Image representation

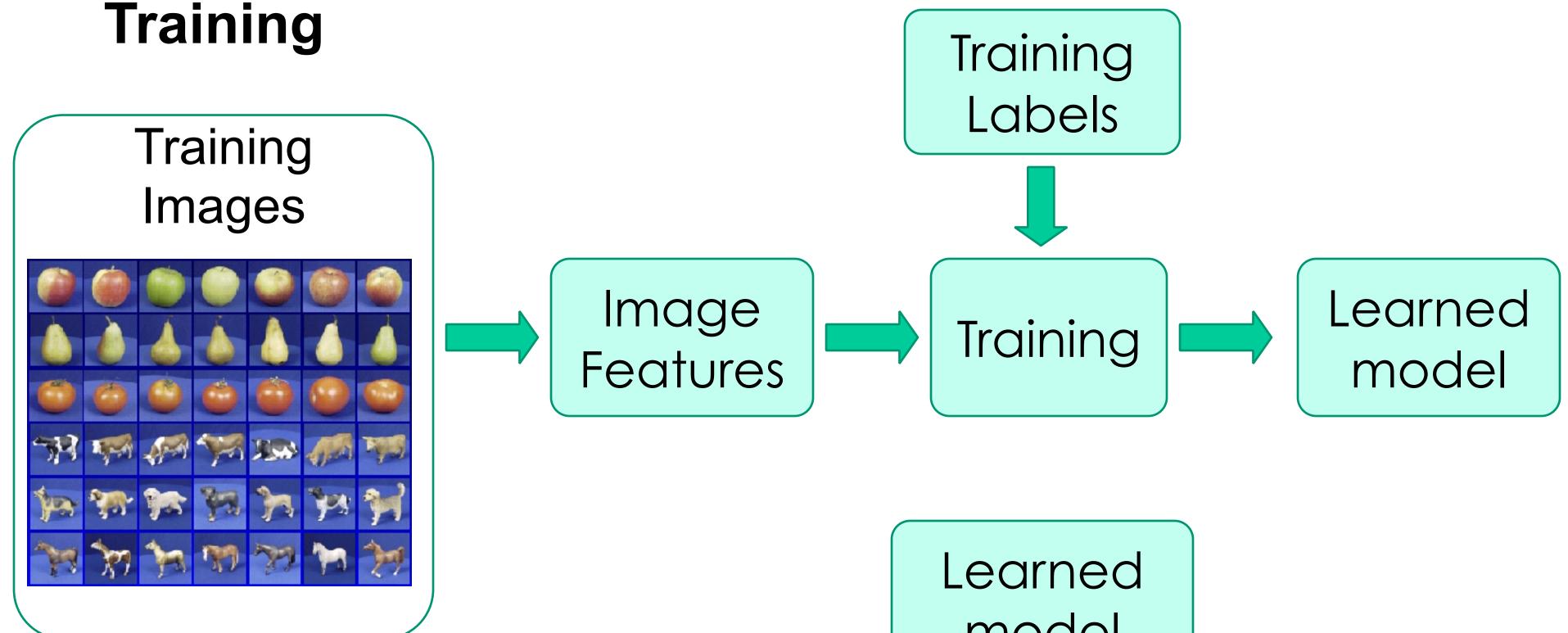
370: Intro to Computer Vision

Subhransu Maji April 22 & 24, 2025

College of **INFORMATION AND COMPUTER SCIENCES**



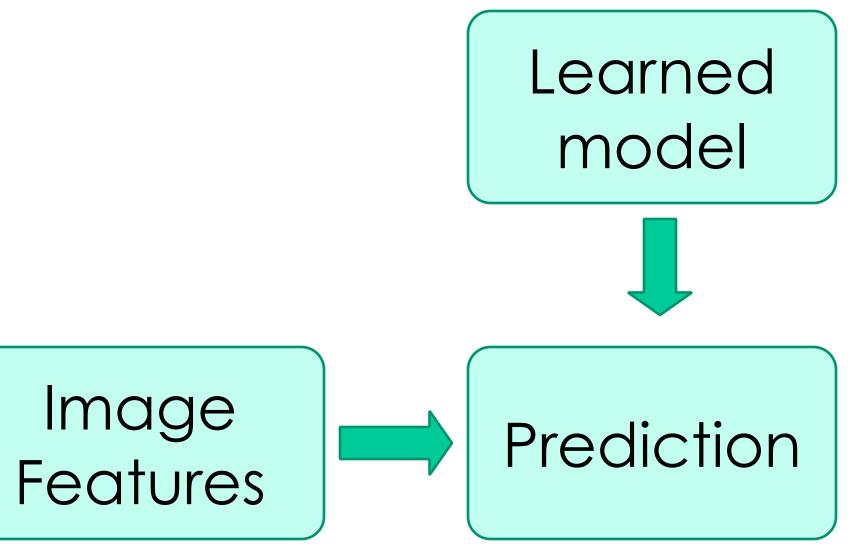
Steps — a classical perspective



Testing







Subhransu Maji – UMass Amherst, Spring 25

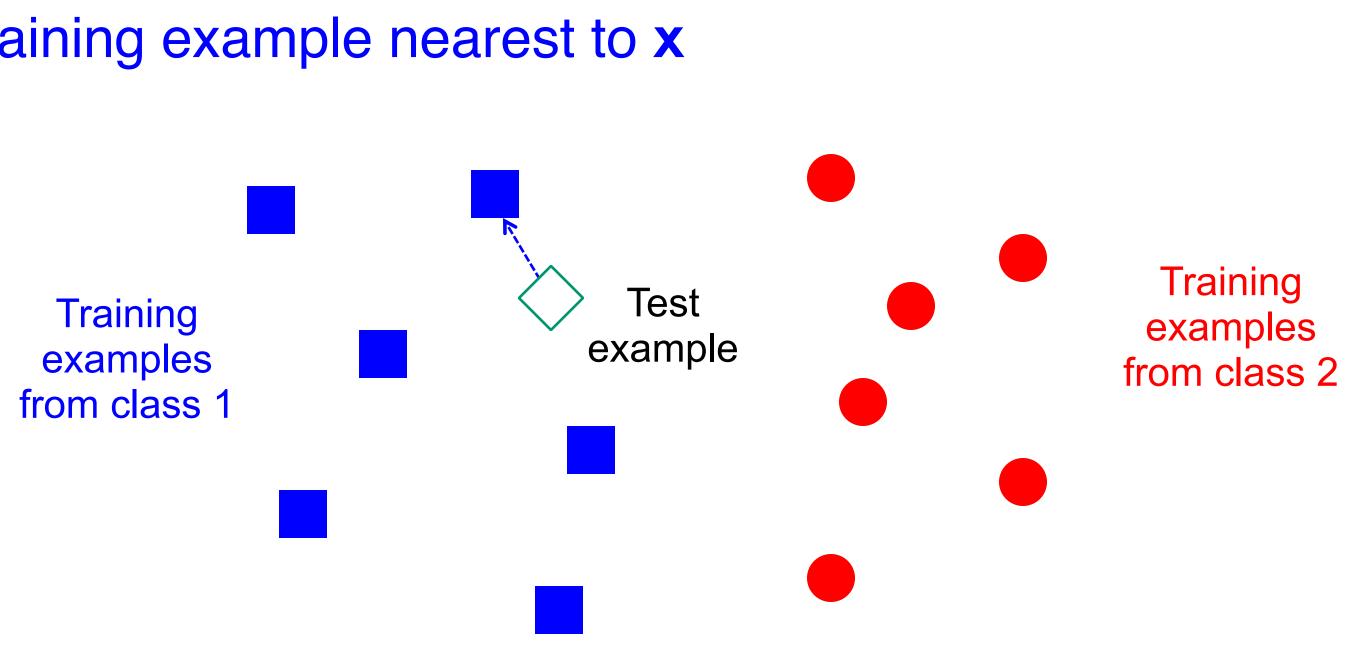
COMPSCI 370

Slide credit: D. Hoiem



Classifiers: Nearest neighbor

 $f(\mathbf{x}) =$ label of the training example nearest to \mathbf{x}



All we need is a distance function for our inputs No training required!

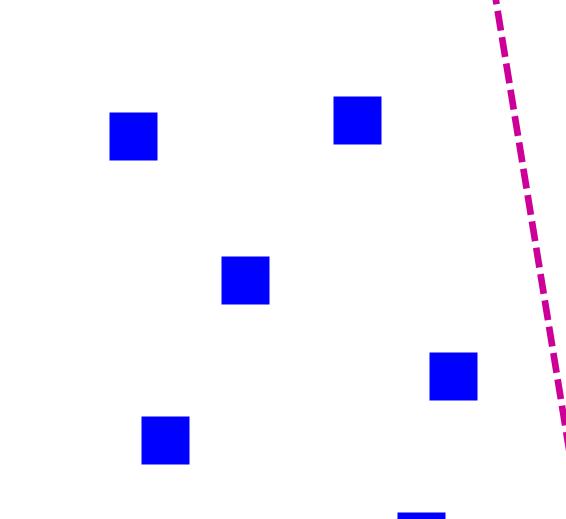
COMPSCI 370



Classifiers: Linear

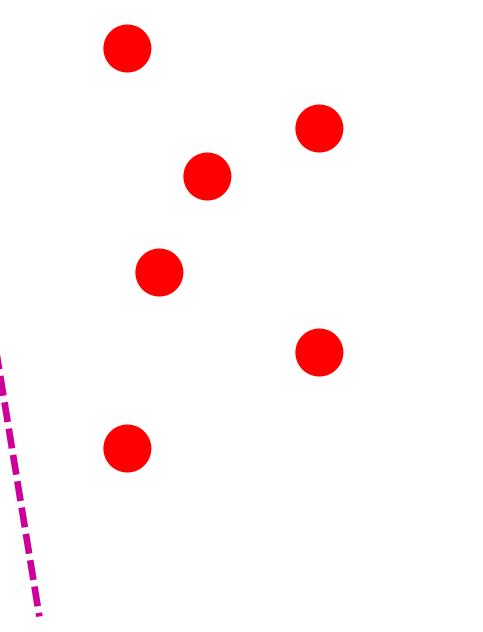
Find a linear function to separate the classes:

 $f(\mathbf{x}) = sign(\mathbf{w} \cdot \mathbf{x} + b)$



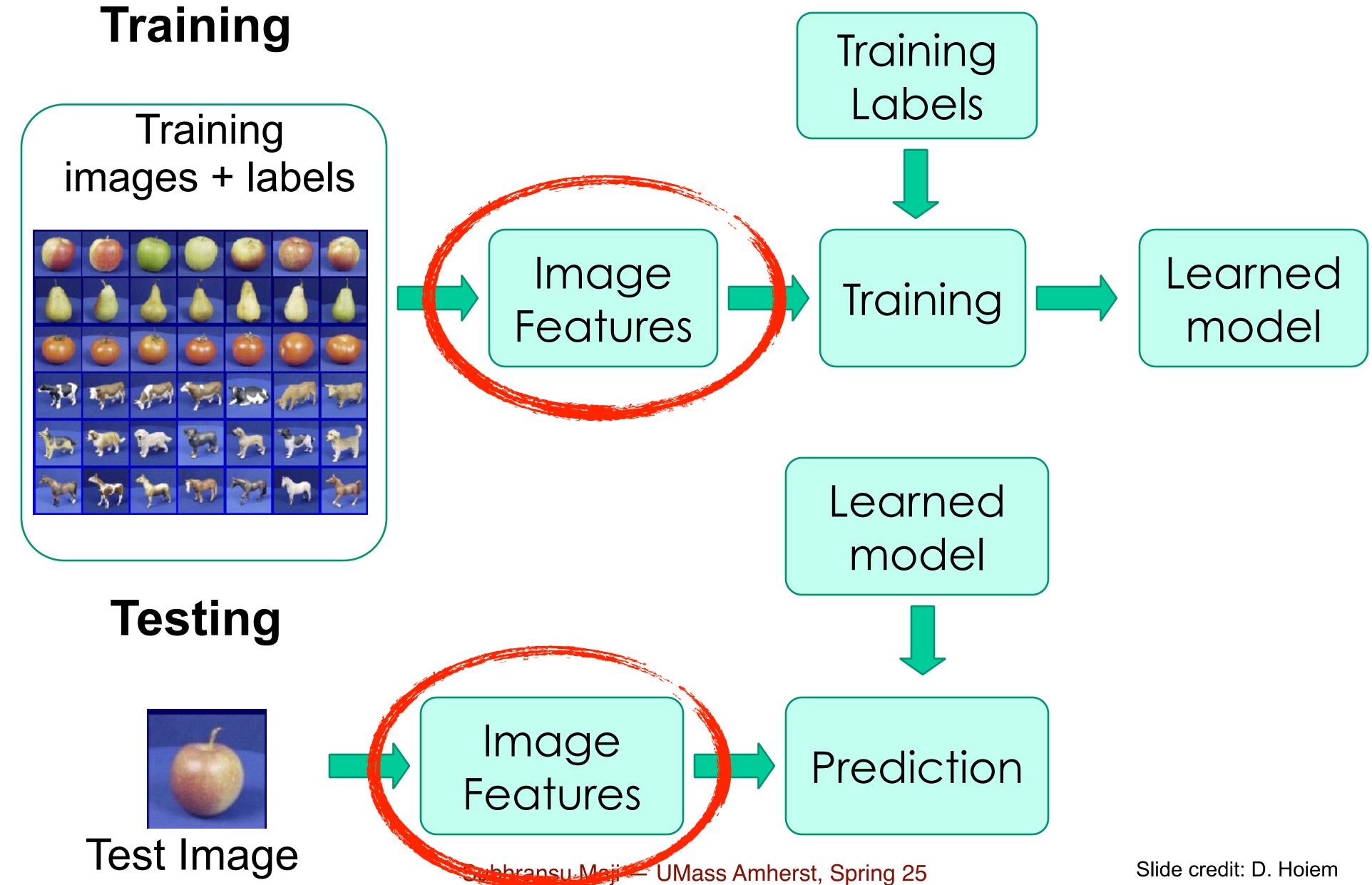
COMPSCI 370







The machine learning approach: today







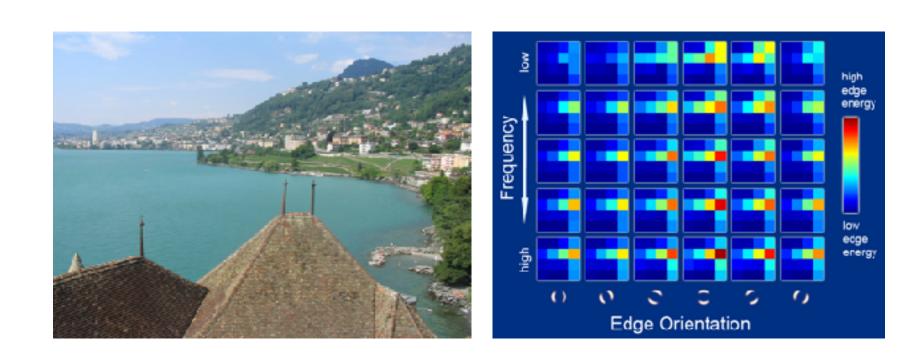


Features (examples)

Raw pixels (and simple functions of raw pixels)

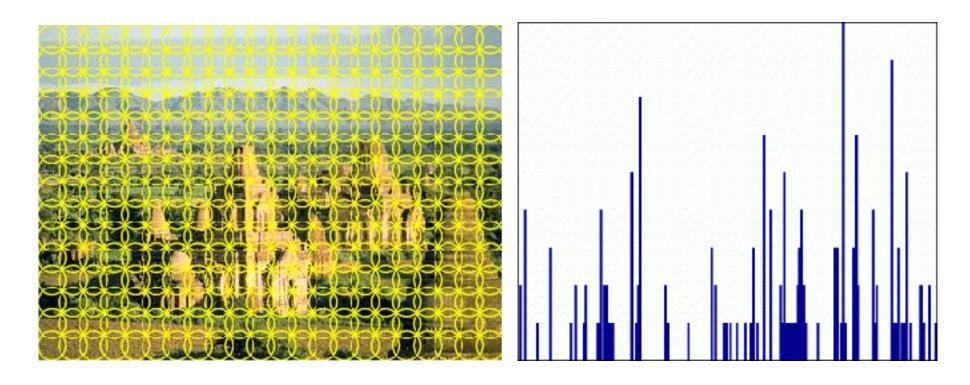


GIST descriptors

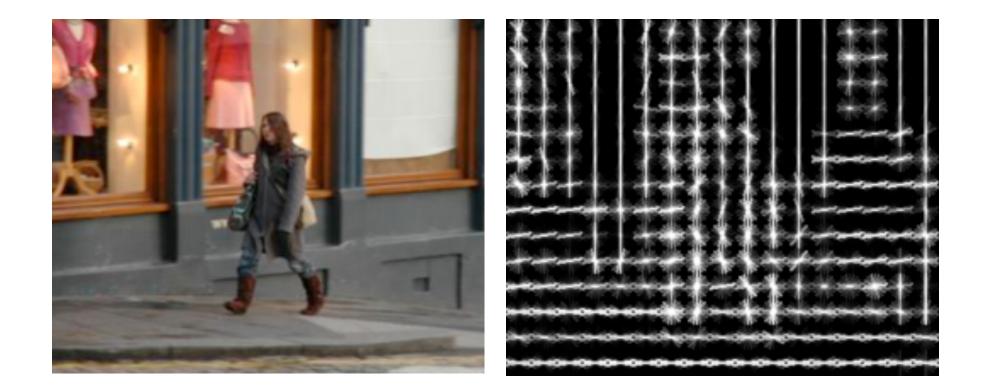


COMPSCI 370

Histograms, bags of features



Histograms of oriented gradients(HOG)





What is a feature map?

