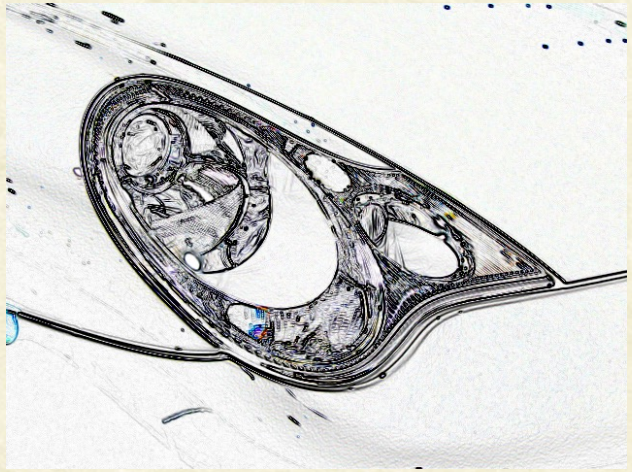
CS7.505: Computer Vision

Spring 2022: Descriptors and Point Matching


Anoop M. Namboodiri

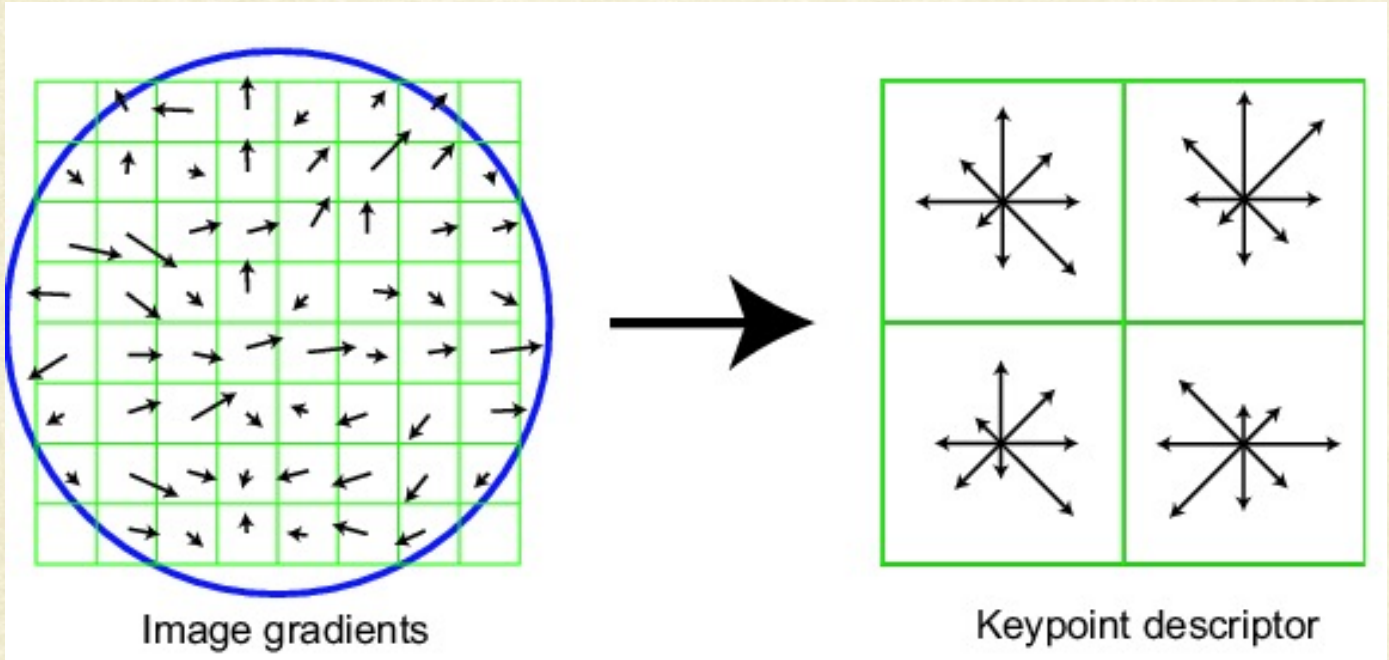
Biometrics and Secure ID Lab, CVIT,

IIIT Hyderabad

Recap: SIFT Descriptor

Full version

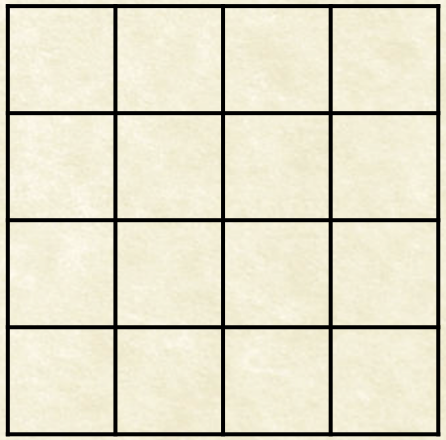
- Divide the 16x16 window into a 4x4 grid of cells (2x2 case shown below)
- Compute an orientation histogram for each cell (Use gaussian weighting)
- 16 cells * 8 orientations = 128 dimensional descriptor



Adapted from slide by David Lowe

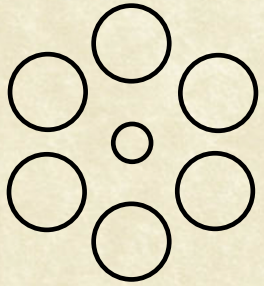
Other methods: Daisy

SIFT

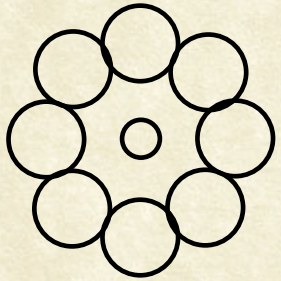


Circular gradient binning

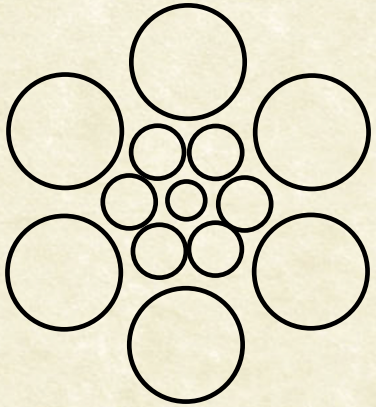
Daisy



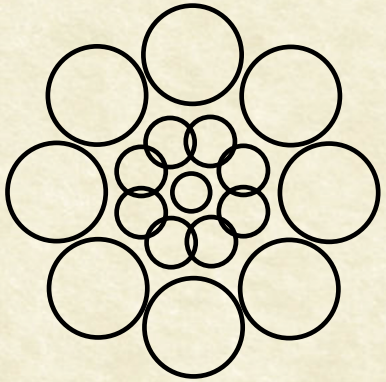
1 Ring 6 Segments



1 Ring 8 Segments



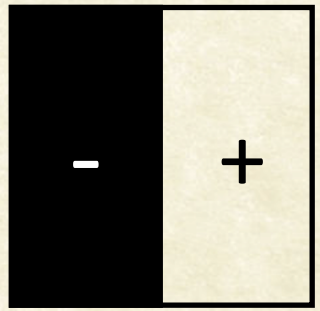
2 Rings 6 Segments



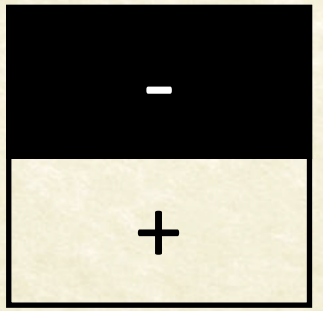
2 Rings 8 Segments

Picking the best DAISY, S. Winder, G. Hua, M. Brown, CVPR 09

Recap: SURF Descriptor



Response in x



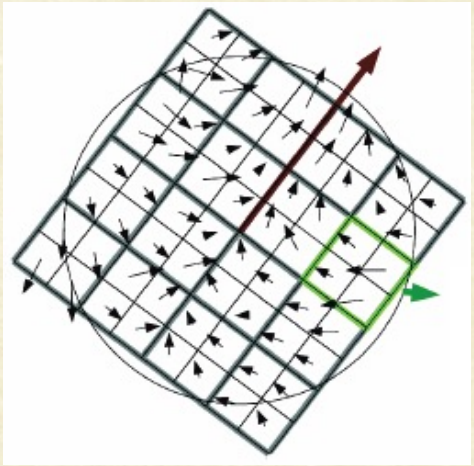
Response in y

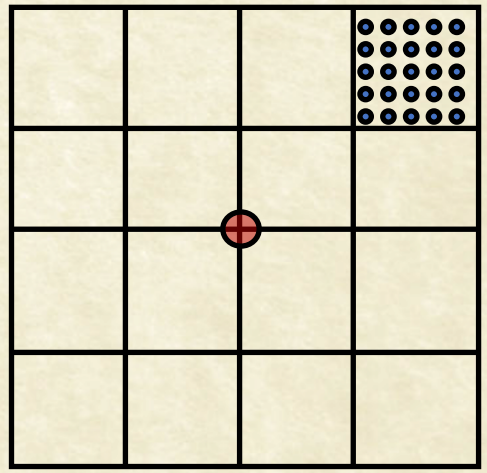
3-Step Process

1. At each interest point, place an oriented window with 4x4 (16 sub-windows), with each sub-window having 5x5 locations
2. Weight the Haar wavelet outputs (***dx*** and ***dy***) using a Gaussian Kernel
3. Within each sub-window, compute:

$$v_{subregion} = \left[\sum dx, \sum dy, \sum |dx|, \sum |dy|\right]$$

- This yields a 64-element descriptor





Other methods: BRIEF

Randomly sample pair of pixels a and b.
1 if a > b, else 0. Store binary vector.

