Lecture:
Visual Bag of Words

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

What we will learn today

- Visual bag of words (BoW)
- Spatial Pyramid Matching
- Naive Bayes

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

What we will learn today

- Visual bag of words (BoW)
- Spatial Pyramid Matching
- Naïve Bayes

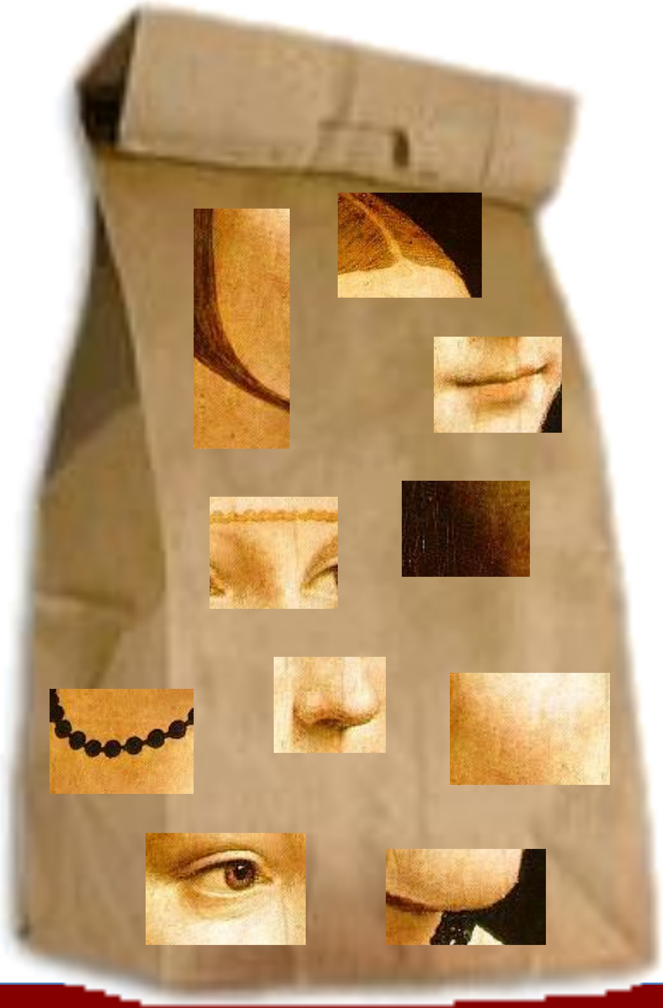
Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Bag of Words Models

Adapted from slides by Rob Fergus and Svetlana Lazebnik

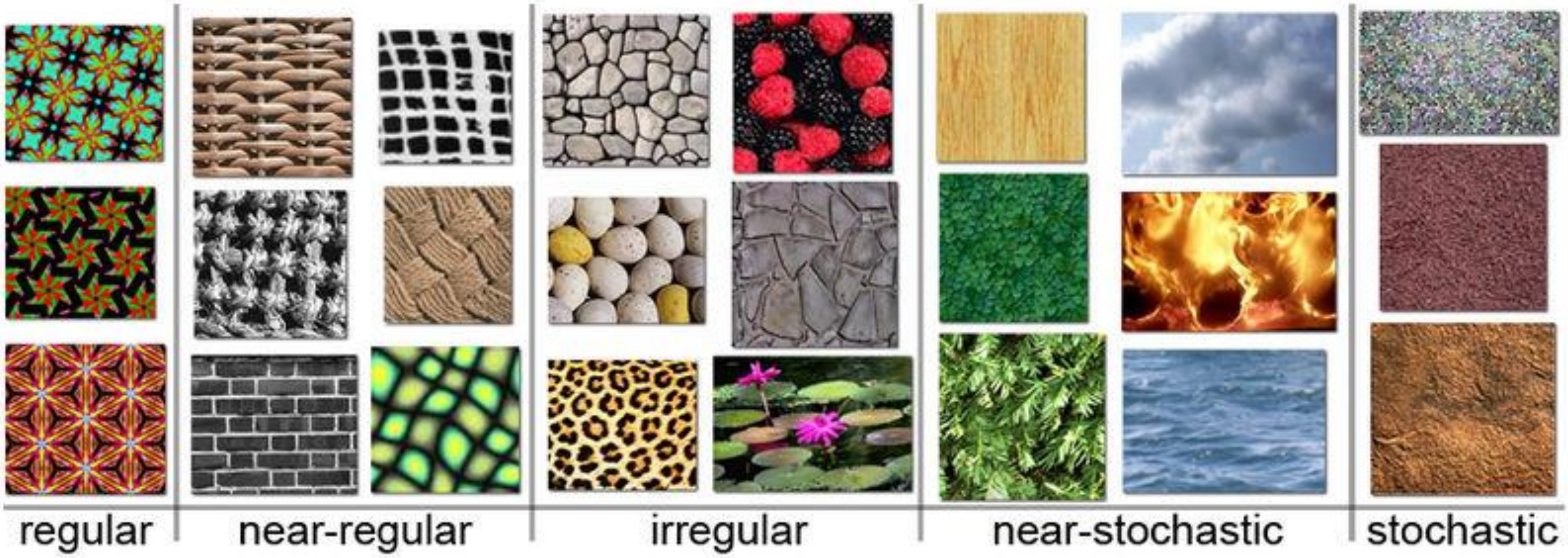
Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Object → Bag of ‘words’



Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Origin 1: Texture Recognition

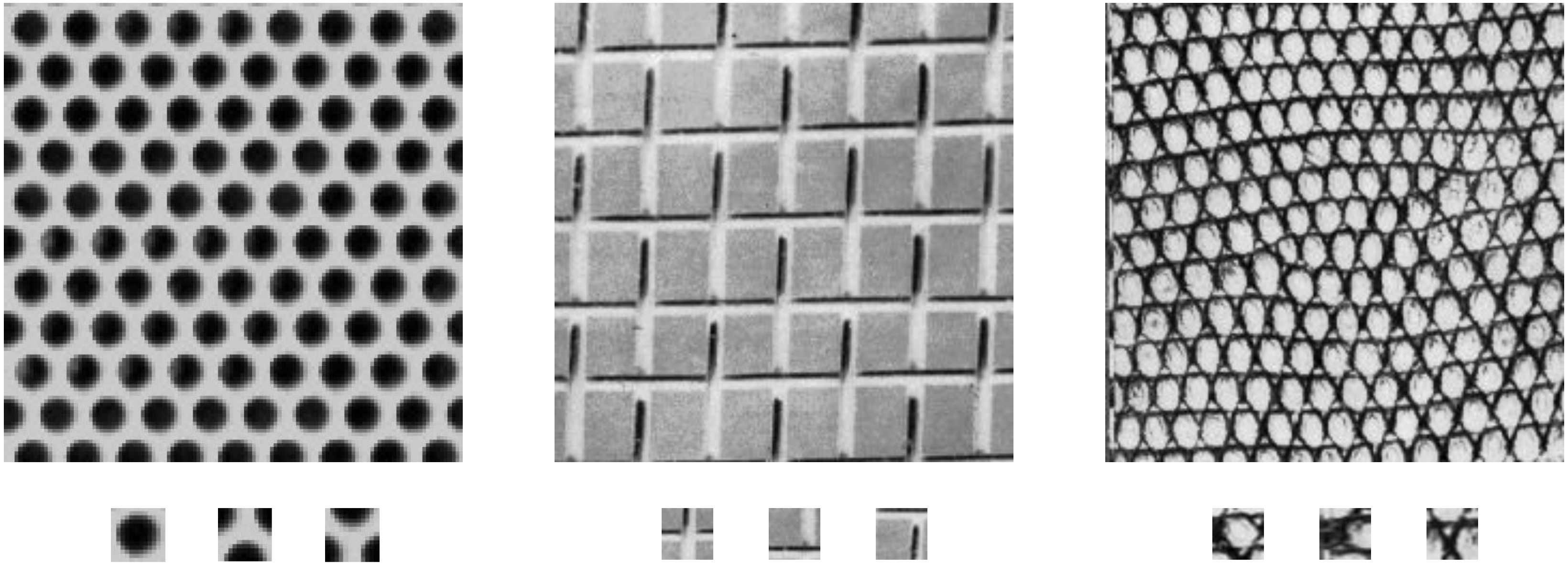


Example textures (from Wikipedia)

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Origin 1: Texture recognition

