# Image recognition

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

#### <sup>2</sup> Make a list of the types of recognition



# <sup>3</sup> Instance recognition: Dam Monument



# <sup>4</sup> Detection: are there people



# <sup>5</sup> Verification: is that John?







# <sup>7</sup> Scene and Context

#### Cityscape Outdoor



#### <sup>8</sup> 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)





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

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

#### 2007-01-23: State of the Union Address George W. Bush (2001-) 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 insurgents iran iran 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 september shia stays strength students succeed sunni tax territories territories threats uphold victory violence violent WaT washington weapons wesley

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



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

2007-01-23: State of the Union Address George W. Bush (2001-)						
abandon choices c deficit c	1962-10-22: Soviet Missiles in Cuba John F. Kennedy (1961-					
expand	aban do	1941 12 08: Request for a Declaration of War				
insurgen	buildu	Franklin D. Roosevelt (1933-45)				
palestinia	declined	abandoning acknowledge aggression aggressors airplanes armaments <b>armed army</b> assault assembly authorizations bombing britain british cheerfully claiming constitution curtail december defeats defending delays <b>democratic dictators</b> disclose				
septemb	halt ha	economic empire endanger facts false forgotten fortunes france freedom fulfilled fullness fundamental gangsters				
violenc	modern	german germany god guam harbor hawaii hemisphere hint hitler hostilities immune improving indies innumerable				
	recessio	invasion islands isolate Japanese labor metals midst midway navy nazis obligation offensive				
	surveill	officially <b>Pacific</b> partisanship patriotism pearl peril perpetrated perpetual philippine preservation privilege reject repaired <b>resisting</b> retain revealing rumors seas soldiers speaks speedy stamina strength sunday sunk supremacy tanks taxes				
		treachery true tyranny undertaken victory War wartime washington				

#### How would you apply this to images?





### Lokale Beeldkenmerken



# Visueel woordenboek







### 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

- ullet
- Locally orderless representation at several levels of resolution  $\bullet$





Lazebnik, Schmid & Ponce (CVPR 2006)



# Spatial pyramid representation Extension of a bag of features

- ullet
- Locally orderless representation at several levels of resolution ullet



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

ullet



#### Scene category dataset



coast

open country

mountain

forest

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

	Weak fe	eatures	Strong features	
	(vocabulary	/ size: 16)	(vocabulary	size: 200)
Level	Single-level	Pyramid	Single-level	Pyramid
$0(1 \times 1)$	$45.3 \pm 0.5$		$72.2 \pm 0.6$	
$1(2 \times 2)$	$53.6 \pm 0.3$	$56.2 \pm 0.6$	$77.9 \pm 0.6$	$79.0\pm0.5$
$2(4 \times 4)$	$61.7 \pm 0.6$	$64.7 \pm 0.7$	$79.4 \pm 0.3$	<b>81.1</b> ±0.3
$3(8 \times 8)$	$63.3 \pm 0.8$	<b>66.8</b> ±0.6	$77.2 \pm 0.4$	$80.7 \pm 0.3$

Slide: S. Lazebnik

suburb

#### Caltech101 dataset

http://www.vision.caltech.edu/Image\_Datasets/Caltech101/Caltech101.html



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

	Weak feat	ures (16)	Strong feat	ures (200)
Level	Single-level	Pyramid	Single-level	Pyramid
0	$15.5 \pm 0.9$		$41.2 \pm 1.2$	
1	$31.4 \pm 1.2$	$32.8 \pm 1.3$	$55.9\pm0.9$	$57.0\pm0.8$
2	$47.2 \pm 1.1$	$49.3 \pm 1.4$	$63.6 \pm 0.9$	<b>64.6</b> $\pm 0.8$
3	52.2 $\pm 0.8$	$54.0 \pm 1.1$	$60.3 \pm 0.9$	$64.6\pm\!0.7$

Slide: S. Lazebnik

### Questions?

#### Drawback to the Codebook Model

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



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