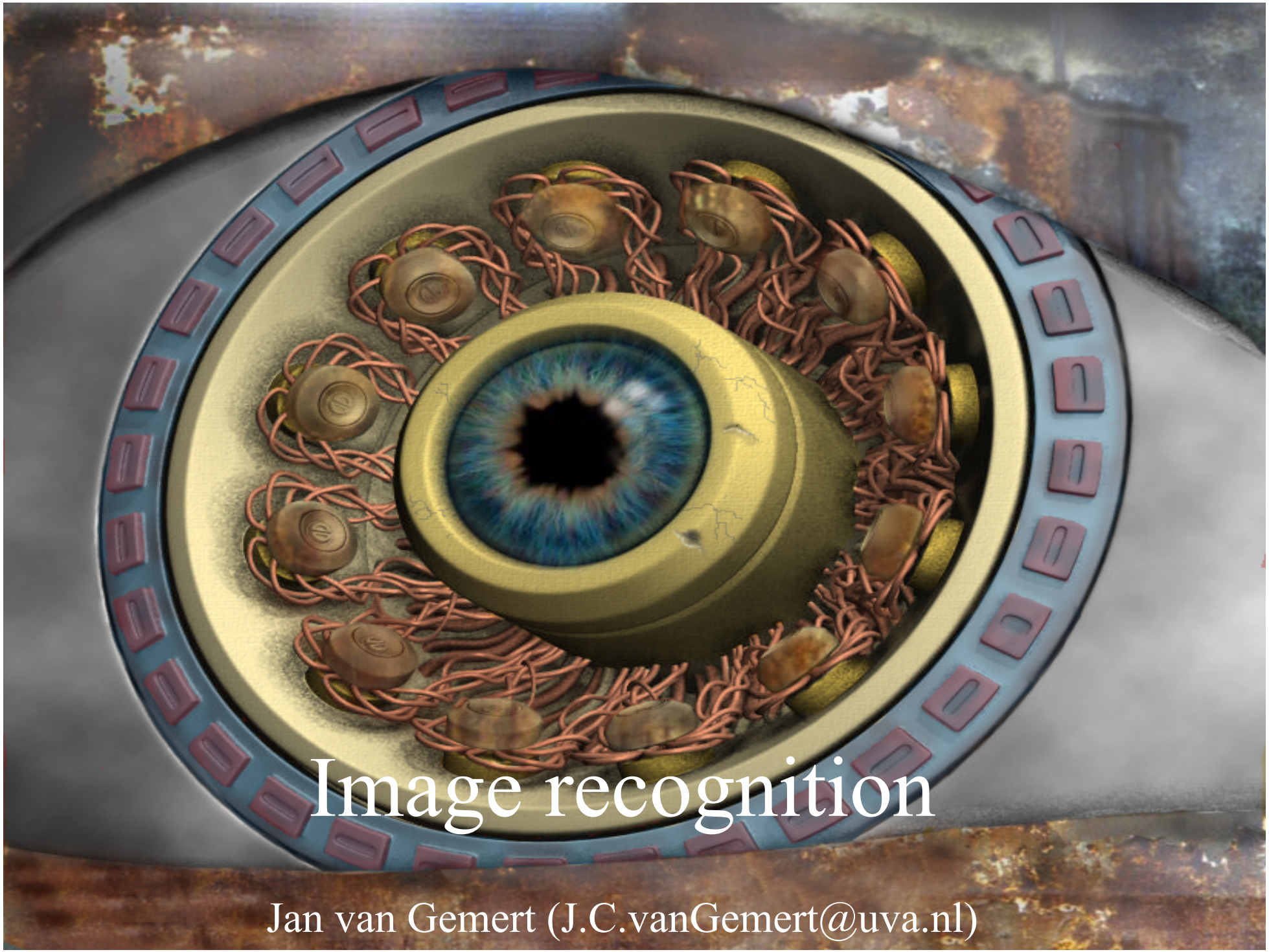


Image recognition

Jan van Gemert (J.C.vanGemert@uva.nl)

Make a list of the types of recognition

Instance recognition: Dam Monument

Detection: are there people

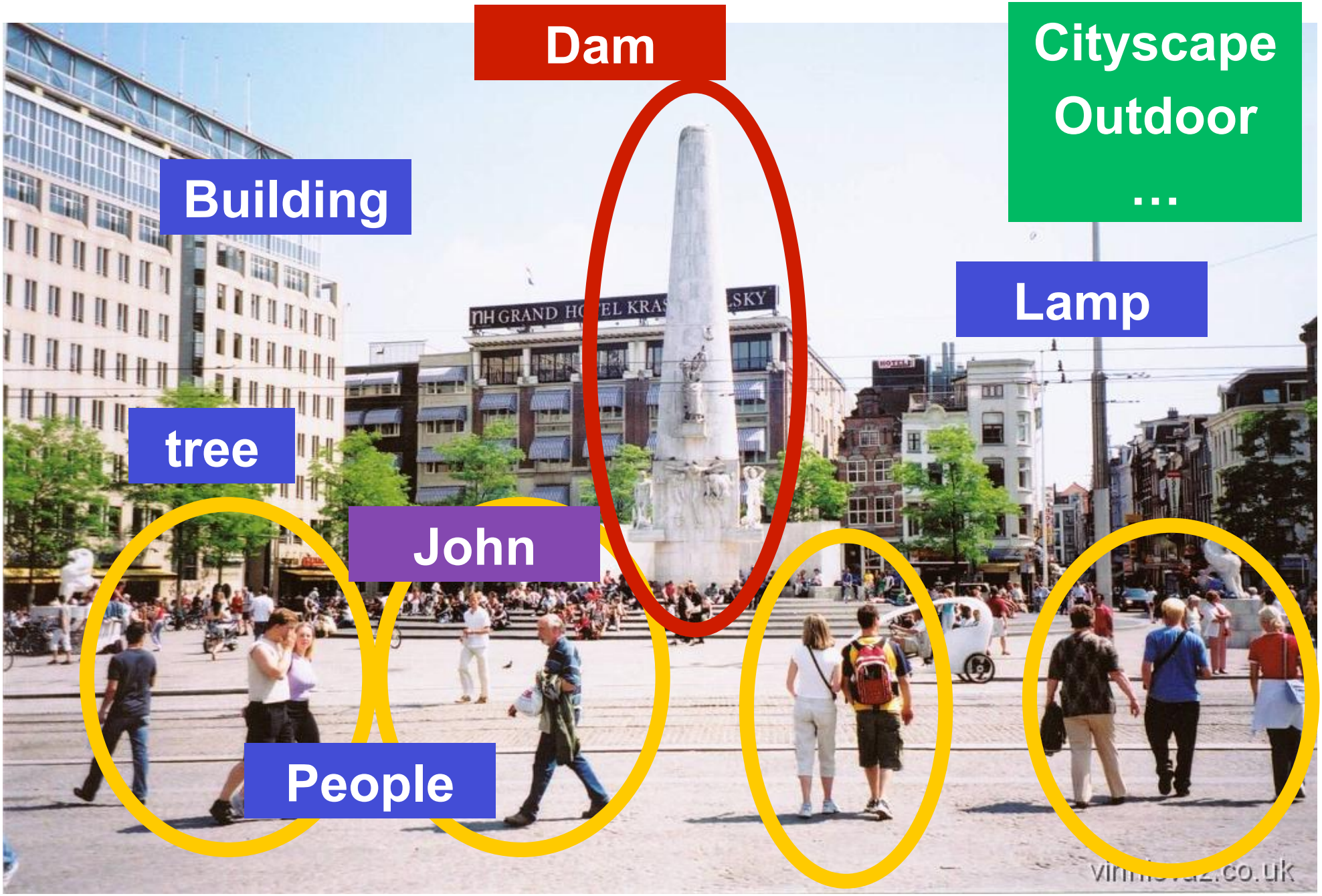
Verification: is that John?

