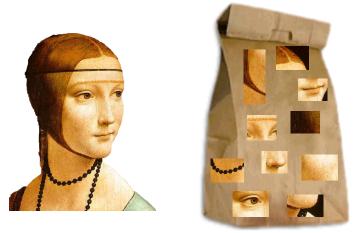


Working with Images: Bag of Visual Words



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

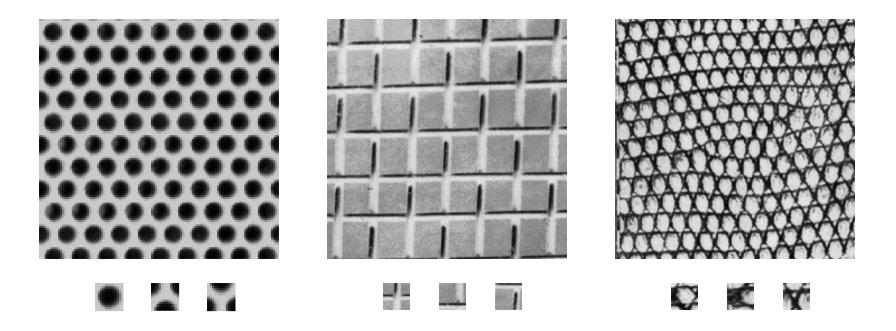
Bag-of-features models





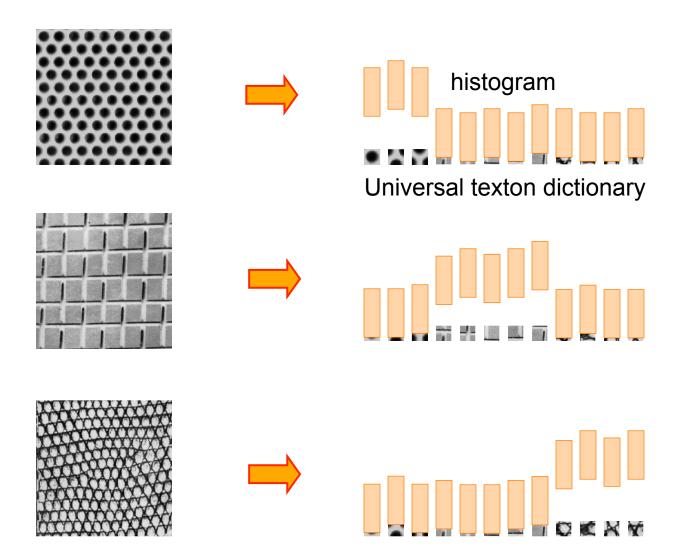
Origin 1: Texture recognition

- 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)



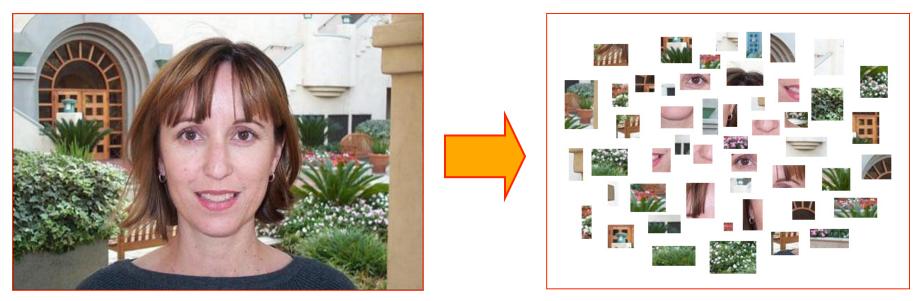
US Presidential Speeches Tag Cloud http://chir.ag/phernalia/preztags/

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

2007-0	1-23: St	ate of the Union Address George W. Bush (2001-)			
abandon choices c deficit c	1962-	1962-10-22: Soviet Missiles in Cuba John F. Kennedy (1961-63)			
expand	abando	1941-12-08: Request for a Declaration of War			
incurren	buildu	Franklin D. Roosevelt (1933-45)			
insurgen palestini	declineo elimina	Edition Edition and a section of the			
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			
	surveil	officially pacific 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			
		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

