

# Computer vision, Feature extraction and Deep Learning

by

Ankit Jha

Resources courtesy: Google, Medium Blogs, Courses from IIT Bombay, MIT, etc.

# Computer Vision

Make computers understand images and video.



What kind of scene?

Where are the cars?

How far is the building?

...

# Vision is really hard

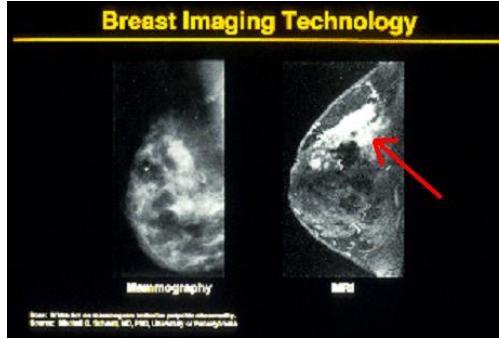
- Vision is an amazing feat of natural intelligence
  - Visual cortex occupies about 50% of Macaque brain
  - More human brain devoted to vision than anything else



# Why computer vision matters



Safety



Health



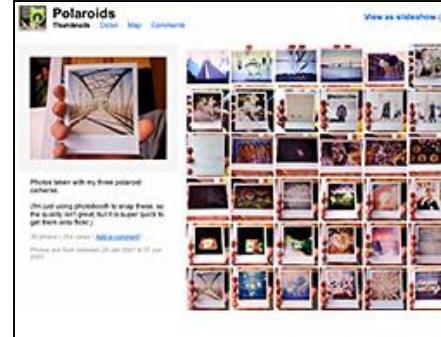
Security



Comfort



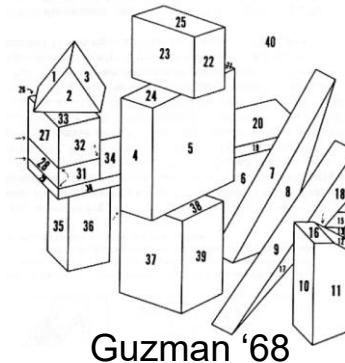
Fun



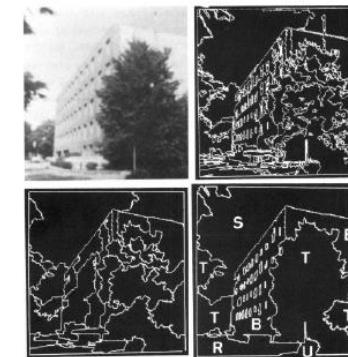
Access

# Ridiculously brief history of computer vision

- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts



Guzman '68



Ohta Kanade '78



Turk and Pentland '91

# How vision is used now

- Examples of state-of-the-art

# Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software



Digit recognition, AT&T labs  
<http://www.research.att.com/~yann/>



License plate readers  
[http://en.wikipedia.org/wiki/Automatic\\_number\\_plate\\_recognition](http://en.wikipedia.org/wiki/Automatic_number_plate_recognition)

# Face detection

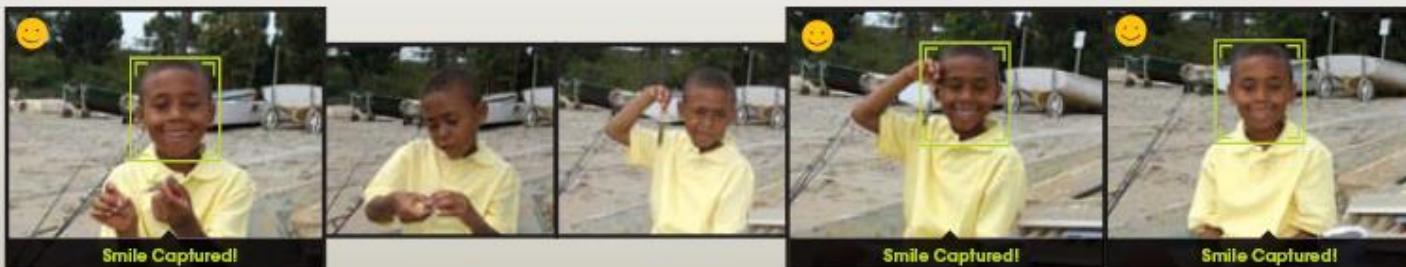
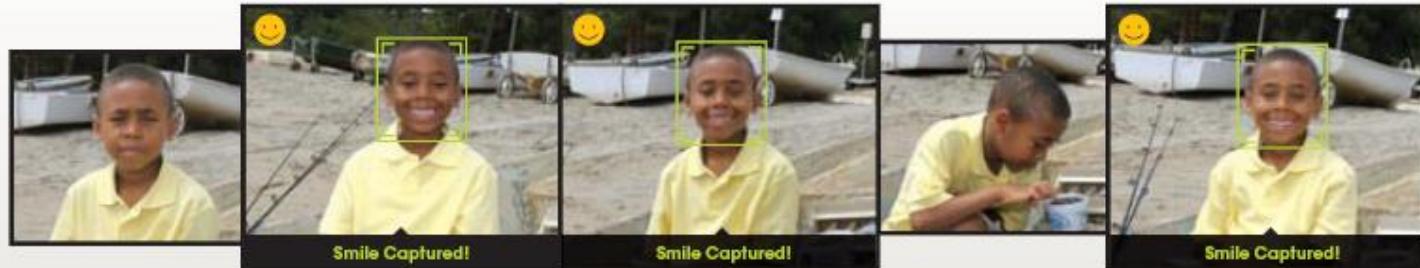


- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, ...

# Smile detection

## The Smile Shutter flow

Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.



[Sony Cyber-shot® T70 Digital Still Camera](#)

# 3D from thousands of images



# Object recognition (in supermarkets)



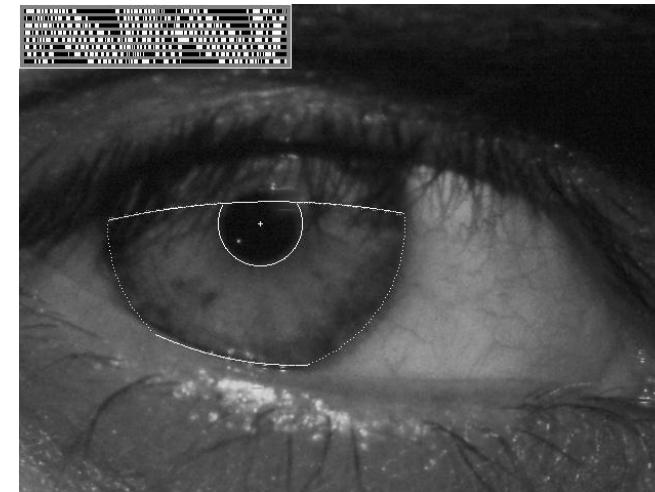
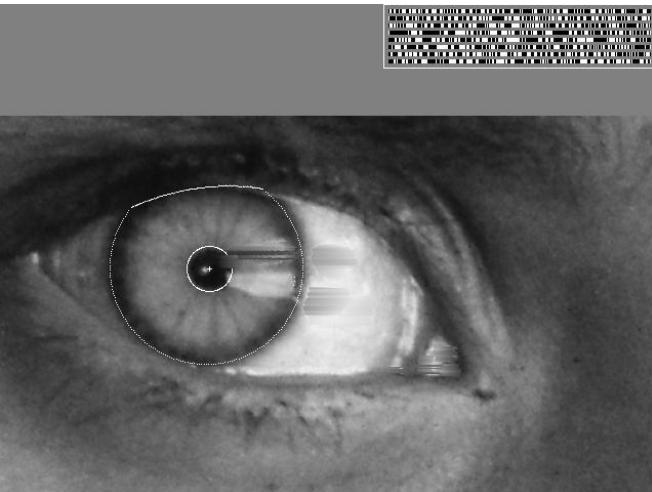
## [LaneHawk by EvolutionRobotics](#)

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk, you are assured to get paid for it...”

# Vision-based biometrics



*“How the Afghan Girl was Identified by Her Iris Patterns”* Read the [story](#)  
[wikipedia](#)



# Login without a password...



Fingerprint scanners on  
many new laptops,  
other devices



Face recognition systems now  
beginning to appear more widely  
<http://www.sensiblevision.com/>

# Object recognition (in mobile phones)



Point & Find, Nokia  
Google Goggles

# Special effects: shape capture



*The Matrix* movies, ESC Entertainment, XYZRGB, NRC

# Special effects: motion capture



*Pirates of the Caribbean*, Industrial Light and Magic

# Sports



*Sportvision* first down line  
Nice [explanation](#) on [www.howstuffworks.com](http://www.howstuffworks.com)

<http://www.sportvision.com/video.html>

# Smart cars

Slide content courtesy of Amnon Shashua

The screenshot shows the Mobileye website homepage. At the top, there are two tabs: "manufacturer products" (highlighted) and "consumer products". Below the tabs is a banner with the slogan "Our Vision. Your Safety." featuring an overhead view of a car with three cameras labeled: "rear looking camera", "forward looking camera", and "side looking camera". To the right of the banner is a "News" section with a list of articles and a thumbnail image of a hand on a steering wheel. Below the banner are three main sections: "EyeQ Vision on a Chip" (with an image of a chip), "Vision Applications" (with an image of a pedestrian crossing), and "AWS Advance Warning System" (with an image of a dashboard display). Each section has a "read more" link.

▶▶ manufacturer products consumer products ◀◀

**Our Vision. Your Safety.**

rear looking camera forward looking camera side looking camera

▶ EyeQ Vision on a Chip

▶ Vision Applications

▶ AWS Advance Warning System

News

- Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System
- Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end

> all news

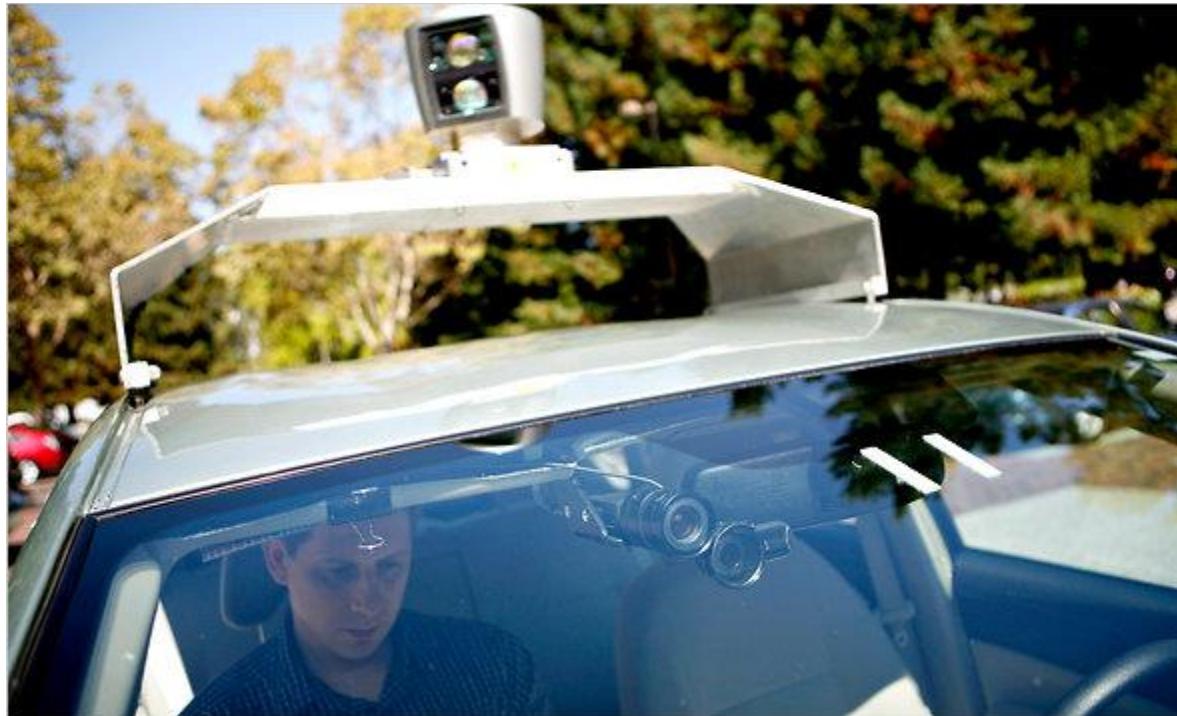
Events

- Mobileye at Equip Auto, Paris, France
- Mobileye at SEMA, Las Vegas, NV

> read more

- Mobileye
  - Vision systems currently in high-end BMW, GM, Volvo models
  - By 2010: 70% of car manufacturers.

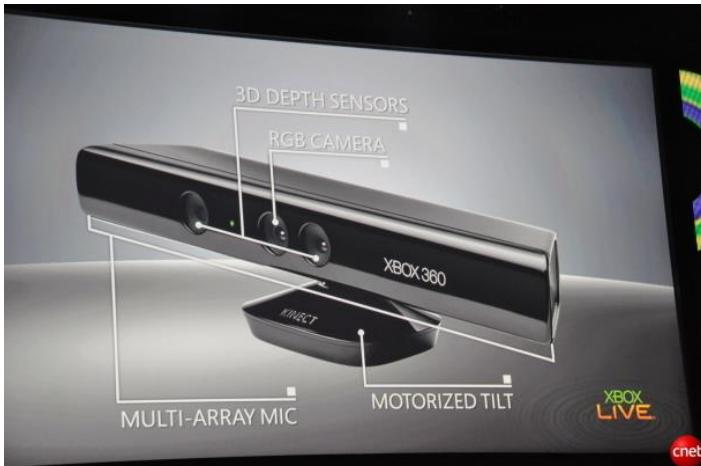
# Google cars



<http://www.nytimes.com/2010/10/10/science/10google.html?ref=artificialintelligence>

# Interactive Games: Kinect

- Object Recognition:  
<http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o>
- Mario: <http://www.youtube.com/watch?v=8CTJL5IUjHg>
- 3D: <http://www.youtube.com/watch?v=7QrnwoO1-8A>
- Robot: <http://www.youtube.com/watch?v=w8BmgtMKFbY>



# Vision in space

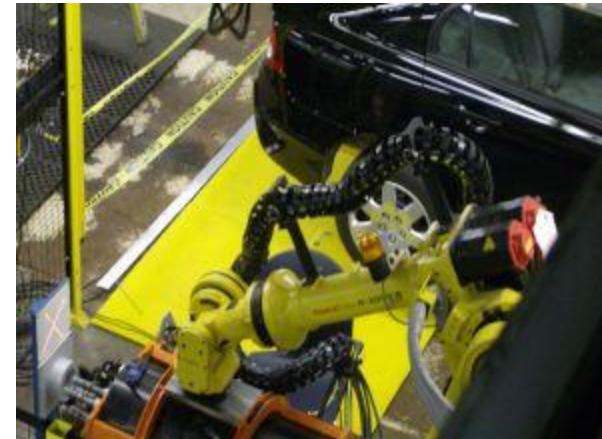


NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

## Vision systems (JPL) used for several tasks

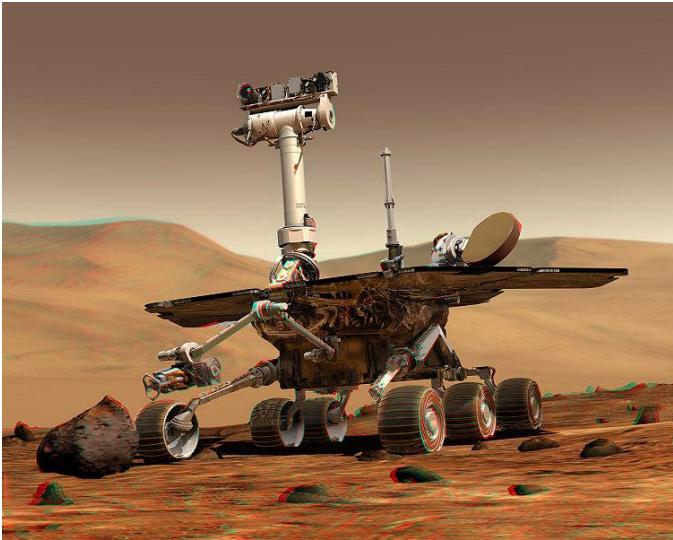
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read “Computer Vision on Mars” by Matthies et al.

# Industrial robots



Vision-guided robots position nut runners on wheels

# Mobile robots



NASA's Mars Spirit Rover  
[http://en.wikipedia.org/wiki/Spirit\\_rover](http://en.wikipedia.org/wiki/Spirit_rover)

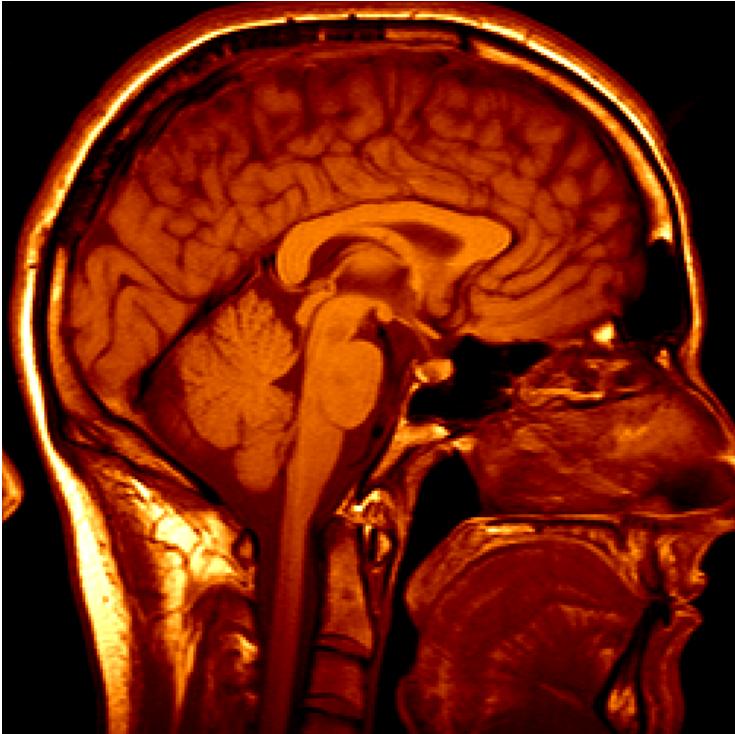


<http://www.robocup.org/>



Saxena et al. 2008  
STAIR at Stanford

# Medical imaging

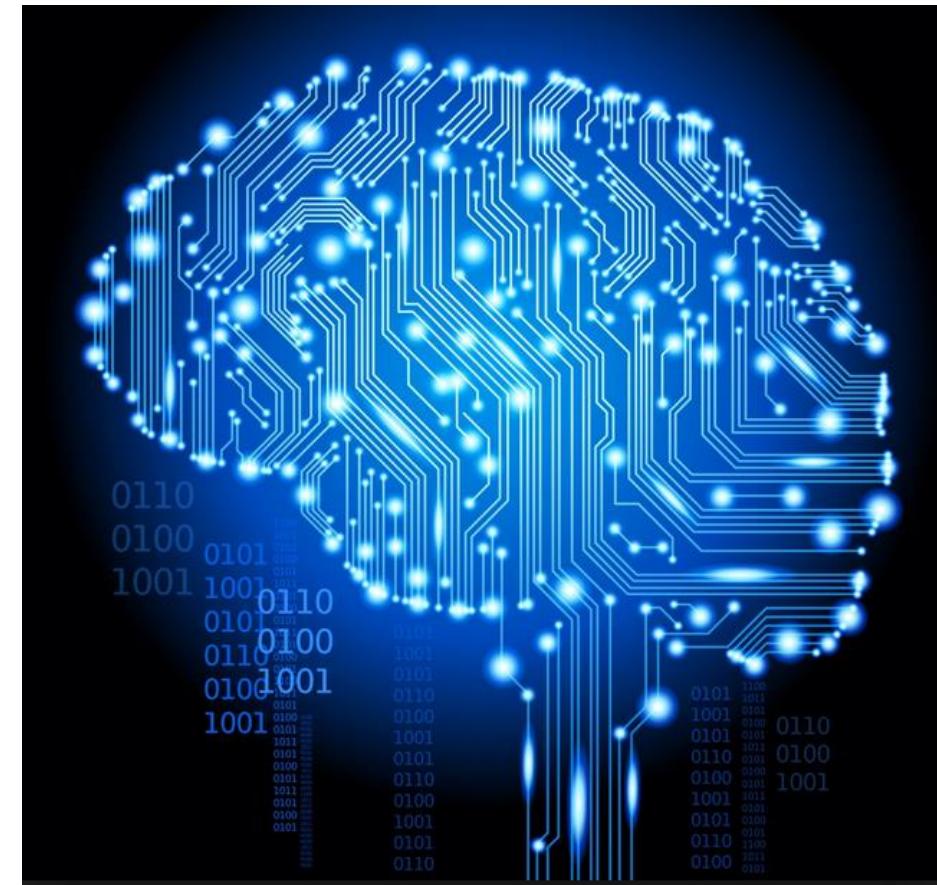


3D imaging  
MRI, CT



Image guided surgery  
Grimson et al., MIT

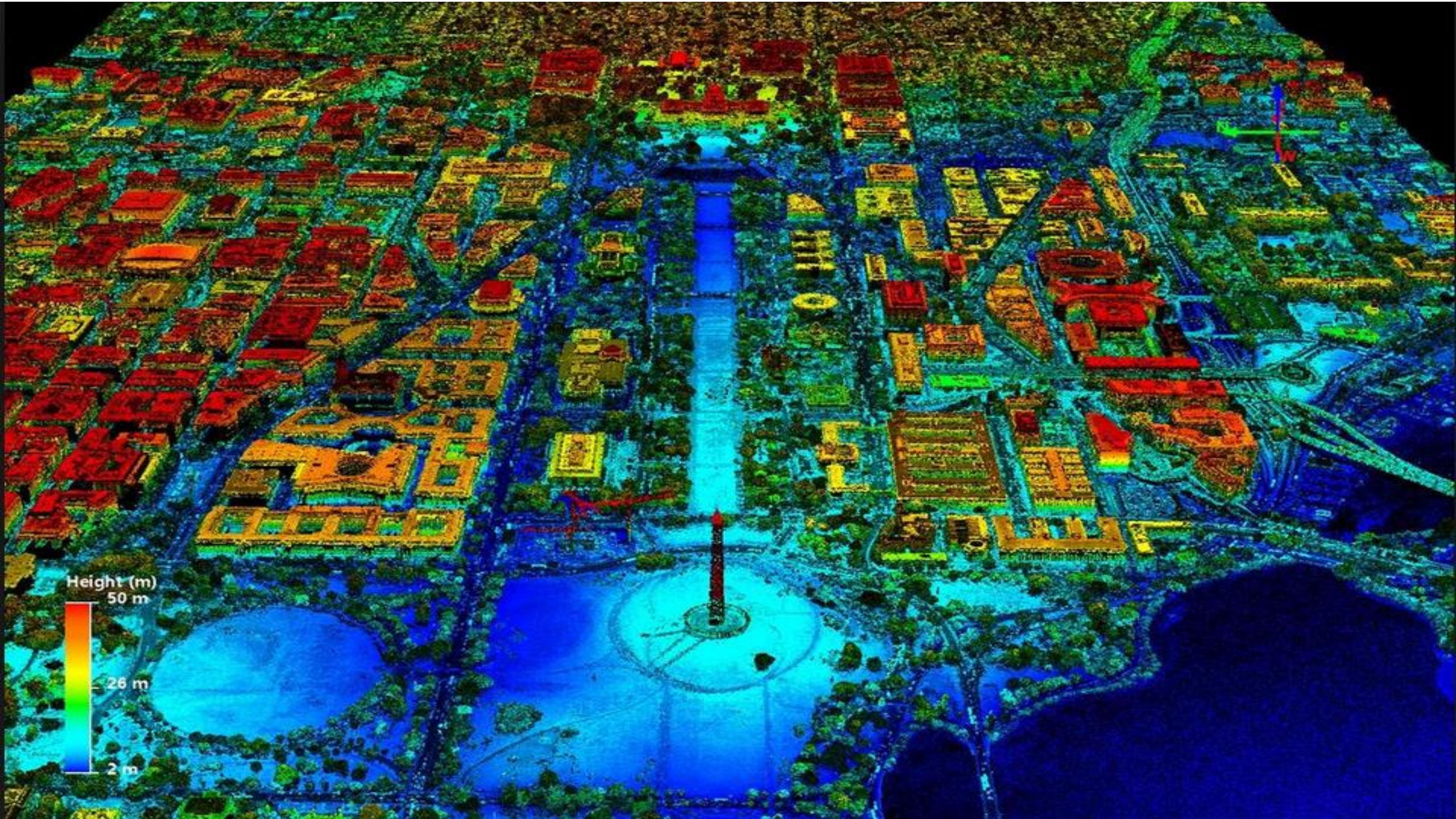
# Deep learning for visual inference

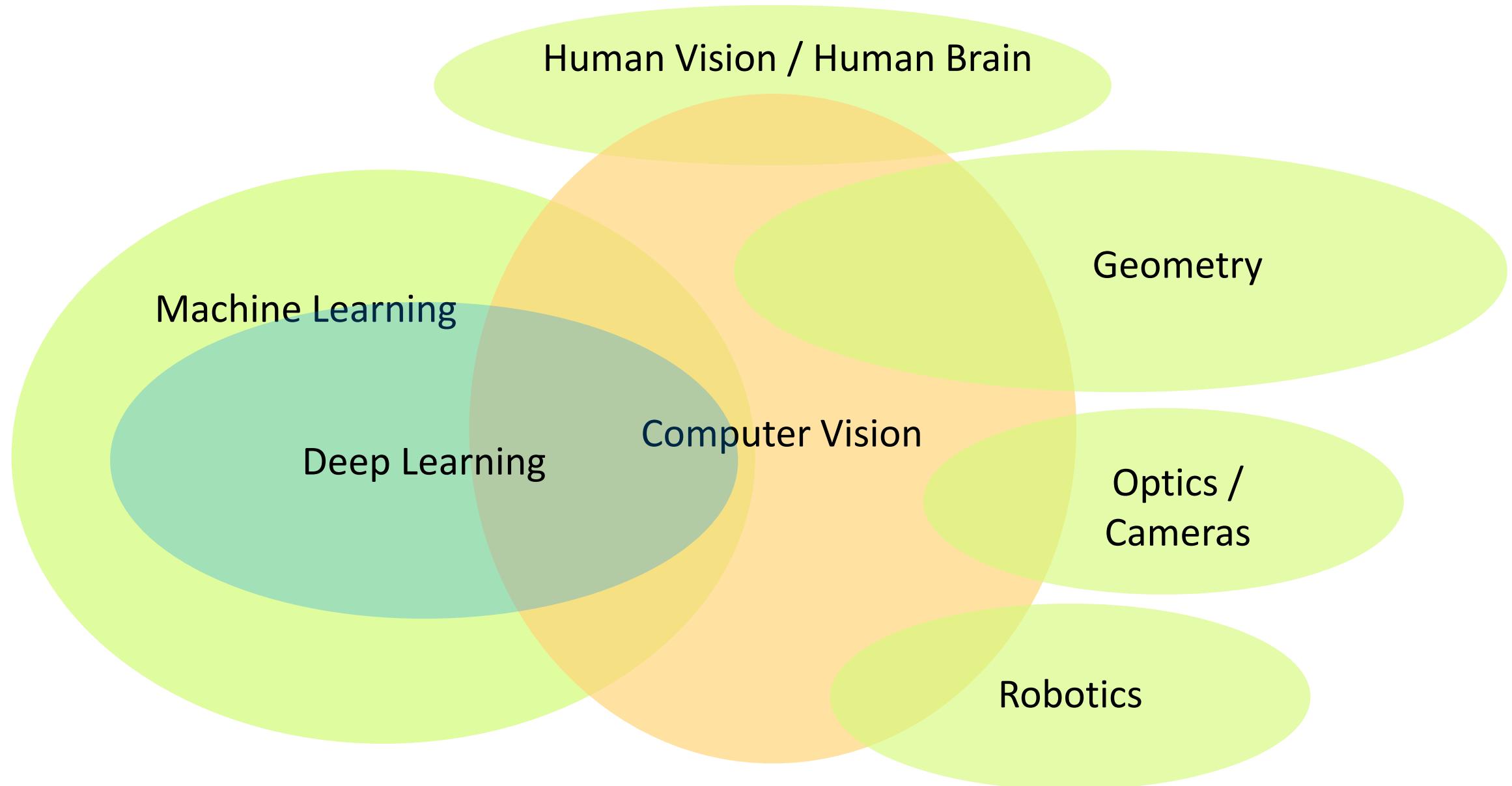




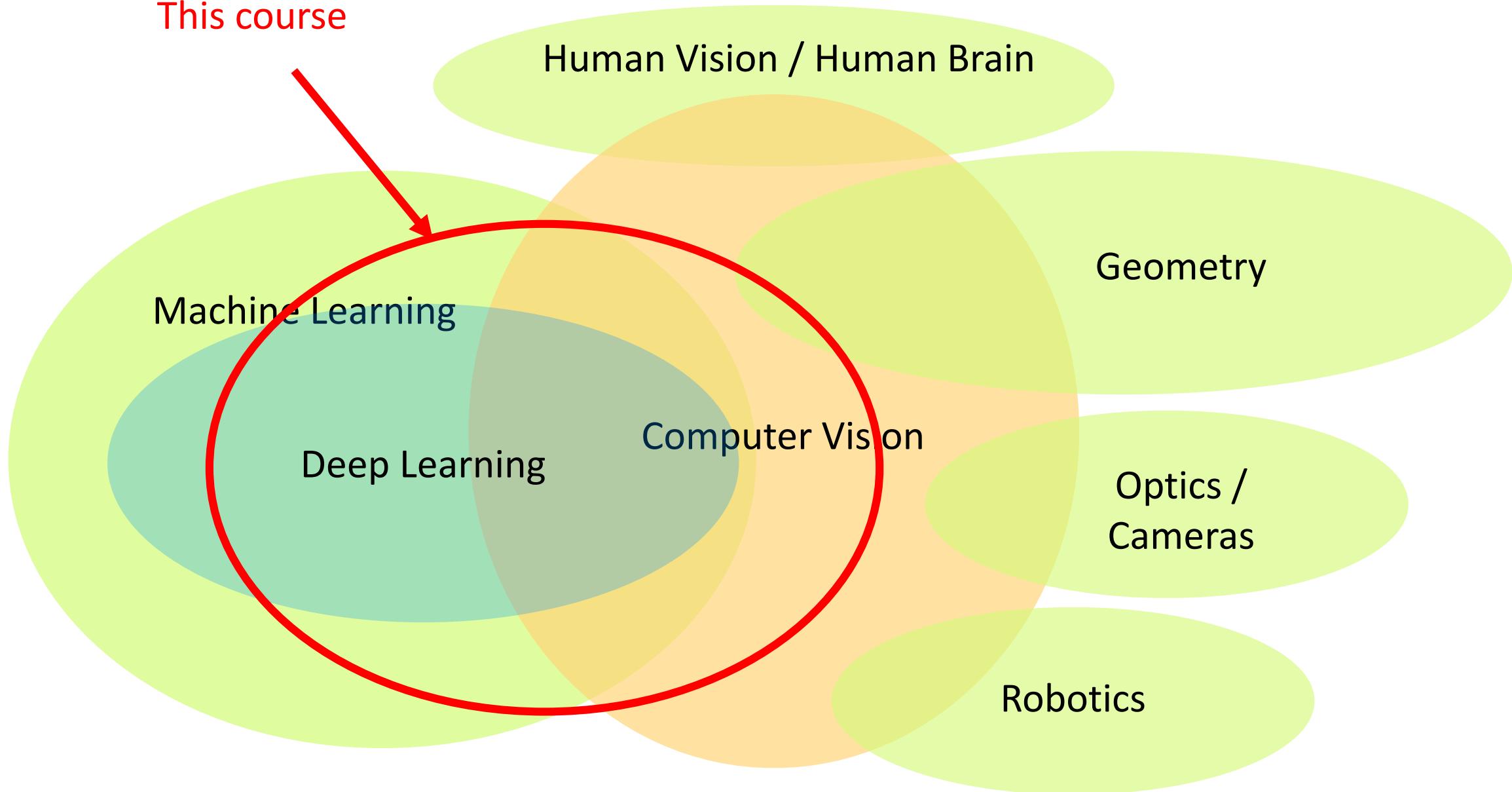








This course



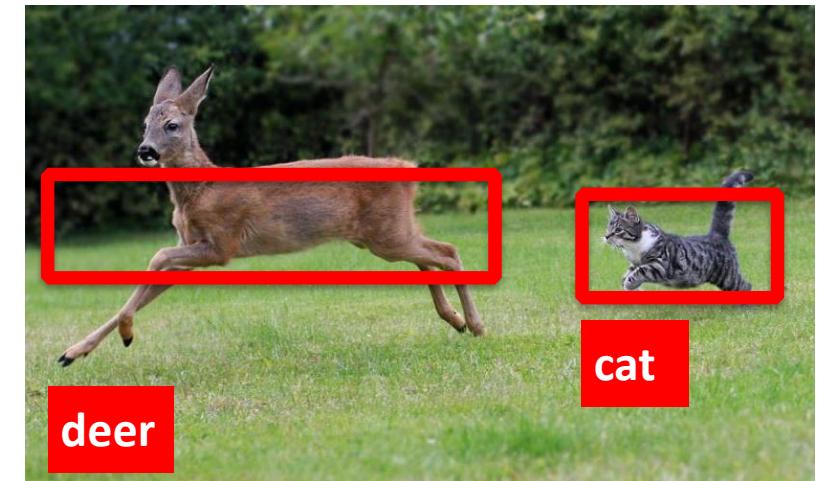
# Relationship with Other Fields

- Image Processing: Image ————— Image



# Relationship with Other Fields

- Computer Vision:      Image    ————— Knowledge



# Relationship with Other Fields

- Computer Graphics: Knowledge ————— Image

Vertices, Locations, Objects,  
Shapes, Colors, Material properties,  
Lighting settings, Camera settings, etc.