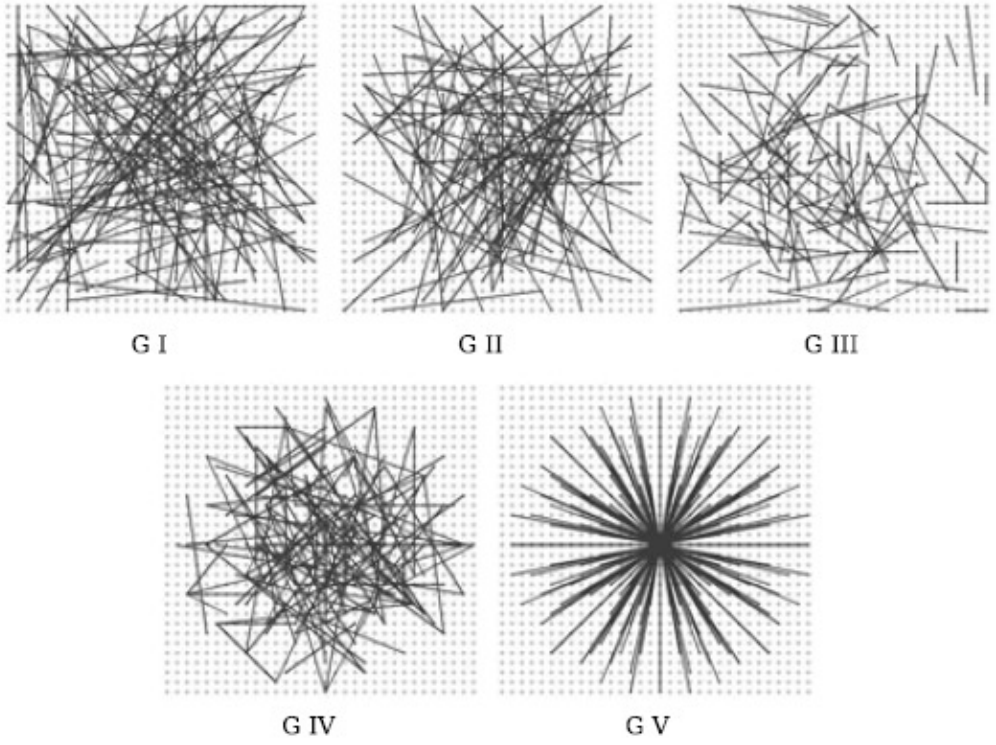
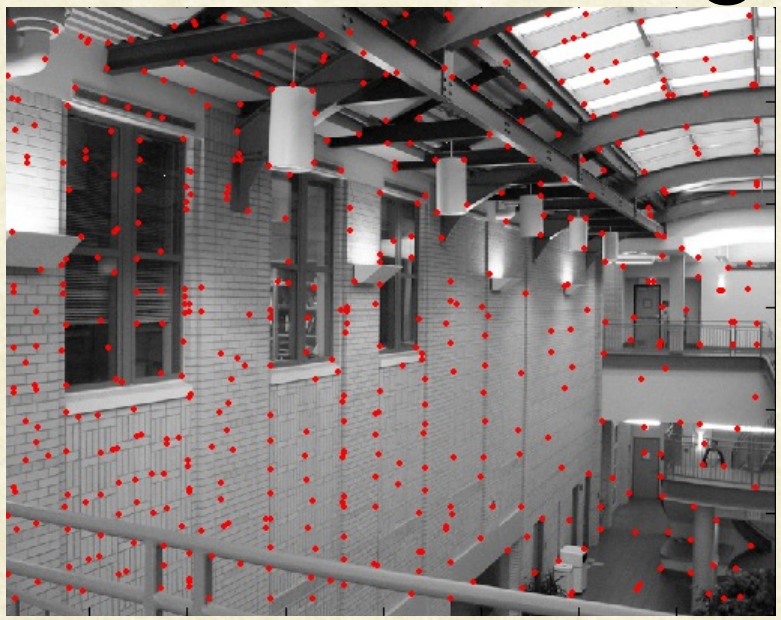
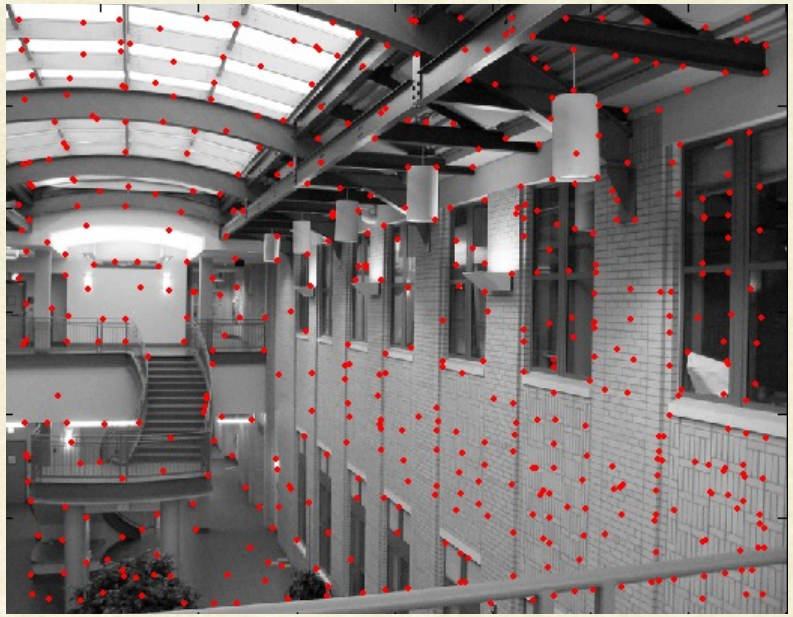


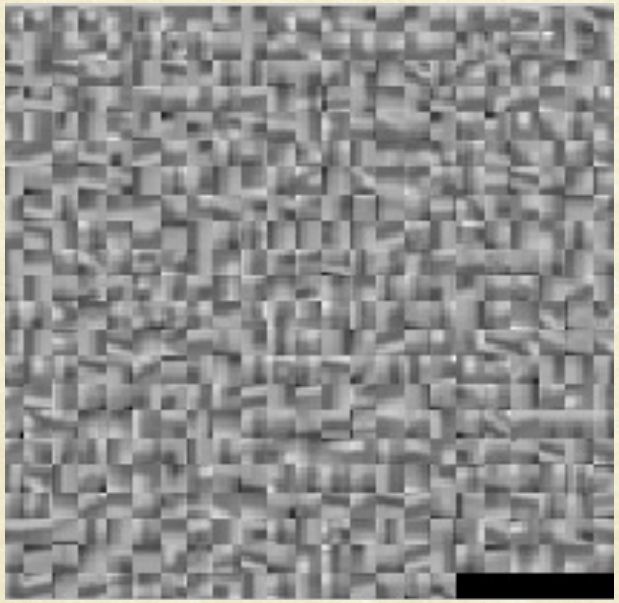
Fig. 2. Different approaches to choosing the test locations. All except the righmost one are selected by random sampling. Showing 128 tests in every image.

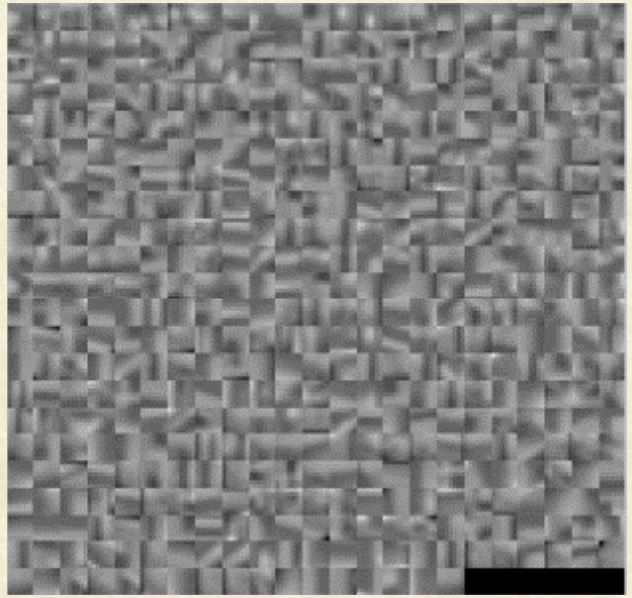
BRIEF: binary robust independent elementary features, Calonder, V Lepetit, C Strecha, ECCV 2010