Object categories

Scene and Context

Cityscape
Outdoor
…

What distinguishes types? Challenges?

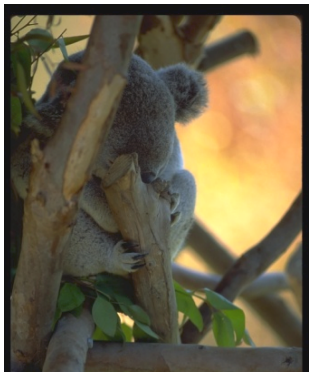


Challenges: robustness

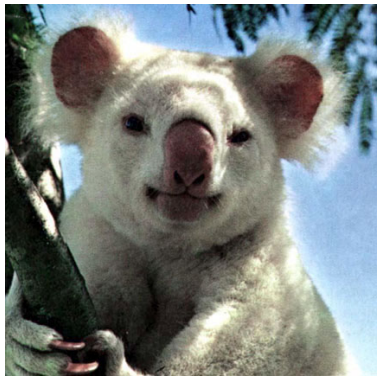
Illumination

Object pose

Clutter



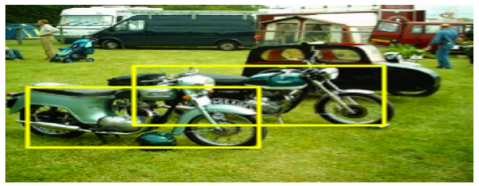
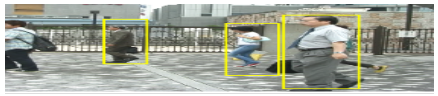
Occlusions



Intra-class appearance

Viewpoint

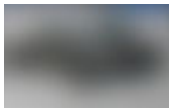
Challenges: robustness

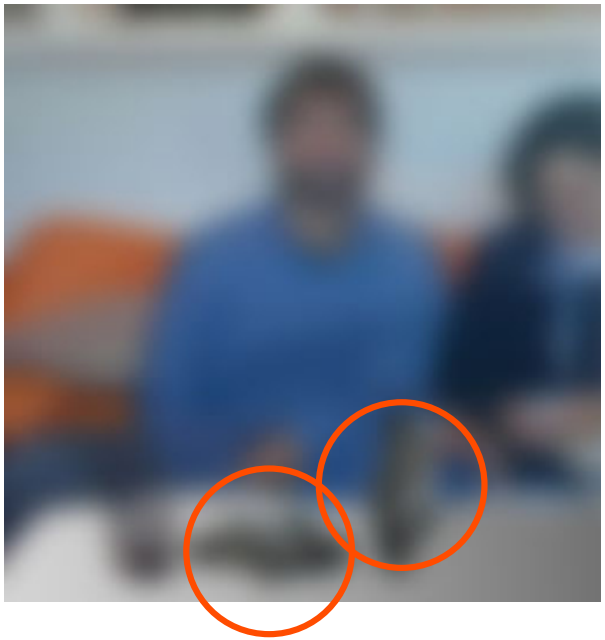



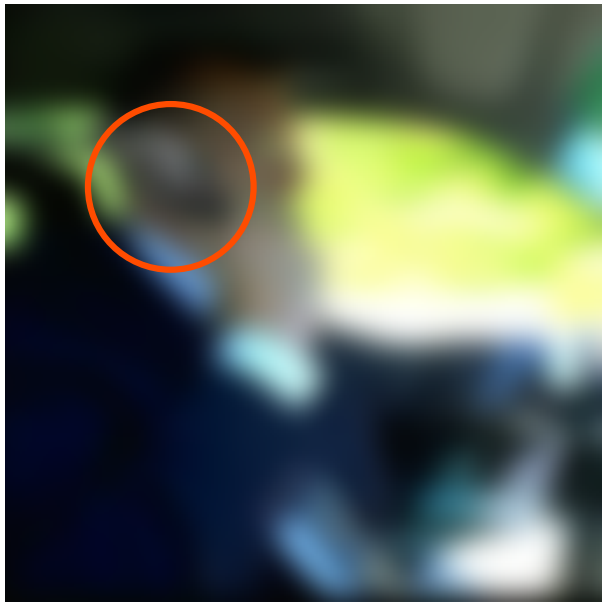


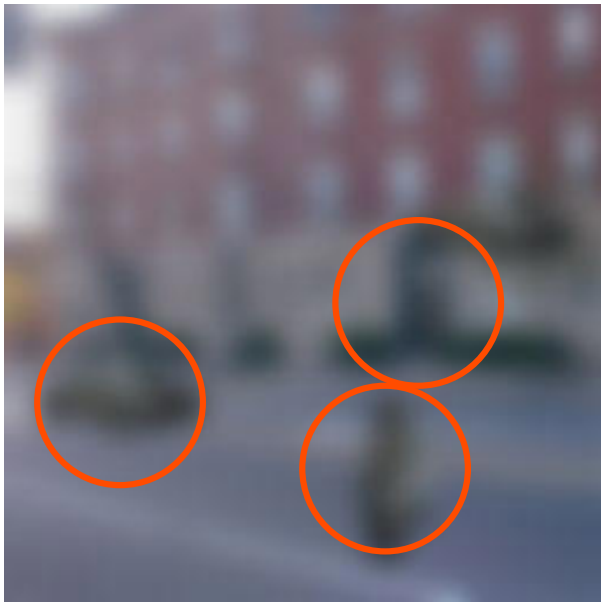
Realistic scenes are crowded, cluttered, have overlapping objects.

Challenges: importance of context









slide credit: Fei-Fei, Fergus & Torralba

Bag-of-features models (Visual words)