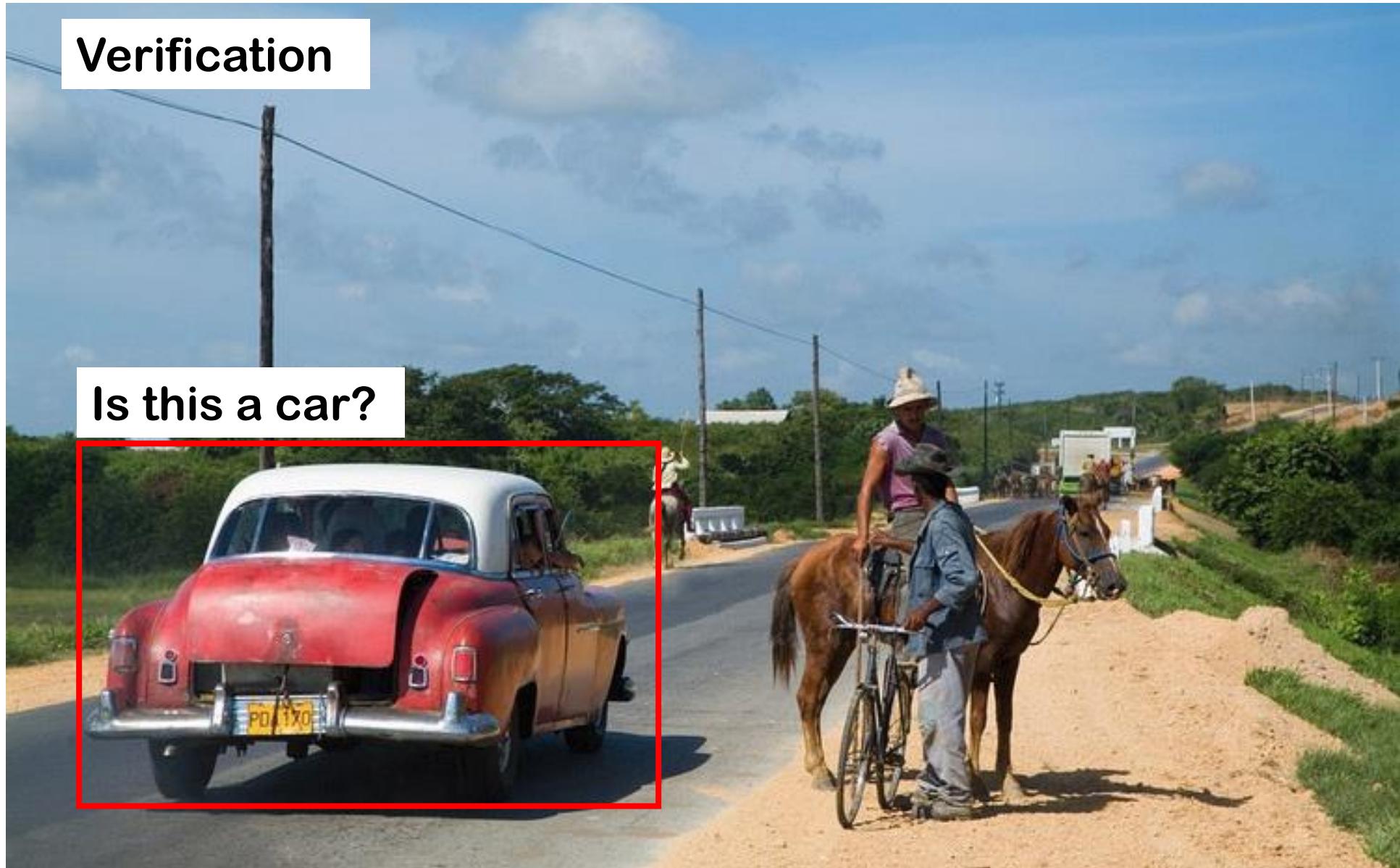


# Visual Recognition?

- What does it mean to “see”?
  - “What” is “where”, Marr 1982
- Get computers to “see”

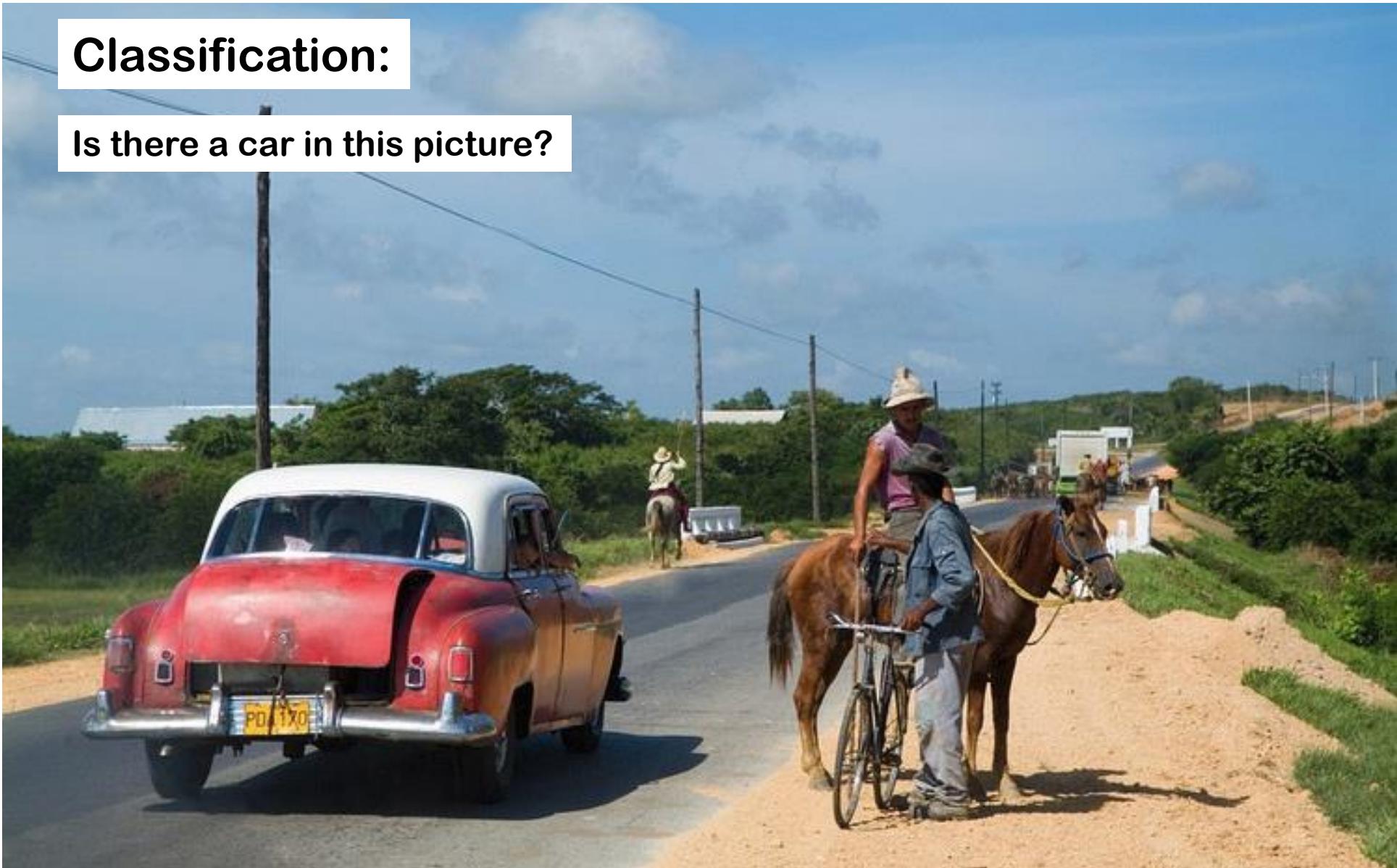
**Verification**

**Is this a car?**



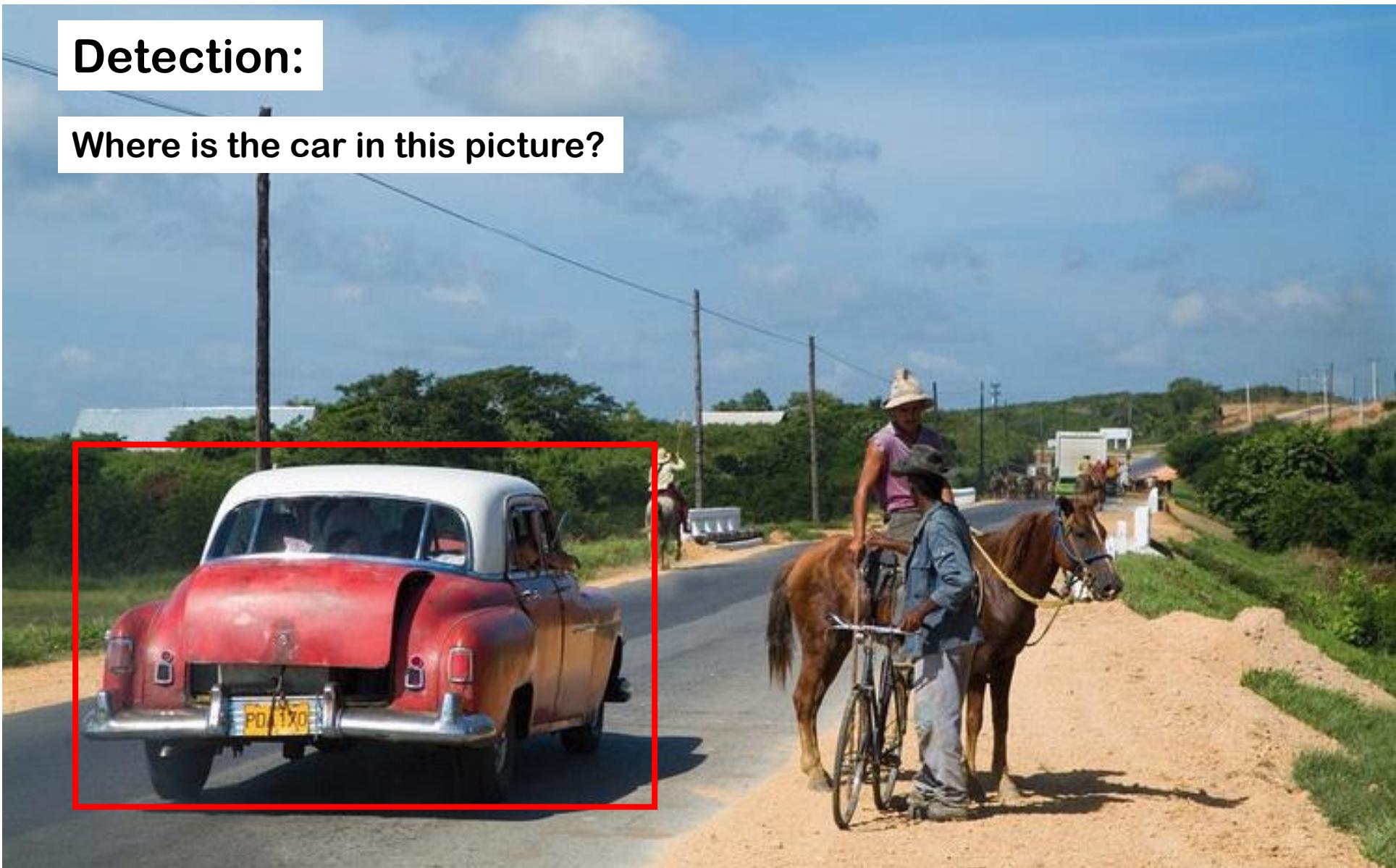
**Classification:**

**Is there a car in this picture?**



**Detection:**

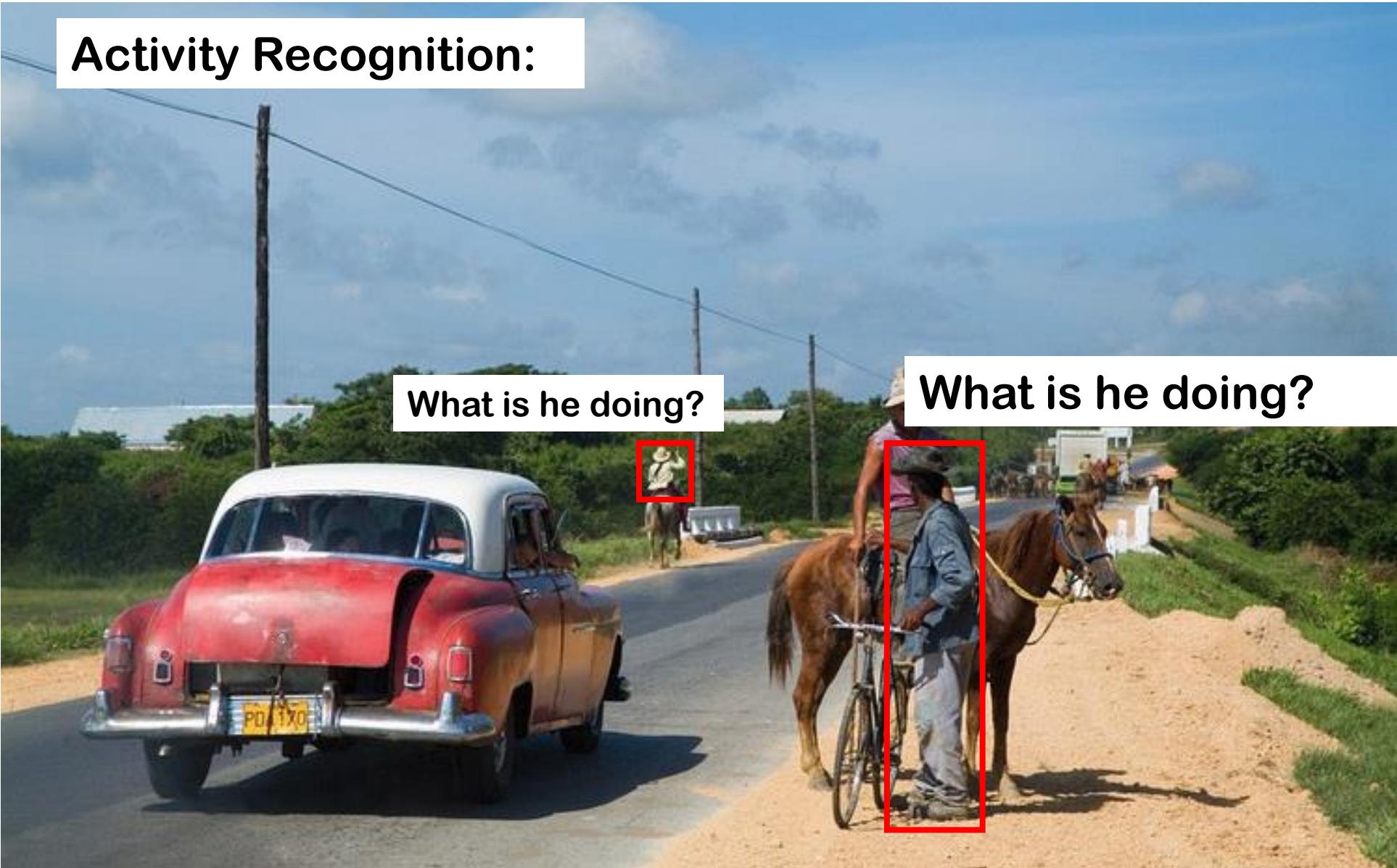
**Where is the car in this picture?**



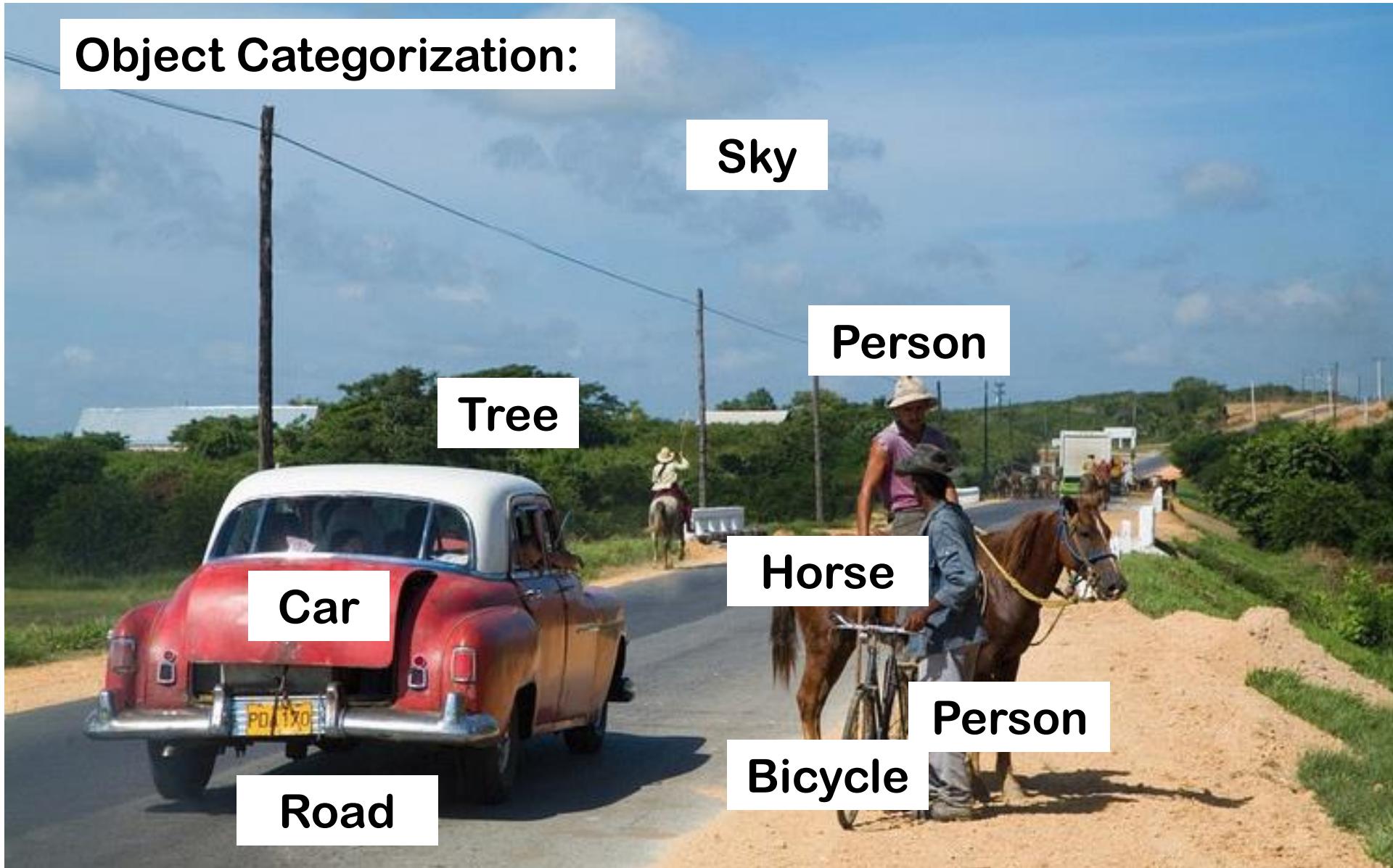
## Pose Estimation:



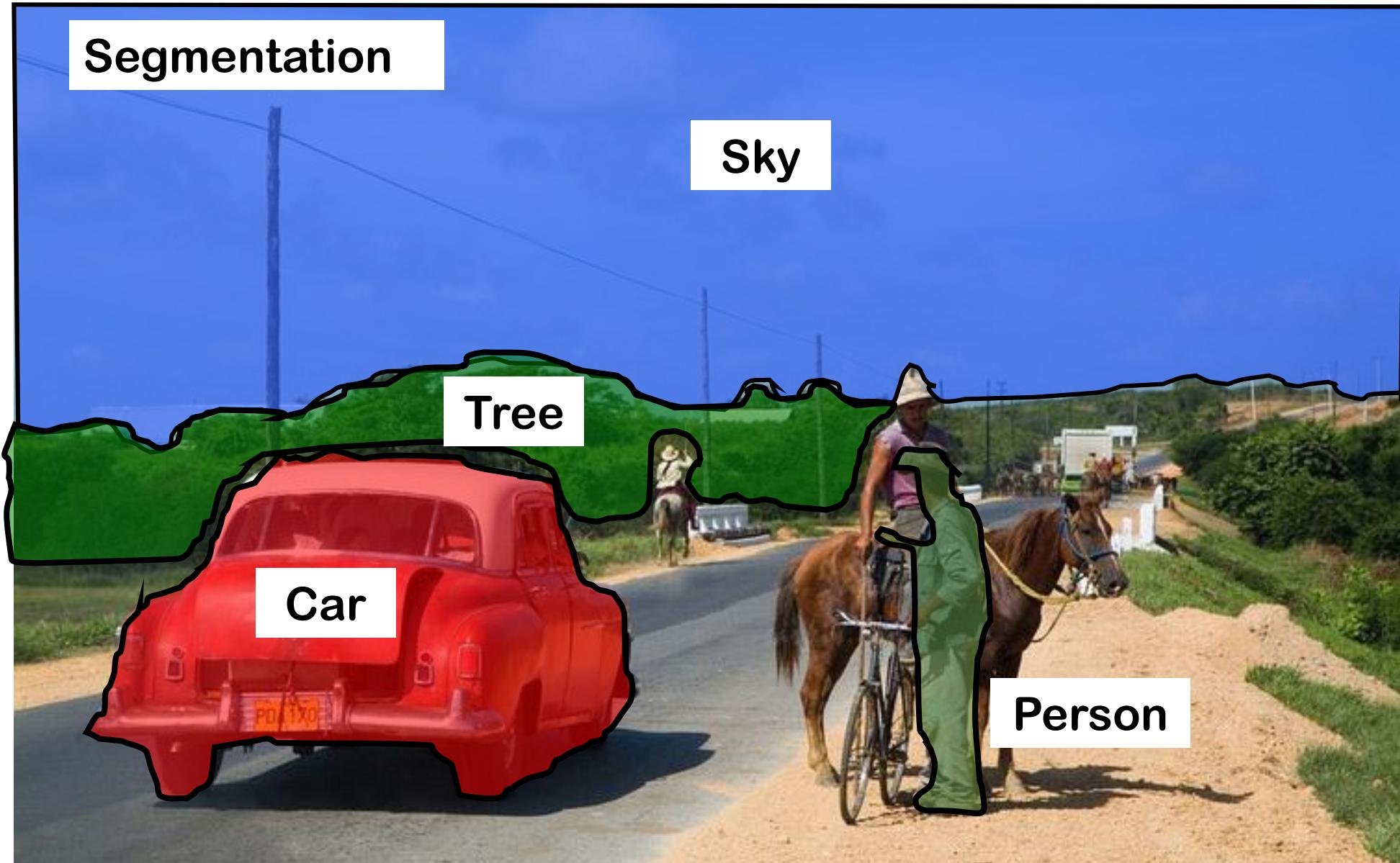
## Activity Recognition:



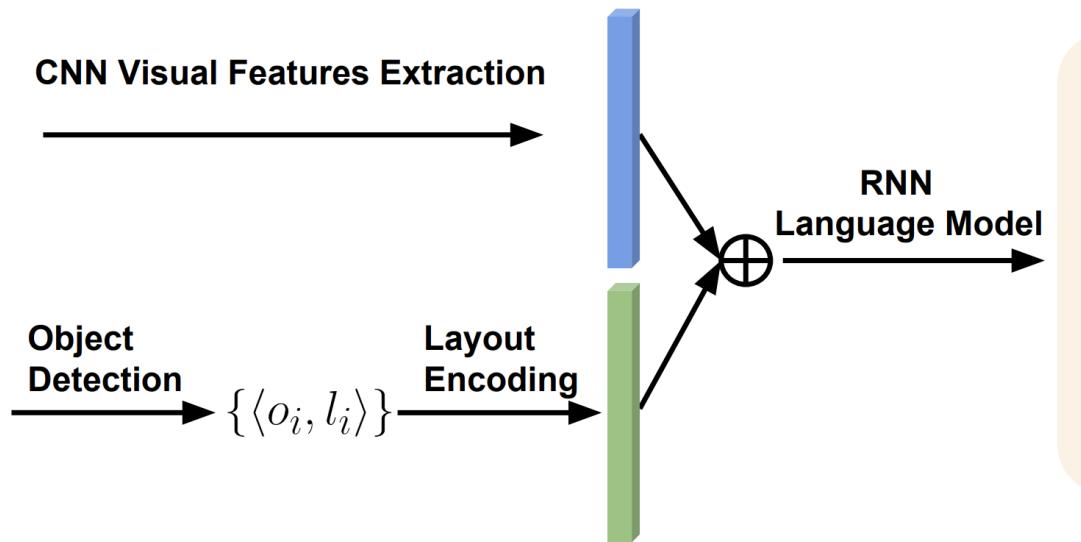
## Object Categorization:



# Segmentation



# Describing Images with Language



a group  
of people  
are flying  
a kite in  
the beach

# Text-to-Image Synthesis: Text2Scene

Input Caption	Real Image	SG2IM	HDGAN	AttnGAN	Text2Scene [no inpainting]	Text2Scene
A room with a <b>TV</b> and some different types of <b>couches</b> .						
A tall <b>monitor</b> is near a <b>keyboard</b> and a <b>mouse</b> .						
a <b>car bridge</b> going <b>over</b> a commuter <b>train</b> .						

# Object recognition

## Is it really so hard?

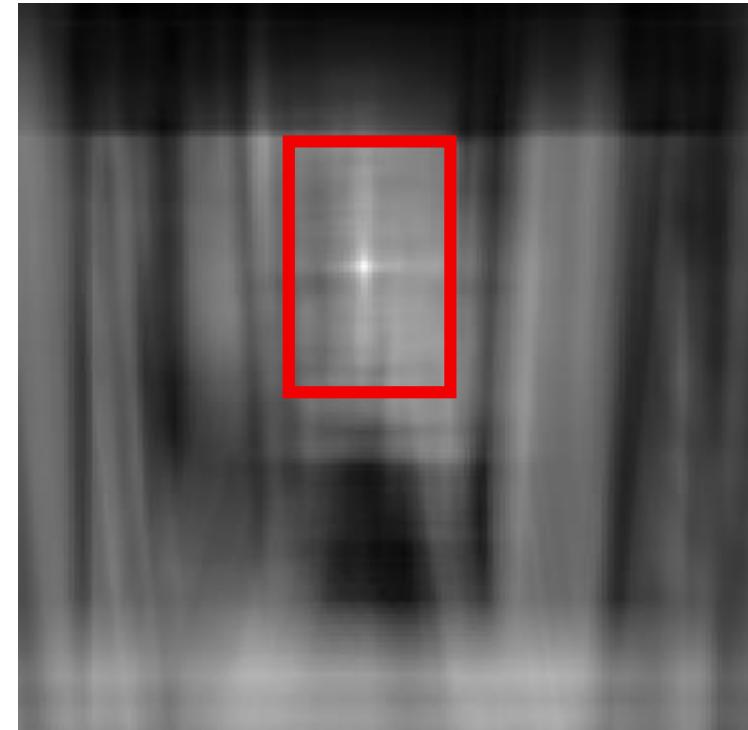
This is a chair



Find the chair in this image



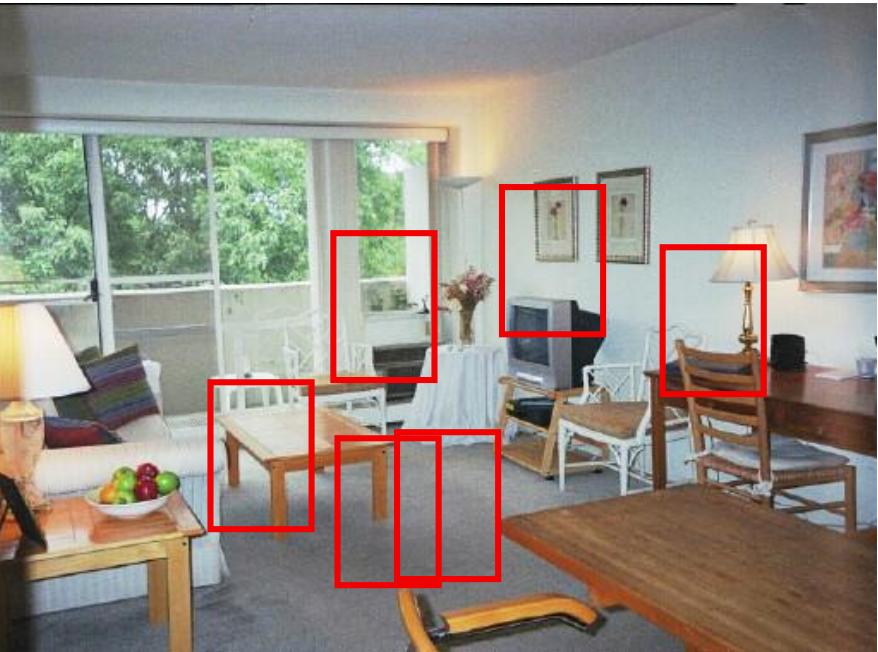
Output of normalized correlation



# Object recognition

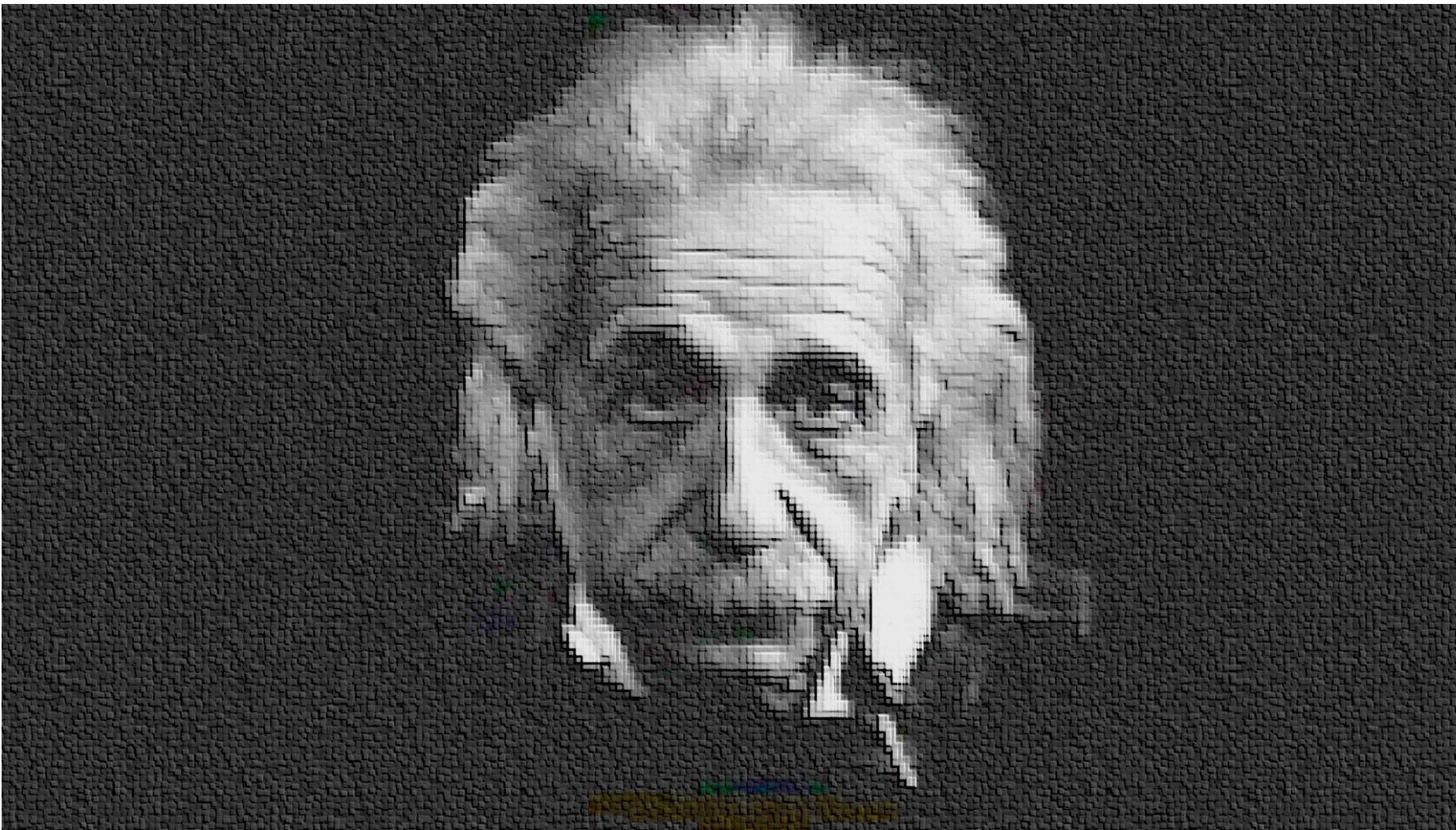
## Is it really so hard?