Any transformation of an image into a new representation Example: transform an image into a binary edge map

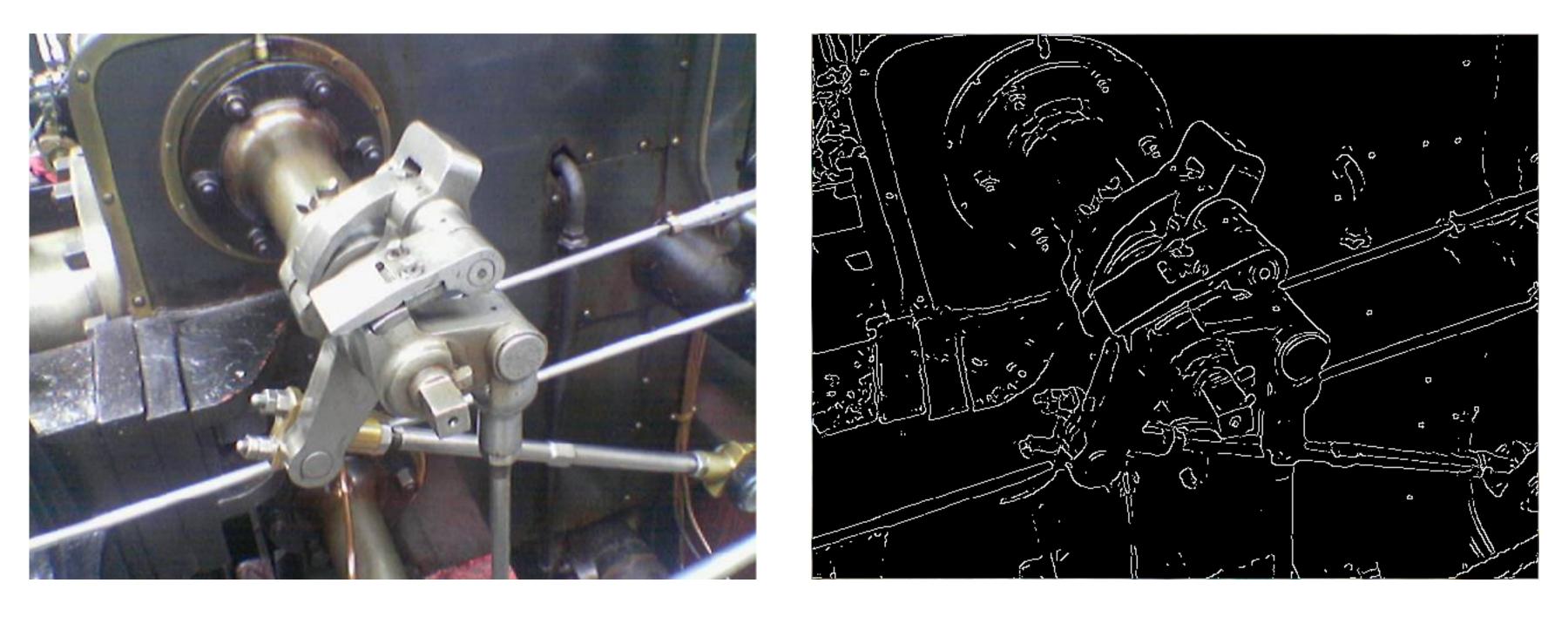


Image source: wikipedia

COMPSCI 370

Subhransu Maji – UMass Amherst, Spring 25



7

Feature map goals

Introduce invariance to nuisance factors

- Illumination changes
- Small translations, rotations, scaling, shape deformations



Preserve useful information: e.g., spatial structure

Image: [Fergus05]

Subhransu Maji – UMass Amherst, Spring 25

COMPSCI 370



Figure 1.3: Variation in appearance due to a change in illumination



We will discuss ...

Two popular image features

- Histogram of Oriented Gradients (HOG)
- Bag of Visual Words (BoVW)

Applications of these features



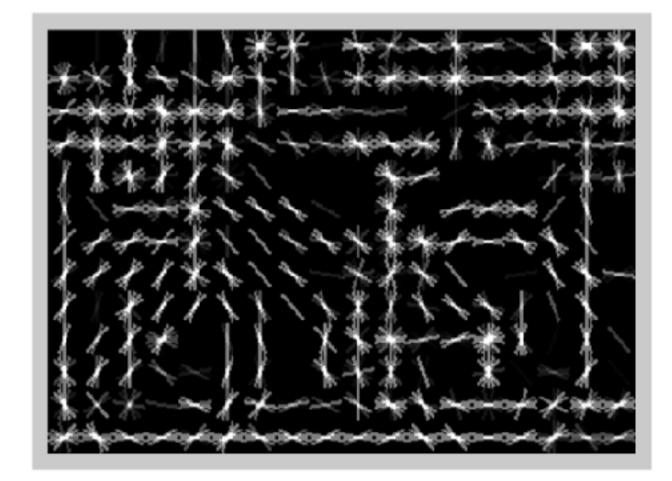
Histogram of Oriented Gradients

Introduced by Dalal and Triggs (CVPR 2005) An extension of the SIFT feature HOG properties:

- Preserves the overall structure of the image
- Provides robustness to illumination and small deformations lacksquare







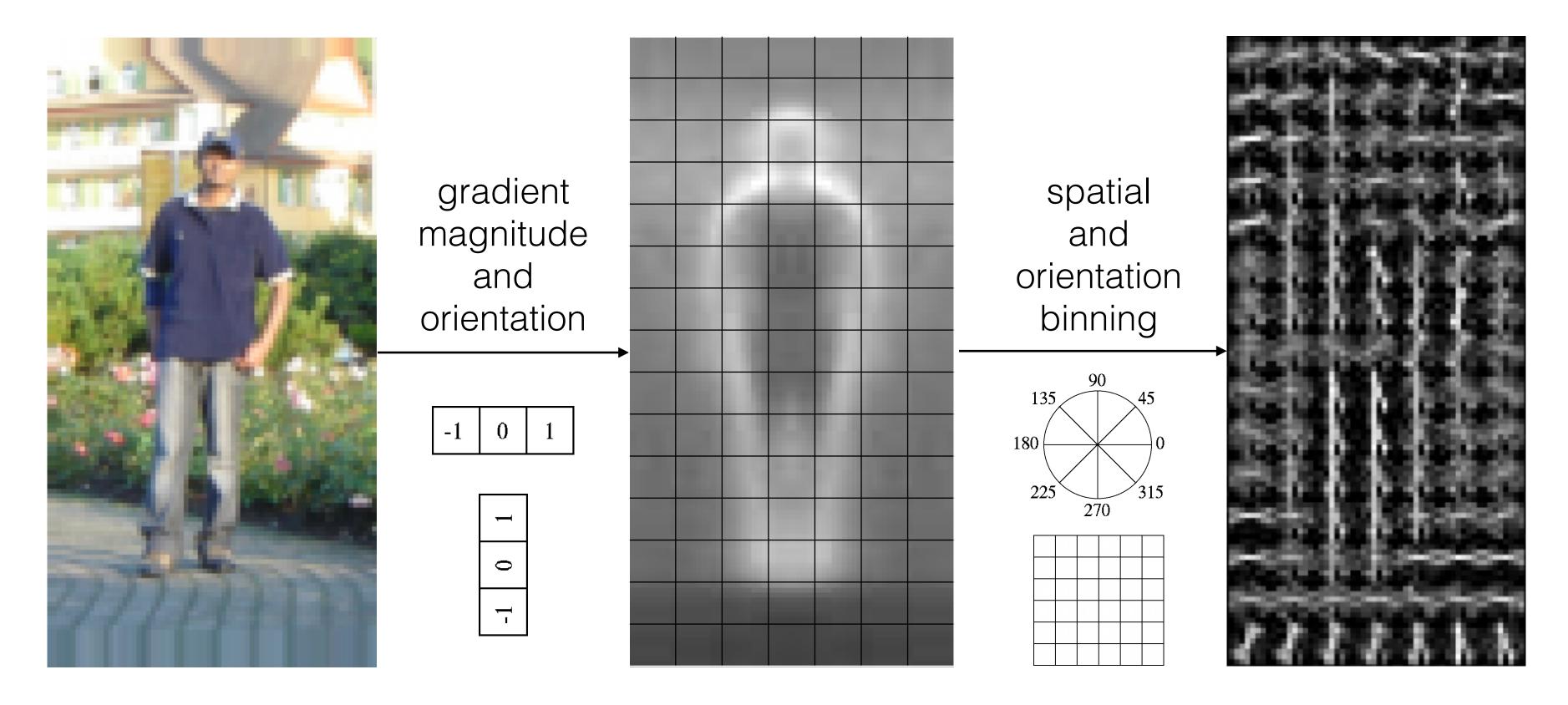
HOG feature





HOG feature: basic idea

Divide the image into blocks Compute histograms of gradients for each regions



image

Subhransu Maji – UMass Amherst, Spring 25

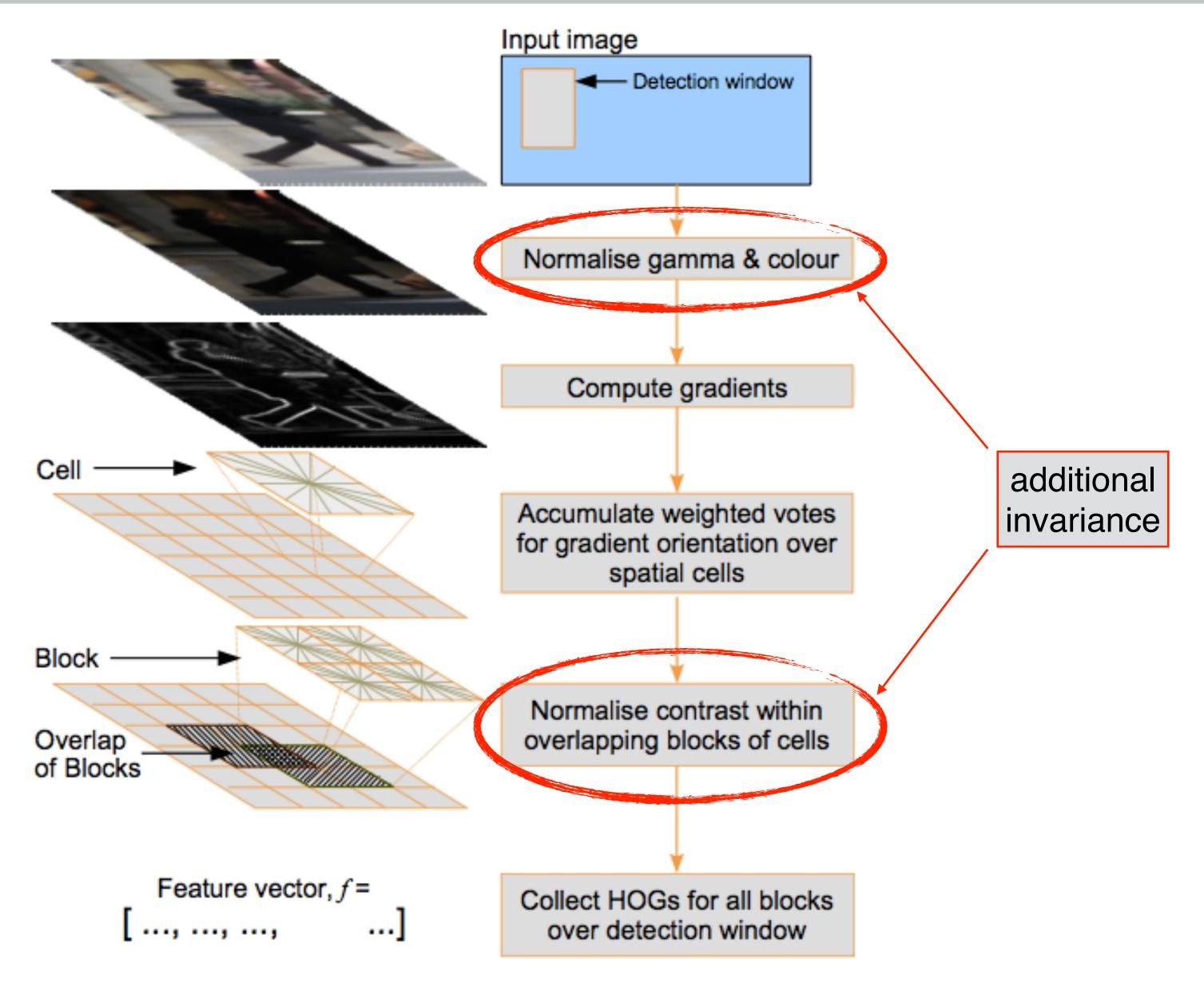
COMPSCI 370

Gradient norm

HOG feature

11

HOG feature: full pipeline



COMPSCI 370

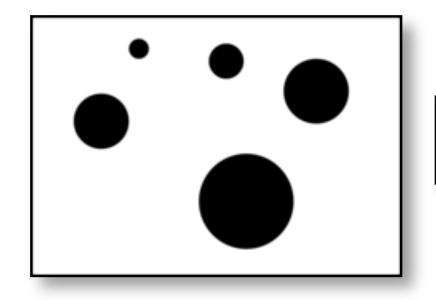


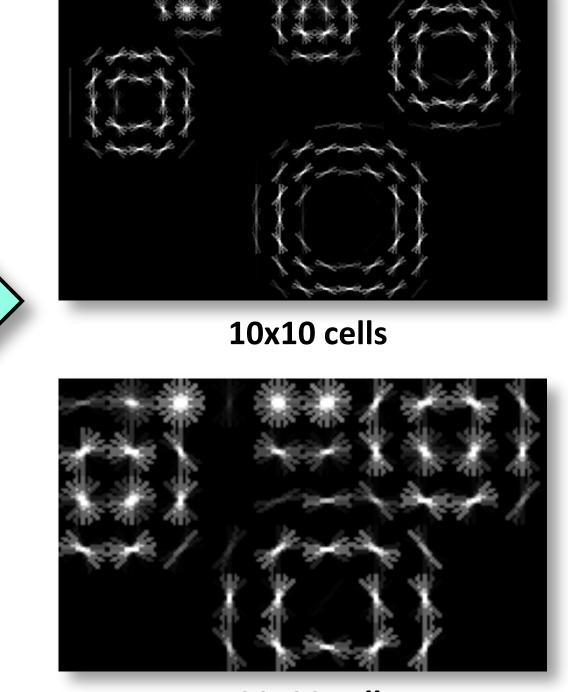


Effect of bin-size

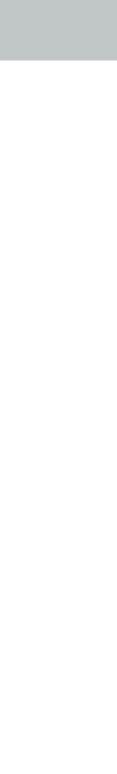
Smaller bin-size: better spatial resolution Larger bin-size: better invariance to deformations Optimal value depends on the object category being modeled

• e.g. rigid vs. deformable objects



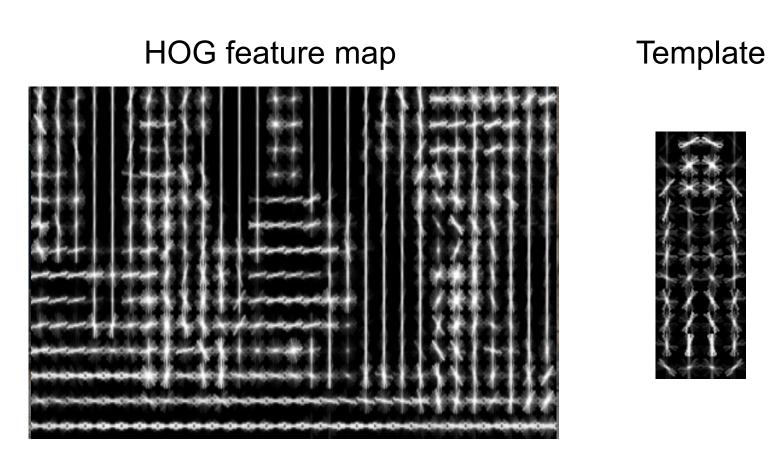


20x20 cells





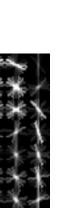
Template matching with HOG



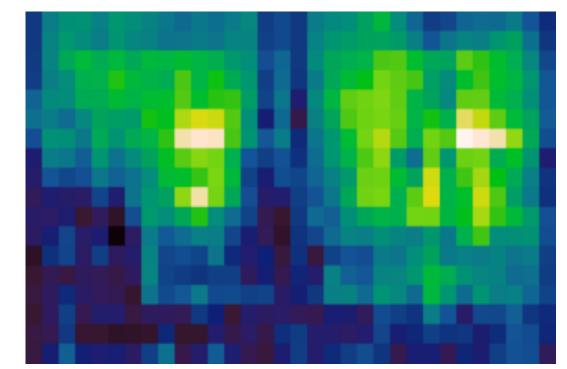
Compute the HOG feature map for the image Convolve the template with the feature map to get score Find peaks of the response map (non-max suppression)