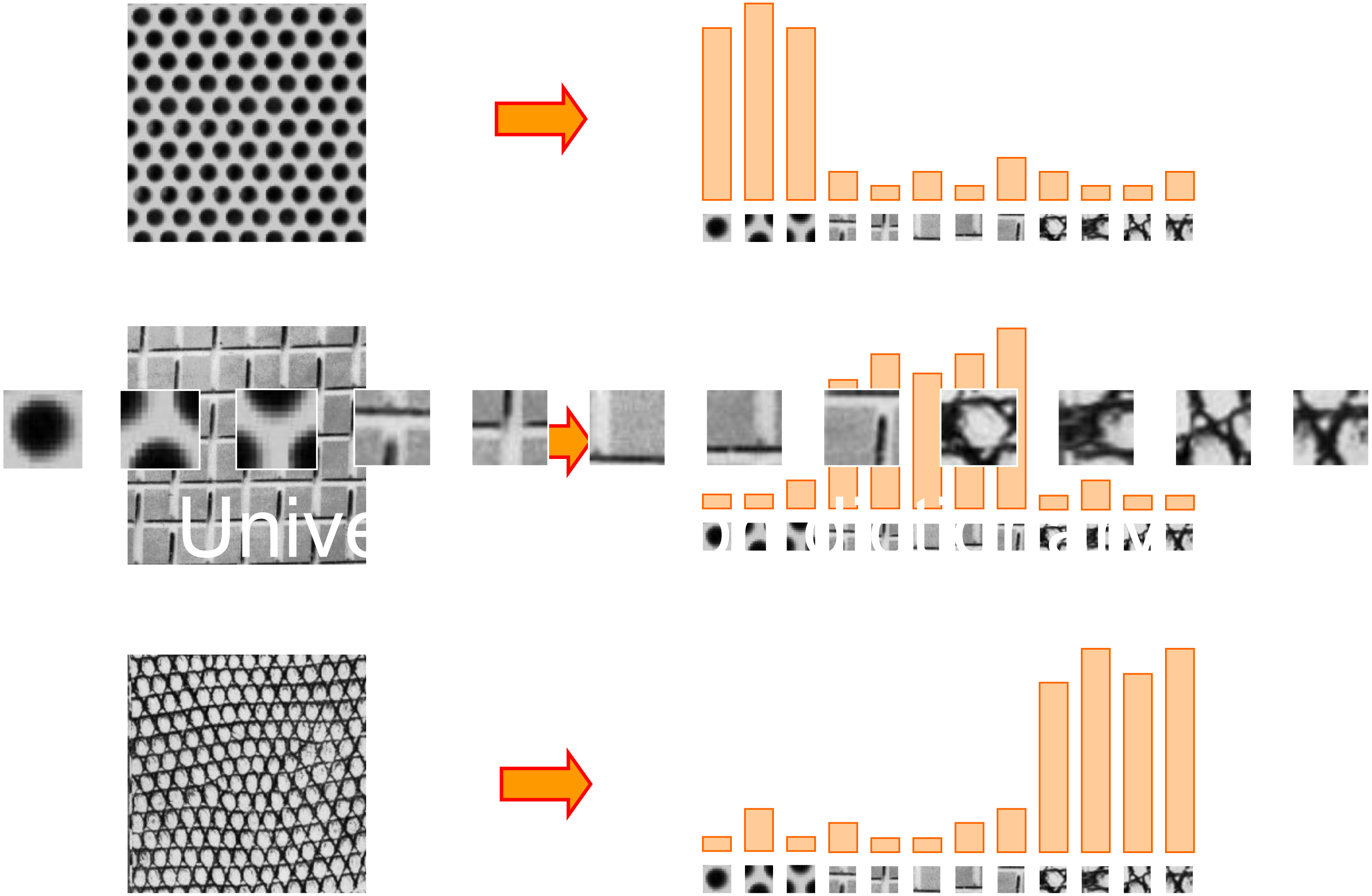
- Texture is characterized by the repetition of basic elements or ***textons***



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

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Origin 1: Texture recognition



Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

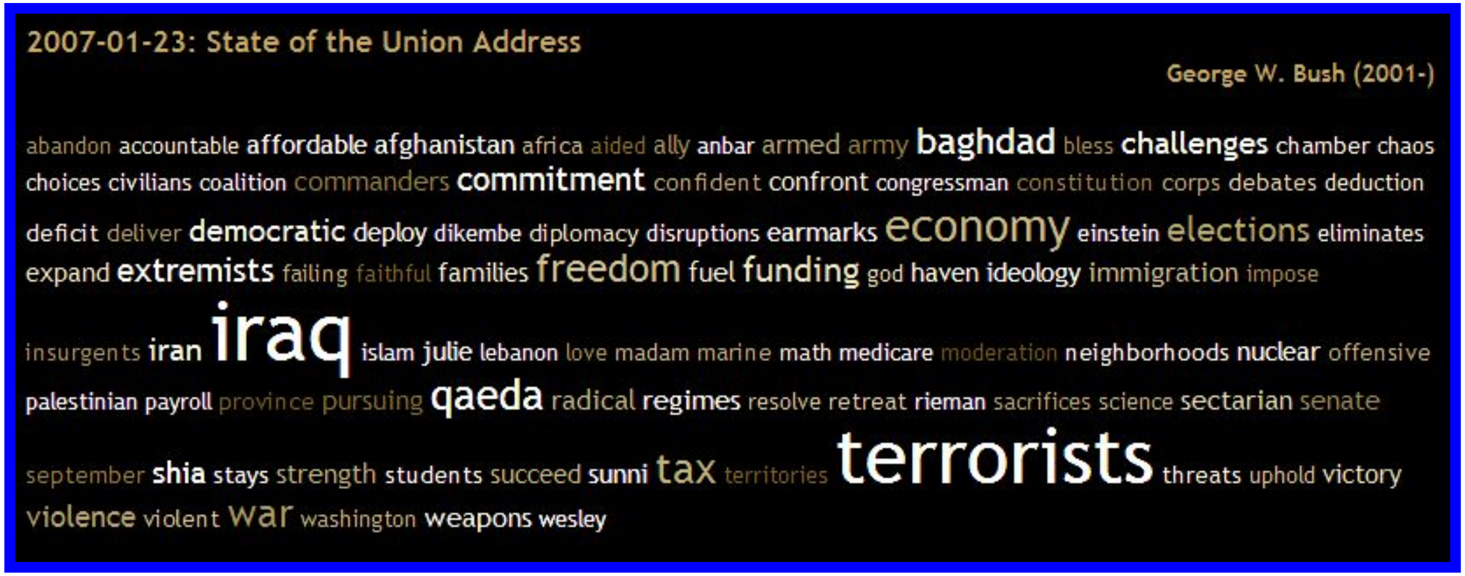
Origin 2: Bag-of-words models

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

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

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-)

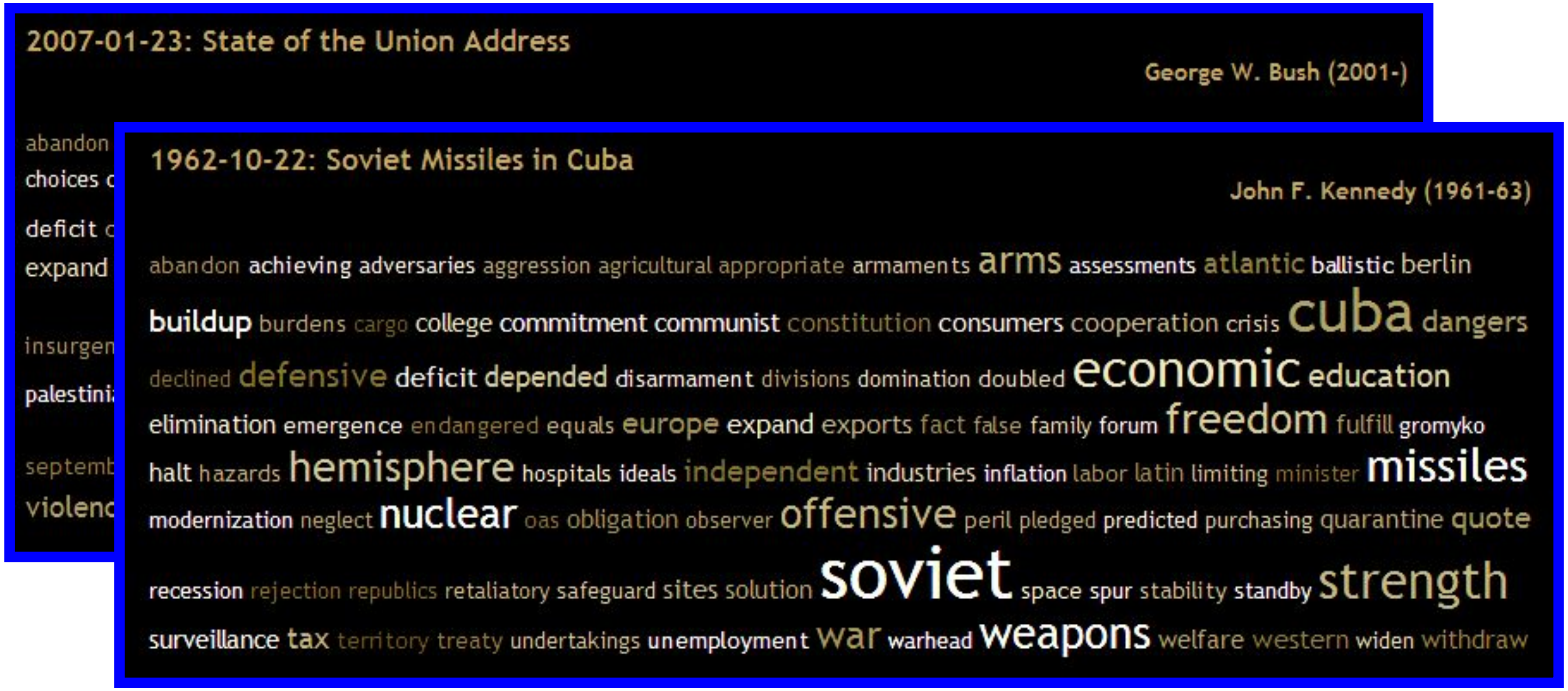
abandon accountable affordable afghanistan africa aided ally anbar armed army baghdad bless challenges chamber chaos choices civilians coalition commanders commitment confident confront congressman constitution corps debates deduction deficit deliver democratic deploy dikembe diplomacy disruptions earmarks 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/`

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

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 c
deficit c
expand
insurgen
palestini
septemb
violenc

1962-10-22: Soviet Missiles in Cuba

John F. Kennedy (1961-63)

