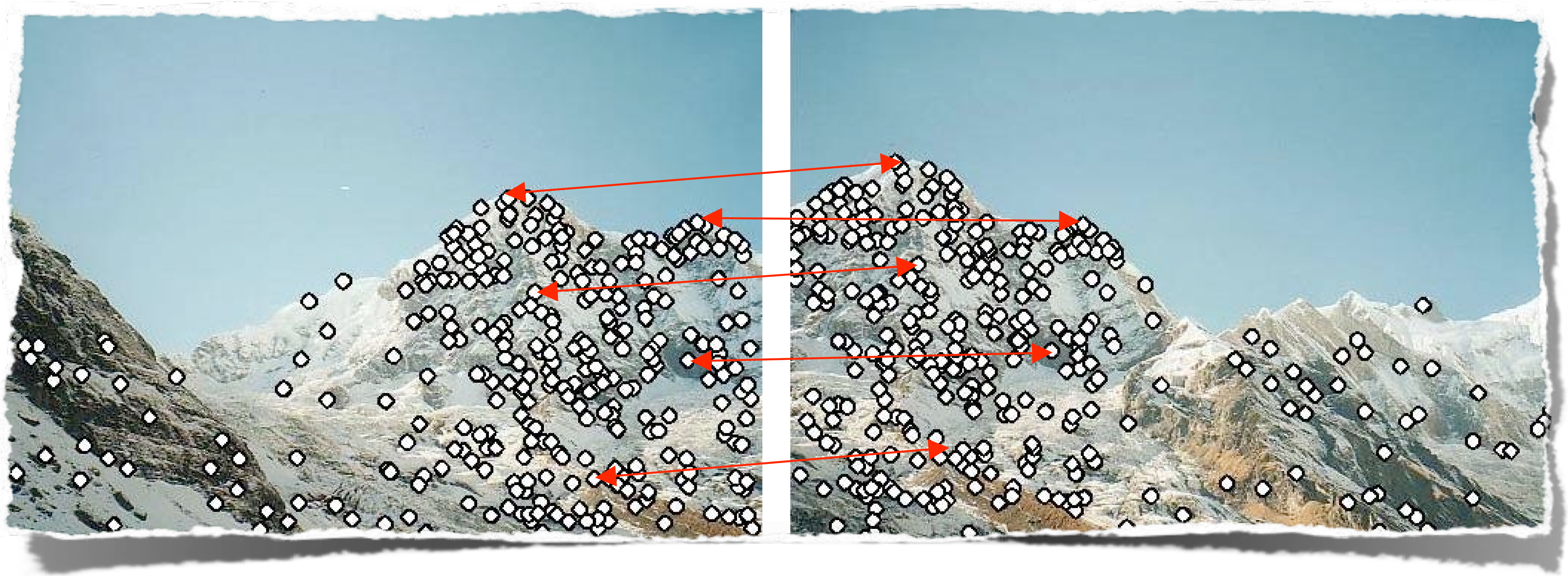
This paper describes a discriminatively trained, multiscale, deformable part model for object detection. Our system achieves a two-fold improvement in average precision over the best performance in the 2006 PASCAL person detection challenge. It also outperforms the best results in the 2007 challenge in ten out of twenty categories. The system relies heavily on deformable parts. While deformable part models have become quite popular, their value had not been

Figure 1. Example detection obtained with the person model. The

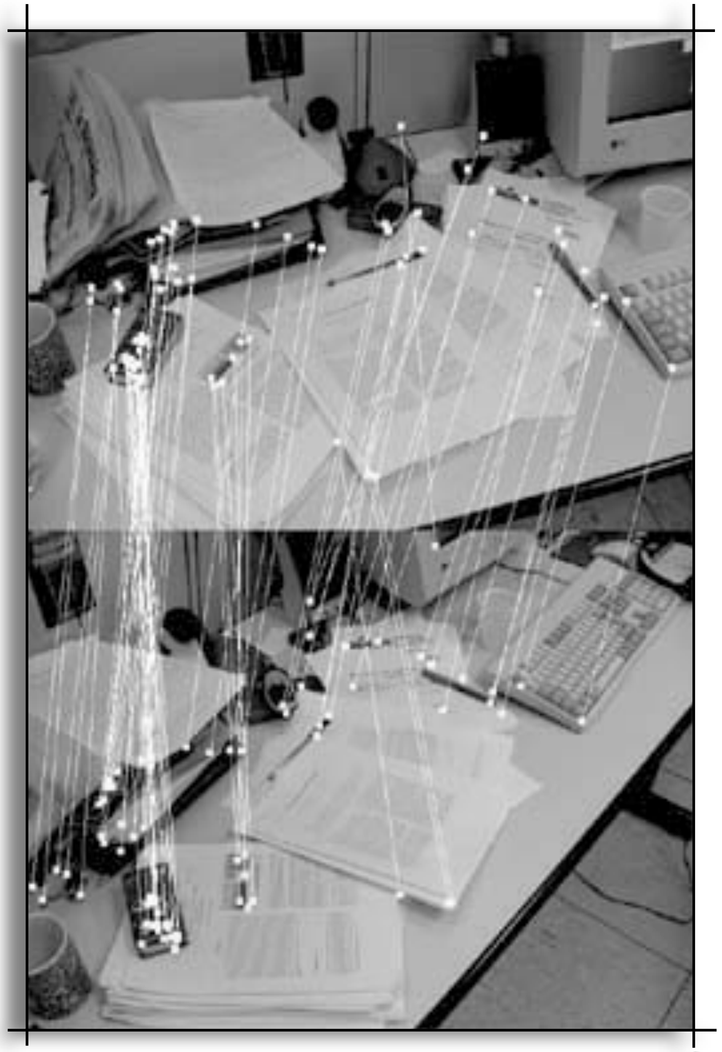
Scale-Invariant Feature Transform (SIFT)

adapted from Kucuktunc

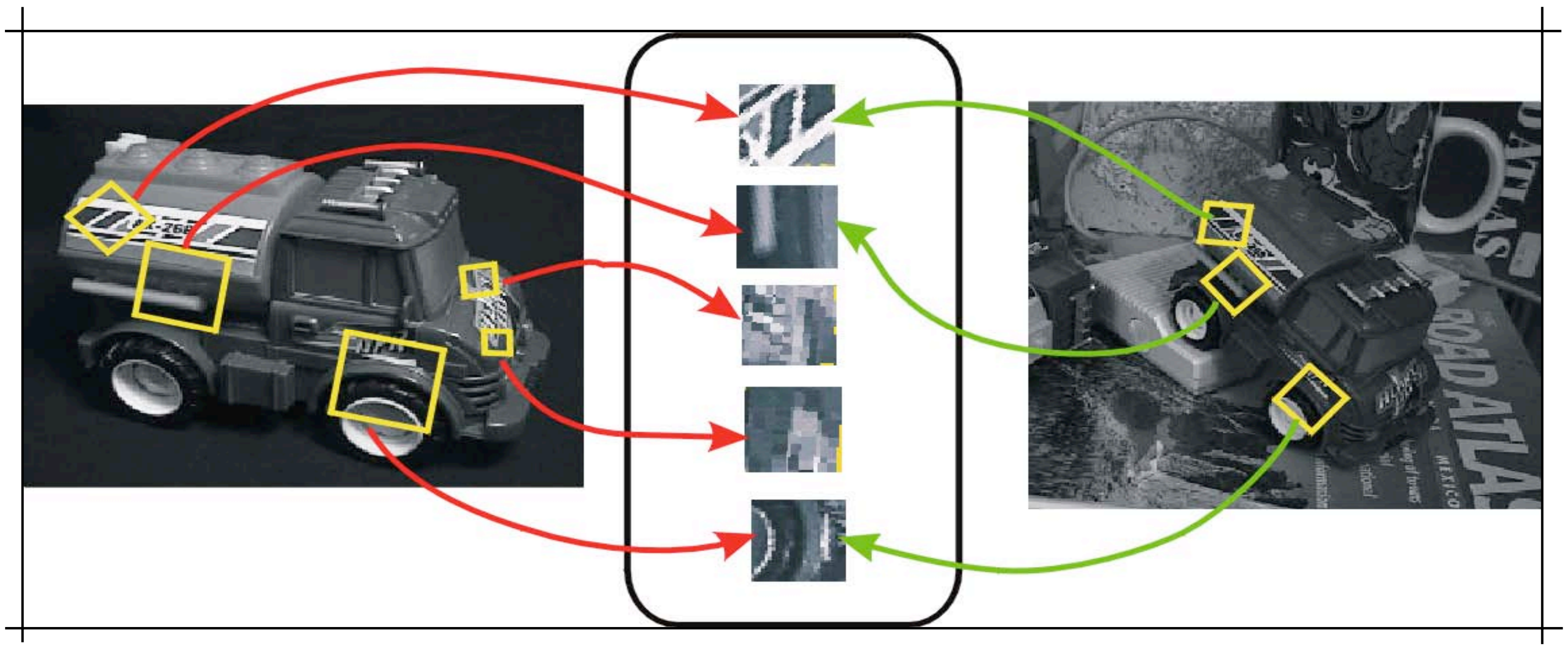
Scale-Invariant Feature Transform (SIFT)



adapted from Brown, ICCV 2003



SIFT *local* features are **invariant...**



adapted from David Lee

like me they are **robust**...

... to changes in illumination, noise, viewpoint, occlusion, etc.

I am sure you want to know **how to build them**

1. find interest points or **"keypoints"**

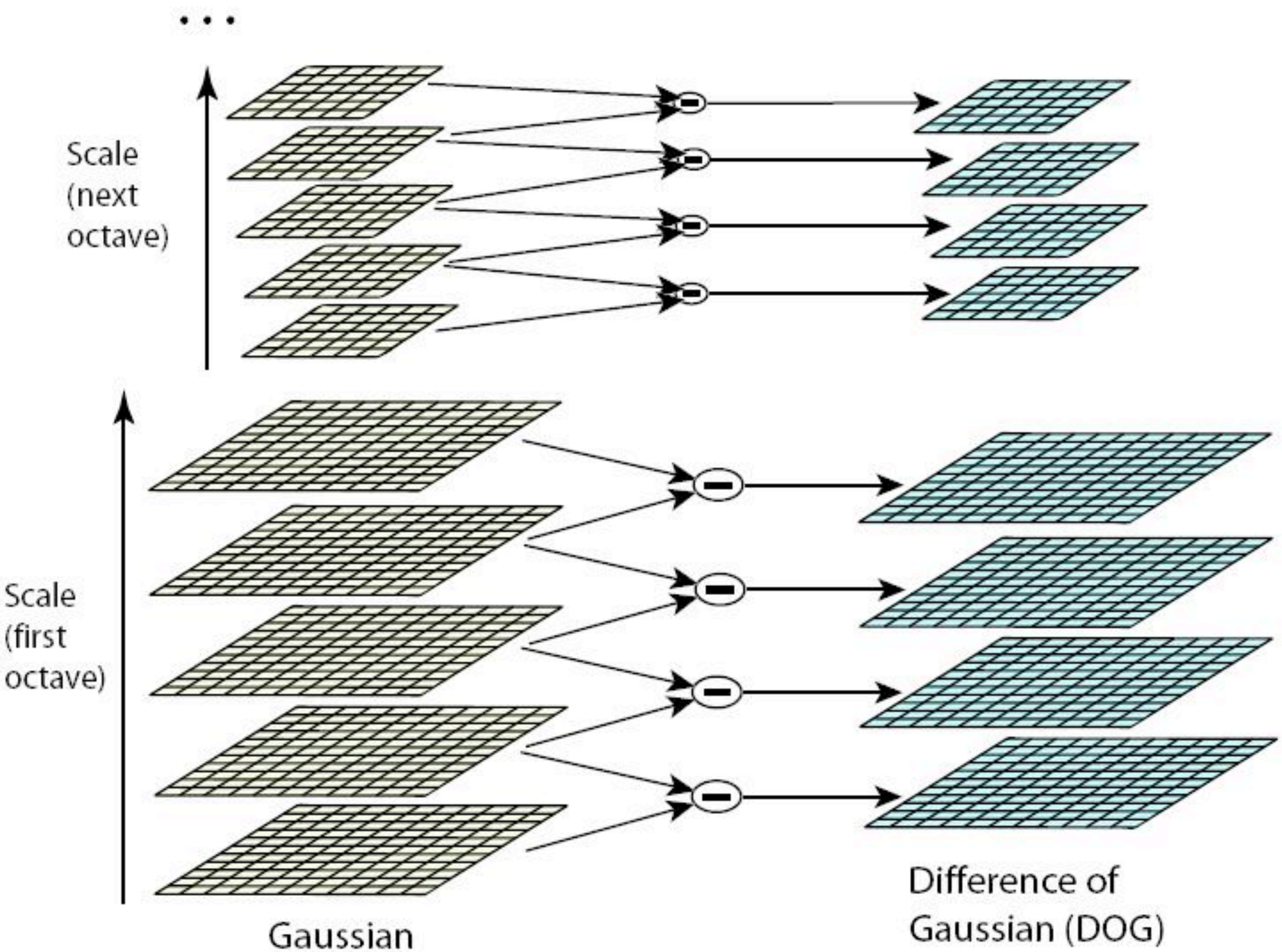
2. find their **dominant orientation**

3. compute their **descriptor**

4. **match** them on other images

1. find interest points or “keypoints”

keypoints are taken as maxima/minima of a **DoG pyramid**



in this settings, extremas are invariant to scale...

a DoG (Difference of Gaussians) pyramid is simple to compute...

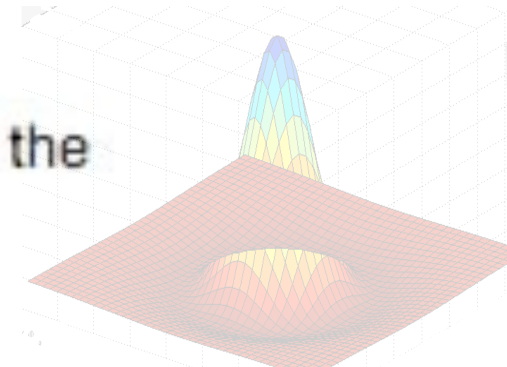
even him can do it!

Specifically, a DoG image $D(x, y, \sigma)$ is given by

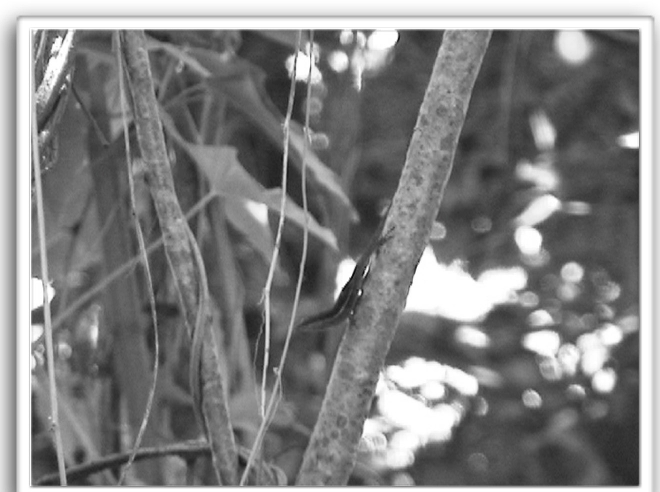
$$D(x, y, \sigma) = L(x, y, k_i\sigma) - L(x, y, k_j\sigma),$$

where $L(x, y, k\sigma)$ is the original image $I(x, y)$ convolved with the Gaussian blur $G(x, y, k\sigma)$ at scale $k\sigma$, i.e.,

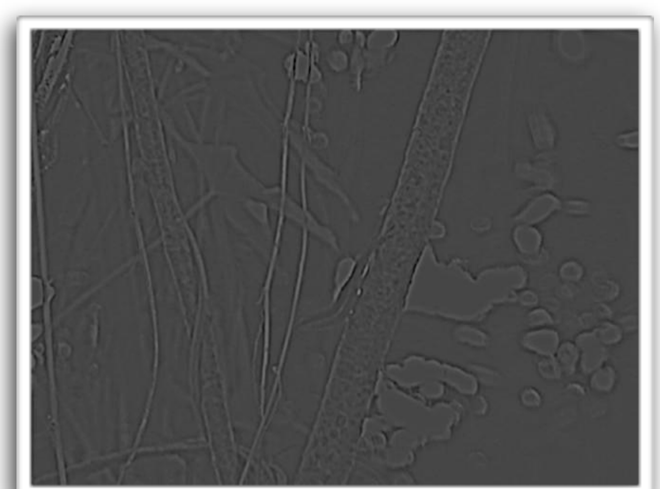
$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y)$$