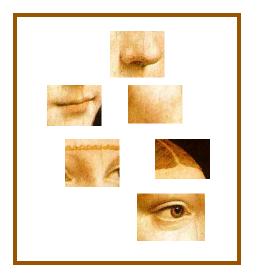
Csurka et al. (2004), Willamowski et al. (2005), Grauman & Darrell (2005), Sivic et al. (2003, 2005)

Bags of features for object recognition



class	bag of features	bag of features	Parts-and-shape model
Class	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





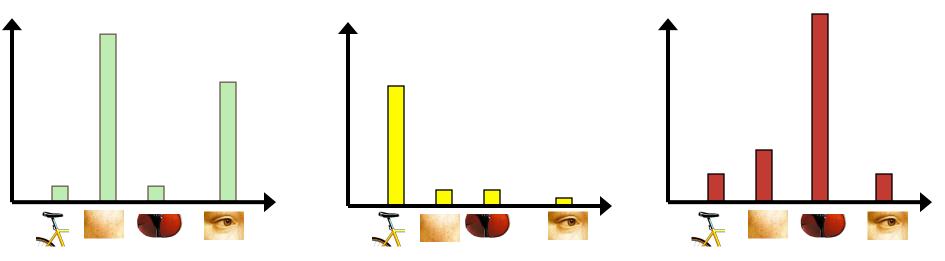


- 1. Extract features
- 2. Learn "visual vocabulary"



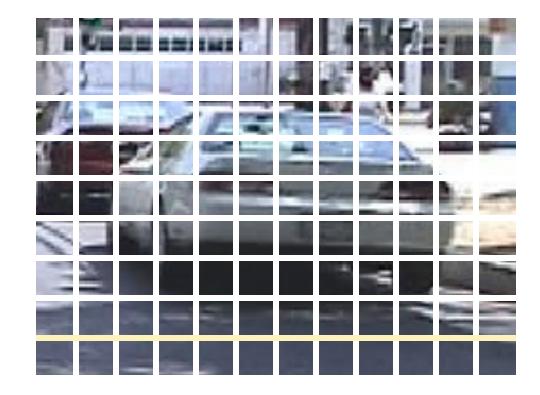
- 1. Extract features
- 2. Learn "visual vocabulary"
- 3. Quantize features using visual vocabulary

- 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

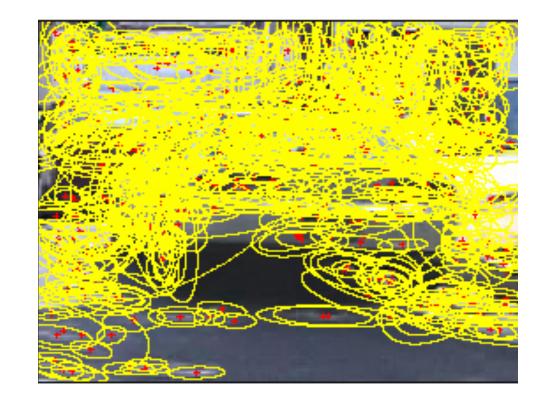


Regular grid

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

Interest point detector

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



Regular grid

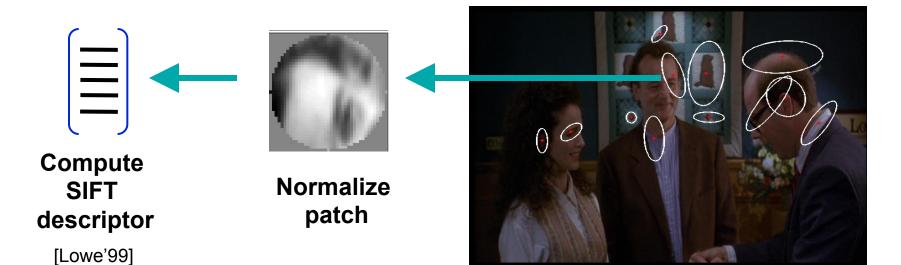
- Vogel & Schiele, 2003
- Fei-Fei & Perona, 2005