Apply it at multiple scales

COMPSCI 370



Detector response map







Multi-scale detection

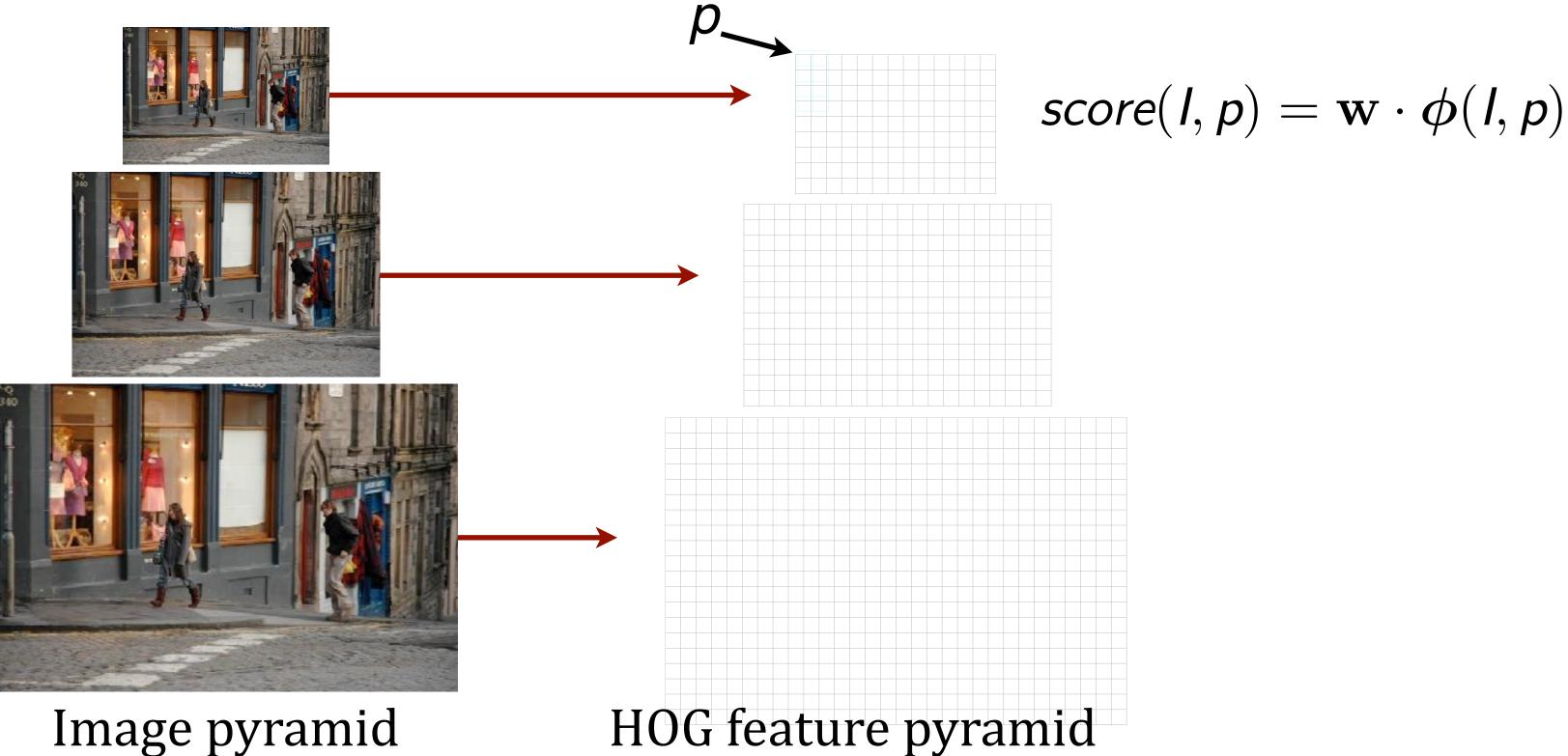


Image pyramid

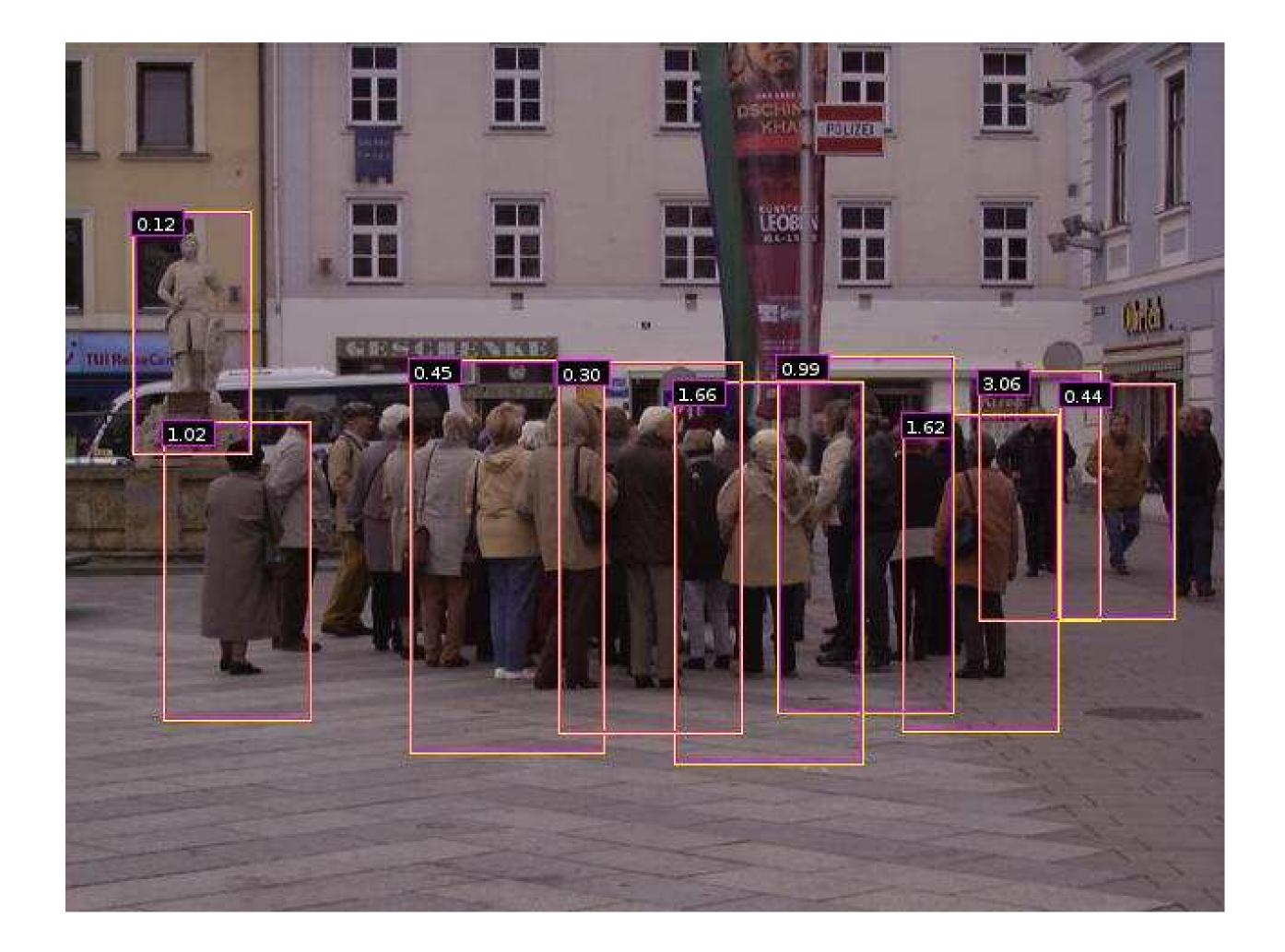
Compute HOG of the whole image at multiple resolutions Score each sub-windows of the feature pyramid

COMPSCI 370





Example detections



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

Subhransu Maji – UMass Amherst, Spring 25

COMPSCI 370





Example detections



Subhransu Maji – UMass Amherst, Spring 25

COMPSCI 370

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



17

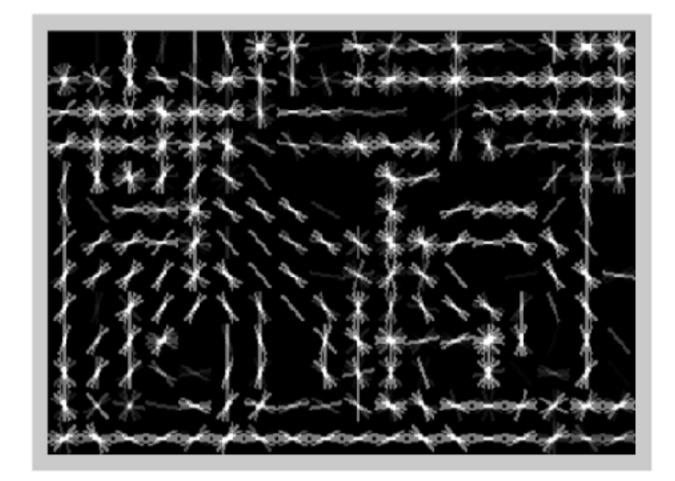
Summary: Histogram of Oriented Gradients

Introduced by Dalal and Triggs (CVPR 2005) An extension of the SIFT feature HOG properties:

- Preserves the overall structure of the image
- Provides robustness to illumination and small deformations lacksquare







HOG feature





We will discuss ...

Two popular image features

- Histogram of Oriented Gradients (HOG)
- Bag of Visual Words (BoVW)

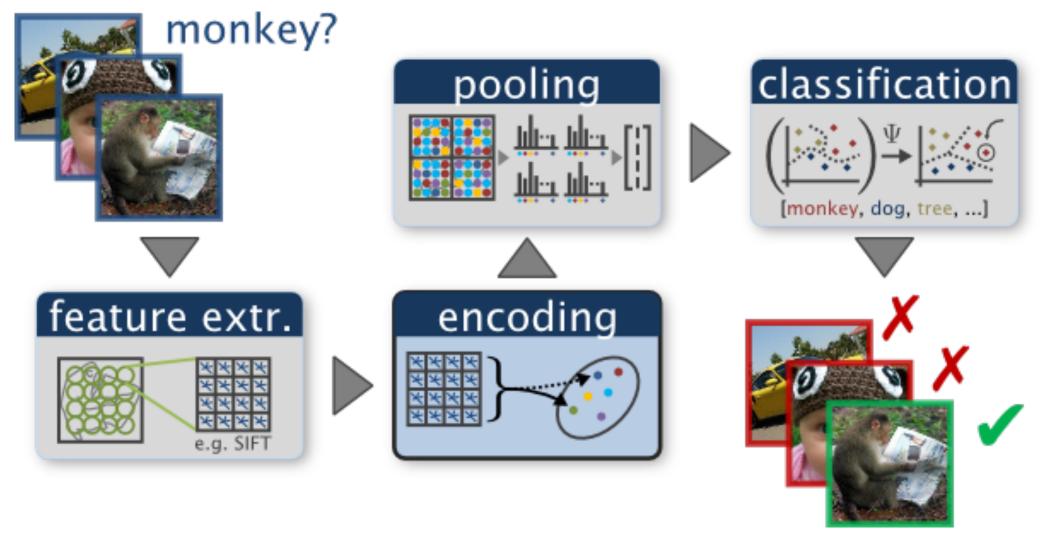




Bag of visual words

Origin and motivation of the "bag of words" model Algorithm pipeline

- Extracting local features ullet
- Learning a dictionary clustering using k-means \bullet
- Encoding methods hard vs. soft assignment \bullet
- Spatial pooling pyramid representations
- Similarity functions and classifiers



COMPSCI 370

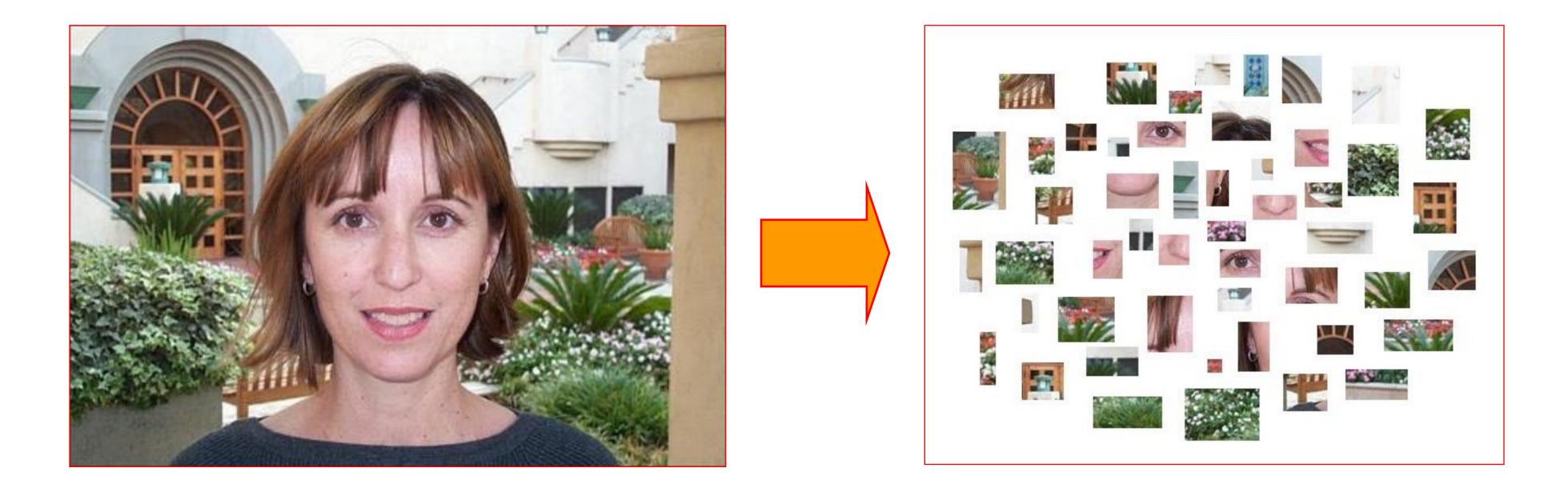


Subhransu Maji – UMass Amherst, Spring 25

Figure from *Chatfield et al.,2011*



Bag of features



Properties:

- Invariance to large translations

Compare this to the HOG feature

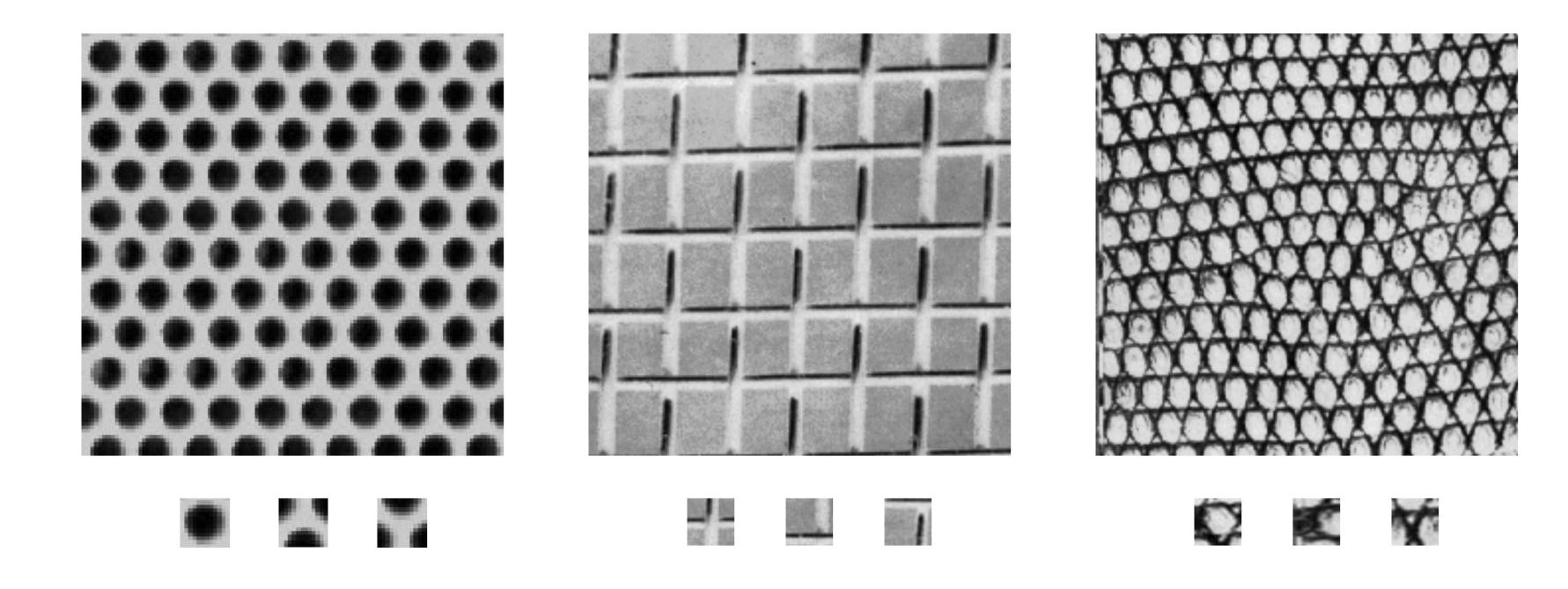
COMPSCI 370

Spatial structure is not preserved



Origin 1: Texture recognition

Texture is characterized by the repetition of basic elements or *textons*



Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

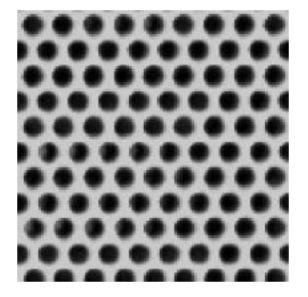
COMPSCI 370

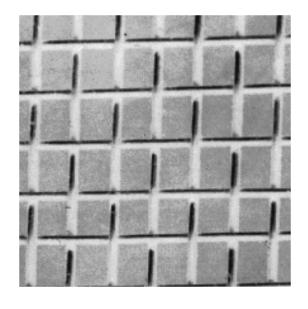
- For stochastic textures, it is the identity of the textons, not their spatial arrangement, that matters

