

Bag-of-features models



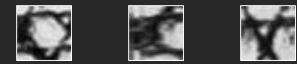
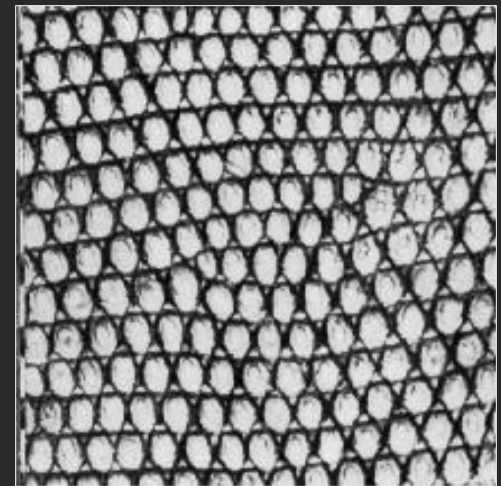
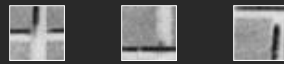
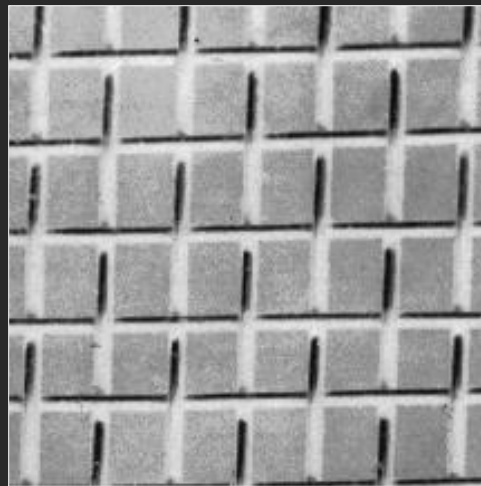
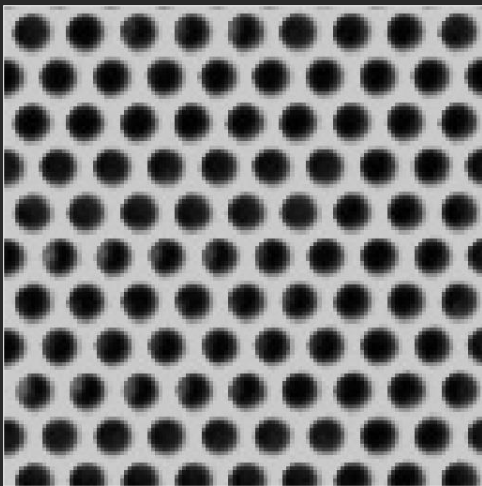
Many slides adapted from Fei-Fei Li, Rob Fergus, and Antonio Torralba

Overview: Bag-of-features models

- Origins and motivation
- Learning visual vocabularies
 - K-means clustering
- Discriminative methods
 - Nearest-neighbor classification
 - Distance functions
 - Support vector machines
 - Kernels
- Generative methods
 - Naïve Bayes
 - Probabilistic Latent Semantic Analysis
- Extensions: incorporating spatial information

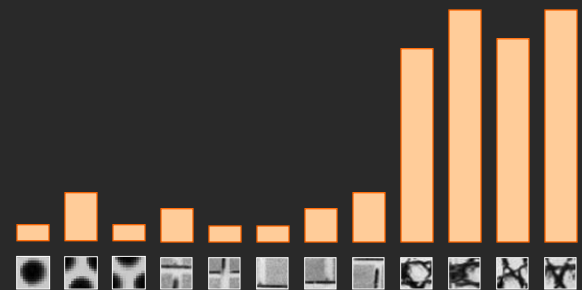
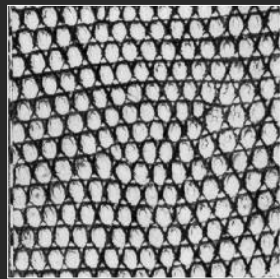
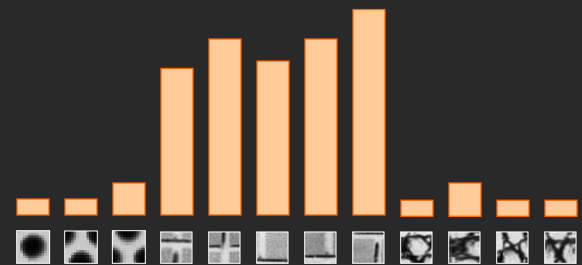
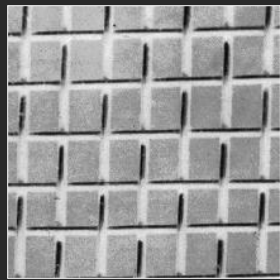
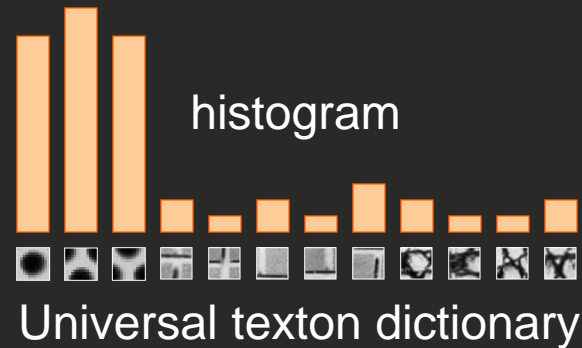
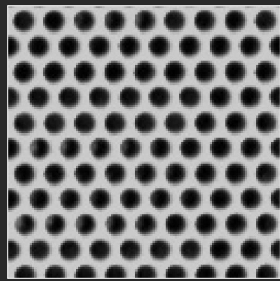
Origin 1: Texture recognition