Interest point detector

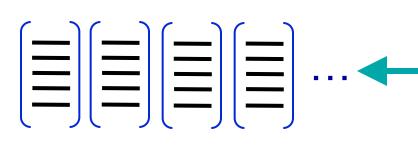
- Csurka et al. 2004
- Fei-Fei & Perona, 2005
- Sivic et al. 2005

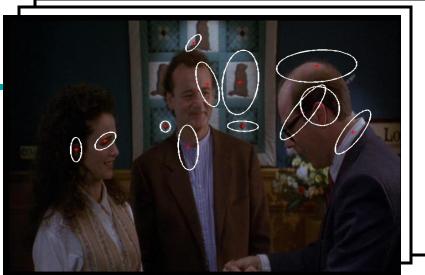
Other methods

- Random sampling (Vidal-Naquet & Ullman, 2002)
- Segmentation based patches (Barnard, Duygulu, Forsyth, de Freitas, Blei, Jordan, 2003)

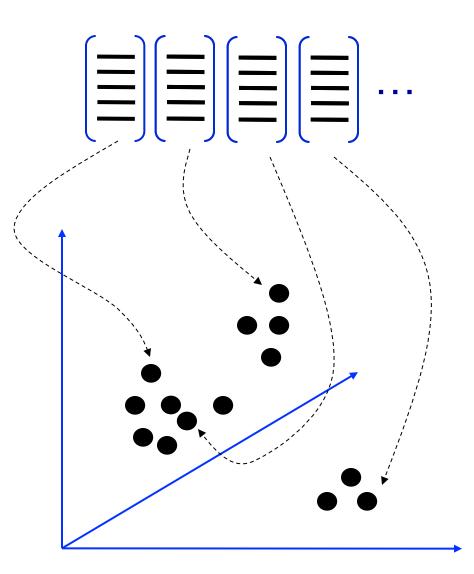


Detect patches [Mikojaczyk and Schmid '02] [Mata, Chum, Urban & Pajdla, '02] [Sivic & Zisserman, '03]