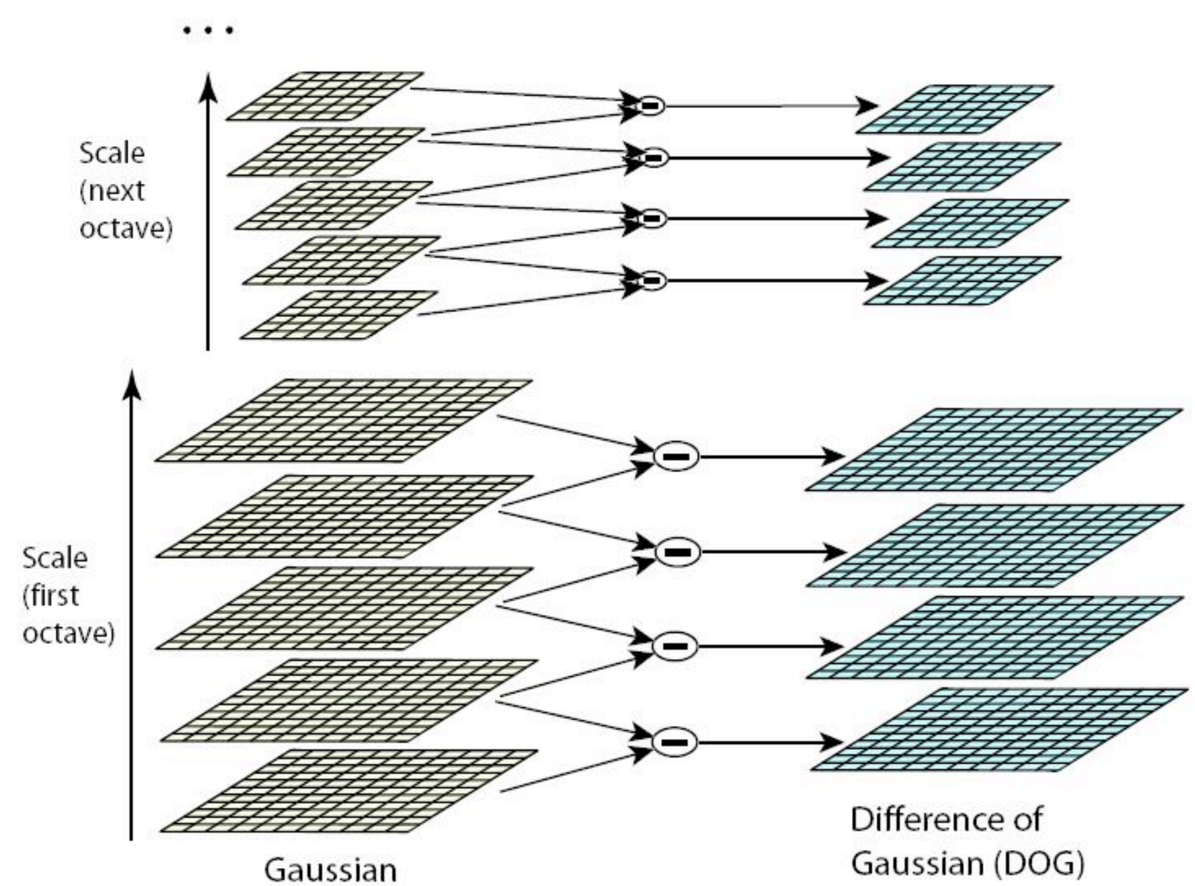
before



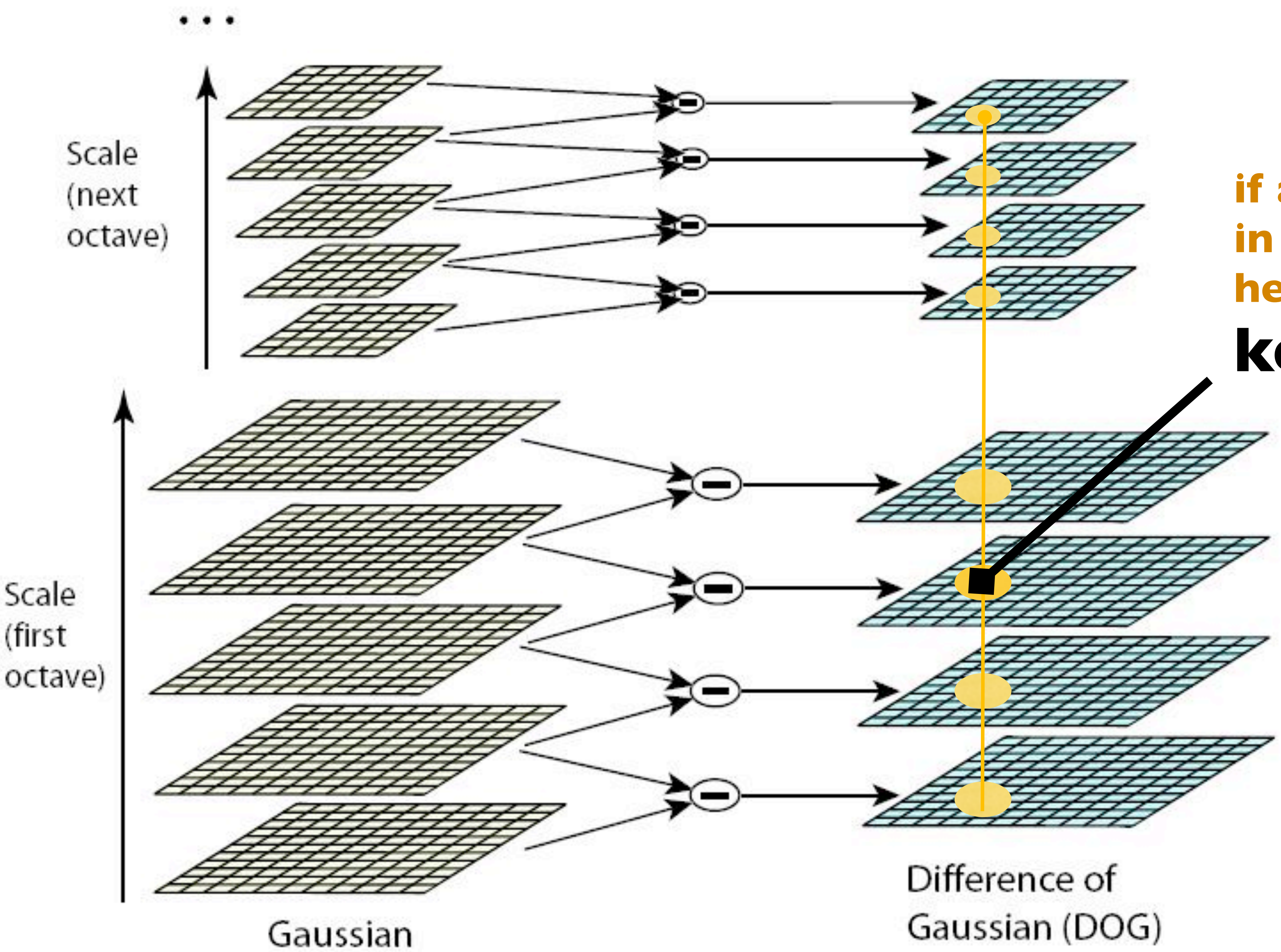
after



adapted from Pallus and Fleishman



then we just have to find **neighborhood extremas** in this **3D DoG space**



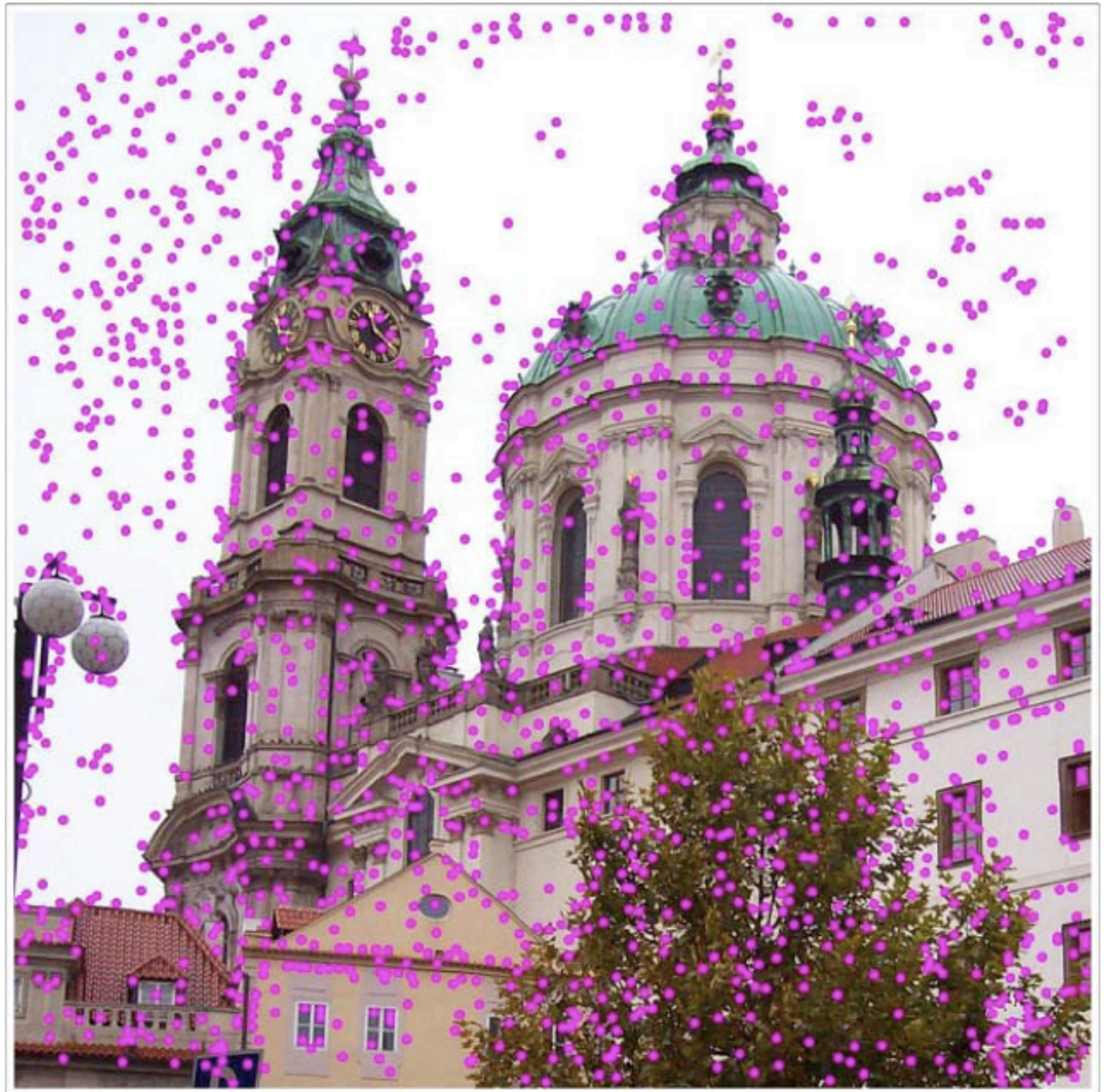
if a pixel is an extrema in its neighboring region he becomes a candidate

keypoint

too many keypoints?

1. remove **low contrast**

2. remove **edges**



adapted from wikipedia

2. find their **dominant orientation**

each selected keypoint is **assigned to one or more "dominant" orientations...**

... this step is important to achieve **rotation invariance**

How?

using the DoG pyramid to achieve **scale invariance:**

a. compute image **gradient magnitude and orientation**

b. build an **orientation histogram**

c. keypoint’s orientation(s) = **peak(s)**

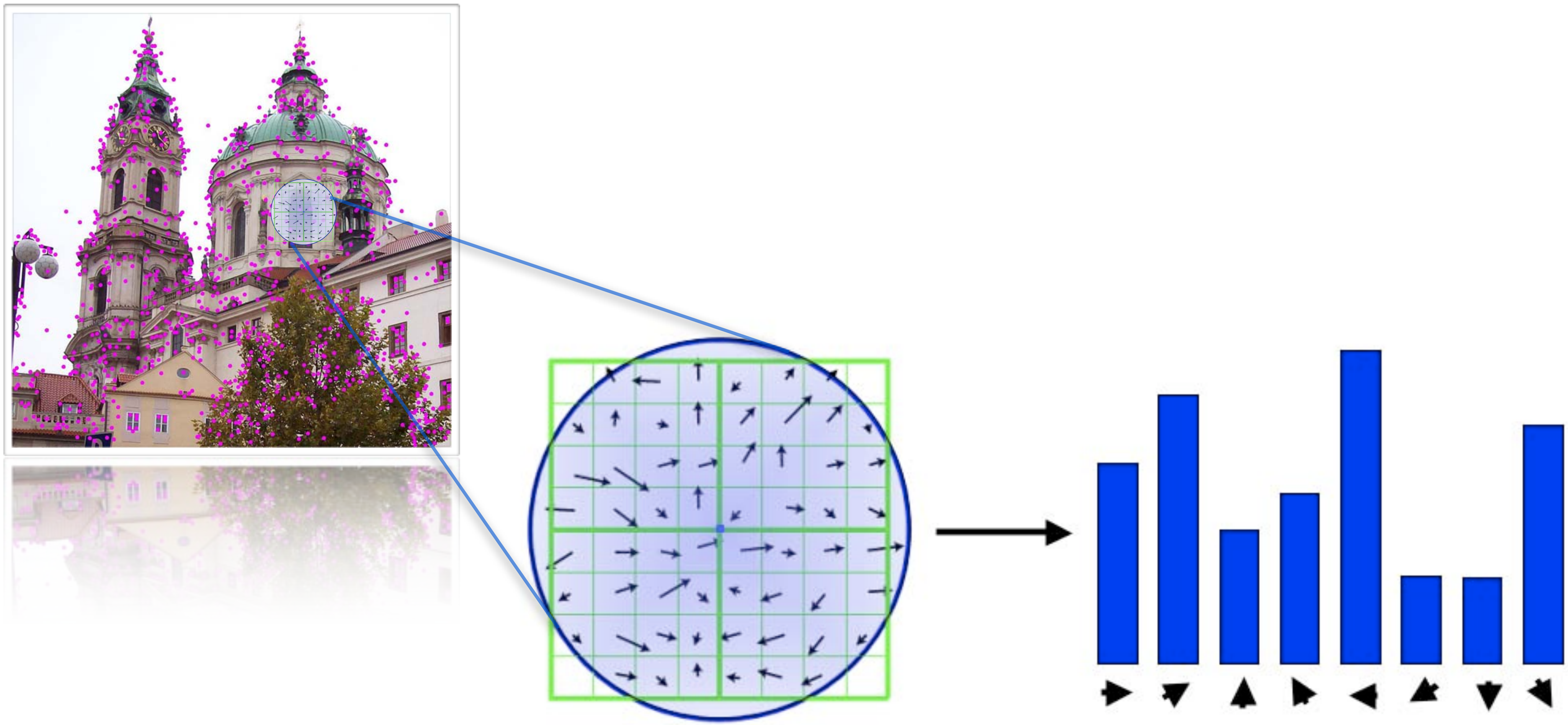
a. compute image gradient magnitude and orientation

First, the Gaussian-smoothed image $L(x, y, \sigma)$ at the keypoint's scale σ is taken so that all computations are performed in a scale-invariant manner. For an image sample $L(x, y)$ at scale σ, the gradient magnitude, $m(x, y)$, and orientation, $\theta(x, y)$, are precomputed using pixel differences:

$$m(x,y) = \sqrt{\left(L(x+1,y) - L(x-1,y)\right)^2 + \left(L(x,y+1) - L(x,y-1)\right)^2}$$

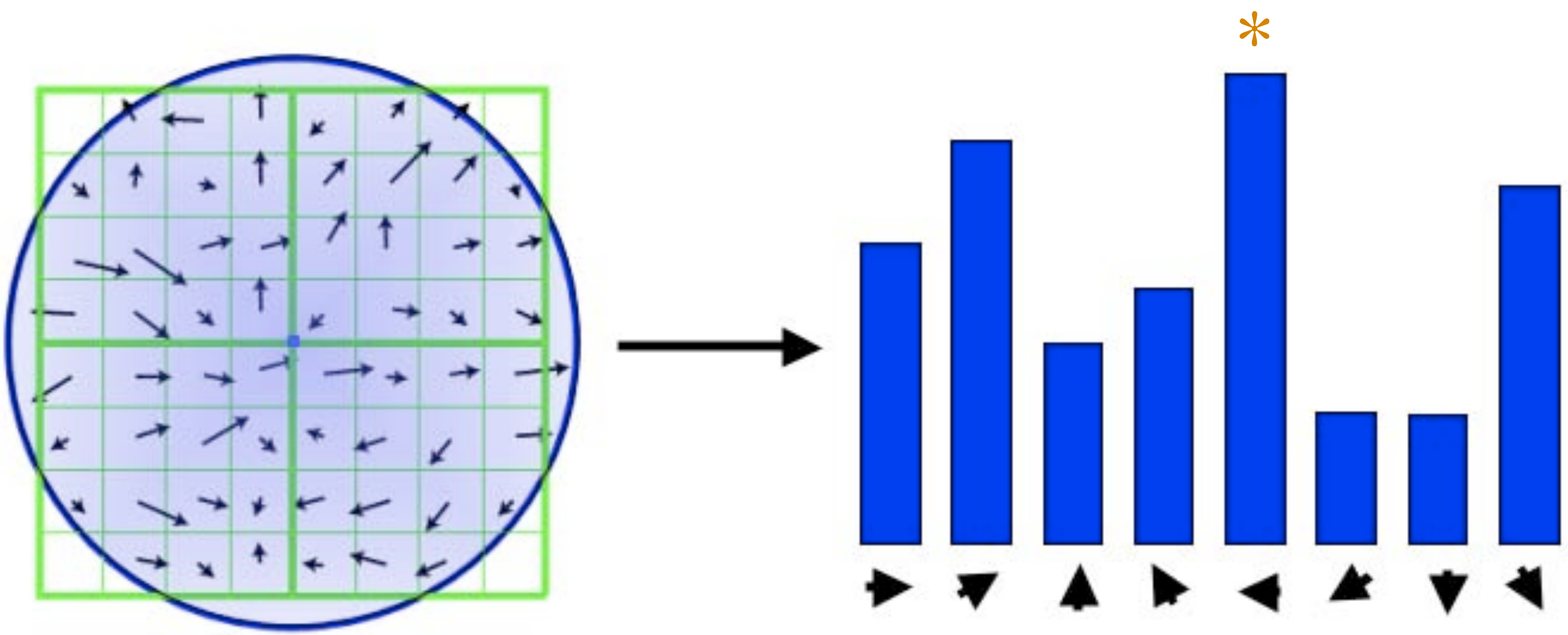
$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right)$$

b. build an orientation histogram



adapted from Ofir Pele

c. keypoint's orientation(s) = peak(s)



* the peak ;-)

3. compute their **descriptor**

SIFT descriptor
= a set of orientation histograms

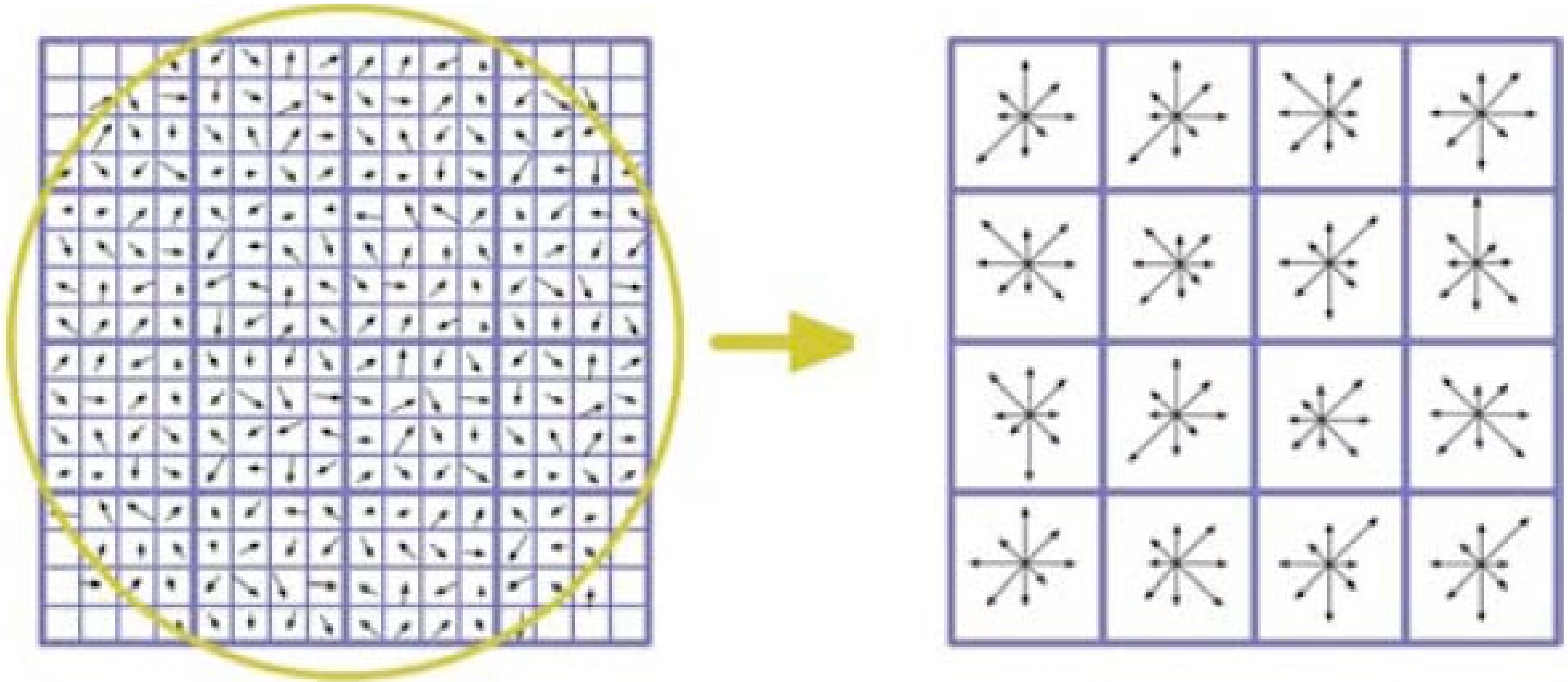


Image Gradients

Key Point Descriptor

16x16 neighborhood of pixel gradients → **4x4 array x 8 bins = 128 dimensions (normalized)**

4. **match** them on other images

How to match?

nearest neighbor
hough transform voting
least-squares fit
etc.

SIFT is great!

\\ invariant to affine transformations

\\ easy to understand

\\ fast to compute

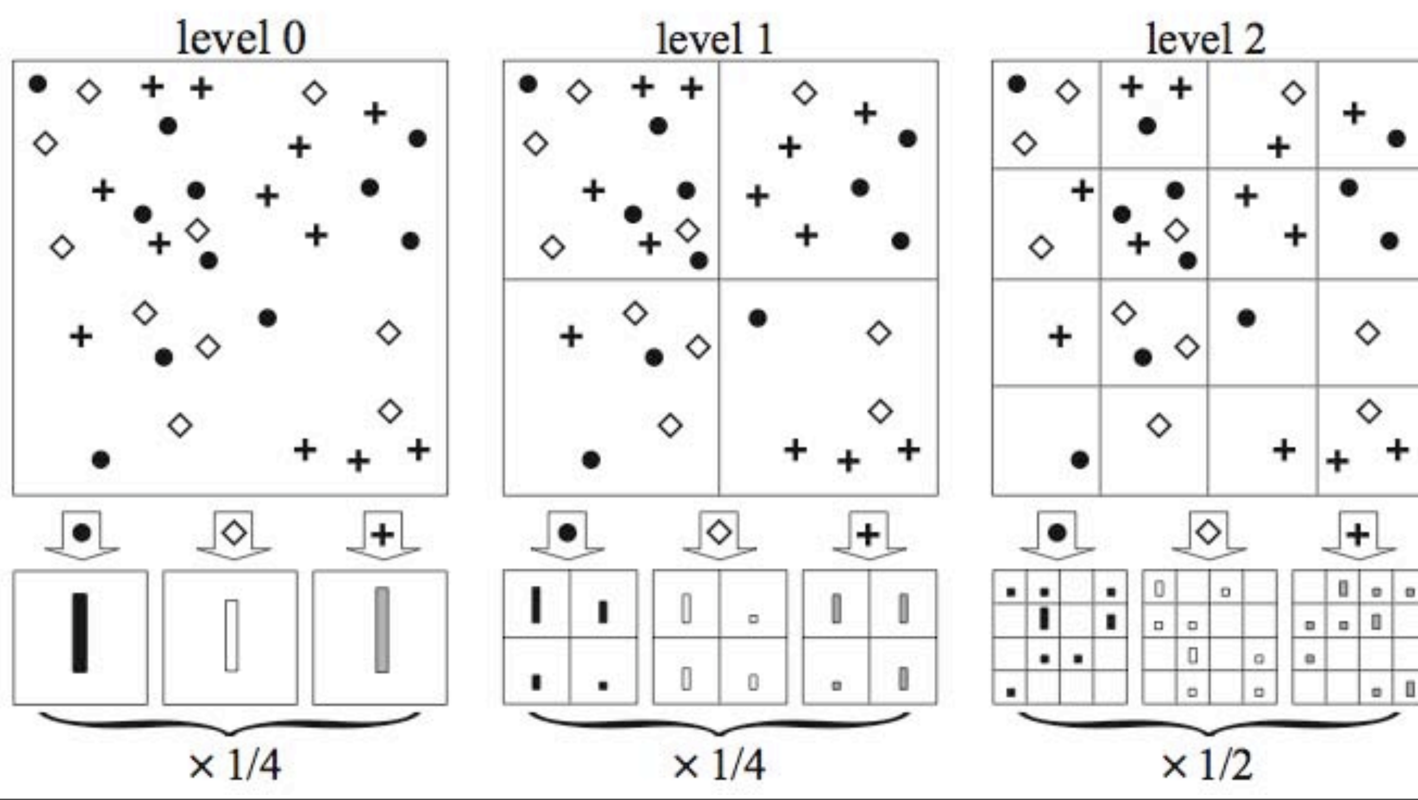
Extension example: Spatial Pyramid Matching using SIFT

Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories

Svetlana Lazebnik[1]
slazebni@uiuc.edu
[1]*Beckman Institute*
University of Illinois

Cordelia Schmid[2]
Cordelia.Schmid@inrialpes.fr
[2]*INRIA Rhône-Alpes*
Montbonnot, France

Jean Ponce[1,3]
ponce@cs.uiuc.edu
[3]*Ecole Normale Supérieure*
Paris, France



CVPR 2006

Lowe
(1999)

Object Recognition from Local Scale-Invariant Features

David G. Lowe
Computer Science Department
University of British Columbia
Vancouver, B.C., V6T 1Z4, Canada
lowe@cs.ubc.ca

Abstract

An object recognition system has been developed that uses a new class of local image features. The features are invariant

translation, scaling, and rotation, and partially invariant to illumination changes and affine or 3D projection. Previous approaches to local feature generation lacked invariance to scale and were more sensitive to projective distortion and

Nalal and Triggs
(2005)

Histograms of Oriented Gradients for Human Detection

Navneet Dalal and Bill Triggs
INRIA Rhône-Alps, 655 avenue de l'Europe, Montbonnot 38334, France
{Navneet.Dalal,Bill.Triggs}@inrialpes.fr,   http://lear.inrialpes.fr

Abstract

We study the question of feature sets for robust visual object recognition, adopting linear SVM based human detection as a test case. After reviewing existing edge and gradient based descriptors, we show experimentally that grids of Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for human detec-

We briefly discuss previous work on human detection in §2, give an overview of our method §3, describe our data sets in §4 and give a detailed description and experimental evaluation of each stage of the process in §5–6. The main conclusions are summarized in §7.

2 Previous Work

Felzenszwalb et al.
(2008)

A Discriminatively Trained, Multiscale, Deformable Part Model

Pedro Felzenszwalb
University of Chicago
pff@cs.uchicago.edu

David McAllester
Toyota Technological Institute at Chicago
mcallester@tti-c.org

Deva Ramanan
UC Irvine
dramanan@ics.uci.edu

Abstract

This paper describes a discriminatively trained, multiscale, deformable part model for object detection. Our system achieves a two-fold improvement in average precision over the best performance in the 2006 PASCAL person detection challenge. It also outperforms the best results in the 2007 challenge in ten out of twenty categories. The system relies heavily on deformable parts. While deformable part models have become quite popular, their value had not been

Figure 1. Example detection obtained with the person model. The

Histograms of Oriented Gradients for Human Detection

Navneet Dalal and Bill Triggs
INRIA Rhône-Alpes, 655 avenue de l'Europe, Montbonnot 38334, France
{Navneet.Dalal,Bill.Triggs}@inrialpes.fr, http://lear.inrialpes.fr

Abstract

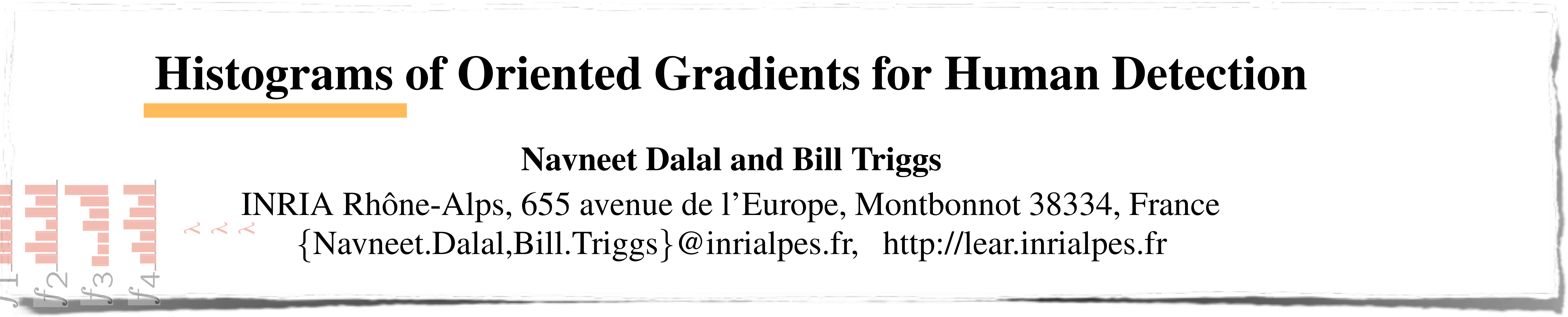
We study the question of feature sets for robust visual object recognition, adopting linear SVM based human detection as a test case. After reviewing existing edge and gradient based descriptors, we show experimentally that grids of Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for human detection. We study the influence of each stage of the computation on performance, concluding that fine-scale gradients, fine orientation binning, relatively coarse spatial binning, and high-quality local contrast normalization in overlapping de-

We briefly discuss previous work on human detection in §2, give an overview of our method §3, describe our data sets in §4 and give a detailed description and experimental evaluation of each stage of the process in §5–6. The main conclusions are summarized in §7.