- 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

Origin 2: Bag-of-words models

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

Origin 2: Bag-of-words models

- 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 **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
september shia stays strength students succeed sunni tax territories **terrorists** threats uphold victory
violence violent **war** washington weapons wesley

US Presidential Speeches Tag Cloud

<http://chir.ag/phernalia/preztags/>

Origin 2: Bag-of-words models

- 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 d

deficit d

expand

insurgen

palestini

septemb

violenc

1962-10-22: Soviet Missiles in Cuba

John F. Kennedy (1961-63)

abandon achieving adversaries aggression agricultural appropriate armaments **arms** assessments atlantic ballistic berlin
buildup burdens cargo college commitment communist constitution consumers cooperation crisis **cuba** dangers
declined **defensive** deficit **depended** disarmament divisions domination doubled **economic** education
elimination emergence endangered equals **europe** expand exports fact false family forum **freedom** fulfill gromyko
halt hazards **hemisphere** hospitals ideals **independent** industries inflation labor latin limiting minister **missiles**
modernization neglect **nuclear** oas obligation observer **offensive** peril pledged predicted purchasing quarantine **quote**
recession rejection republics retaliatory safeguard sites solution **soviet** space spur stability standby **strength**
surveillance **tax** territory treaty undertakings unemployment **war** warhead **weapons** welfare western widen withdraw

US Presidential Speeches Tag Cloud

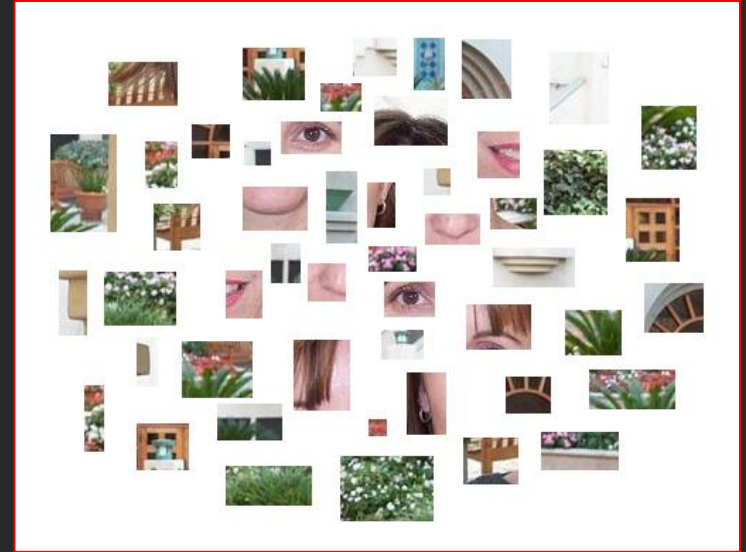
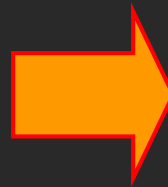
<http://chir.ag/phernalia/preztags/>