Bag-of-words model in text

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

Bag-of-words model in text

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

Bag-of-words model in text

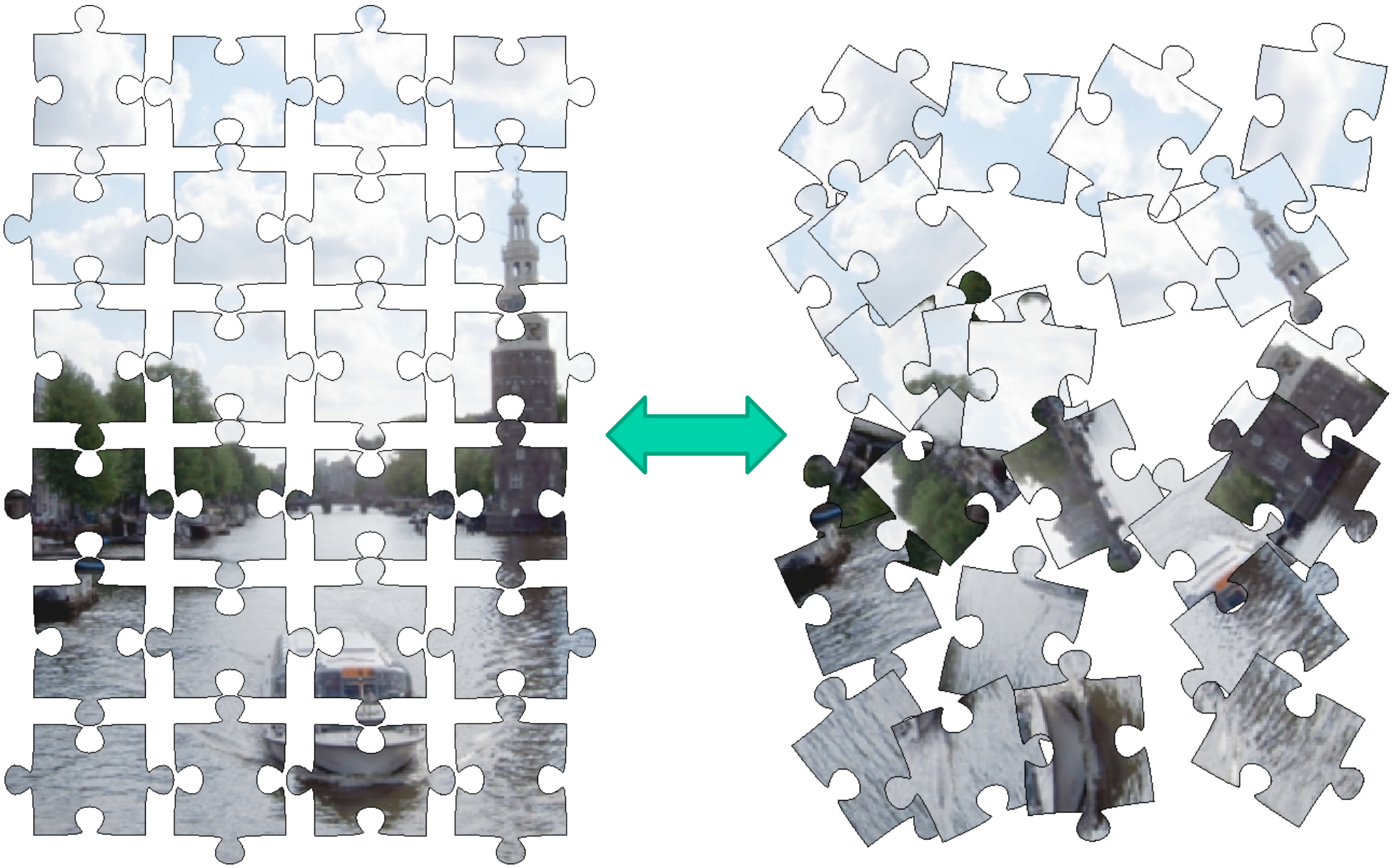
- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

Bag-of-words model in text

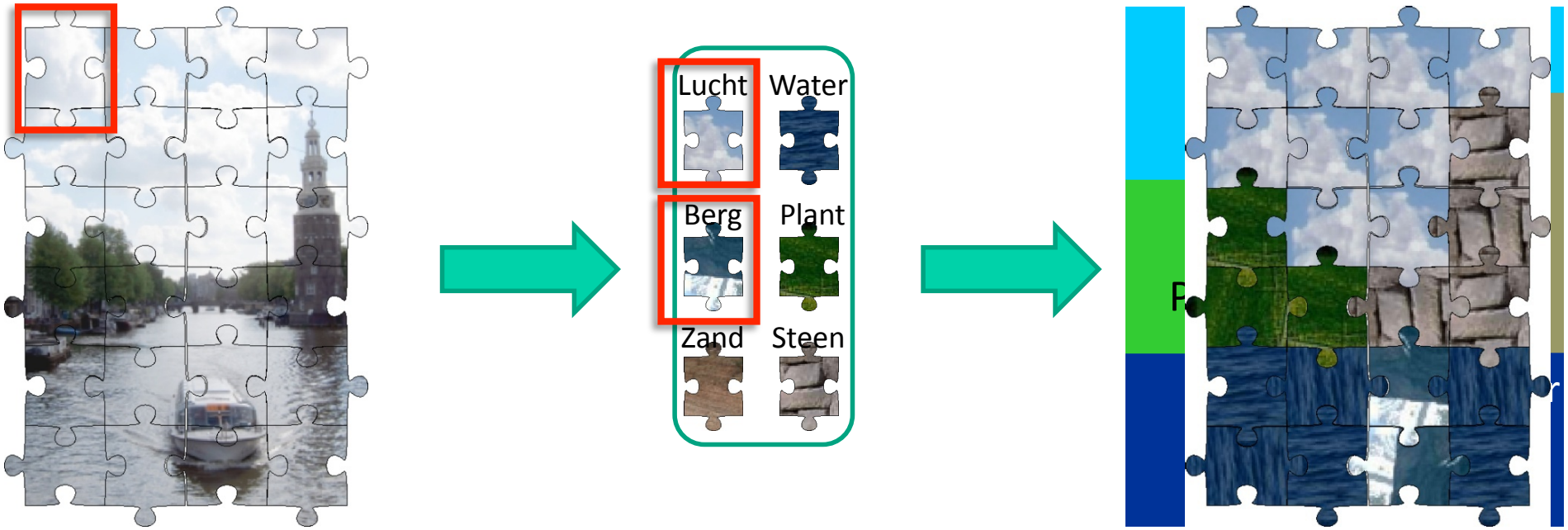
- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

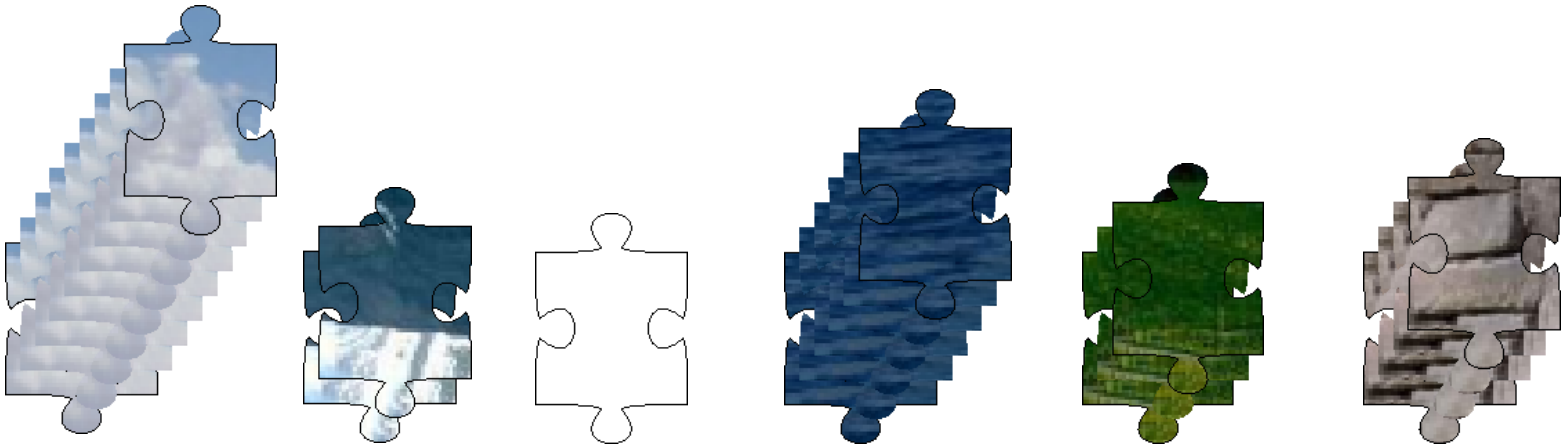
How would you apply this to images?