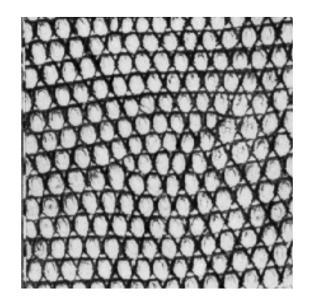




Origin 1: Texture recognition

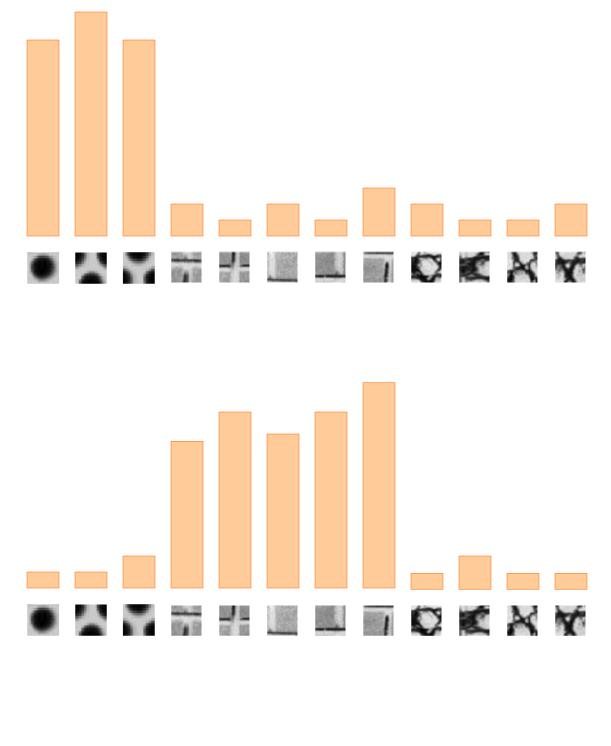


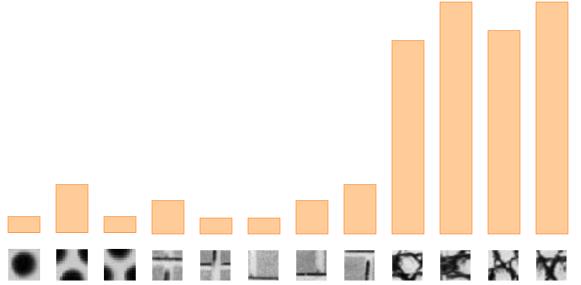






COMPSCI 370







Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

COMPSCI 370



Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

2007-01-23: State of the Union Address

abandon accountable affordable afghanistan africa aided ally anbar armed army baghdad bless challenges chamber chaos choices civilians coalition commanders commitment confident confront congressman constitution corps debates deduction deficit deliver democratic deploy dikembe diplomacy disruptions earmarks economy einstein elections eliminates expand extremists failing faithful families freedom fuel funding god haven ideology immigration impose nsurgents iran Iraq islam julie lebanon love madam marine math medicare moderation neighborhoods nuclear offensive palestinian payroll province pursuing **qaeda** radical regimes resolve retreat rieman sacrifices science sectarian senate eptember shia stays strength students succeed sunni tax territories territories threats uphold victory

violence violent War washington weapons wesley

George W. Bush (2001-)



20

Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

07-0 1	1-23: State of the Union Address
ndon ices c	1962-10-22: Soviet Missiles in Cuba
icit d band (abandon achieving adversaries aggression agricultur
urgen	buildup burdens cargo college commitment con
estinia	declined defensive deficit depended disarma
	elimination emergence endangered equals europ
temb	halt hazards hemisphere hospitals ideals i
lenc	modernization neglect nuclear oas obligation of
	recession rejection republics retaliatory safeguard site

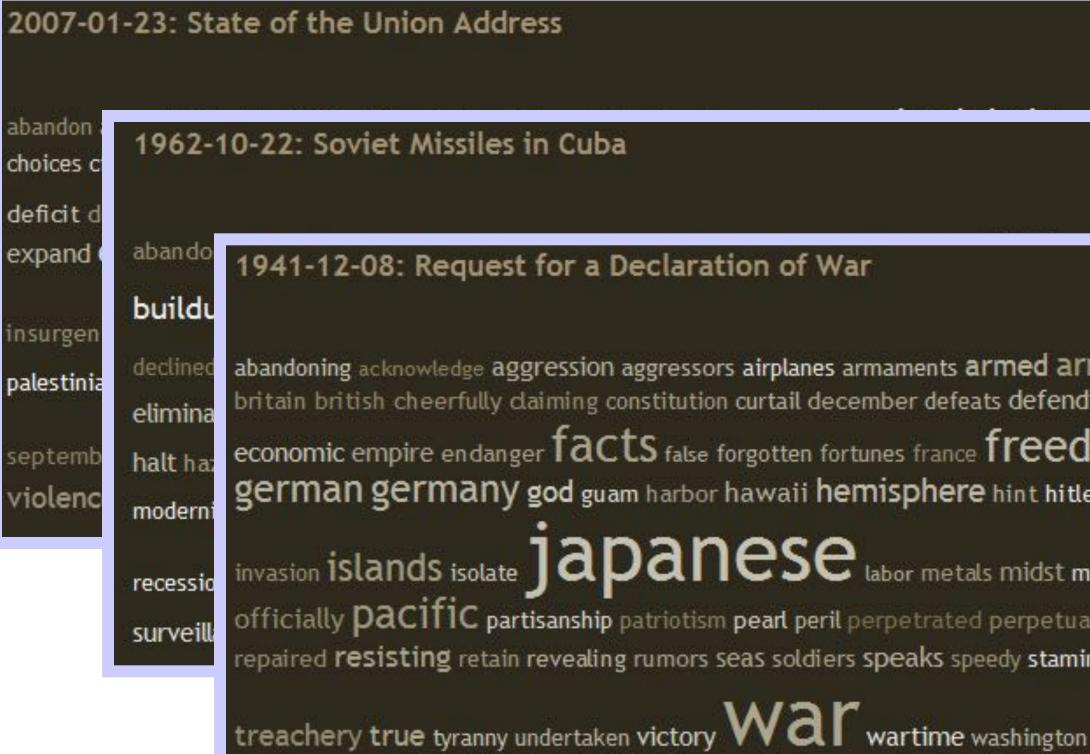
ral appropriate armaments **arms** assessments atlantic ballistic berlin mmunist constitution consumers cooperation crisis CUDA dangers ament divisions domination doubled **economic education** pe expand exports fact false family forum freedom fulfill gromyko ndependent industries inflation labor latin limiting minister missiles observer **offensive** peril pledged **predicted** purchasing quarantine **quote** tes solution Soviet space spur stability standby strength surveillance tax territory treaty undertakings unemployment War warhead Weapons welfare western widen withdraw

George W. Bush (2001-)

John F. Kennedy (1961-63)



Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)



COMPSCI 370

George	W. Bush	(2001-)
		(,

John F. Kennedy (1961-63)

Franklin D. Roosevelt (1933-45)

abandoning acknowledge aggression aggressors airplanes armaments armed army assault assembly authorizations bombing britain british cheerfully claiming constitution curtail december defeats defending delays democratic dictators disclose

economic empire endanger facts false forgotten fortunes france freedom fulfilled fullness fundamental gangsters german germany god guam harbor hawaii hemisphere hint hitler hostilities immune improving indies innumerable

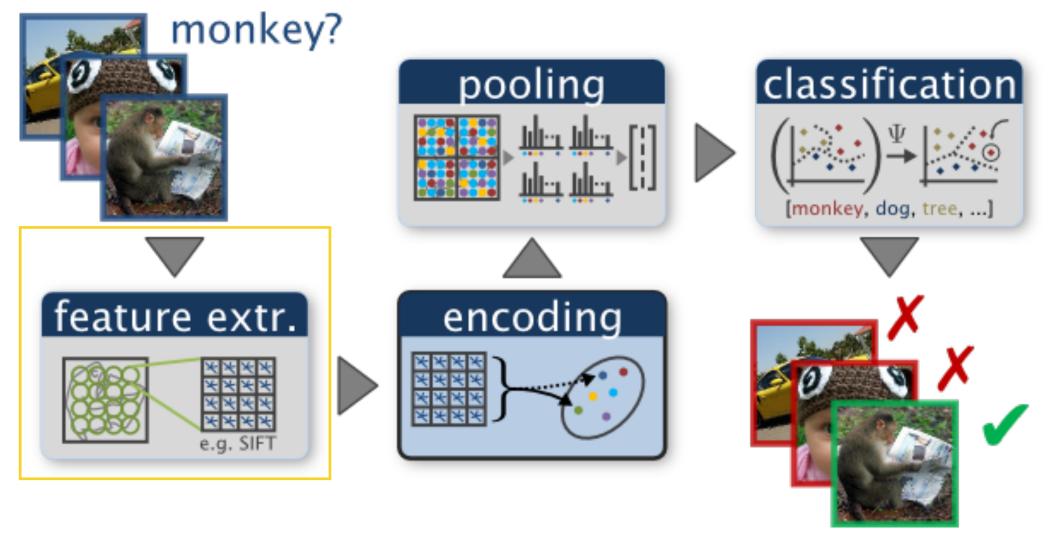
invasion islands isolate] apanese labor metals midst midway navy nazis obligation offensive officially **DACIFIC** partisanship patriotism pearl peril perpetrated perpetual philippine preservation privilege reject repaired **resisting** retain revealing rumors seas soldiers speaks speedy stamina strength sunday sunk supremacy tanks taxes



Lecture outline

Origin and motivation of the "bag of words" model Algorithm pipeline

- Extracting local features
- Learning a dictionary clustering using k-means \bullet
- Encoding methods hard vs. soft assignment \bullet
- Spatial pooling pyramid representations
- Similarity functions and classifiers



COMPSCI 370

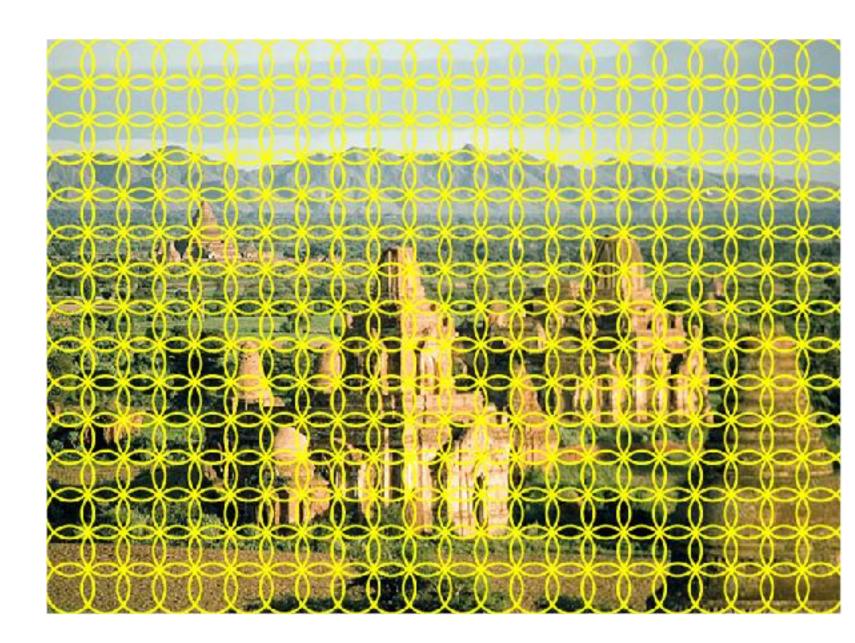
Subhransu Maji – UMass Amherst, Spring 25

Figure from *Chatfield et al.,2011*



Local feature extraction

Regular grid or interest regions



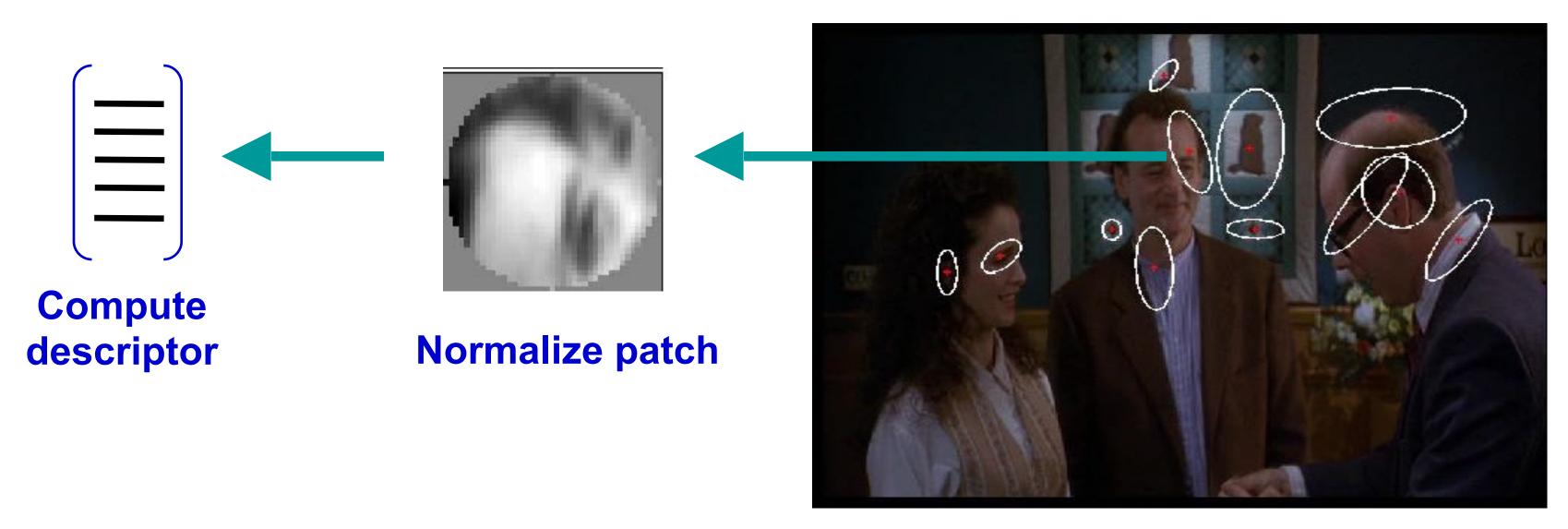
COMPSCI 370



corner detector



Local feature extraction



Choices of descriptor:

- Pixels
- SIFT

. . .