Origin 2: Bag-of-words models

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



Bags of features for object recognition



face, flowers, building

- Works pretty well for image-level classification

Bags of features for object recognition

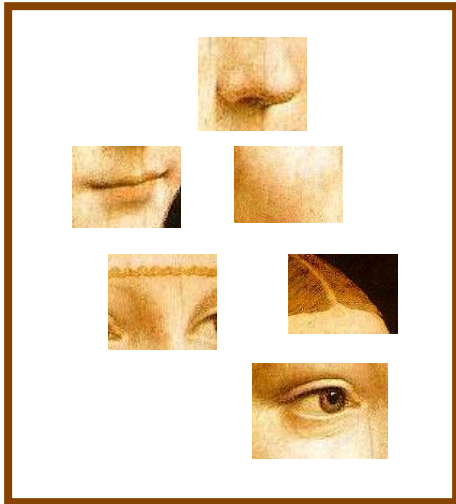
Caltech6 dataset



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

Bag of features: outline

1. Extract features



Bag of features: outline

1. Extract features
2. Learn “visual vocabulary”

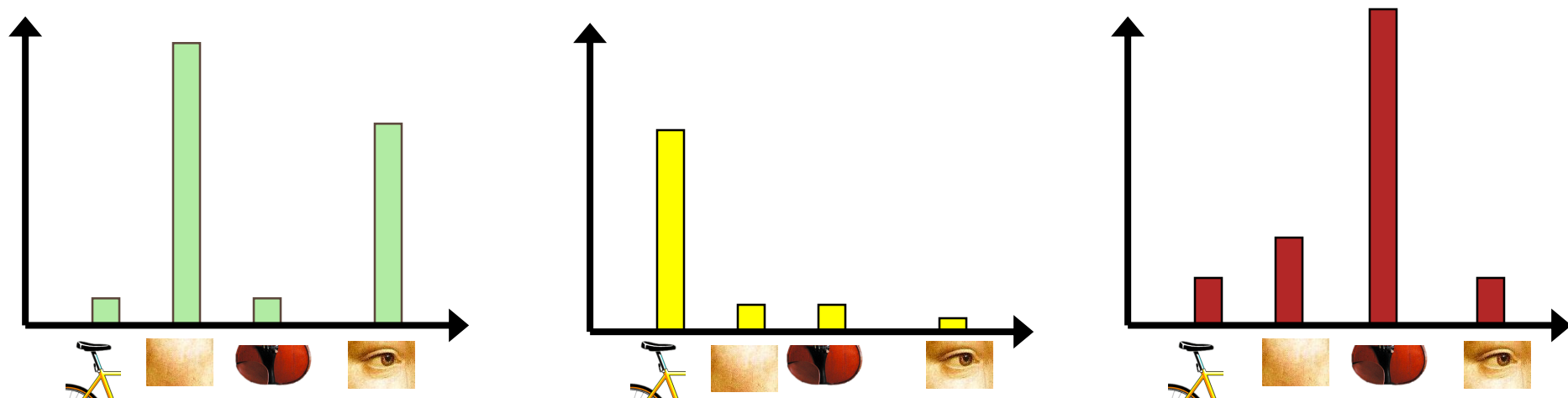


Bag of features: outline

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary

Bag of features: outline

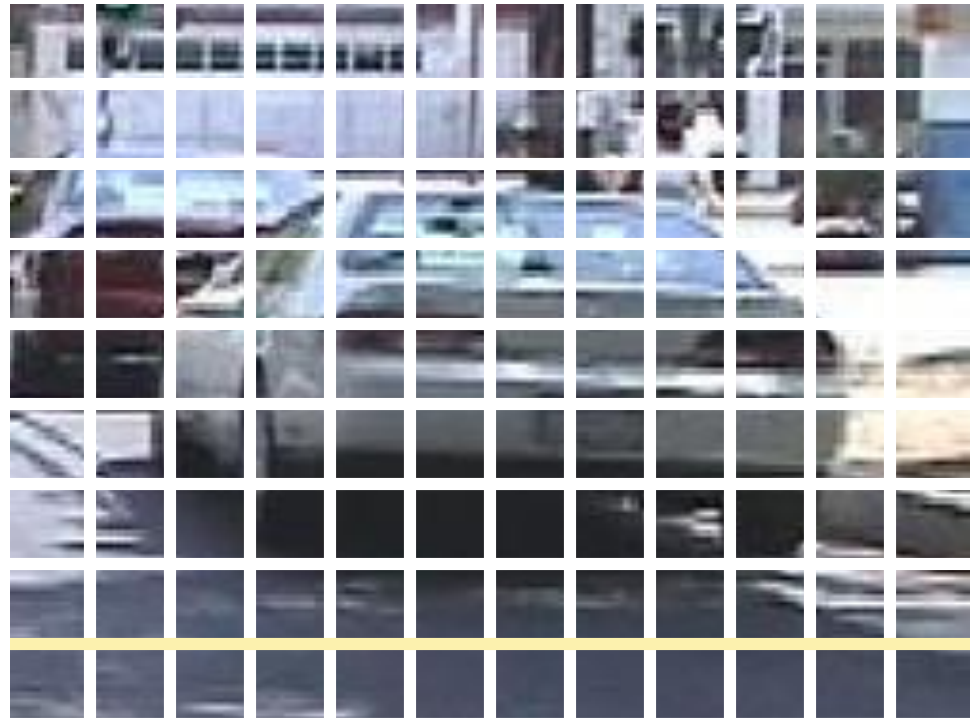
1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
4. Represent images by frequencies of “visual words”