Find the chair in this image

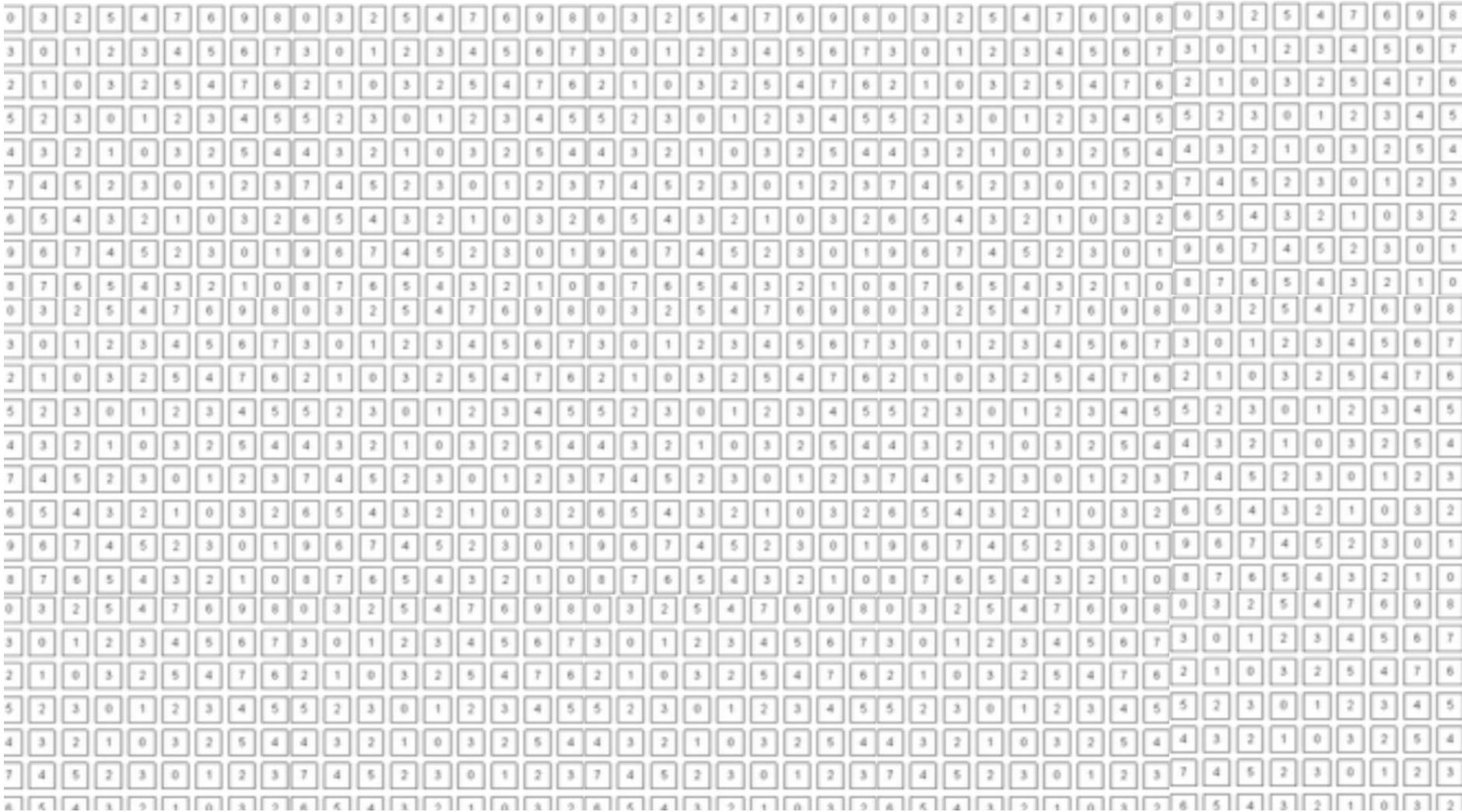


Pretty much garbage  
Simple template matching is not going to make it

This is an image to us



# This is an image to a computer



# Challenges 1: view point variation



Michelangelo 1475-1564



slide by F



## Challenges 2: illumination



slide credit: S. Ullman

# Challenges 3: occlusion



Magritte, 1957

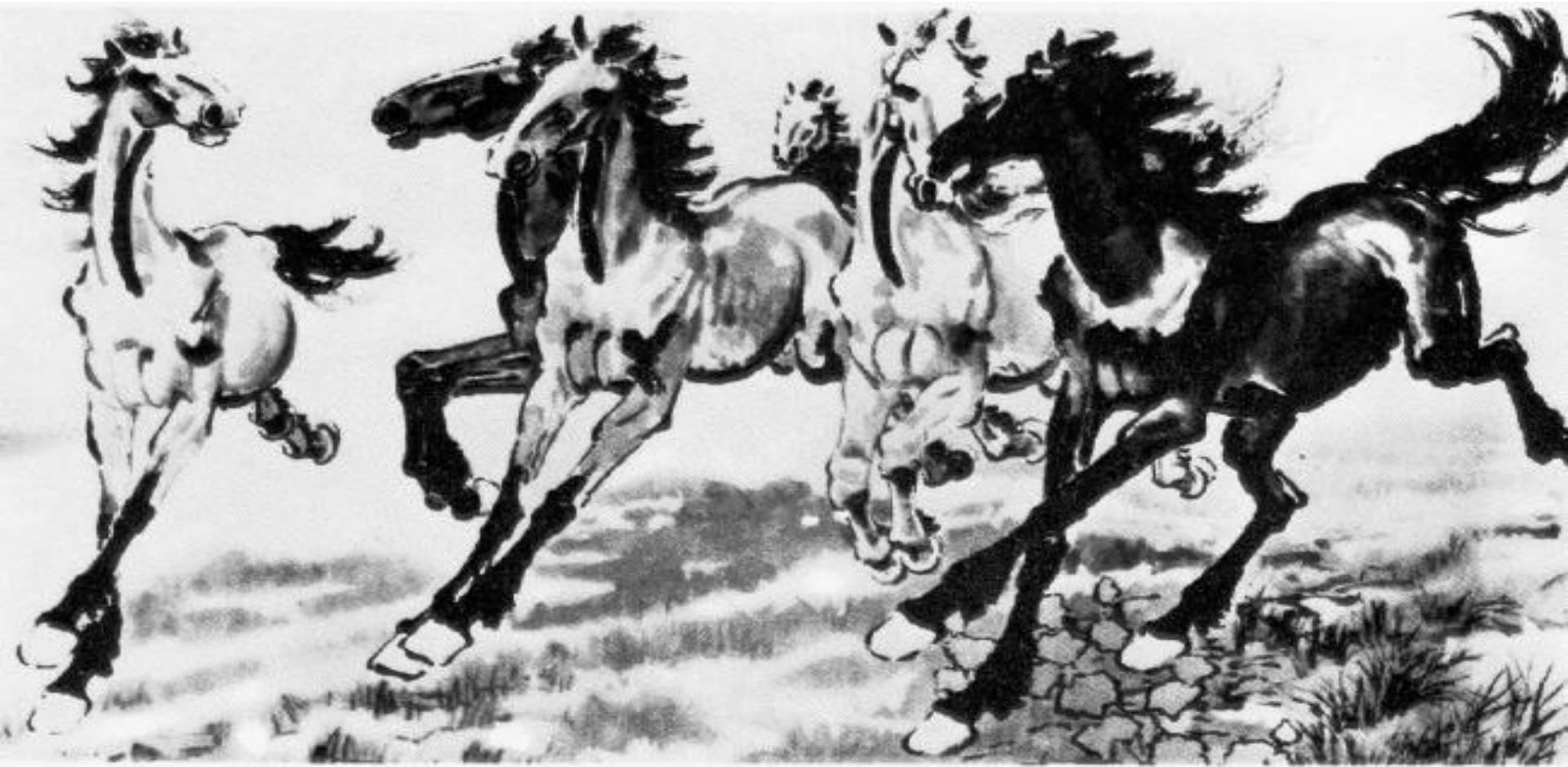
slide by Fei Fei, Fergus & Torralba

## Challenges 4: scale



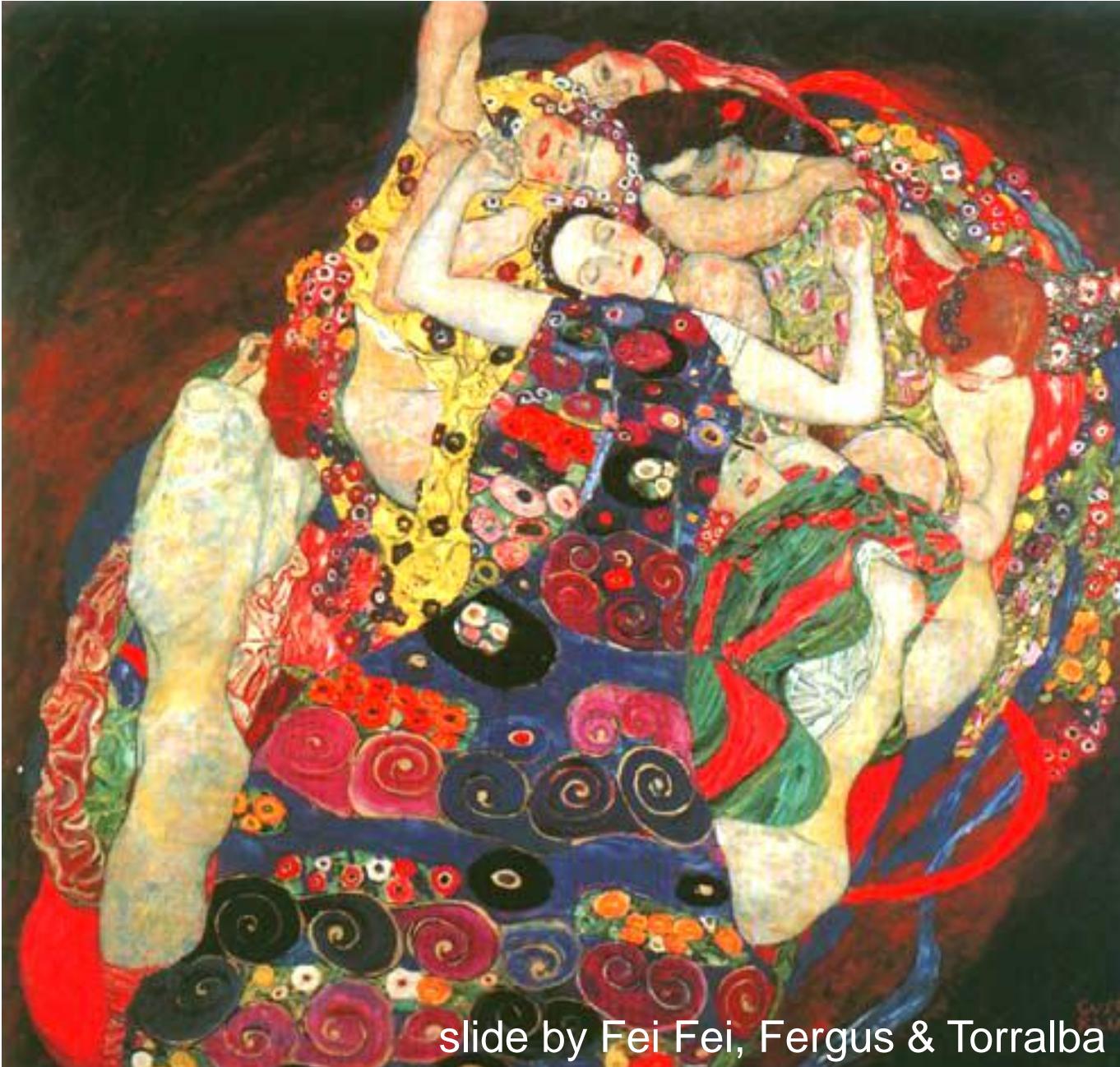
slide by Fei Fei, Fergus & Torralba

# Challenges 5: deformation



Xu, Beihong 1943

## Challenges 6: background clutter



Klimt, 1913

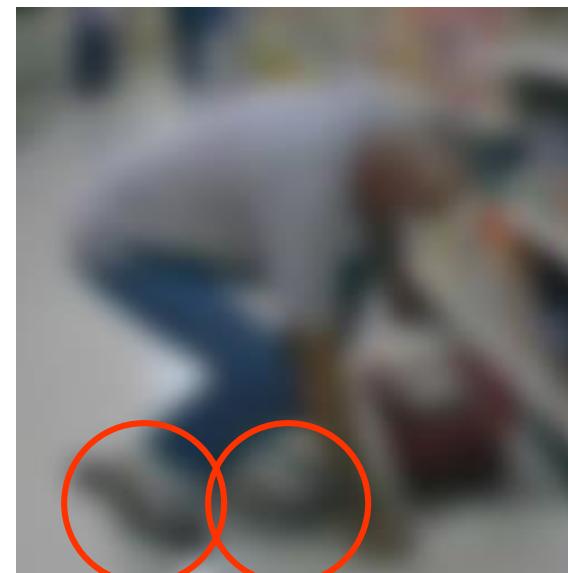
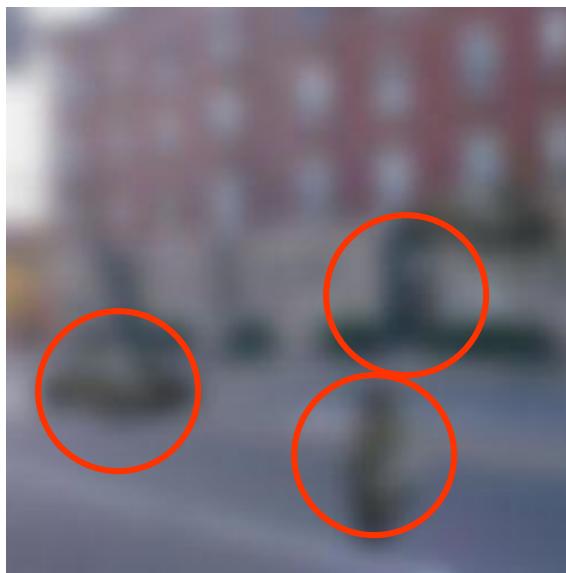
slide by Fei Fei, Fergus & Torralba

# Challenges 7: object intra-class variation



slide by Fei-Fei, Fergus & Torralba

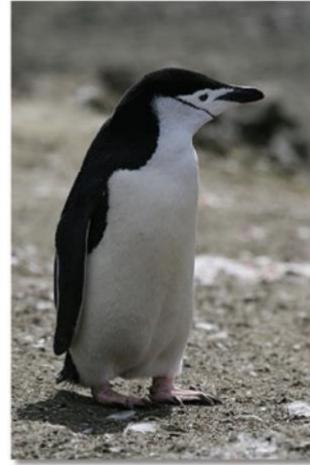
# Challenges 8: local ambiguity



# Challenge 9 – labels are ambiguous

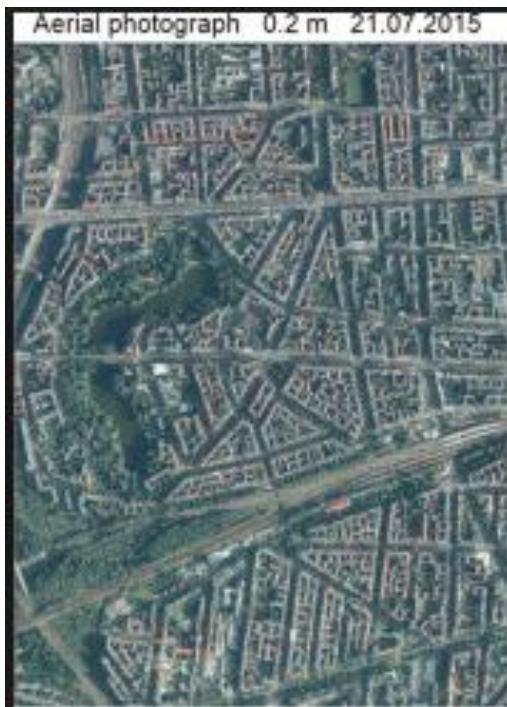


Superordinates: animal, vertebrate  
Basic Level: bird  
Entry Level: bird  
Subordinates: American robin

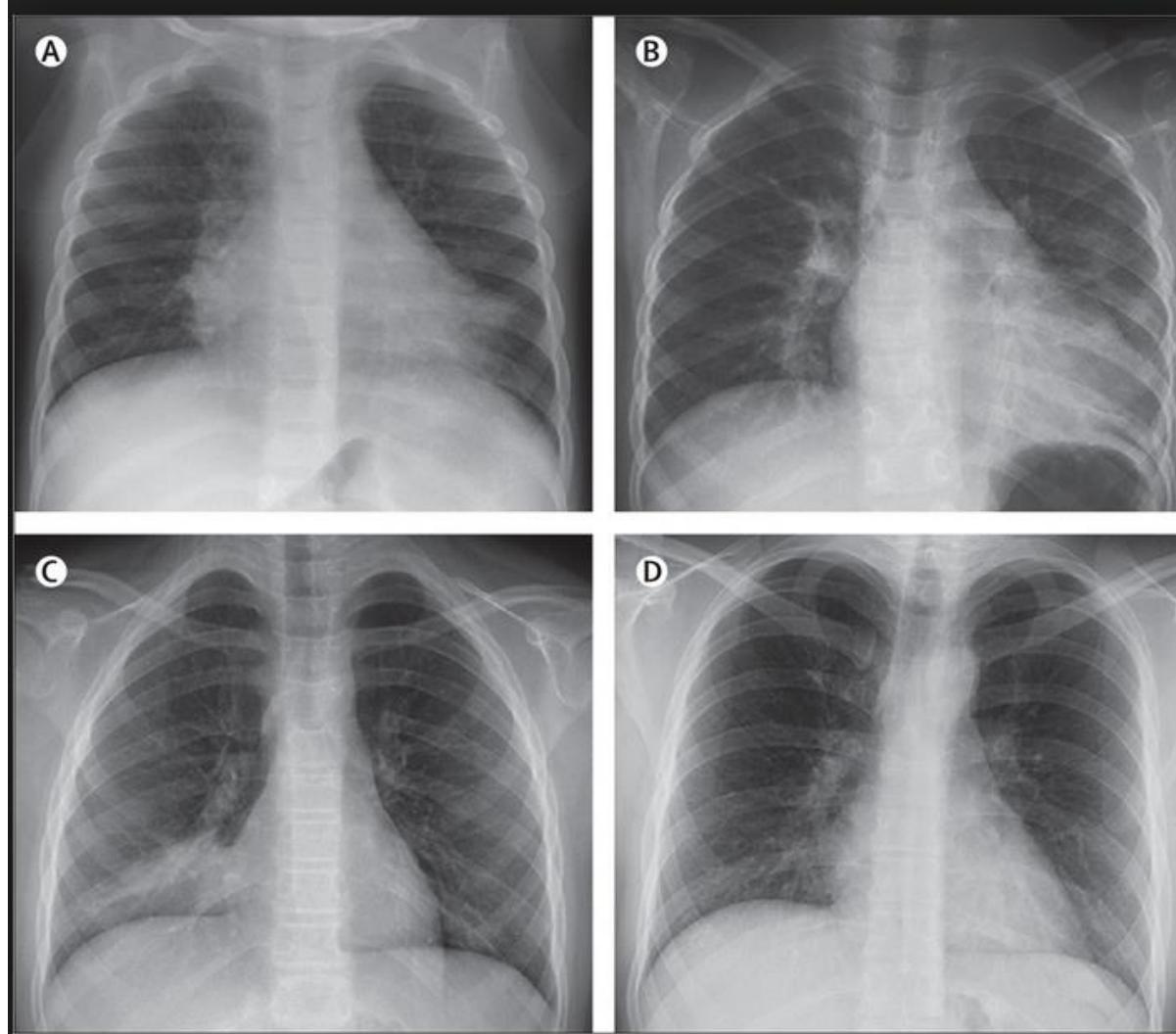


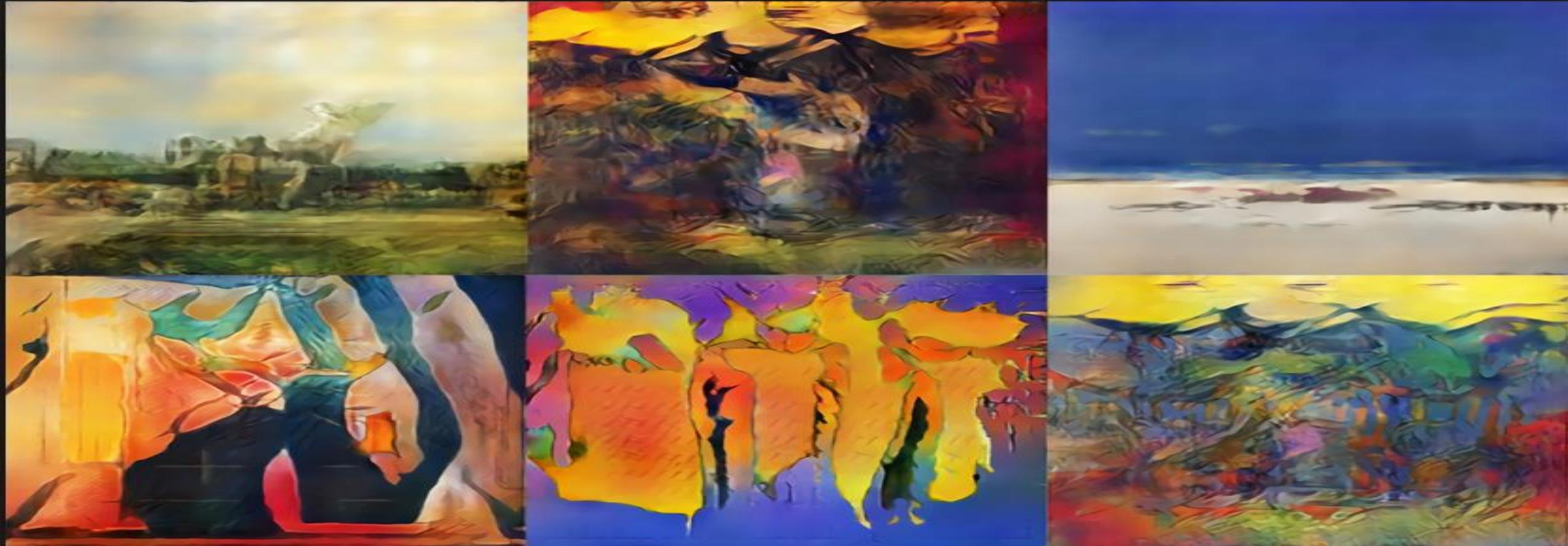
Superordinates: animal, vertebrate  
Basic Level: bird  
Entry Level: penguin  
Subordinates: Chinstrap penguin

# Challenge 10 – different spatial resolution



# Challenge 11 – medical images

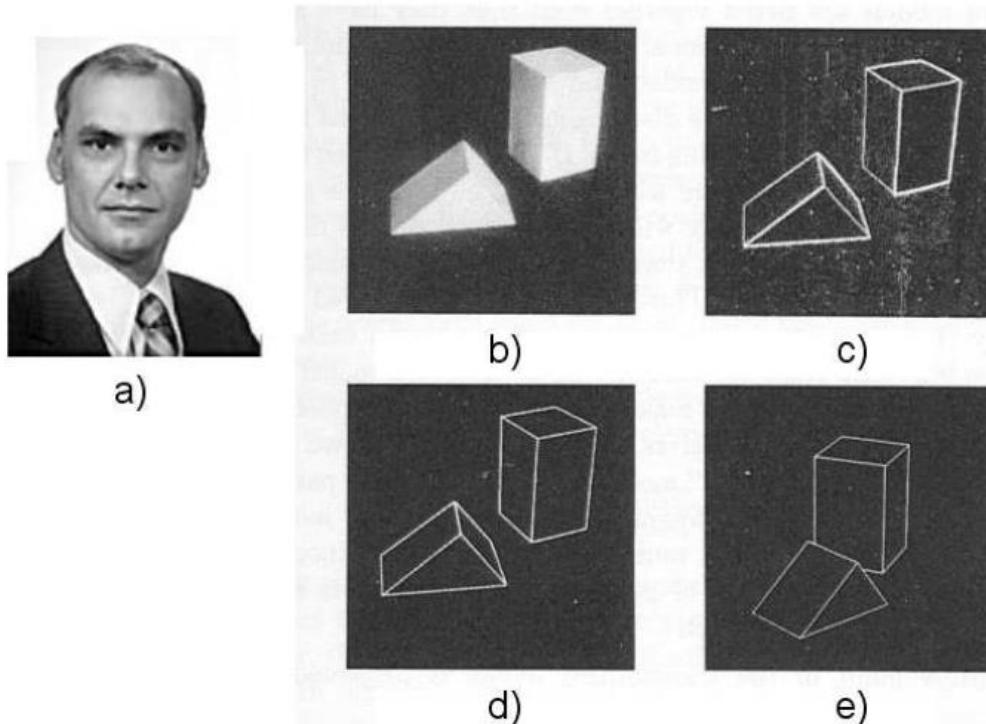




# History of ideas in recognition

- **1960s – early 1990s: the geometric era**
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

# Recognition as an alignment problem: Block world

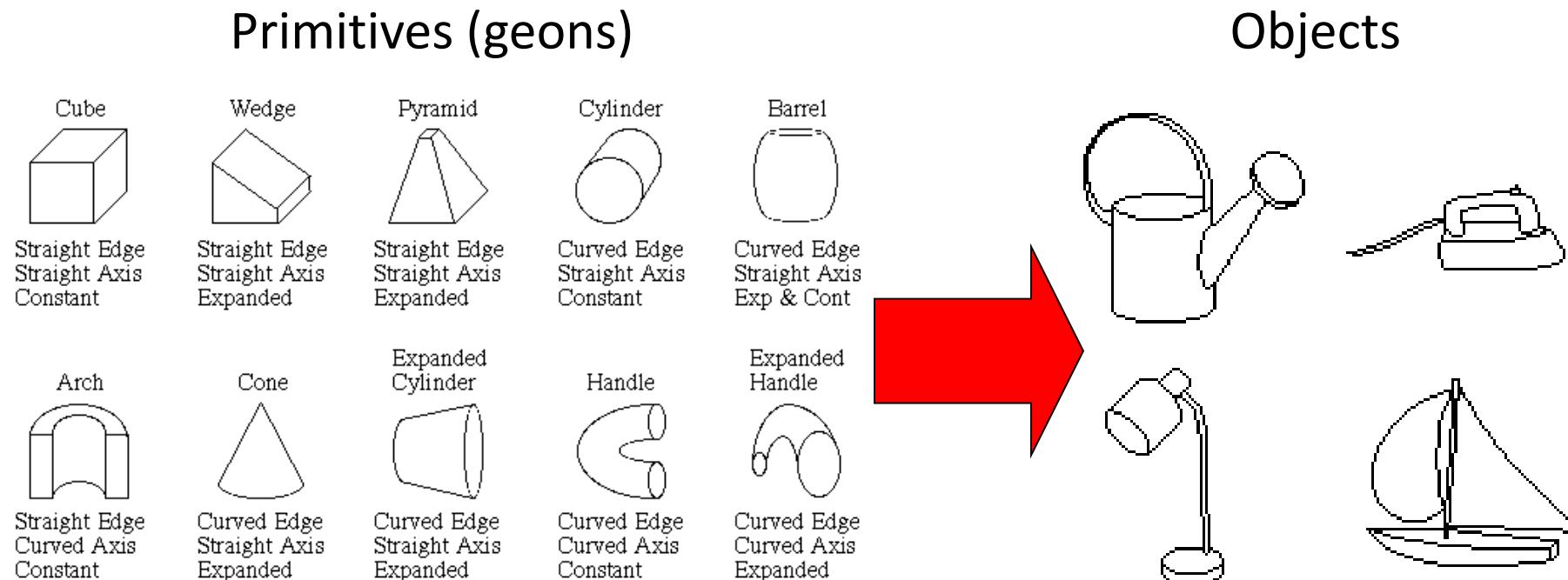


L. G. Roberts, [Machine Perception of Three Dimensional Solids](#), Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

**Fig. 1.** A system for recognizing 3-d polyhedral scenes. a) L.G. Roberts. b) A blocks world scene. c) Detected edges using a  $2 \times 2$  gradient operator. d) A 3-d polyhedral description of the scene, formed automatically from the single image. e) The 3-d scene displayed with a viewpoint different from the original image to demonstrate its accuracy and completeness. (b) - e) are taken from [64] with permission MIT Press.)

# Recognition by components

Biederman (1987)

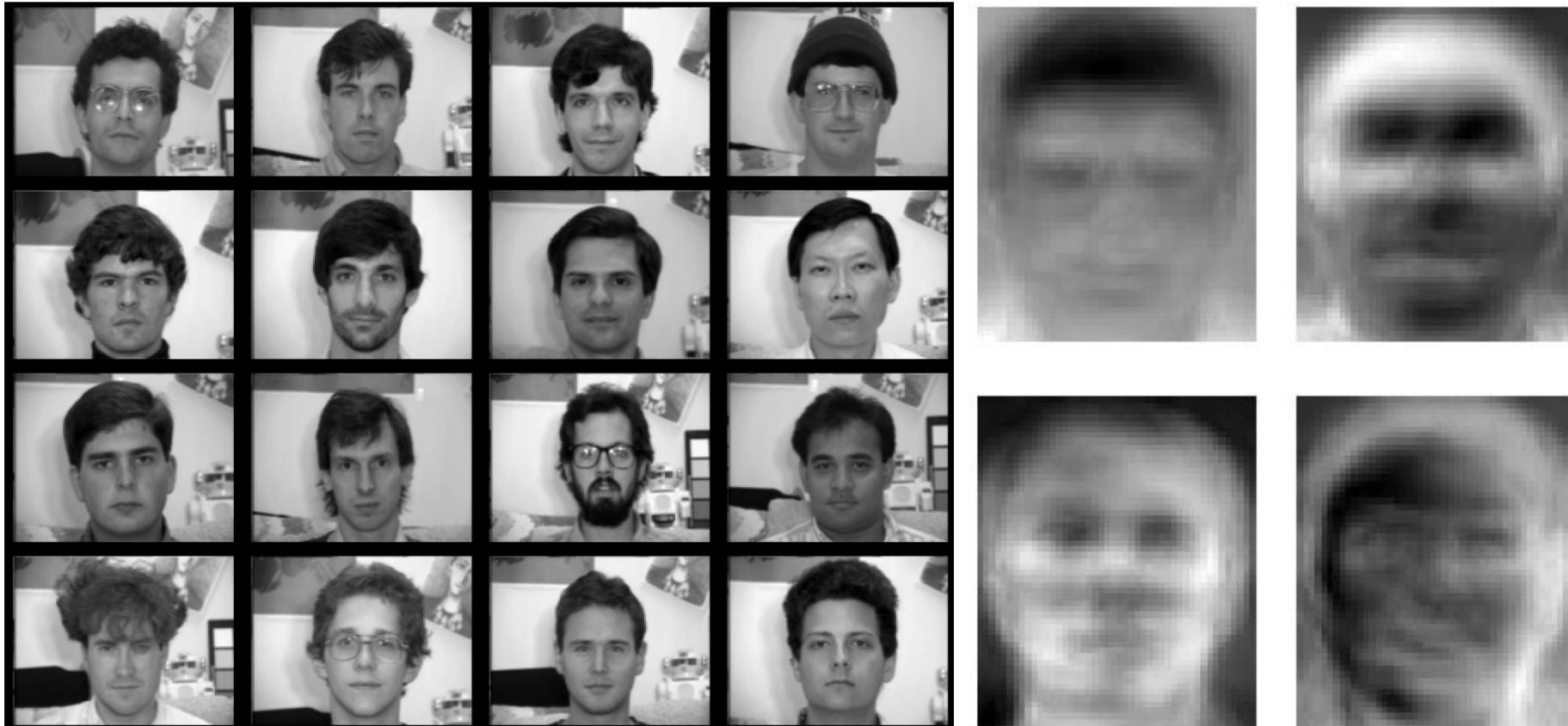


[http://en.wikipedia.org/wiki/Recognition\\_by\\_Components\\_Theory](http://en.wikipedia.org/wiki/Recognition_by_Components_Theory)

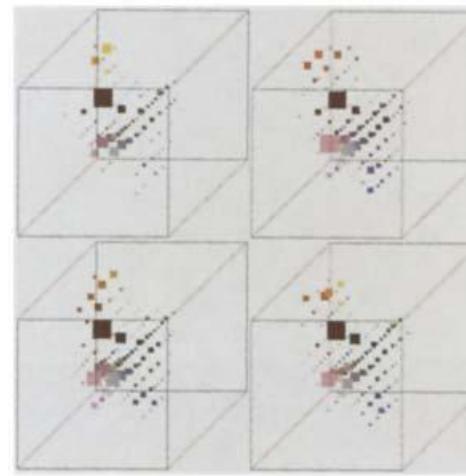
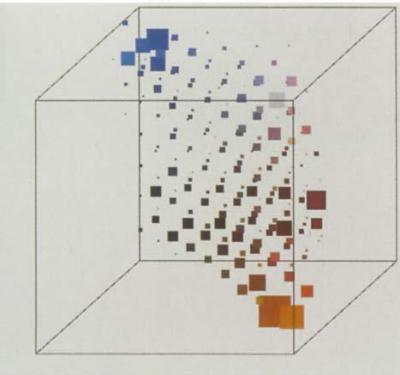
# History of ideas in recognition

- 1960s – early 1990s: the geometric era
- **1990s: appearance-based models**
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

# Eigenfaces (Turk & Pentland, 1991)



# Color Histograms



Swain and Ballard, [Color Indexing](#), IJCV 1991.