2 Previous Work

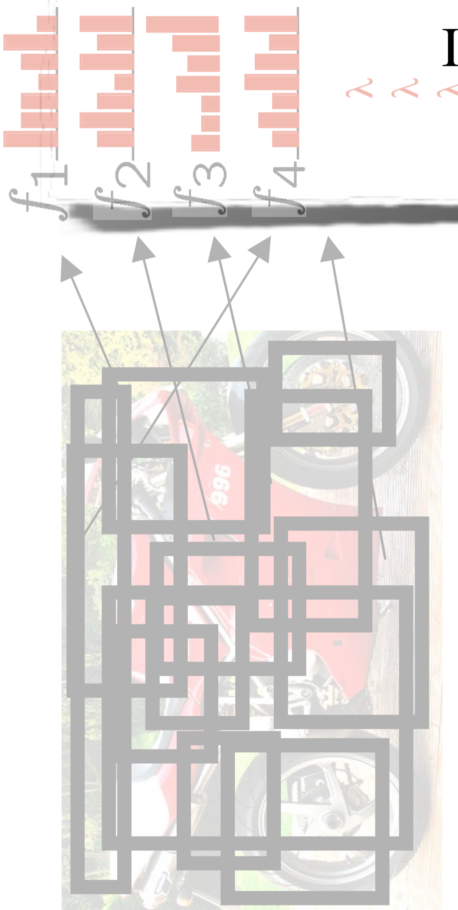
There is an extensive literature on object detection, but here we mention just a few relevant papers on human detection [18, 17, 22, 16, 20]. See [6] for a survey. Papageorgiou *et al* [18] describe a pedestrian detector based on a polynomial

first of all, let me put this paper in context



Histograms of Oriented Gradients for Human Detection

Navneet Dalal and Bill Triggs
INRIA Rhône-Alpes, 655 avenue de l'Europe, Montbonnot 38334, France
{Navneet.Dalal,Bill.Triggs}@inrialpes.fr, http://lear.inrialpes.fr



Swain & Ballard 1991 - Color Histograms

Schiele & Crowley 1996 - Receptive Fields Histograms

Lowe 1999 - SIFT

Schneiderman & Kanade 2000 - Localized Histograms of Wavelets

Leung & Malik 2001 - Texton Histograms

Belongie et al. 2002 - Shape Context

Dalal & Triggs 2005 - Dense Orientation Histograms

...

histograms of local image measurement have been quite successful

Histograms of Oriented Gradients for Human Detection

features

Navneet Dalal and Bill Triggs
INRIA Rhône-Alps, 655 avenue de l'Europe, Montbonnot 38334, France
{Navneet.Dalal,Bill.Triggs}@inrialpes.fr, http://lear.inrialpes.fr

Gravrila & Philomen 1999 - **Edge Templates + Nearest Neighbor**

Papageorgiou & Poggio 2000, Mohan et al. 2001, DePoortere et al. 2002 - **Haar Wavelets + SVM**

Viola & Jones 2001 - **Rectangular Differential Features + AdaBoost**

Mikolajczyk et al. 2004 - **Parts Based Histograms + AdaBoost**

Ke & Sukthankar 2004 - **PCA-SIFT**

...

tons of "feature sets" have been proposed

Histograms of Oriented Gradients for Human Detection

Navneet Dalal and Bill Triggs
INRIA Rhône-Alps, 655 avenue de l'Europe, Montbonnot 38334, France
{Navneet.Dalal,Bill.Triggs}@inrialpes.fr, http://lear.inrialpes.fr

difficult!

Wide variety of **articulated poses**

Variable **appearance/clothing**

Complex **backgrounds**

Unconstrained **illuminations**

Occlusions

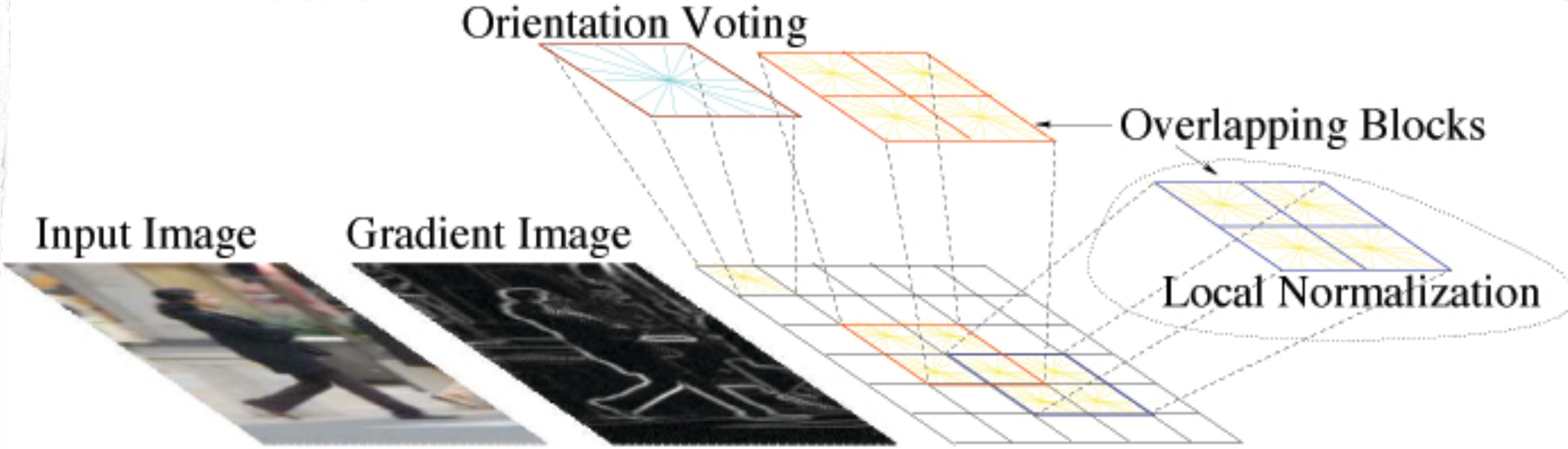
Different **scales**

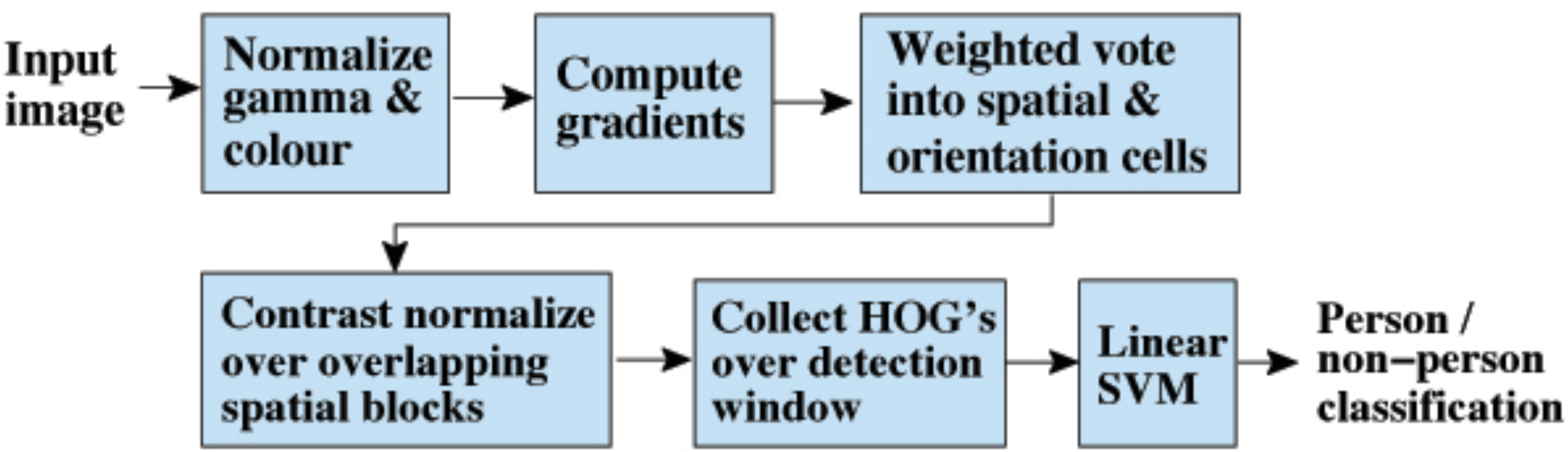
...

localizing humans in images is a challenging task...

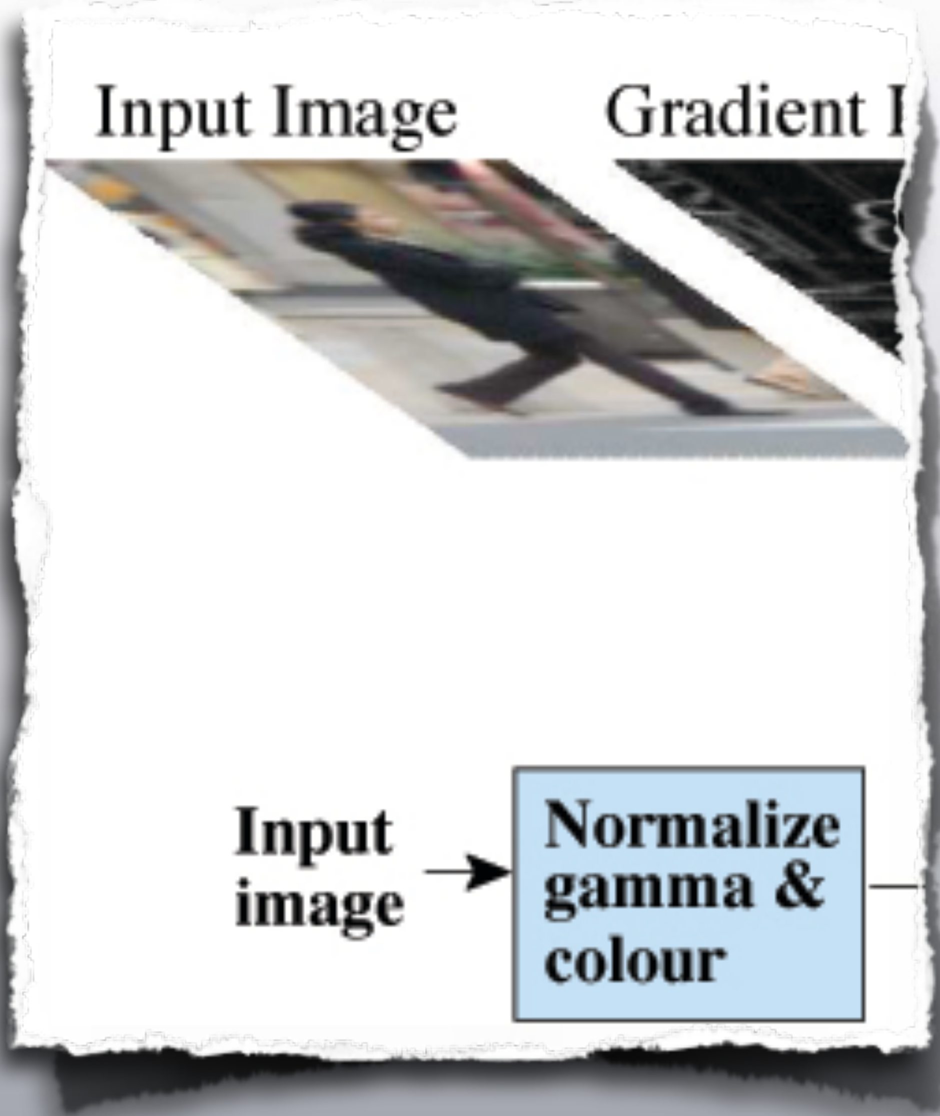
Approach

- robust feature set (HOG)
- simple classifier (linear SVM)
- fast detection (sliding window)

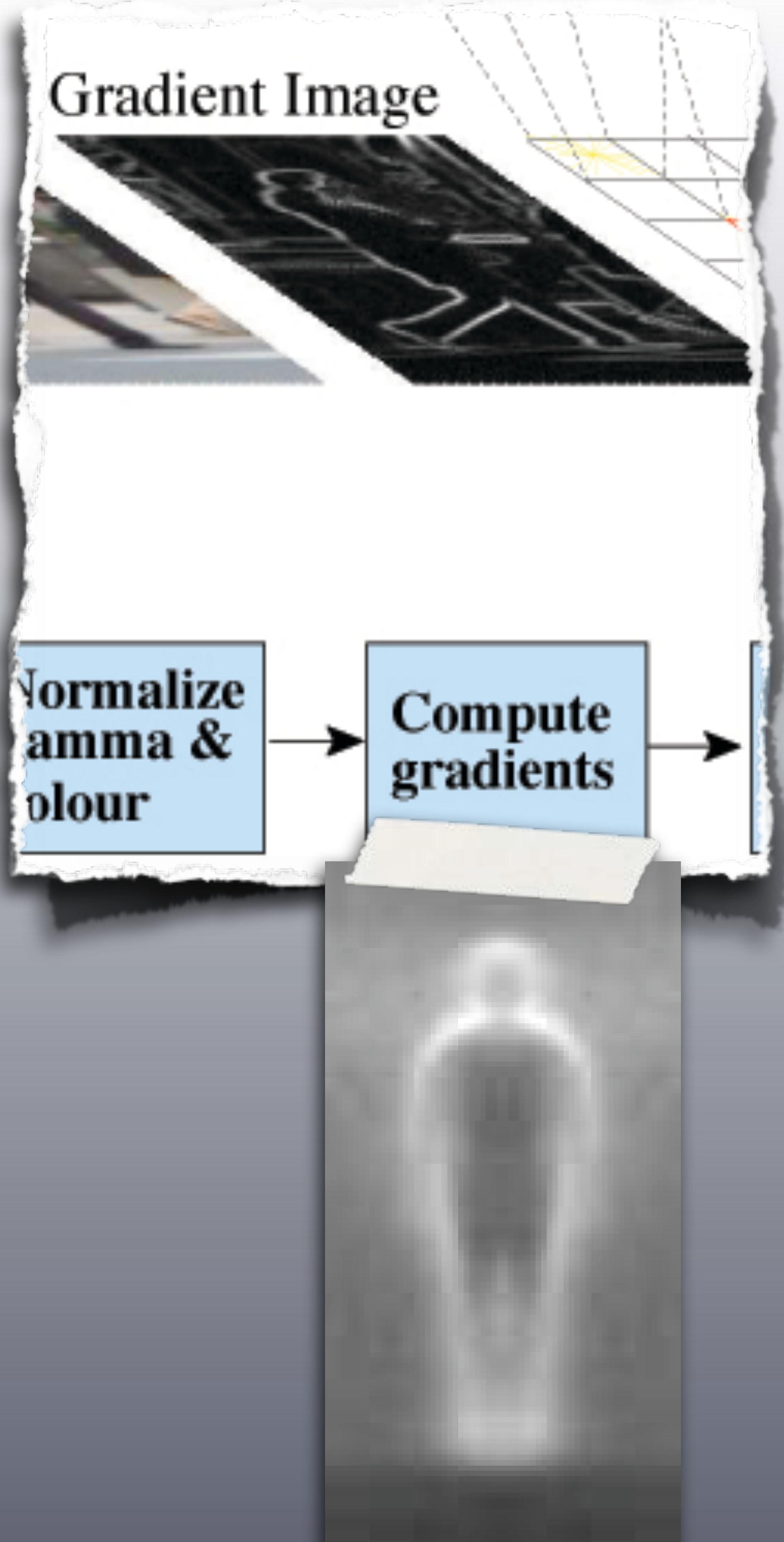




adapted from Bill Triggs



- Gamma normalization
- Space: RGB, LAB or Gray
- Method: SQRT or LOG



- Filtering with **simple masks**

-1	0	1

centered *

-1	1

uncentered

1	-8	0	8	-1

cubic-corrected

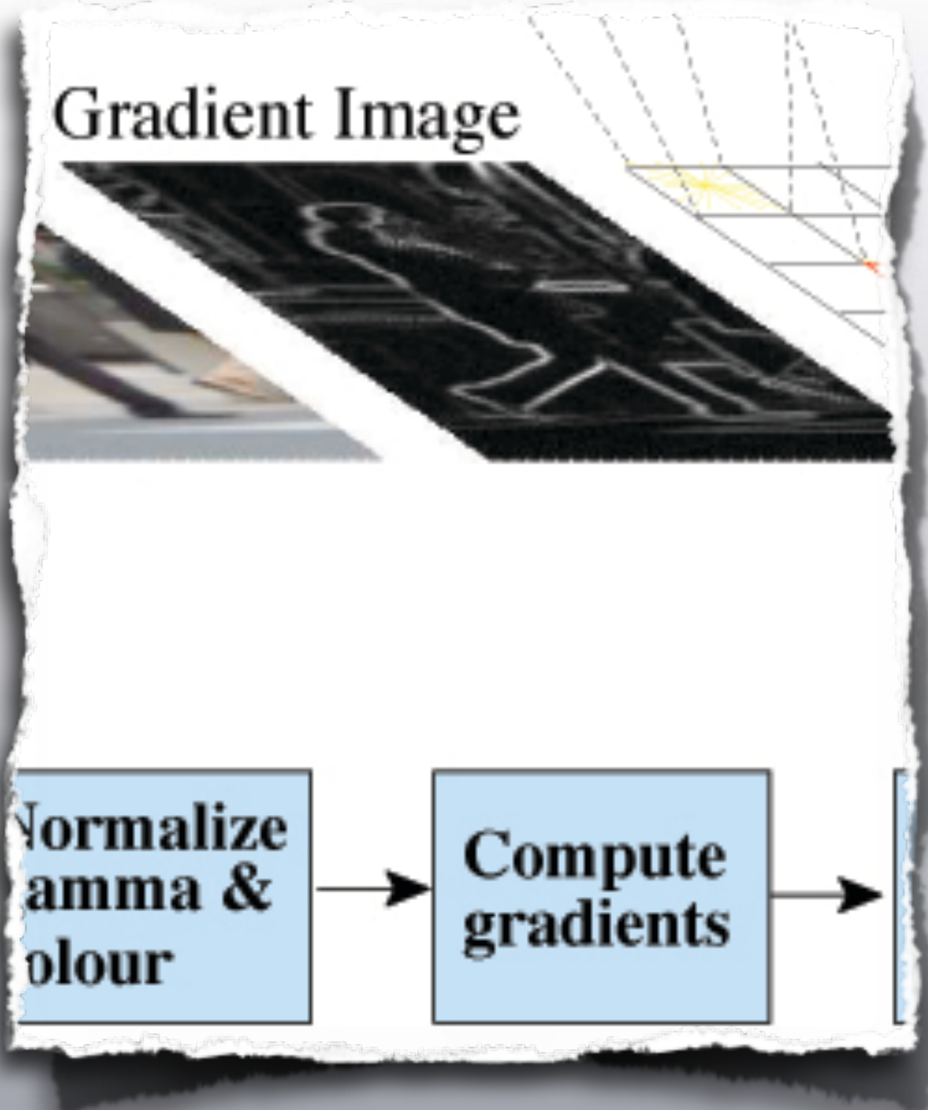
0	1
-1	0

diagonal

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

Sobel

**** centered performs the best***



remember SIFT ?

