## **Bag-of-features models**





Many slides adapted from Fei-Fei Li, Rob Fergus, and Antonio Torralba

## **Overview: Bag-of-features models**

- Origins and motivation
- Image representation
  - Feature extraction
  - Visual vocabularies
- Discriminative methods
  - Nearest-neighbor classification
  - Distance functions
  - Support vector machines
  - Kernels
- Generative methods
  - Naïve Bayes
  - Probabilistic Latent Semantic Analysis
- Extensions: incorporating spatial information

## Origin 1: Texture recognition

- Texture is characterized by the repetition of basic elements or *textons*
- For stochastic textures, it is the identity of the textons, not their spatial arrangement, that matters



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

## Origin 1: Texture recognition



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

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

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



US Presidential Speeches Tag Cloud http://chir.ag/phernalia/preztags/

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



US Presidential Speeches Tag Cloud http://chir.ag/phernalia/preztags/

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 (1961-63)			
expand	aban do	1941-12-08: Request for a Declaration of War			
insurgen	buildı	Franklin D. Roosevelt (1933-45)			
palestinia	declined elimina	abandoning acknowledge aggression aggressors airplanes armaments <b>armed army</b> assault assembly <b>authorizations bombing</b> britain british cheerfully claiming constitution curtail december defeats defending delays democratic dictators disclose economic empire endanger <b>facts</b> false forgotten fortunes france <b>freedom</b> fulfilled fullness fundamental gangsters <b>german germany god</b> guam harbor hawaii <b>hemisphere</b> bint hitler hostilities immune improving indies innumerable			
septemt violenc	halt ha				
vioterie	modern				
	recession invasion ISLANDS isolate Japannesse labor metals midst midway Navy nazis obligation offensive				
	surveil	repaired <b>resisting</b> retain revealing rumors seas soldiers speaks speedy stamina strength sunday sunk supremacy tanks taxes			
		treachery true tyranny undertaken victory Wartime washington			

US Presidential Speeches Tag Cloud http://chir.ag/phernalia/preztags/

## Bags of features for object recognition



#### face, flowers, building

• Works pretty well for image-level classification

Csurka et al. (2004), Willamowski et al. (2005), Grauman & Darrell (2005), Sivic et al. (2003, 2005)

## Bags of features for object recognition

#### Caltech6 dataset



class	bag of features	bag of features	Parts-and-shape model
01055	Zhang et al. (2005)	Willamowski et al. (2004)	Fergus et al. (2003)
airplanes	98.8	97.1	90.2
cars (rear)	98.3	98.6	90.3
cars (side)	95.0	87.3	88.5
faces	100	99.3	96.4
motorbikes	98.5	98.0	92.5
spotted cats	97.0		90.0

1. Extract features







- 1. Extract features
- 2. Learn "visual vocabulary"



- 1. Extract features
- 2. Learn "visual vocabulary"
- 3. Quantize features using visual vocabulary

- 1. Extract features
- 2. Learn "visual vocabulary"
- 3. Quantize features using visual vocabulary
- Represent images by frequencies of "visual words"



### Regular grid

- Vogel & Schiele, 2003
- Fei-Fei & Perona, 2005



### Regular grid

- Vogel & Schiele, 2003
- Fei-Fei & Perona, 2005

#### Interest point detector

- Csurka et al. 2004
- Fei-Fei & Perona, 2005
- Sivic et al. 2005



### Regular grid

- Vogel & Schiele, 2003
- Fei-Fei & Perona, 2005

### Interest point detector

- Csurka et al. 2004
- Fei-Fei & Perona, 2005
- Sivic et al. 2005

#### Other methods

- Random sampling (Vidal-Naquet & Ullman, 2002)
- Segmentation-based patches (Barnard et al. 2003)





Compute SIFT descriptor [Lowe'99]

Normalize patch



#### **Detect patches**

[Mikojaczyk and Schmid '02] [Mata, Chum, Urban & Pajdla, '02] [Sivic & Zisserman, '03]





## 2. Learning the visual vocabulary



# 2. Learning the visual vocabulary



Slide credit: Josef Sivic

# 2. Learning the visual vocabulary



Slide credit: Josef Sivic

## K-means clustering

 Want to minimize sum of squared Euclidean distances between points x<sub>i</sub> and their nearest cluster centers m<sub>k</sub>

$$D(X,M) = \sum (x_i - m_k)^2$$

cluster k point i in cluster k

Algorithm:

- Randomly initialize K cluster centers
- Iterate until convergence:
  - Assign each data point to the nearest center
  - Recompute each cluster center as the mean of all points assigned to it

## From clustering to vector quantization

- Clustering is a common method for learning a visual vocabulary or codebook
  - Unsupervised learning process
  - Each cluster center produced by k-means becomes a codevector
  - Codebook can be learned on separate training set
  - Provided the training set is sufficiently representative, the codebook will be "universal"
- The codebook is used for quantizing features
  - A vector quantizer takes a feature vector and maps it to the index of the nearest codevector in a codebook
  - Codebook = visual vocabulary
  - Codevector = visual word

## **Example visual vocabulary**



Fei-Fei et al. 2005

## Image patch examples of visual words



































Sivic et al. 2005

## Visual vocabularies: Issues

- How to choose vocabulary size?
  - Too small: visual words not representative of all patches
  - Too large: quantization artifacts, overfitting
- Generative or discriminative learning?
- Computational efficiency
  - Vocabulary trees (Nister & Stewenius, 2006)



## **3. Image representation**



codewords

## Image classification

 Given the bag-of-features representations of images from different classes, how do we learn a model for distinguishing them?