Svetlana Lazebnik

# History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- **Mid-1990s: sliding window approaches**
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

# Sliding window approaches



# History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- **Late 1990s: local features**
- Early 2000s: parts-and-shape models
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

# Local features for object instance recognition



D. Lowe (1999, 2004)

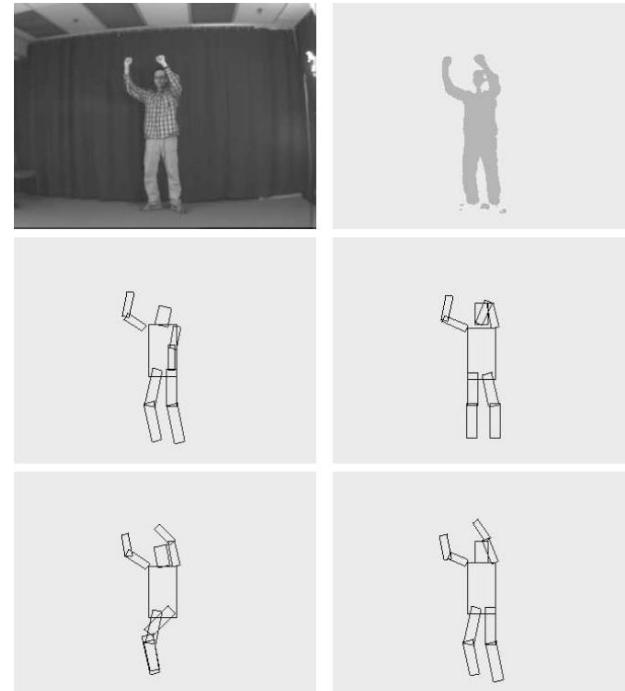
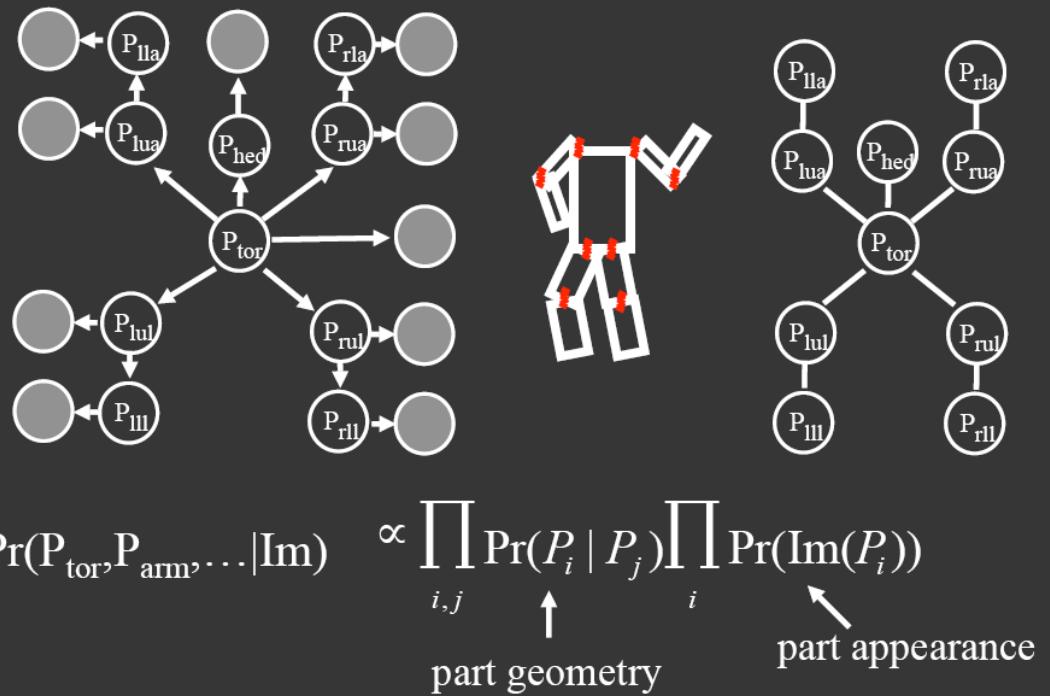
# History of ideas in recognition

- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- **Early 2000s: parts-and-shape models**
- Mid-2000s: bags of features
- Present trends: combination of local and global methods, data-driven methods, context

# Representing people

## Pictorial structure model

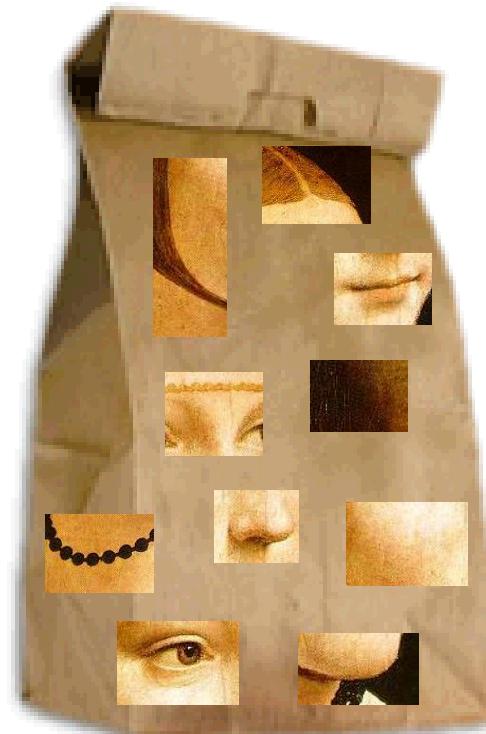
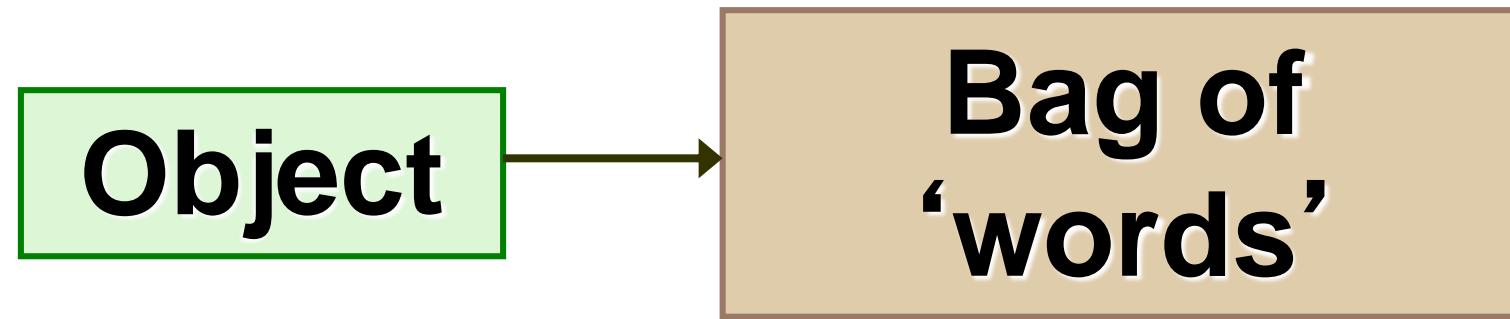
Fischler and Elschlager(73), Felzenszwalb and Huttenlocher(00)



# History of ideas in recognition

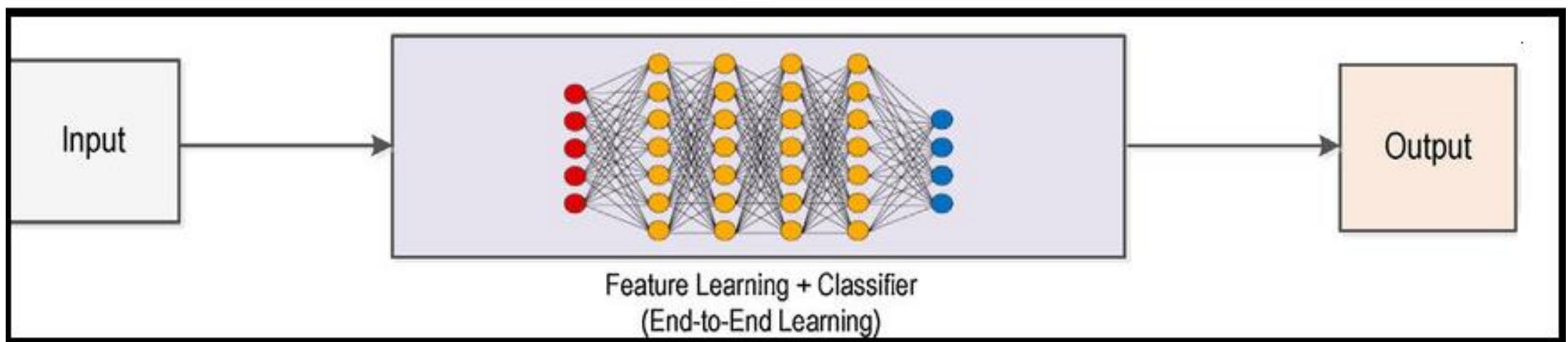
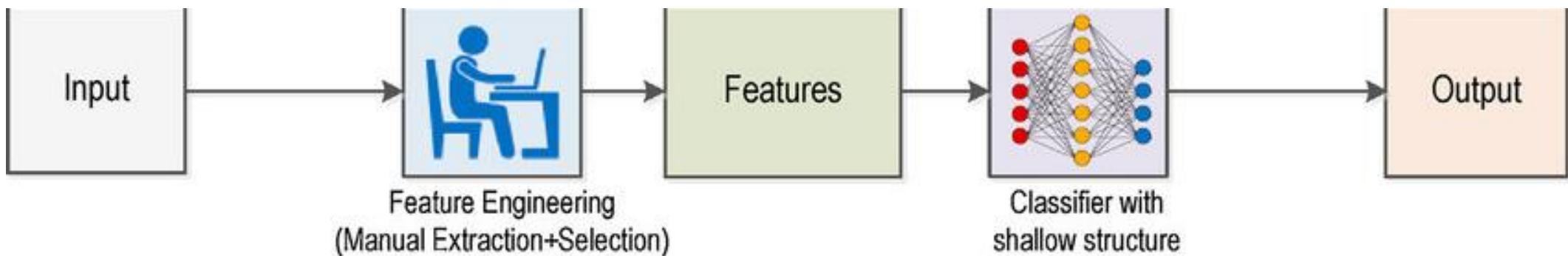
- 1960s – early 1990s: the geometric era
- 1990s: appearance-based models
- Mid-1990s: sliding window approaches
- Late 1990s: local features
- Early 2000s: parts-and-shape models
- **Mid-2000s: bags of features**
- Present trends: combination of local and global methods, data-driven methods, context

# Bag-of-words models



# Deep Learning and Vision

- Deep Learning has been a great disruption into the field of Computer Vision. Has made a lot of new things work!
- Many deep learning methods being applied to vision these days.
- This is a deep learning course for visual inference. We will briefly review some pre-deep learning methods, and then mostly deep learning.



# Objectives

- Understanding foundational concepts for representation learning using neural networks
- Becoming familiar with state-of-the-art models for tasks such as image classification, object detection, image segmentation, scene recognition, etc.
- Obtain practical experience in the implementation of visual recognition models using deep learning.

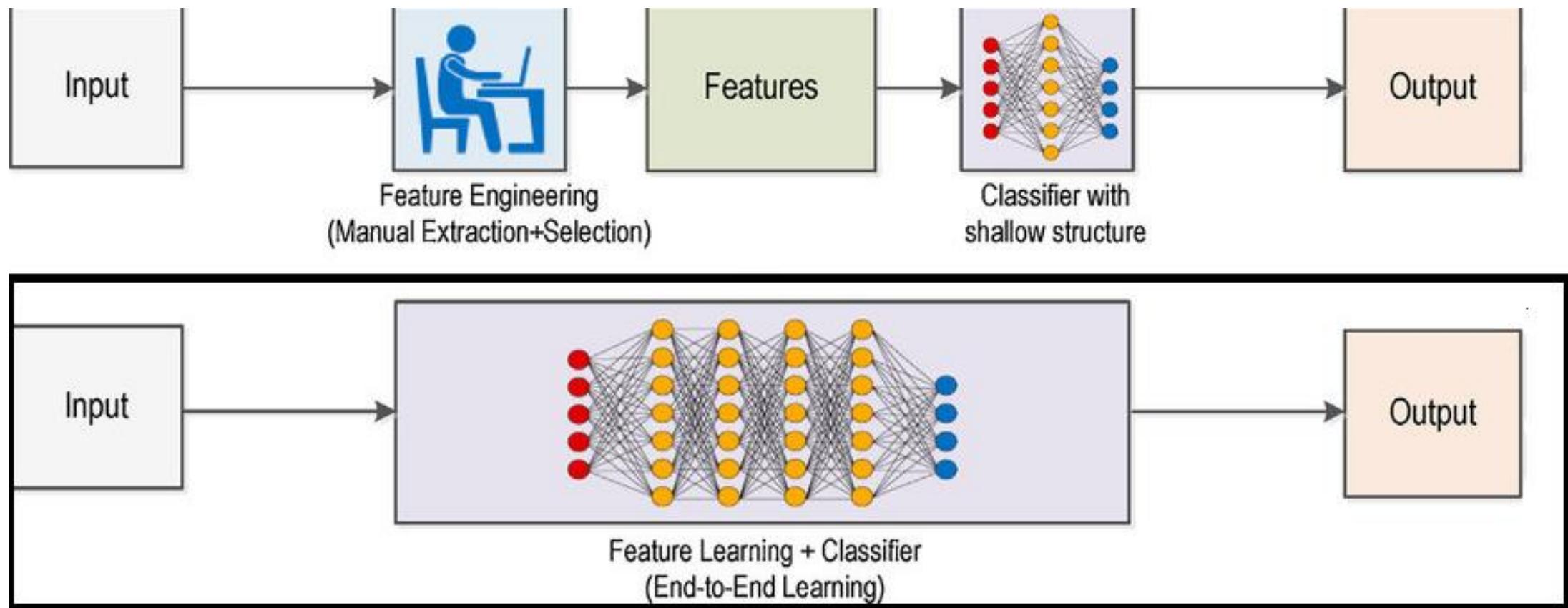
# Resources

- Online lectures (Stanford, MIT, NYU, IITM)
- Deep learning book by GoodFellow
- <http://neuralnetworksanddeeplearning.com/>
- Deep Learning Methods and Applications by Deng & Yu
- <https://www.purestorage.com/content/dam/purestorage/pdf/whitepapers/oreilly-deep-learning-book.pdf>
- Papers

# Pytorch, Tensorflow, Keras

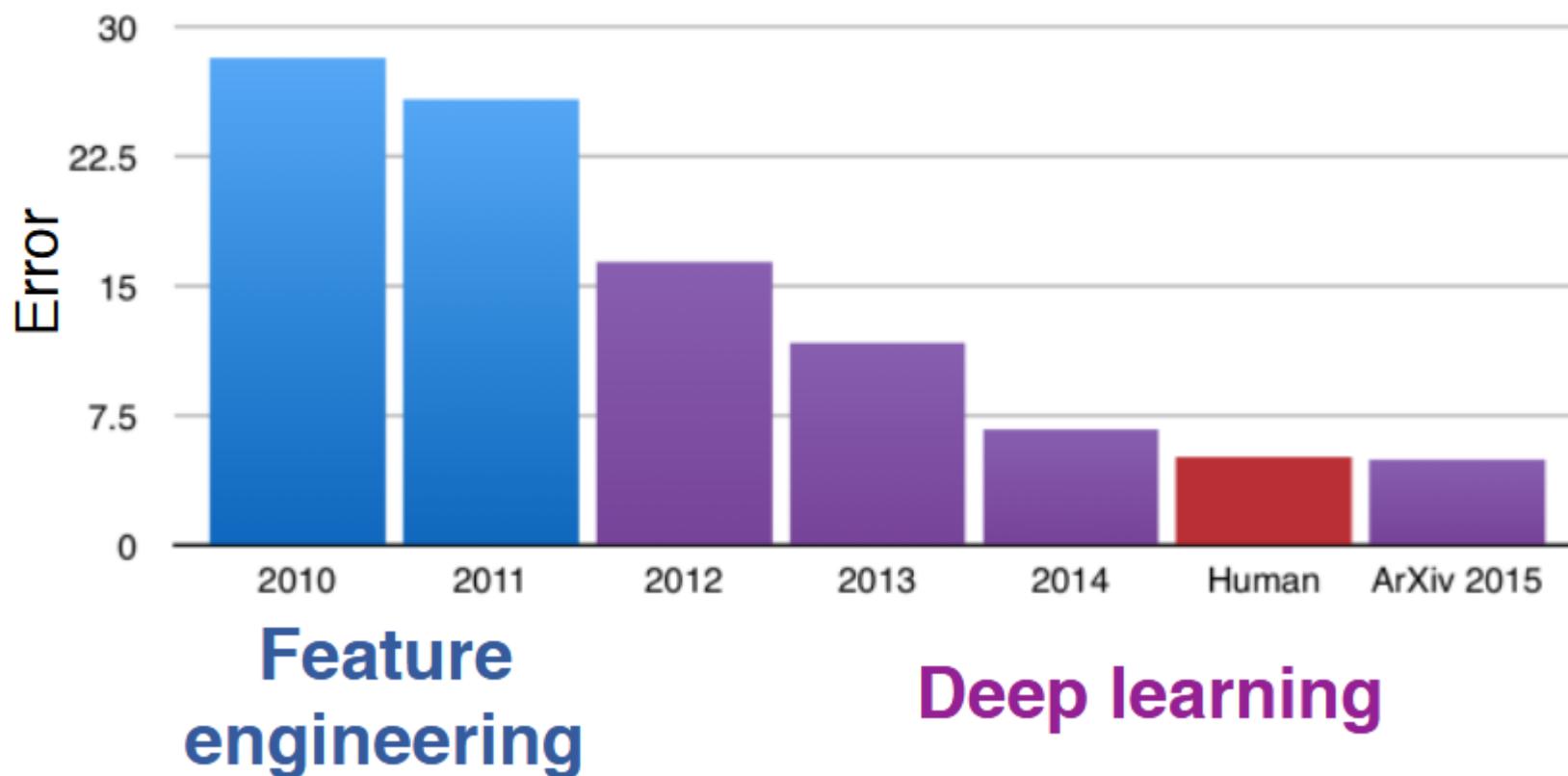
- Pytorch – zero to GAN
  - <https://www.youtube.com/watch?v=GlsG-ZUy0MY>
- Keras by deepLizard
  - <https://www.youtube.com/watch?v=qFJeN9V1Zsl>
- Tensorflow tutorial
  - <https://www.youtube.com/watch?v=tPYj3fFJGjk>

# Traditional ML vs Deep Learning (Recap)



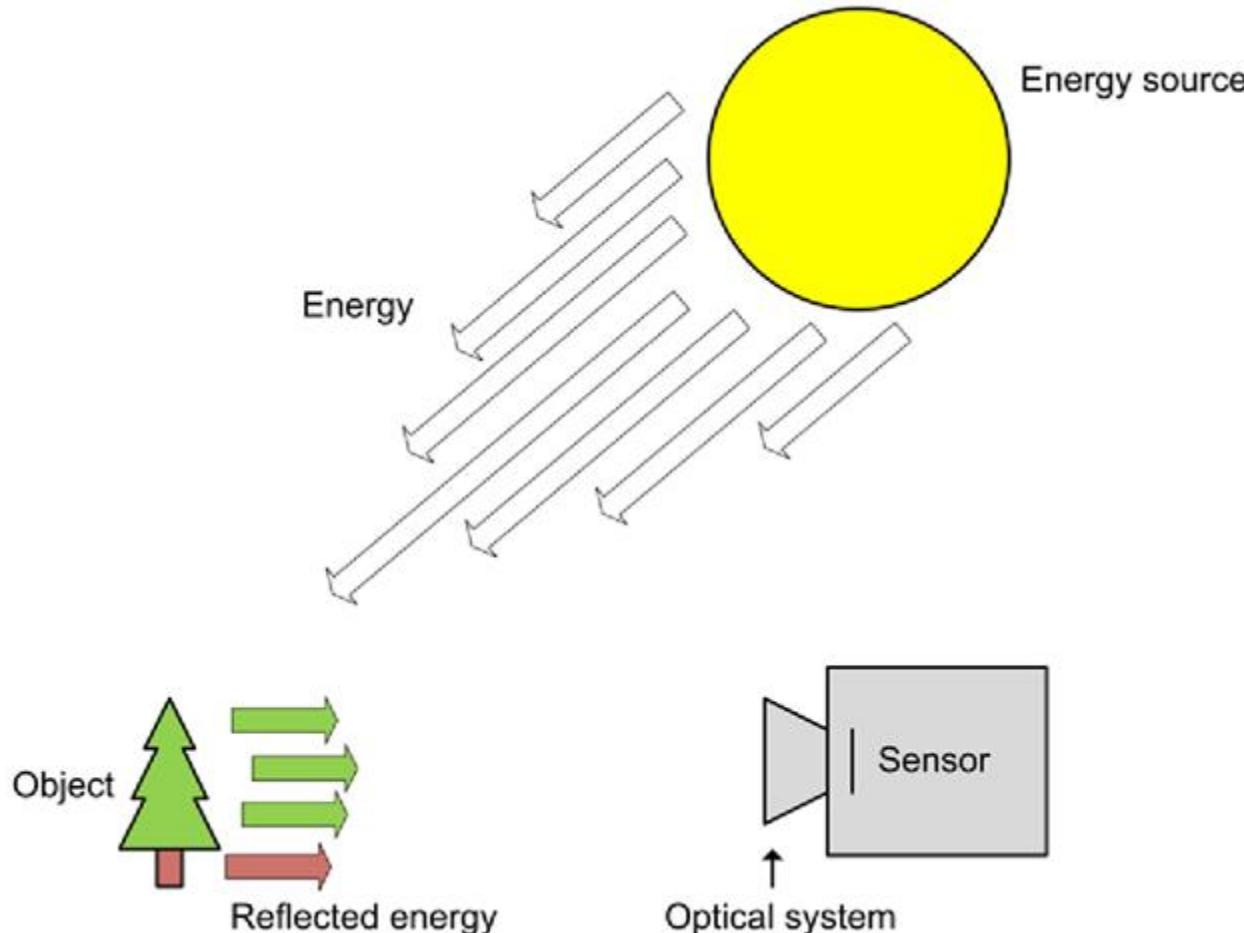
# Feature learning

Object classification accuracy on ImageNet (ILSVRC)

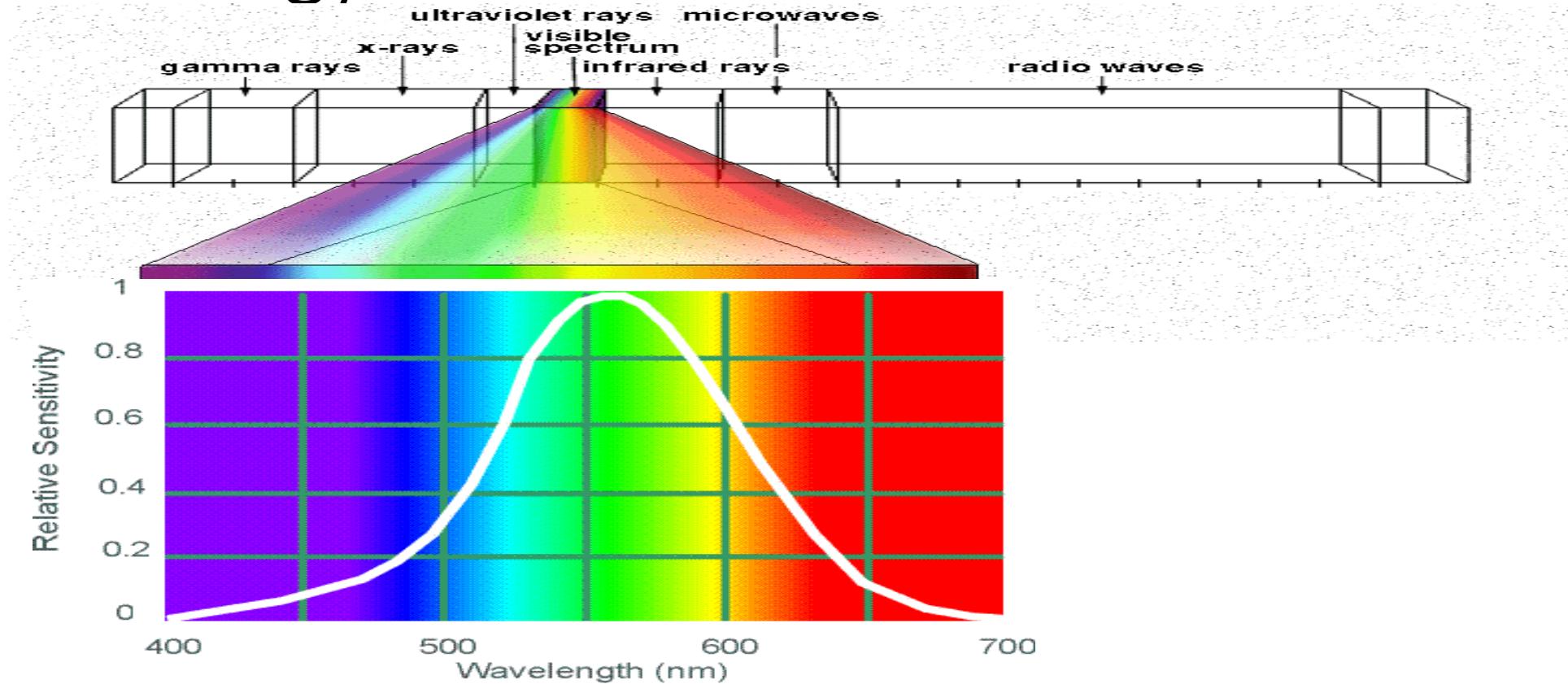


# Image acquisition

$$z = f(x, y)$$

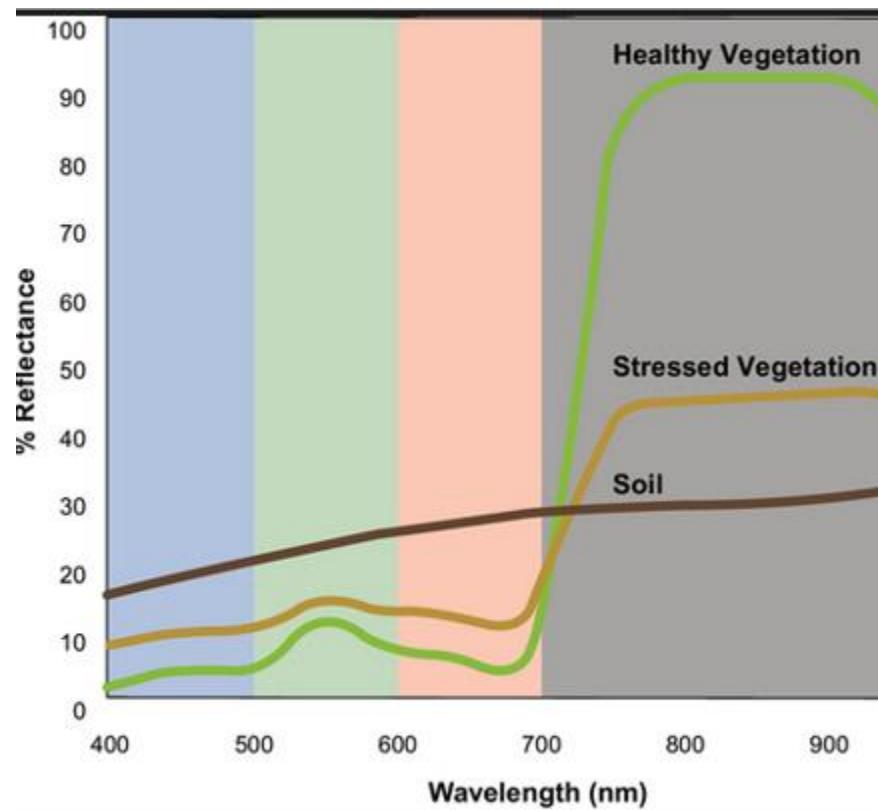


# Electromagnetic Spectrum – characterizing the energy



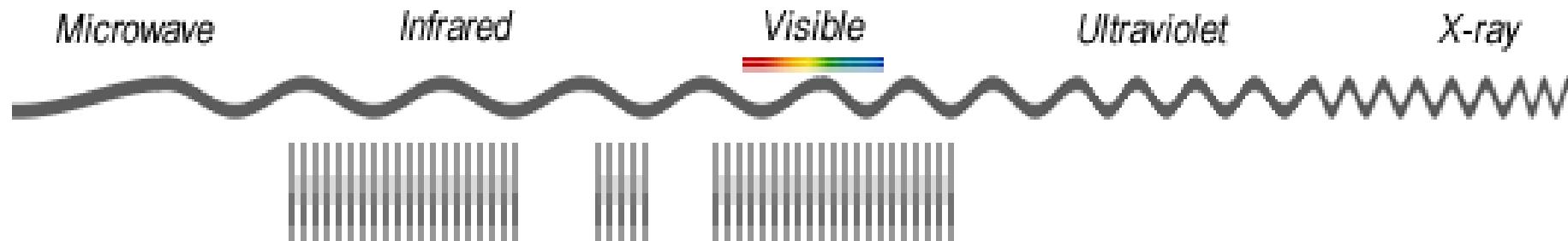
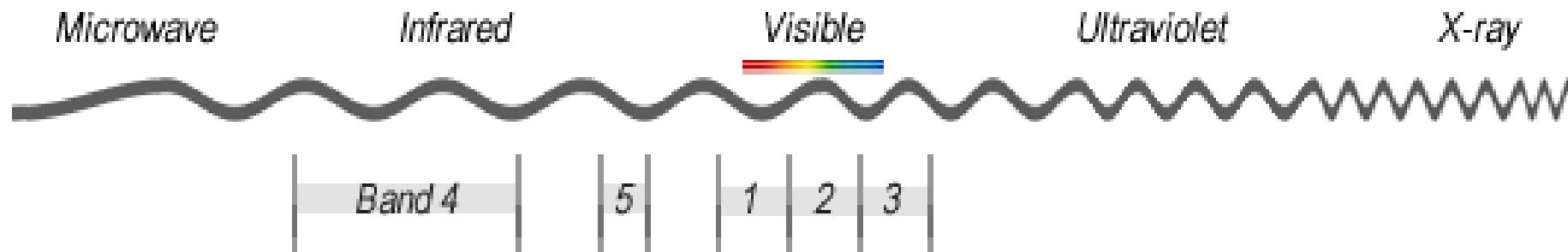
Human Luminance Sensitivity Function

# Beyond visible spectra – satellite images

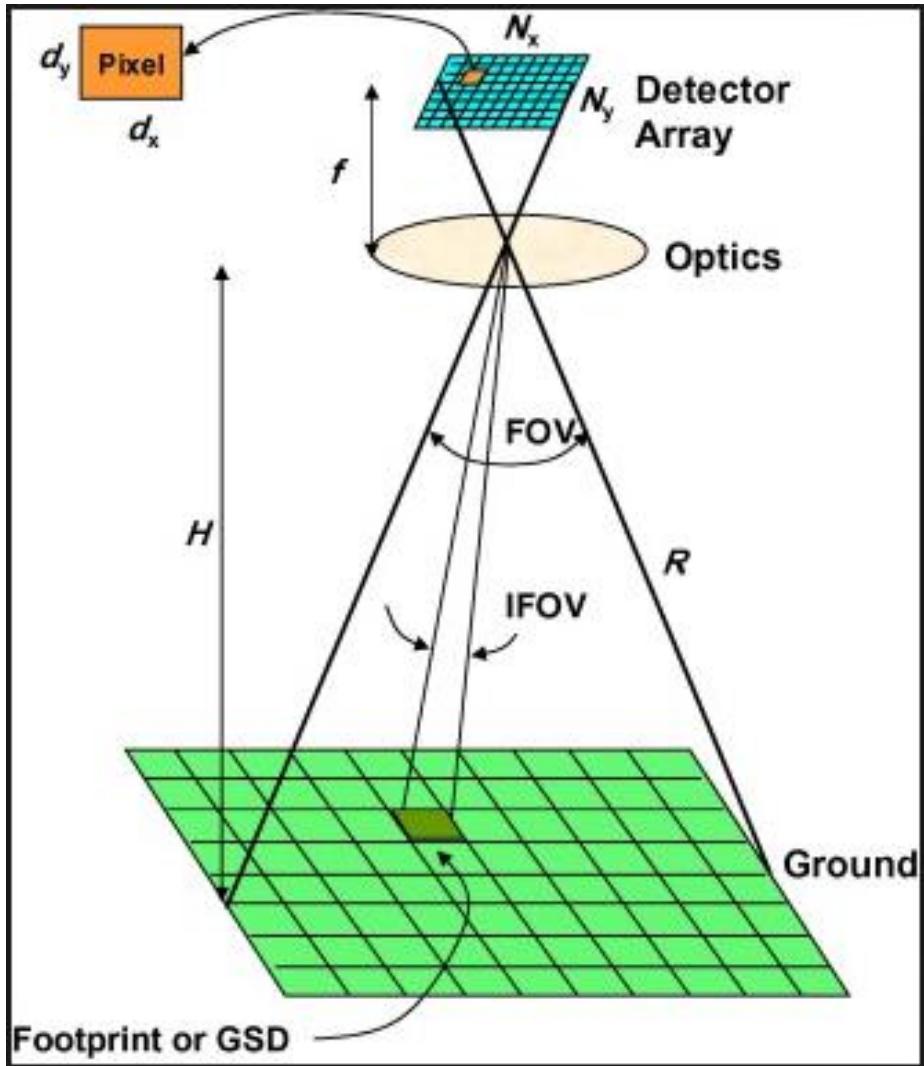


The Infra-red region is better in characterizing the classes than the visible region

# Multi-spectral and hyper-spectral



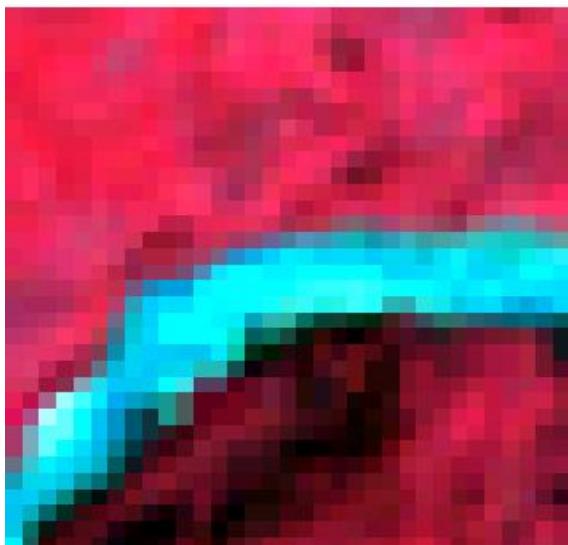
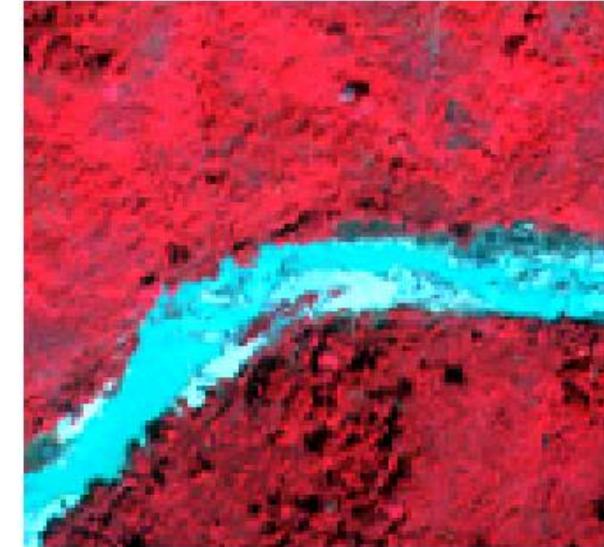
# Spatial resolution



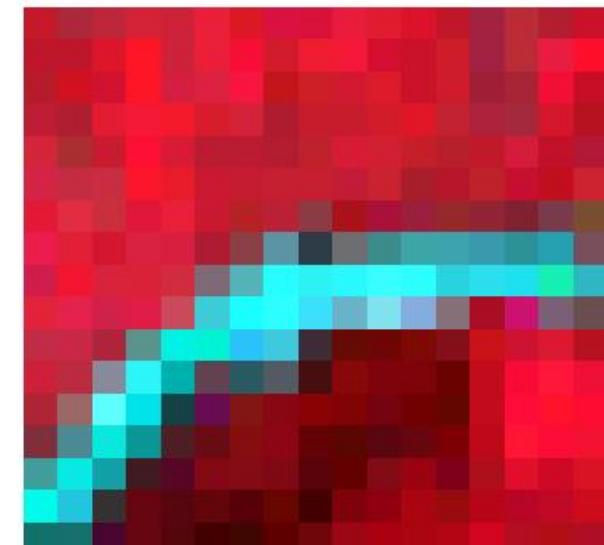
IKONOS 1m panchromatic merge 1m resolution