- **Filtering with simple masks**

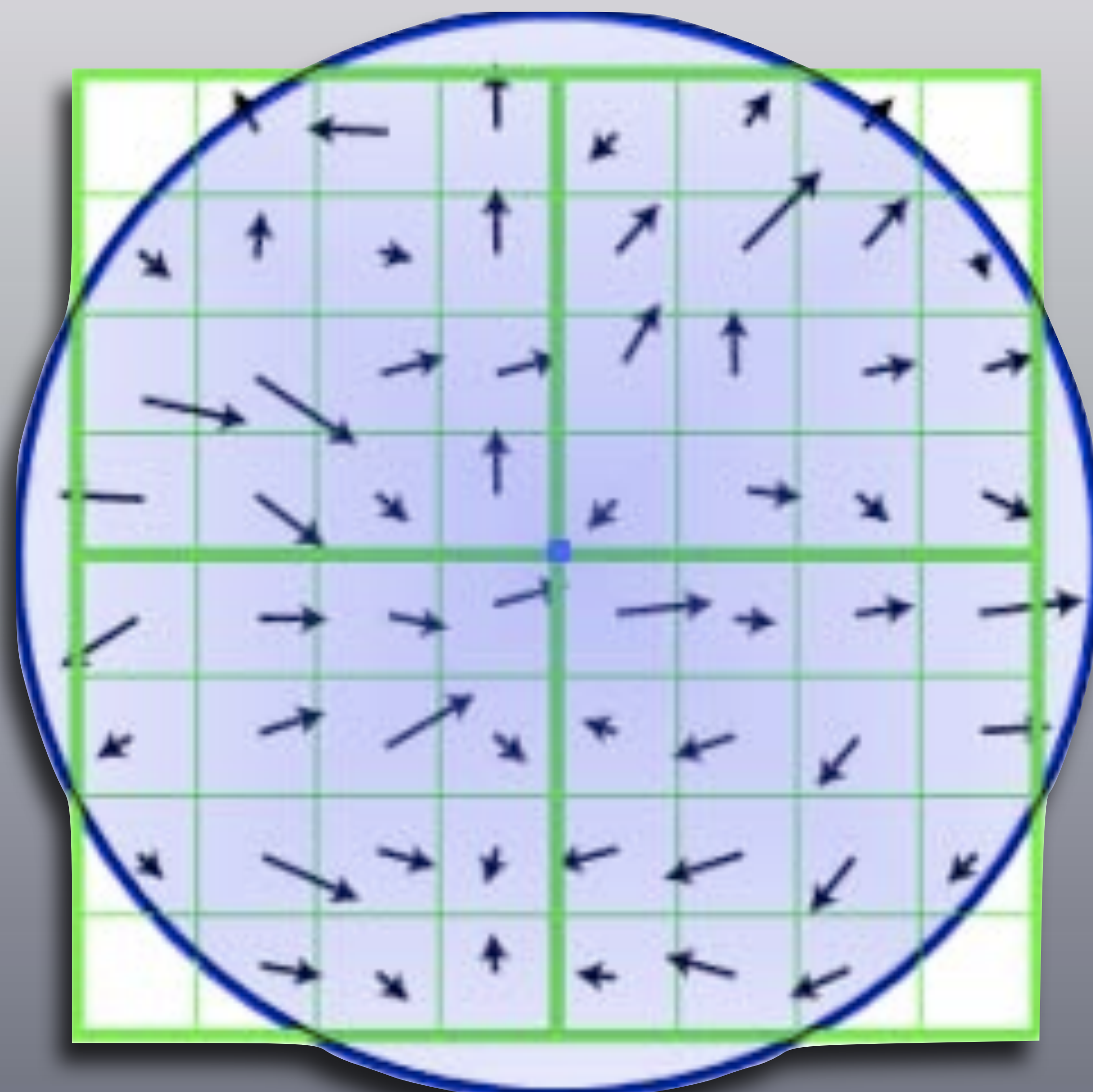
-1	0	1

centered

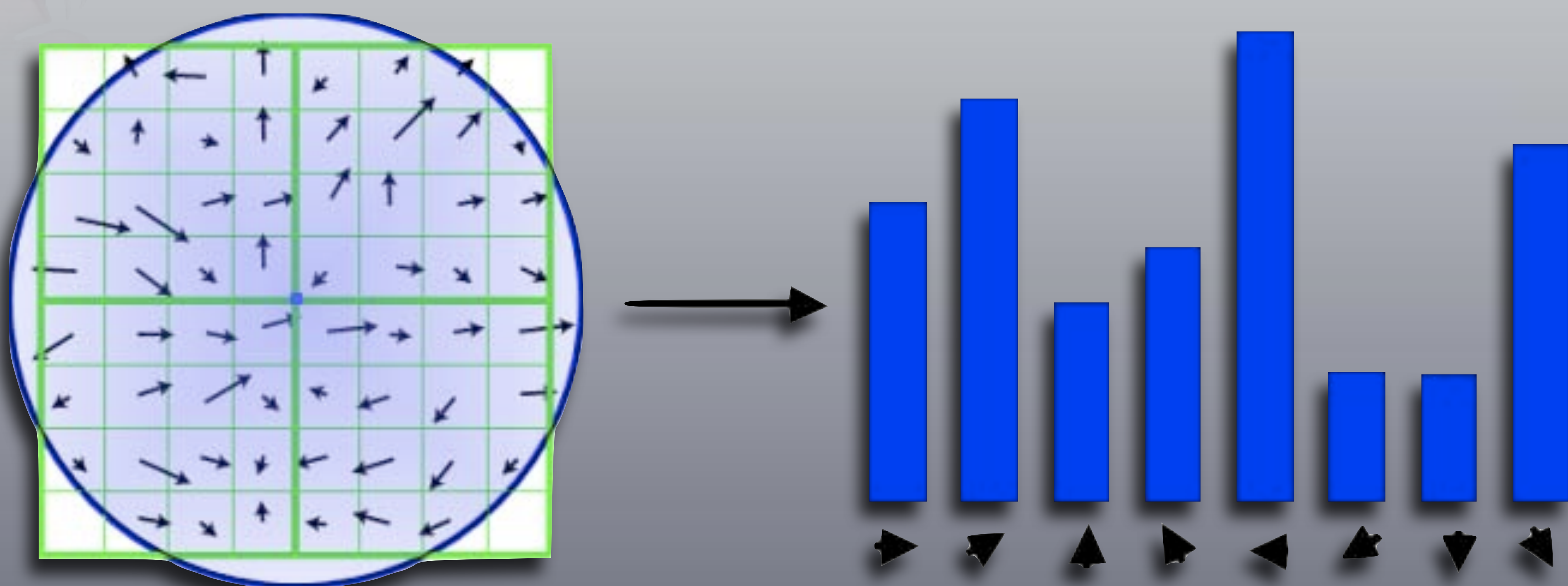
$$m(x,y) = \sqrt{\left(L(x+1,y) - L(x-1,y)\right)^2 + \left(L(x,y+1) - L(x,y-1)\right)^2}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right)$$

...after filtering, each "pixel" represents an **oriented gradient**...

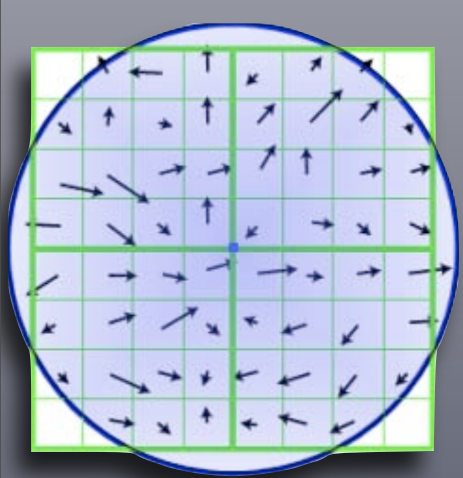


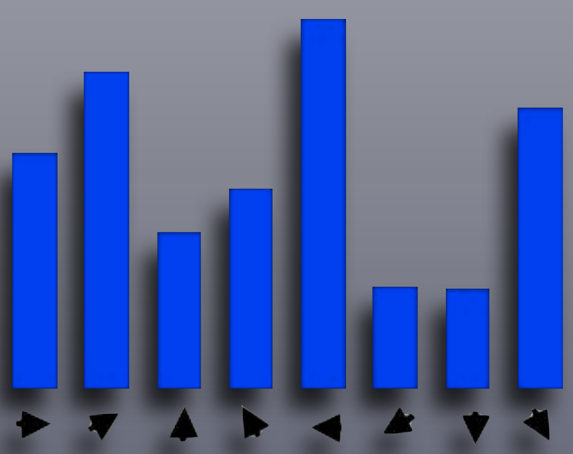
...pixels are regrouped in "cells", they cast a weighted vote for an **orientation histogram**...

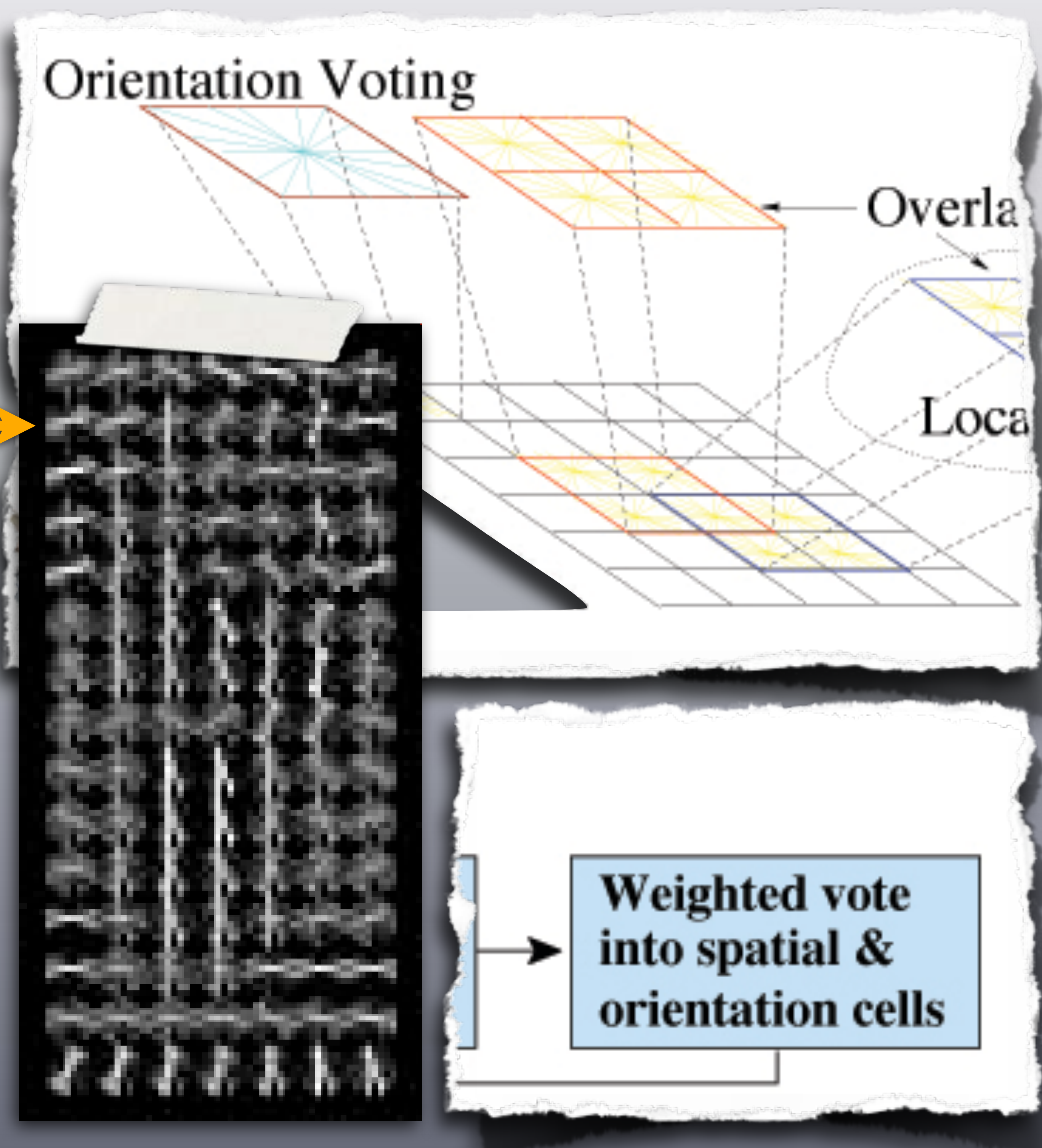


HOG (Histogram of Oriented Gradients)

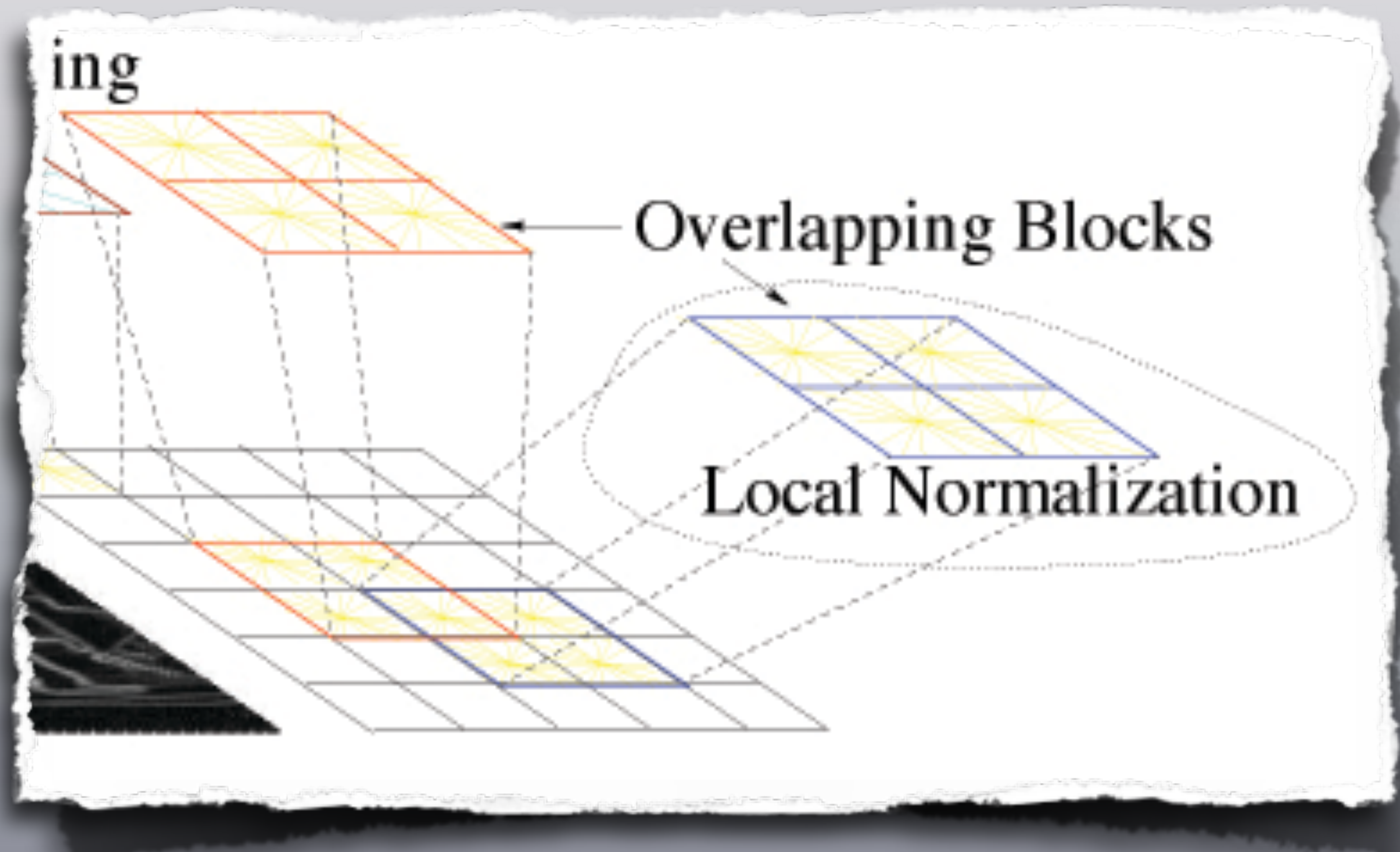
a window can be represented like that

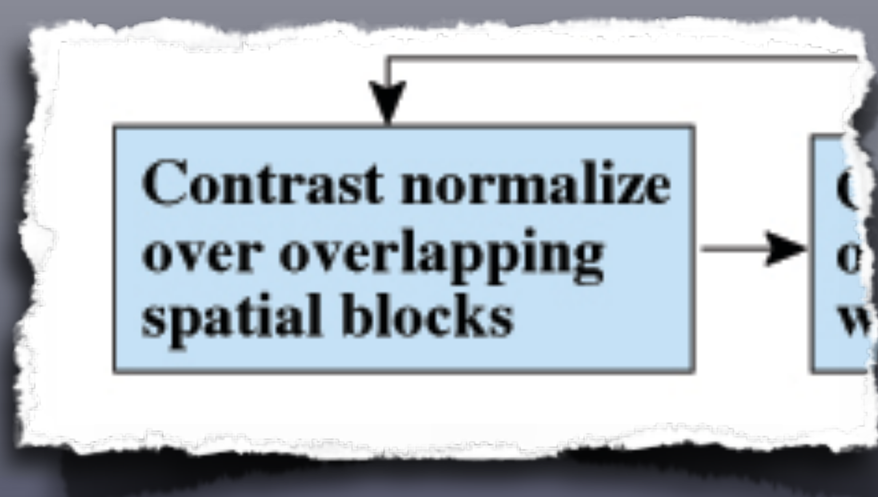




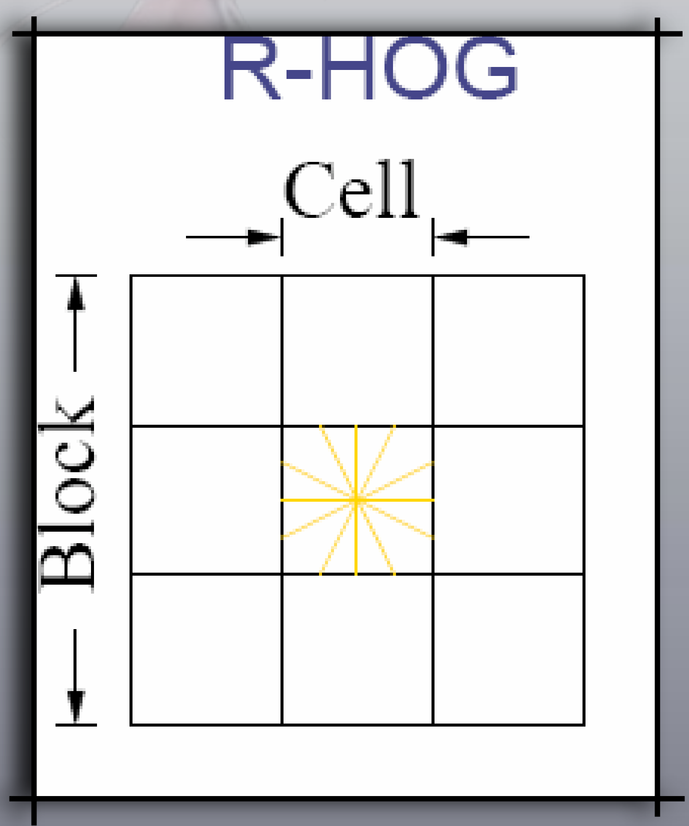


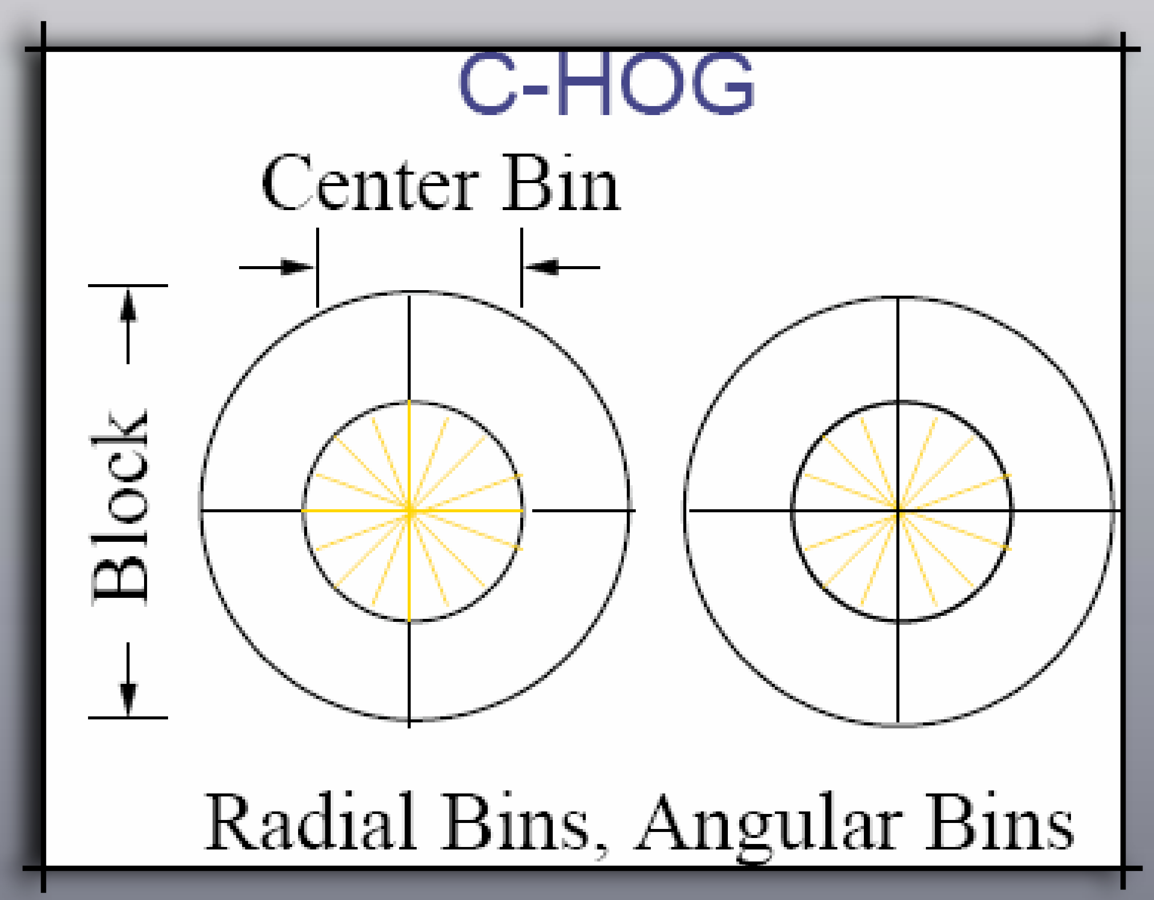
then, cells are locally **normalized** using overlapping "blocks"





they used two types of blocks





- rectangular
- similar to SIFT (but dense)

- circular
- similar to Shape Context

and four different types of block **normalization**

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

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

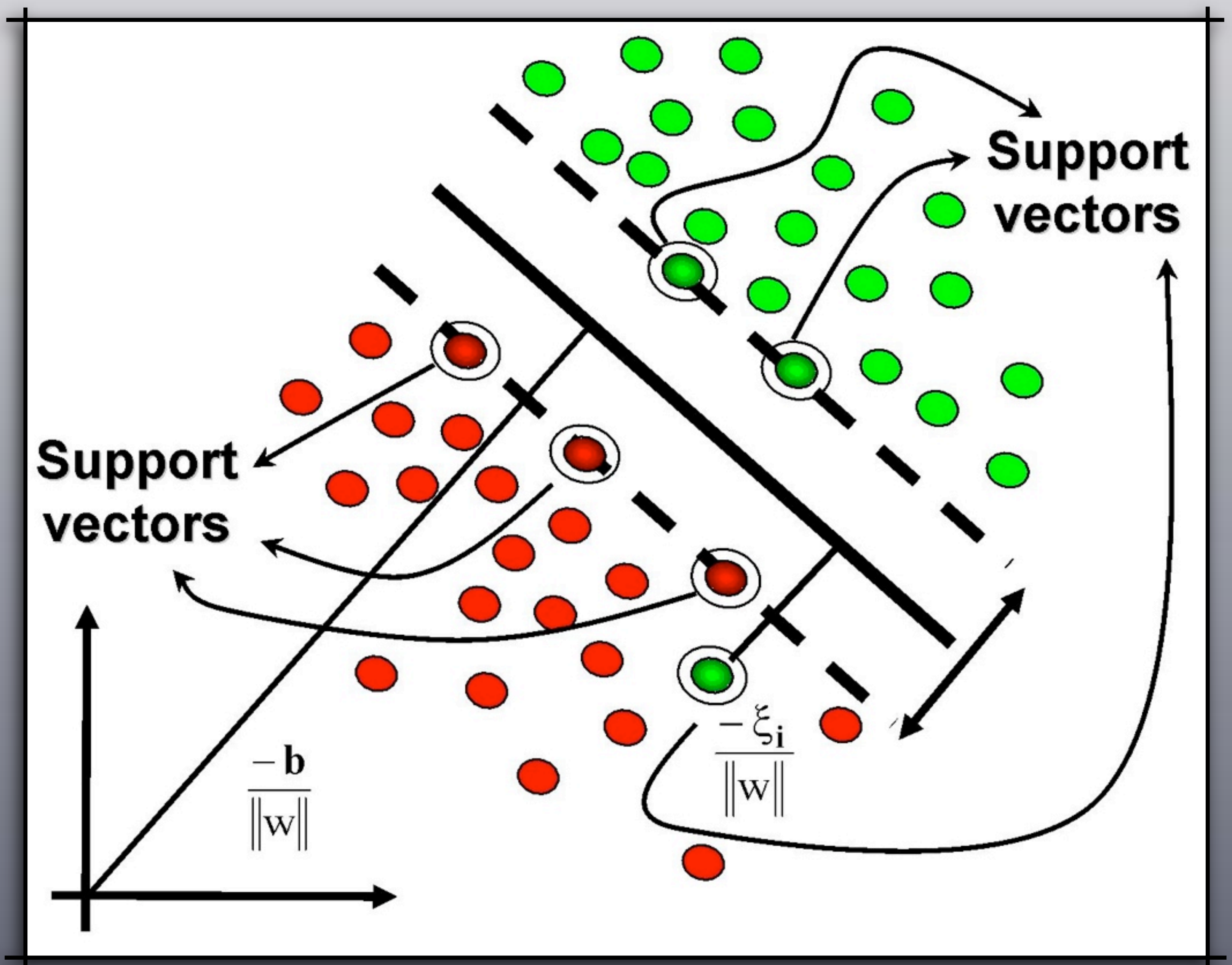
$$L1-norm : v \longrightarrow v/(||v||_1+\epsilon)$$