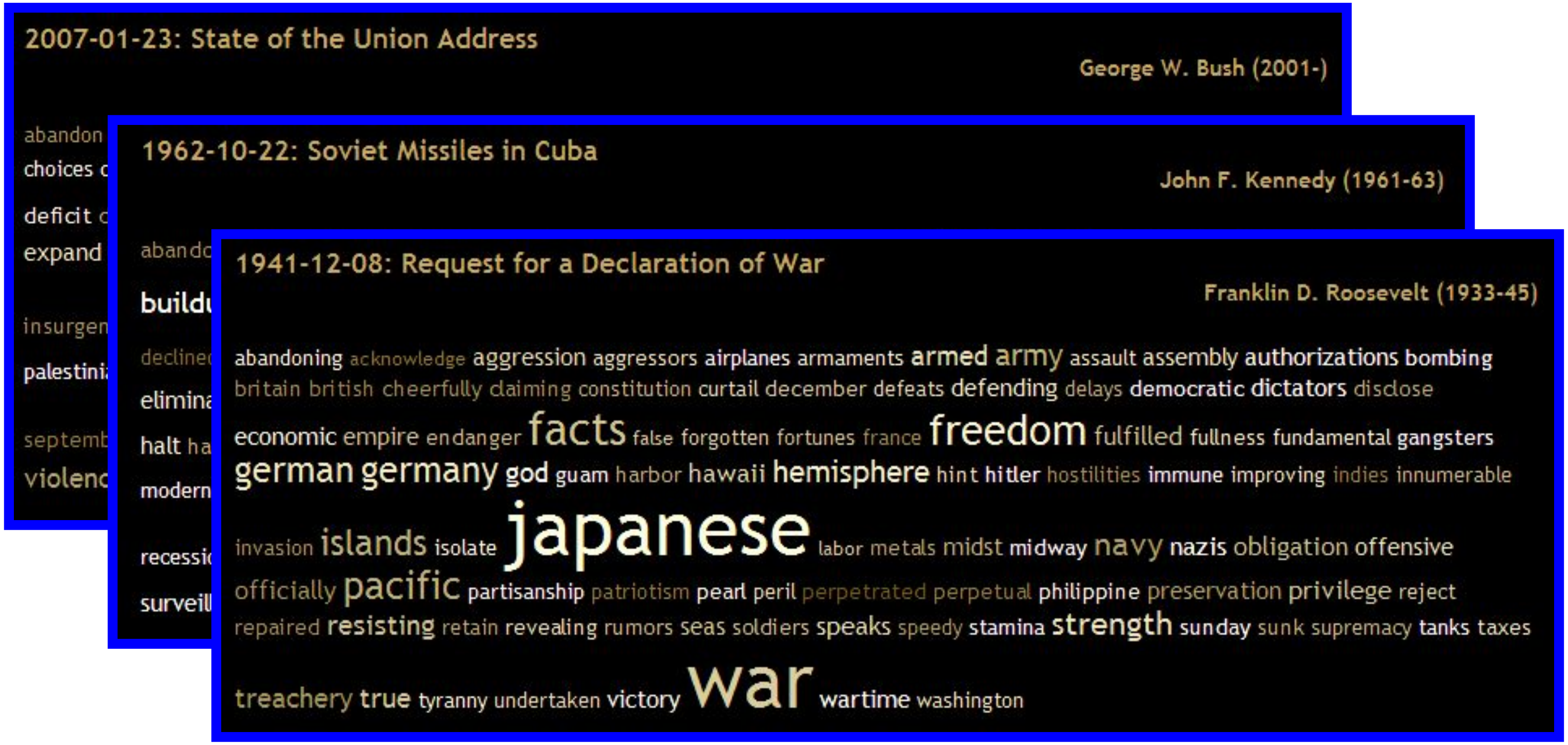
abandon achieving adversaries aggression agricultural appropriate armaments arms assessments atlantic ballistic berlin buildup burdens cargo college commitment communist constitution consumers cooperation crisis cuba dangers declined defensive deficit depended disarmament divisions domination doubled 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/`

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

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 c
deficit c
expand
insurgen
palestini
septemb
violenc

1962-10-22: Soviet Missiles in Cuba

John F. Kennedy (1961-63)

abando
buildu
decline
elimina
halt ha
modern
recessi
surveil

1941-12-08: Request for a Declaration of War

Franklin D. Roosevelt (1933-45)

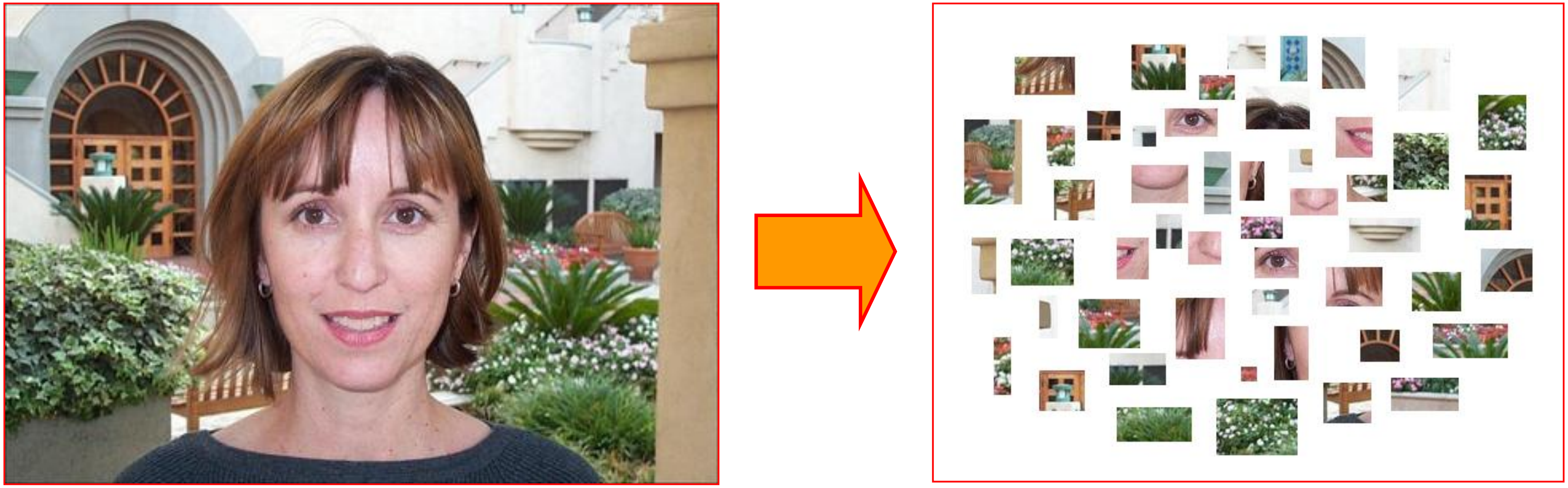
abandoning acknowledge aggression aggressors airplanes armaments armed army assault assembly authorizations bombing britain british cheerfully claiming constitution curtail december defeats defending delays democratic dictators disclose economic empire endanger facts false forgotten fortunes france freedom fulfilled fullness fundamental gangsters german germany god guam harbor hawaii hemisphere hint hitler hostilities immune improving indies innumerable invasion islands isolate japanese labor metals midst midway navy nazis obligation offensive 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 war wartime washington

US Presidential Speeches Tag Cloud

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

Adapted from slides by Juan Carlos Niebles and Ranjay Krishna

Bags of features for object recognition



face, flowers, building

- Works pretty well for image-level classification and for recognizing object *instances*

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

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
	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

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Bag of features

- First, take a bunch of images, extract features, and build up a “dictionary” or “visual vocabulary” – a list of common features
- Given a new image, extract features and build a histogram – for each feature, find the closest visual word in the dictionary

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Bag of features: outline

1. Extract features

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Bag of features: outline

1. Extract features
2. Learn "visual vocabulary"

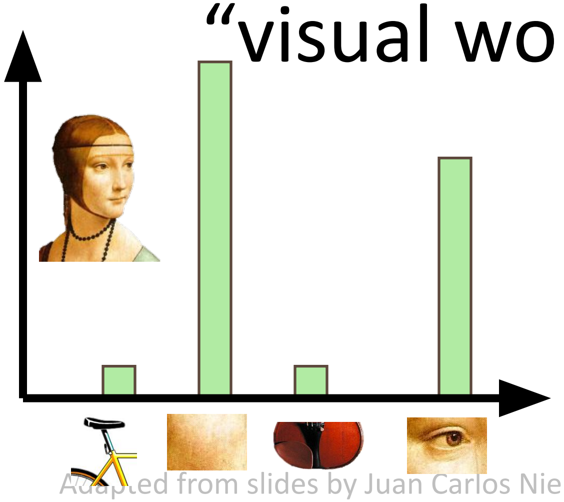
Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

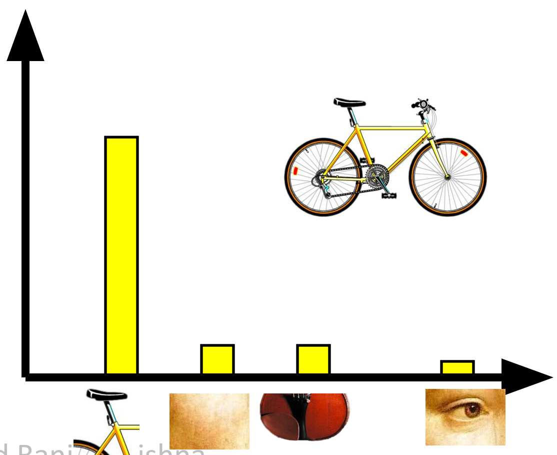
Bag of features: outline

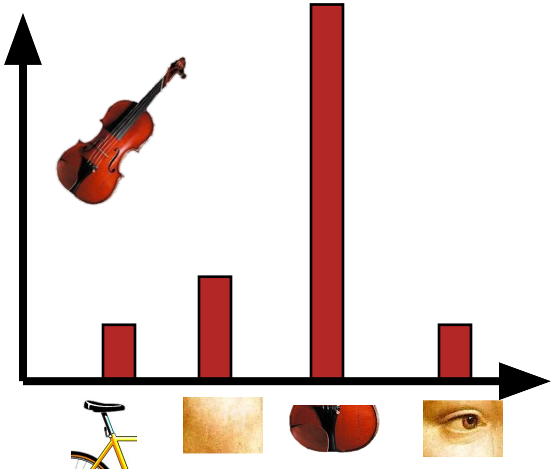
1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary

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"