Lokale Beeldkenmerken



Visueel woordenboek

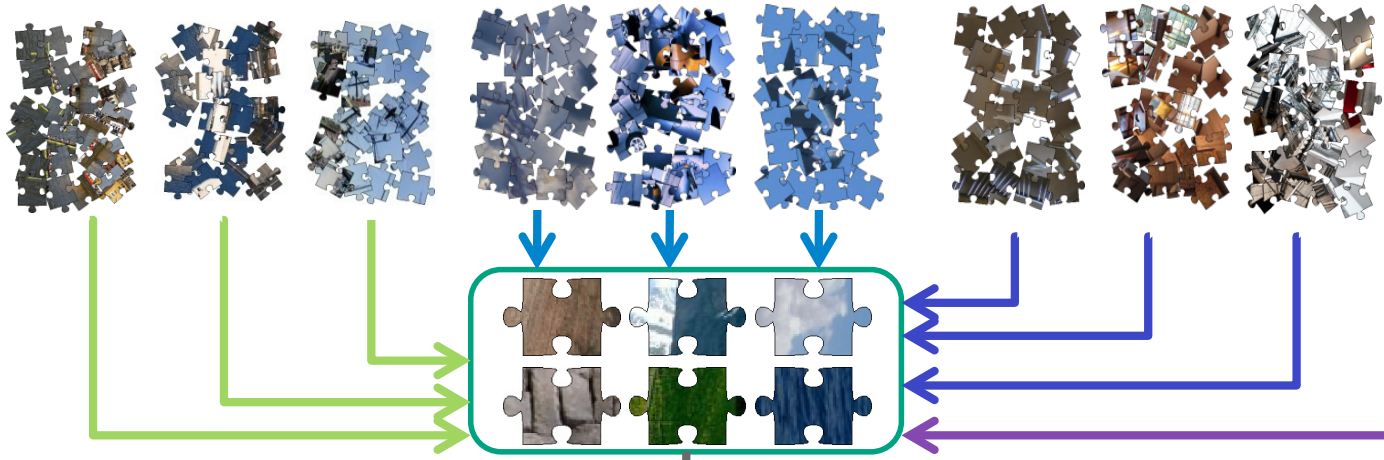


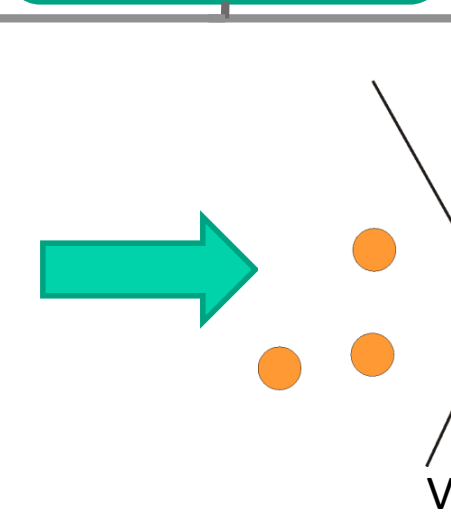


Lucht	Berg	Zand	Water	Plant	Steen
9	2	0	6	3	4

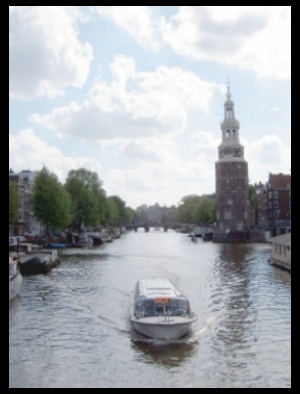
Hele beeld gerepresenteerd als vector van 6 getallen

Leren





Classificeren



Test matrix

Example of Visual Words

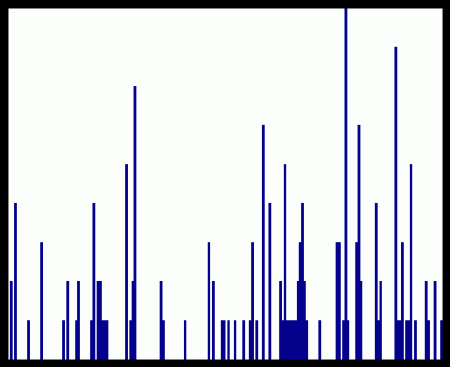
Questions?

- Advantages?
  - Translation invariant
  - Robust matching of visual word instances
  - Fixed length feature vector
  - Good performance

- Disadvantages?
  - Loss of spatial position and object configuration
  - Quantization loses information

Spatial pyramid representation

- Extension of a bag of features
- Locally orderless representation at several levels of resolution