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

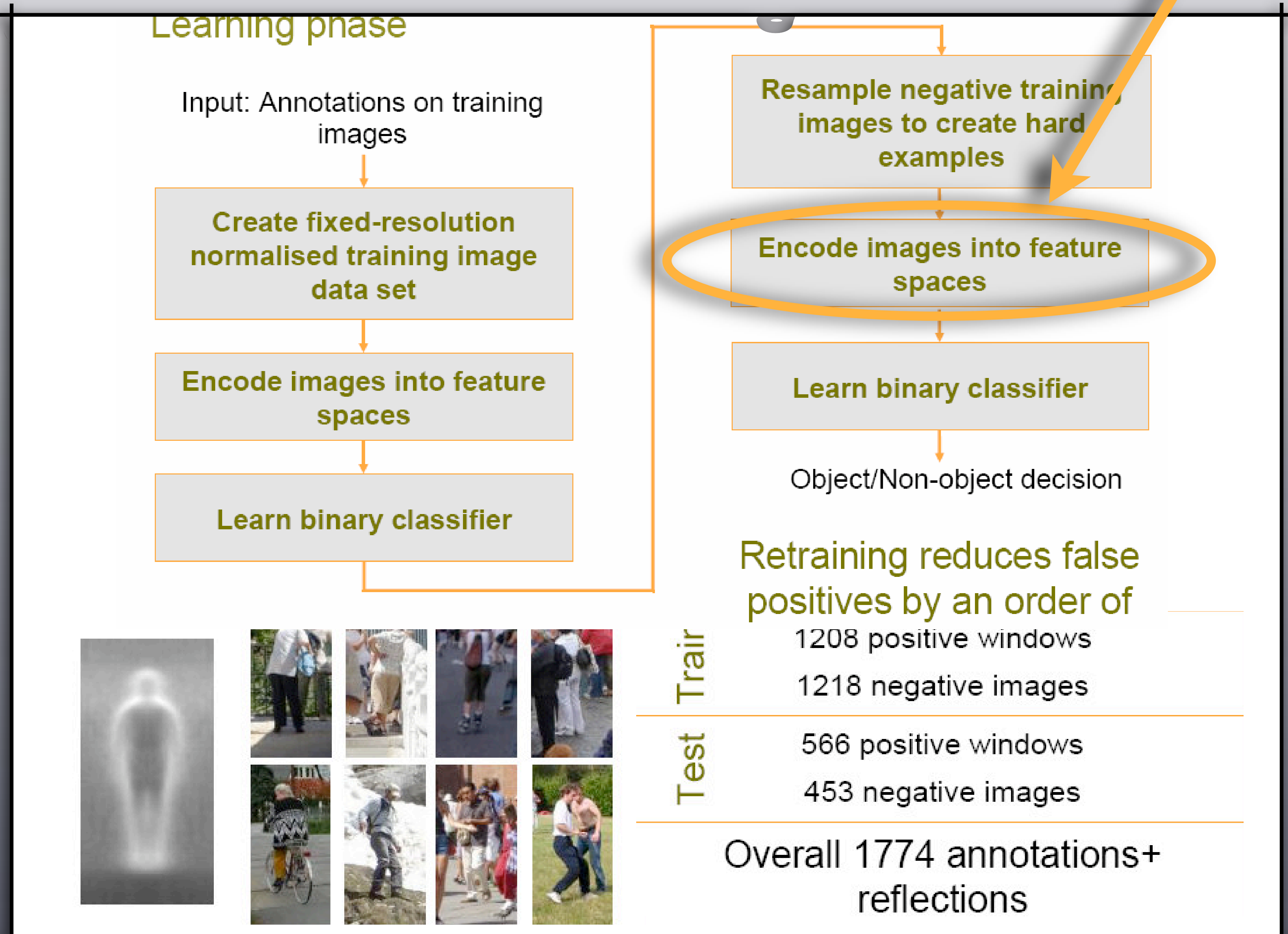
like SIFT, they gain **invariance**...

...to illuminations, small deformations, etc.

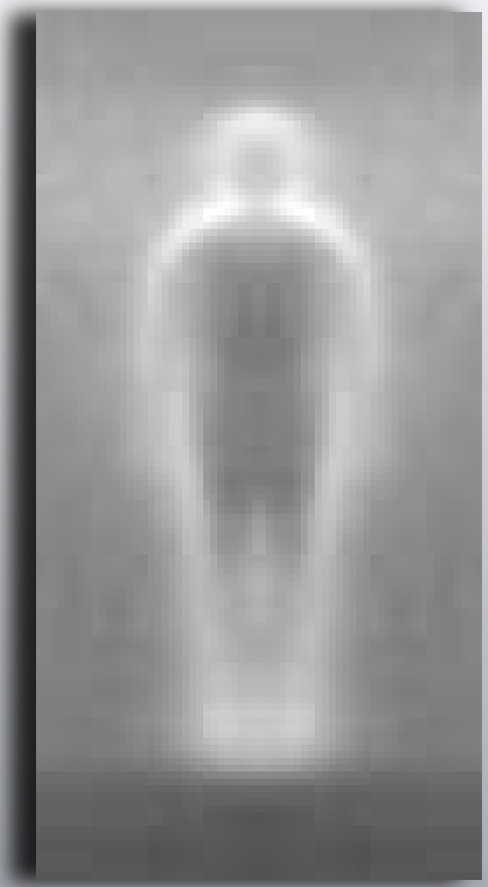
finally, a **sliding window** is classified by a simple **linear SVM**



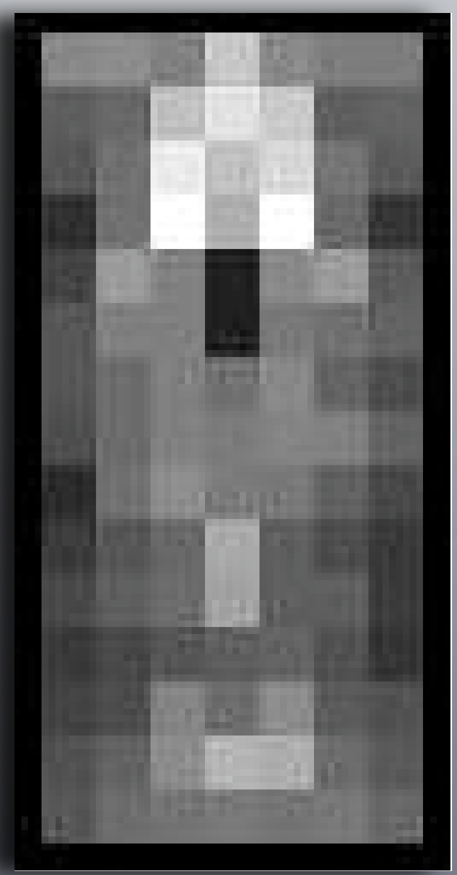
during the learning phase, the algorithm “looked” for hard examples



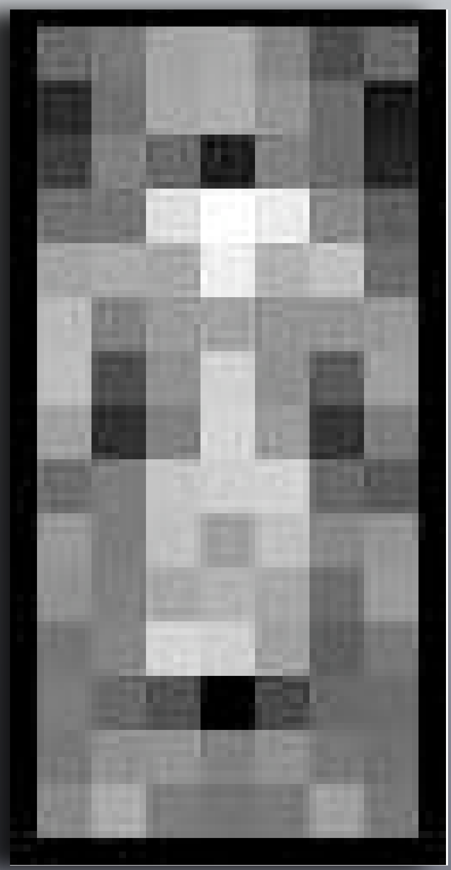
adapted from Martial Hebert



average gradients

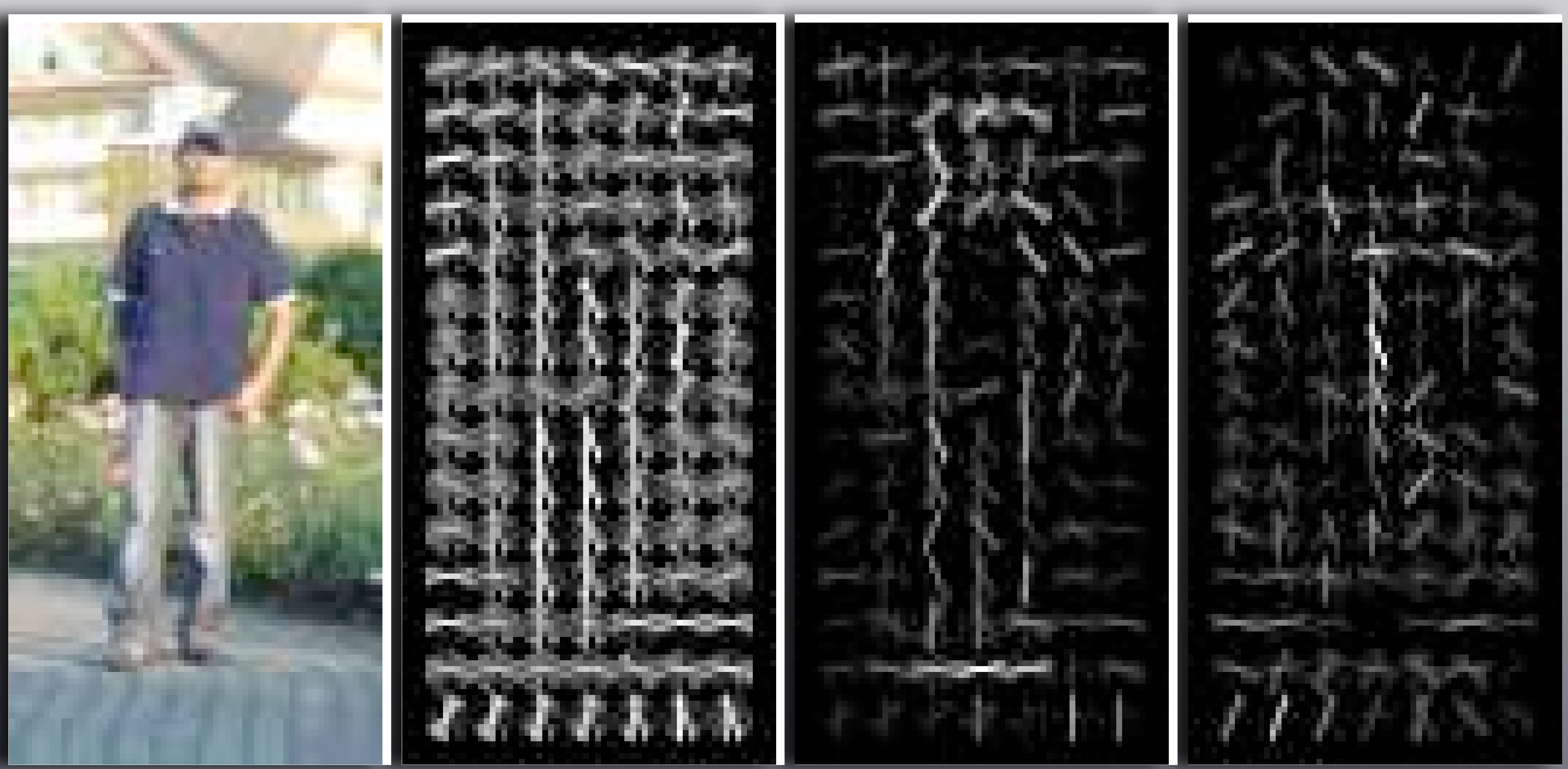


positive weights



negative weights

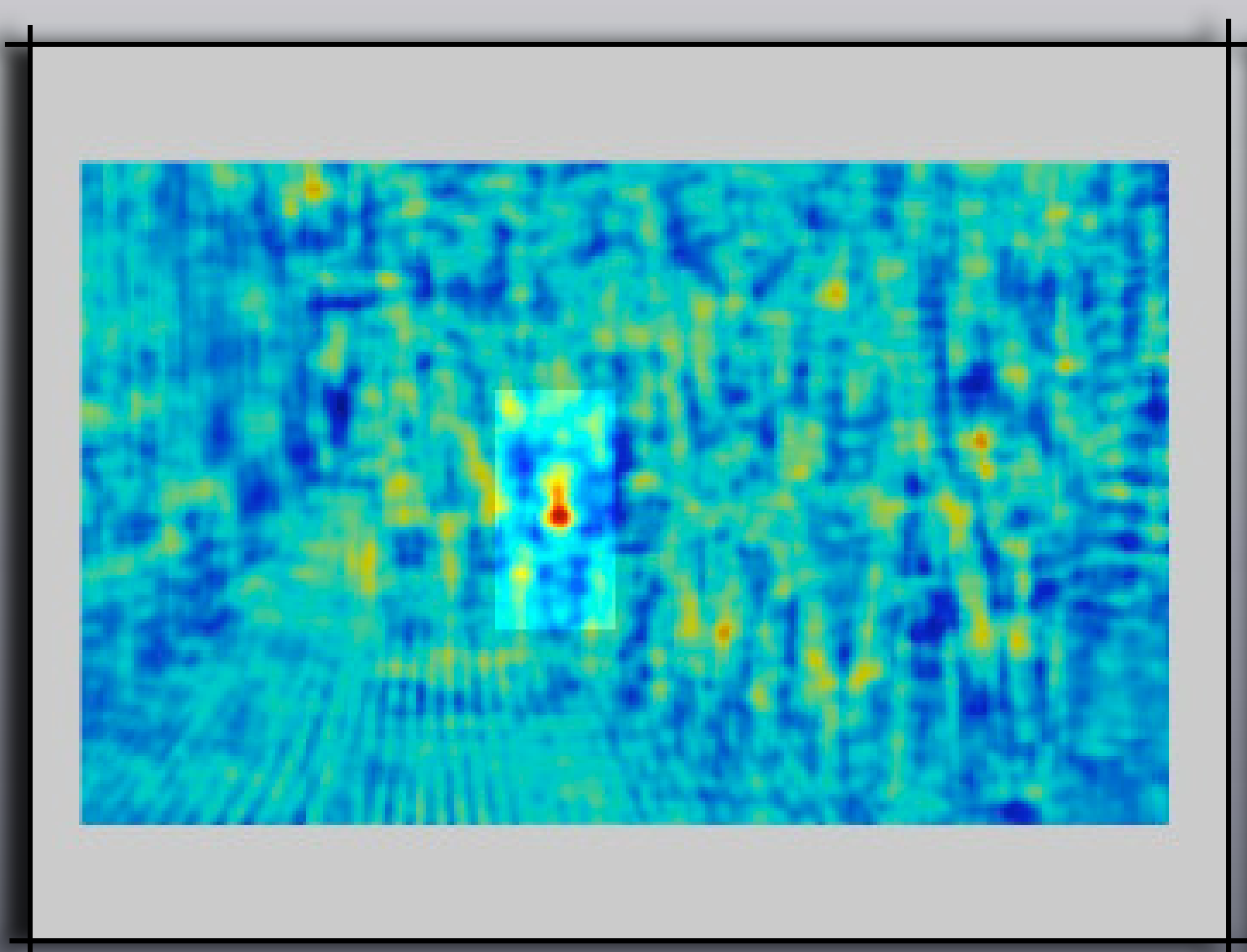
Example



Example

adapted from Bill Triggs

Example



adapted from Martial Hebert

Results

90% @ 1e-5 FPPW

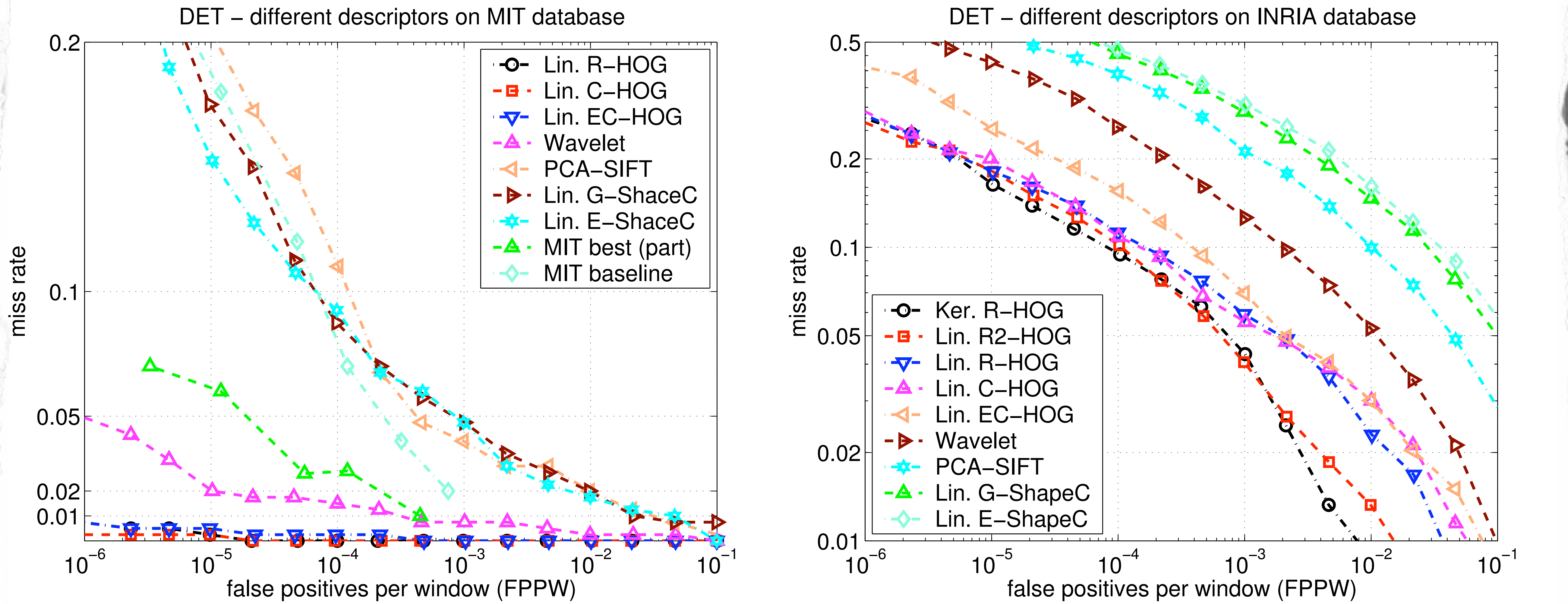


Figure 3. The performance of selected detectors on (left) MIT and (right) INRIA data sets. See the text for details.

not good

good

Experiments

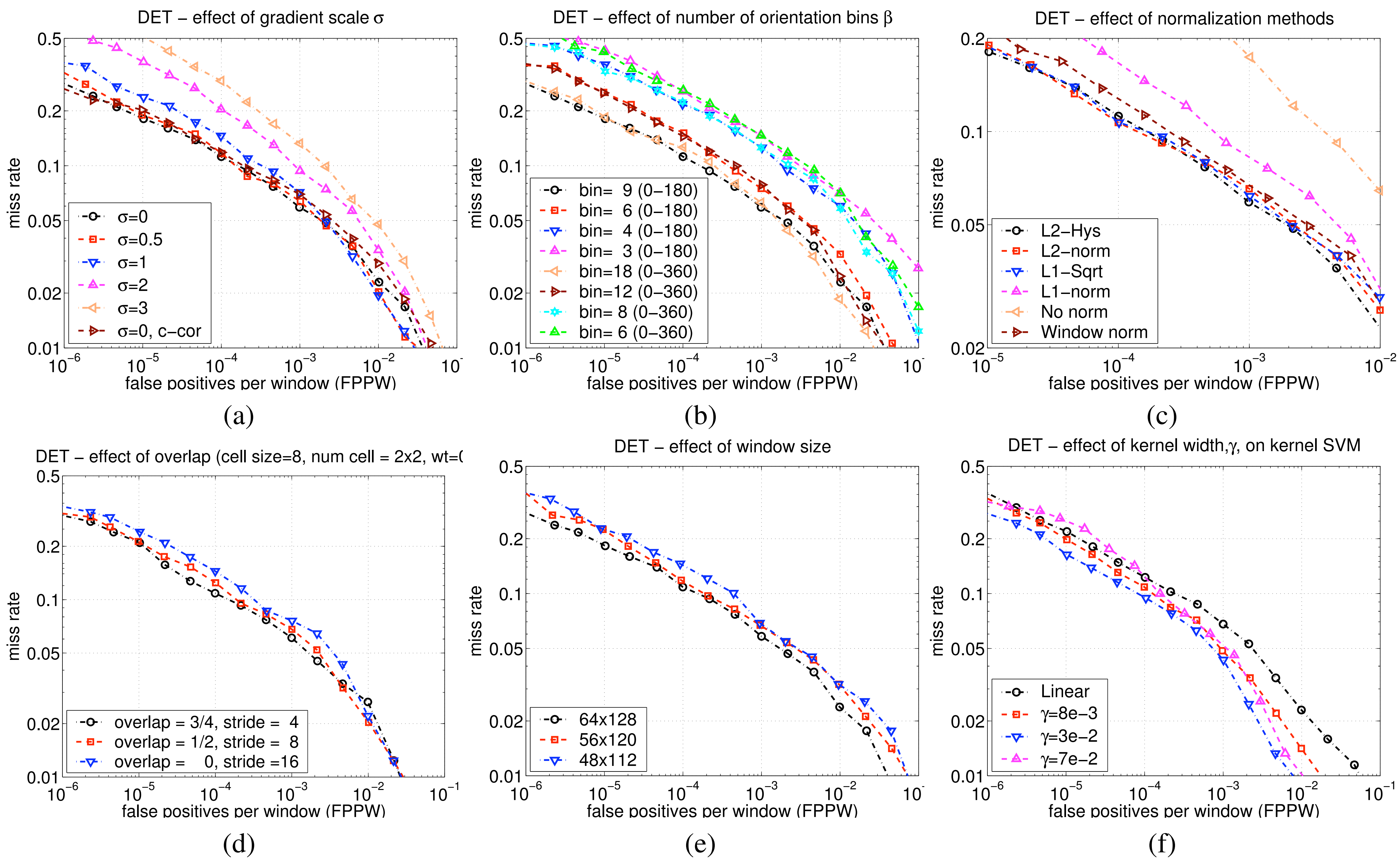


Figure 4. For details see the text. (a) Using fine derivative scale significantly increases the performance. ('c-cor' is the 1D cubic-corrected point derivative). (b) Increasing the number of orientation bins increases performance significantly up to about 9 bins spaced over 0°–180°. (c) The effect of different block normalization schemes (see §6.4). (d) Using overlapping descriptor blocks decreases the miss rate by around 5%. (e) Reducing the 16 pixel margin around the 64×128 detection window decreases the performance by about 3%. (f) Using a Gaussian kernel SVM, $\exp(-\gamma\|\mathbf{x_1}-\mathbf{x_2}\|^2)$, improves the performance by about 3%.