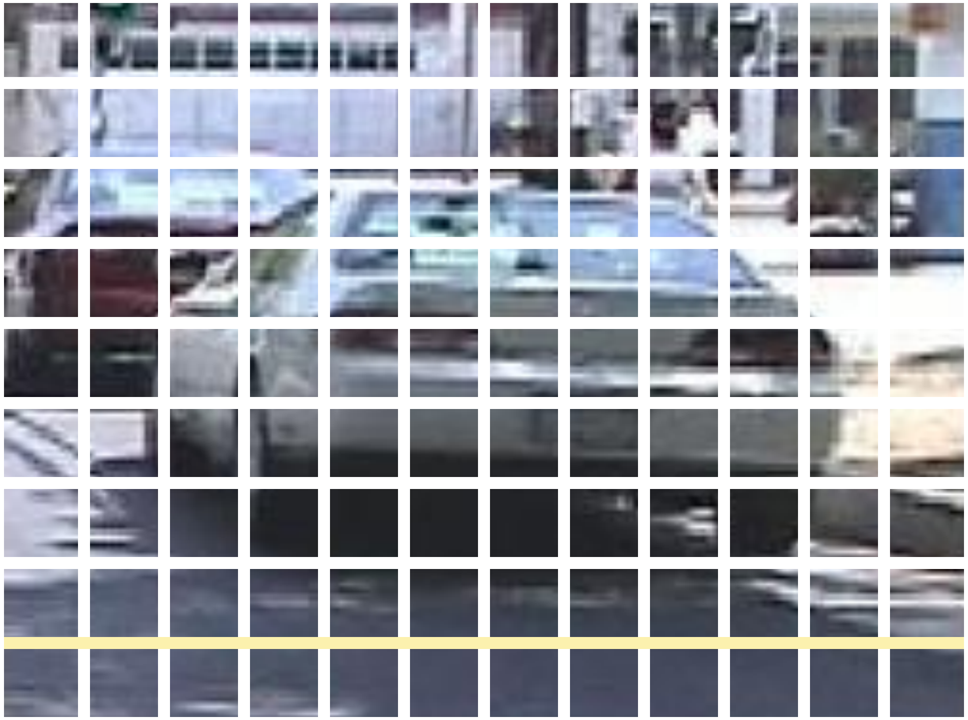




Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

1. Feature extraction

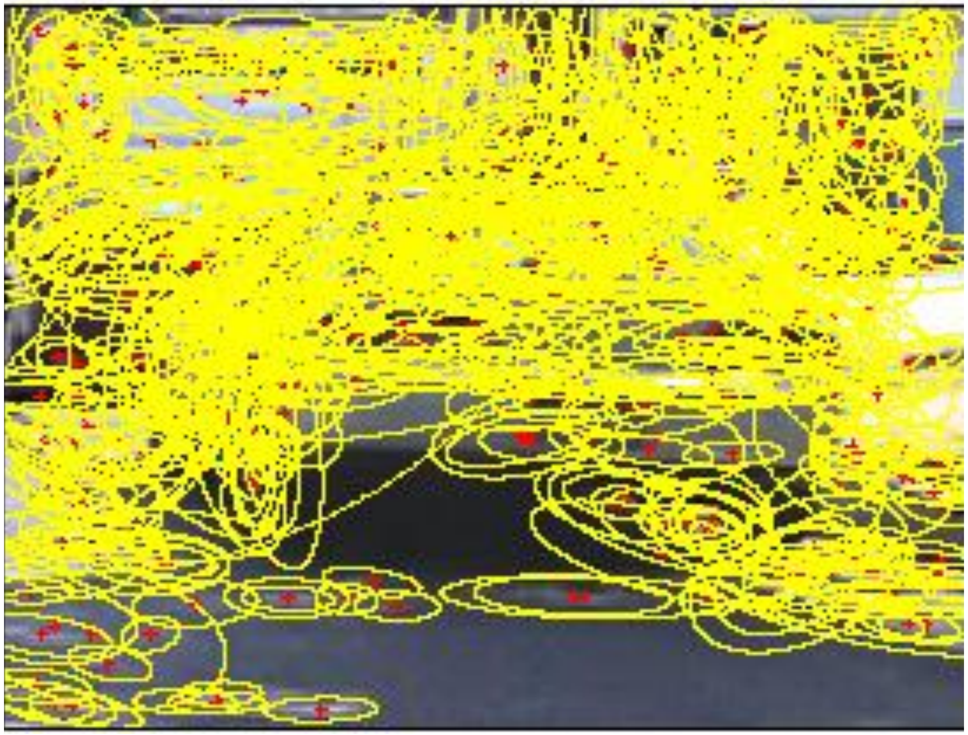
- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005



Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

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



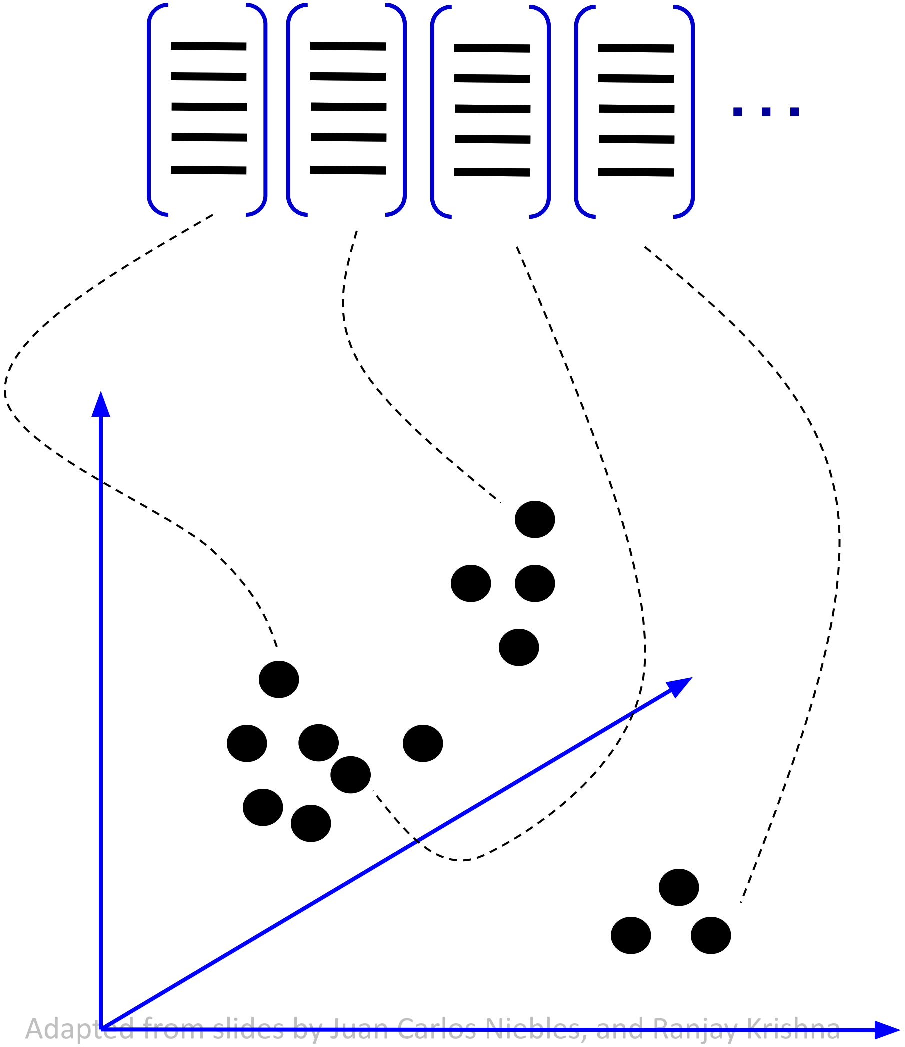
Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

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 et al. 2003)

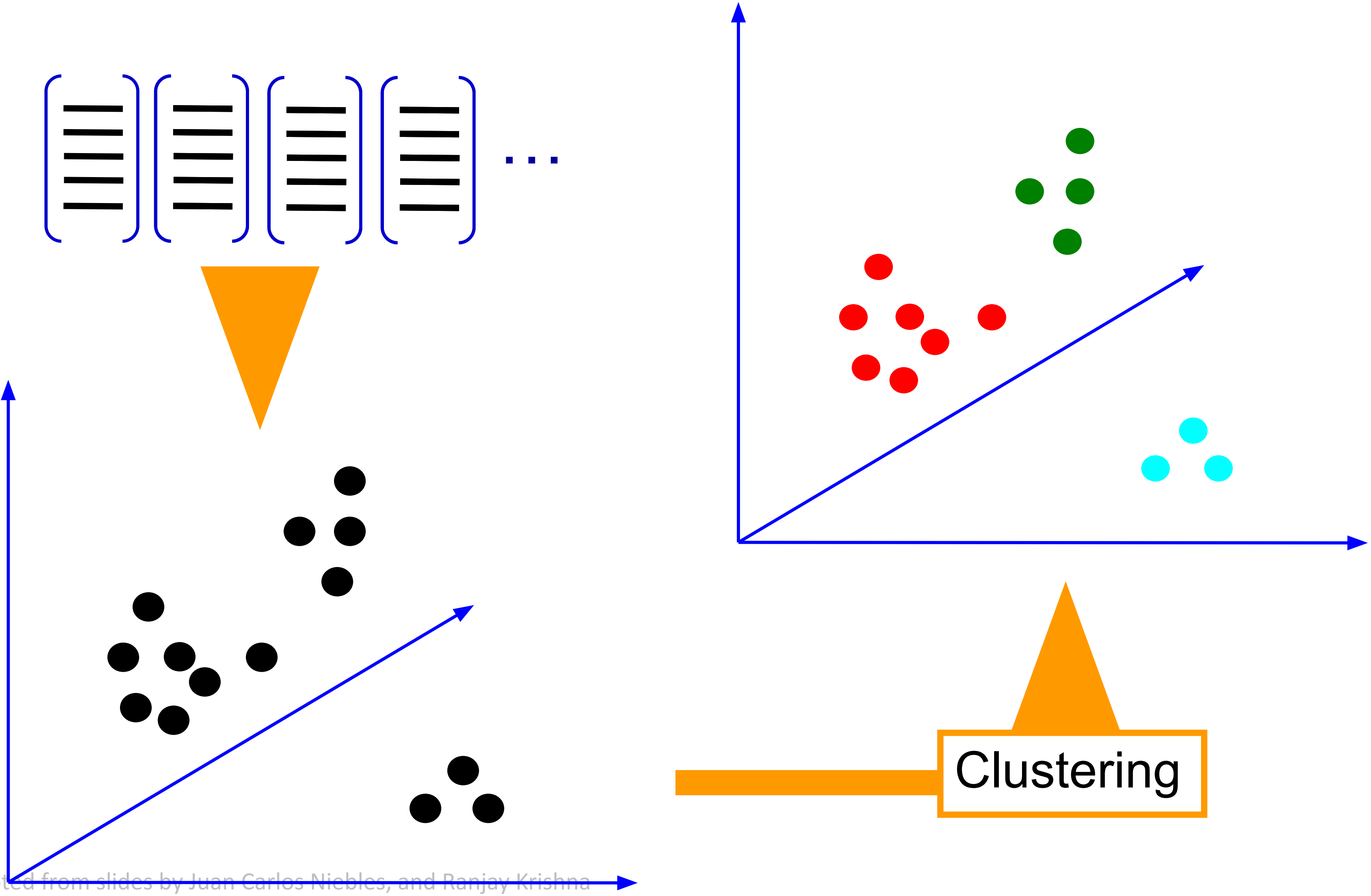
Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