1. Feature extraction

Regular grid

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



1. Feature extraction

Regular grid

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

Interest point detector

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



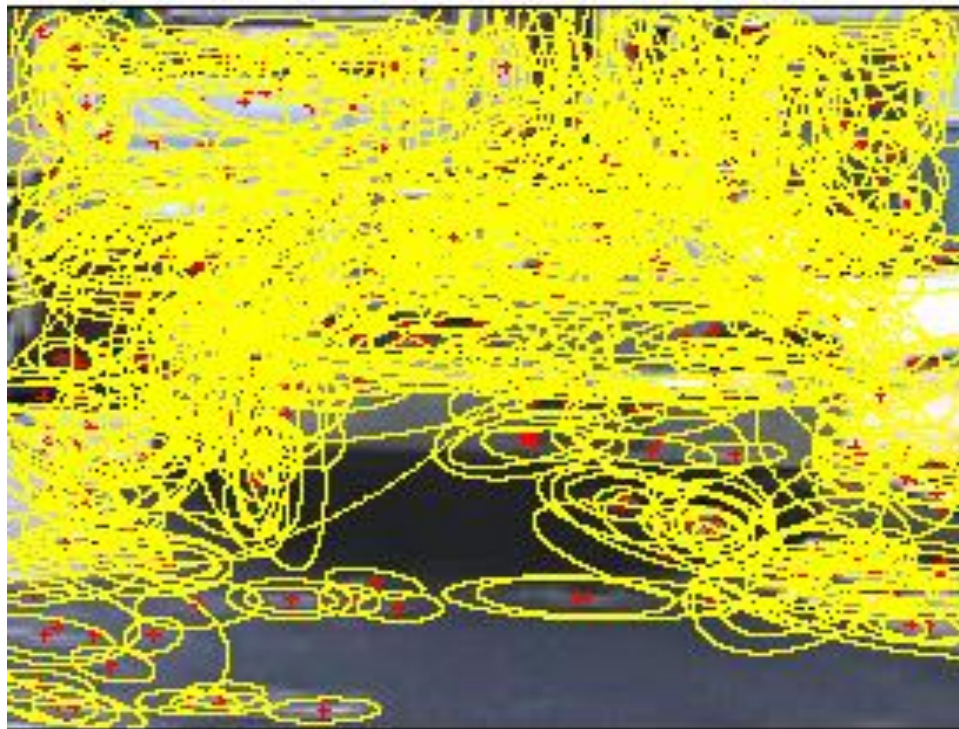
1. Feature extraction

Regular grid

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

Interest point detector

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



1. Feature extraction

Regular grid

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

Interest point detector

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

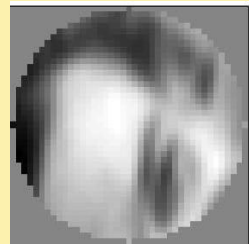
Other methods

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

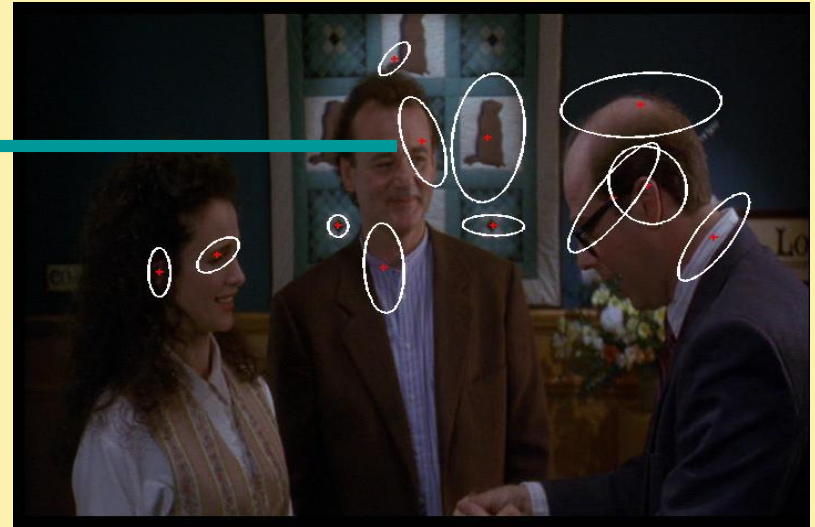
1. Feature extraction



**Compute
SIFT
descriptor**
[Lowe'99]



**Normalize
patch**



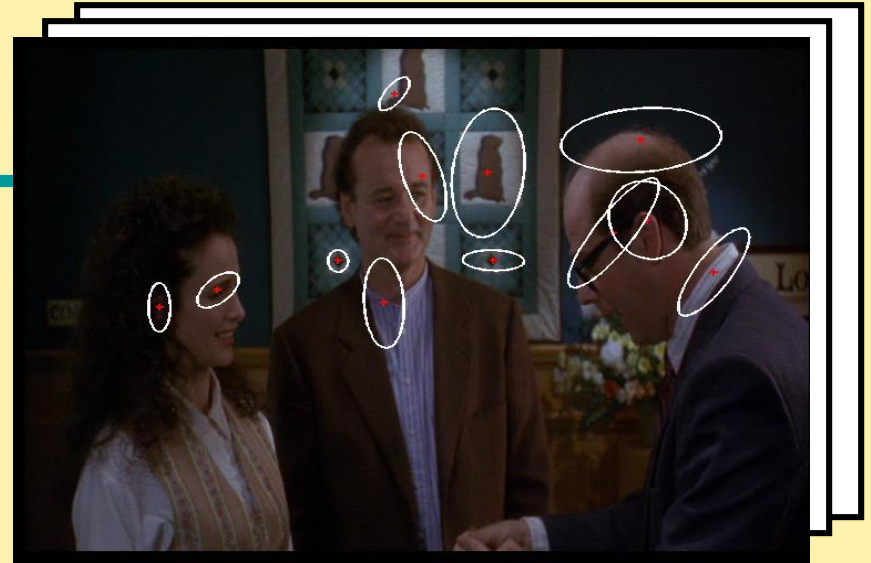
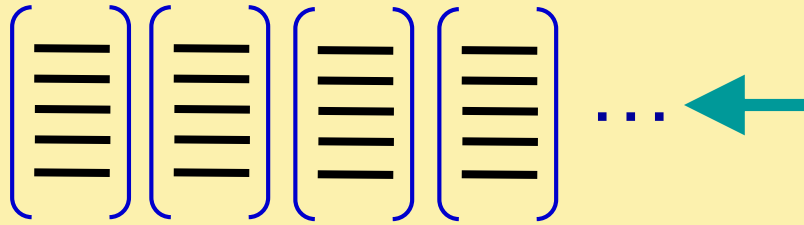
Detect patches

[Mikojaczyk and Schmid '02]

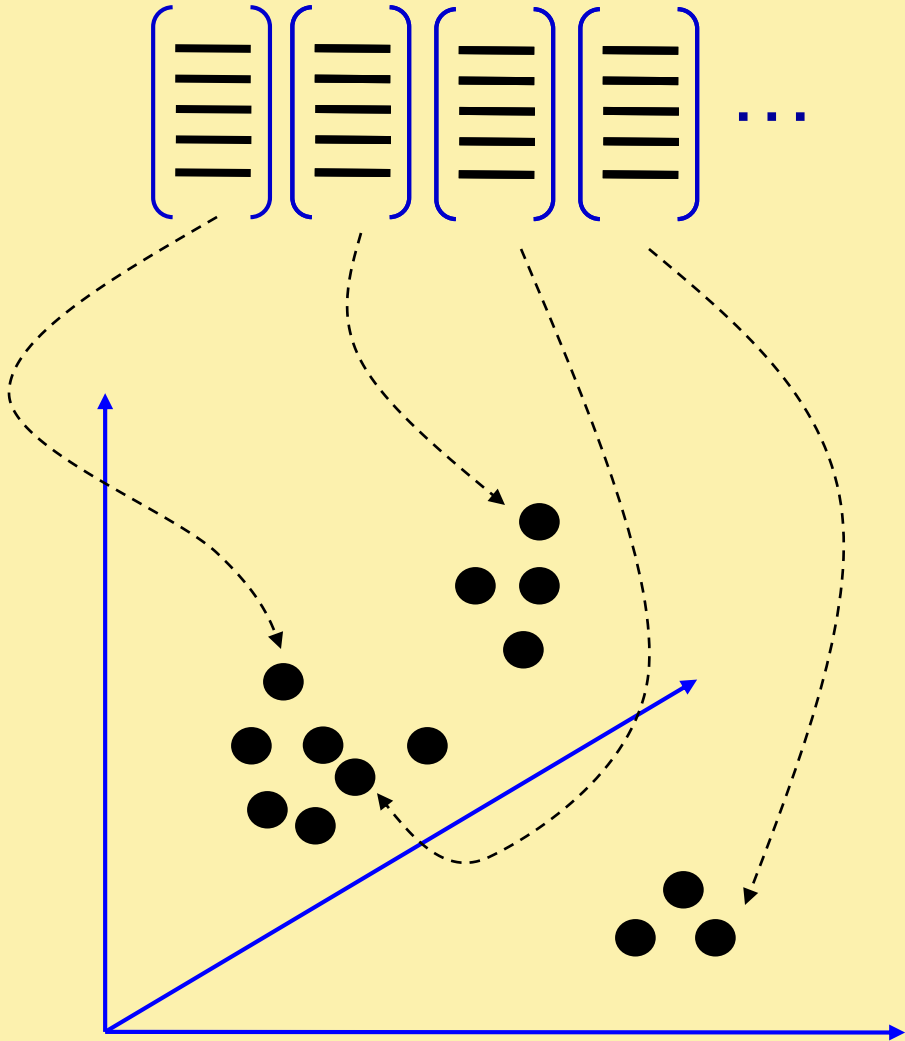
[Mata, Chum, Urban & Pajdla, '02]