Feature matching

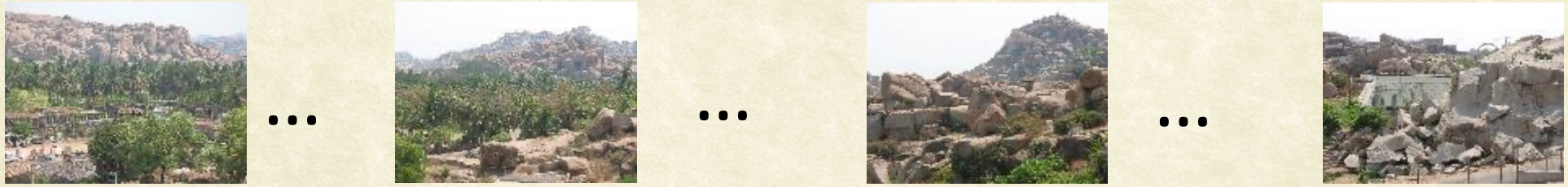








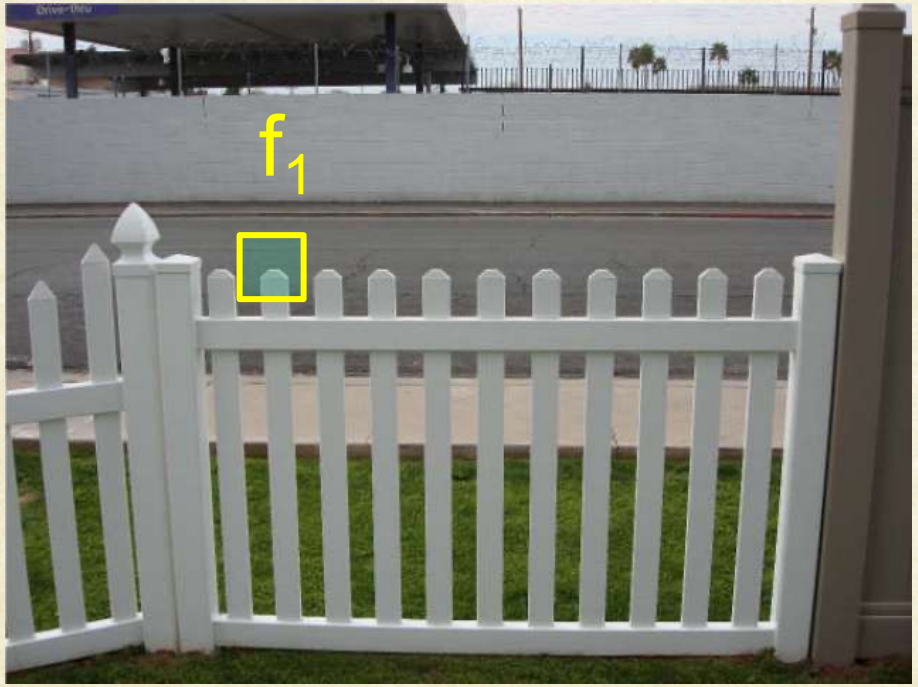
Mosaicing Example



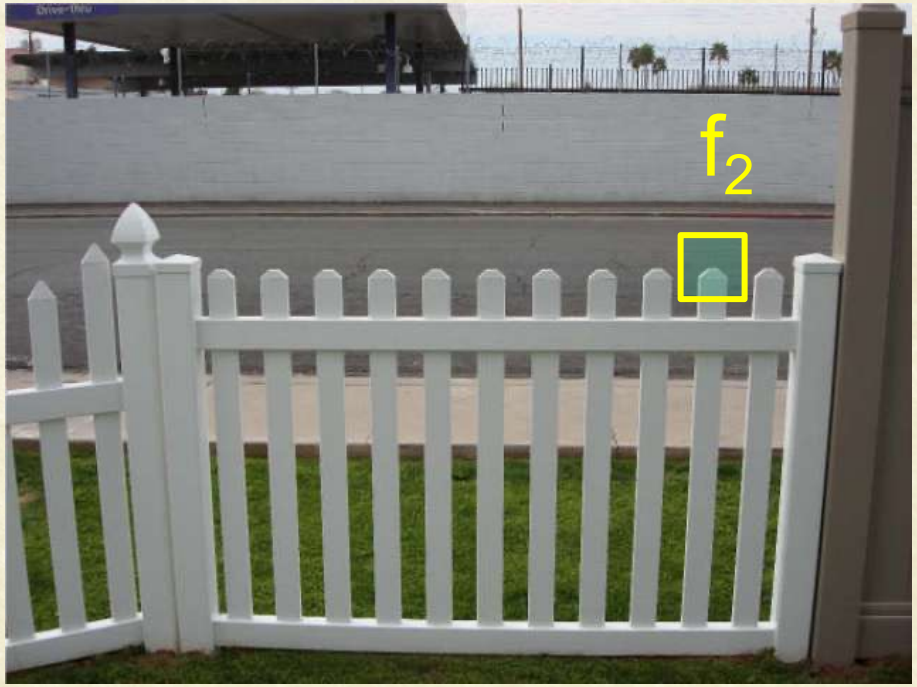
Computing Feature Distance

How to define the difference between two features f_1, f_2?

- Simple approach is SSD(f_1, f_2)
  - Sum of square differences (SSD) between entries of the two descriptors
  - Can give good scores to very ambiguous (bad) matches



I_1

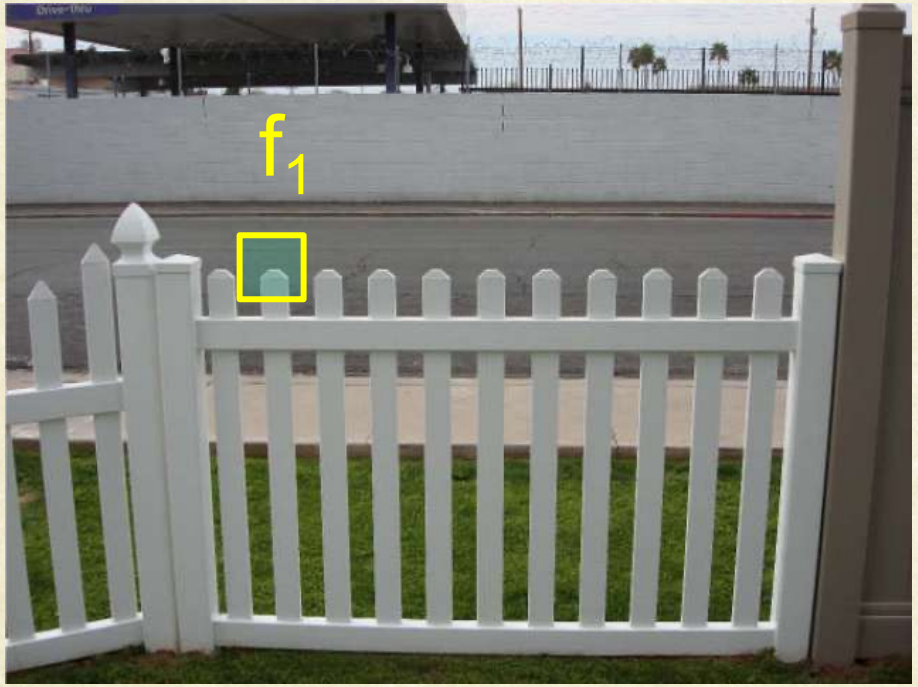


I_2

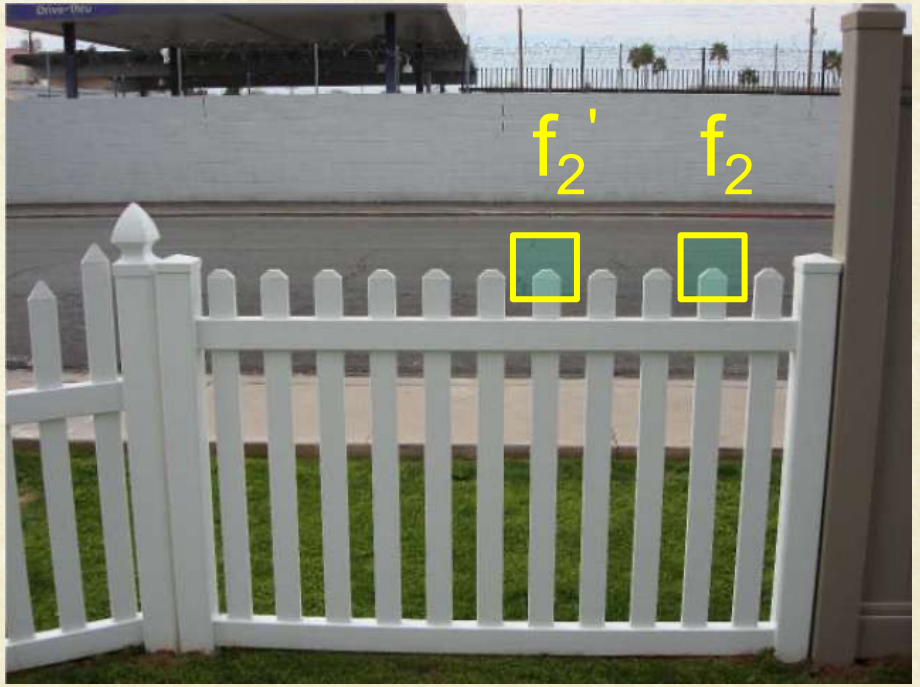
Computing Feature Distance

How to define the difference between two features f_1, f_2?