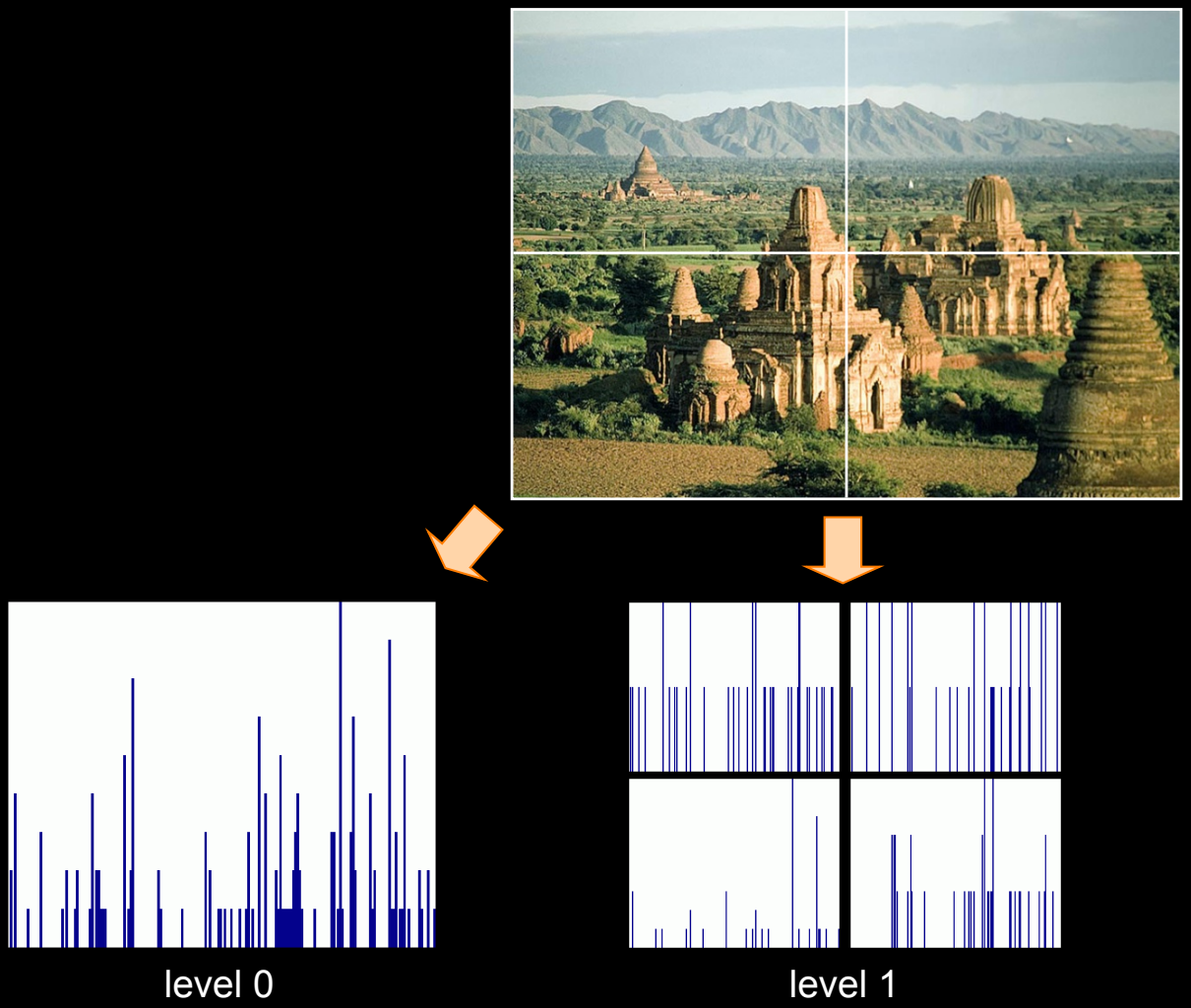
level 0

Lazebnik, Schmid & Ponce (CVPR 2006)

Slide: S. Lazebnik

Spatial pyramid representation

- Extension of a bag of features
- Locally orderless representation at several levels of resolution

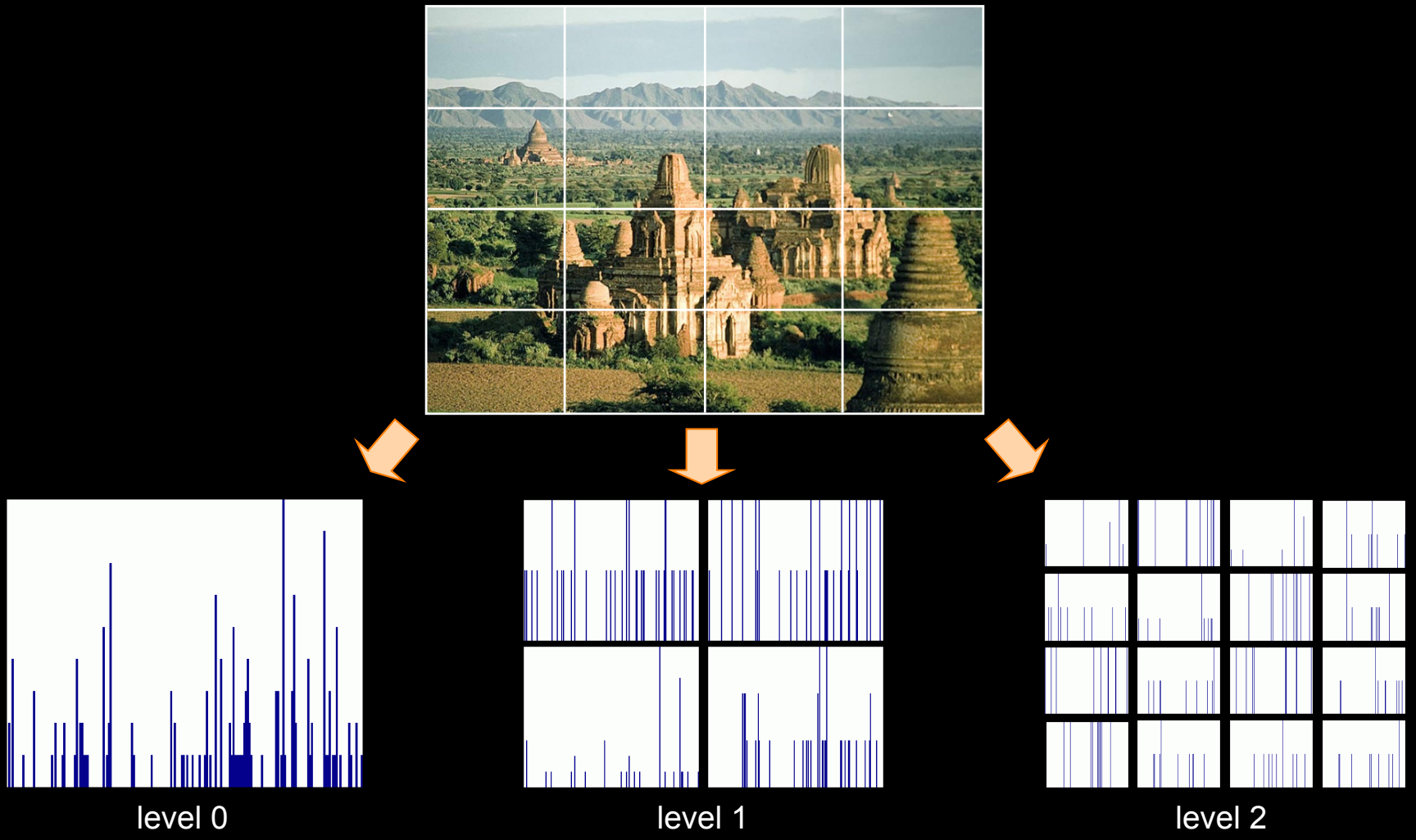


Lazebnik, Schmid & Ponce (CVPR 2006)

Slide: S. Lazebnik

Spatial pyramid representation

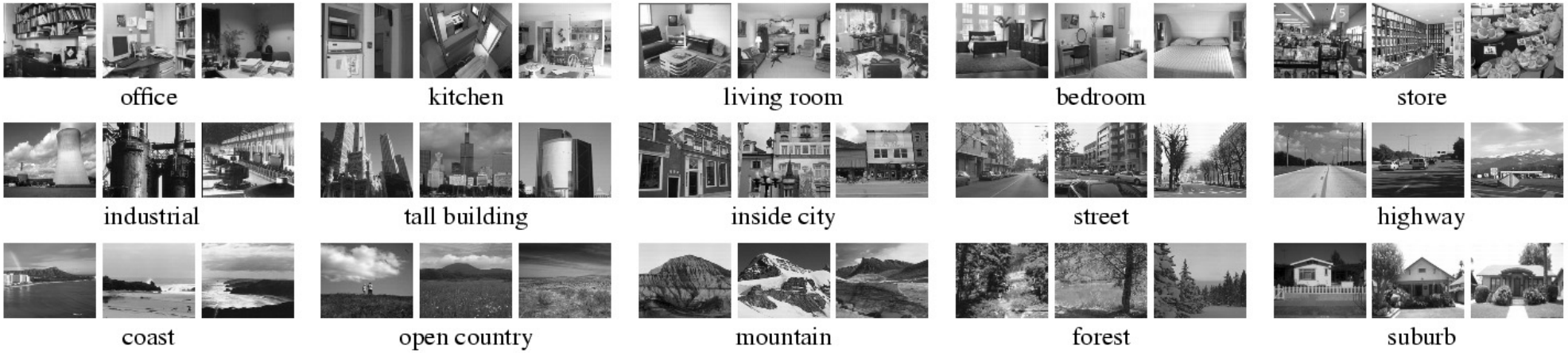
- Extension of a bag of features
- Locally orderless representation at several levels of resolution



Lazebnik, Schmid & Ponce (CVPR 2006)