[Sivic & Zisserman, '03]

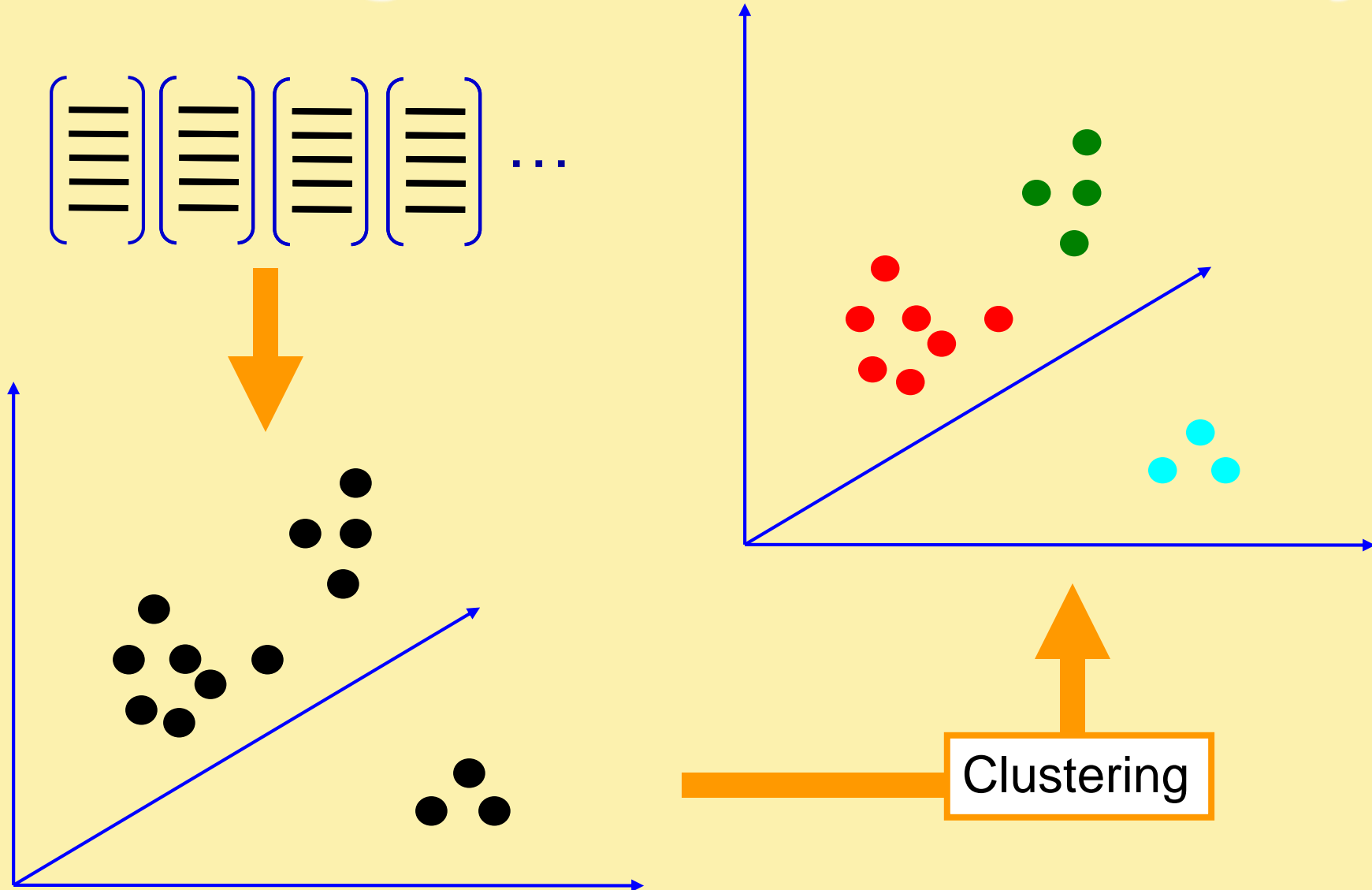
1. Feature extraction



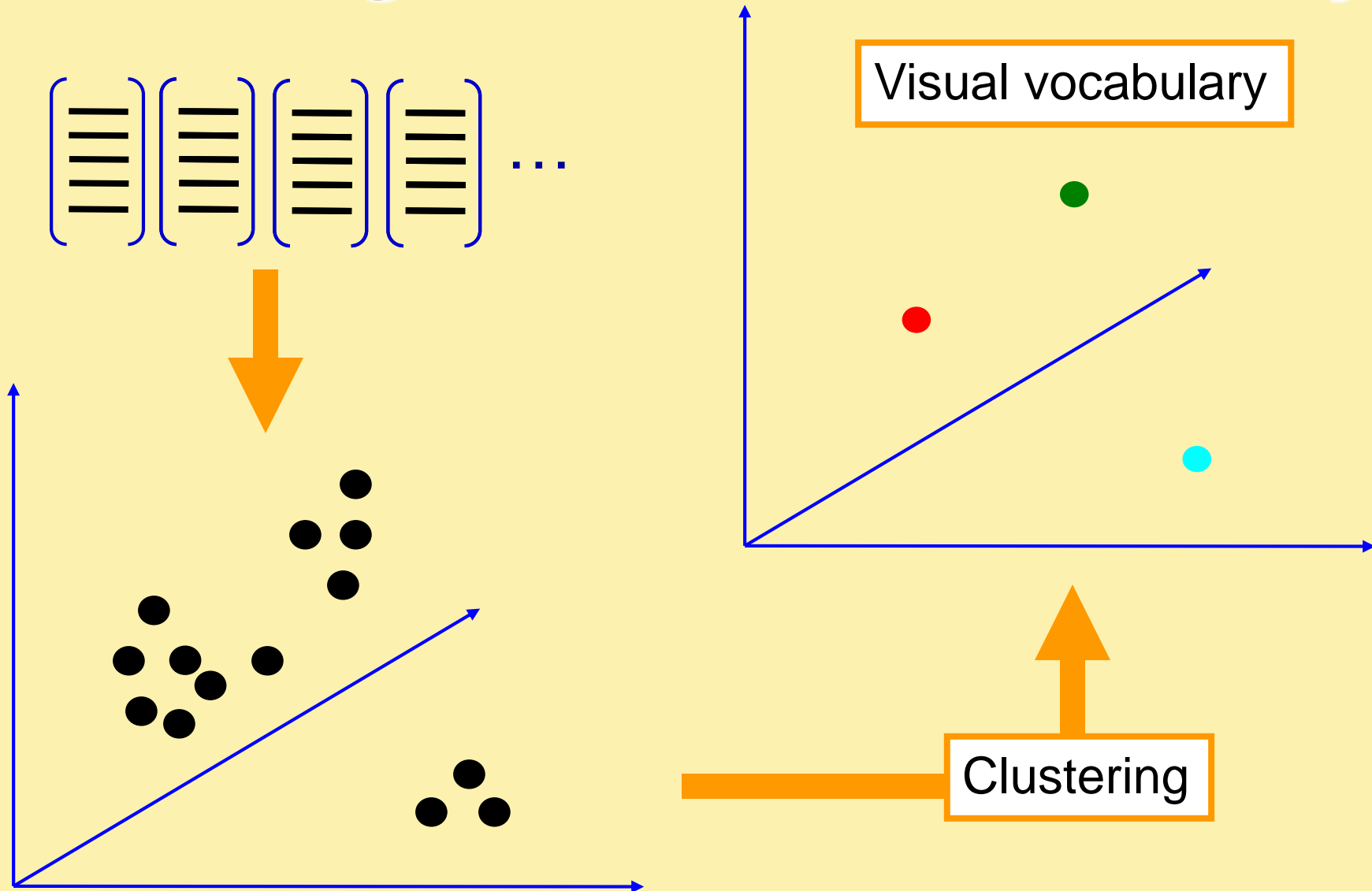
2. Learning the visual vocabulary



2. Learning the visual vocabulary



2. Learning the visual vocabulary



K-means clustering

- Want to minimize sum of squared Euclidean distances between points x_i and their nearest cluster centers m_k

$$D(X, M) = \sum_{\text{cluster } k} \sum_{\substack{\text{point } i \text{ in} \\ \text{cluster } k}} (x_i - m_k)^2$$