- Shape context
- Geometric blur

COMPSCI 370

Detect patches

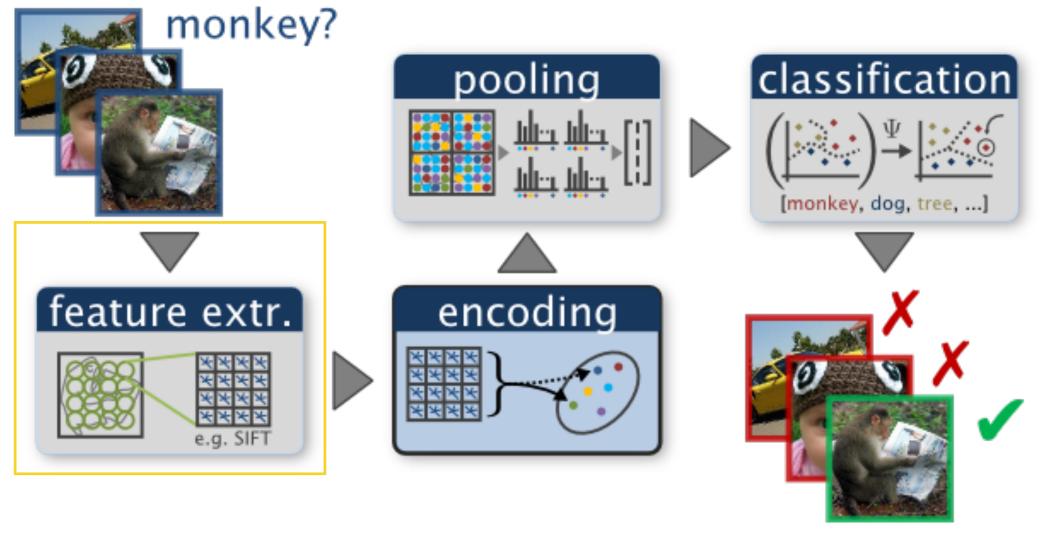
Subhransu Maji – UMass Amherst, Spring 25



Lecture outline

Origin and motivation of the "bag of words" model Algorithm pipeline

- Extracting local features \bullet
- Learning a dictionary clustering using k-means \bullet
- Encoding methods hard vs. soft assignment
- Spatial pooling pyramid representations
- Similarity functions and classifiers



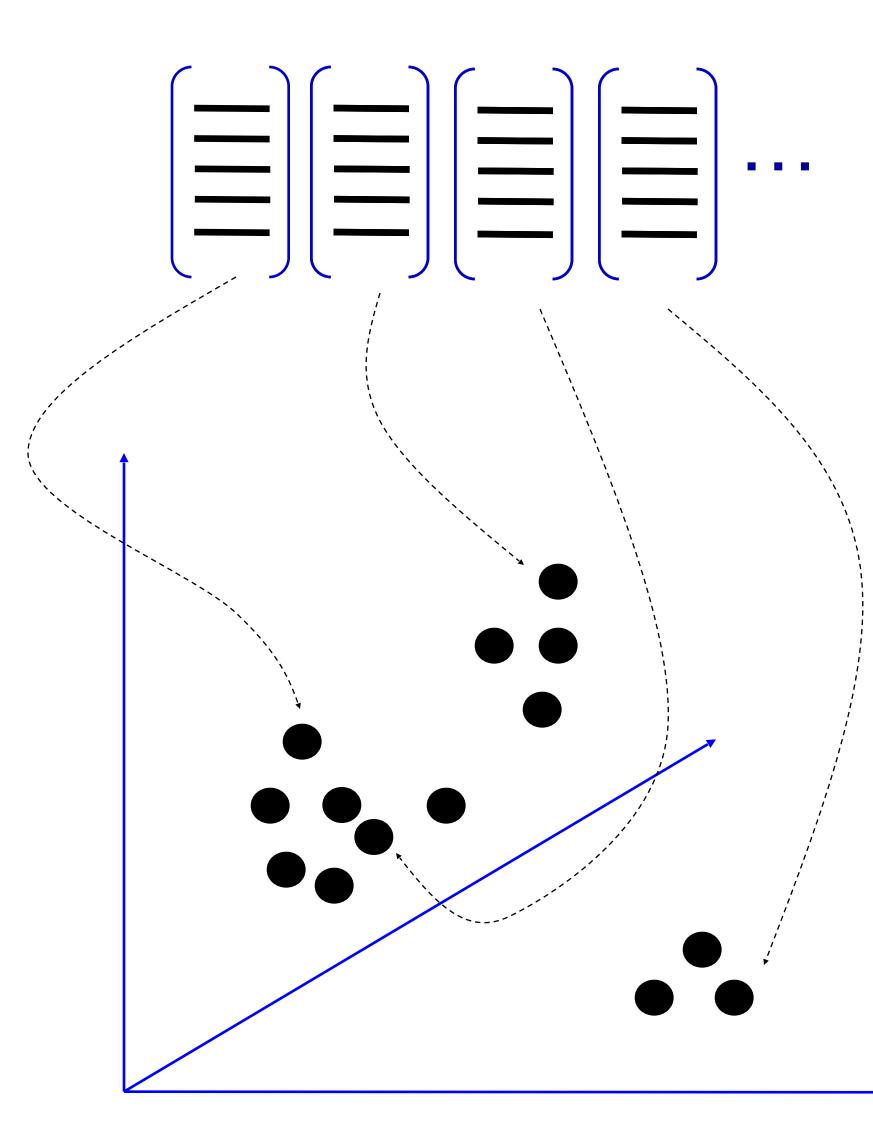
COMPSCI 370

Subhransu Maji – UMass Amherst, Spring 25

Figure from *Chatfield et al.,2011*



Learning a dictionary

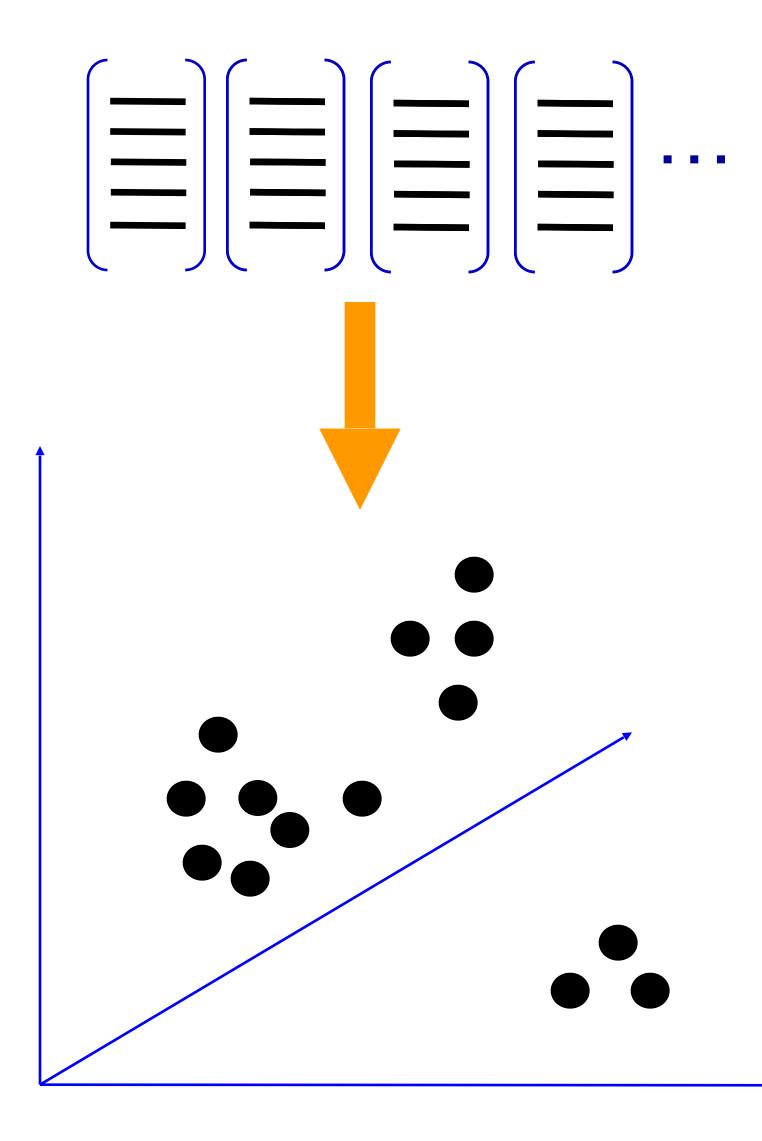


COMPSCI 370

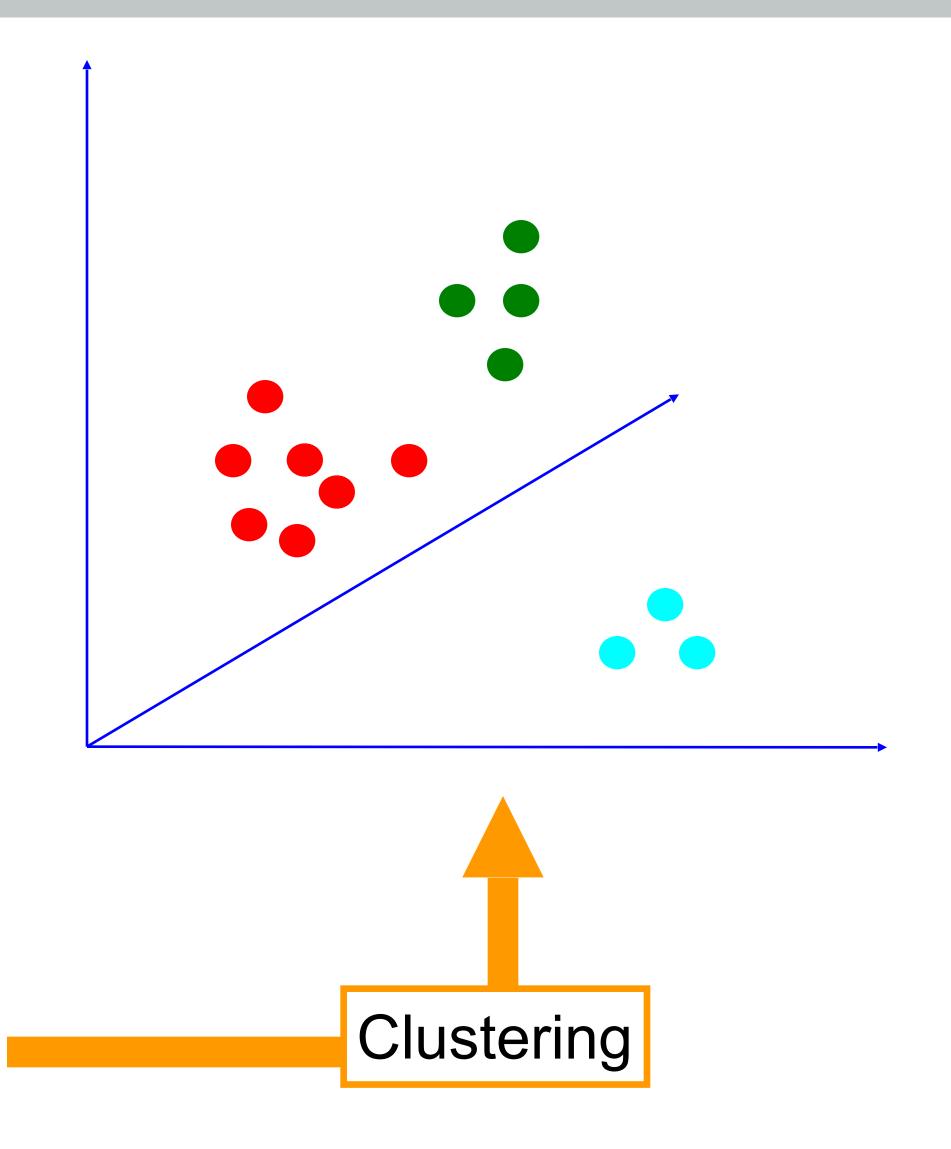
Subhransu Maji – UMass Amherst, Spring 25



Learning a dictionary



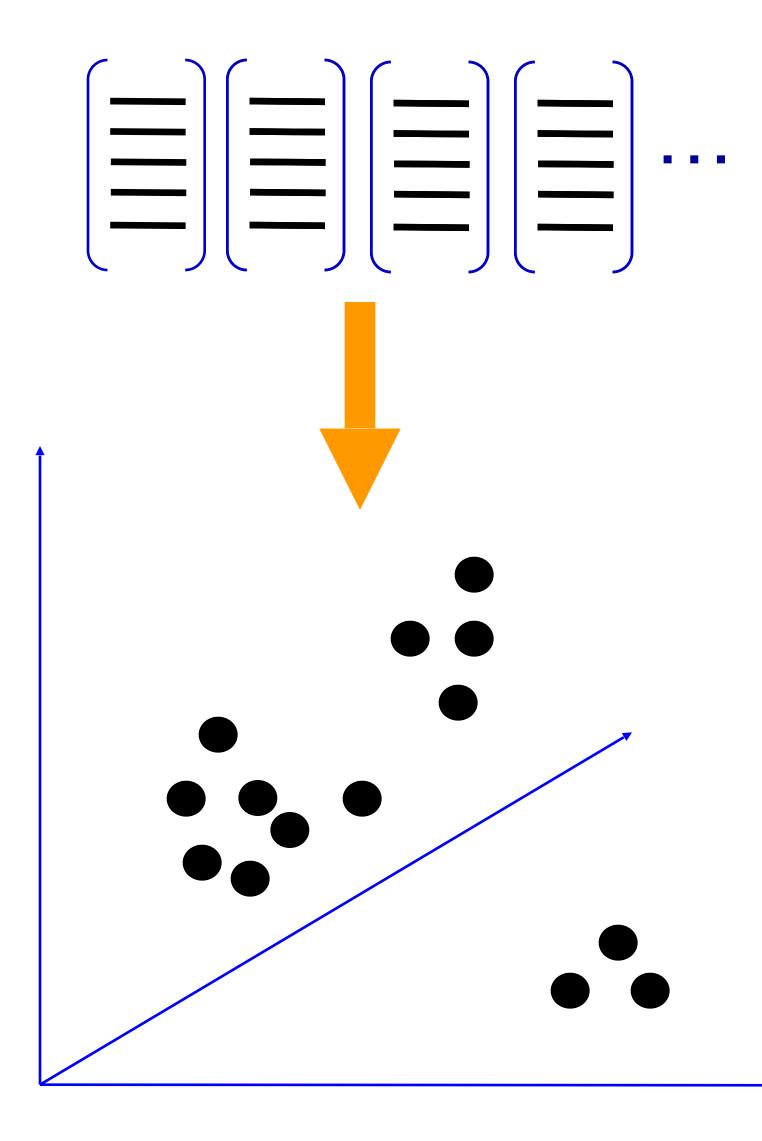
COMPSCI 370



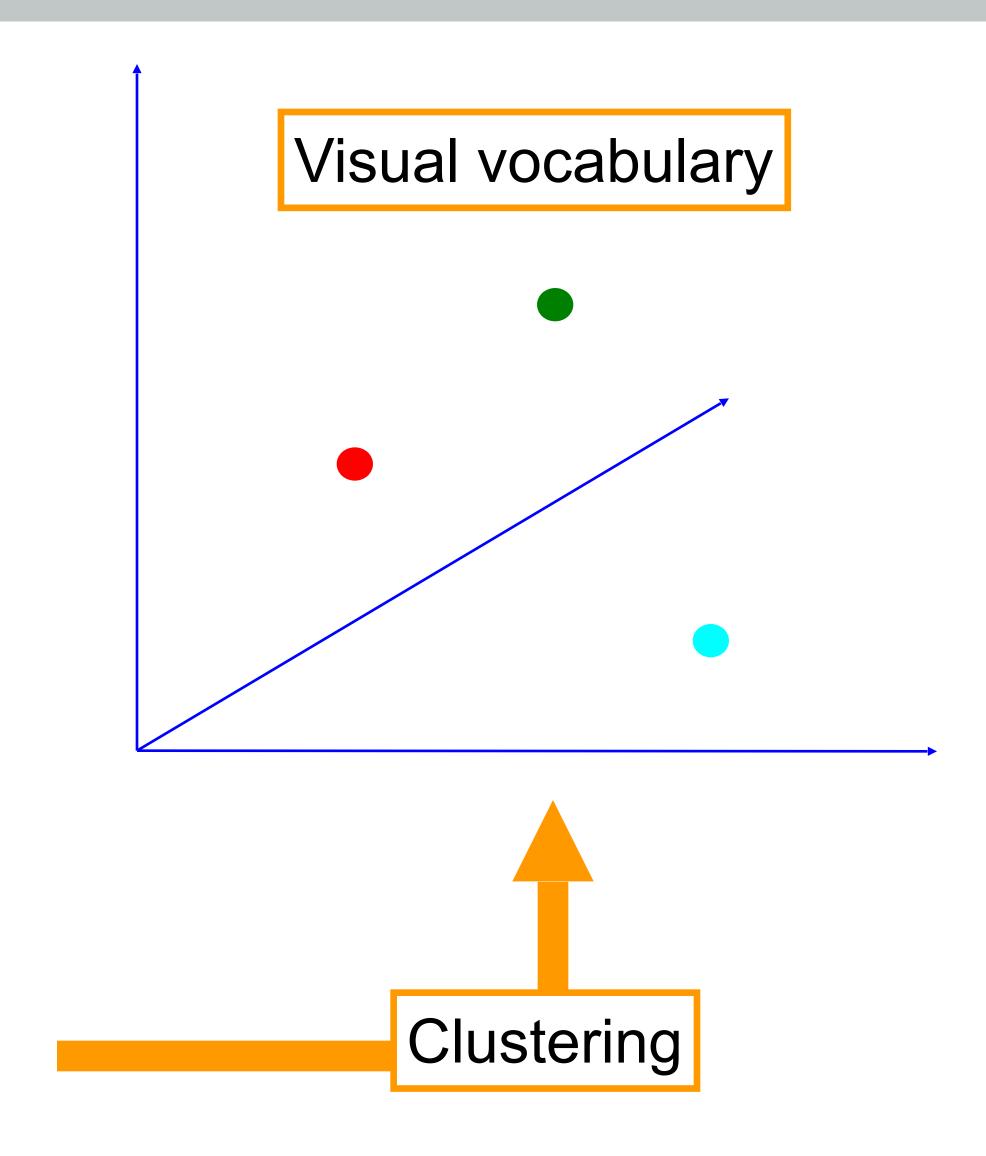
Subhransu Maji – UMass Amherst, Spring 25