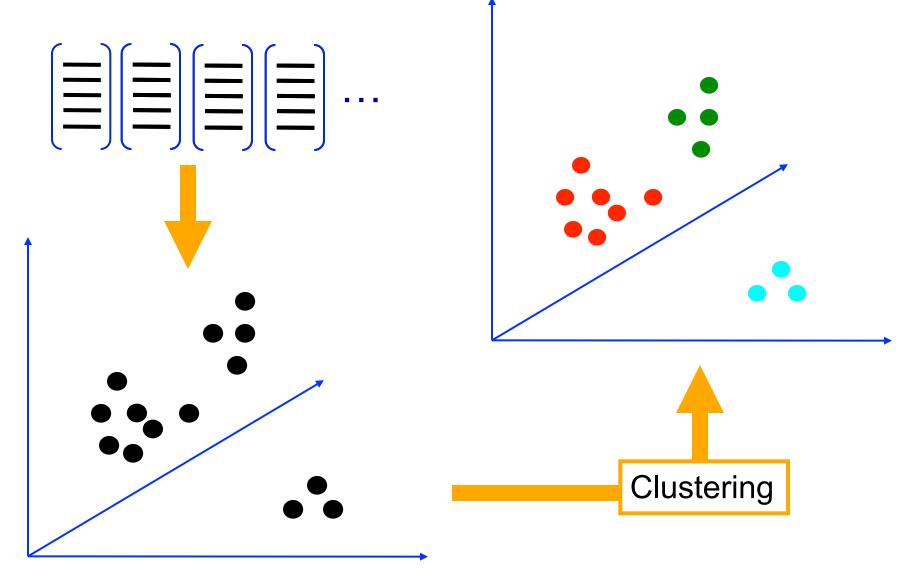




2. Learning the visual vocabulary

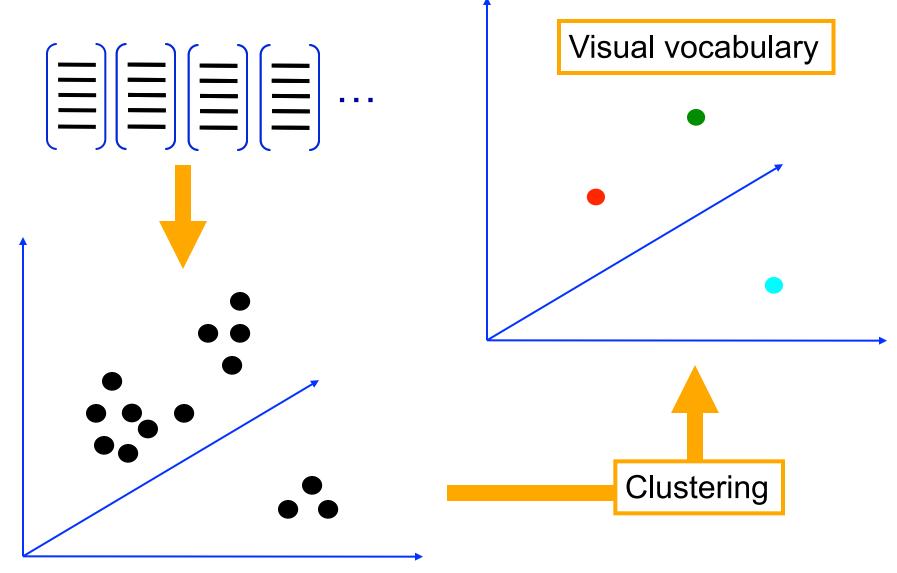


2. Learning the visual vocabulary



Slide credit: Josef Sivic

2. Learning the visual vocabulary

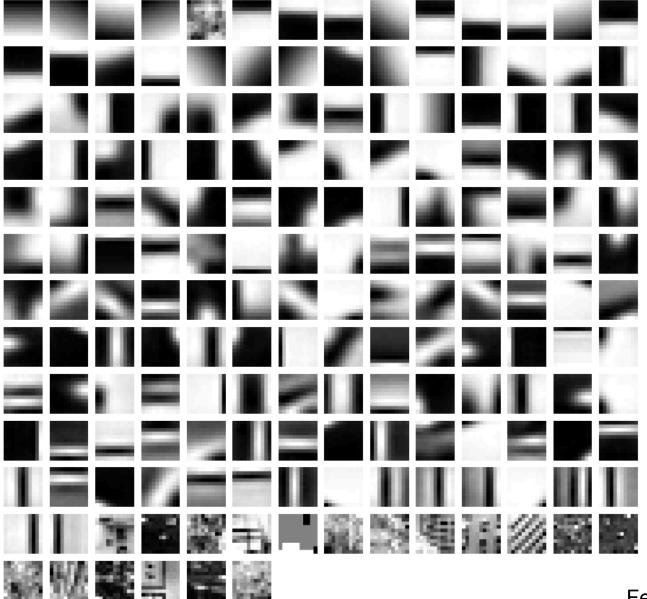


Slide credit: Josef Sivic

From clustering to vector quantization

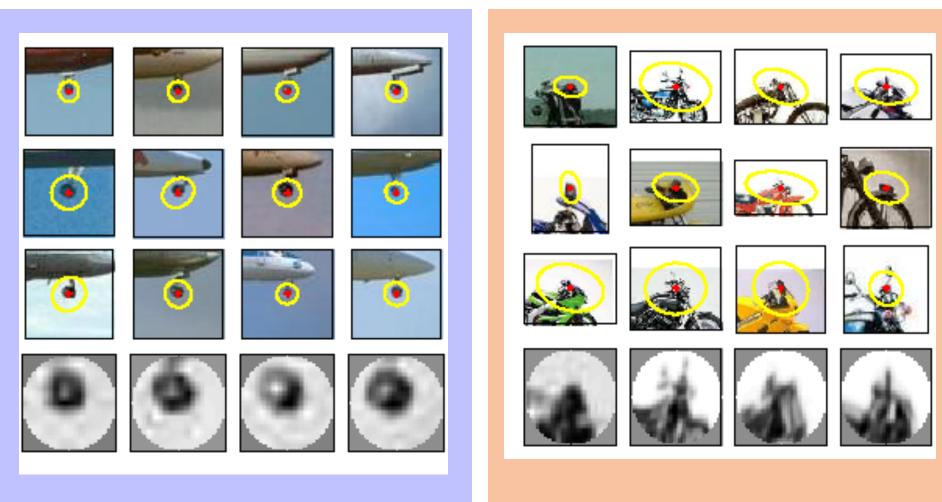
- Clustering is a common method for learning a visual vocabulary or codebook
 - Unsupervised learning process
 - Each cluster center produced by k-means becomes a codevector
 - Codebook can be learned on separate training set
 - Provided the training set is sufficiently representative, the codebook will 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