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

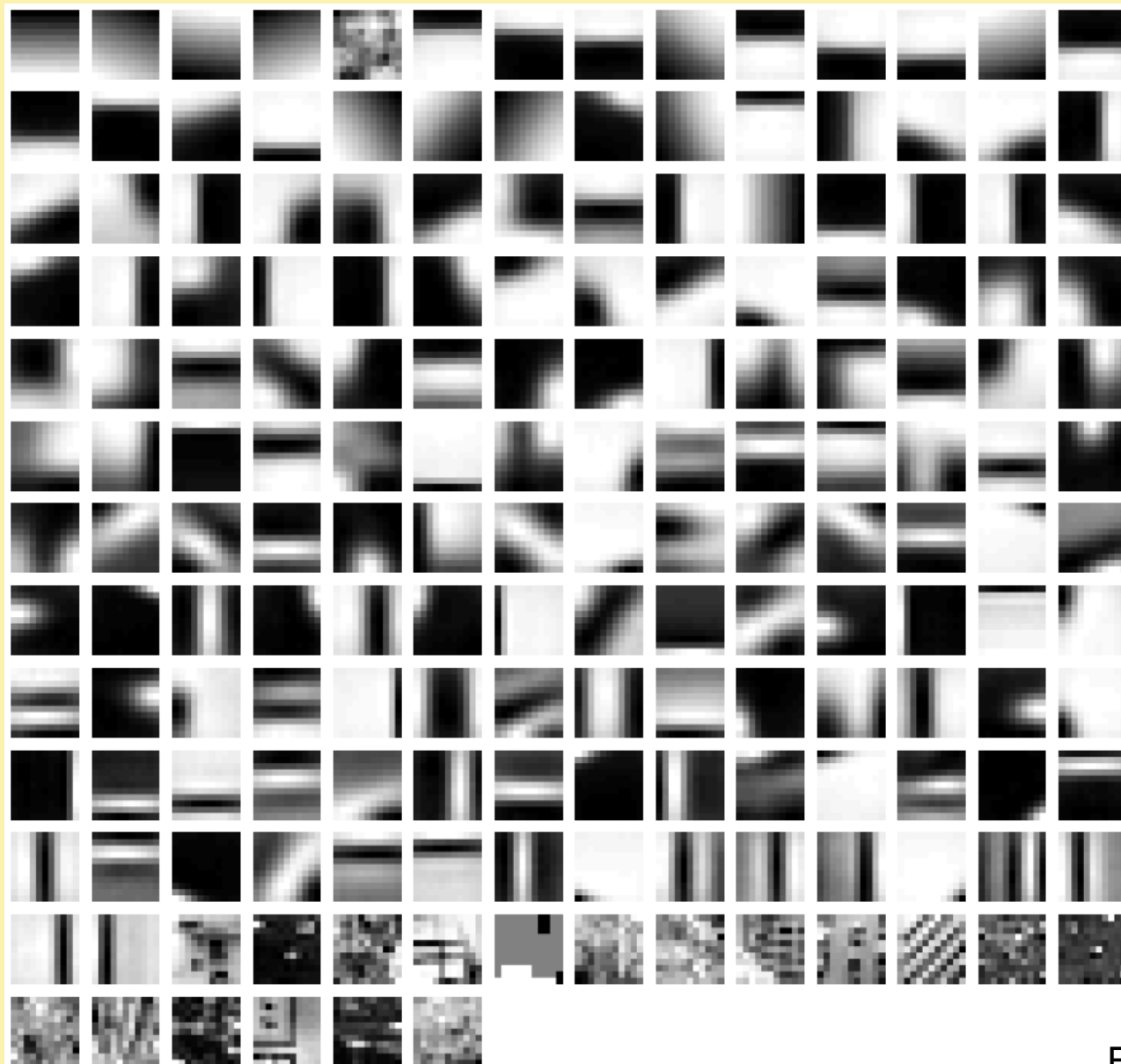