Learning a dictionary



COMPSCI 370



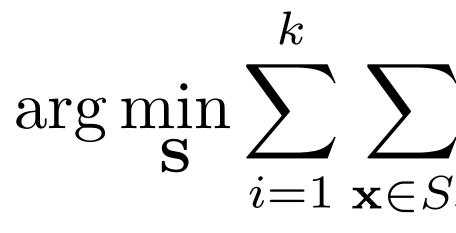
Subhransu Maji – UMass Amherst, Spring 25



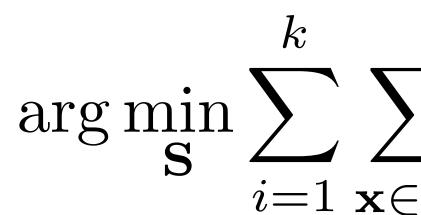
Lloyd's algorithm for k-means

Initialize k centers by picking k points randomly among all the points Repeat till convergence (or max iterations)

• Assign each point to the nearest center (assignment step)



Estimate the mean of each group (update st



$$\sum_{i=1}^{n} ||\mathbf{x} - \mu_i||^2$$

$$\sum_{i \in S_i} ||\mathbf{x} - \mu_i||^2$$



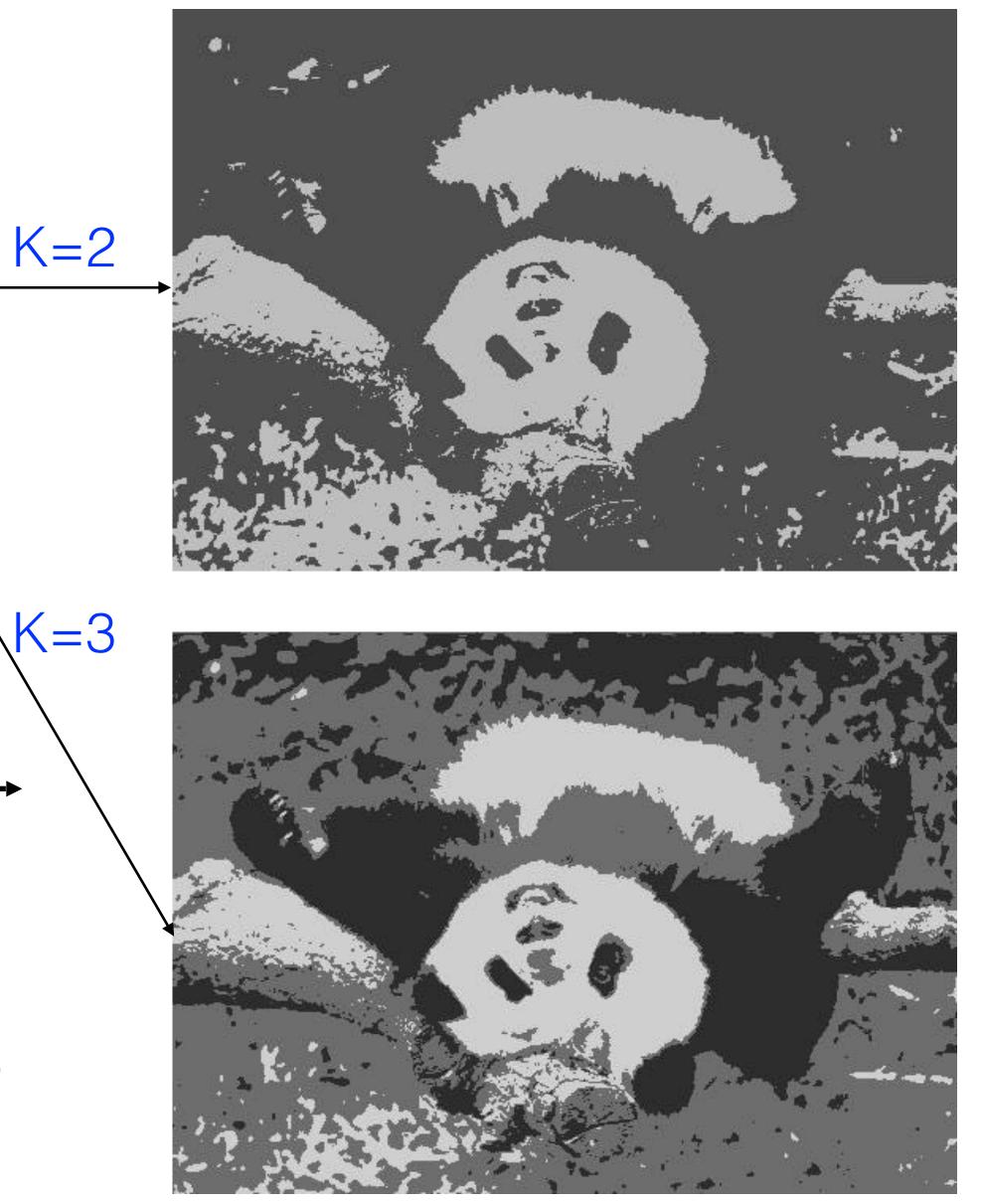
k-means for image segmentation



Grouping pixels based on **intensity** similarity

feature space: intensity value (1D)

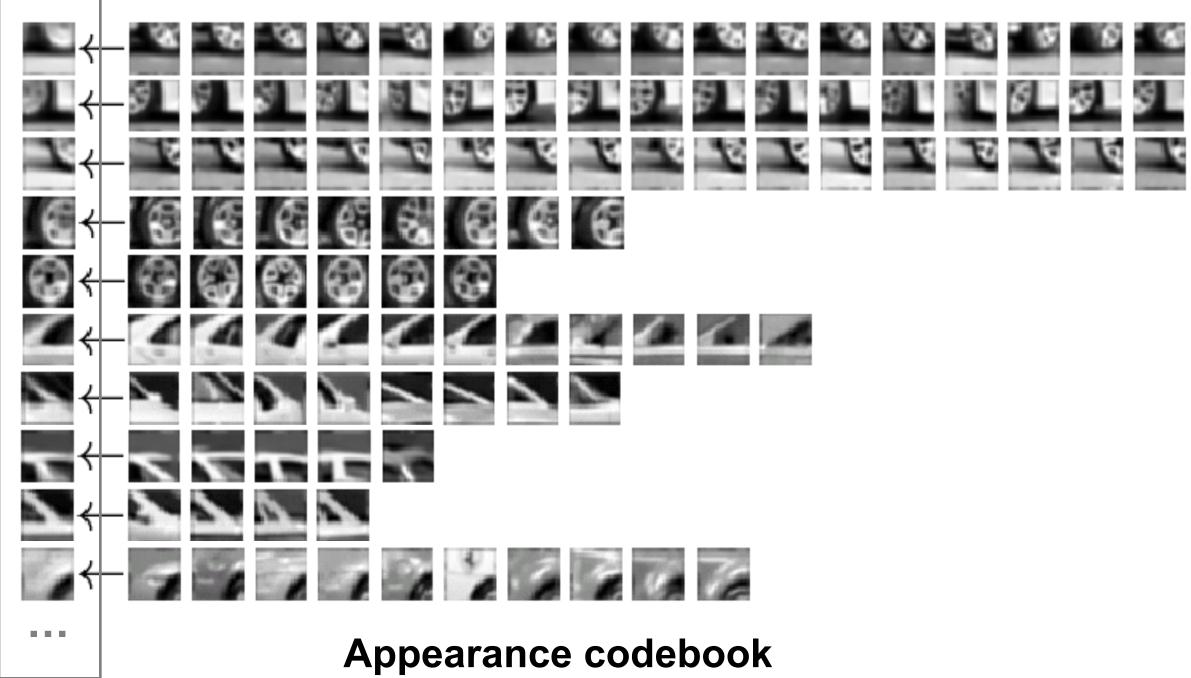
COMPSCI 370





Example codebook





Subhransu Maji – UMass Amherst, Spring 25

COMPSCI 370

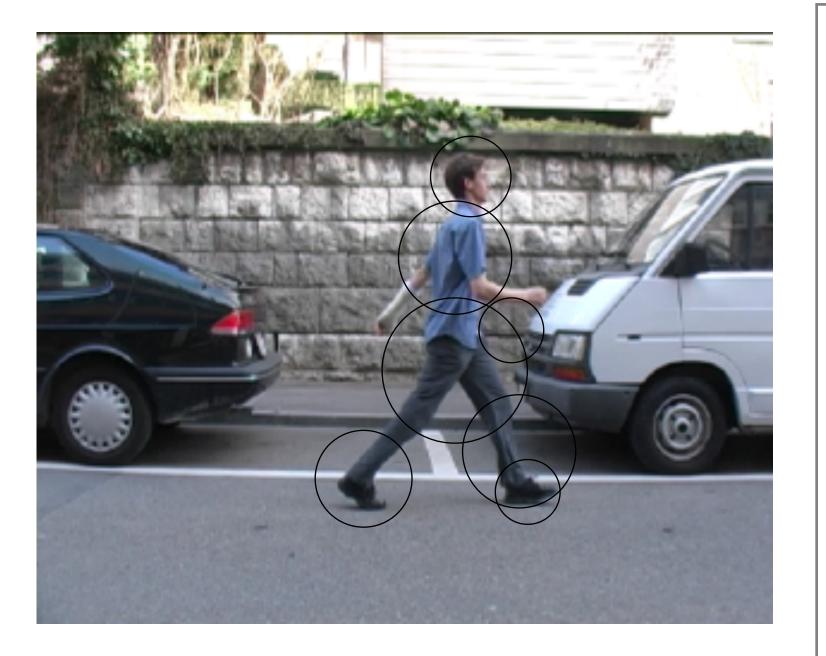




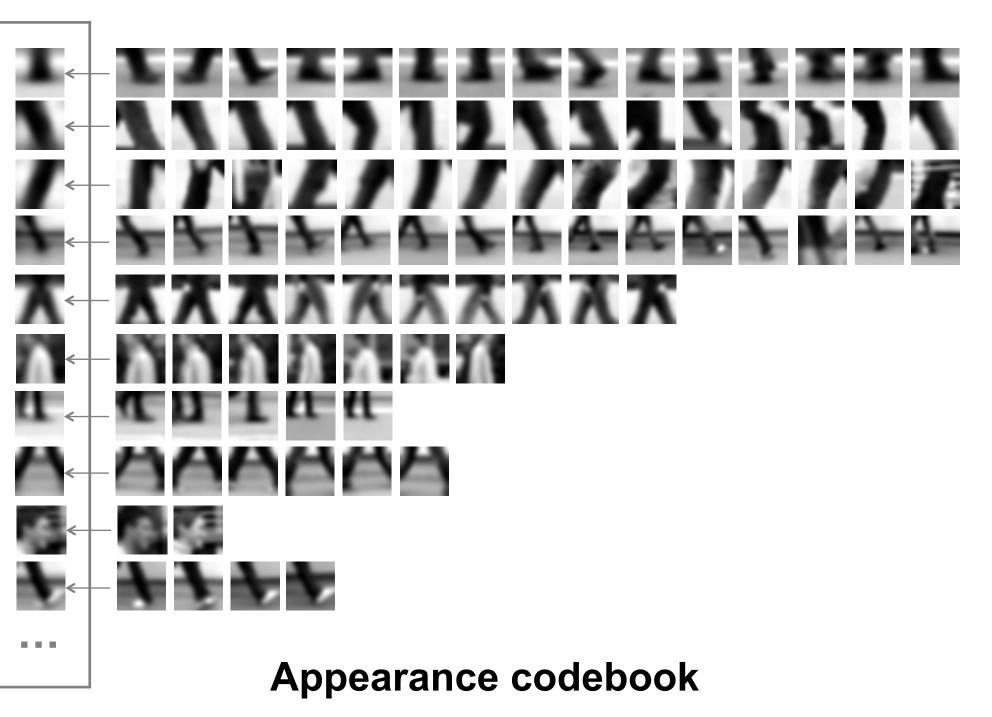
Source: B. Leibe



Another codebook





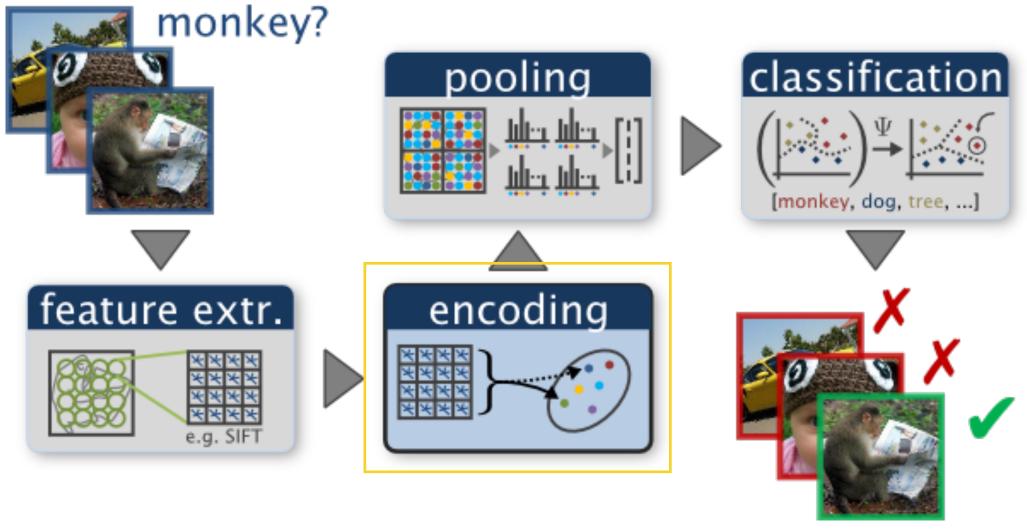


Source: B. Leibe



Origin and motivation of the "bag of words" model Algorithm pipeline

- Extracting local features
- Learning a dictionary clustering using k-means \bullet
- Encoding methods hard vs. soft assignment •
- Spatial pooling pyramid representations
- Similarity functions and classifiers



COMPSCI 370

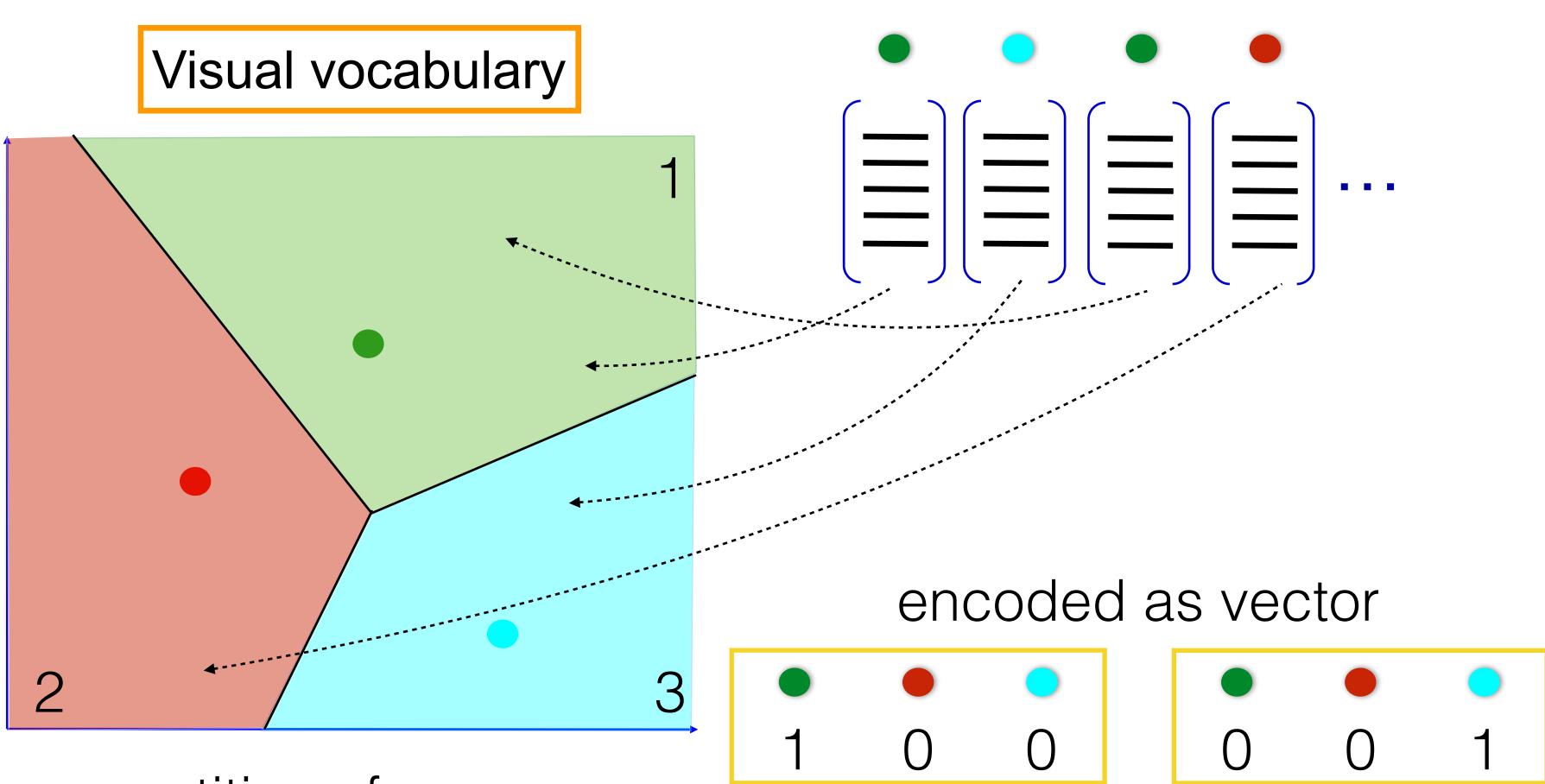
Subhransu Maji – UMass Amherst, Spring 25