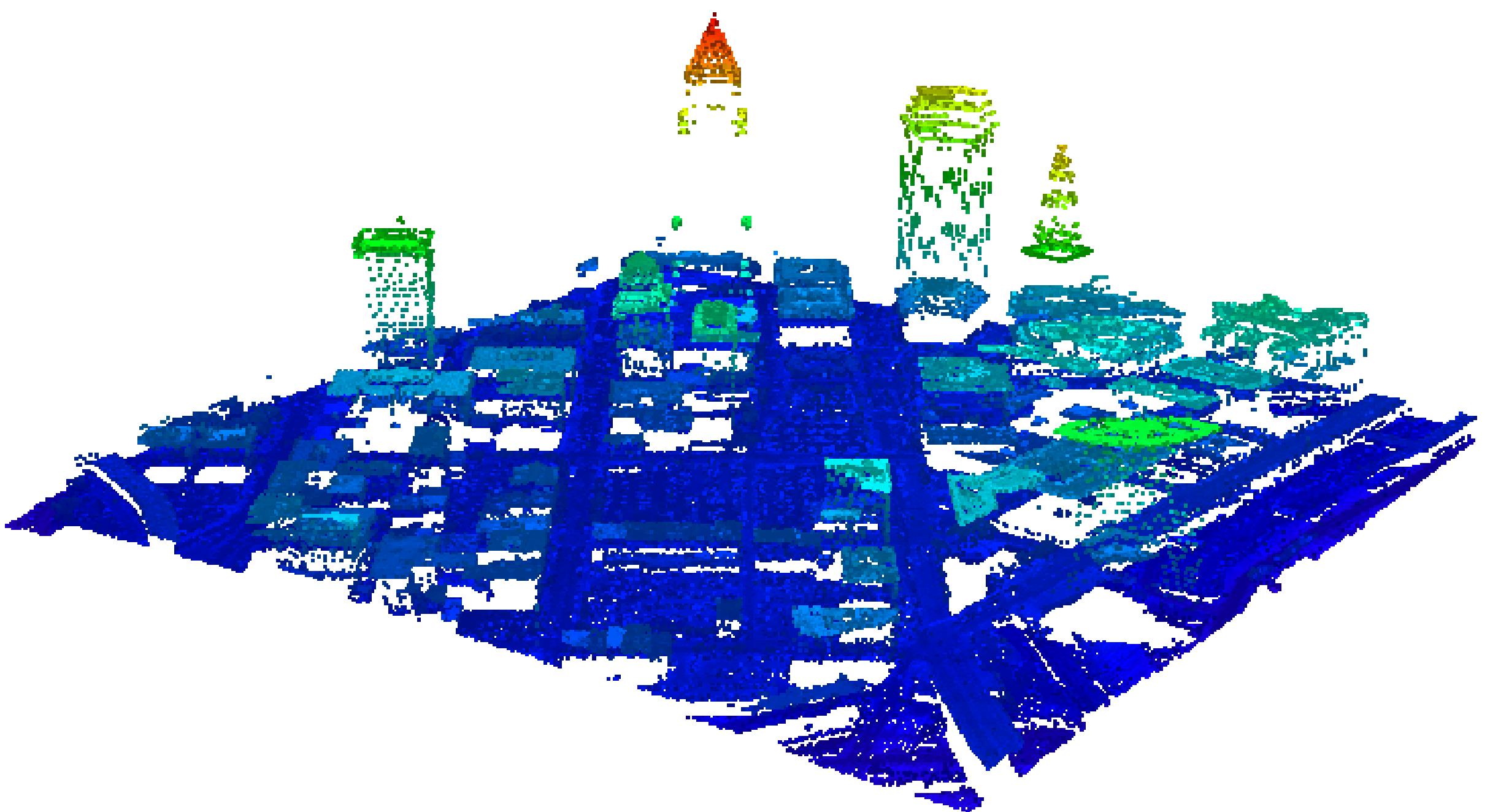
IKONOS multispectral 4m resolution



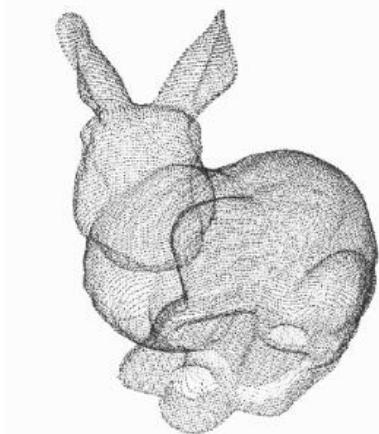
ASTER 15m resolution



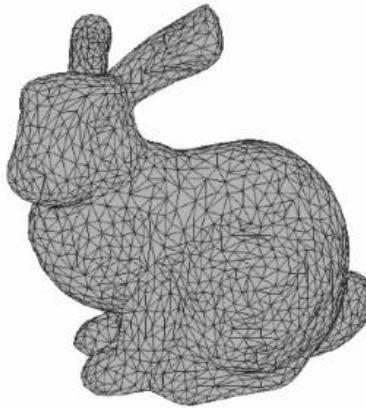
Landsat ETM+ 30m resolution



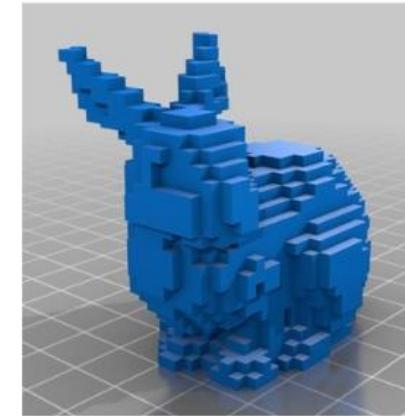
# Different rep. of 3D information in images



Point Cloud



Mesh



Volumetric



Projected View  
RGB(D)

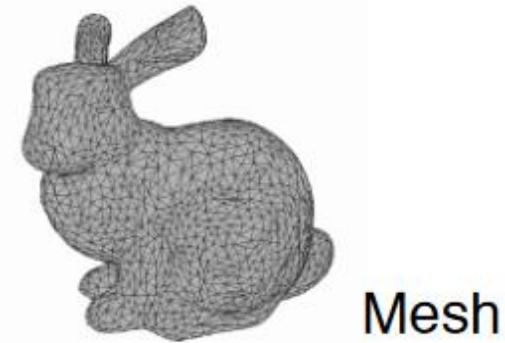
...

# Point cloud is close to raw sensor data

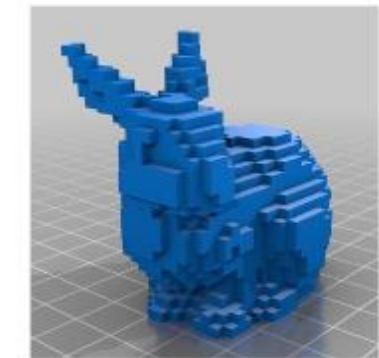
## Point cloud is canonical



**Point Cloud**



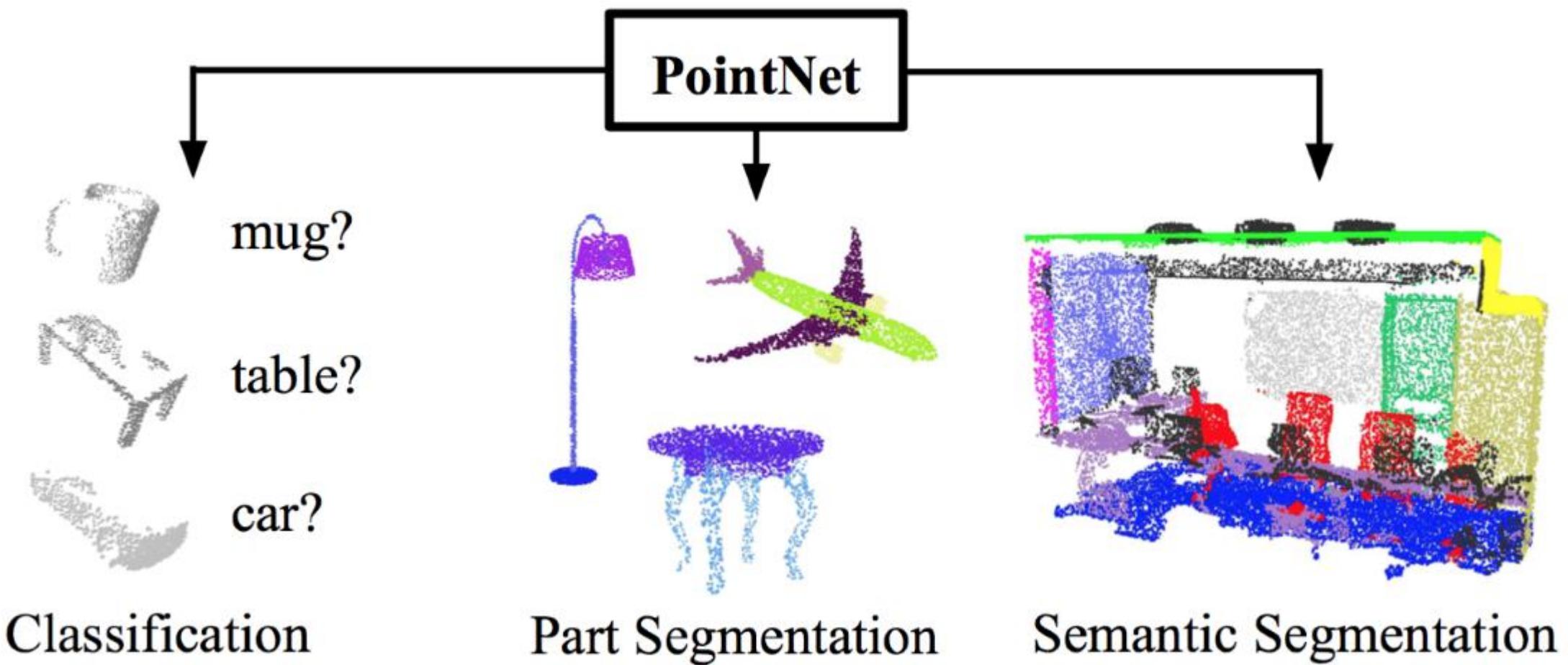
Mesh



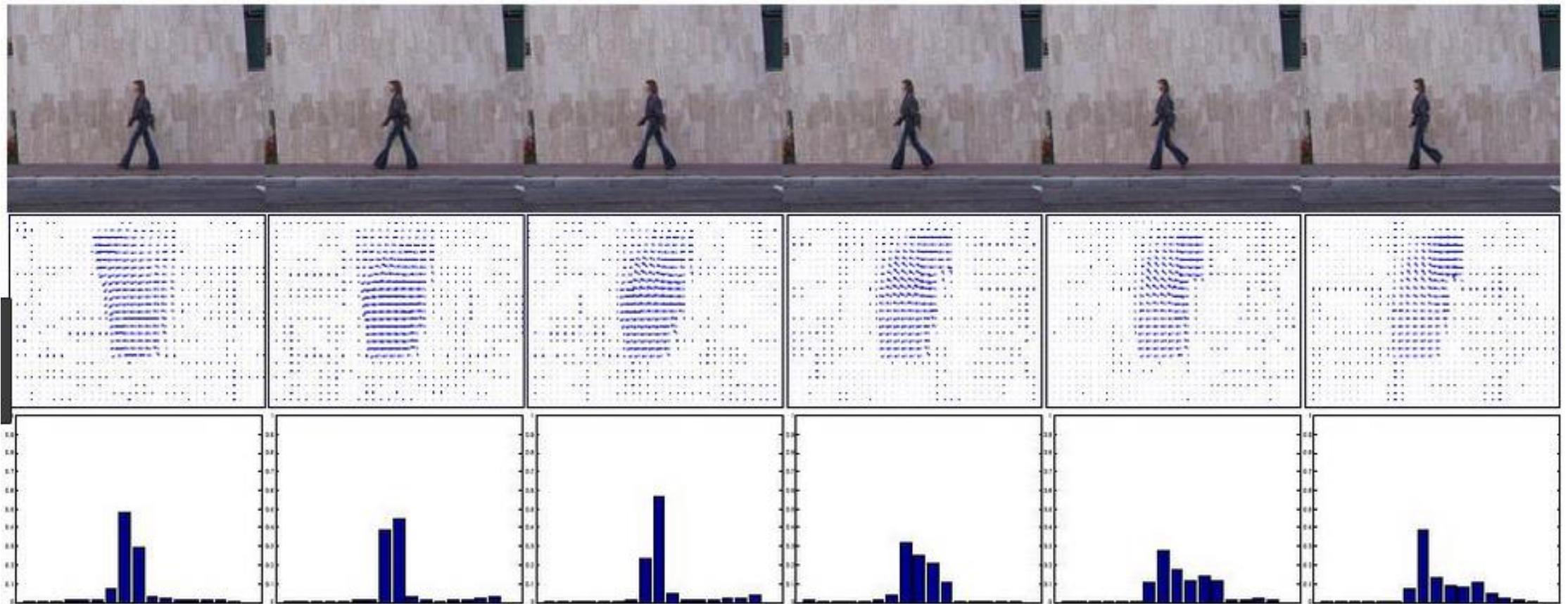
Volumetric



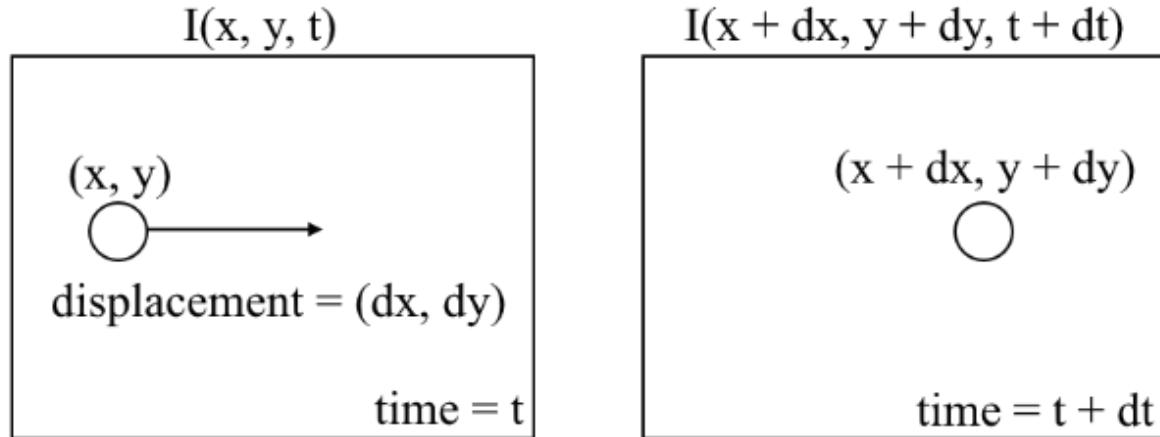
Depth Map



# Video data – appearance + motion (optical flow)



# Optical flow - a quick review



$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t)$$

$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \delta t + \dots$$

$$\Rightarrow \frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \delta t = 0$$

$$\frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t} = 0$$

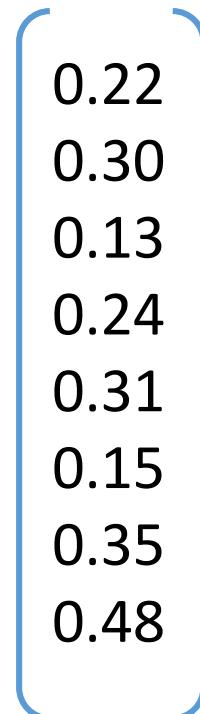
# Color coding the optical flow





# What are Image Features?

Image



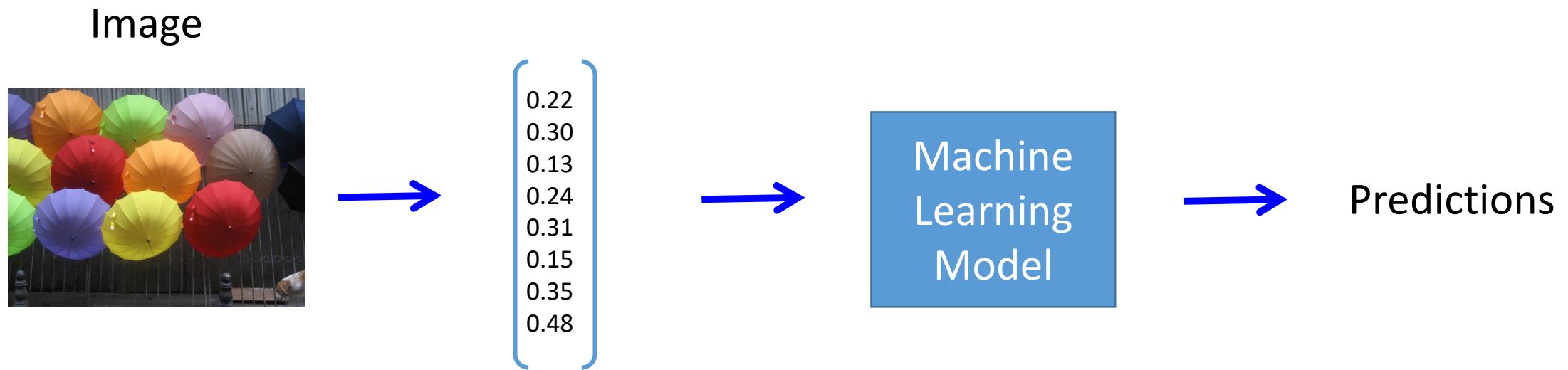
Color histogram?

Maximum color on sub-  
areas of the image?

Any statistics on the  
input image?

The output of some  
image processing on  
the input image?

# Why are they useful?



As inputs to a machine learning model

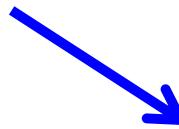
# Why are they useful?



[0.22  
0.30  
0.13  
0.24  
0.31  
0.15  
0.35  
0.48]

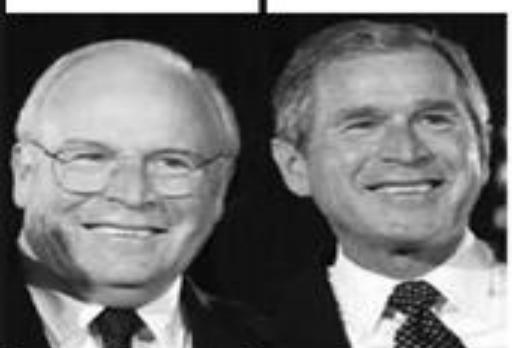


[0.24  
0.34  
0.23  
0.27  
0.63  
0.15  
0.25  
0.48]



Distance function  
(e.g. Euclidean  
distance)

To compare images (i.e. retrieve similar images)

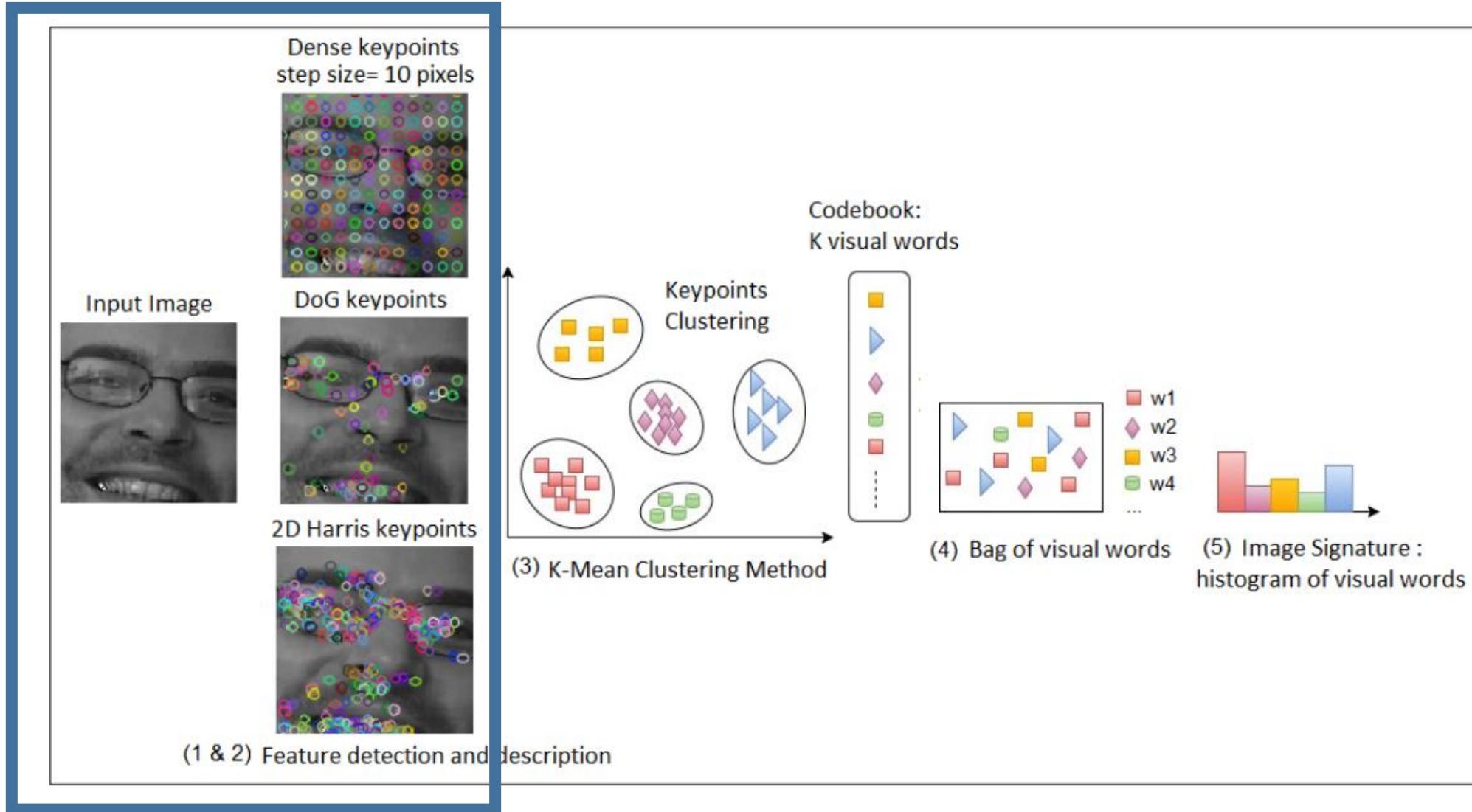


Recognition by Global Features

Images Reconstructed by Low Frequency Fourier Coefficients

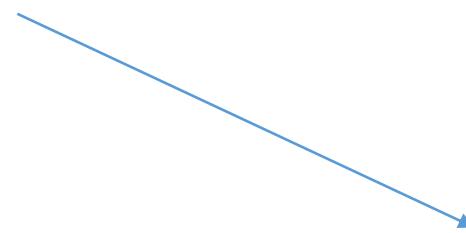


# Feature detection and description



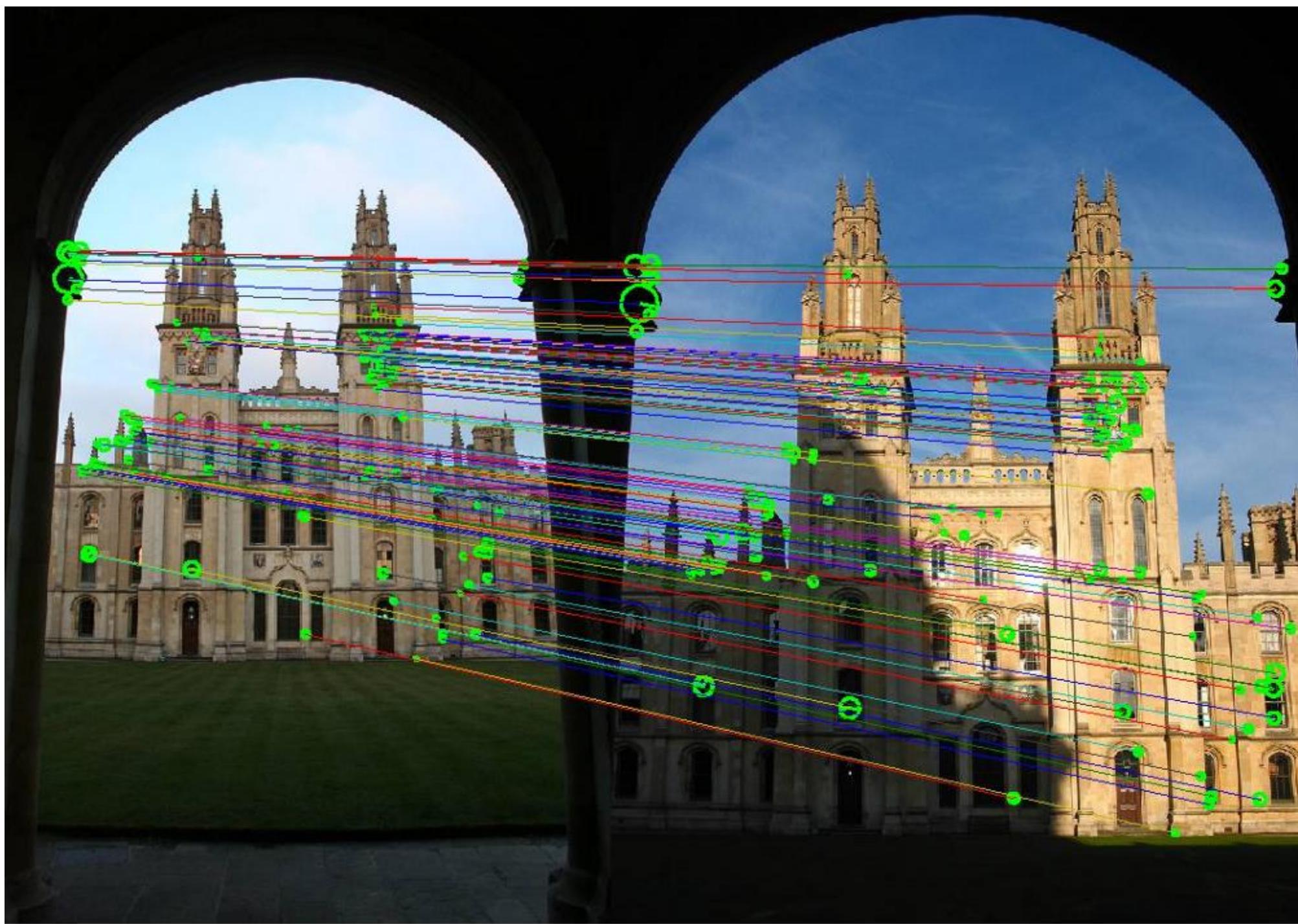
# Visual features

- Color histogram
- Edge & boundary feature
- Shape feature
- Texture feature
- Interests points based feature
- **Deep features**



Invariant to transformations

- Scale
- Rotation
- Translation
- Affine transformation
- Illumination change
- Some more



# Image Features: Color



# Image Features: Color



80 million tiny images: a large dataset for  
non-parametric object and scene recognition

Antonio Torralba, Rob Fergus and William T. Freeman

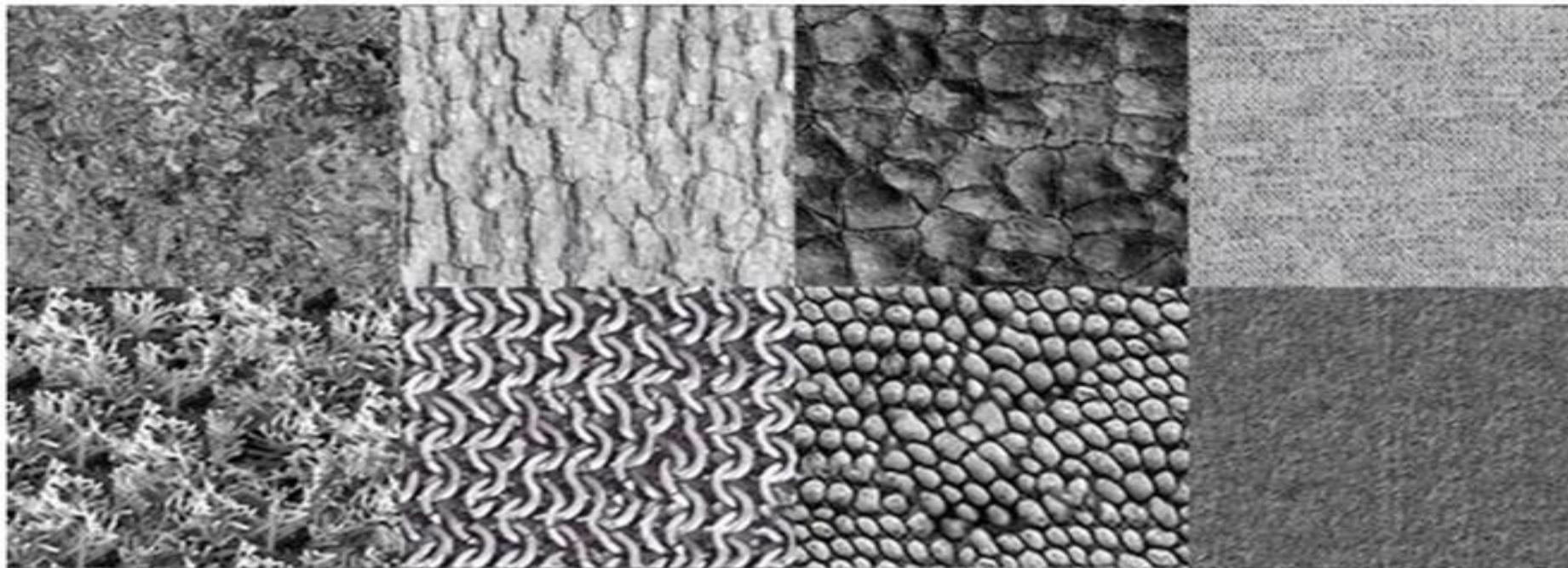
# Color often not a powerful feature



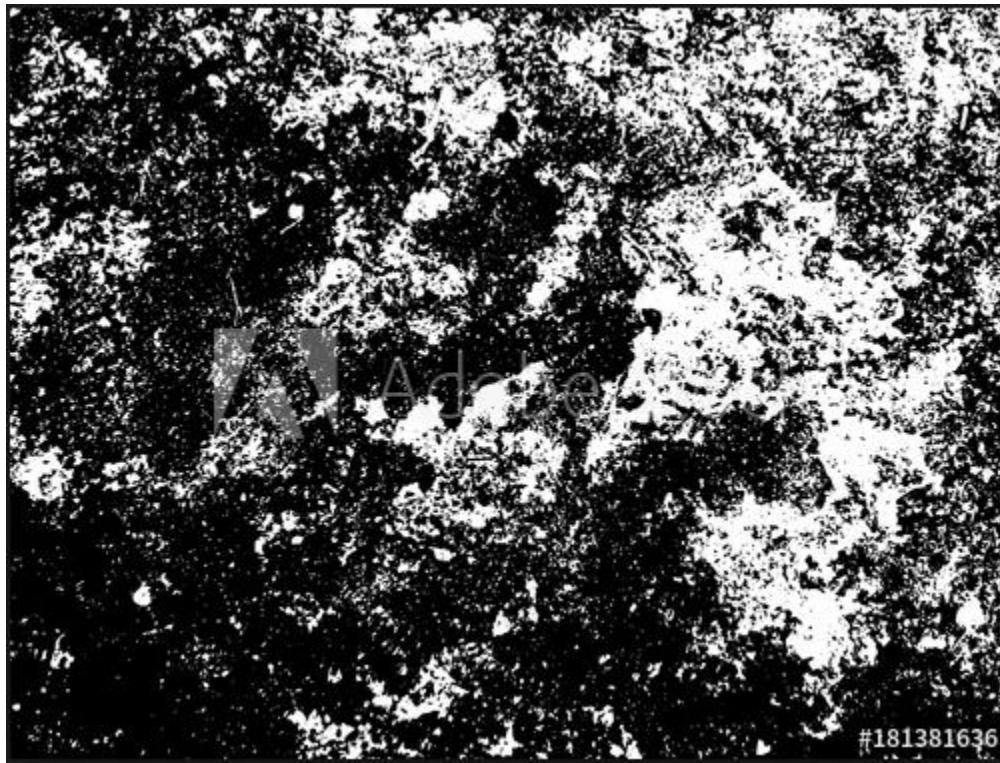
However, these are all images of people but the colors in each image are very different.