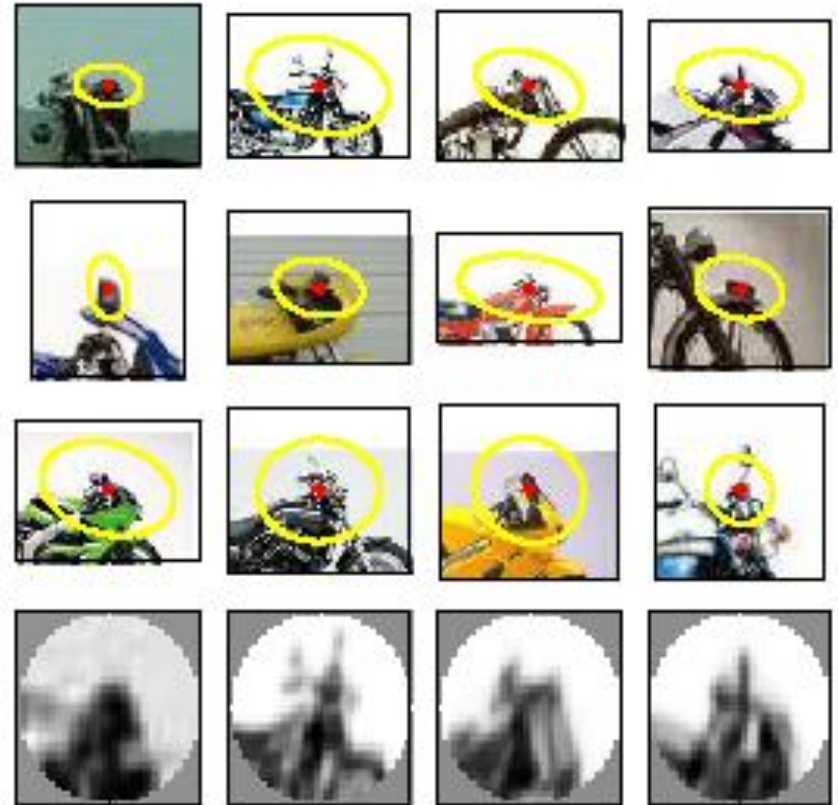
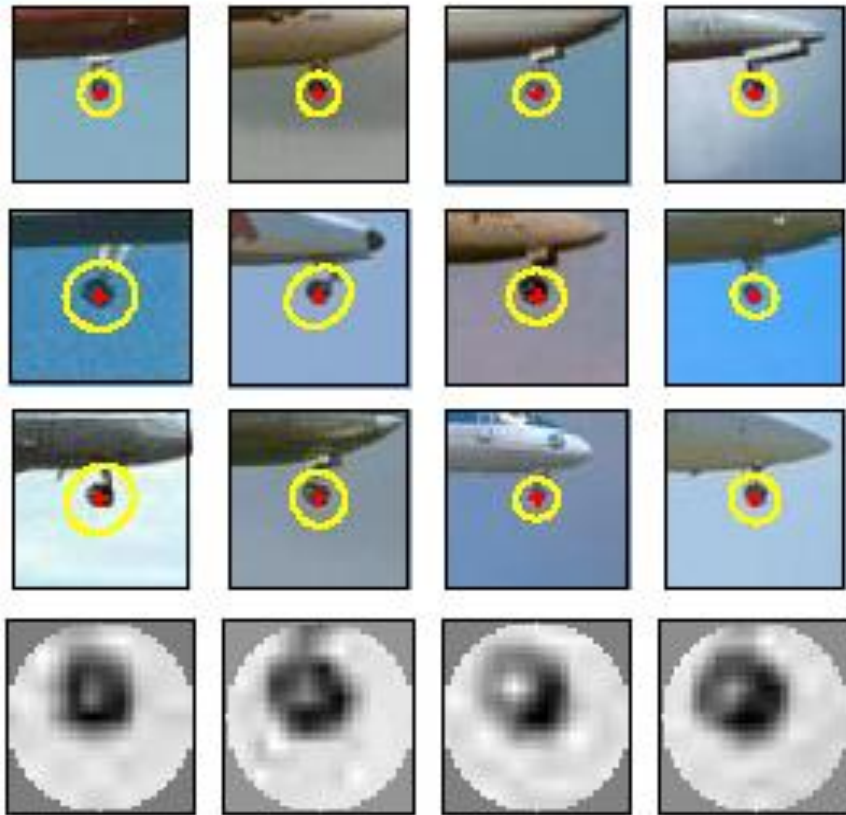
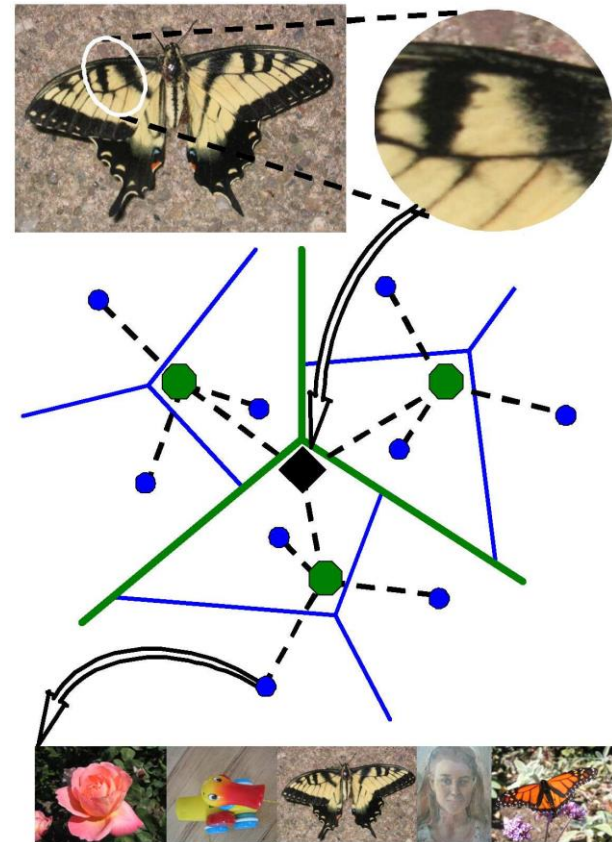