Figure from *Chatfield et al.,2011*



Encoding methods

Assigning words to features



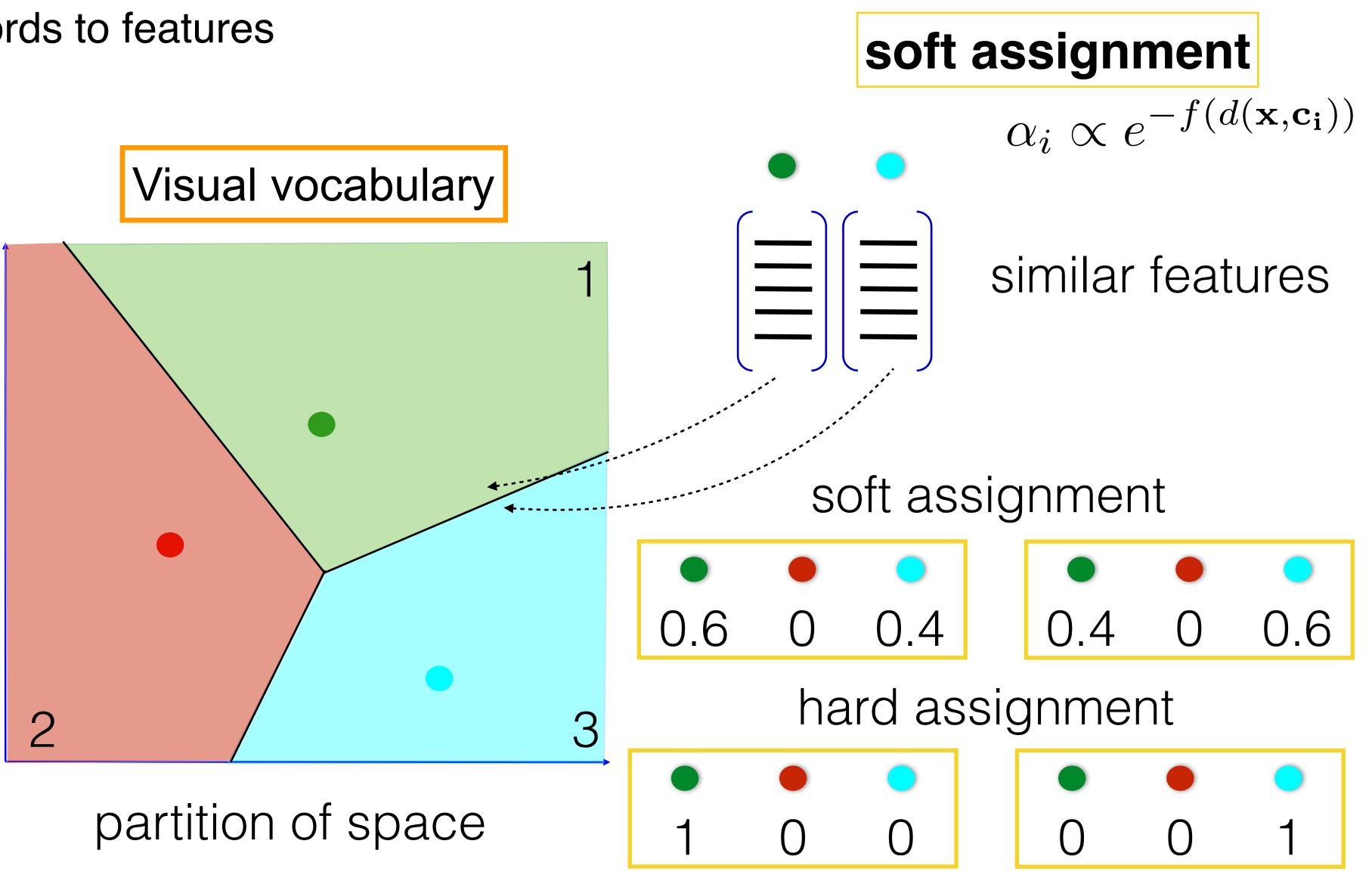
partition of space

COMPSCI 370



Encoding methods

Assigning words to features



COMPSCI 370

Subhransu Maji – UMass Amherst, Spring 25

41

Encoding considerations

What should be the size of the dictionary?

- Too small: don't capture the variability of the dataset
- Too large: have too few points per cluster

Speed of embedding

- Exact nearest neighbor is slow if the dictionary is large
- Approximate nearest neighbor techniques
 - Search trees organize data in a tree
 - Hashing create buckets in the feature space



Origin and motivation of the "bag of words" model Algorithm pipeline

- Extracting local features
- Learning a dictionary clustering using k-means \bullet
- Encoding methods hard vs. soft assignment \bullet
- Spatial pooling pyramid representations •
- Similarity functions and classifiers

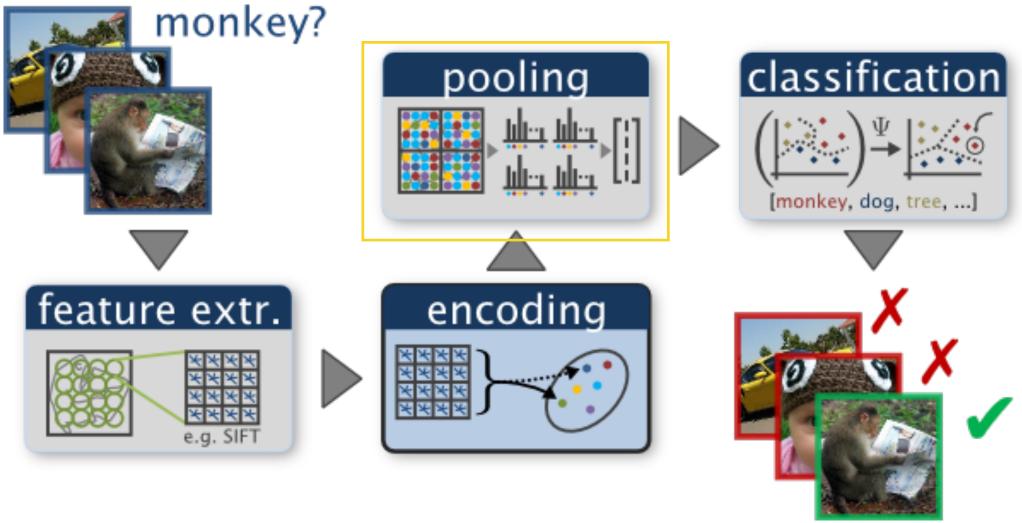


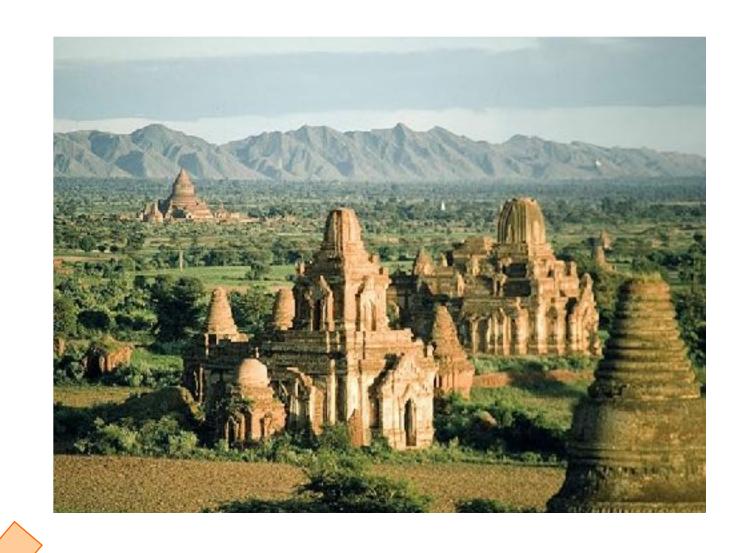
Figure from Chatfialder Mali, 201 Mass Amherst, Spring 25