2. Learning the visual vocabulary



Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

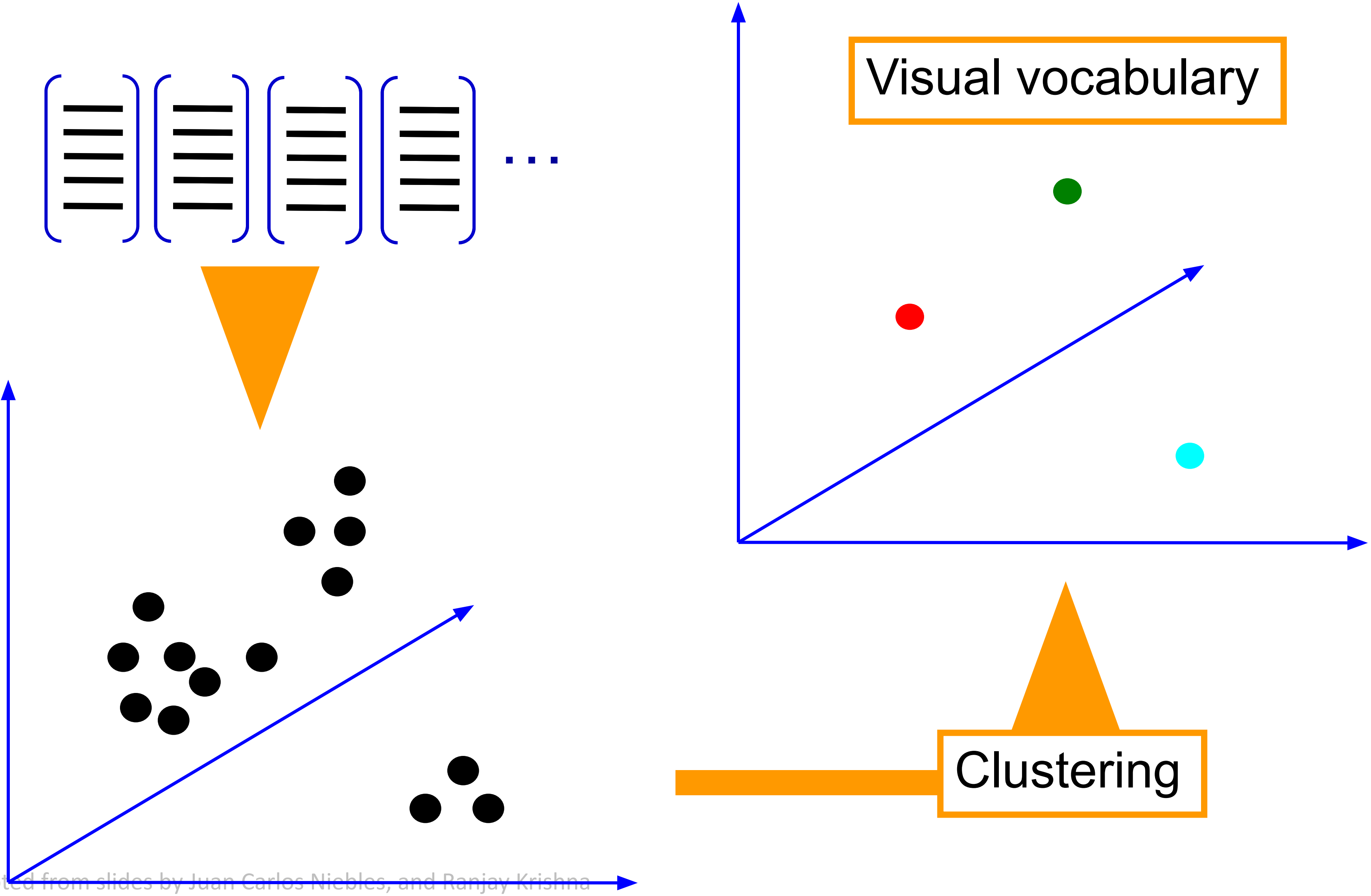
2. Learning the visual vocabulary



Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Slide credit: Josef Sivic

2. Learning the visual vocabulary



Visual vocabulary

Clustering

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Slide credit: Josef Sivic

K-means clustering recap

- 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

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

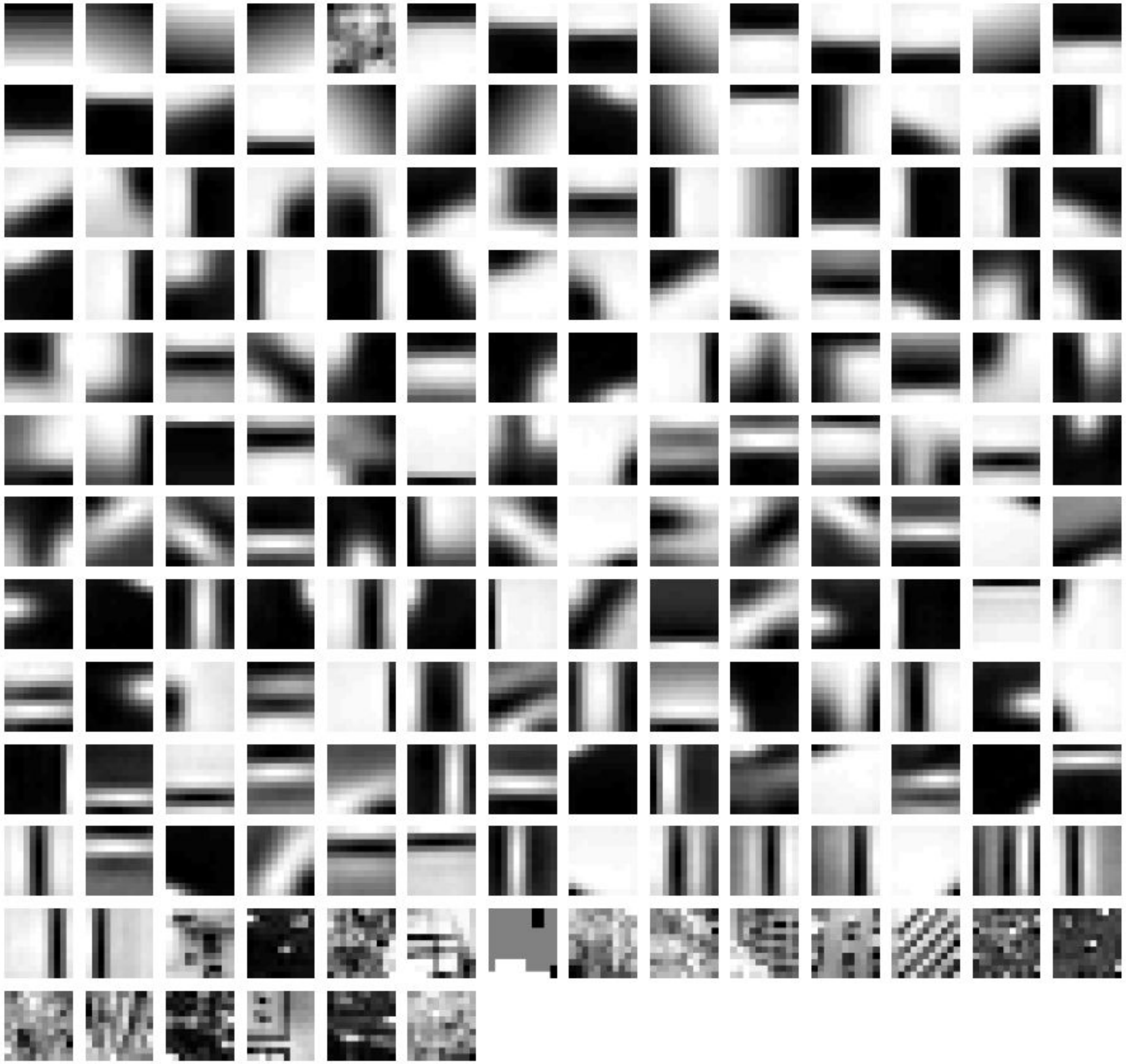
From clustering to vector quantization

- 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

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

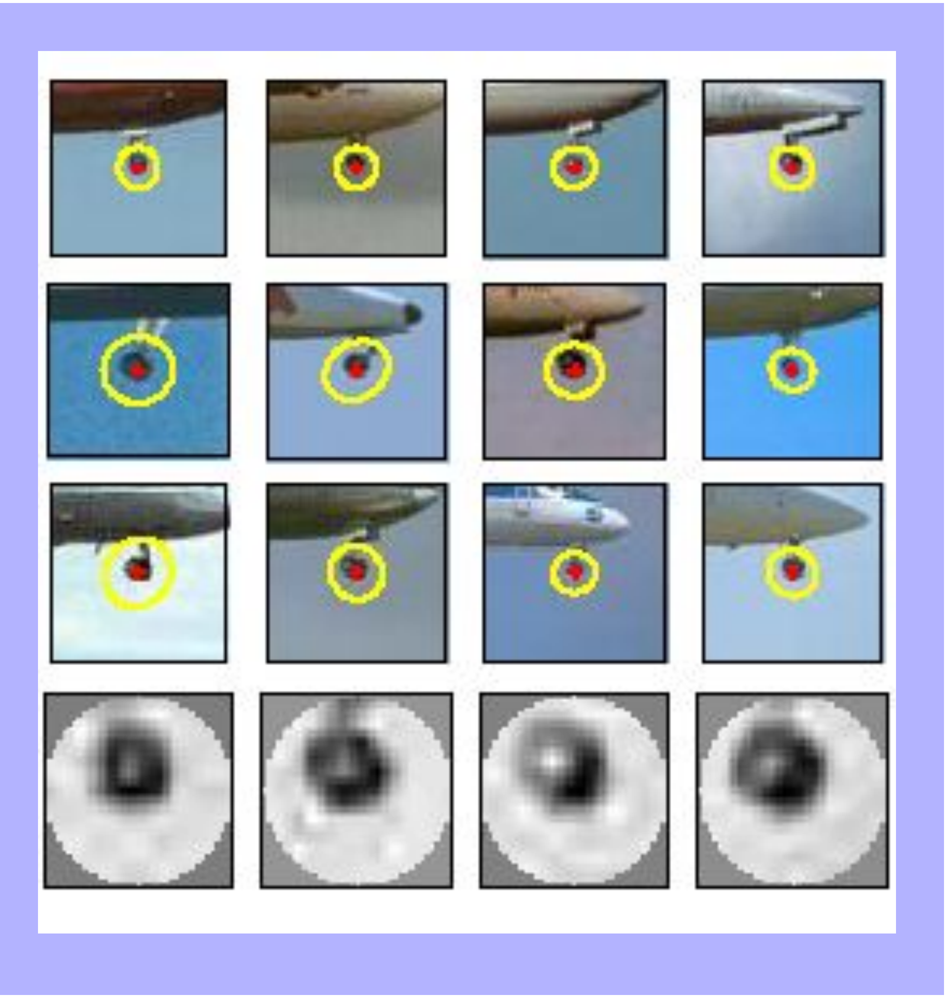
Example visual vocabulary

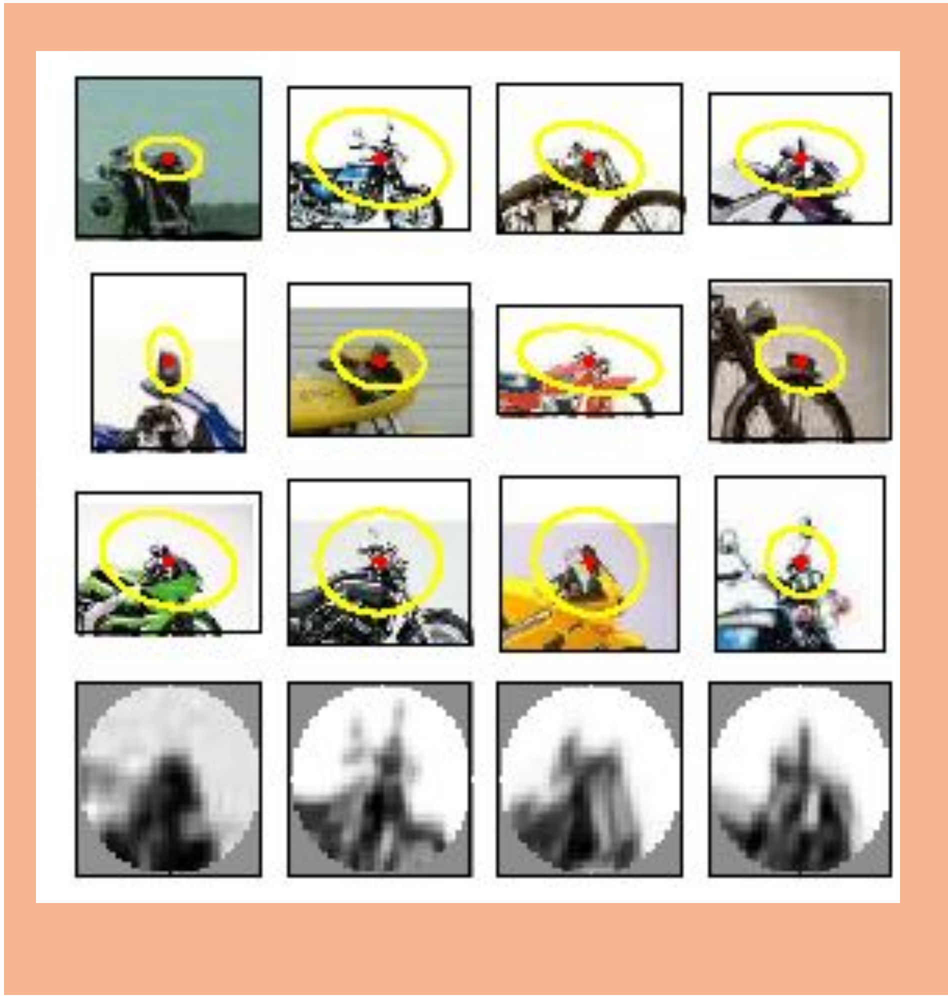


Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Fei-Fei et al. 2005

Image patch examples of visual words



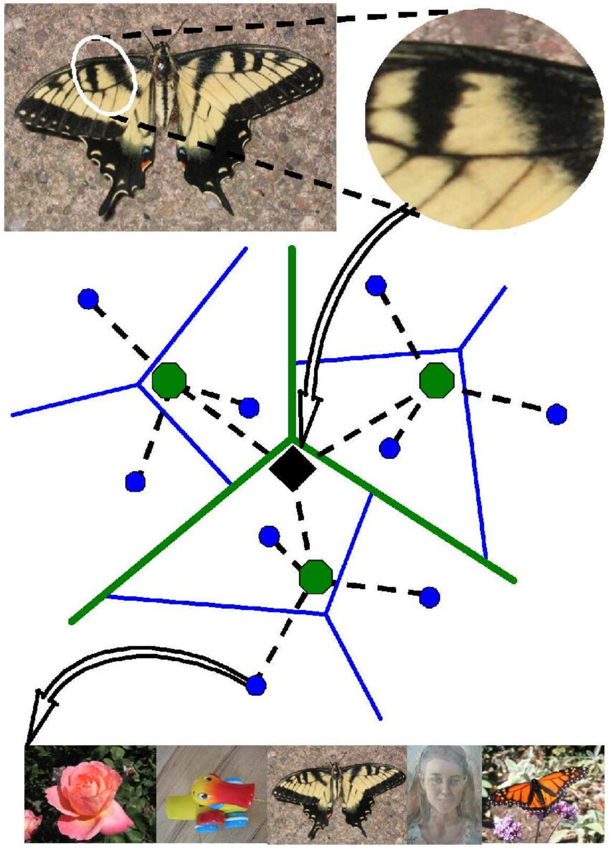


Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Sivic et al. 2005

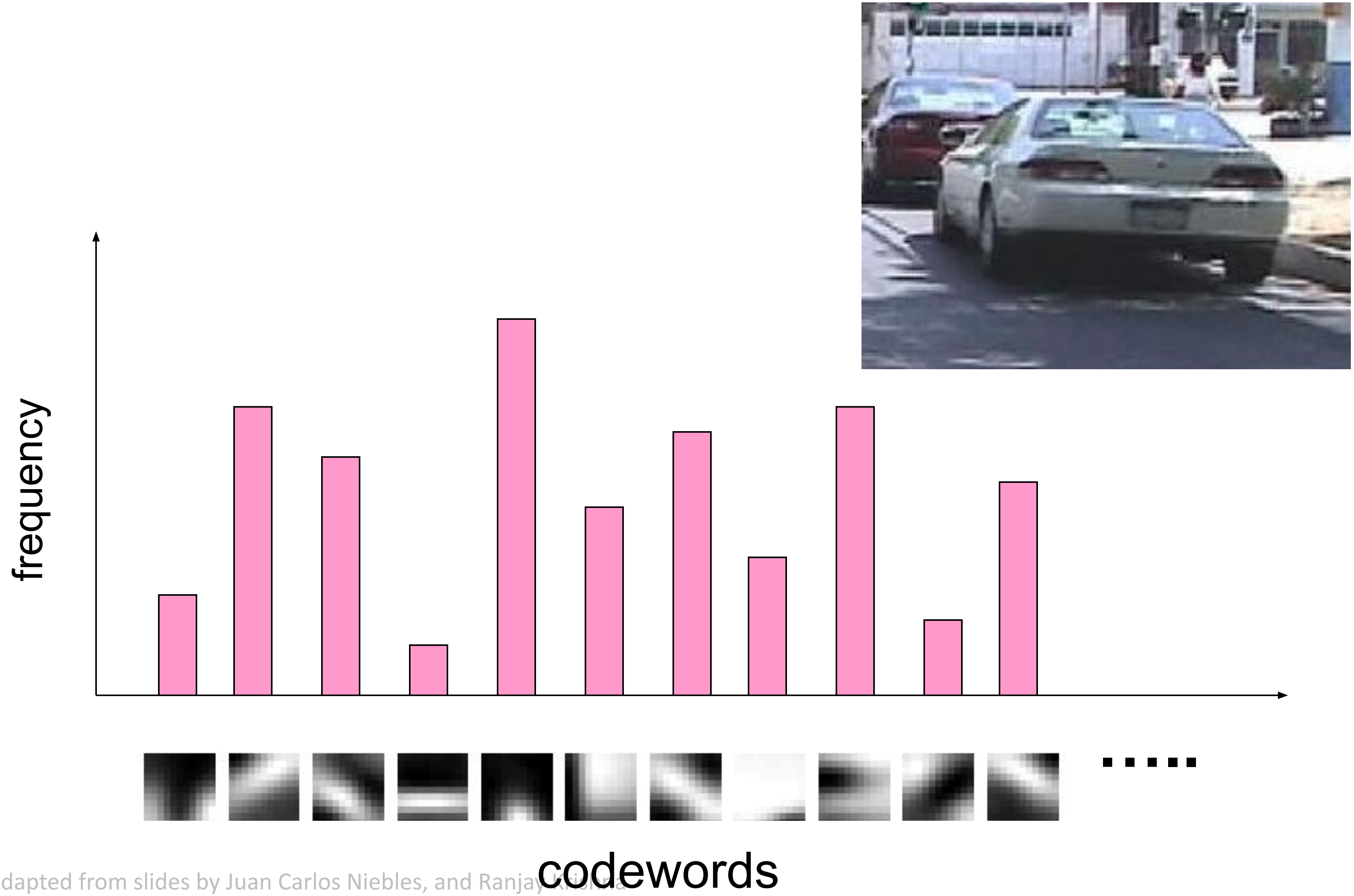
Visual vocabularies: Issues

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



Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

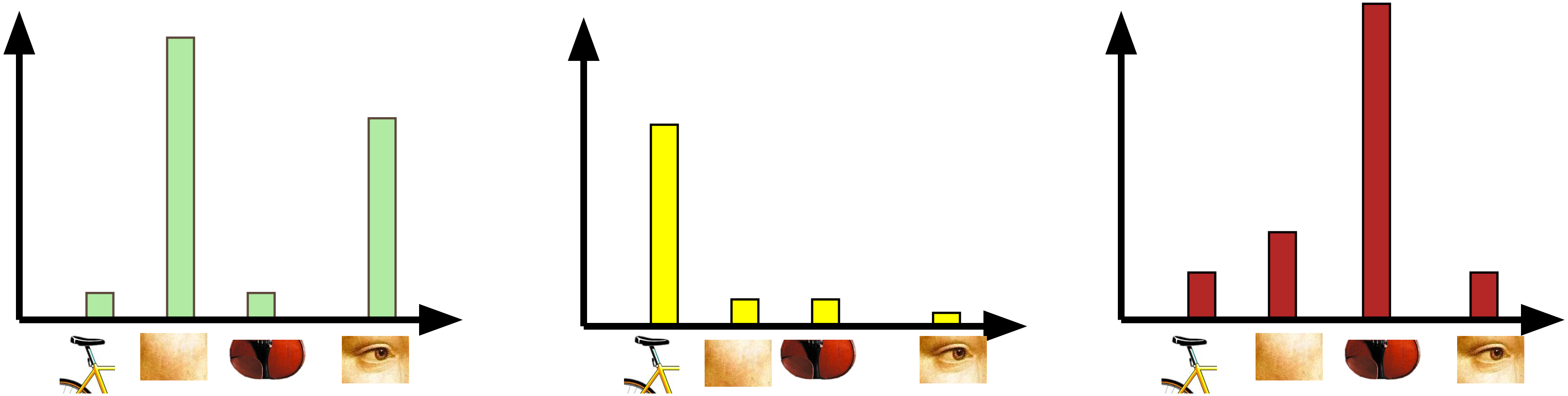