- Better approach: Ratio of Distances = SSD(f_1, f_2) / SSD(f_1, f_2')
  - f_2 is best SSD match to f_1 in I_2 ; f_2' is 2nd best SSD match to f_1 in I_2
  - Gives large values (~1) for ambiguous matches

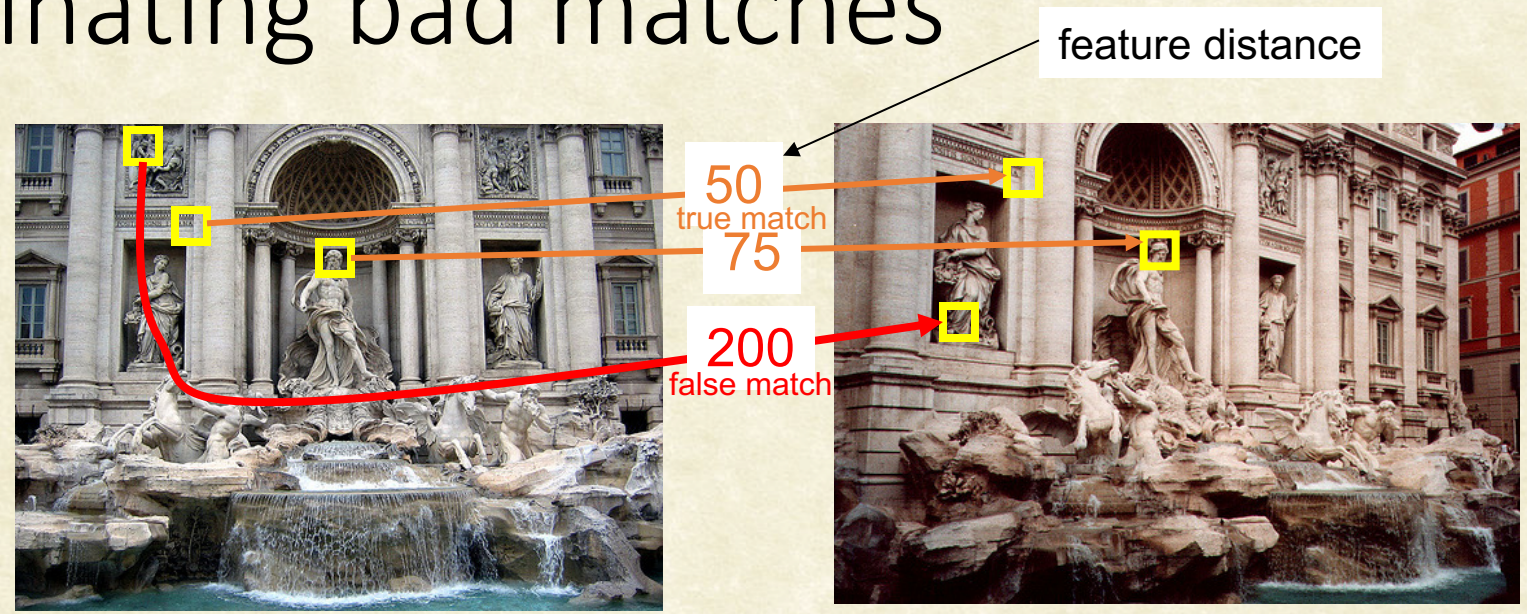


I_1



I_2

Eliminating bad matches



Throw out features with distance > threshold

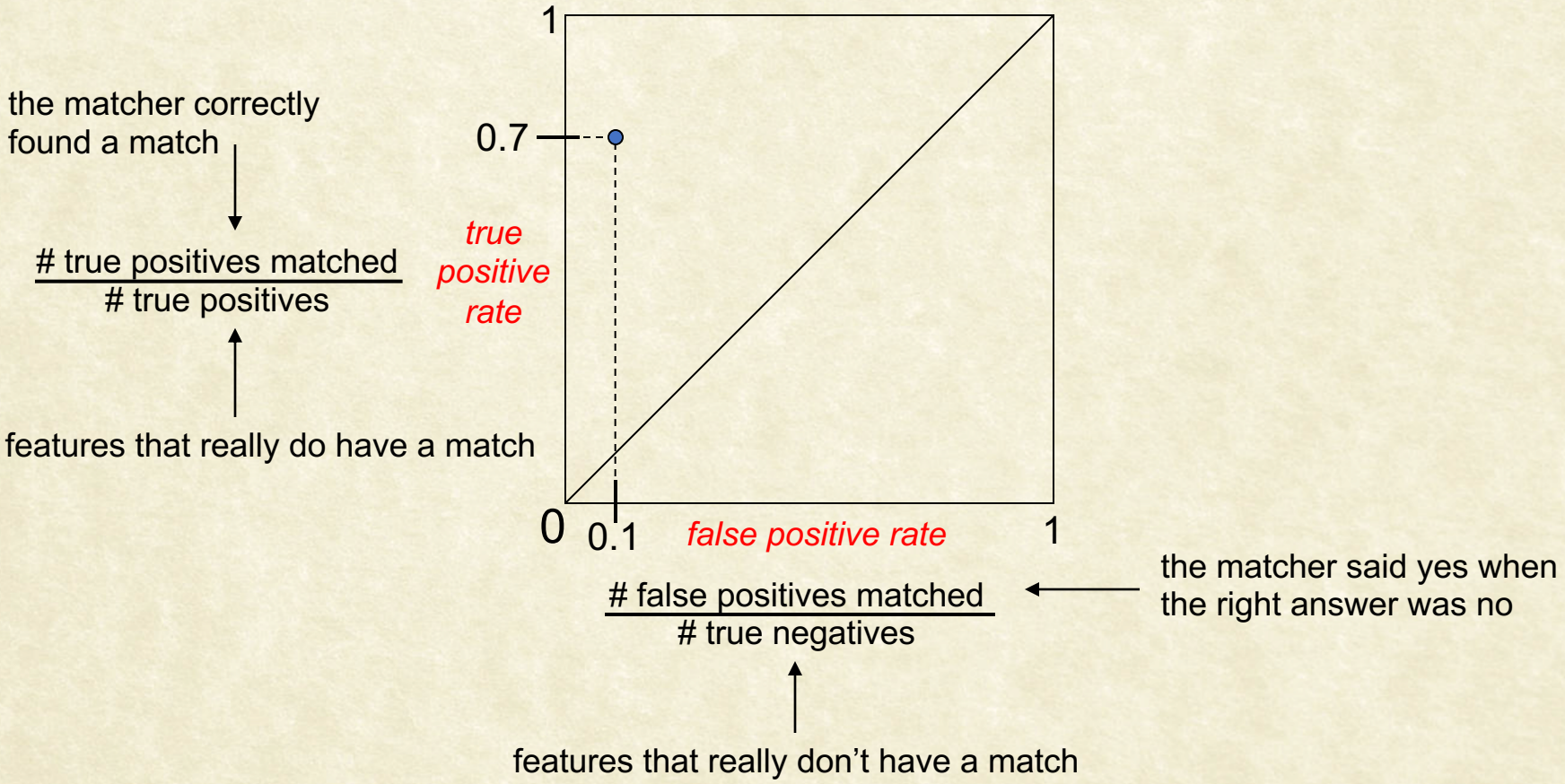
- How to choose the threshold?

The distance threshold affects performance

- True positives = # of detected matches that are correct
  - Suppose we want to maximize these—how to choose threshold?
- False positives = # of detected matches that are incorrect
  - Suppose we want to minimize these—how to choose threshold?

Evaluating the results

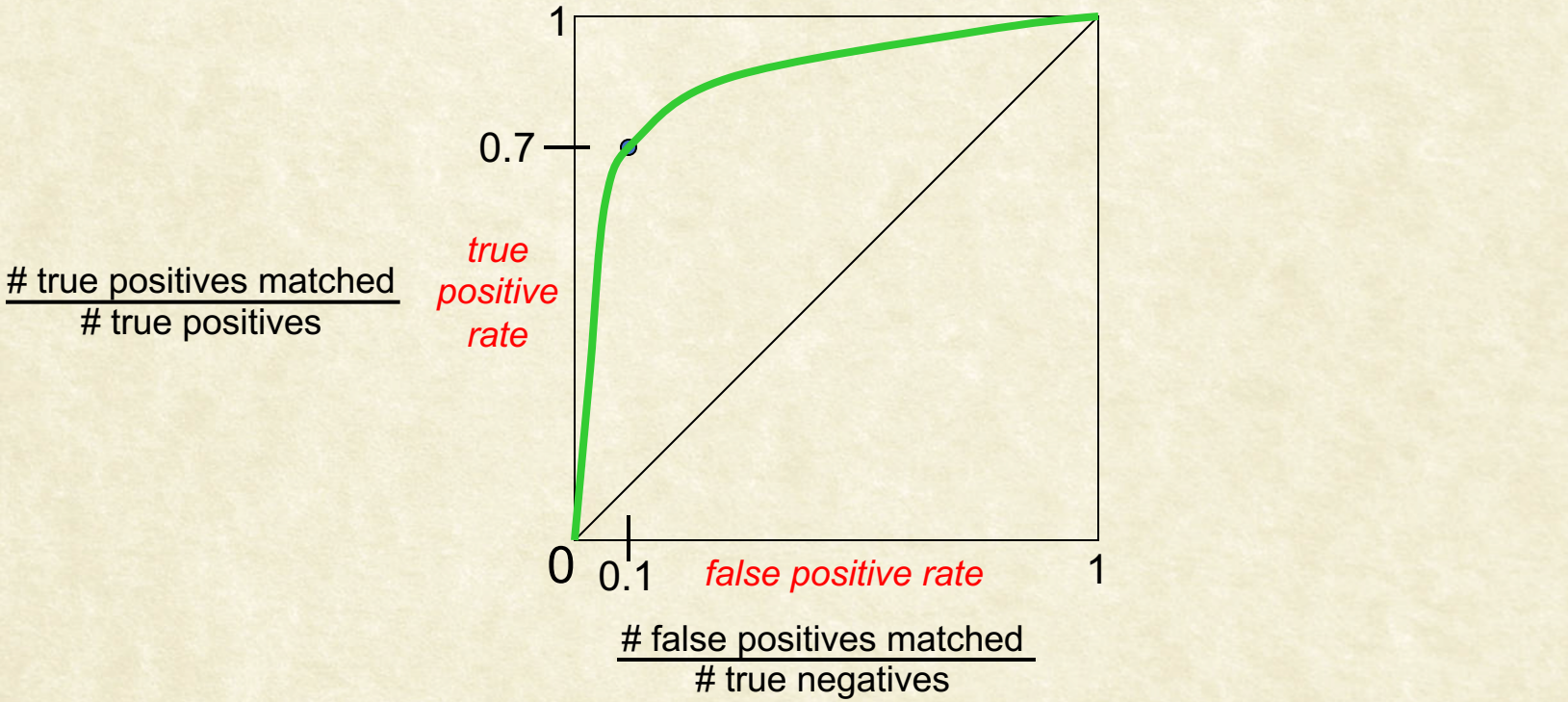
How can we measure the performance of a feature matcher?