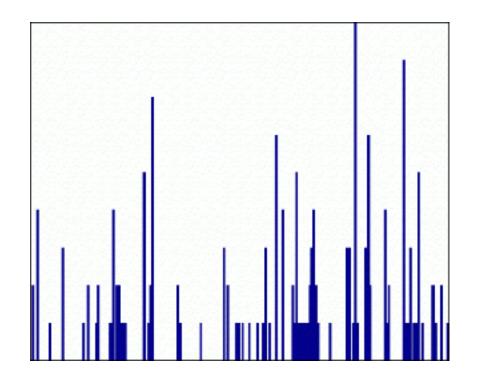
COMPSCI 370



Spatial pyramids

pooling: aggregate features within a region







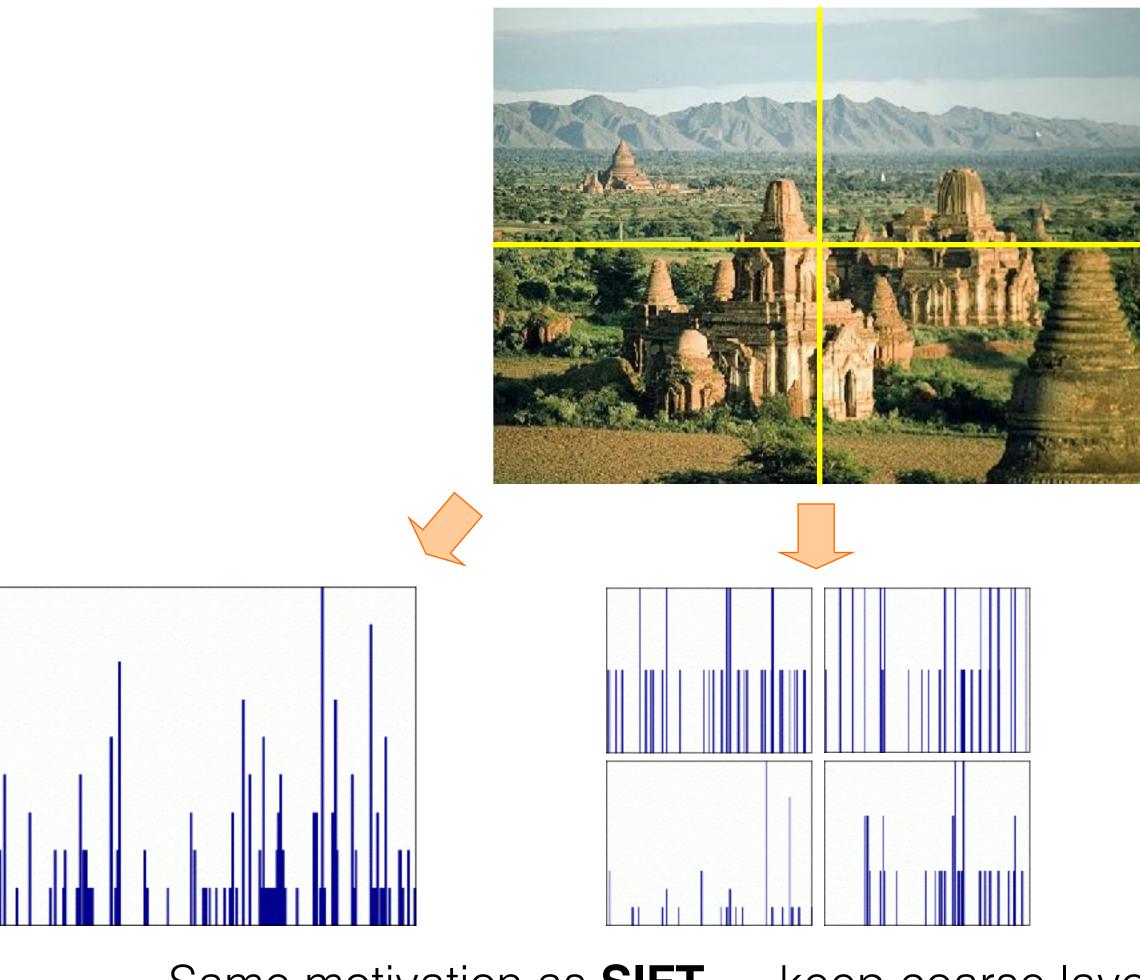
COMPSCI 370

Lazebnik, Schmid & Ponce (CVPR 2006) Subhransu Maji — UMass Amherst, Spring 25



Spatial pyramids

pooling: aggregate features within a region



Lazebnik, Schmid & Ponce (CVPR 2006) Subhransu Maji — UMass Amherst, Spring 25

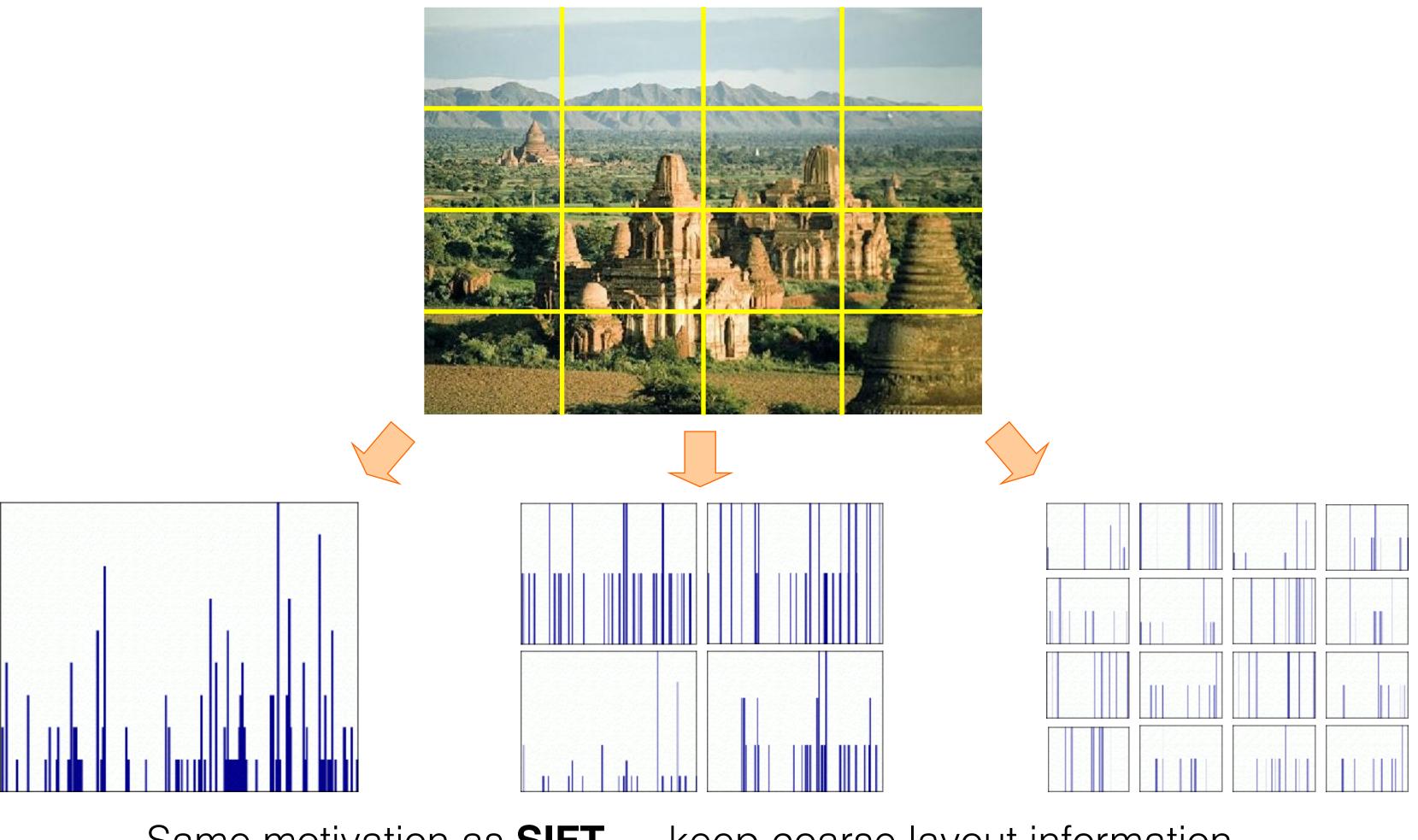
COMPSCI 370

Same motivation as **SIFT** — keep coarse layout information



Spatial pyramids

pooling: aggregate features within a region



Lazebnik, Schmid & Ponce (CVPR 2006) Subhransu Maji — UMass Amherst, Spring 25

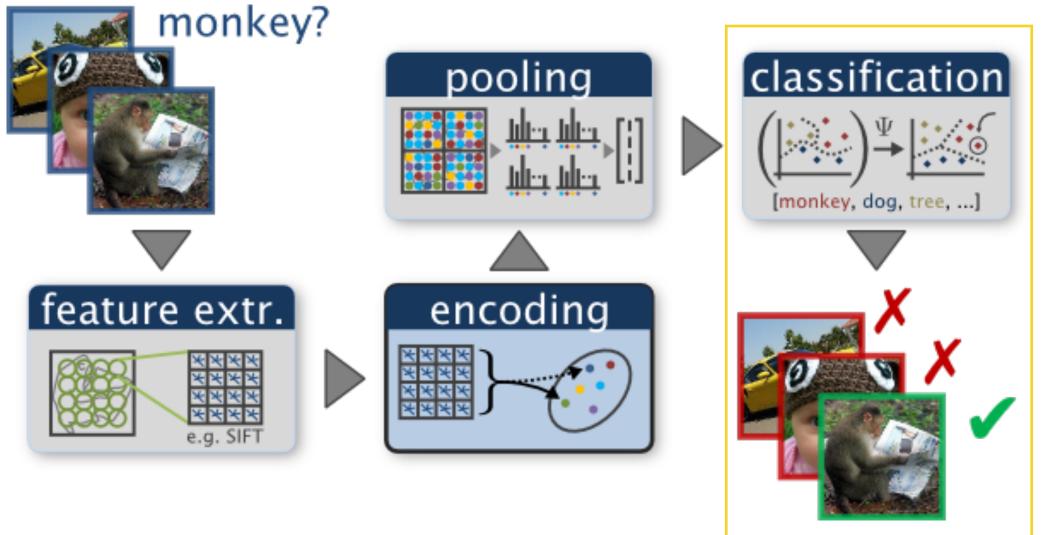
COMPSCI 370

Same motivation as **SIFT** — keep coarse layout information



Origin and motivation of the "bag of words" model Algorithm pipeline

- Extracting local features
- Learning a dictionary clustering using k-means \bullet
- Encoding methods hard vs. soft assignment \bullet
- Spatial pooling pyramid representations
- Similarity functions and classifiers •



COMPSCI 370

Subhransu Maji – UMass Amherst, Spring 25

Figure from *Chatfield et al.,2011*



Bags of features representation

Ι

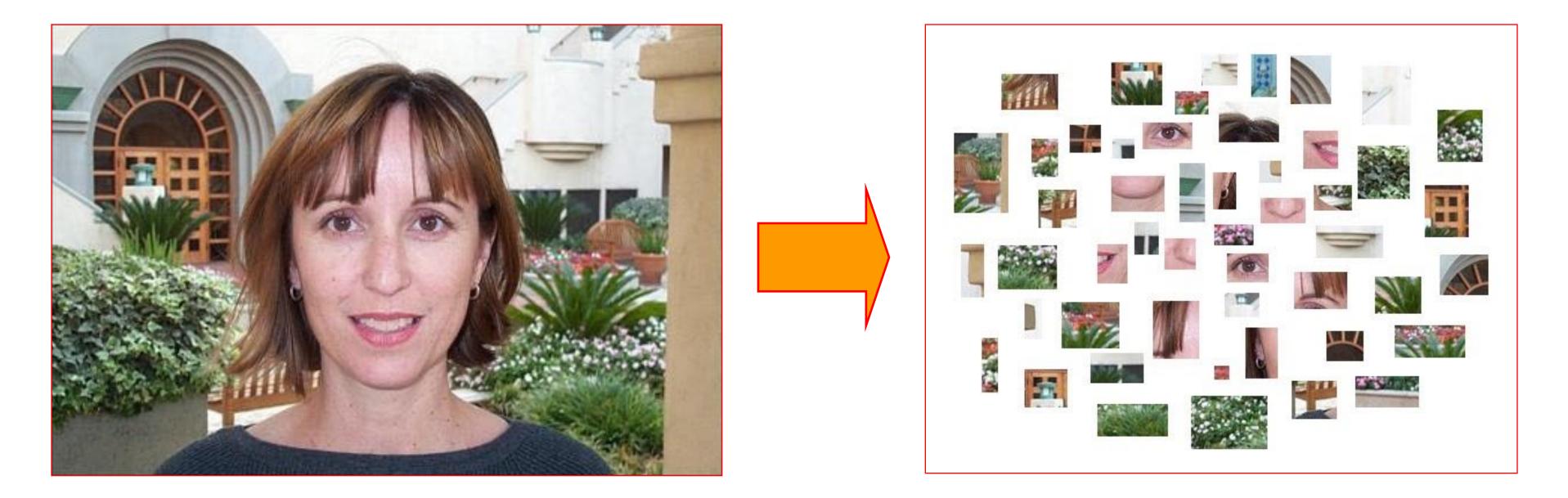


image similarity = feature similarity

COMPSCI 370

 $\mathbf{h} = \Phi(I)$



Similarity functions and classifiers

Euclidean distance:

 $D(\mathbf{h}_1, \mathbf{h}_2) = \mathbf{1}$

L1 distance:

 $D(\mathbf{h}_1, \mathbf{h}_2) =$

Use k-NN classifiers with these distances, or linear classifiers

COMPSCI 370

$$\sum_{i=1}^{N} (\mathbf{h}_1(i) - \mathbf{h}_2(i))^2$$

$$\sum_{i=1}^{N} |\mathbf{h}_1(i) - \mathbf{h}_2(i)|$$



Origin and motivation of the "bag of words" model Algorithm pipeline

- Extracting local features
- Learning a dictionary clustering using k-means \bullet
- Encoding methods hard vs. soft assignment \bullet
- Spatial pooling pyramid representations
- Similarity functions and classifiers

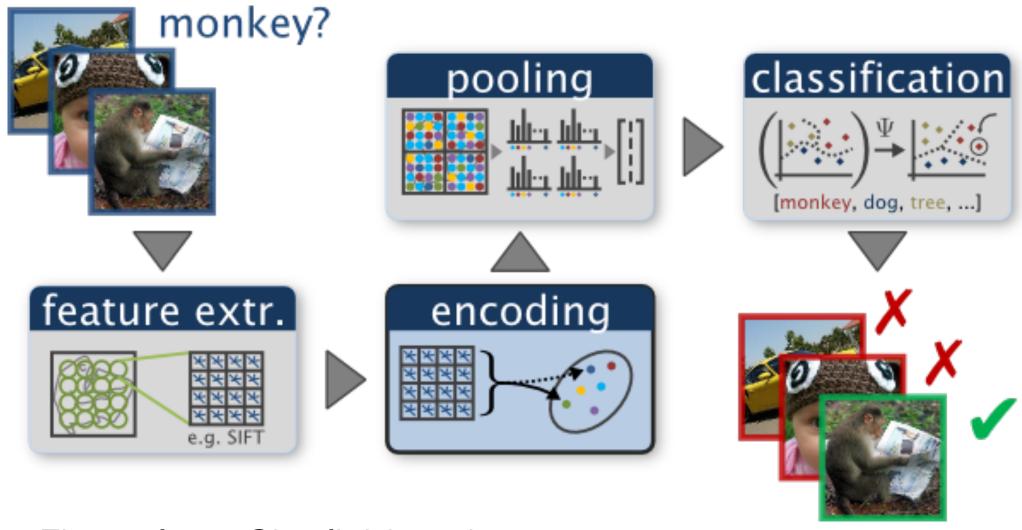


Figure from Chatfialder Mali, 201 Mass Amherst, Spring 25

COMPSCI 370

Putting it all together

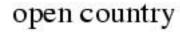


Results: scene category dataset





coast



Multi-class classification results (100 training images per class)

	Weak features		Strong features	
	(vocabulary size: 16)		(vocabulary size: 200)	
Level	Single-level	Pyramid	Single-level	Pyramid
$0(1 \times 1)$	45.3 ± 0.5		72.2 ± 0.6	
$1(2 \times 2)$	53.6 ± 0.3	56.2 ± 0.6	77.9 ± 0.6	79.0 ± 0.5
$2(4 \times 4)$	61.7 ± 0.6	64.7 ± 0.7	79.4 ± 0.3	81.1 ±0.3
3 (8 × 8)	63.3 ± 0.8	66.8 ±0.6	77.2 ± 0.4	80.7 ± 0.3

COMPSCI 370



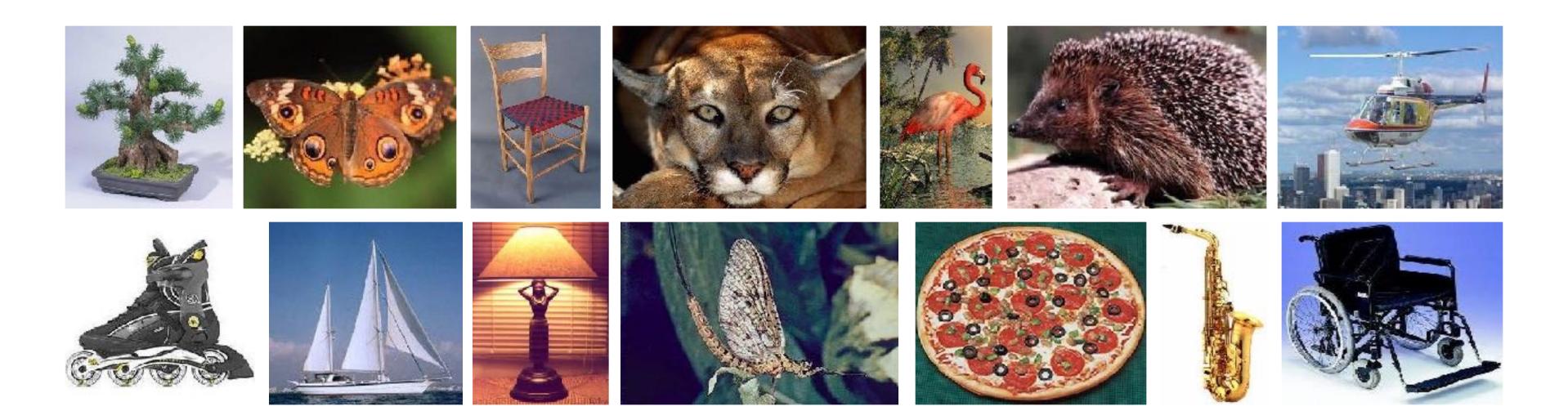
mountain

forest

suburb



Results: Caltech-101 dataset



Multi-class classification results (30 training images per class)

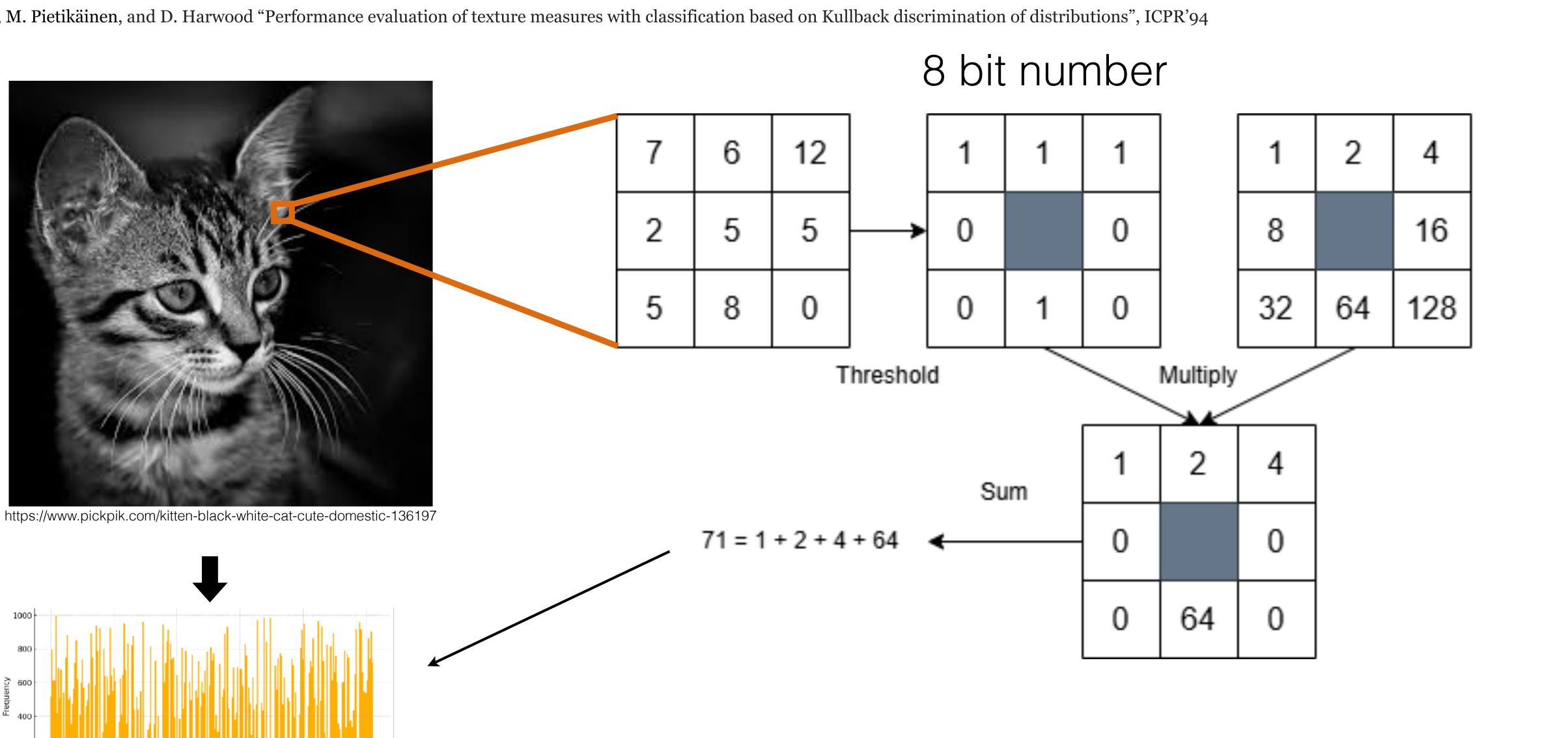
	Weak features (16)		Strong features (200)	
Level	Single-level	Pyramid	Single-level	Pyramid
0	15.5 ± 0.9		41.2 ± 1.2	
1	31.4 ± 1.2	32.8 ± 1.3	55.9 ± 0.9	57.0 ± 0.8
2	47.2 ± 1.1	49.3 ± 1.4	63.6 ± 0.9	64.6 ±0.8
3	52.2 ± 0.8	$\textbf{54.0} \pm 1.1$	60.3 ± 0.9	64.6 ± 0.7

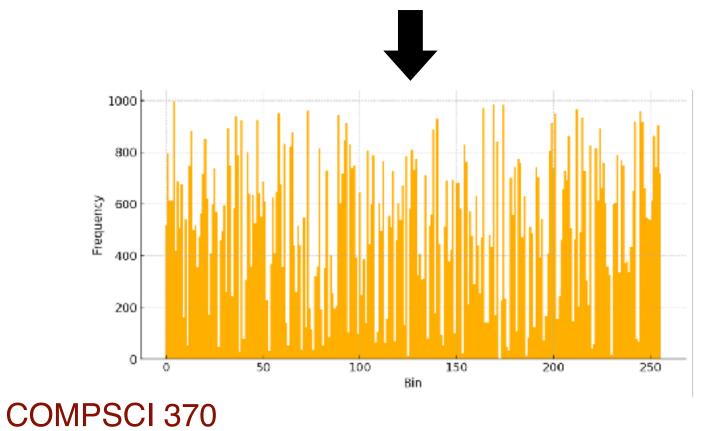
COMPSCI 370



Local binary pattern — homework

T. Ojala, M. Pietikäinen, and D. Harwood "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", ICPR'94







Further thoughts and readings ...

- All about embeddings (detailed experiments and code)
 - BMVC 2011
 - <u>http://www.robots.ox.ac.uk/~vgg/research/encoding_eval/</u>
 - linear coding (LLC)

• K. Chatfield et al., The devil is in the details: an evaluation of recent feature encoding methods,

• Includes discussion of advanced embeddings such as Fisher vector representations and locally