# Texture feature

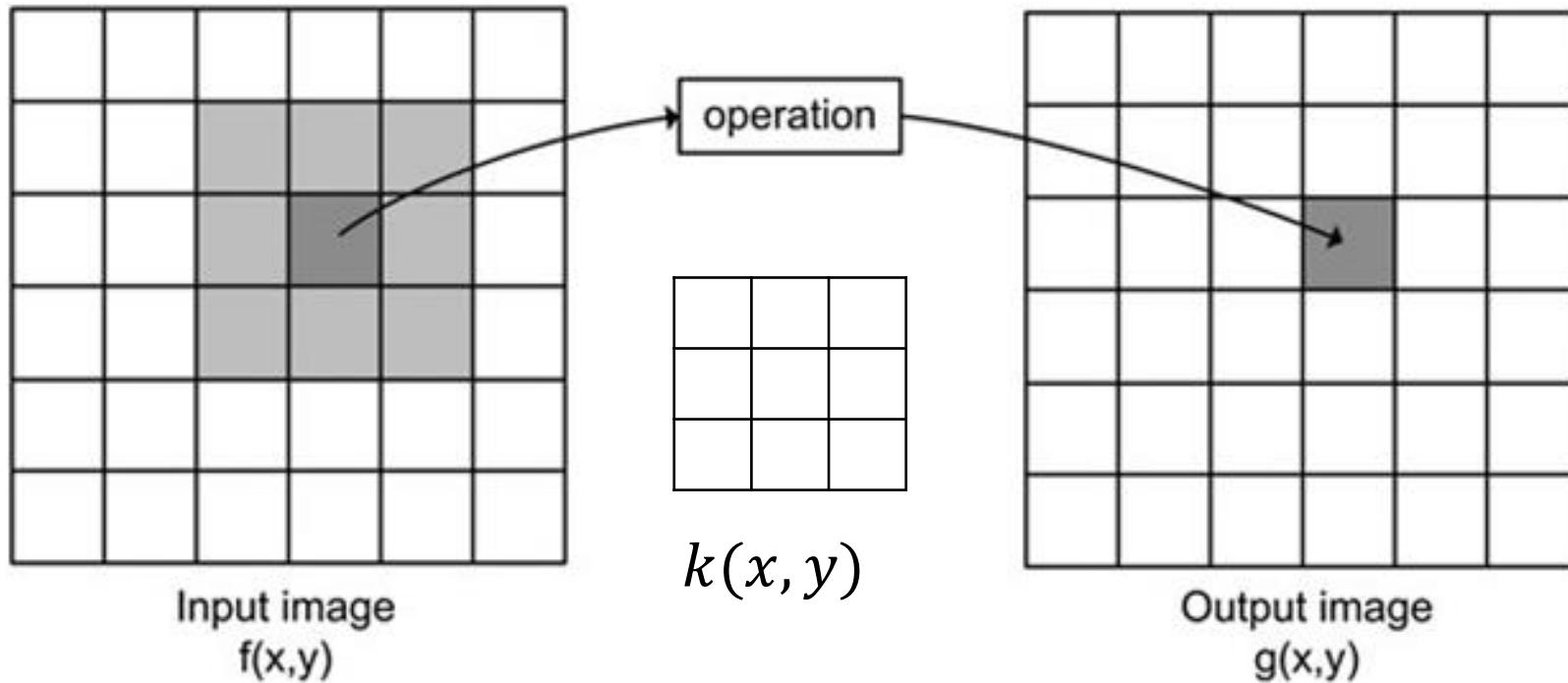
**Texture is a description of the spatial arrangement of color or intensities in an image or a selected region of an image.**



But textures are not always easy to quantify



# Image filtering: Convolution operator



$$g(x, y) = \sum_v \sum_u k(u, v) f(x - u, y - v)$$

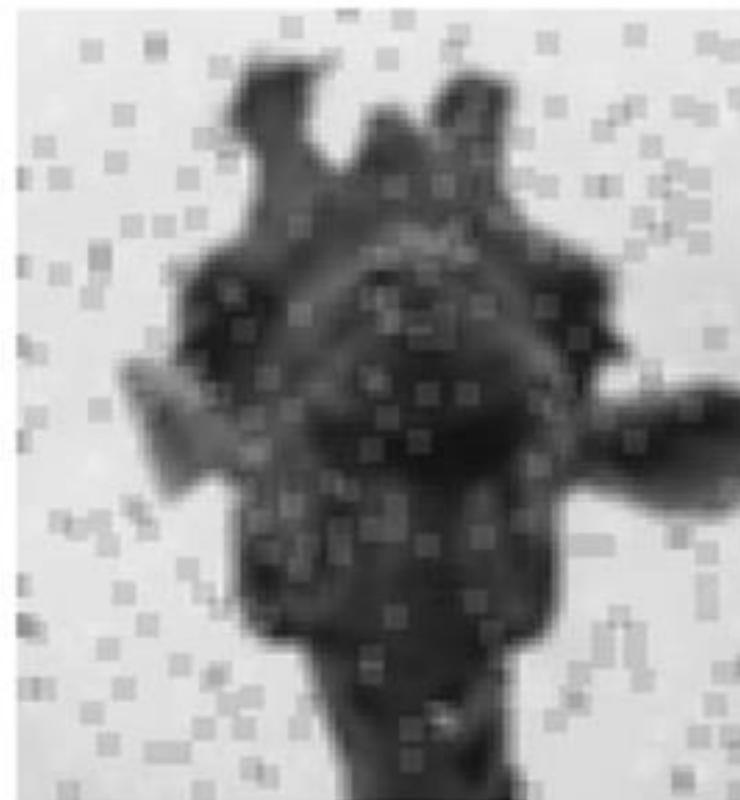
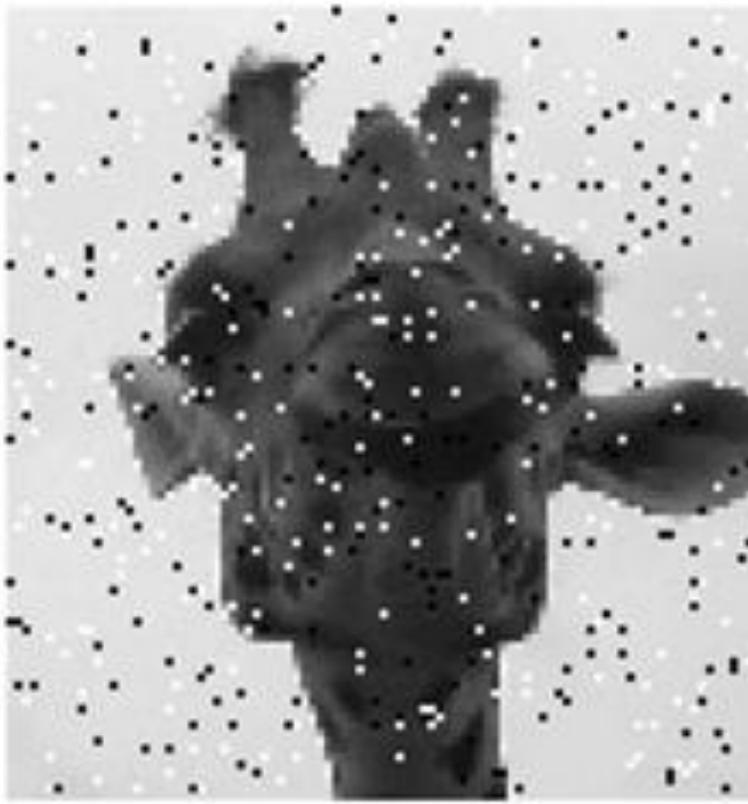
Input image \* Weights → Output image

4	5	7	6	6
3	2	8	0	7
6	7	7	1	5
3	0	1	1	1
4	3	2	1	7

\*

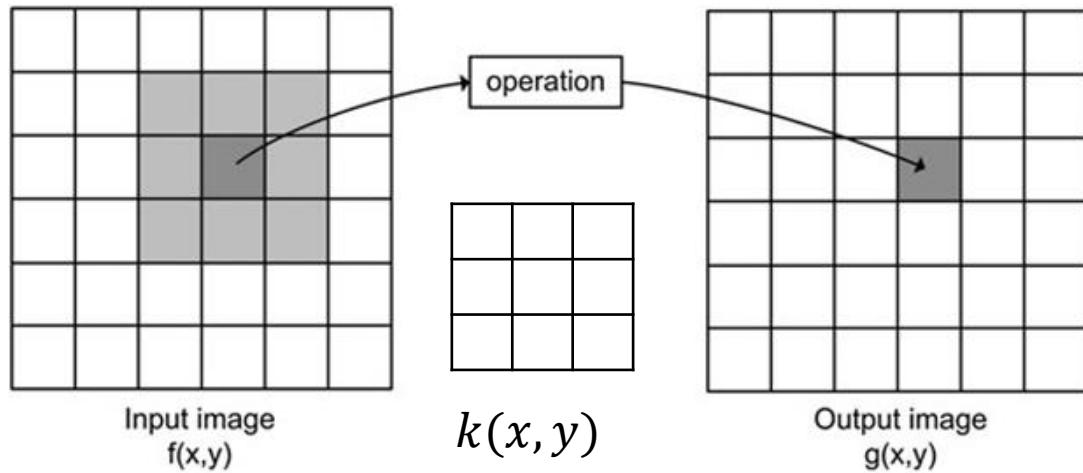
0	0	0
1	0	1
0	0	0


# Image filtering: e.g. Mean Filter



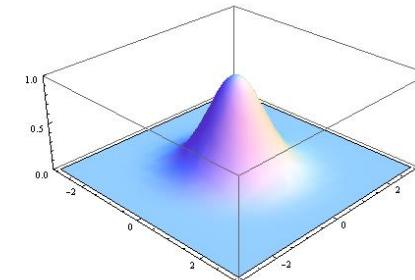
# Image filtering: Convolution operator

Important filter: gaussian filter (gaussian blur)



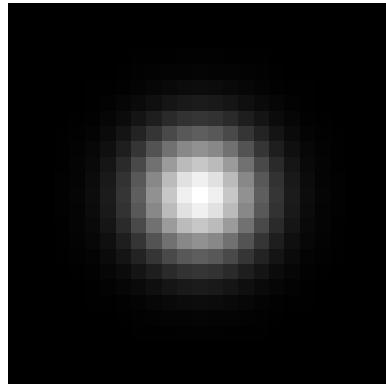
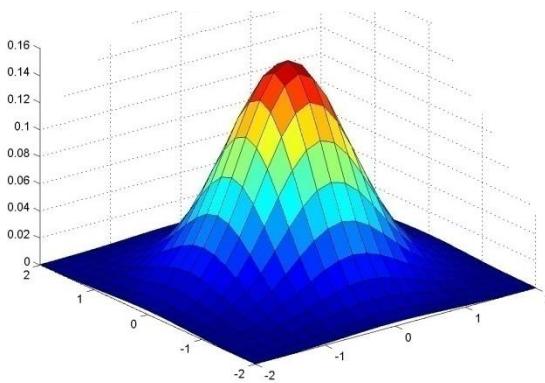
$$k(x, y) =$$

1/16	1/8	1/16
1/8	1/4	1/8
1/16	1/8	1/16



# Important filter: Gaussian

- Weight contributions of neighboring pixels by nearness



0.003	0.013	0.022	0.013	0.003
0.013	0.059	0.097	0.059	0.013
0.022	0.097	0.159	0.097	0.022
0.013	0.059	0.097	0.059	0.013
0.003	0.013	0.022	0.013	0.003

$5 \times 5, \sigma = 1$

$$G_\sigma = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

# Image filtering: Convolution operator

e.g. gaussian filter (gaussian blur)



# Practice with linear filters



0	0	0
0	1	0
0	0	0

?

Original

# Practice with linear filters



Original

0	0	0
0	1	0
0	0	0



Filtered  
(no change)

# Practice with linear filters



0	0	0
0	0	1
0	0	0

?

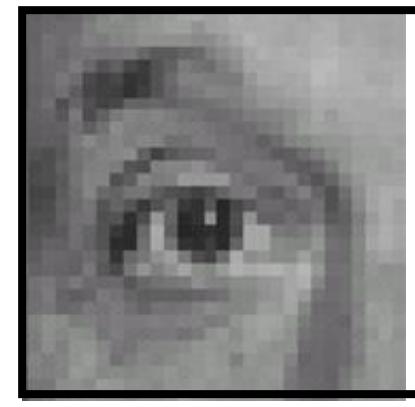
Original

# Practice with linear filters



Original

0	0	0
0	0	1
0	0	0



Shifted left  
By 1 pixel

# Practice with linear filters



Original

$$\begin{array}{|c|c|c|} \hline 0 & 0 & 0 \\ \hline 0 & 2 & 0 \\ \hline 0 & 0 & 0 \\ \hline \end{array}$$

$$- \frac{1}{9} \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array}$$

?

(Note that filter sums to 1)

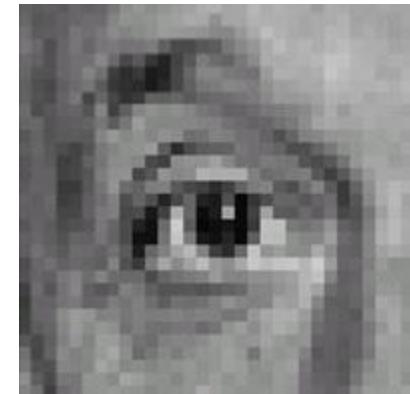
# Practice with linear filters



Original

$$\begin{array}{|c|c|c|} \hline 0 & 0 & 0 \\ \hline 0 & 2 & 0 \\ \hline 0 & 0 & 0 \\ \hline \end{array}$$

$$- \frac{1}{9} \begin{array}{|c|c|c|} \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array}$$



## Sharpening filter

- Accentuates differences with local average

# Key properties of linear filters

## Linearity:

$$\text{imfilter}(I, f_1 + f_2) = \text{imfilter}(I, f_1) + \text{imfilter}(I, f_2)$$

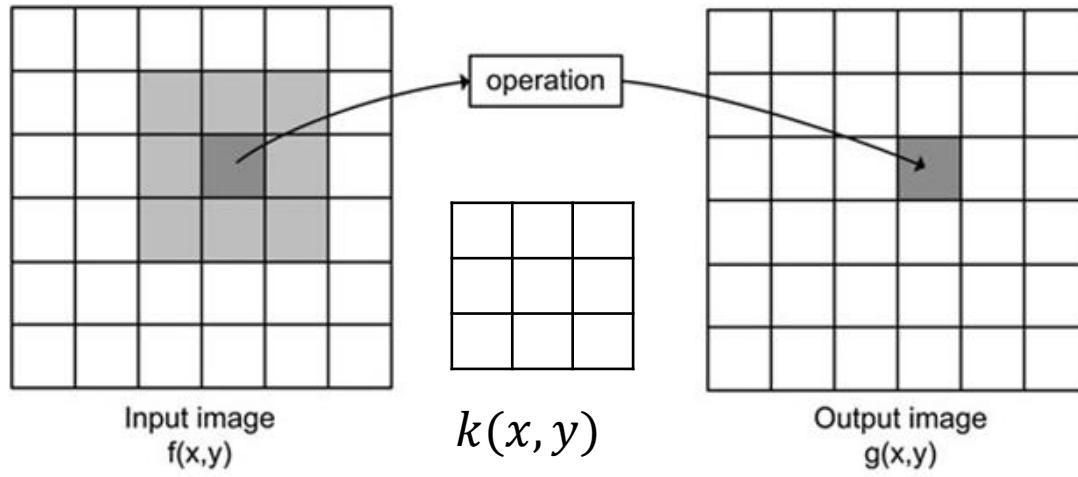
**Shift invariance:** same behavior regardless of pixel location

$$\text{imfilter}(I, \text{shift}(f)) = \text{shift}(\text{imfilter}(I, f))$$

Any linear, shift-invariant operator can be represented as a convolution

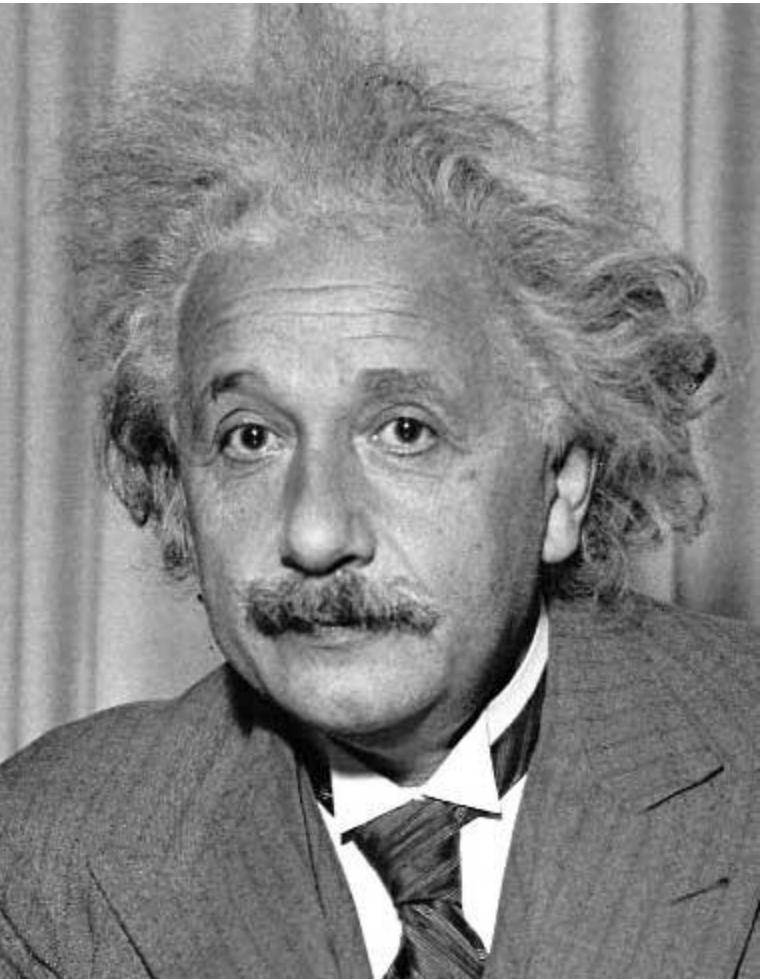
# Image filtering: Convolution operator

## Important Filter: Sobel operator



$$k(x, y) =$$

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

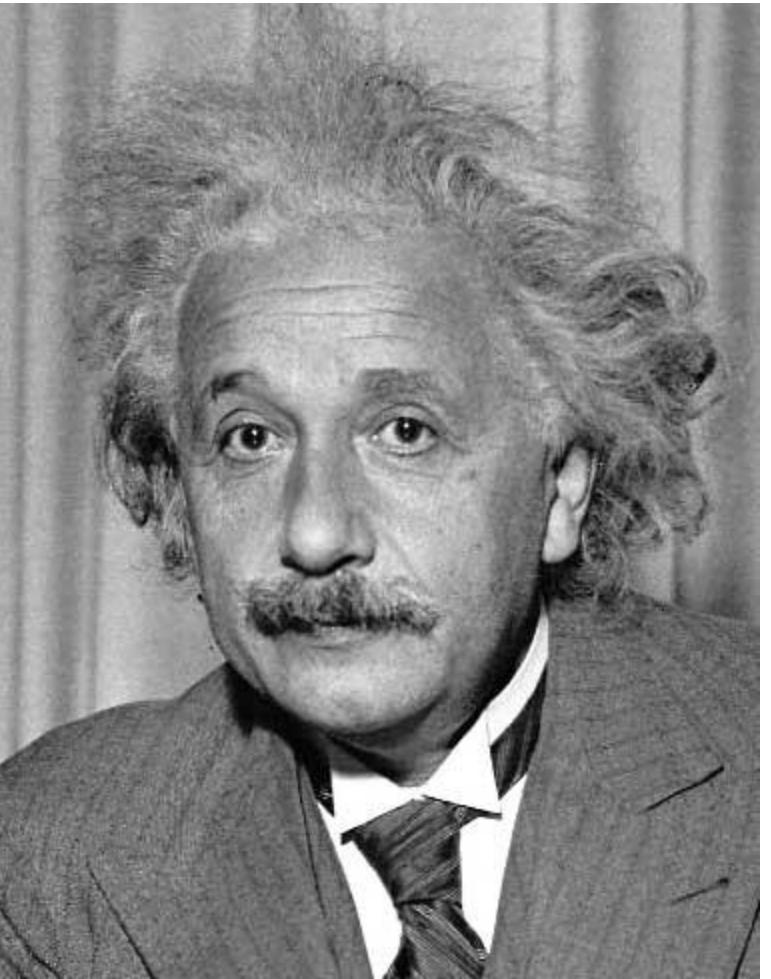


1	0	-1
2	0	-2
1	0	-1

Sobel



Vertical Edge  
(absolute value)



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

Sobel



Horizontal Edge  
(absolute value)

Sobel operators are equivalent to 2D partial derivatives of the image

- Vertical sobel operator – Partial derivative in X (width)
- Horizontal sobel operator – Partial derivative in Y (height)
- Can compute magnitude and phase at each location
- Useful for detecting edges

Sobel filters are (approximate) partial derivatives of the image

Let  $f(x, y)$  be your input image, then the partial derivative is:

$$\frac{\partial f(x, y)}{\partial x} = \lim_{h \rightarrow 0} \frac{f(x + h, y) - f(x, y)}{h}$$

Also:

$$\frac{\partial f(x, y)}{\partial x} = \lim_{h \rightarrow 0} \frac{f(x + h, y) - f(x - h, y)}{2h}$$

But digital images are not continuous, they are discrete

Let  $f[x, y]$  be your input image, then the partial derivative is:

$$\Delta_x f[x, y] = f[x + 1, y] - f[x, y]$$

Also:  $\Delta_x f[x, y] = f[x + 1, y] - f[x - 1, y]$

But digital images are not continuous, they are discrete

Let  $f[x, y]$  be your input image, then the partial derivative is:

$$\Delta_x f[x, y] = f[x + 1, y] - f[x, y]$$

$$k(x, y) =$$

-1	1
----	---

Also:

$$\Delta_x f[x, y] = f[x + 1, y] - f[x - 1, y]$$

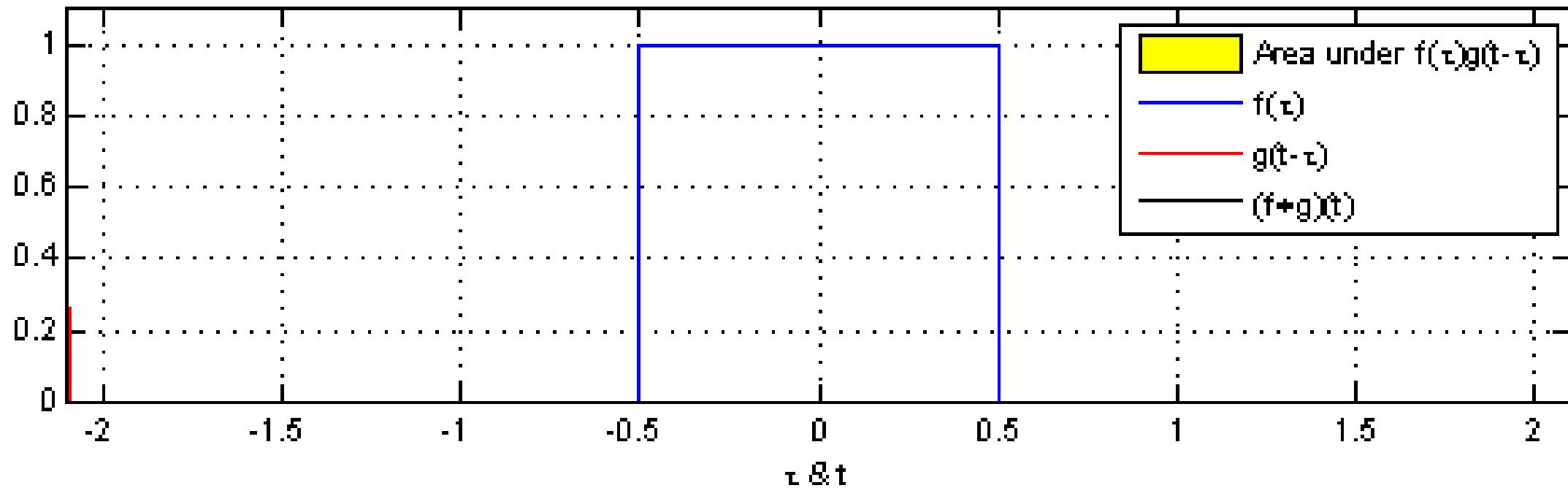
$$k(x, y) =$$

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

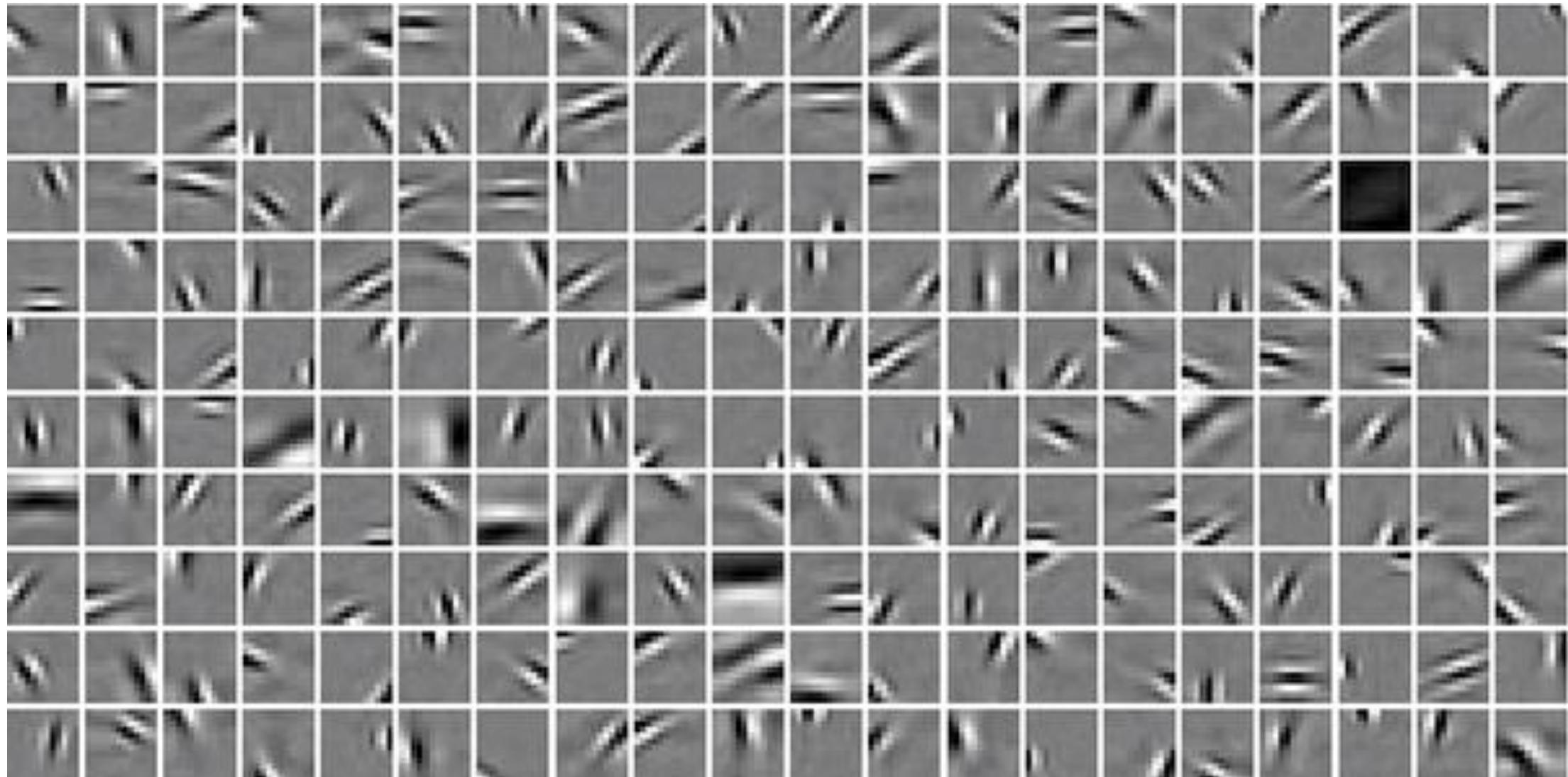
# Sobel Operators Smooth in Y and then Differentiate in X

$$k(x, y) = \begin{matrix} 1 \\ 2 \\ 1 \end{matrix} * \begin{matrix} 1 & 0 & -1 \end{matrix} = \begin{matrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{matrix}$$

Similarly to differentiate in Y



# Gabor filters

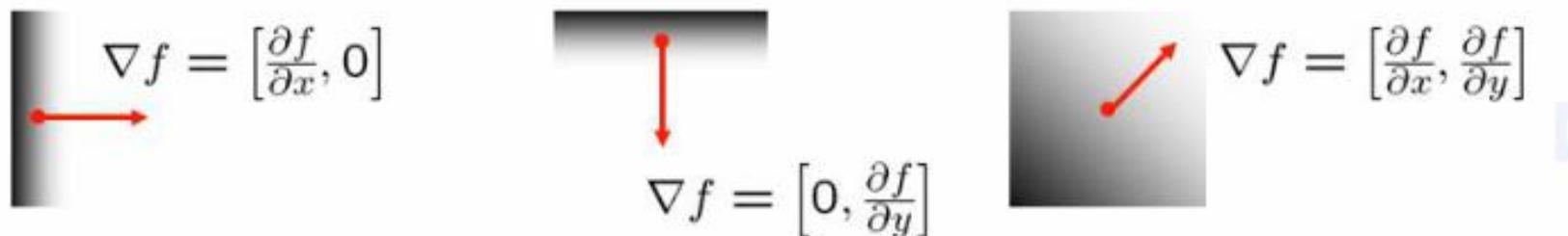


# Image gradient in a nutshell

The gradient of an image:

$$\nabla f = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right]$$

The gradient points in the direction of most rapid change in intensity



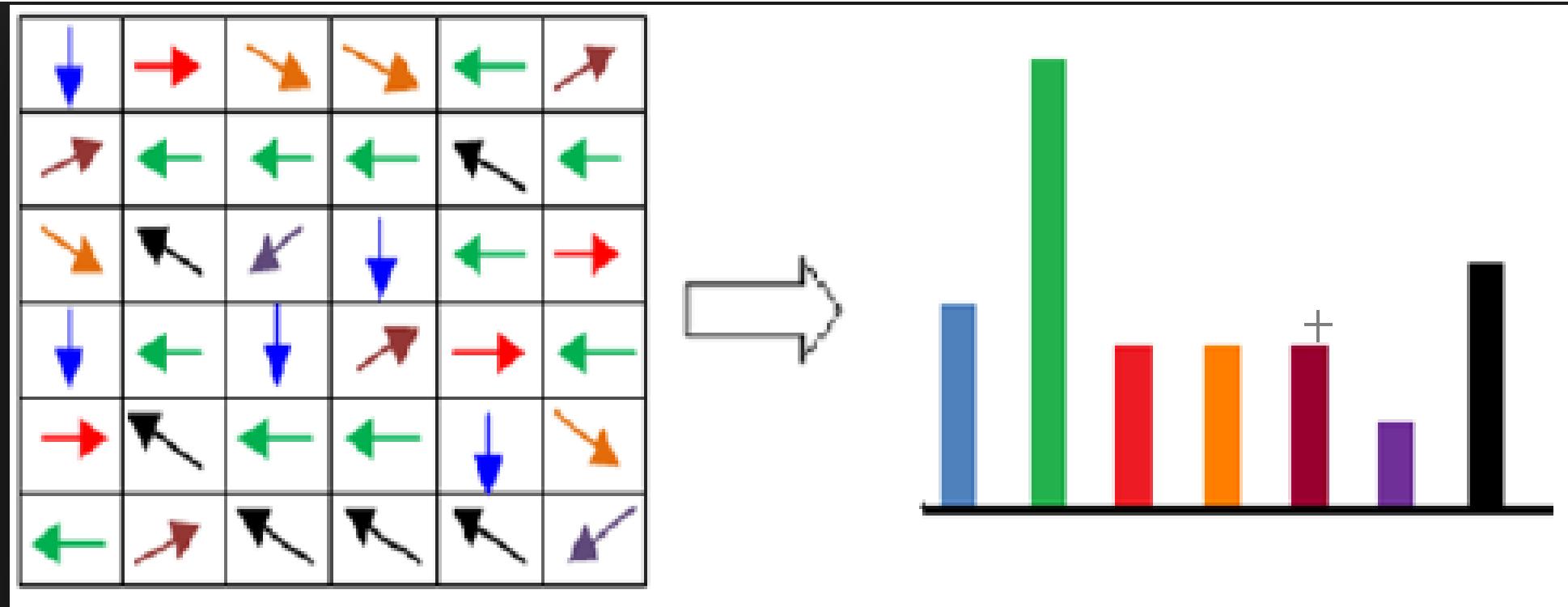
The gradient direction (orientation of edge normal) is given by:

$$\theta = \tan^{-1} \left( \frac{\partial f}{\partial y} / \frac{\partial f}{\partial x} \right)$$

The *edge strength* is given by the **gradient magnitude**

$$\| \nabla f \| = \sqrt{\left( \frac{\partial f}{\partial x} \right)^2 + \left( \frac{\partial f}{\partial y} \right)^2}$$

# Gradient histogram

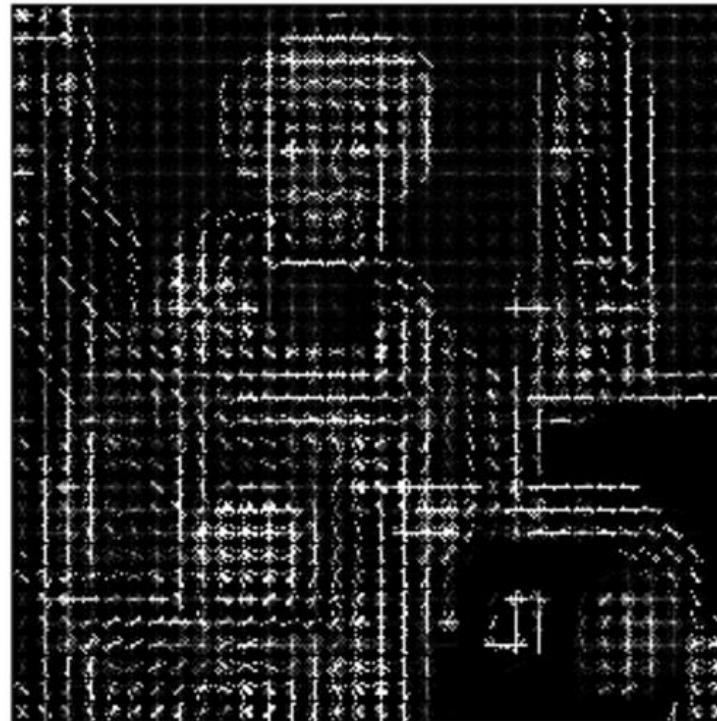


# Image Features: HoG

Input image

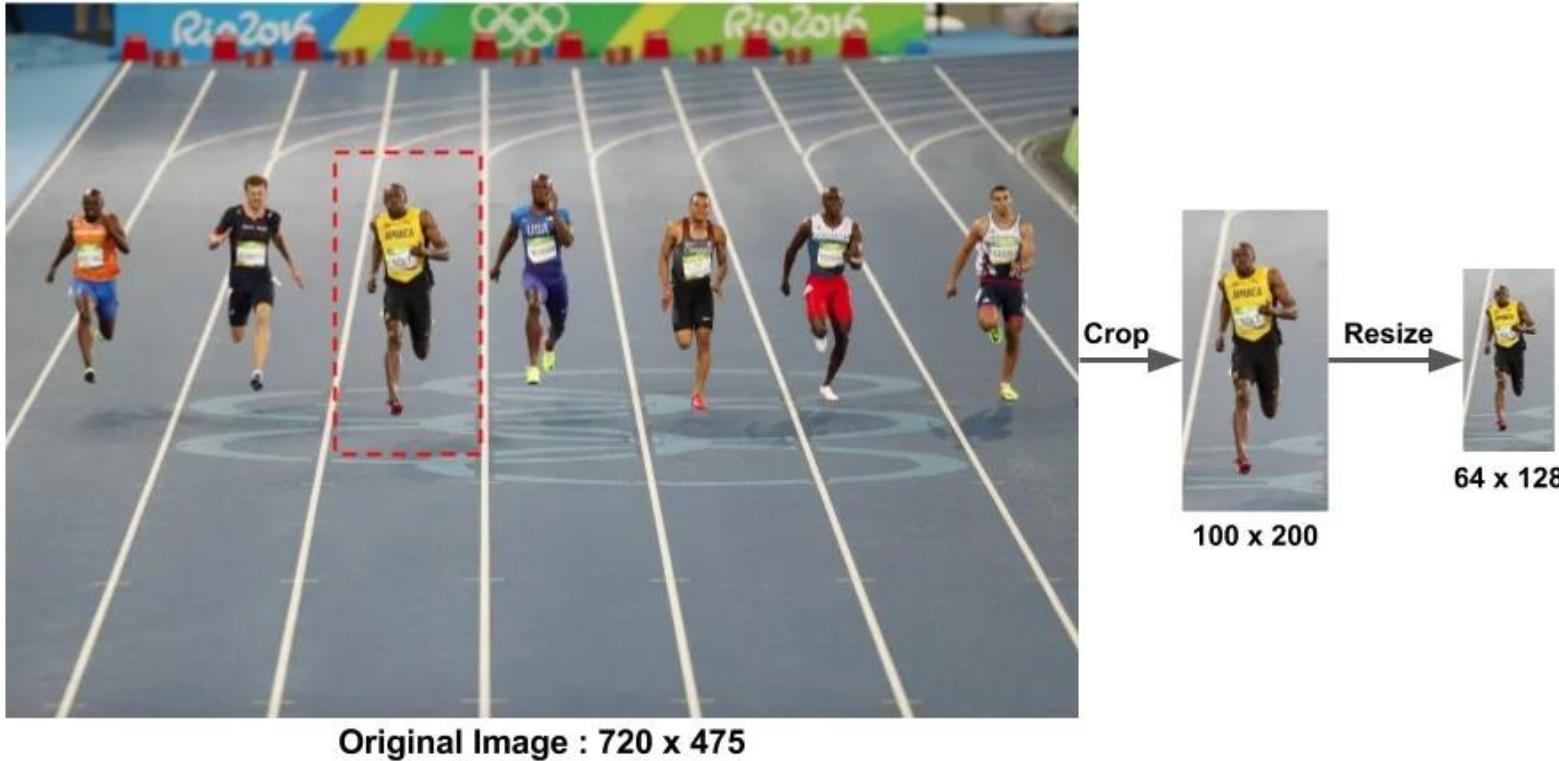


Histogram of Oriented Gradients



Paper by Navneet Dalal & Bill Triggs presented at CVPR 2005 for detecting people.