3. Image representation



Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Image classification

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



Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

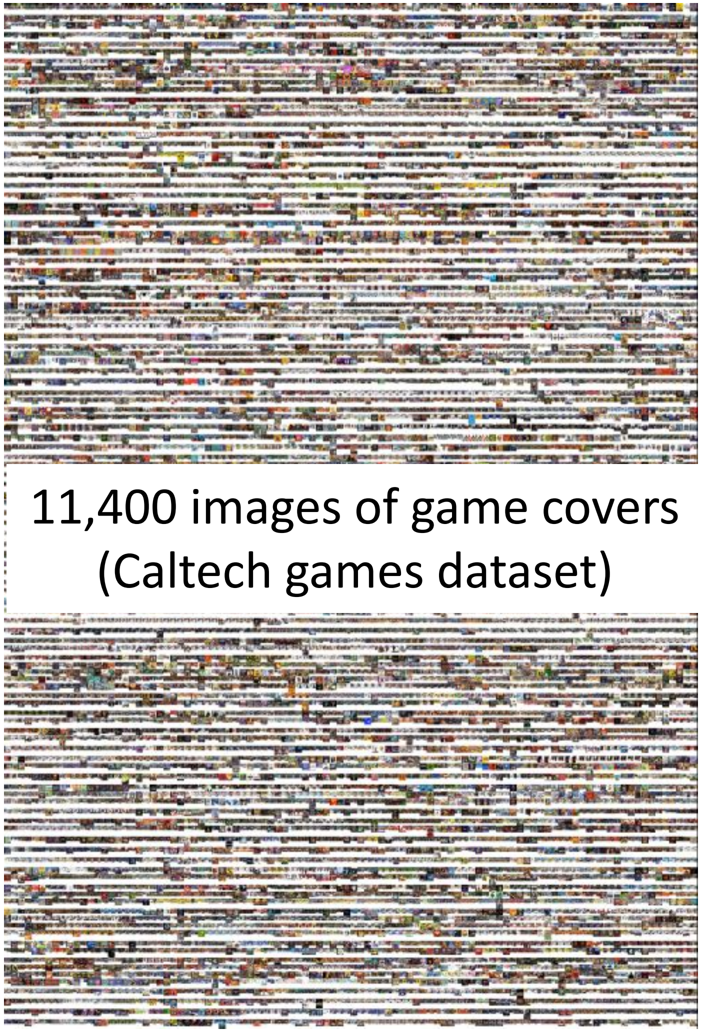
Uses of BoW representation

- Treat as feature vector for standard classifier
  - e.g k-nearest neighbors, support vector machine
- Cluster BoW vectors over image collection
  - Discover visual themes

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Large-scale image matching

- Bag-of-words models have been useful in matching an image to a large database of object *instances*



11,400 images of game covers (Caltech games dataset)

how do I find this image in the database?

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Large-scale image search

Build the database:

- Extract features from the database images
- Learn a vocabulary using k-means (typical k: 100,000)
- Compute *weights* for each word
- Create an inverted file mapping words □ images

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Weighting the words

- Just as with text, some visual words are more discriminative than others

the, and, or vs. ***cow, AT&T, Cher***