Evaluating the results

How can we measure the performance of a feature matcher?

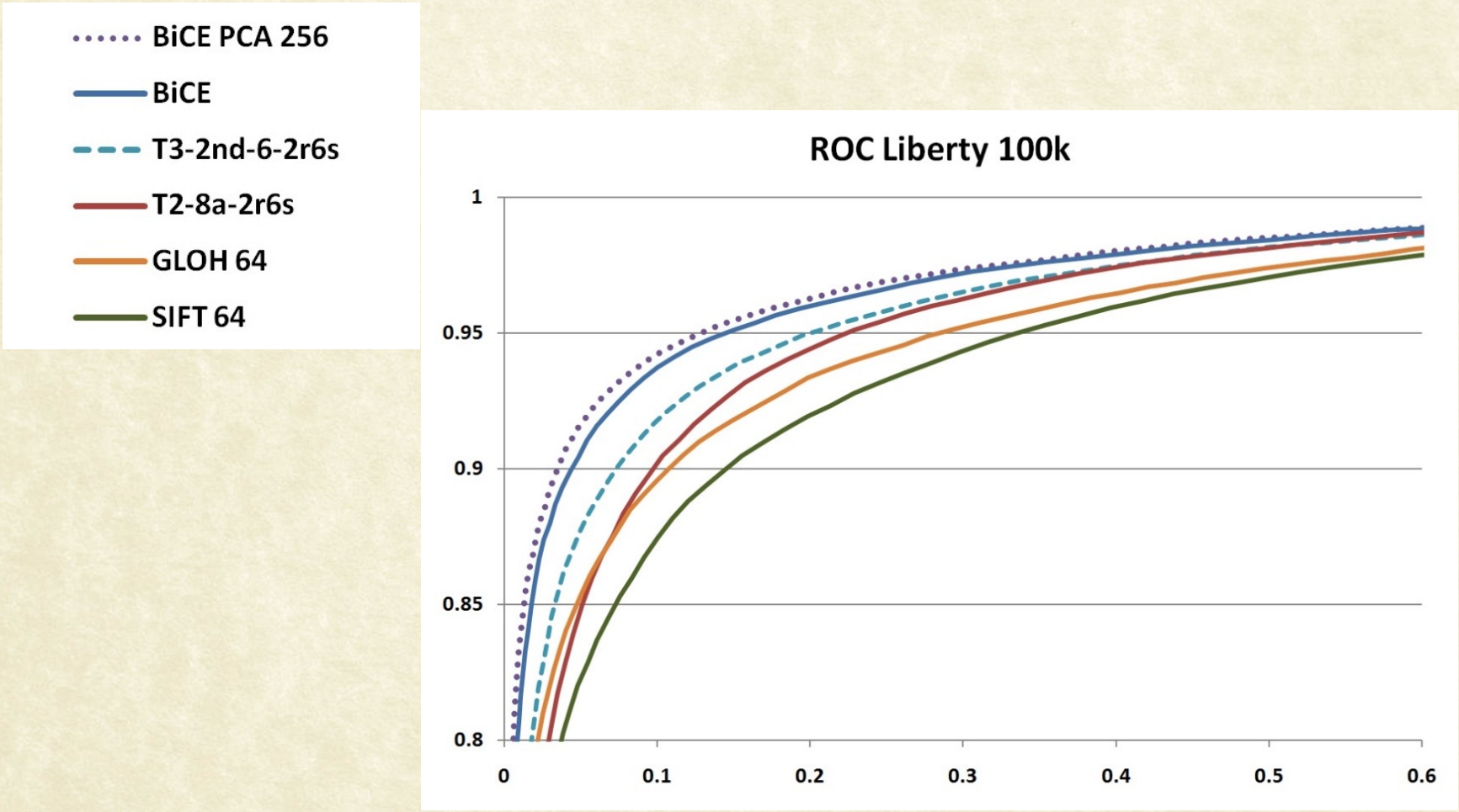
ROC curve **("Receiver Operator Characteristic")**



ROC Curves

- Generated by counting # current/incorrect matches, for different threholds
- Want to maximize area under the curve (AUC)
- Useful for comparing different feature matching methods
- For more info: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

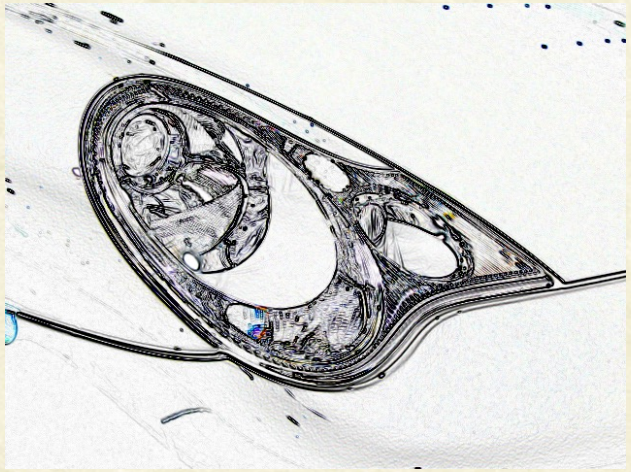
Some actual ROC curves



Questions?

CS7.505: Computer Vision

Spring 2022: Object Retrieval: Bag of Visual Words


Anoop M. Namboodiri

Biometrics and Secure ID Lab, CVIT,

IIIT Hyderabad

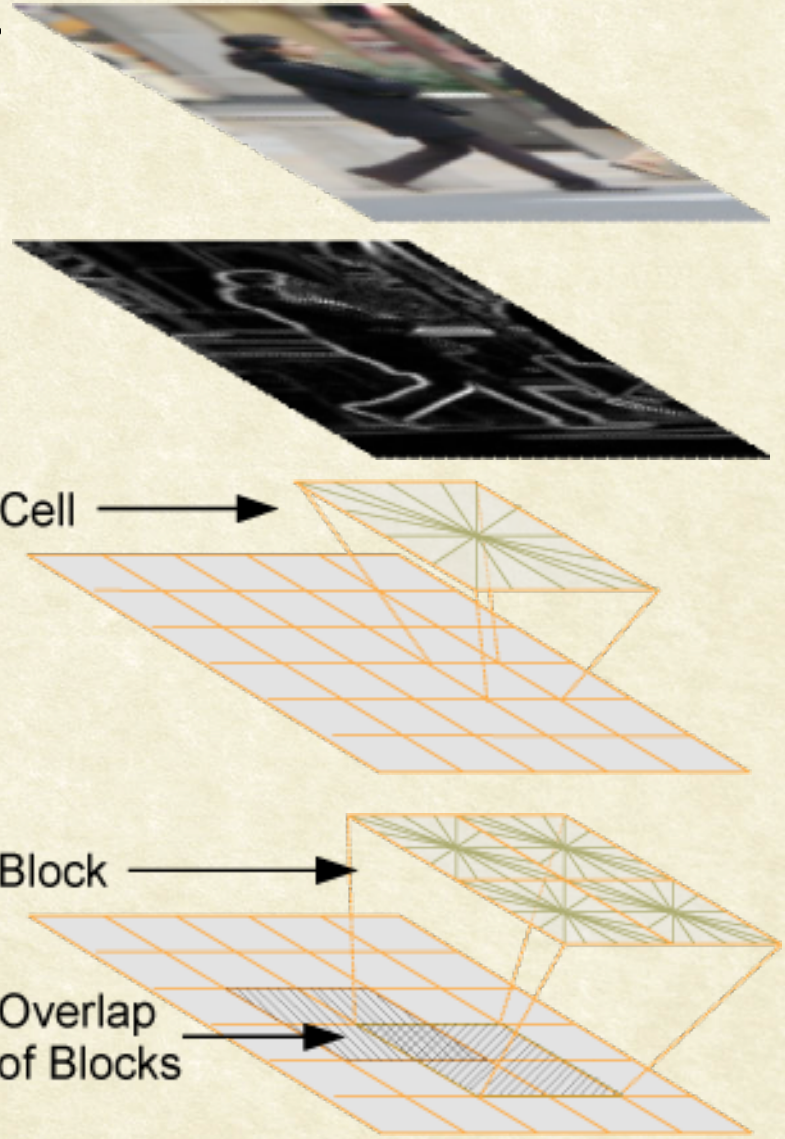
Recap: HoG Descriptor

1. Compute gradients on an image region of 64x128 pixels

2. Compute histograms on ‘cells’ of typically 8x8 pixels (i.e. 8x16 cells)

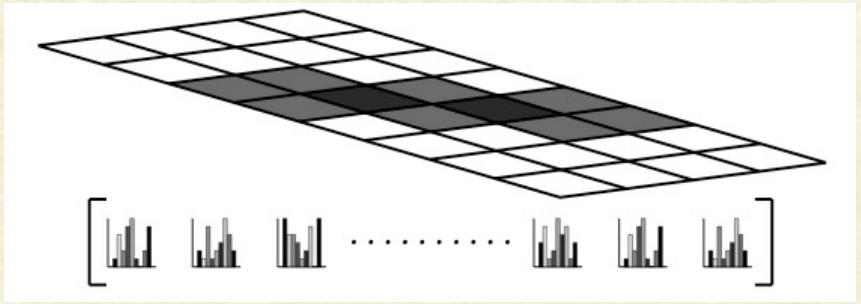
3. Normalize histograms within overlapping blocks of cells (typically 2x2 cells, i.e. 7x15 blocks)