Experiments

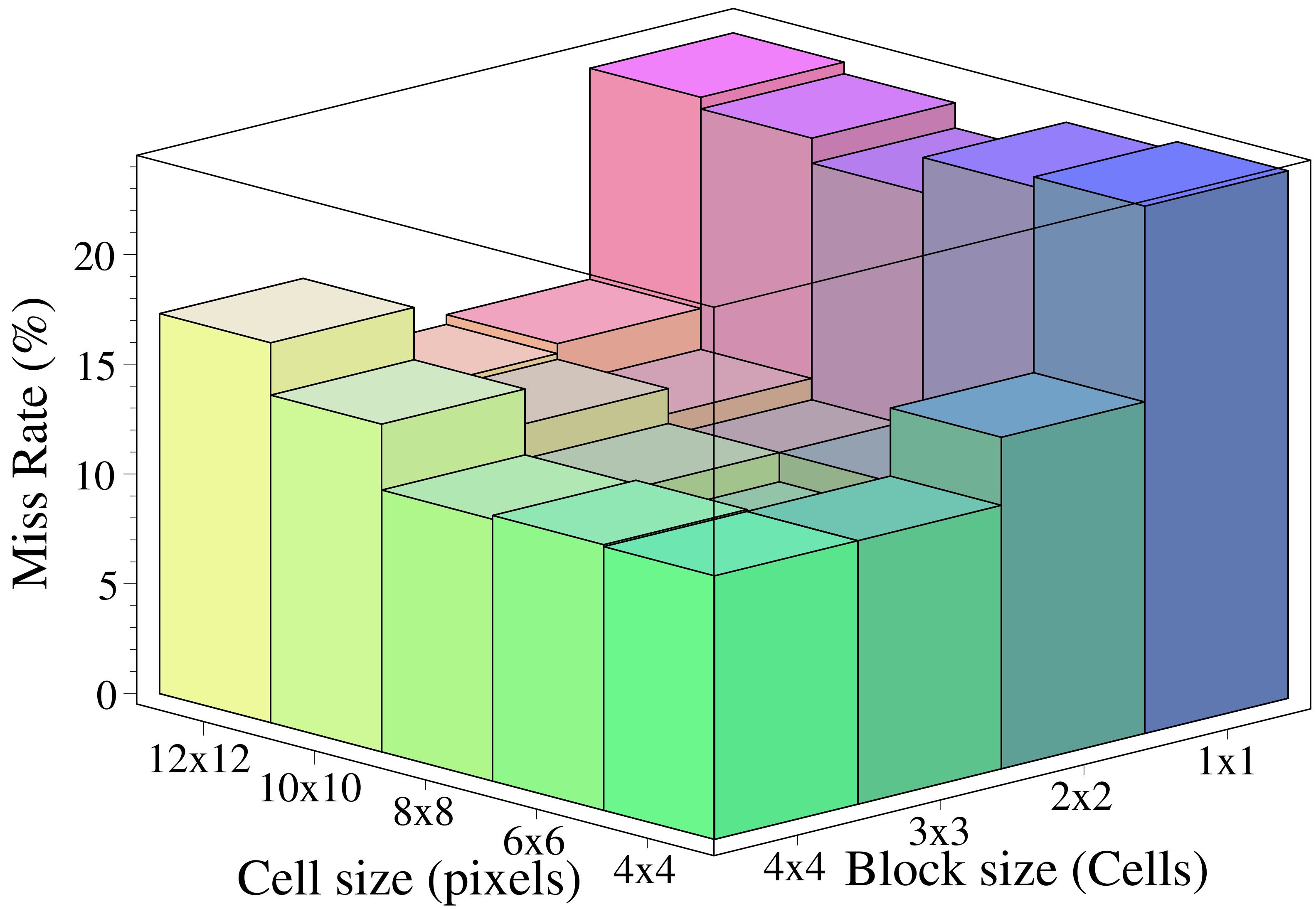


Figure 5. The miss rate at 10^{-4} FPPW as the cell and block sizes change. The stride (block overlap) is fixed at half of the block size. 3×3 blocks of 6×6 pixel cells perform best, with 10.4% miss rate.

Further Development

- Detection on Pascal VOC (2006)
- Human Detection in Movies (ECCV 2006)
- US Patent by MERL (2006)
- Stereo Vision HoG (ICVES 2008)

Extension example:
Pyramid HoG++

Representing shape with a spatial pyramid kernel

Anna Bosch
University of Girona
Computer Vision Group
17003 Girona, Spain
aboschr@eia.udg.es

Andrew Zisserman
University of Oxford
Robotics Research Group
OX1 3PJ Oxford, UK
az@robots.ox.ac.uk

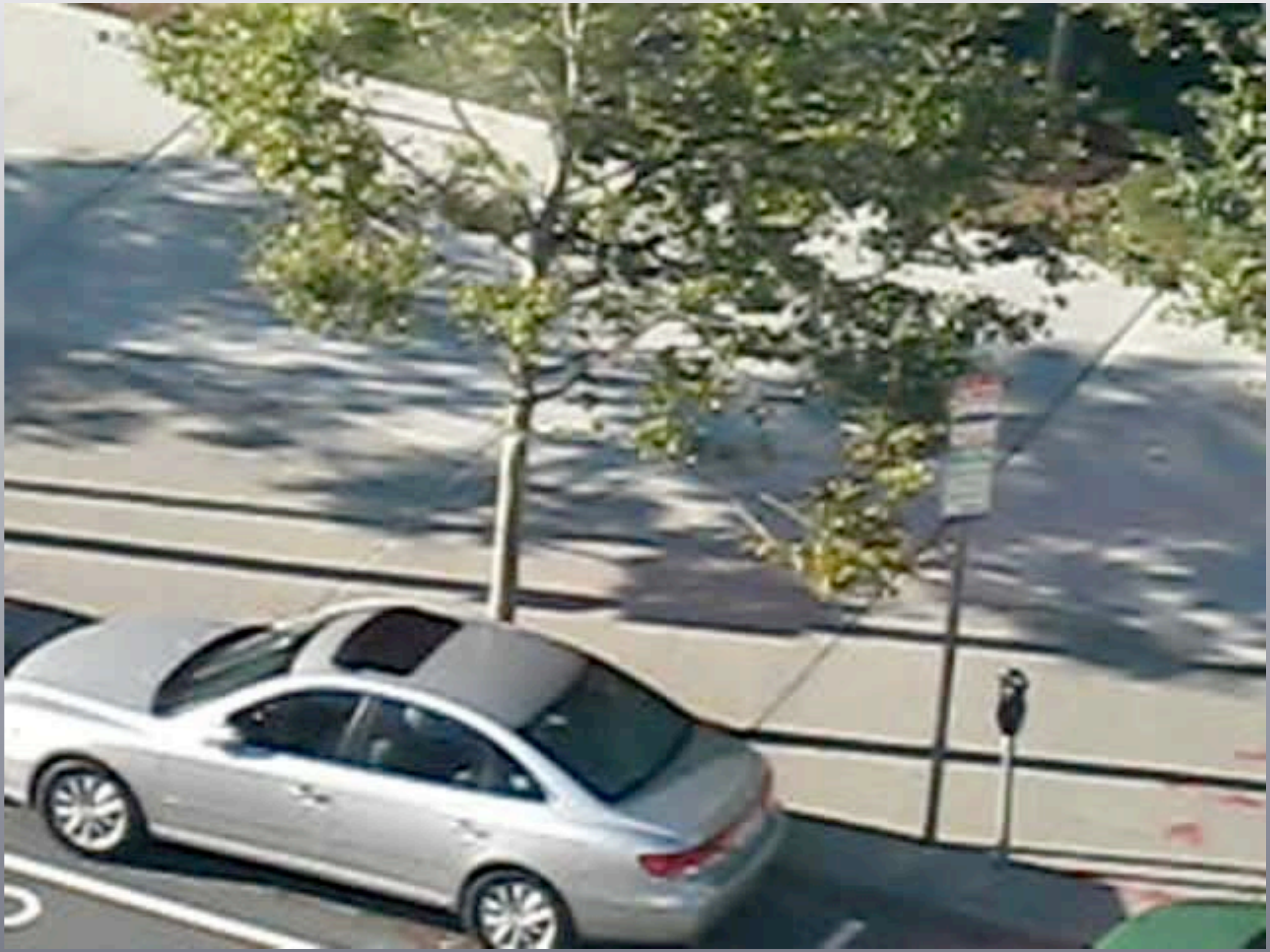
Xavier Munoz
University of Girona
Computer Vision Group
17003 Girona, Spain
xmunoz@eia.udg.es

Image classification using ROIs and Multiple Kernel Learning

Anna Bosch · Andrew Zisserman · Xavier Munoz

A simple demo...

A simple demo...



so, it doesn’t work ?!?

no no, it works...

...it just doesn’t work well...

Lowe
(1999)

Object Recognition from Local Scale-Invariant Features

David G. Lowe
Computer Science Department
University of British Columbia
Vancouver, B.C., V6T 1Z4, Canada
lowe@cs.ubc.ca

Abstract

An object recognition system has been developed that uses a new class of local image features. The features are invariant to image scaling, translation, and rotation, and partially in-

translation, scaling, and rotation, and partially invariant to illumination changes and affine or 3D projection. Previous approaches to local feature generation lacked invariance to scale and were more sensitive to projective distortion and

Nalal and Triggs
(2005)

Histograms of Oriented Gradients for Human Detection

Navneet Dalal and Bill Triggs
INRIA Rhône-Alps, 655 avenue de l'Europe, Montbonnot 38334, France
{Navneet.Dalal,Bill.Triggs}@inrialpes.fr, http://lear.inrialpes.fr

Abstract

We study the question of feature sets for robust visual object recognition, adopting linear SVM based human detection as a test case. After reviewing existing edge and gradient based descriptors, we show experimentally that grids of Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for human detec-

We briefly discuss previous work on human detection in §2, give an overview of our method §3, describe our data sets in §4 and give a detailed description and experimental evaluation of each stage of the process in §5–6. The main conclusions are summarized in §7.

2 Previous Work

Felzenszwalb et al.
(2008)

A Discriminatively Trained, Multiscale, Deformable Part Model

Pedro Felzenszwalb
University of Chicago
pff@cs.uchicago.edu

David McAllester
Toyota Technological Institute at Chicago
mcallester@tti-c.org

Deva Ramanan
UC Irvine
dramanan@ics.uci.edu

Abstract

This paper describes a discriminatively trained, multiscale, deformable part model for object detection. Our system achieves a two-fold improvement in average precision over the best performance in the 2006 PASCAL person detection challenge. It also outperforms the best results in the 2007 challenge in ten out of twenty categories. The system relies heavily on deformable parts. While deformable part models have become quite popular, their value had not been

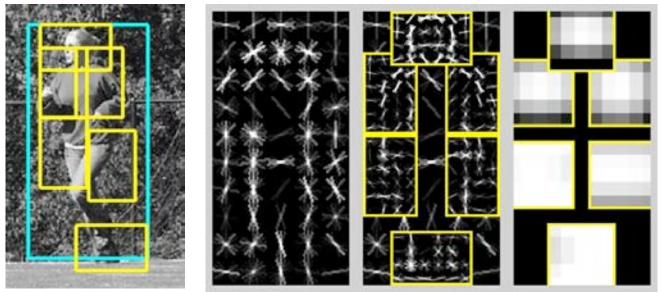
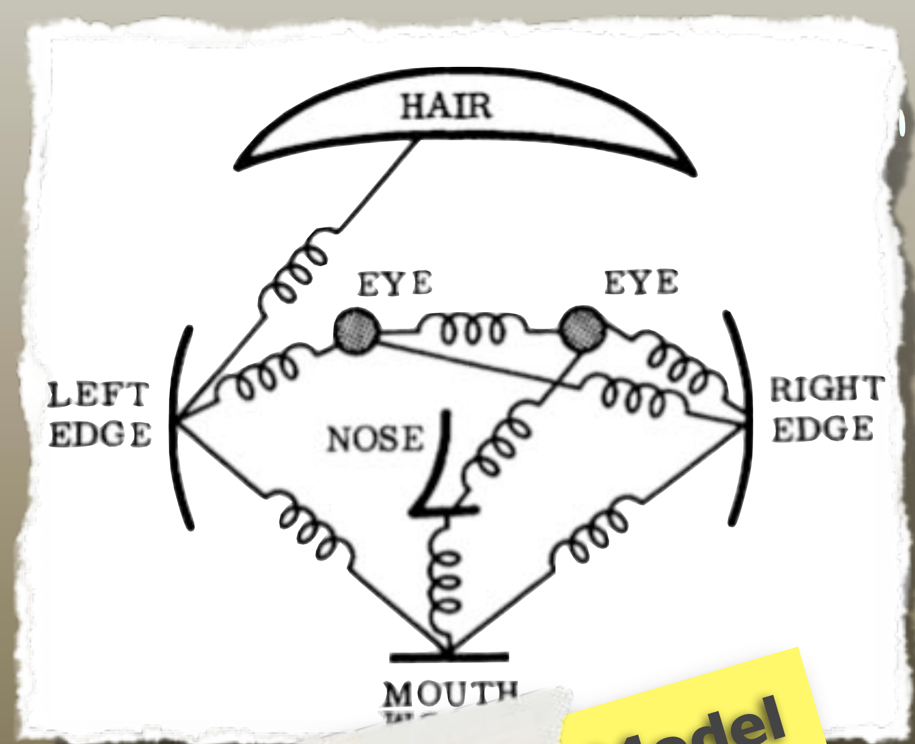


Figure 1. Example detection obtained with the person model. The

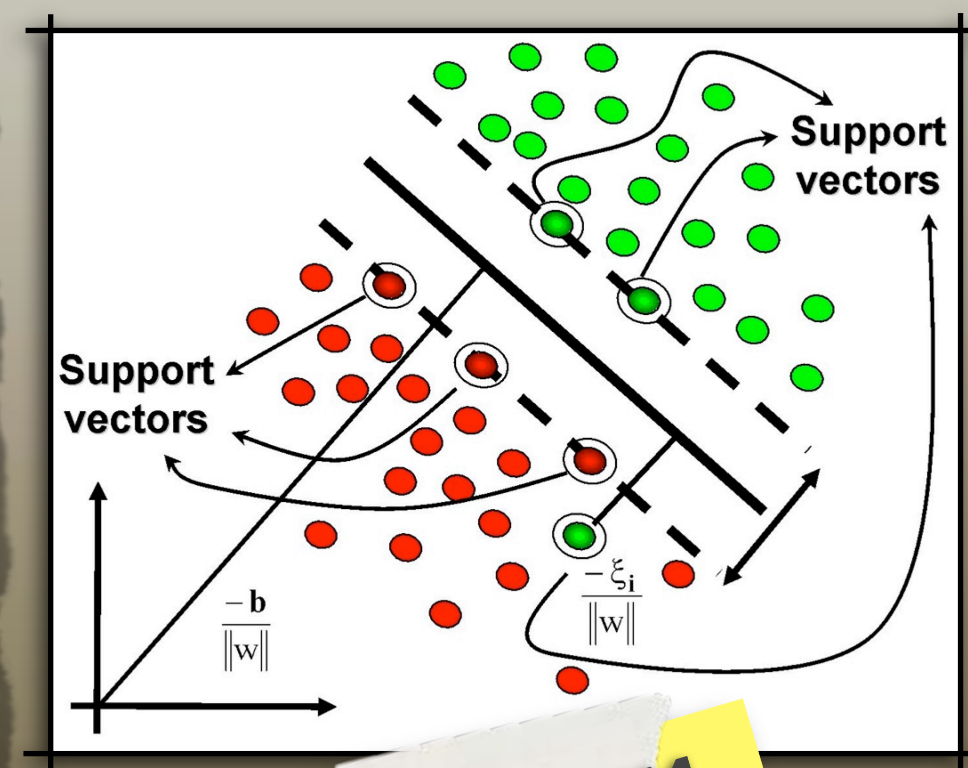
This paper describes one of the best algorithm in object detection...

They used the following methods:

HOG Features



Deformable Part Model

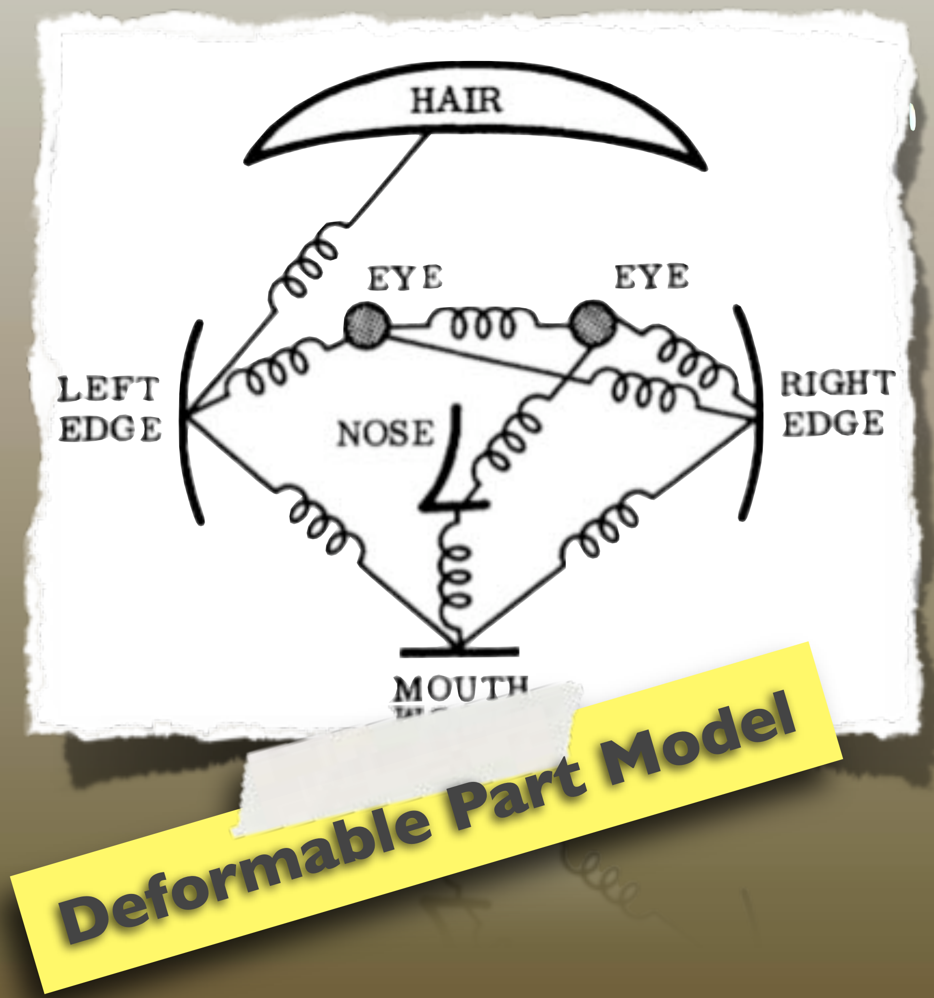


Latent SVM

They used the following methods:

Introduced by
Dalal & Triggs (2005)

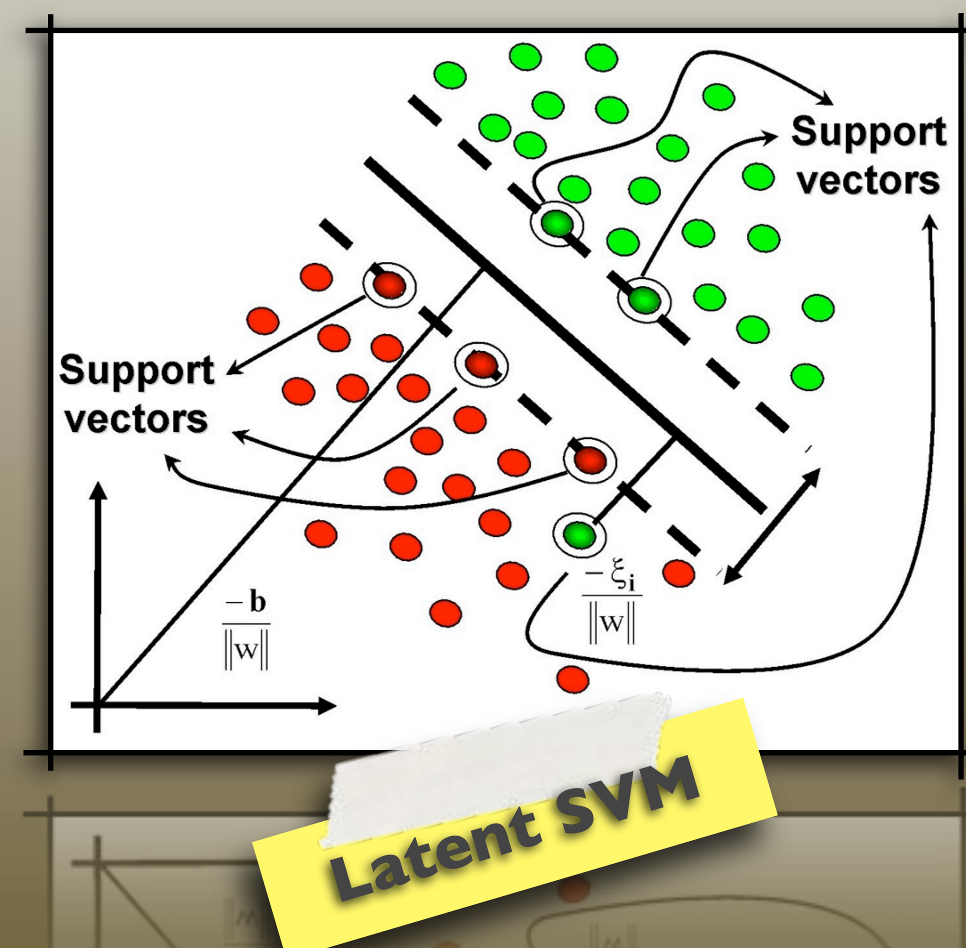
They used the following methods:



**Introduced by
Fischler & Elschlager (1973)**

They used the following methods:

Introduced by the authors




HOG Features

Model Overview

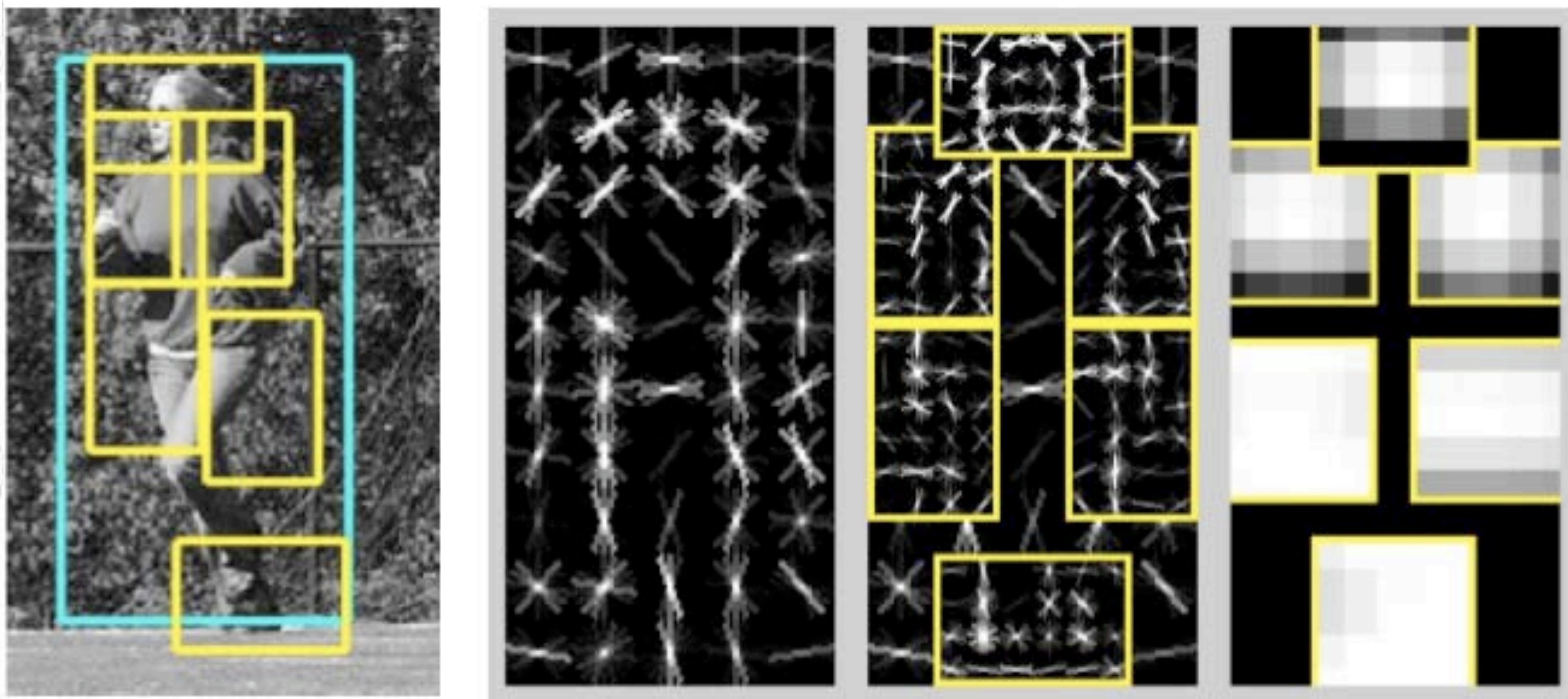
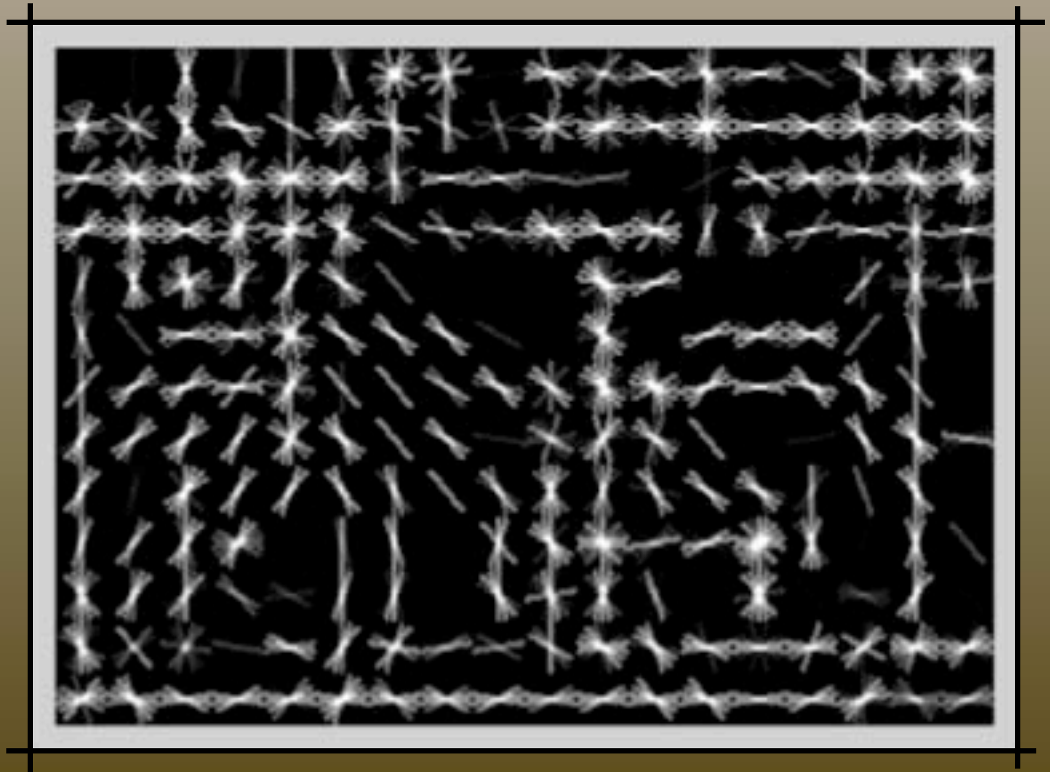


Figure 1. Example detection obtained with the person model. The model is defined by a coarse template, several higher resolution part templates and a spatial model for the location of each part.

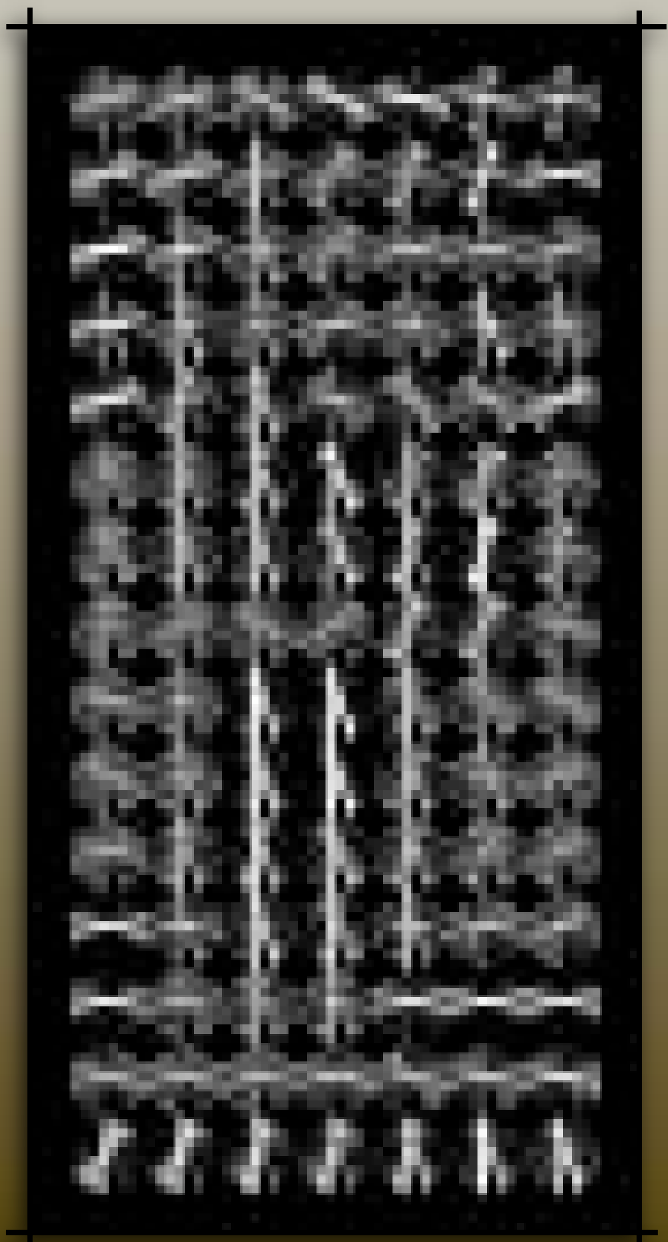
HOG Features

// 8x8 pixel blocks window

// features computed at different resolutions (pyramid)



HOG Pyramid



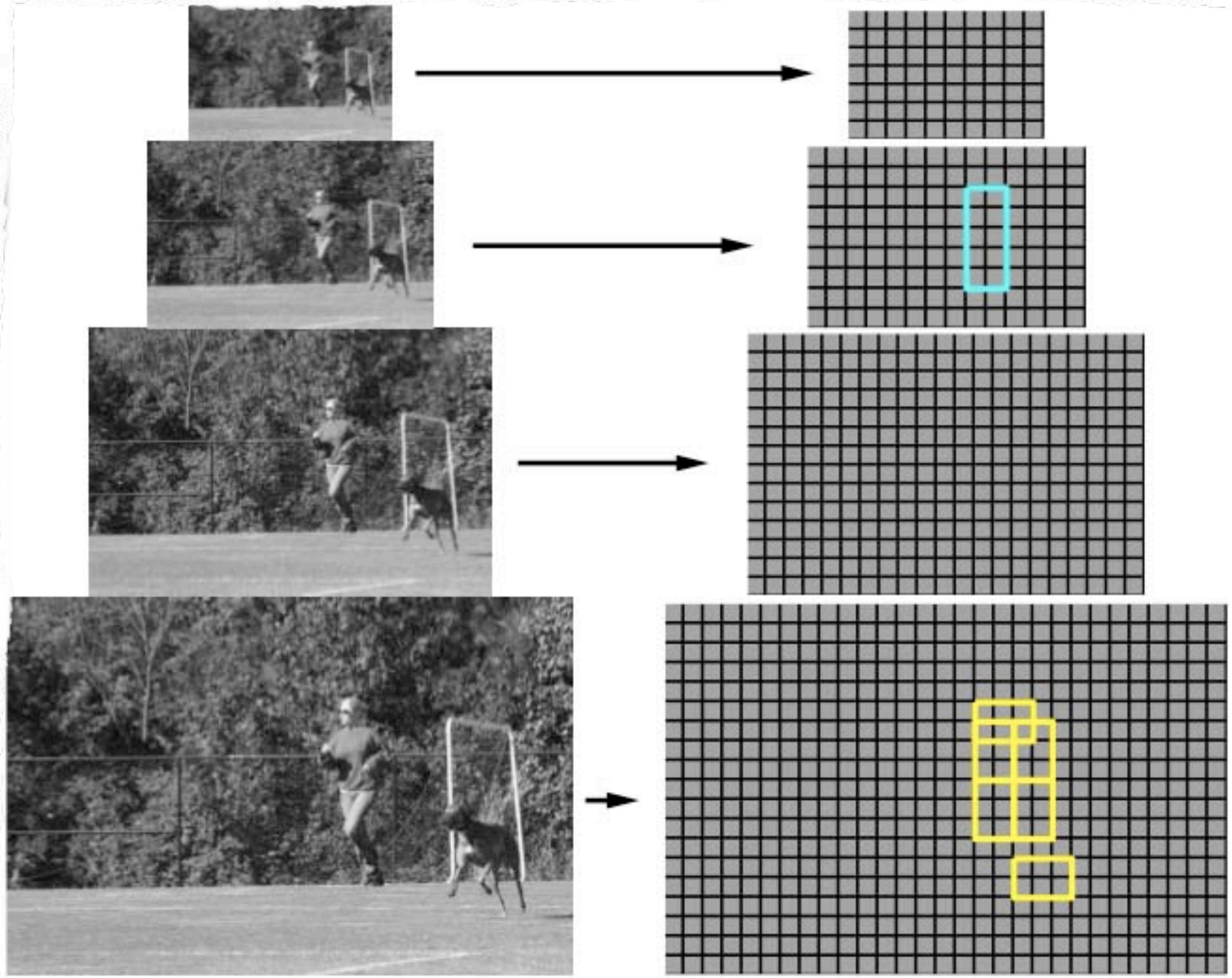
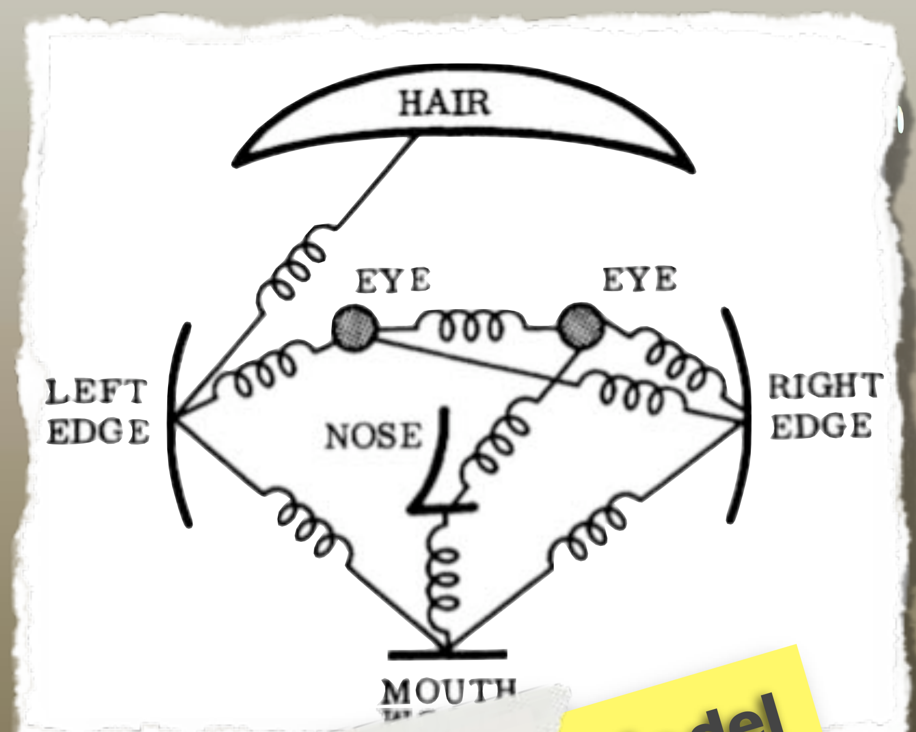
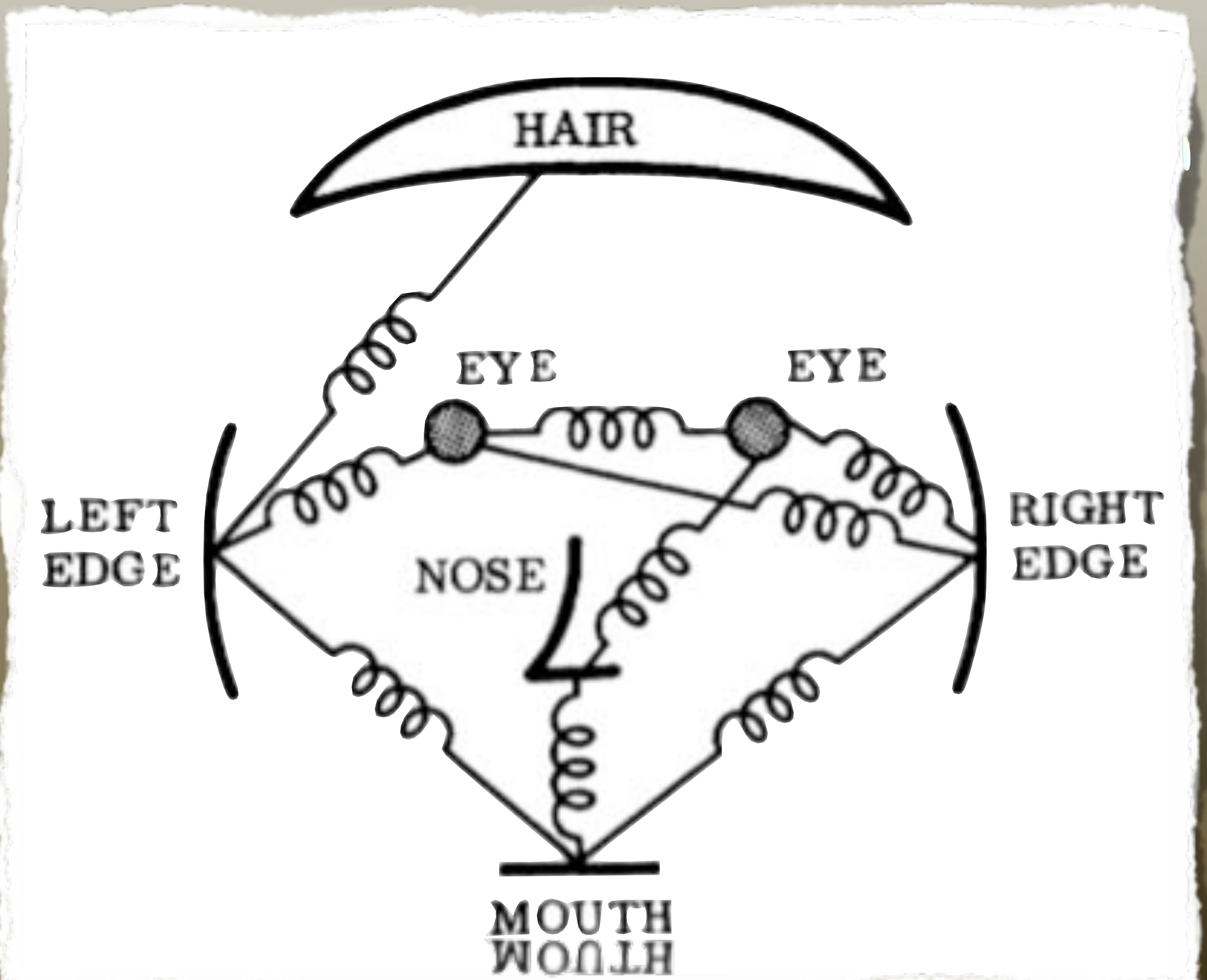


Figure 2. The HOG feature pyramid and an object hypothesis defined in terms of a placement of the root filter (near the top of the pyramid) and the part filters (near the bottom of the pyramid).



Deformable Part Model

Deformable Part Model



// each part is a **local property**

// **springs** capture spatial relationships

// here, the springs can be "**negative**"

Deformable Part Model

detection score =

sum of filter responses - deformation cost

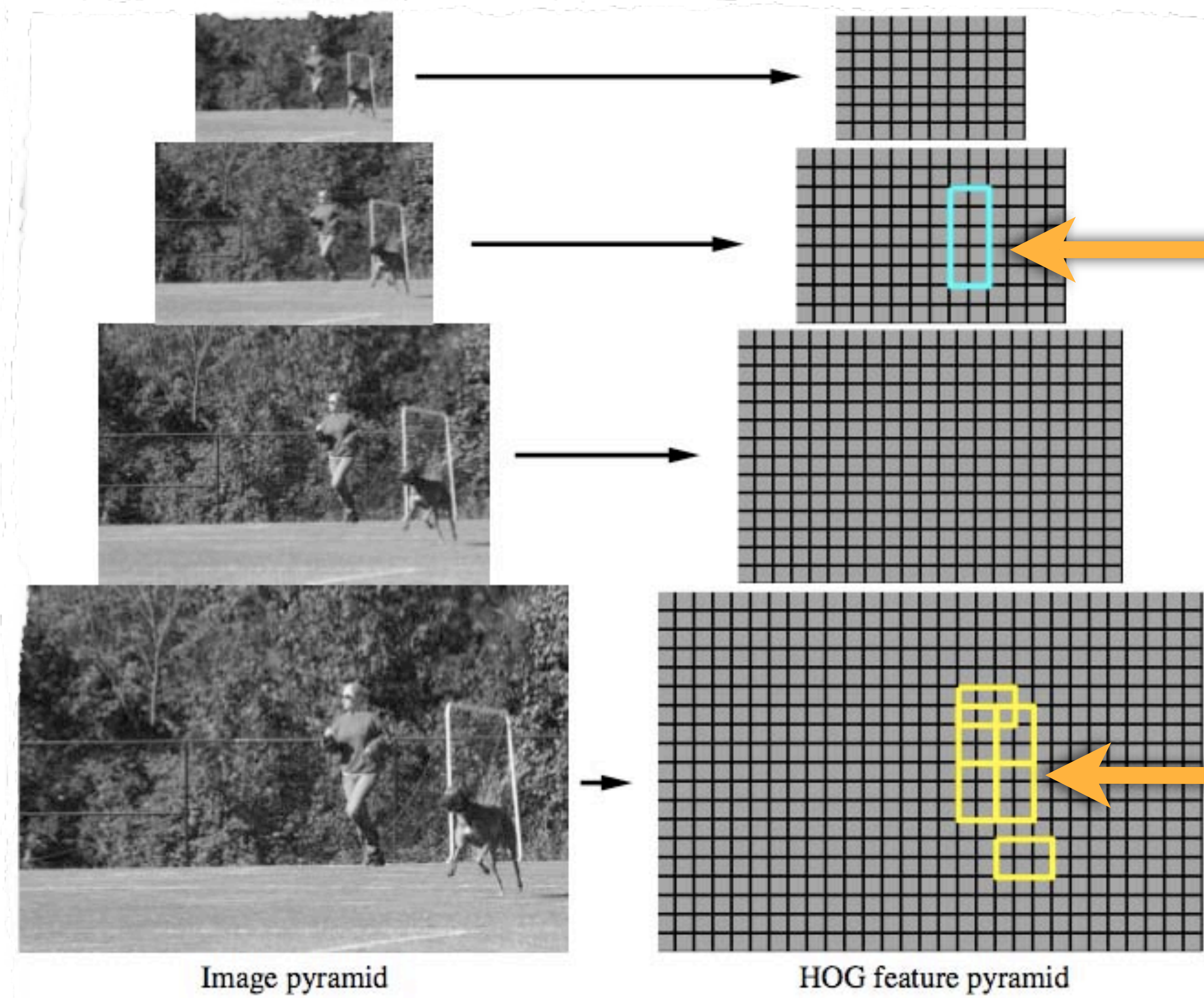
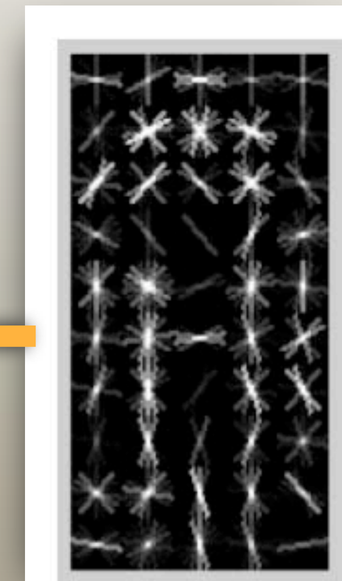
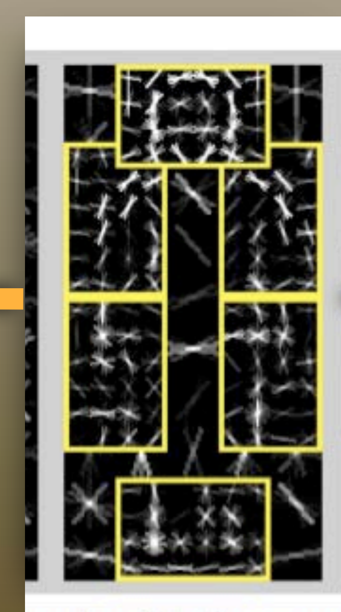