Slide: S. Lazebnik

Scene category dataset



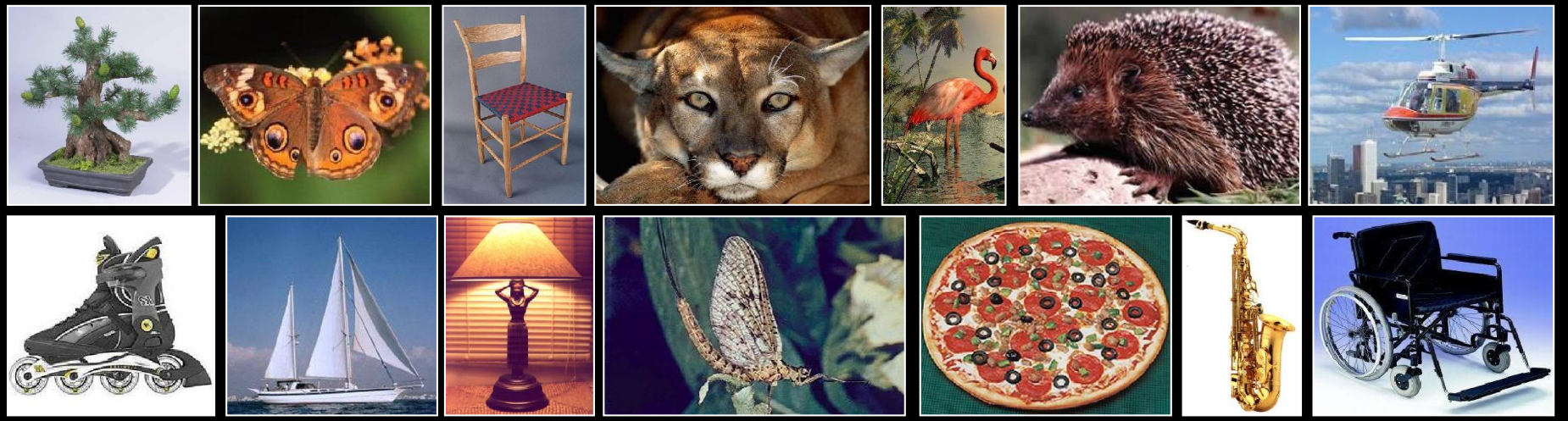
Multi-class classification results
(100 training images per class)

	Weak features (vocabulary size: 16)		Strong features (vocabulary size: 200)	
Level	Single-level	Pyramid	Single-level	Pyramid
0 (1×1)	45.3 ±0.5		72.2 ±0.6	
1 (2×2)	53.6 ±0.3	56.2 ±0.6	77.9 ±0.6	79.0 ±0.5
2 (4×4)	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

Slide: S. Lazebnik

Caltech101 dataset

http://www.vision.caltech.edu/Image_Datasets/Caltech101/Caltech101.html



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

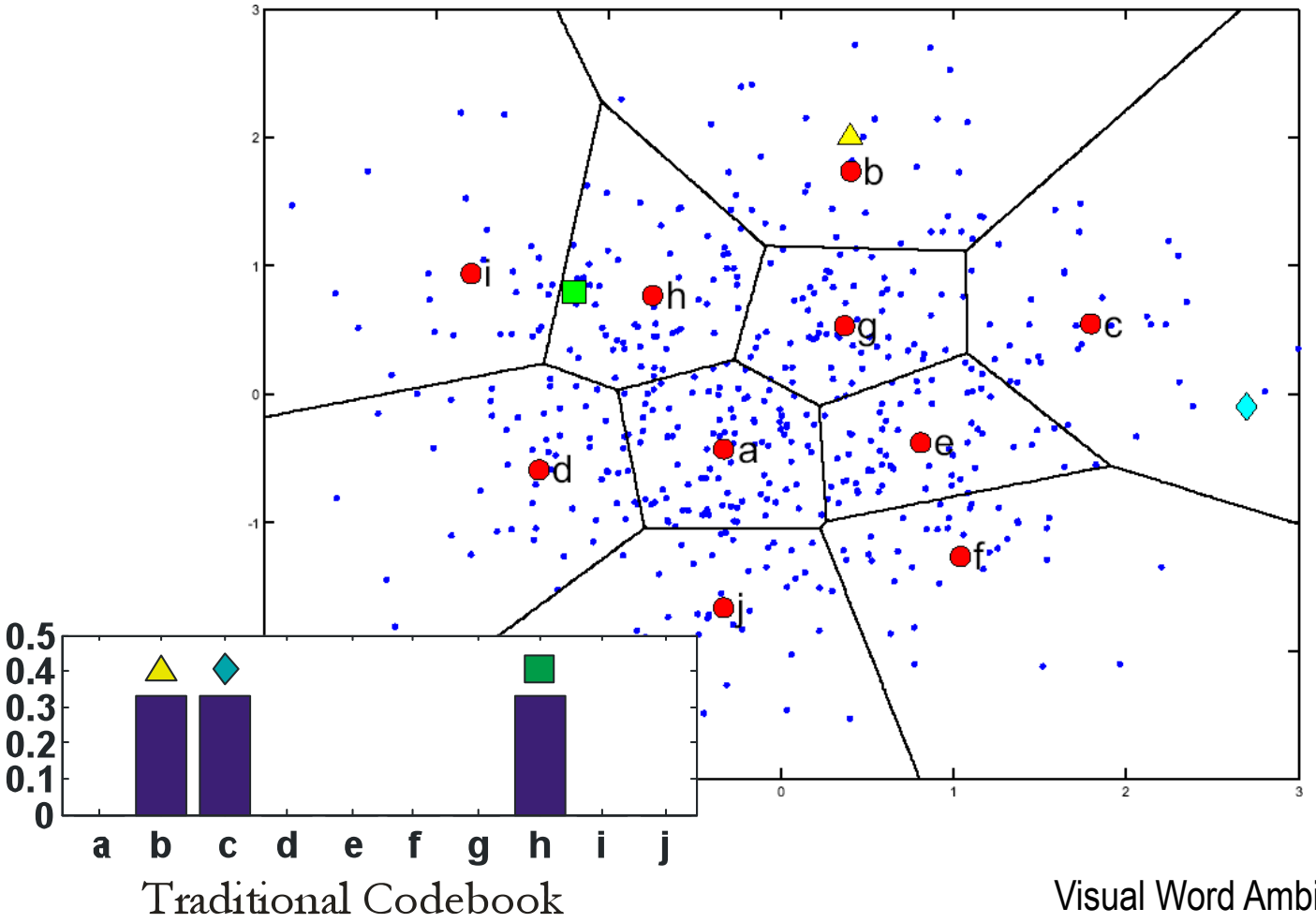
	Weak features (16)		Strong features (200)	
Level	Single-level	Pyramid	Single-level	Pyramid
0	15.5 ±0.9		41.2 ±1.2	
1	31.4 ±1.2	32.8 ±1.3	55.9 ±0.9	57.0 ±0.8
2	47.2 ±1.1	49.3 ±1.4	63.6 ±0.9	**64.6** ±0.8
3	52.2 ±0.8	**54.0** ±1.1	60.3 ±0.9	64.6 ±0.7

Slide: S. Lazebnik

Questions?