4. Concatenate histograms



Recap: HoG Descriptor

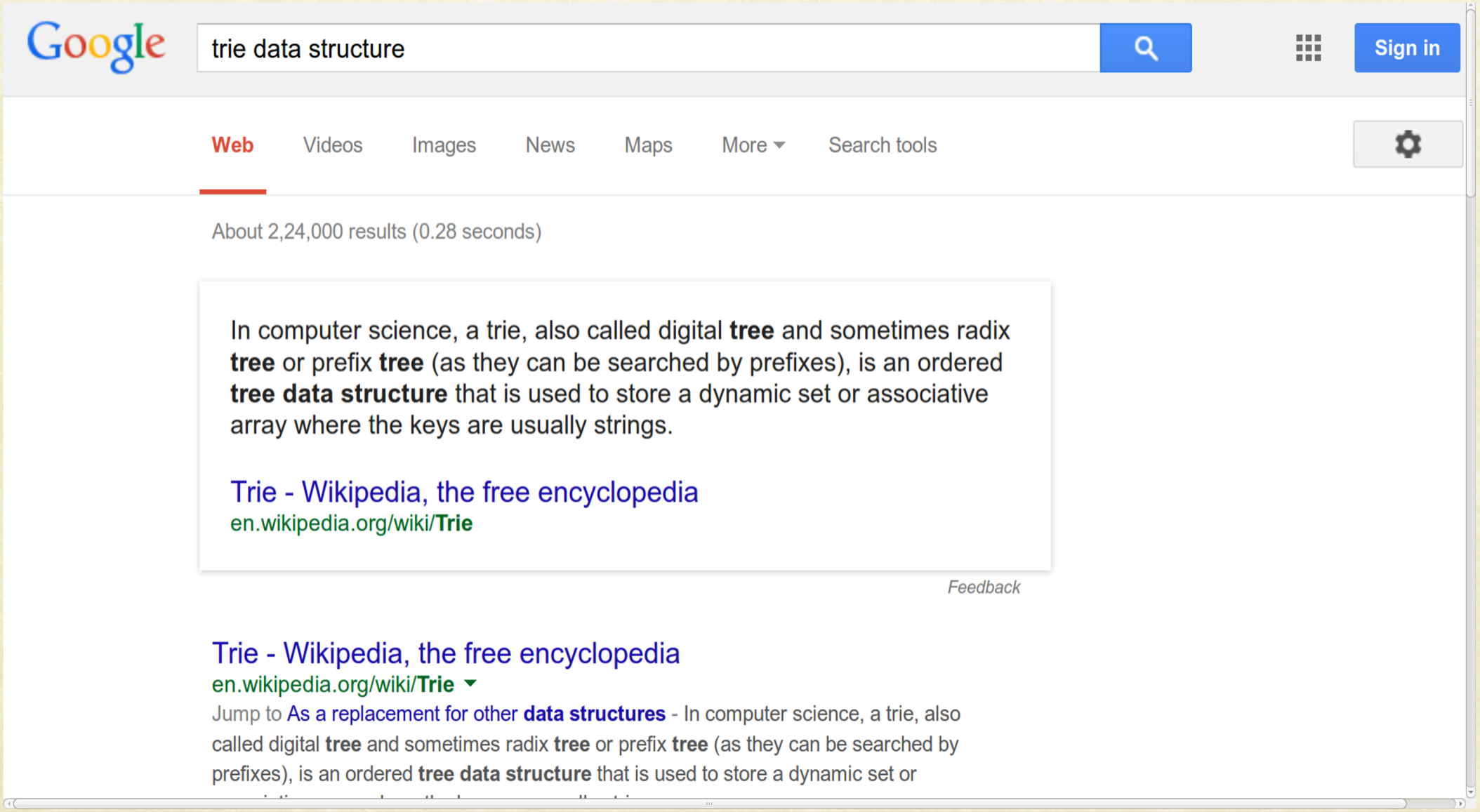
- Concatenation of Blocks



420 x 9-dim Histograms = 3780 dim

- Visualization:

Origins: Trie [from Re'trie'val]



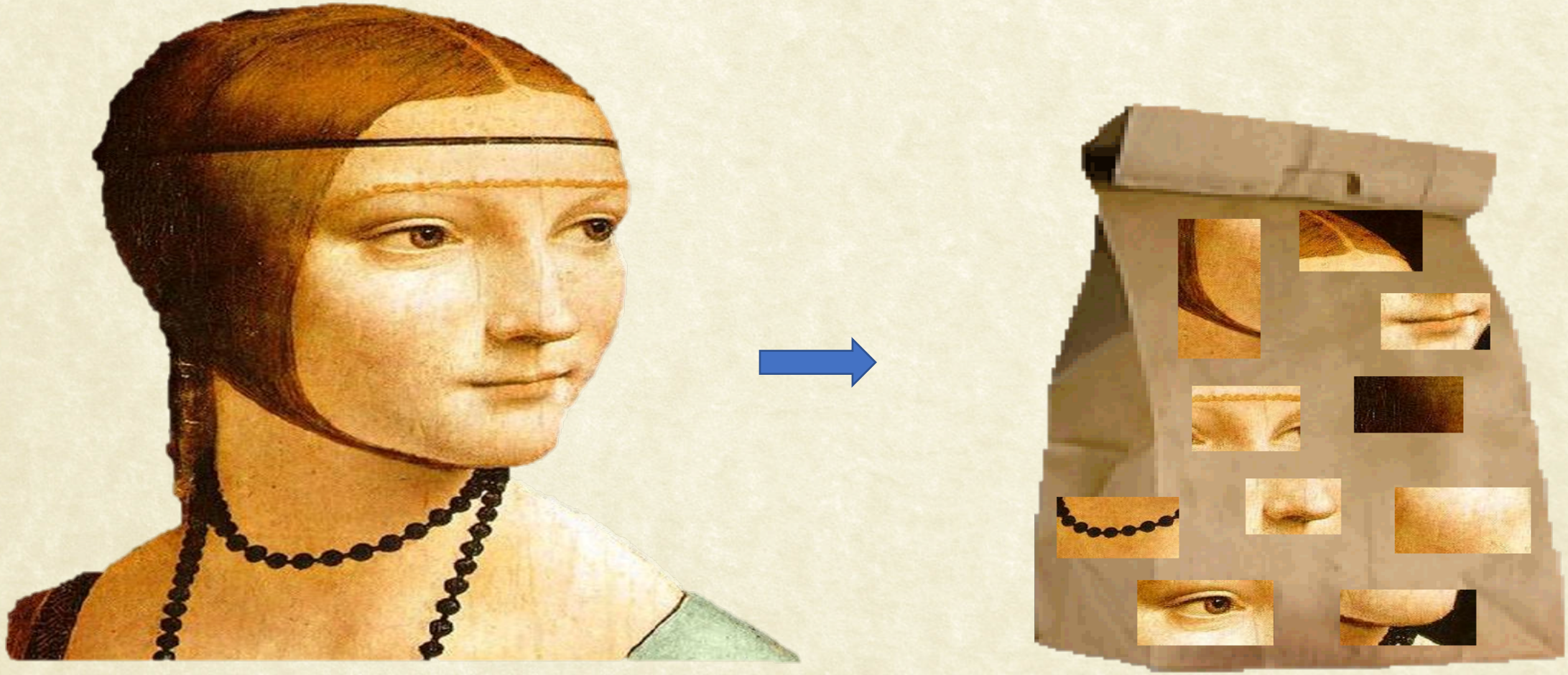
Motivation

- How to “summarize” a document like Google ?
- How do we use “words” in a document to summarize, based on the search word, and assign importance to sentences ?
- Word frequency, unimportant words, document as histogram of words ?
- What information is lost ?
  - Spatial.
  - Sentence construction (syntax/grammar).
  - ...