# Image Features: HoG



# Image Features: HoG

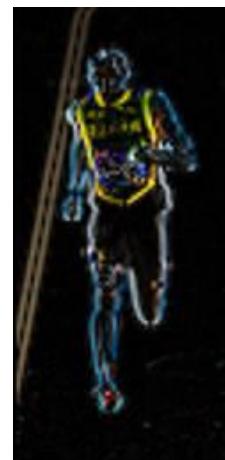
Compute gradients



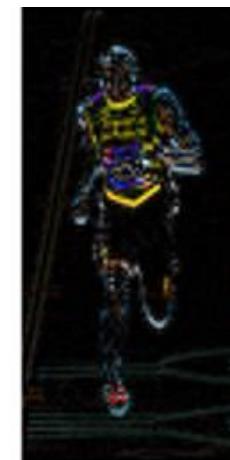
\*

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

-1
0
1



$I_x$



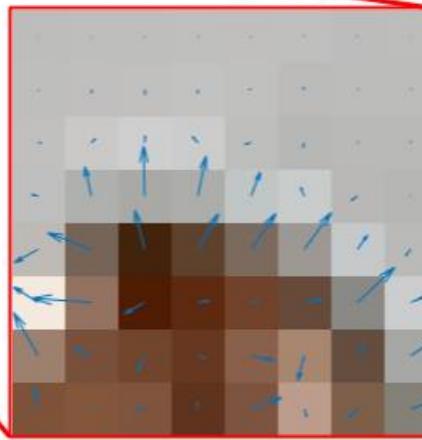
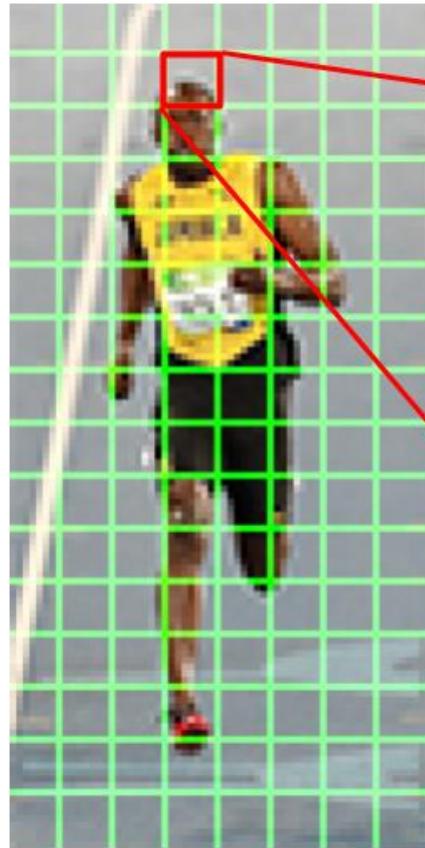
$I_y$

$$\sqrt{I_x^2 + I_y^2}$$



# Image Features: HoG

We will aggregate gradient magnitude and directions in 8x8 pixel regions



2	3	4	4	3	4	2	2
5	11	17	13	7	9	3	4
11	21	23	27	22	17	4	6
23	99	165	135	85	32	26	2
91	155	133	136	144	152	57	28
98	196	76	38	26	60	170	51
165	60	60	27	77	85	43	136
71	13	34	23	108	27	48	110

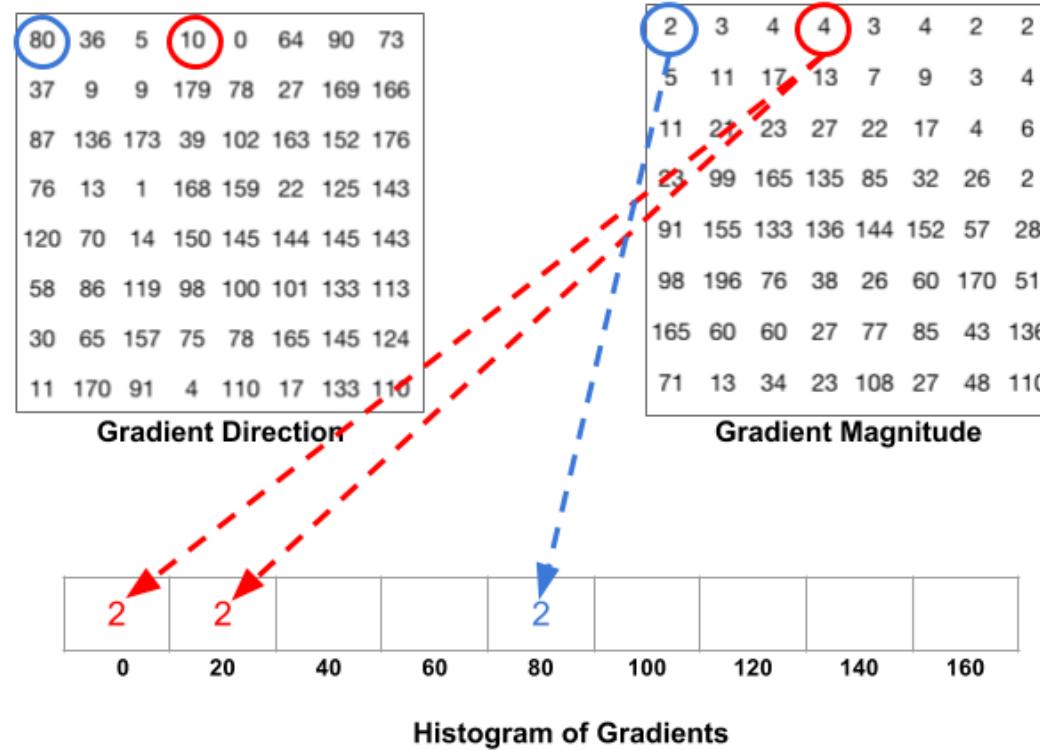
Gradient Magnitude

80	36	5	10	0	64	90	73
37	9	9	179	78	27	169	166
87	136	173	39	102	163	152	176
76	13	1	168	159	22	125	143
120	70	14	150	145	144	145	143
58	86	119	98	100	101	133	113
30	65	157	75	78	165	145	124
11	170	91	4	110	17	133	110

Gradient Direction

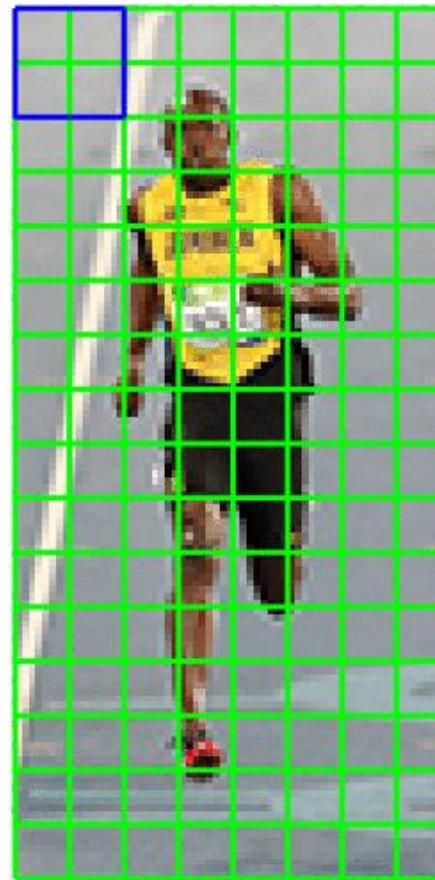
# Image Features: HoG

Compute a histogram  
with 9 bins for angles  
from 0 to 180



# Image Features: HoG

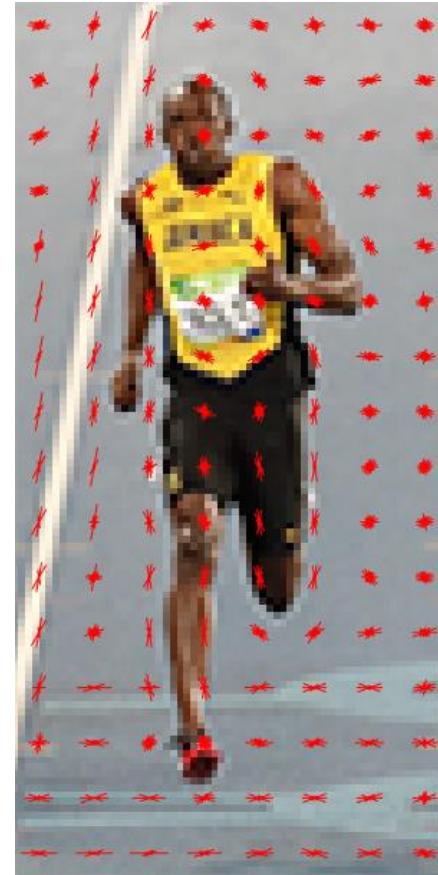
Normalize histograms  
with respect to  
histograms of adjacent  
neighbors.



# Image Features: HoG

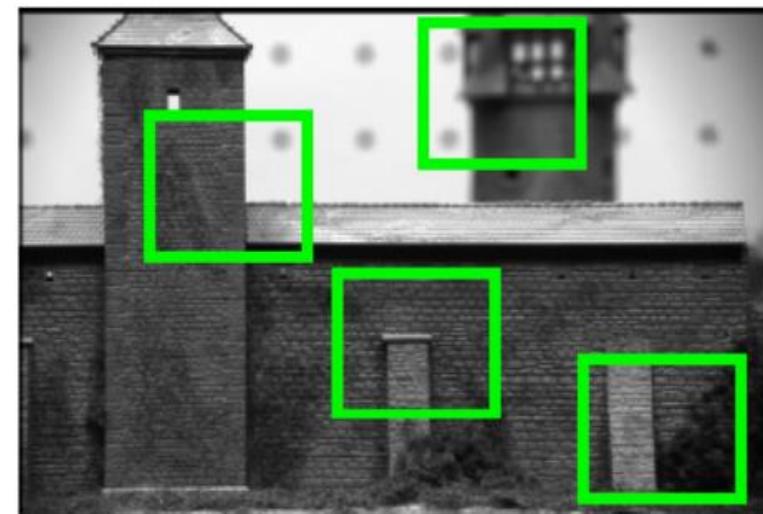
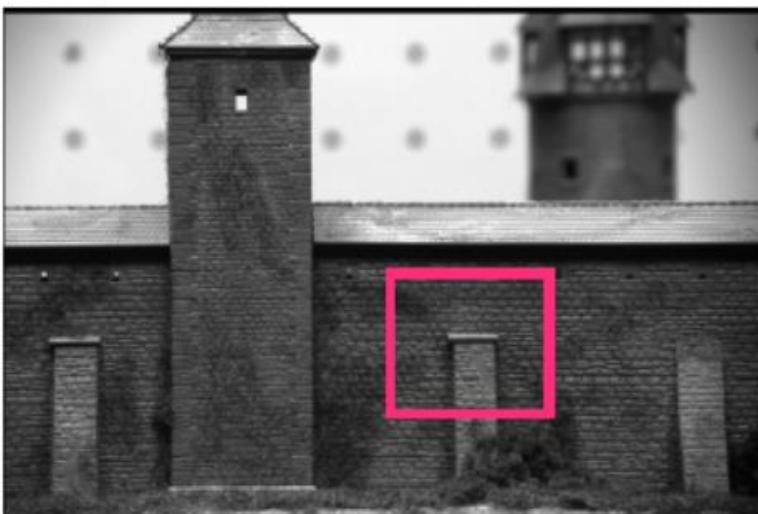
Image (or image region) represented by a vector containing all the histograms.

In this case how long is that vector?



# Interest points - motivation

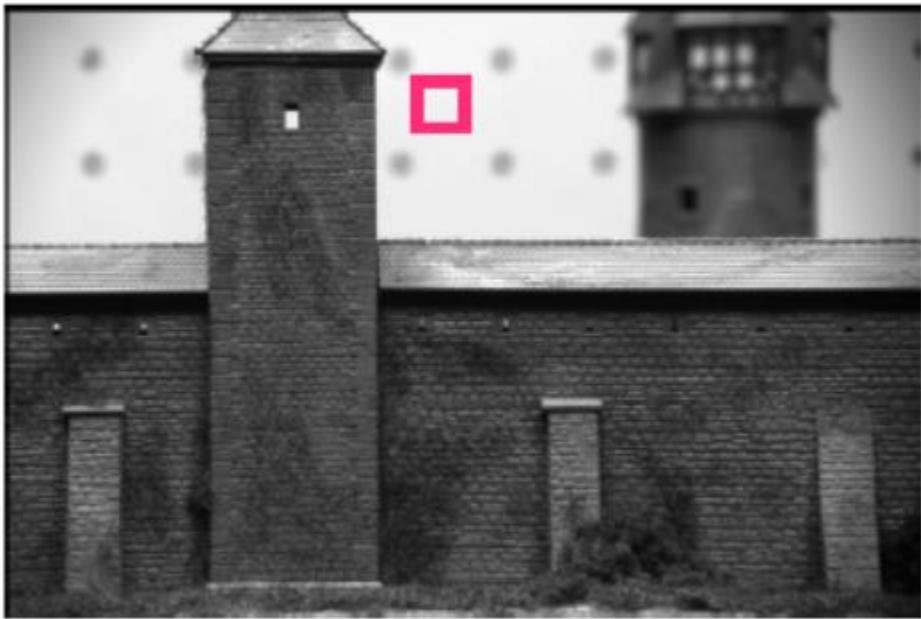
Elements to be matched are image patches of fixed size



Task: find the best (most similar) patch in a second image



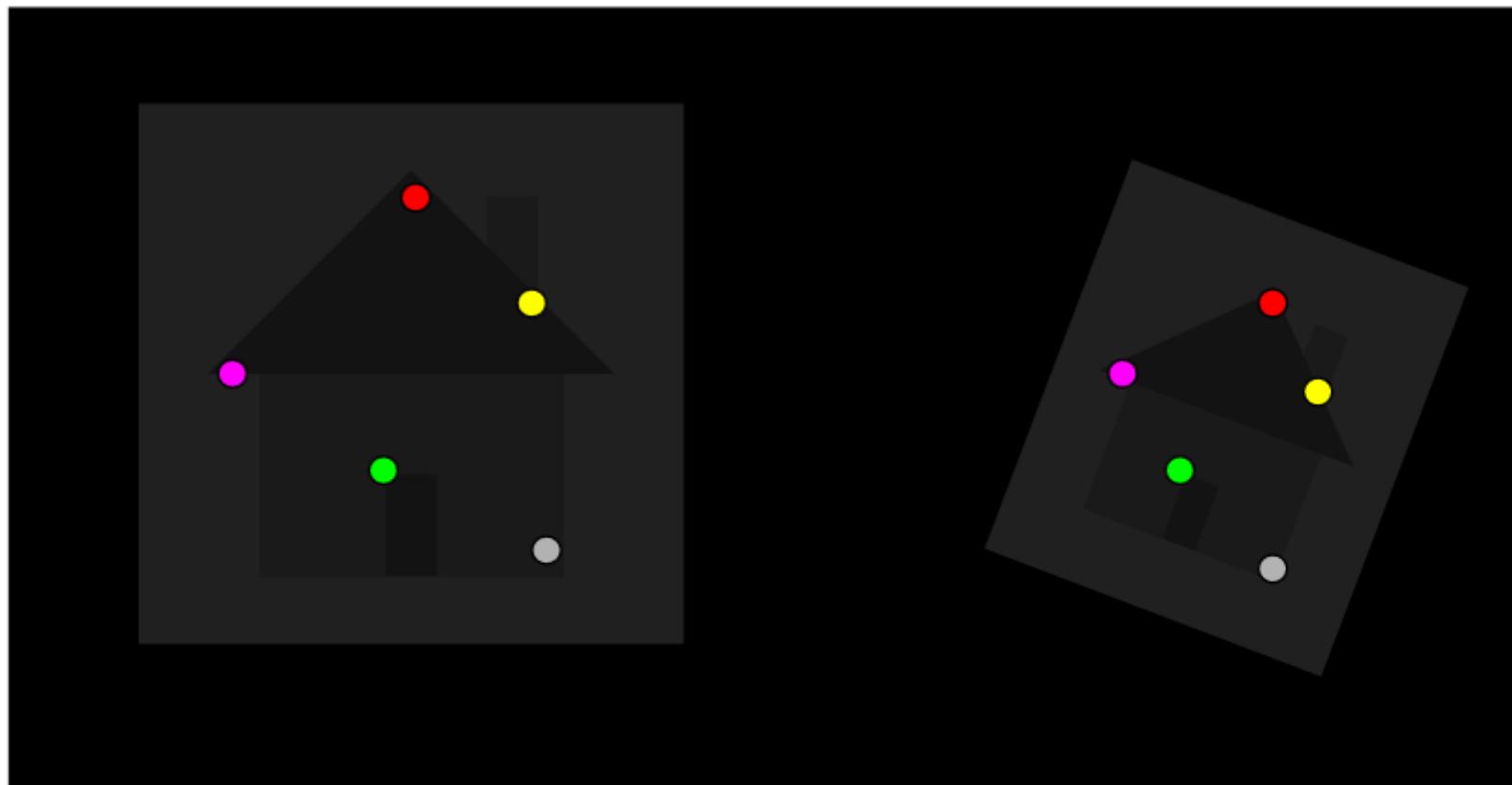
# Interest points - motivation



?  
=



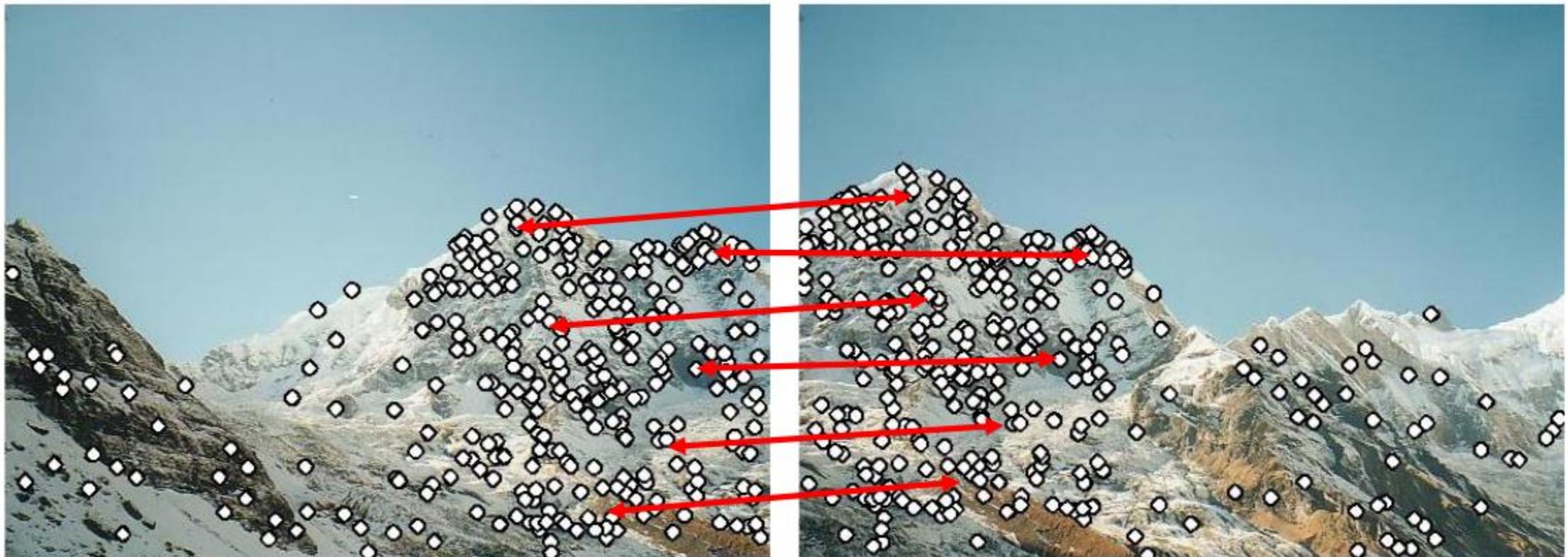
Corresponding points (or features) between images enable the estimation of parameters describing geometric transforms between the images.



- How do we combine these two images?

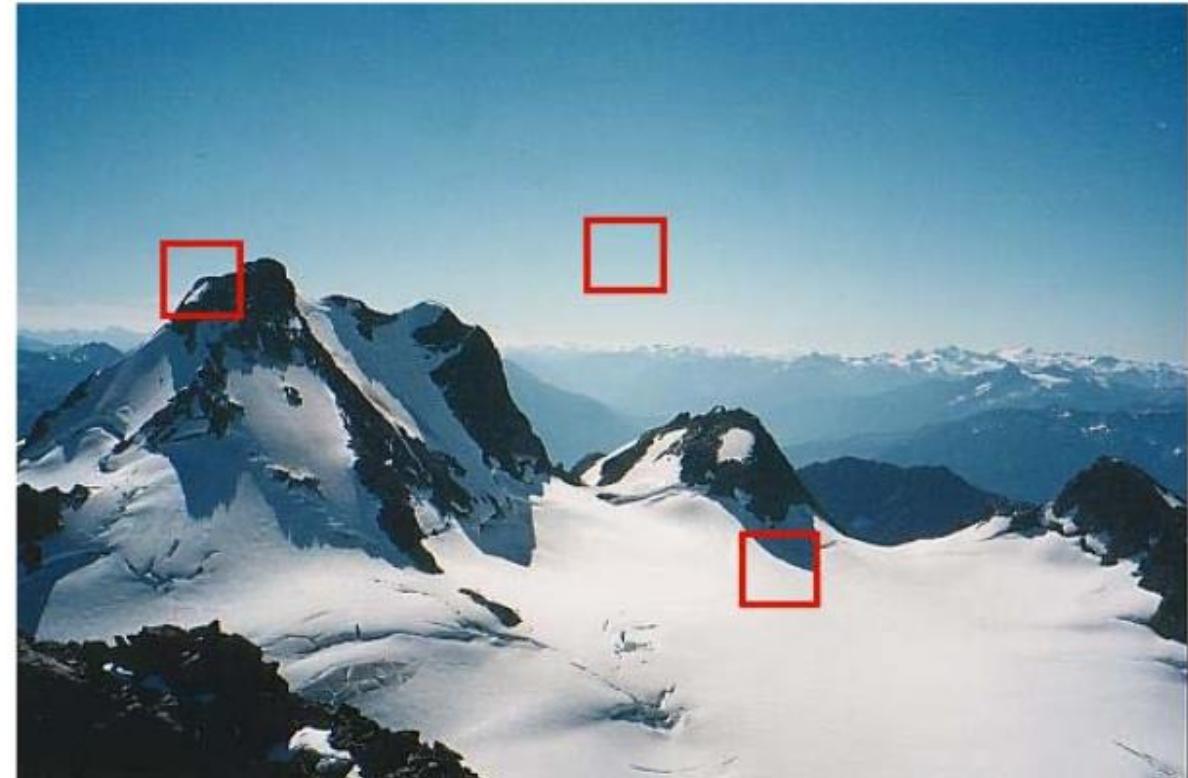
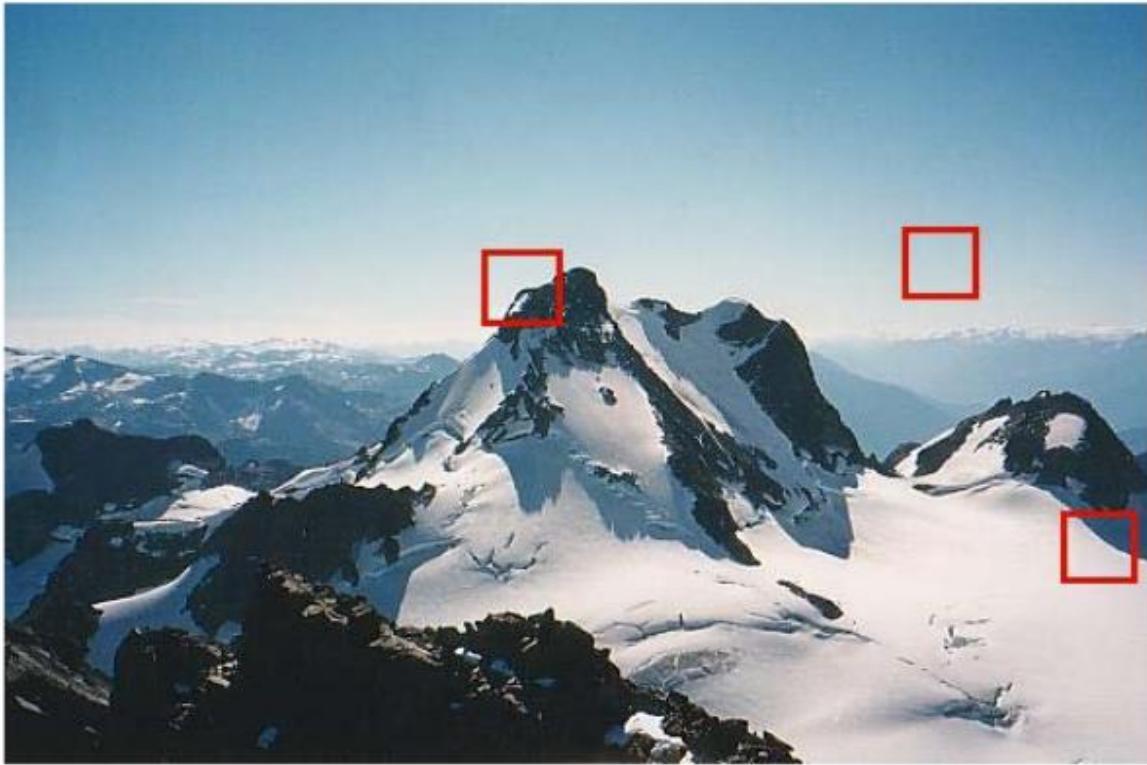


# Panorama stitching



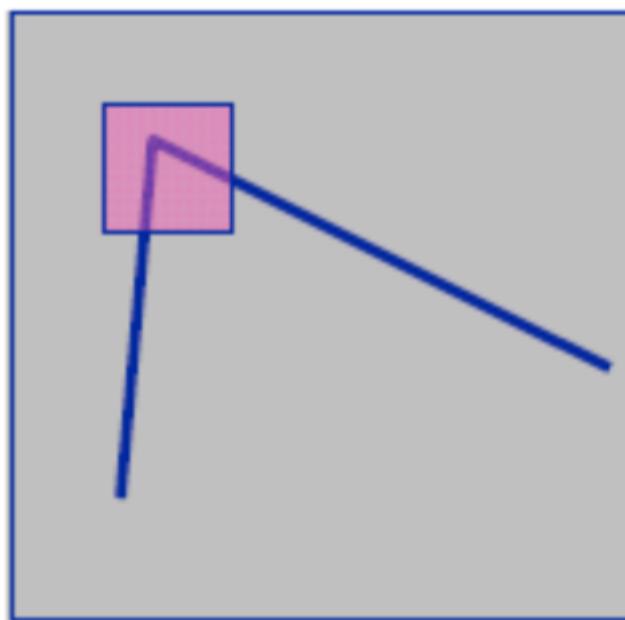


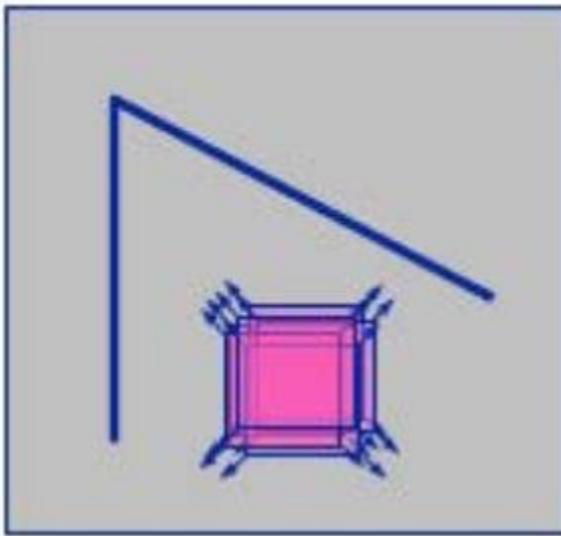
# Corners



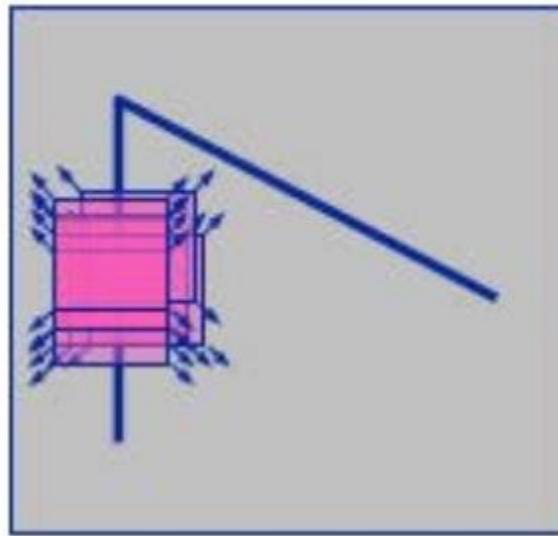
We should easily recognize the point by looking through a small window

Shifting a window in *any direction* should give *a large change* in response

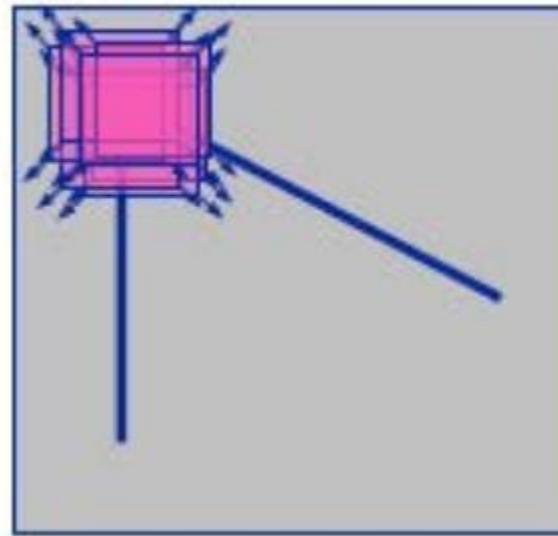




“flat” region:  
no change in  
all directions



“edge”:  
no change along  
the edge direction



“corner”:  
significant change  
in all directions

Change of intensity for the shift  $[u, v]$ :

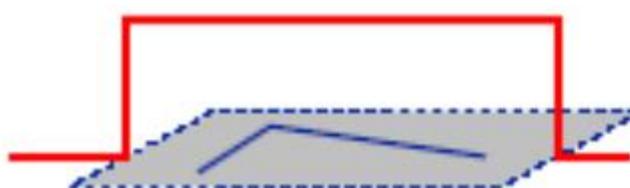
$$E(u, v) = \sum_{x, y} w(x, y)[I(x + u, y + v) - I(x, y)]^2$$