Drawback to the Codebook Model

- Hard assignment of features to codewords
  - Codeword Uncertainty / Codeword Plausibility



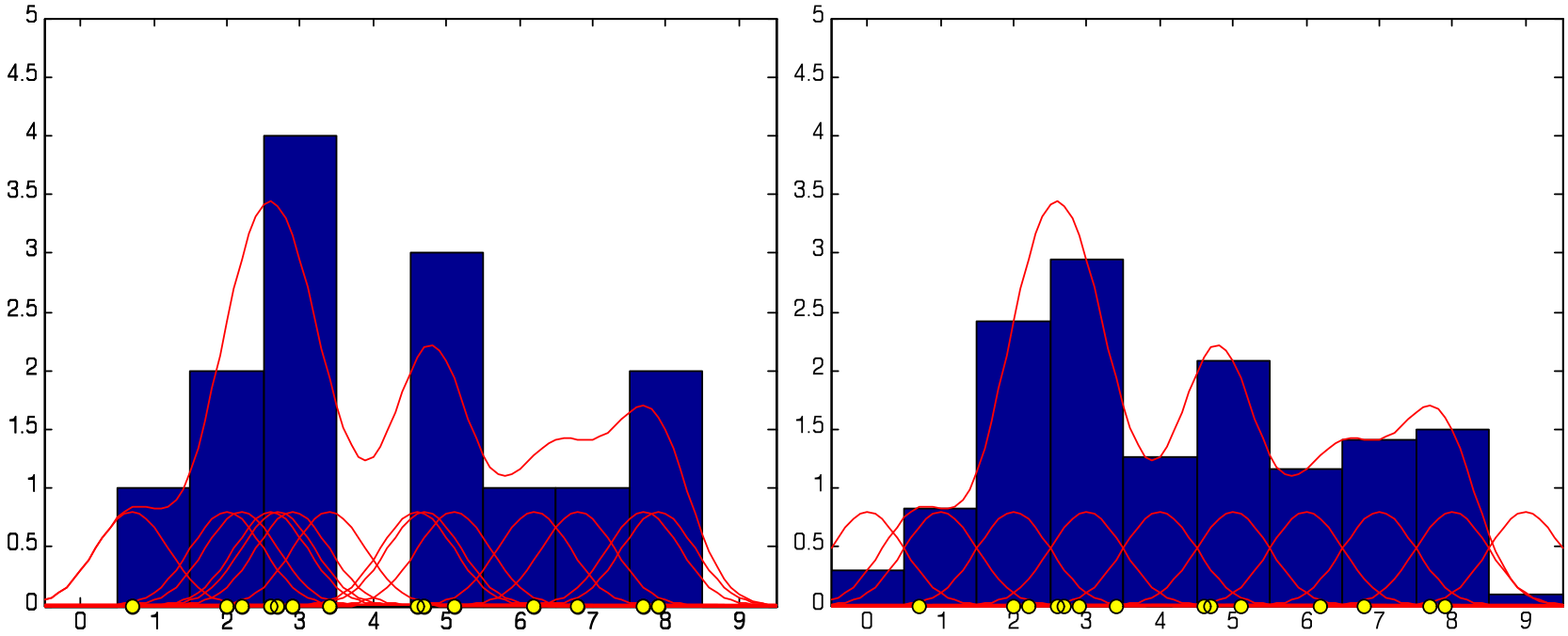
Traditional Codebook

Visual Word Ambiguity, TPAMI 2010

Silverman and Green. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

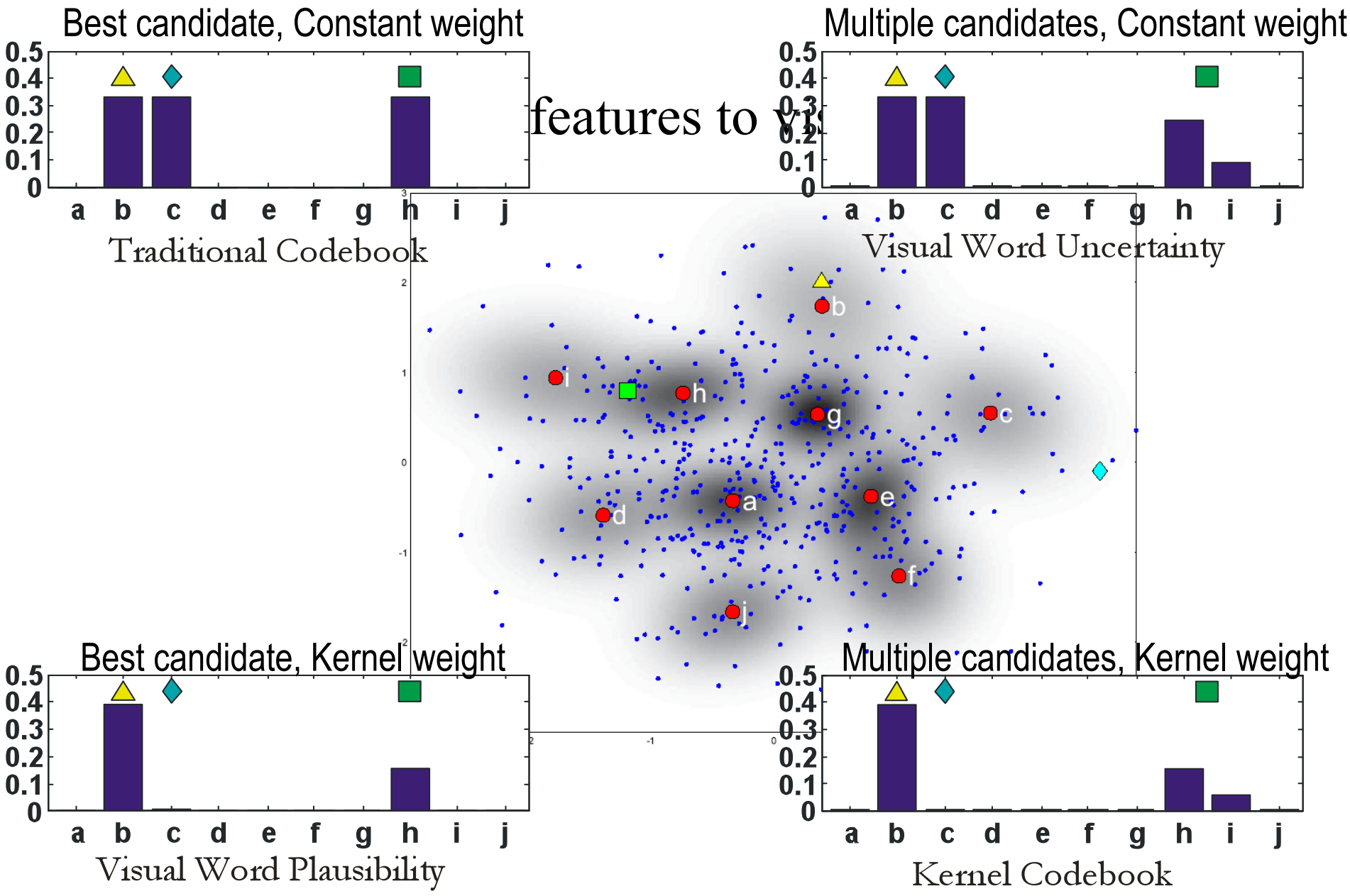
Kernel Codebooks

- Codebook model is a histogram
  - Problem: Hard assignment to histogram bins
  - a solution: Kernel density estimation