Motivation

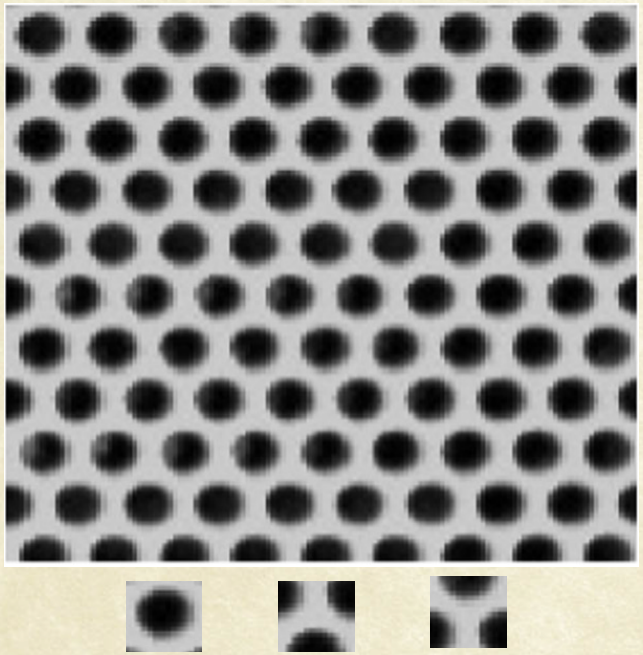
- Related problems ?
  - Comparing two document images.
    - Are these two documents on the same topic ?
    - Are these two documents on “similar” topics ?
      - What do you mean by “similarity” ?
  - Searching for a piece of text or paragraph.
    - Giving collection of words in “quotes”.
  - How can we do this with images ?

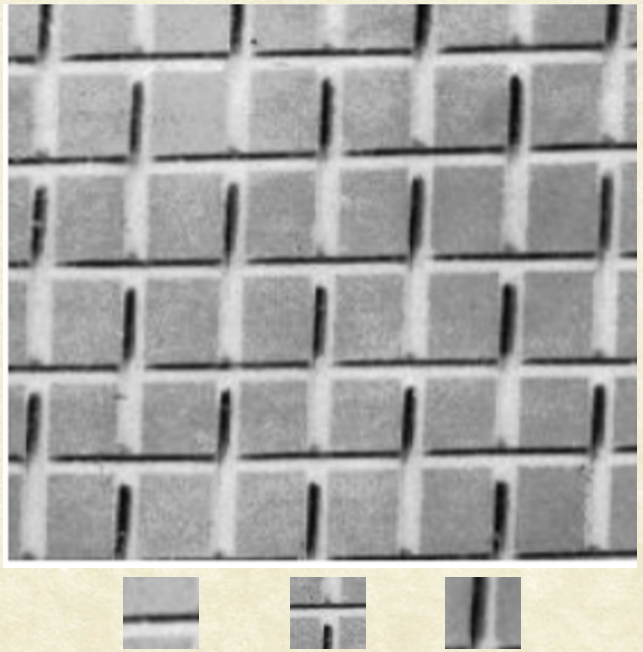
Bag of Features Model

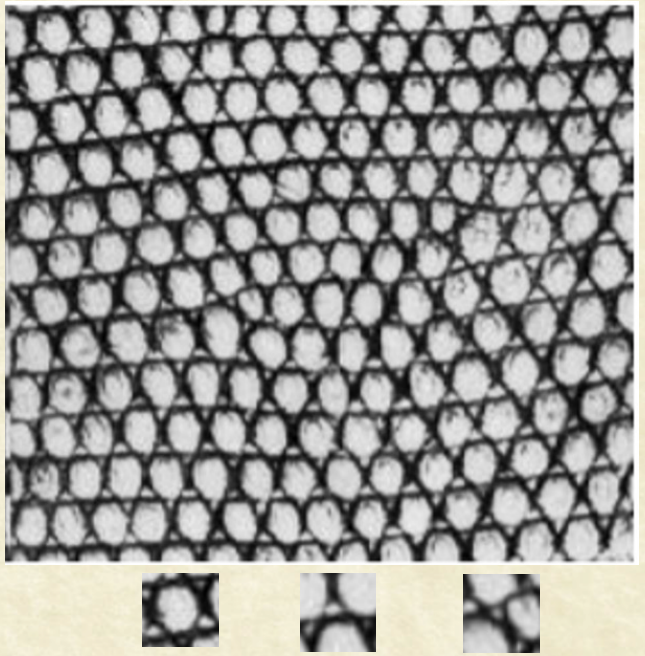


Origin 1: Texture Recognition

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

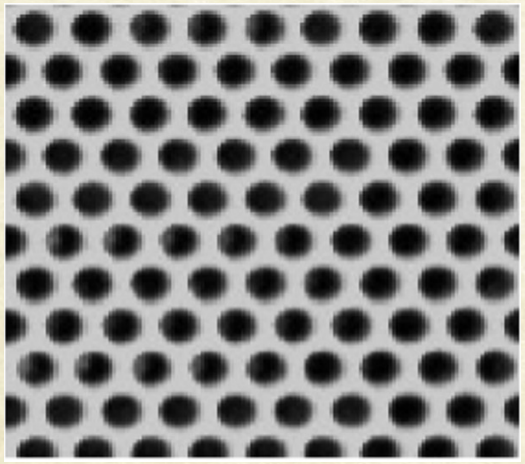


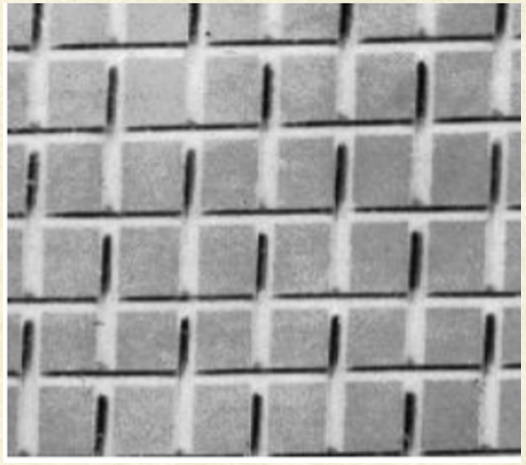


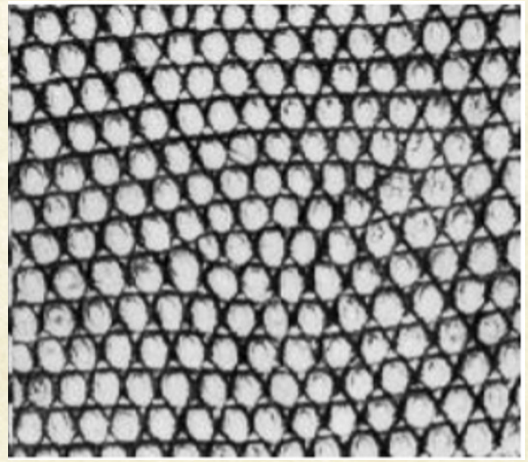


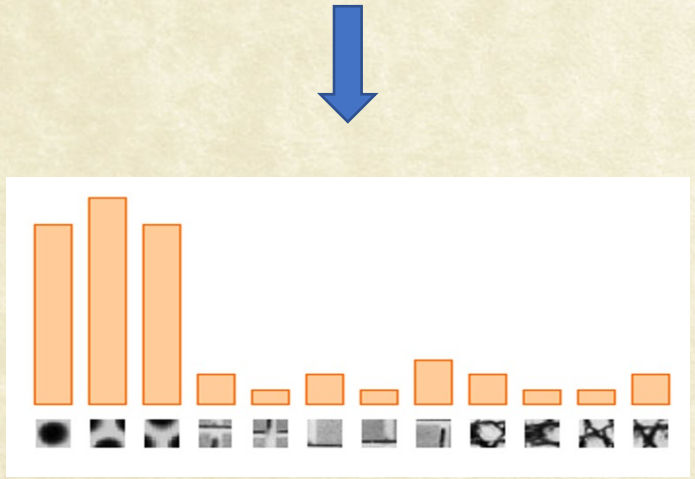
Julesz 1981; Cula & Dana 2001; Leung & Malik 2001; Mori et al. 2001; Varma & Zisserman 2003,03; Lazebnik, Schmid & Ponce 2003

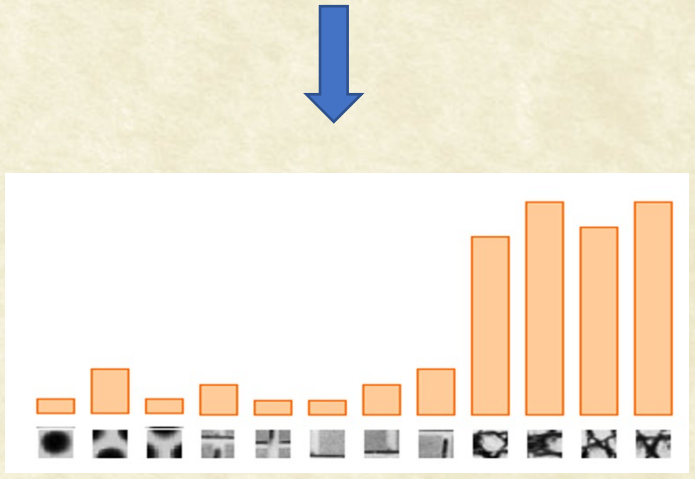
Origin 1: Texture Recognition











Julesz 1981; Cula & Dana 2001; Leung & Malik 2001; Mori et al. 2001; Varma & Zisserman 2003,03; Lazebnik, Schmid & Ponce 2003