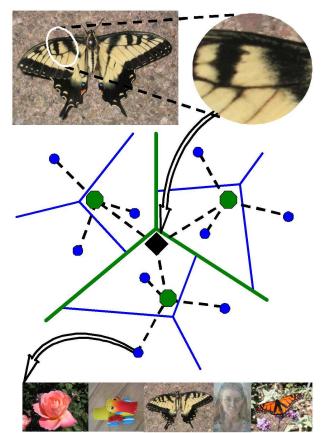
Fei-Fei et al. 2005

Image patch examples of visual words



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

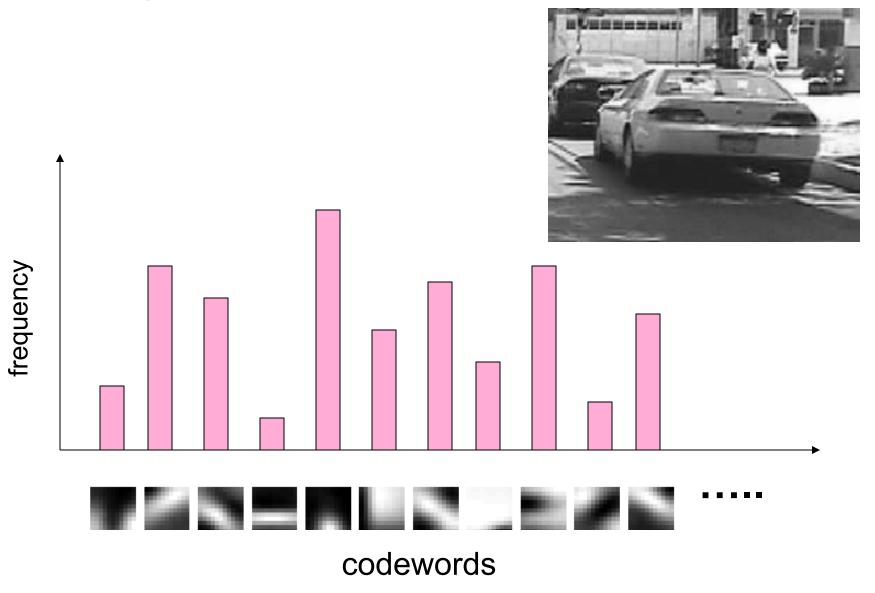


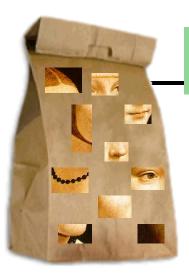
Image classification

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









Weakness of the model

No rigorous geometric information of the object components

It's intuitive to most of us that objects are made of parts – no such information

Not extensively tested yet for

- View point invariance
- Scale invariance

Segmentation and localization unclear