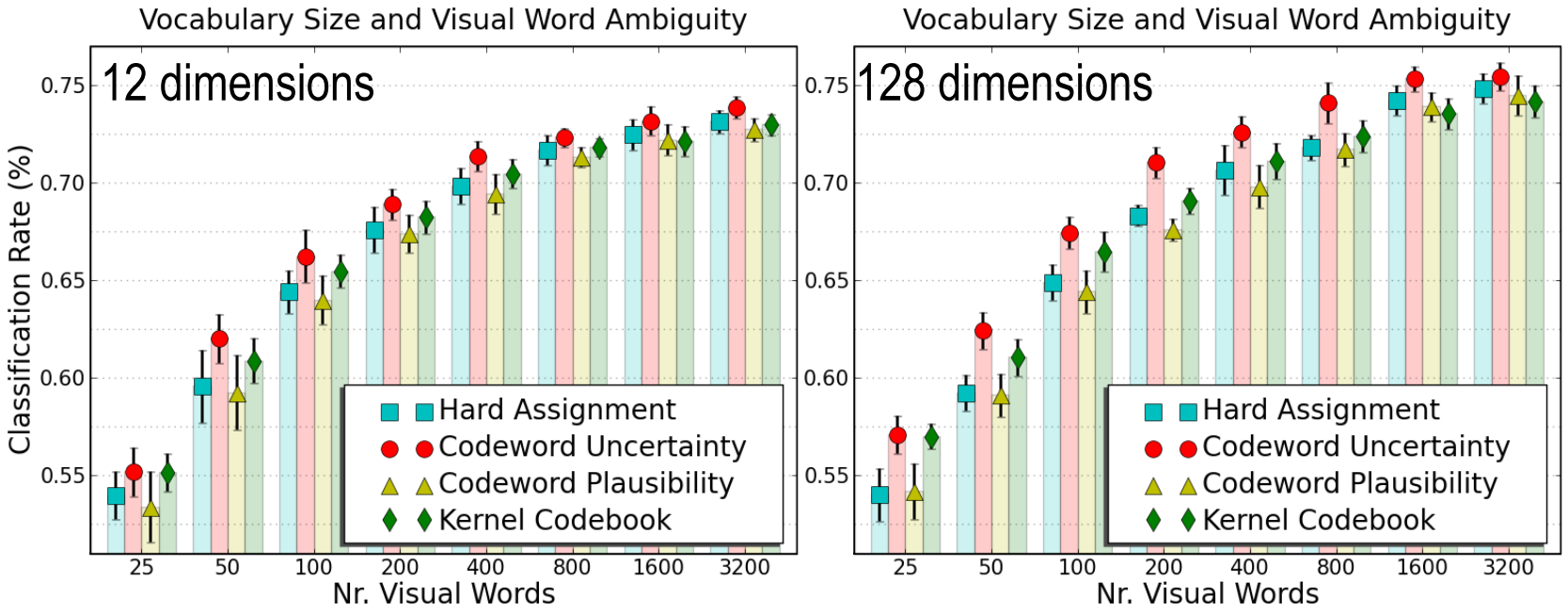


  - Symmetric Kernel allows kernel codewords

Ambiguity Modeling

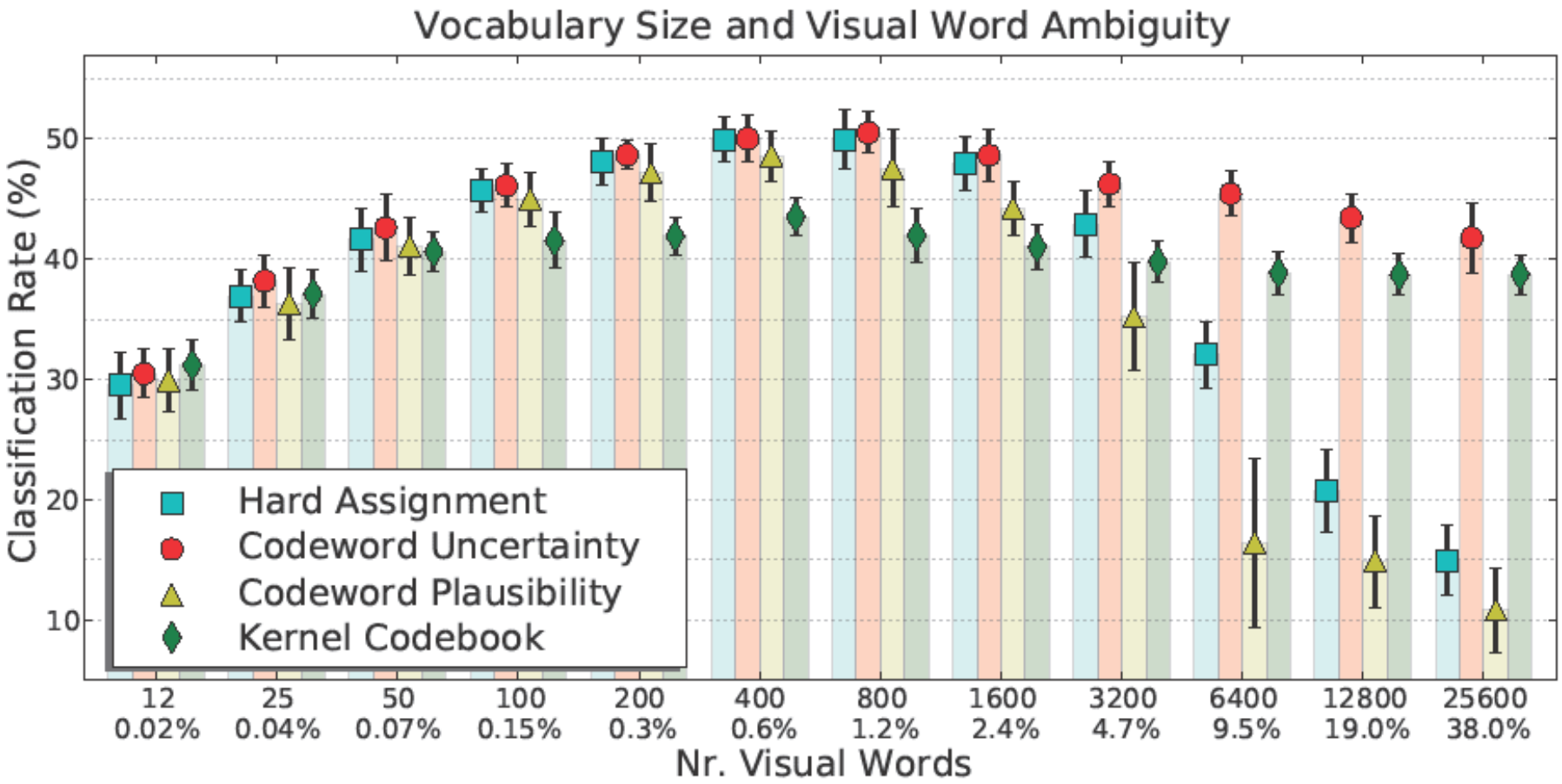


Results Scene-15 Dataset



- Ambiguity has more effect in high-dimensions
- Codeword Plausibility always performs worst
- Codeword Uncertainty always performs best
- Ambiguity has more effect for small vocabularies
- Performance increases asymptotically with vocabulary size

Results Scene-15 Dataset



- Hard assignment completely fails for too large vocabularies
- Soft assignment stays relatively constant

Questions?

Summary

- Types of Recognition, challenges
- Bag of Features, Visual-words
- Spatial-Pyramid
- Soft-assignment