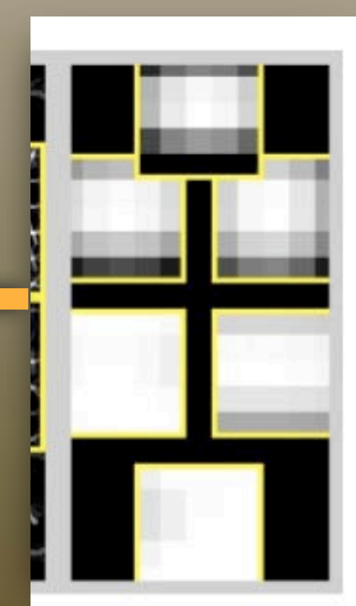
Figure 2. The HOG feature pyramid and an object hypothesis defined in terms of a placement of the root filter (near the top of the pyramid) and the part filters (near the bottom of the pyramid).



root filter



part filters



deformable model

Deformable Part Model

score of a placement

$$\sum_{i=0}^{n} F_i \cdot \phi(H, p_i) + \sum_{i=1}^{n} a_i \cdot (\tilde{x}_i, \tilde{y}_i) + b_i \cdot (\tilde{x}_i^2, \tilde{y}_i^2), \quad (1)$$

filters

feature vector (at position p in the pyramid H)

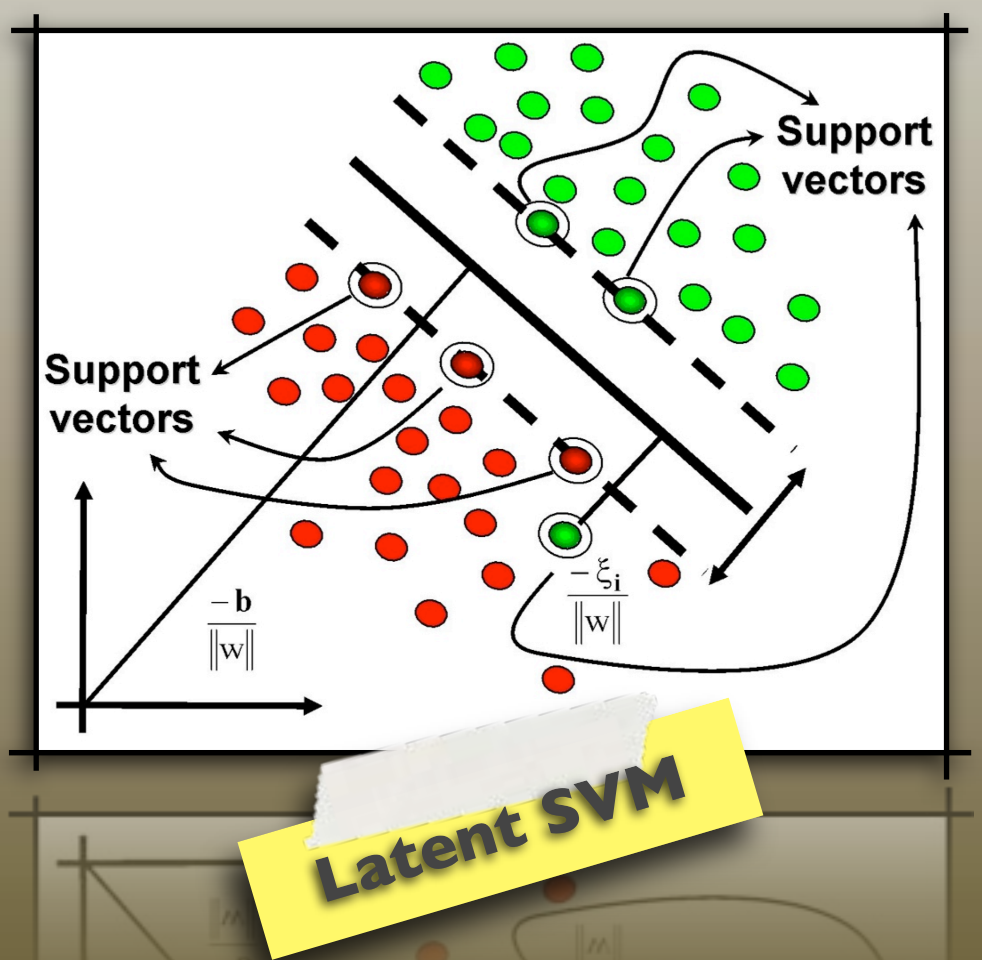
coefficients of a quadratic function on the placement

position relative to the root location

The score of a placement z can be expressed in terms of the dot product, $\beta \cdot \psi(H, z)$, between a vector of model parameters β and a vector $\psi(H, z)$,

$$\beta = (F_0, \ldots, F_n, a_1, b_1 \ldots, a_n, b_n).$$

$$\psi(H, z) = (\phi(H, p_0), \phi(H, p_1), \ldots \phi(H, p_n), \\ \tilde{x}_1, \tilde{y}_1, \tilde{x}_1^2, \tilde{y}_1^2, \ldots, \tilde{x}_n, \tilde{y}_n, \tilde{x}_n^2, \tilde{y}_n^2,).$$


Support vectors
Support vectors
$\frac{-b}{\|w\|}$
$\frac{-\xi_i}{\|w\|}$
Latent SVM

Latent SVM

The score of a placement z can be expressed in terms of the dot product, $\beta \cdot \psi(H, z)$, between a vector of model parameters β and a vector $\psi(H, z)$,

$$\beta = (F_0, \ldots, F_n, a_1, b_1 \ldots, a_n, b_n).$$

$$\psi(H, z) = (\phi(H, p_0), \phi(H, p_1), \ldots \phi(H, p_n), \tilde{x}_1, \tilde{y}_1, \tilde{x}_1^2, \tilde{y}_1^2, \ldots, \tilde{x}_n, \tilde{y}_n, \tilde{x}_n^2, \tilde{y}_n^2,).$$

filters and deformation parameters

features

part displacements

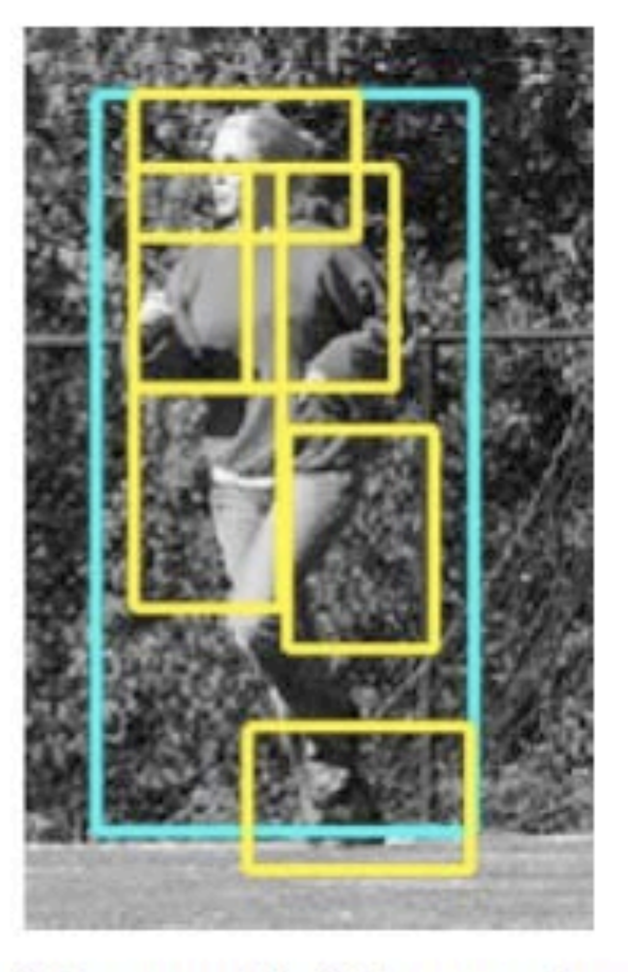
A latent SVM is defined as follows. We assume that each example x is scored by a function of the form,

$$f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z), \tag{2}$$

where β is a vector of model parameters and z is a set of latent values. For our deformable models we define $\Phi(x, z) = \psi(H(x), z)$ so that $\beta \cdot \Phi(x, z)$ is the score of placing the model according to z.

In analogy to classical SVMs we would like to train β from labeled examples $D = (\langle x_1, y_1 \rangle, \ldots, \langle x_n, y_n \rangle)$ by optimizing the following objective function,

$$\beta^*(D) = \underset{\beta}{\operatorname{argmin}}\, \lambda ||\beta||^2 + \sum_{i=1}^{n} \max(0, 1 - y_i f_\beta(x_i)). \tag{3}$$



Latent SVM

$$f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z)$$

Note that $f_\beta(x)$ as defined in (2) is a maximum of functions each of which is linear in β. Hence $f_\beta(x)$ is convex in β. This implies that the hinge loss $\max(0, 1 - y_i f_\beta(x_i))$ is convex in β when $y_i = -1$. That is, the loss function is convex in β for negative examples. We call this property of the loss function *semi-convexity*.

1. Holding β fixed, optimize the latent values for the positive examples $z_i = \text{argmax}_{z \in Z(x_i)} \beta \cdot \Phi(x, z)$.

2. Holding $\{z_i\}$ fixed for positive examples, optimize β by solving the convex problem defined above.

Bonus

// Data Mining **Hard** Negatives

// Model Initialization

Results

Pascal VOC 2006

	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	per
Our rank	3	1	2	1	1	2	2	4	1	1	1	4	2	2	1
Our score	.180	**.411**	.092	**.098**	**.249**	.349	.396	.110	**.155**	**.165**	**.110**	.062	.301	.337	**.2**
Darmstadt							.301								
INRIA Normal	.092	.246	.012	.002	.068	.197	.265	.018	.097	.039	.017	.016	.225	.153	.
INRIA Plus	.136	.287	.041	.025	.077	.279	.294	.132	.106	.127	.067	.071	**.335**	.249	.0
IRISA		.281					.318	.026	.097	.119			.289	.227	.
MPI Center	.060	.110	.028	.031	.000	.164	.172	.208	.002	.044	.049	.141	.198	.170	.0
MPI ESSOL	.152	.157	**.098**	.016	.001	.186	.120	**.240**	.007	.061	.098	**.162**	.034	.208	.1
Oxford	**.262**	.409				**.393**	**.432**							**.375**	
TKK	.186	.078	.043	.072	.002	.116	.184	.050	.028	.100	.086	.126	.186	.135	.0

Results

Models learned

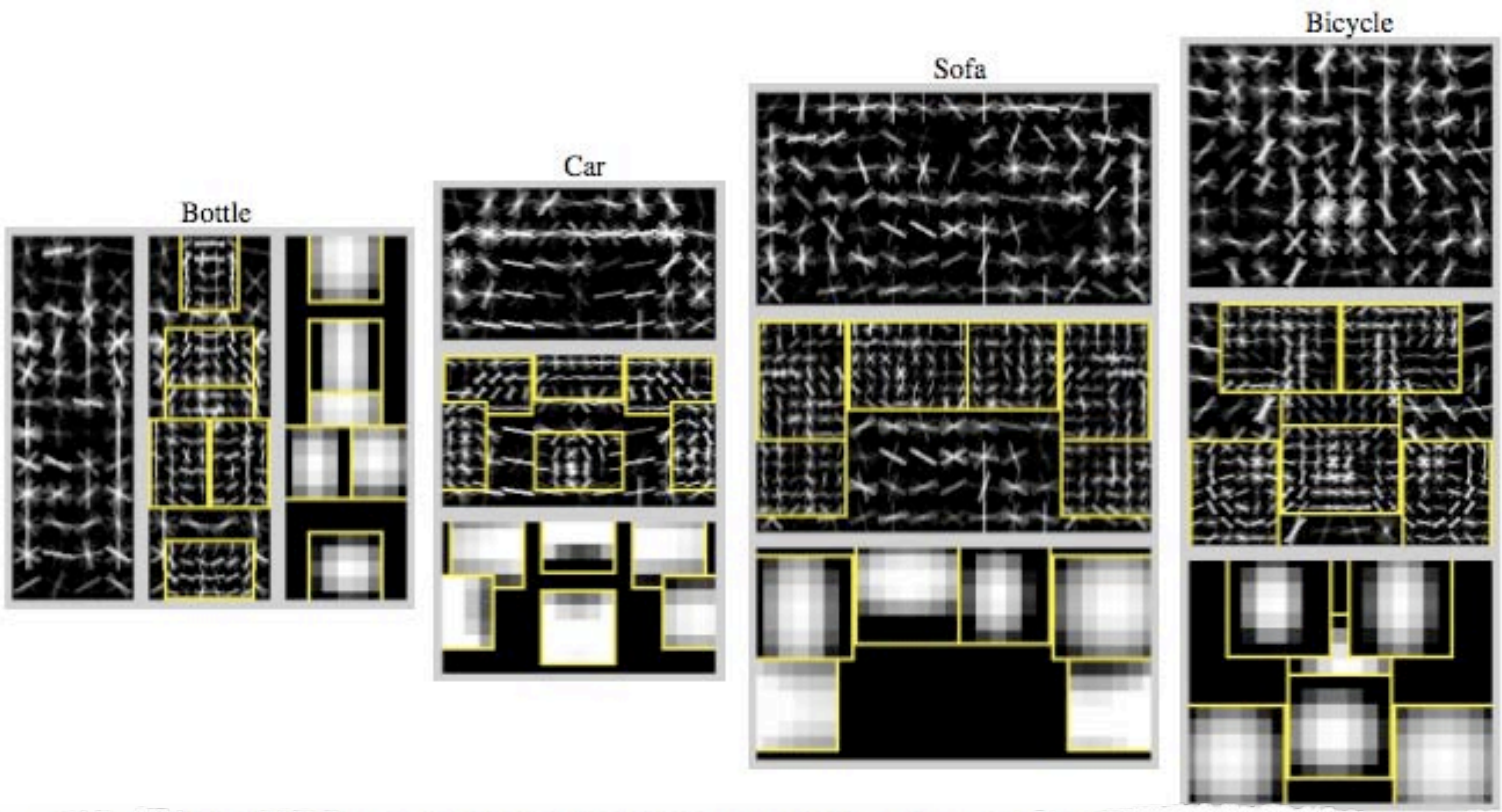
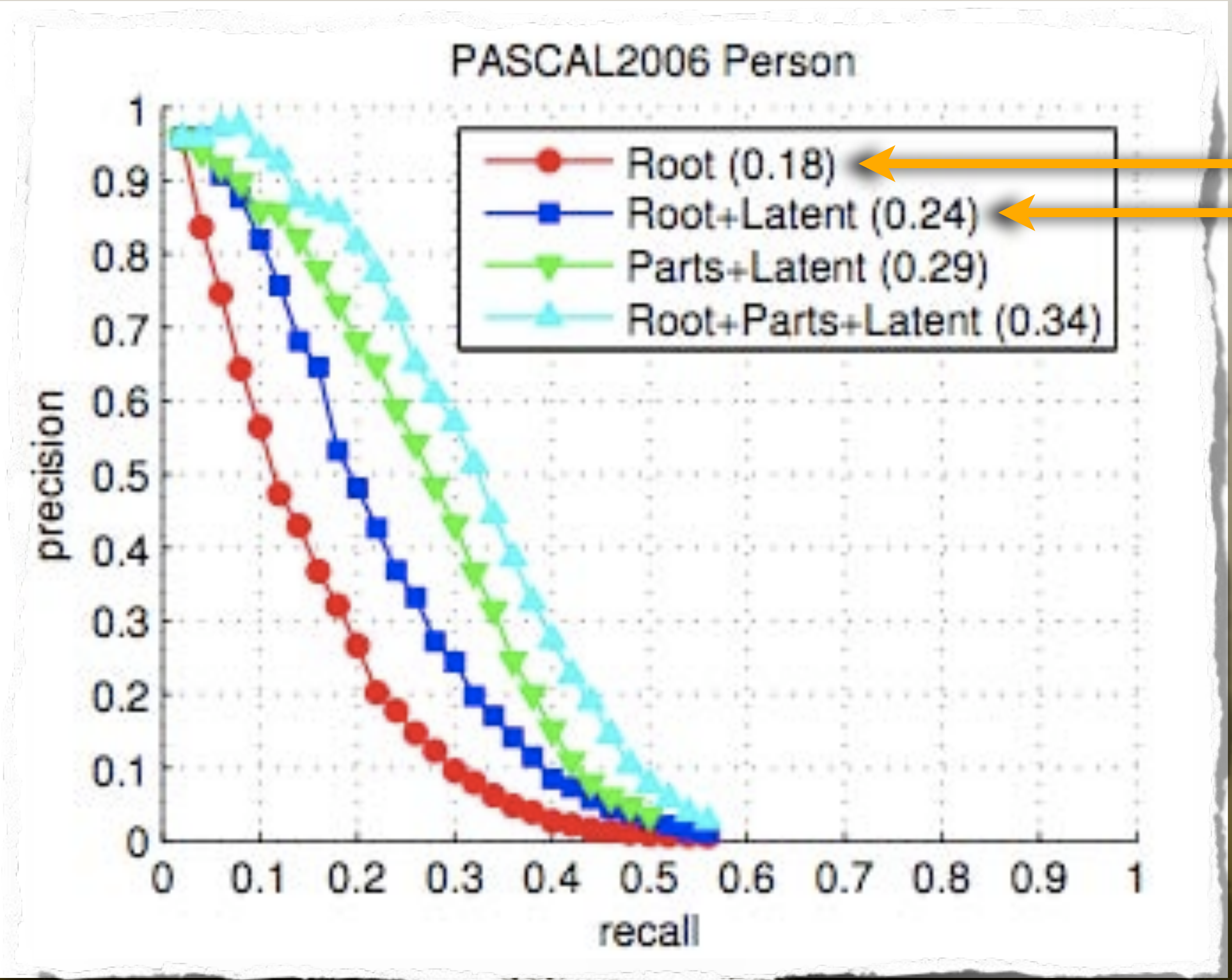


Figure 4. Some models learned from the PASCAL VOC 2007 dataset. We show the total energy in each orientation of the HOG cells in

Experiments



~ Dalal’s model

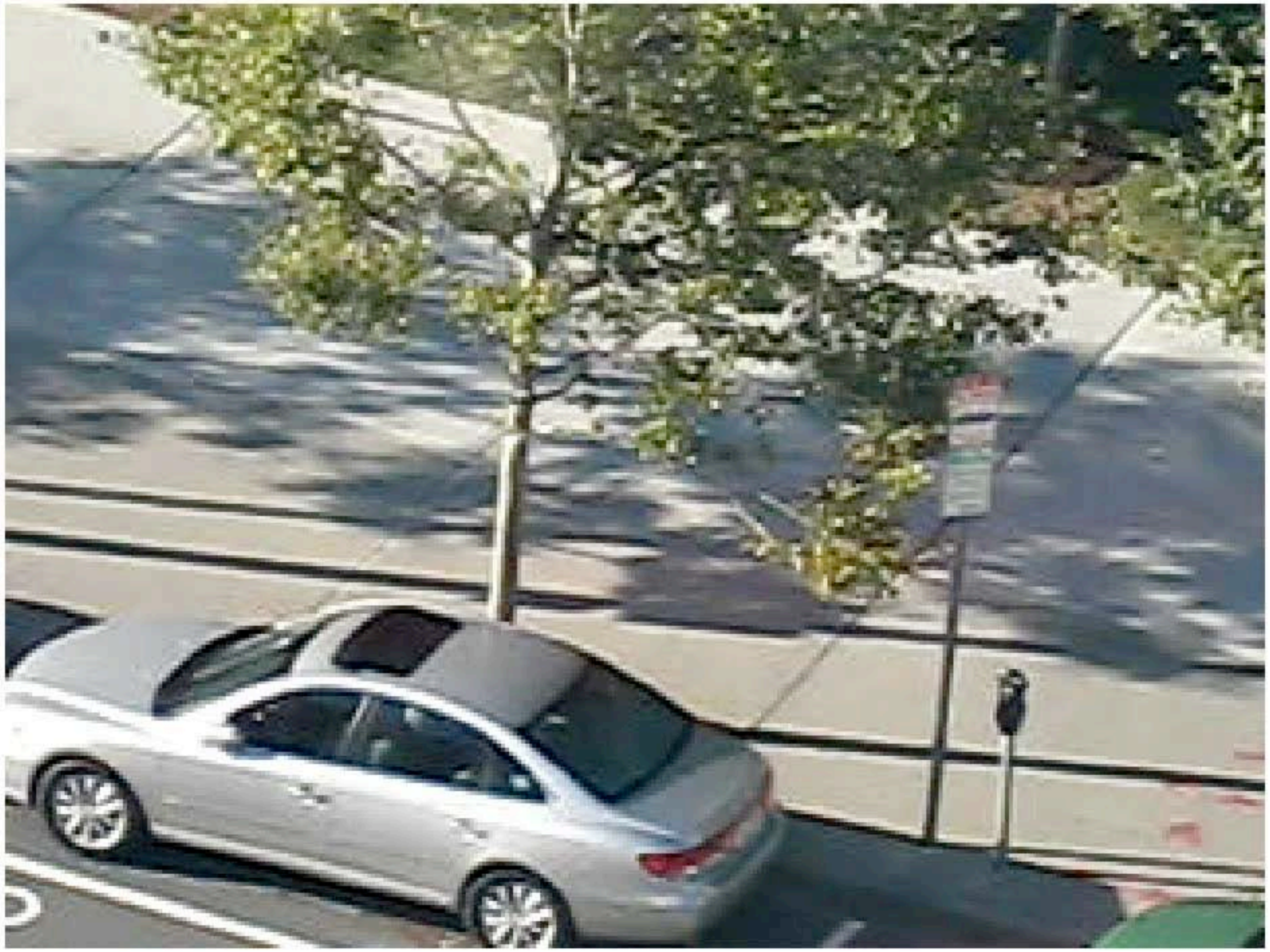
~ Dalal’s + LSVM

Examples

errors

A simple demo...

A simple demo...



Conclusions

so, it **doesn't work** ?!?

no no, it **works...**

...it just **doesn't work well...**

...or there is a **problem with the seat-computer interface...**

Conclusion

"The aim of computer vision is to overfit to our visual world"

-- remark by Antonio Torralba (after his third beer)

http://www.cs.cmu.edu/~efros/courses/LBMV07/


"Thank You"
CALL AGAIN
and bring
a Friend!
MIT
VOX