- the bigger fraction of the documents a word appears in, the less useful it is for matching
  - e.g., a word that appears in *all* documents is not helping us

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

TF-IDF weighting

- Instead of computing a regular histogram distance, we'll weight each word by it's *inverse document frequency*

- inverse document frequency (IDF) of word *j* =

$$\log \frac{\text{number of documents}}{\text{number of documents in which } j \text{ appears}}$$

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

TF-IDF weighting

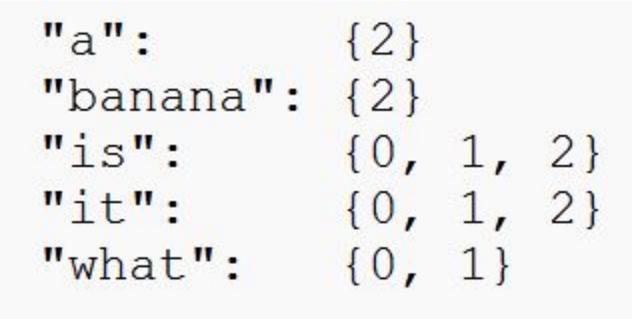
- To compute the value of bin *j* in image *I*:

term frequency of *j* in *I* **x** *inverse document frequency* of *j*

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Inverted file

- Each image has ~1,000 features
- We have ~100,000 visual words
  - Each histogram is extremely sparse (mostly zeros)
- Inverted file
  - mapping from words to documents



```
"a":      {2}
"banana": {2}
"is":     {0, 1, 2}
"it":     {0, 1, 2}
"what":   {0, 1}
```

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Inverted file

- Can quickly use the inverted file to compute similarity between a new image and all the images in the database
  - Only consider database images whose bins overlap the query image