Window  
function

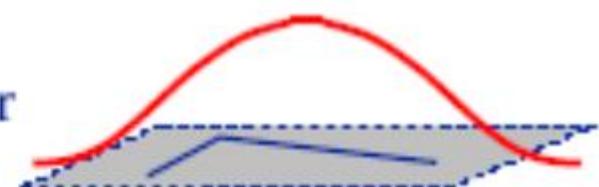
Shifted  
intensity

Intensity

Window function  $w(x, y) =$



or



$$E(u, v) = \sum_{x,y} w(x, y)[I(x + u, y + v) - I(x, y)]^2$$

$$E(u, v) \approx \sum_{x,y} [I(x, y) + uI_x + vI_y - I(x, y)]^2$$

$$E(u, v) \approx \sum_{x,y} u^2 I_x^2 + 2uv I_x I_y + v^2 I_y^2$$

$$E(u, v) \approx [u \quad v] \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

Cornerness

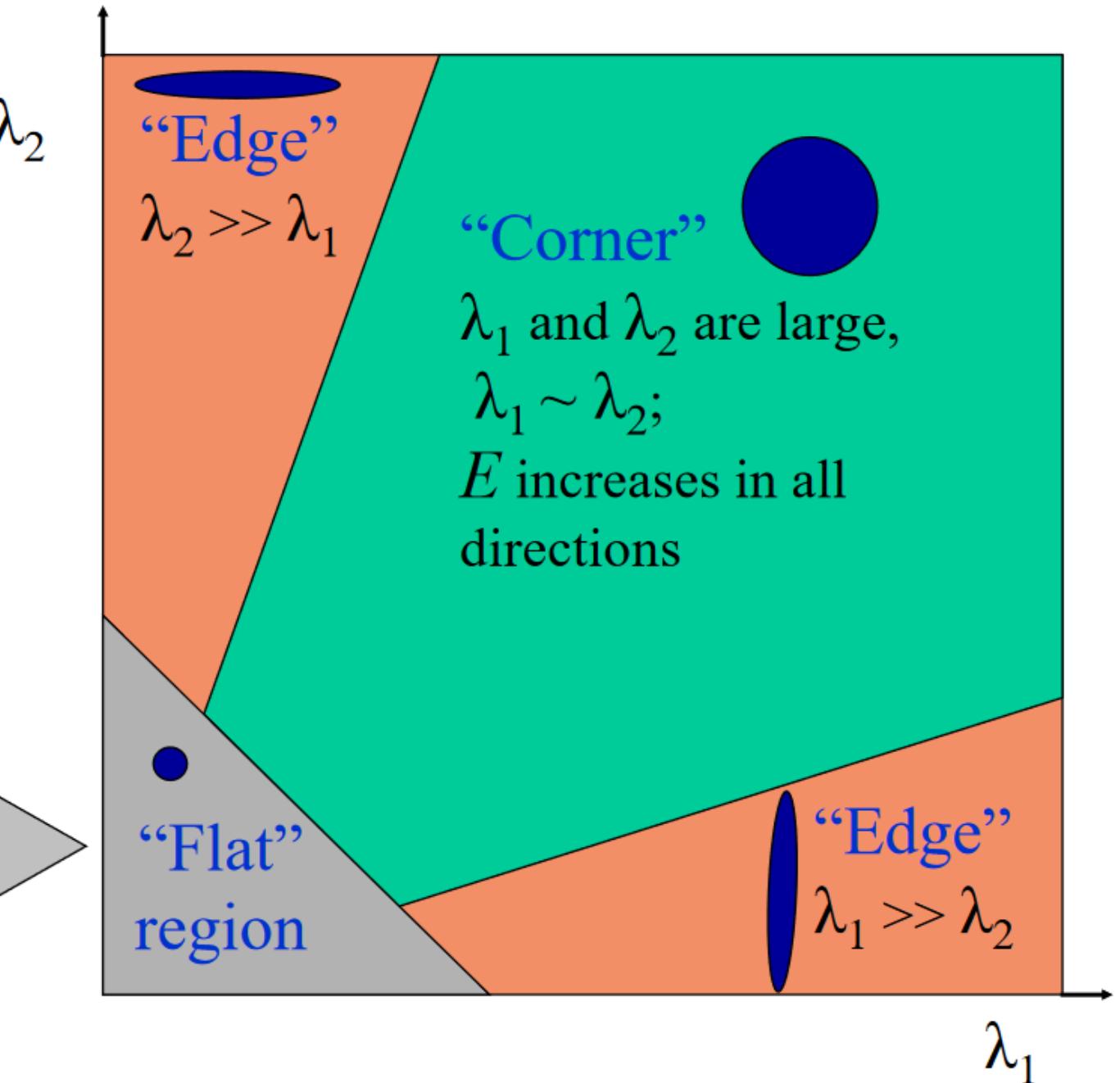
$$R = \det M - k(\text{trace } M)^2$$

$$\det M = \lambda_1 \lambda_2$$

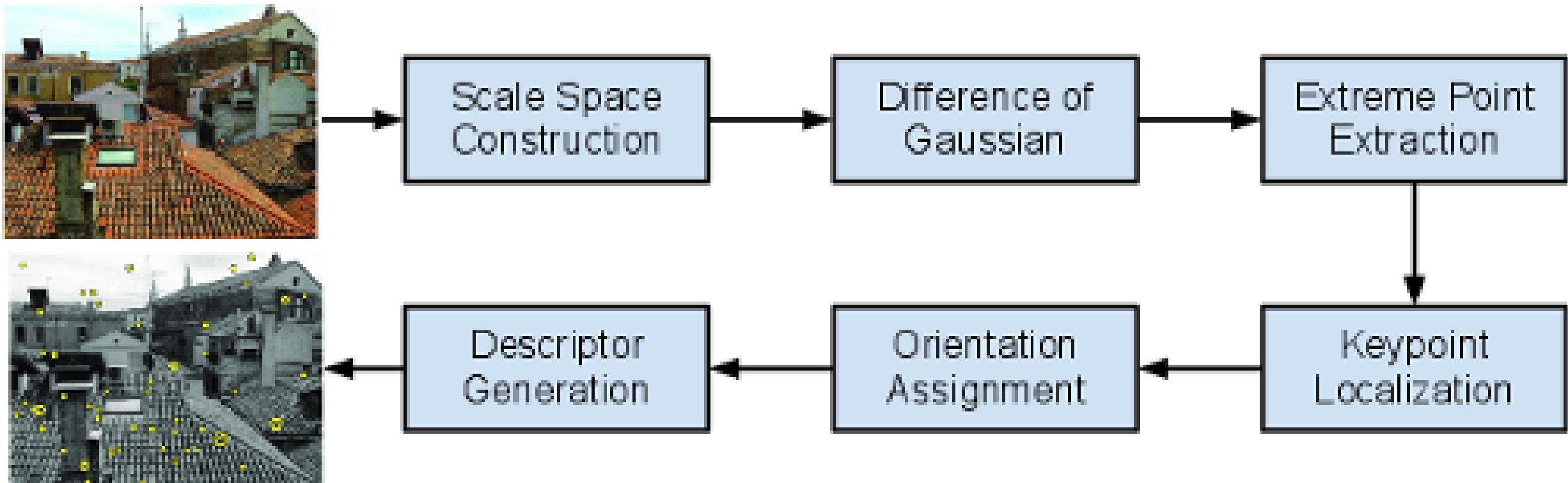
$$\text{trace } M = \lambda_1 + \lambda_2$$

Classification of  
image points using  
eigenvalues of  $A_W$ :

$\lambda_1$  and  $\lambda_2$  are small;  
 $S_W$  is almost constant  
in all directions



# The state of the art interest point detector - SIFT



Won't cover in the class – Pl refer to David Lowe paper for details

## Given:

- positive training images containing an object class, and



- negative training images that don't



## Classify:

- a test image as to whether it contains the object class or not

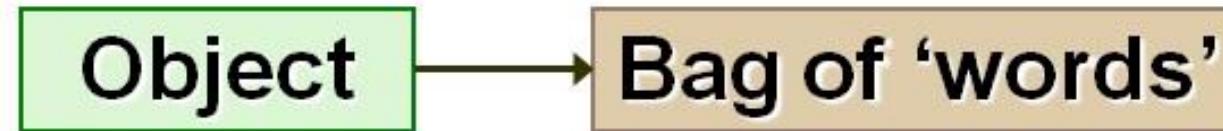


?

# Visual words

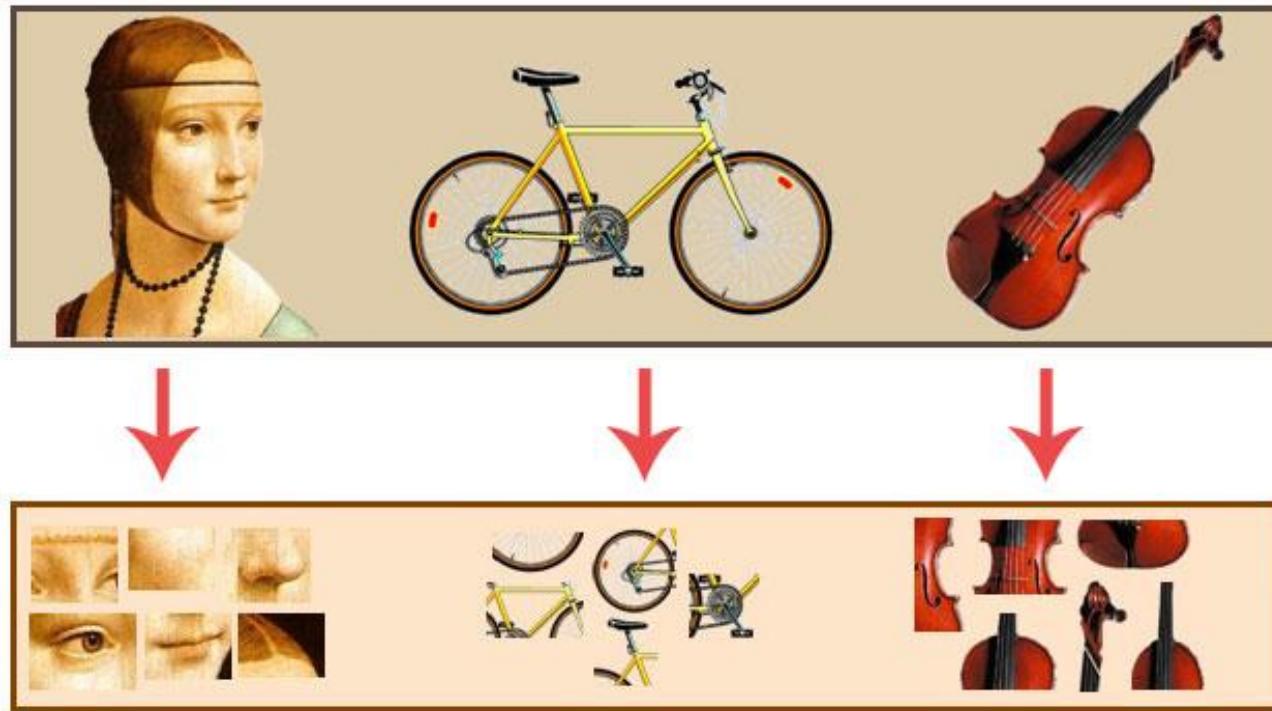


# Image Features: Bag of (Visual) Words Representation

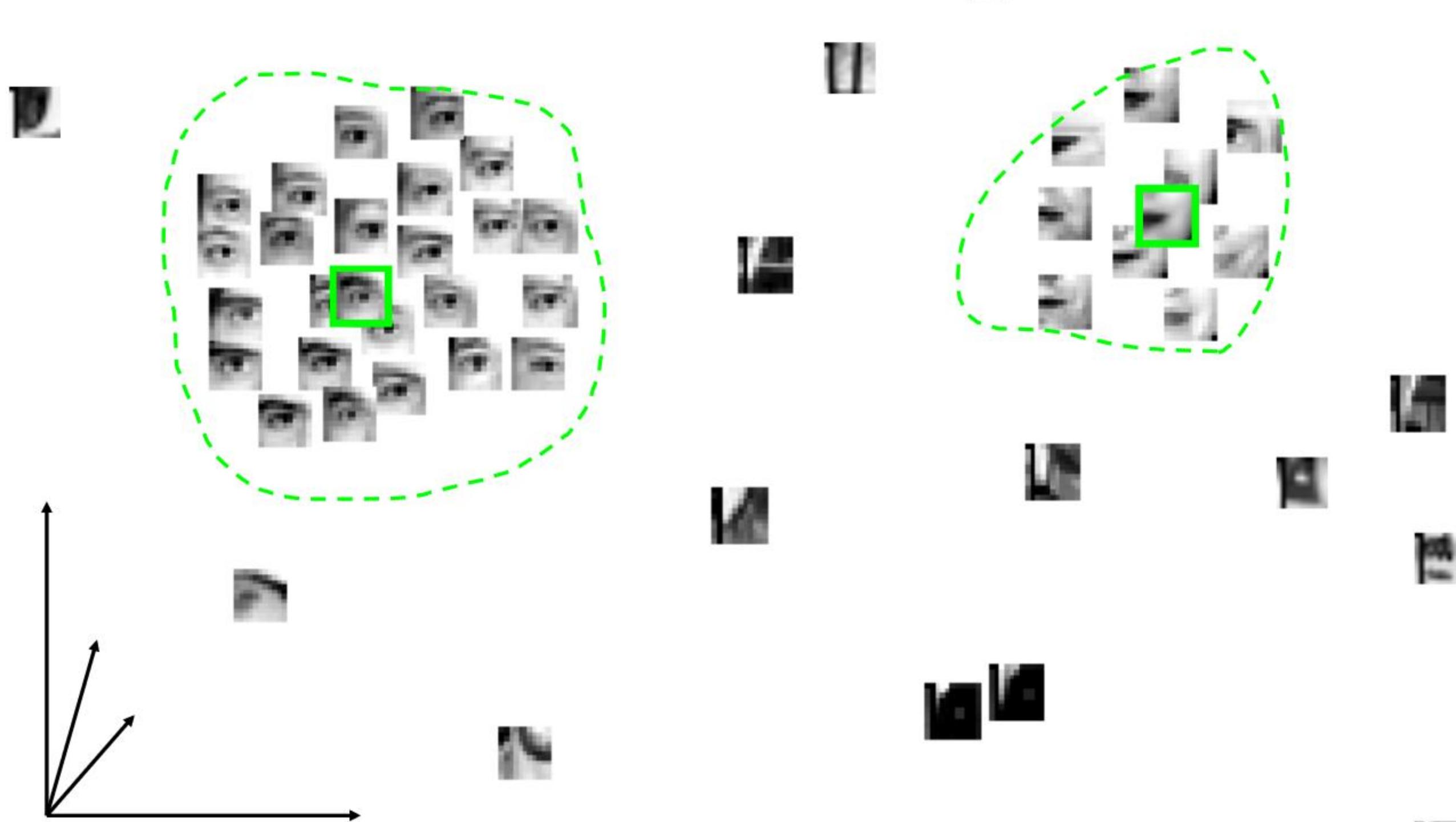


Bag of Features (read more in Gabriela Csurka's ECCV 2004 paper)

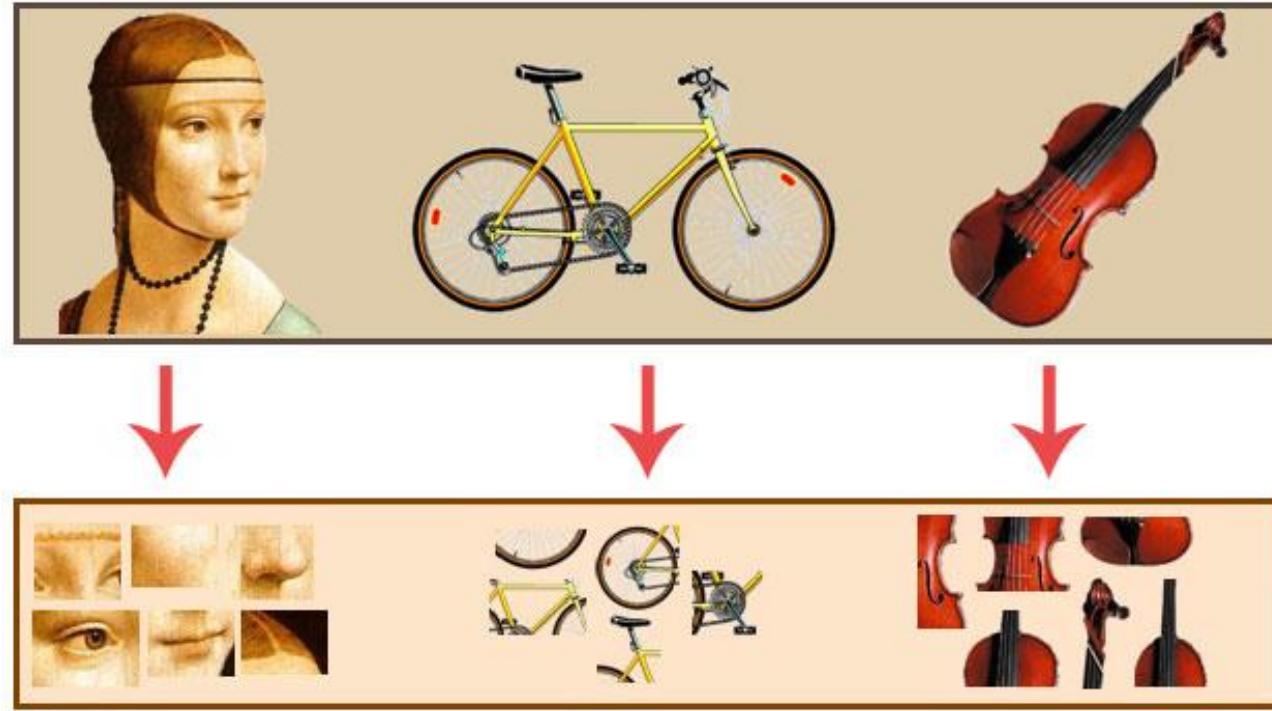
Extract SIFT  
Feature  
Descriptors



Intuition: sample a bunch of pieces (“words”) from various parts of the image. Images of the same object are more likely to share similar pieces



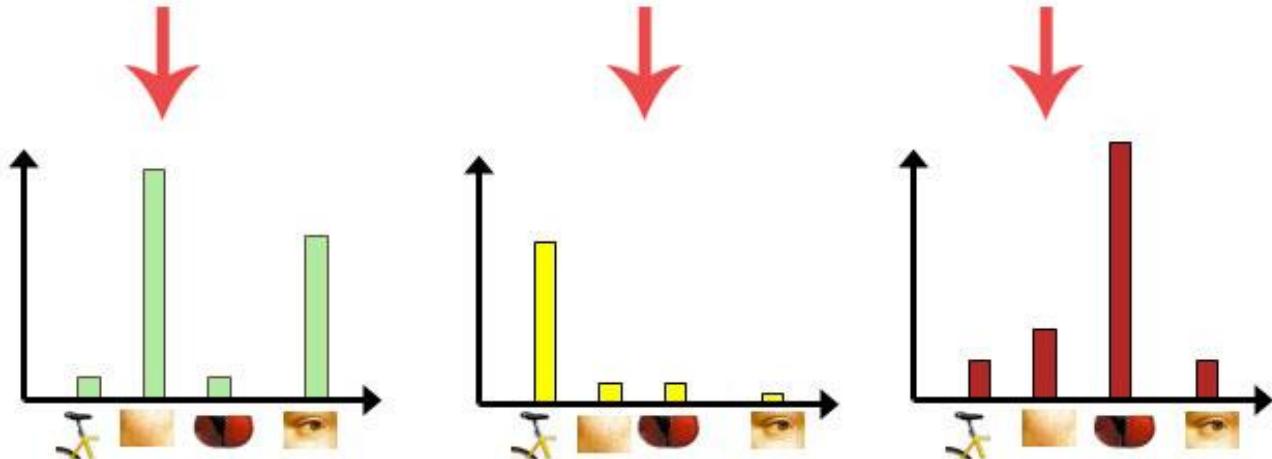
Extract SIFT  
Feature  
Descriptors



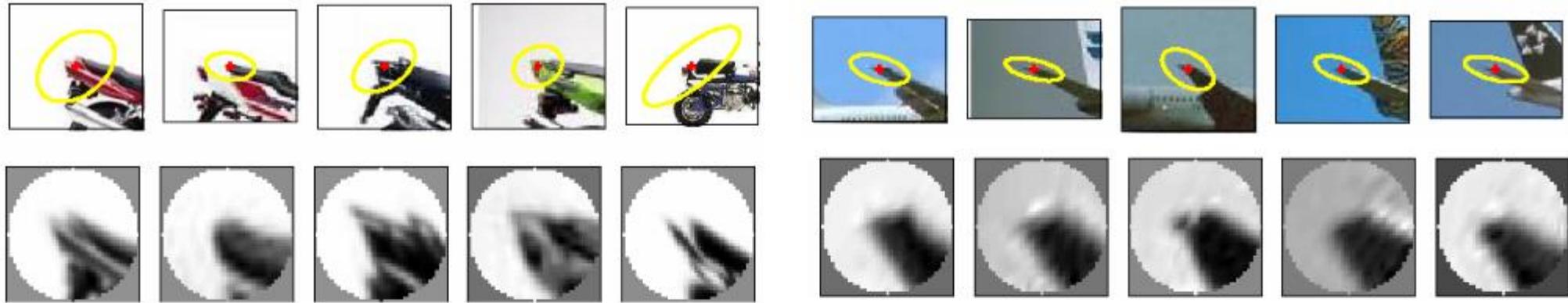
Create Dictionary  
of Distinctive SIFT  
Features



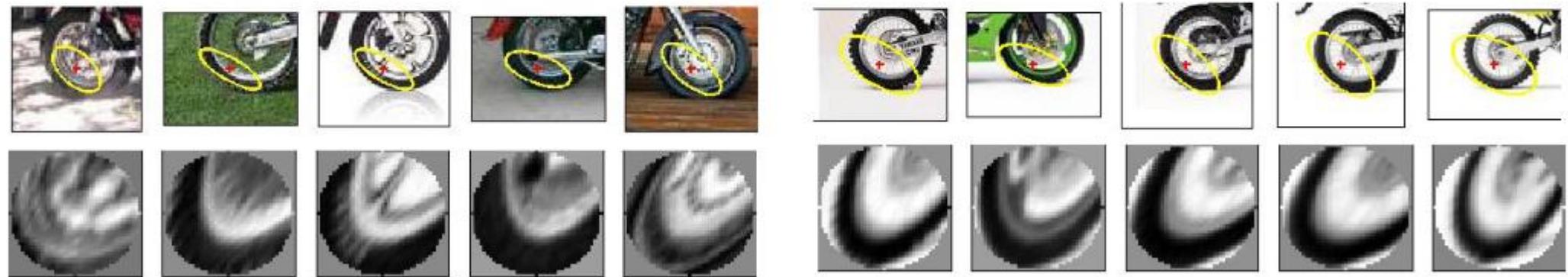
Extract SIFT  
Feature  
Descriptors



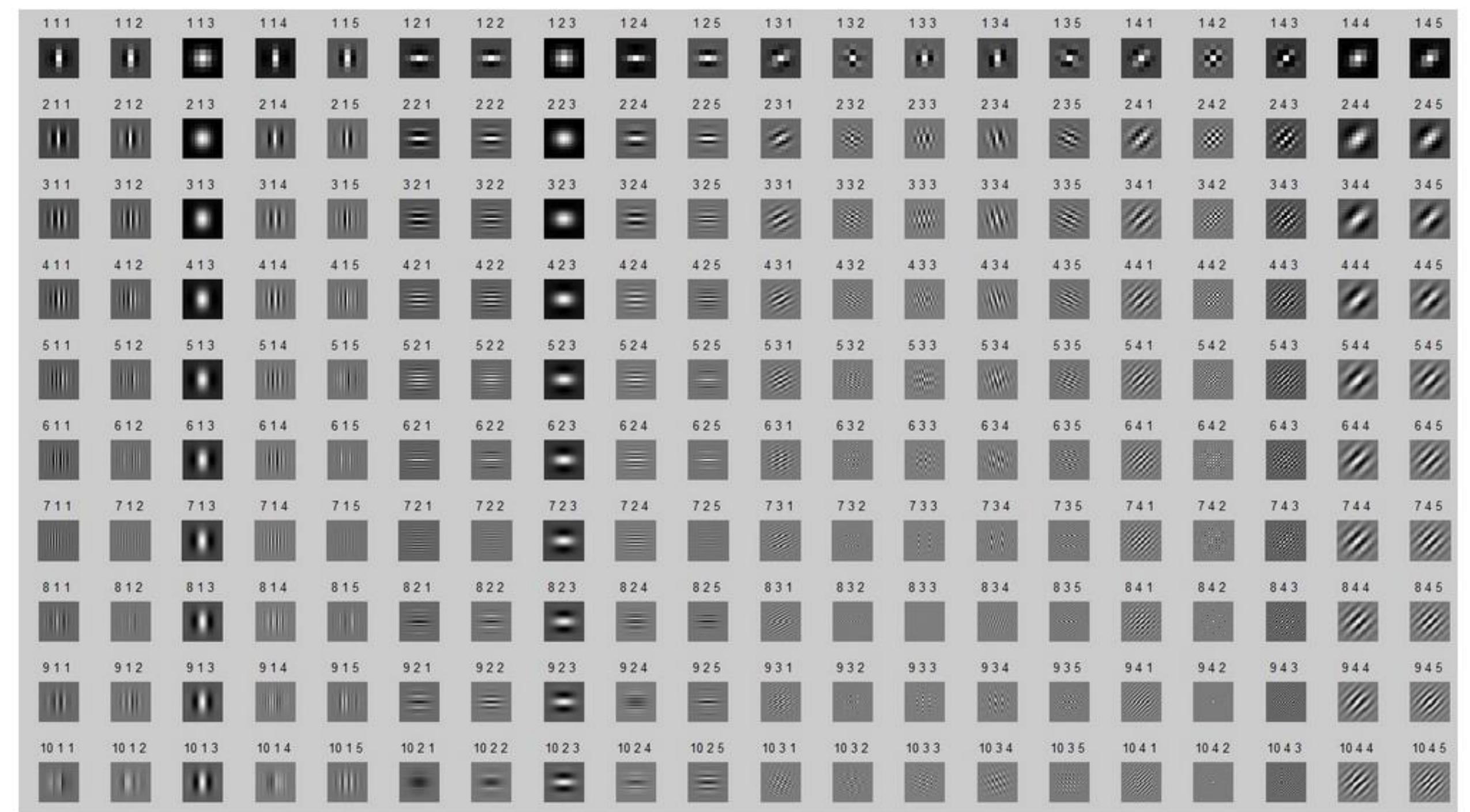
Compute  
Histograms of  
Features



**Visual Polysemy:** Single visual word occurring on different (but locally similar) parts on different object categories.

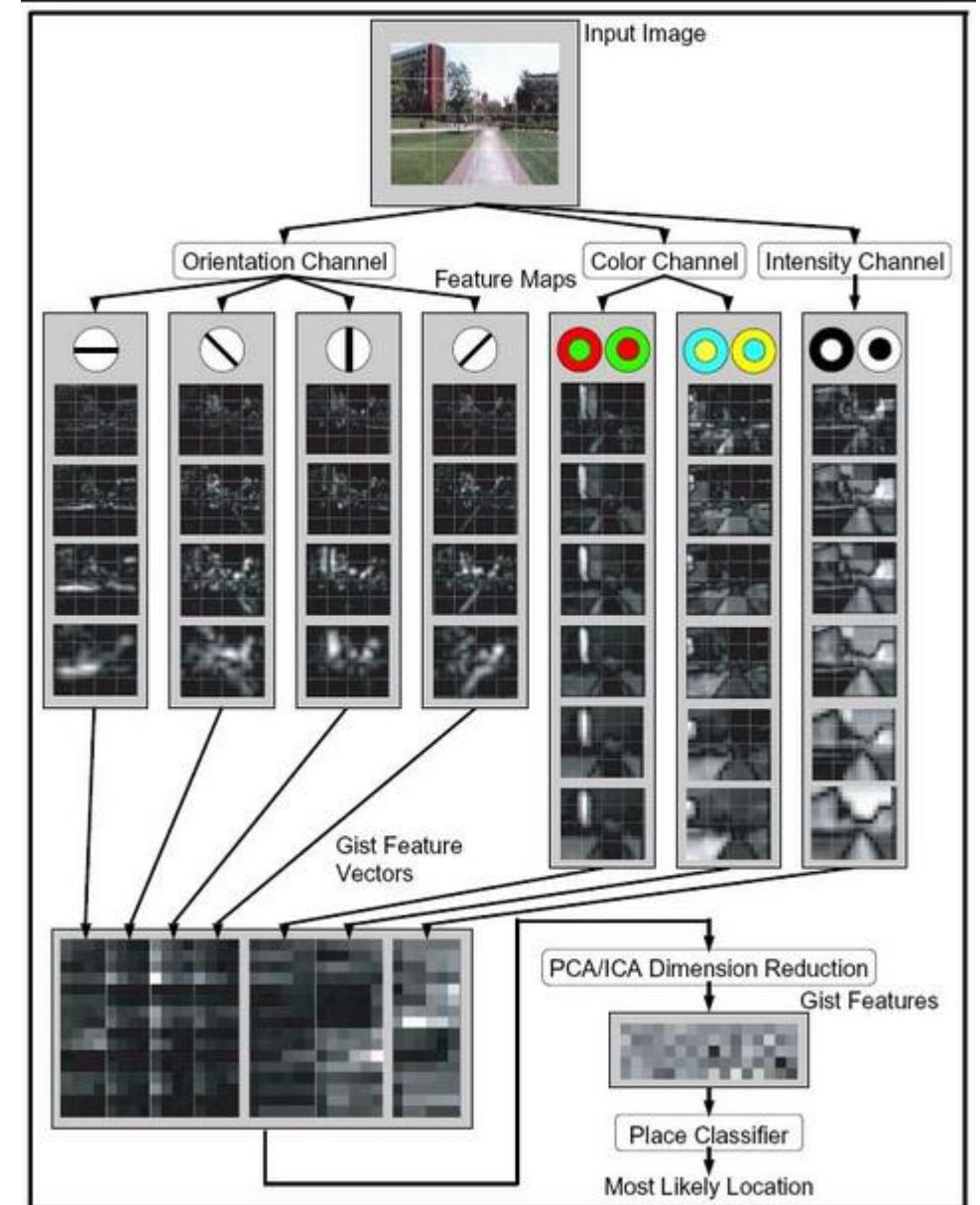


**Visual Synonyms:** Two different visual words representing a similar part of an object (wheel of a motorbike).



# GIST

For more, see <http://ilab.usc.edu/siagian/Research/Gist/Gist.html>

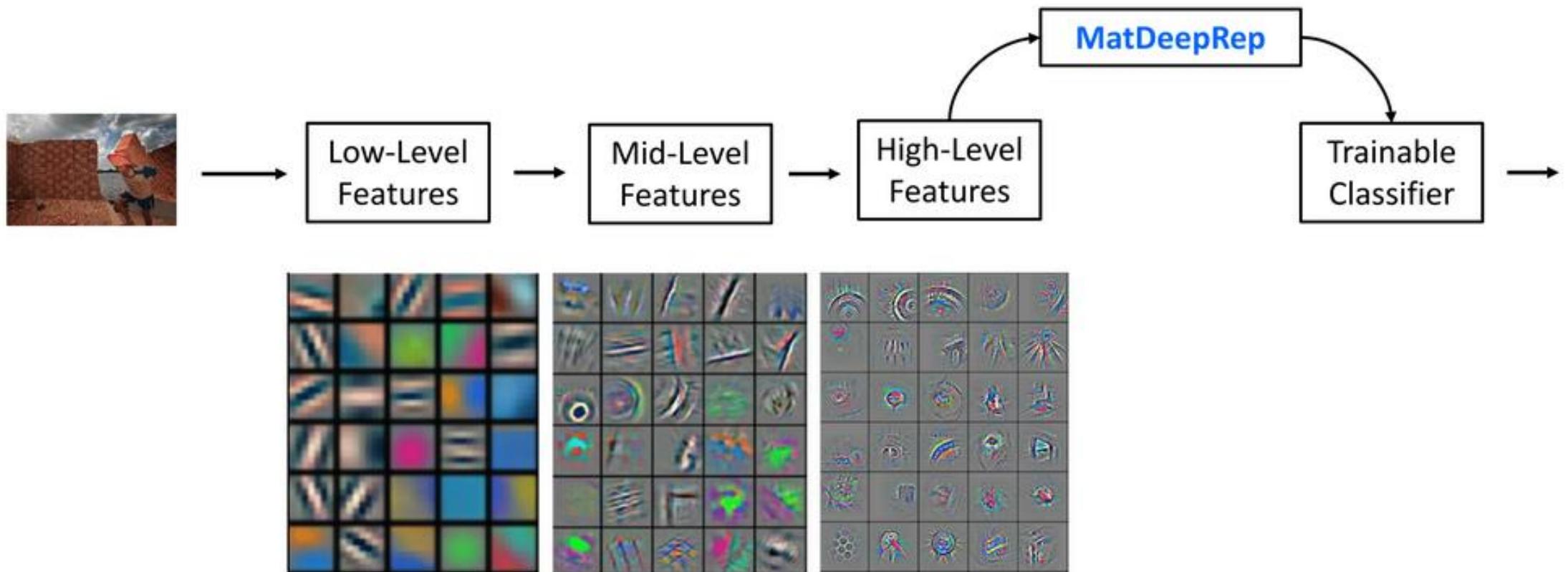


# Idea of low, mid and high level features

- Edge, point, line – low level
  - Visual words – mid level
  - Object feature – high level (more semantic meanings)
- 
- ✓ However, we have different set of techniques for extracting such features before deep learning
  - ✓ Those techniques are not optimized
  - ✓ No idea on which feature is good for which data/task

**But, this gives the notion of hierarchical features!**

# In deep learning



# Suggested reading

1940

*PROCEEDINGS OF THE IRE*

*November*

## What the Frog's Eye Tells the Frog's Brain\*

J. Y. LETTVIN†, H. R. MATORANA‡, W. S. McCULLOCH||, SENIOR MEMBER, IRE,  
AND W. H. PITTS||

# Suggested readings

## Recent Advances in Features Extraction and Description Algorithms: A Comprehensive Survey

Ehab Salahat, *Member, IEEE*, and Murad Qasaimeh, *Member, IEEE*