Image patch examples of visual words



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

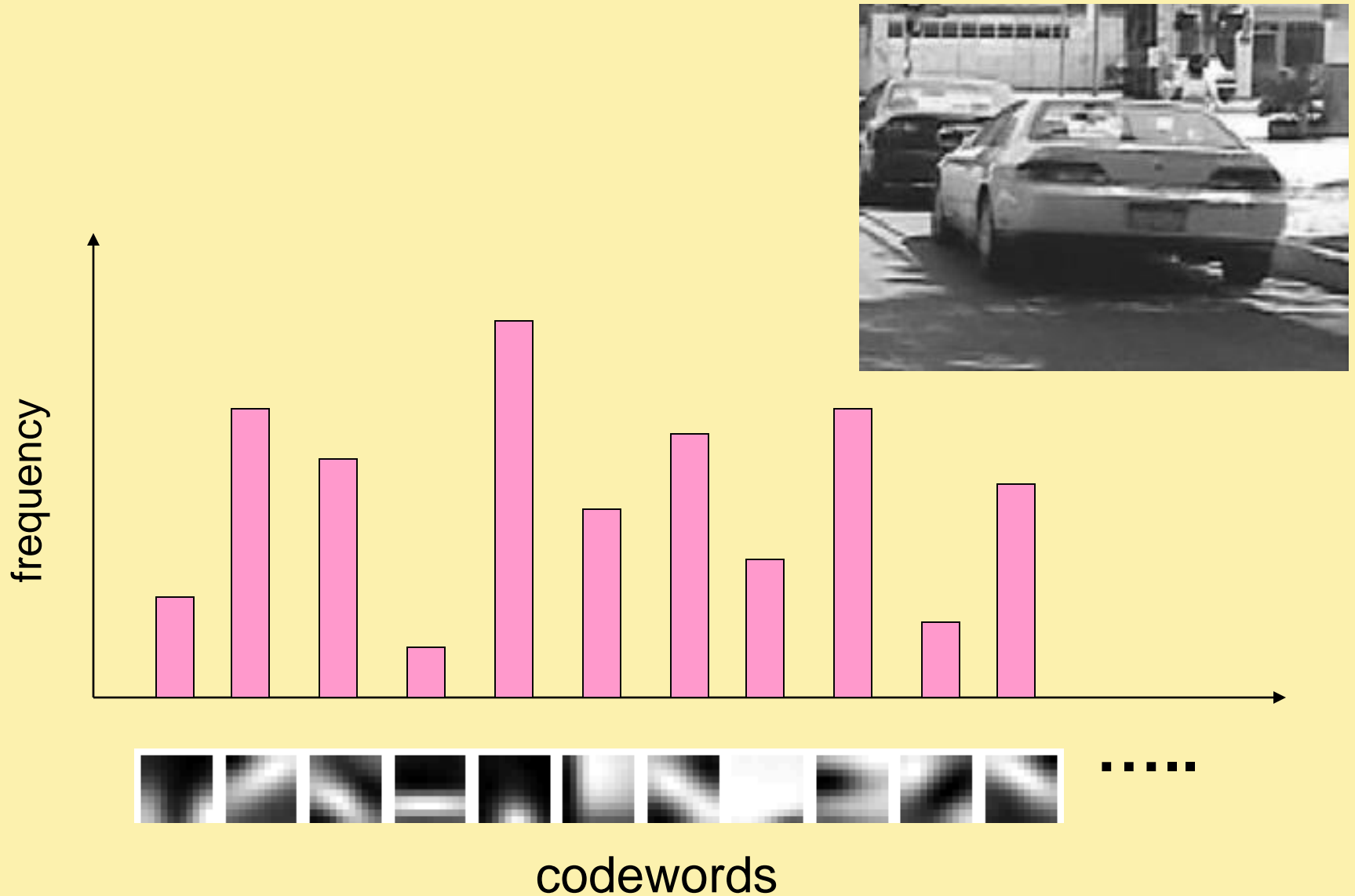


Image classification

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