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

Large-scale image search

query image

top 6 results

- Cons:
  - performance degrades as the database grows

Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

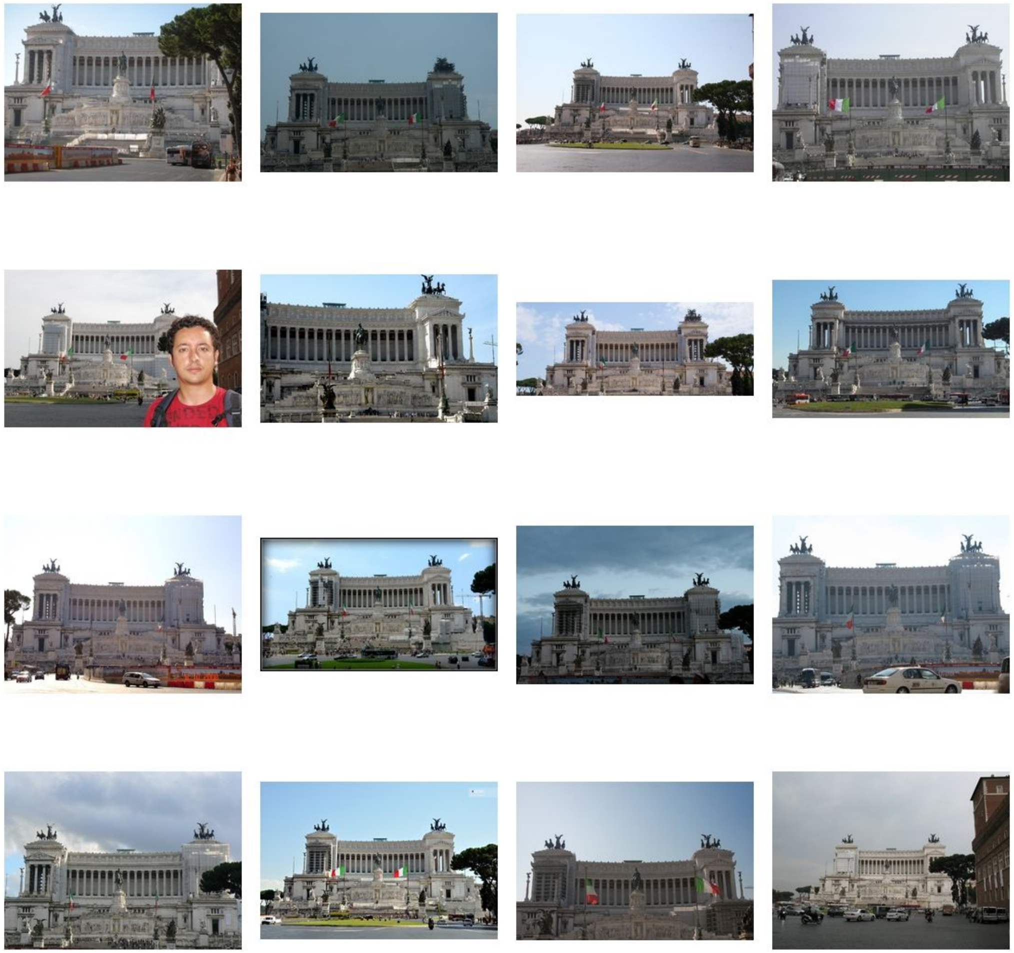
Large-scale image search

- Pros:
  - Works well for CD covers, movie posters
  - Real-time performance possible

real-time retrieval from a database of 40,000 CD covers

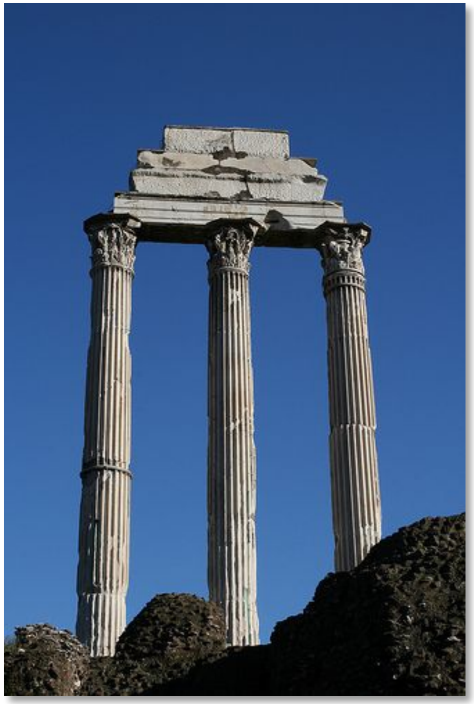
Nister & Stewenius, **Scalable Recognition with a Vocabulary Tree**

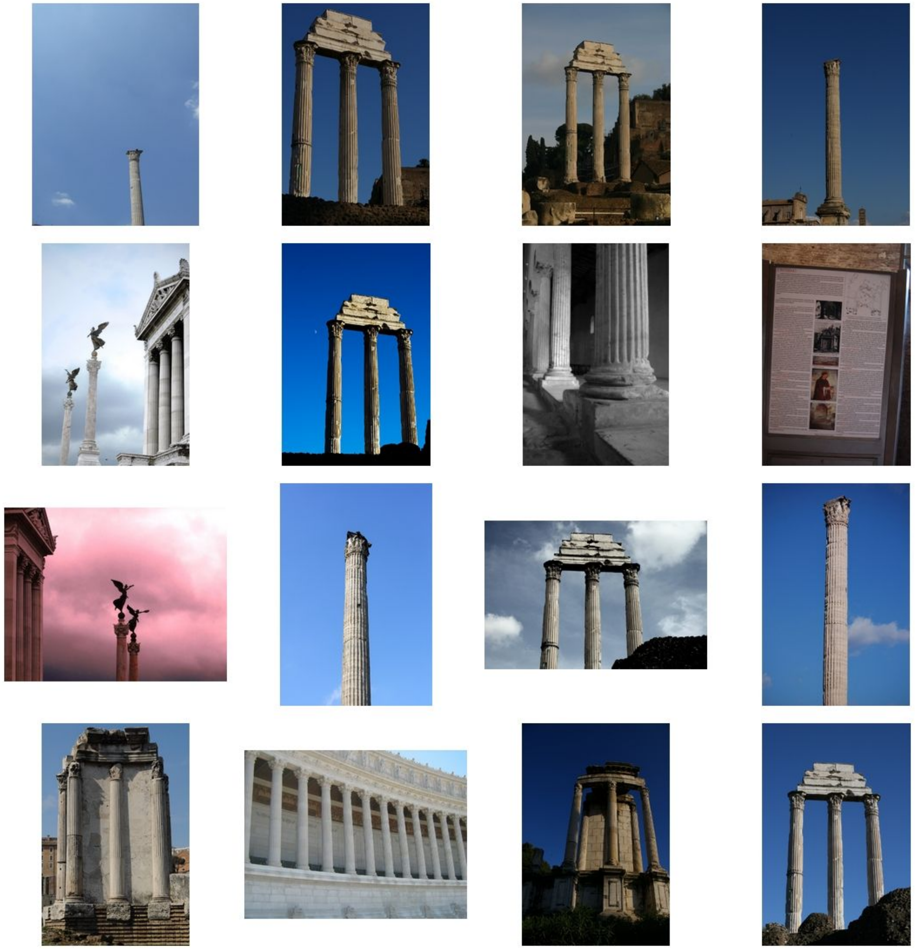
Example bag-of-words matches



Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

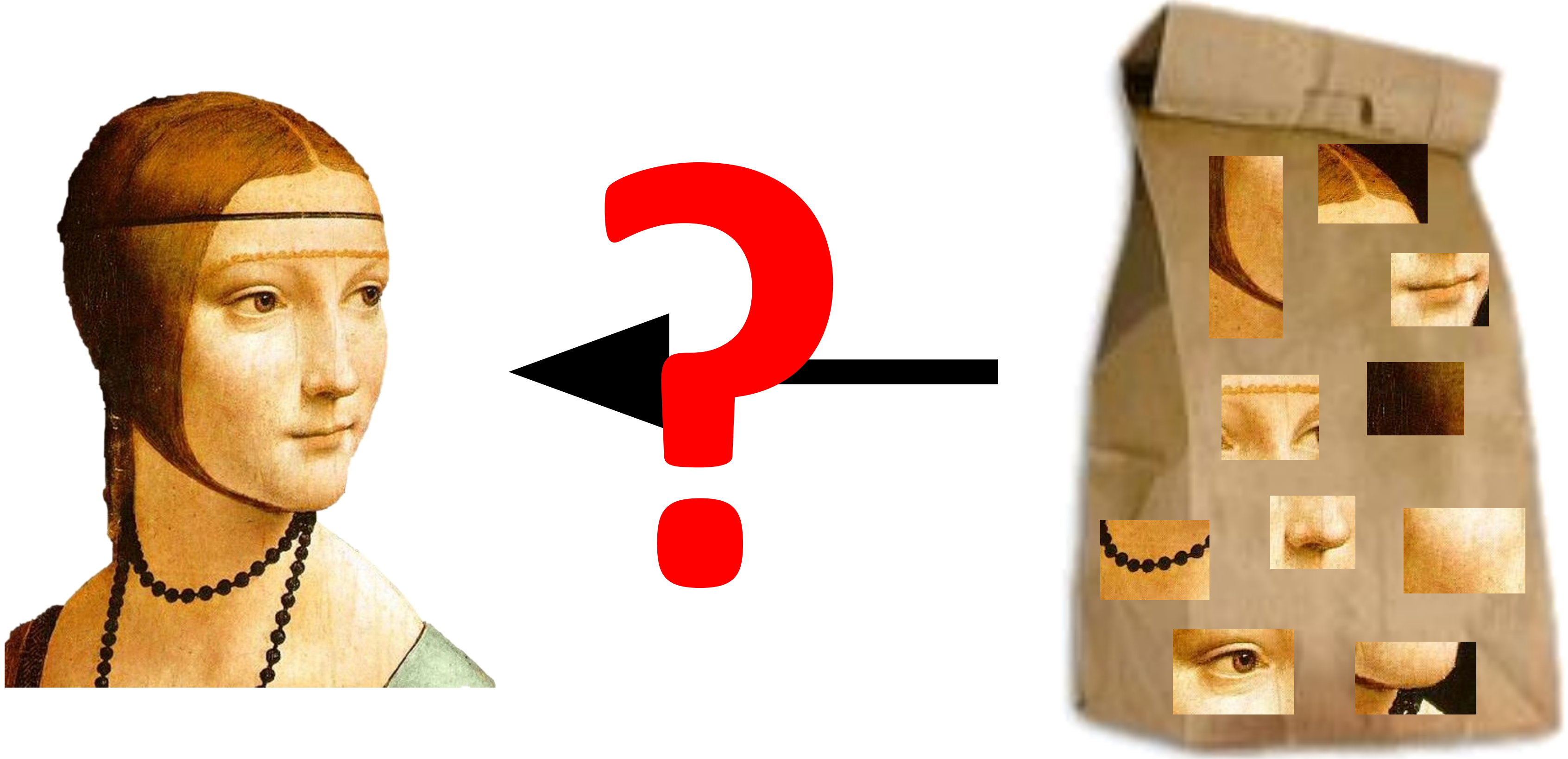
Example bag-of-words matches





Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna

What about spatial info?



Adapted from slides by Juan Carlos Niebles, and Ranjay Krishna