Functions for Comparing Histograms

- L-1 Distance:

$$D(h_1, h_2) = \sum_{i=1}^{N} |h_1(i) - h_2(i)|$$

- χ^2 Distance:

$$D(h_1, h_2) = \sum_{i=1}^{N} \frac{(h_1(i) - h_2(i))^2}{h_1(i) + h_2(i)}$$

- Quadratic Distance (cross-bin)

$$D(h_1, h_2) = \sum_{i,j} A_{ij} \, (h_1(i) - h_2(j))^2$$

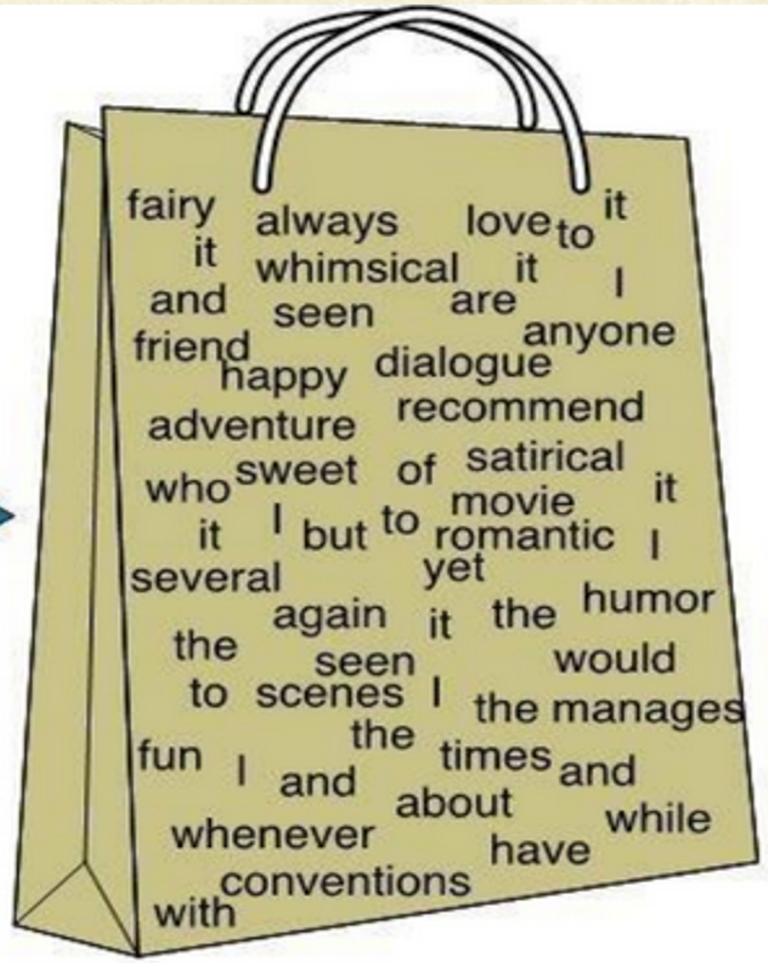
Jan Puzicha, Yossi Rubner, Carlo Tomasi, Joachim M. Buhmann:
Empirical Evaluation of Dissimilarity Measures for Color and Texture, ICCV 1999

Bag of Words - Text Domain

- Orderless documentation representation, frequencies of words from a dictionary.

Bag of Words

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!



it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1
...	...

Bag of Words Histogram

Dictionary

1. Soviet
2. Cuba
3. Missile
4. Japanese
5. Germany
6. War
7. Iraq
8. Terrorists
9. Freedom
10. Commitment

- Orderless document representation; frequencies of words from a dictionary
- Classification to determine document category

Tag Clouds


Slide
just
colleague
within
text
see
Person
check
visible
View
friend
linked
problem
new
design
PDF
time
makes
comments
click
change
Next
go
use
apply
best
made
Just
sends
messages
master
choose
smaller
left
win
MS
increase
screen
theme
business
layout
slide
comment
size
one
need
feedback
last point
e-book
every
contest
answer
top
wonderful
box
India
presentation
people
add
Automatically
delete
key
Click
presentations
applied
Master
particular
Now
make
pixels
also
audience
well
designer
Comment
using
PowerPoint
colour
know
list
second
adds
now
easy
feature
PPT
Rs
option
slides
book
want
logo
might
tab
reflect
going
open
Manvendra
Go
two
New
back
Review
carry
create
small
type
right
font
view
blue
title
layouts
first
element

Difference between Features & Words

Words

- Dictionary/Vocab.
- Meaning.
- Finite/Precise.
- Language known.

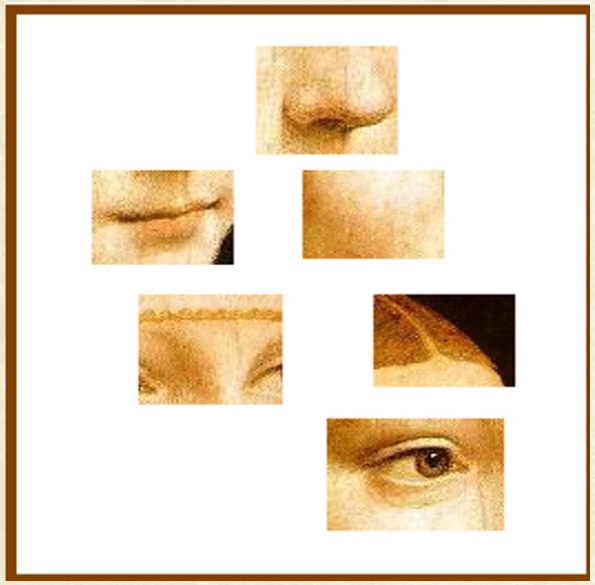
Features

- Dictionary/Vocab ?
- Meaning ?
- Finite/Precise ?
- What Language ?

How do we get “visual” words ?

Bag of Features for Image Classification

1. Extract Features



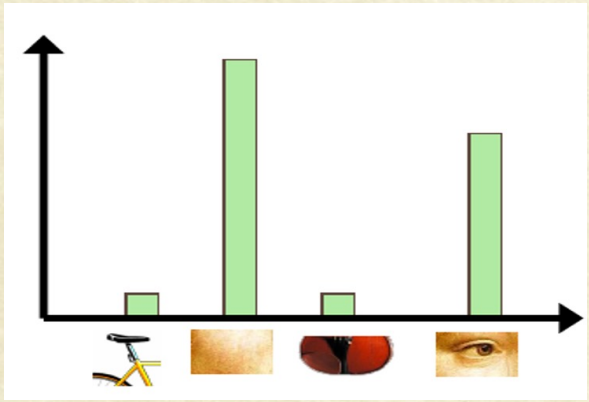
2. Learn “Visual Vocabulary”

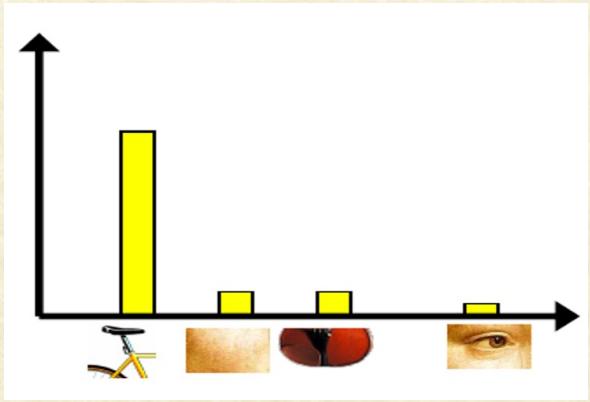
Bag of Features for Image Classification

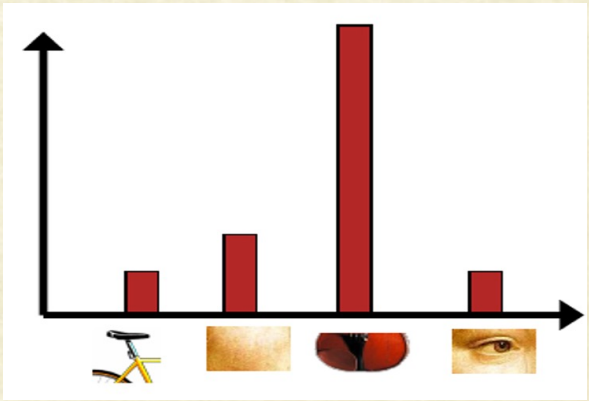
1. Extract Features

2. Learn “Visual Vocabulary”

3. Quantize Features using Visual Vocabulary
4. Represent Images by Frequencies of Visual Words

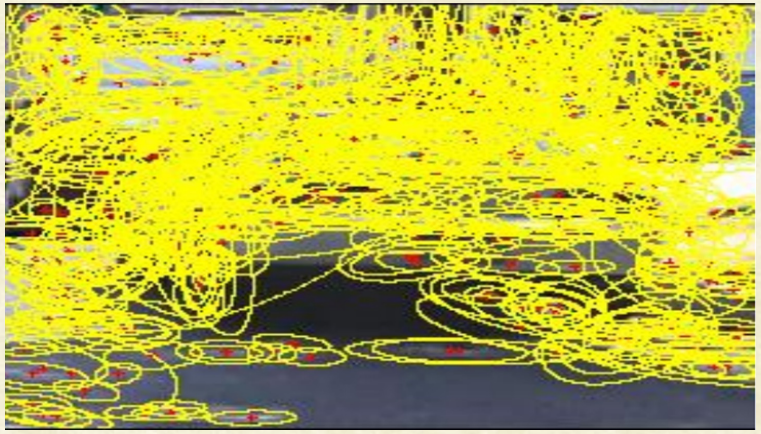


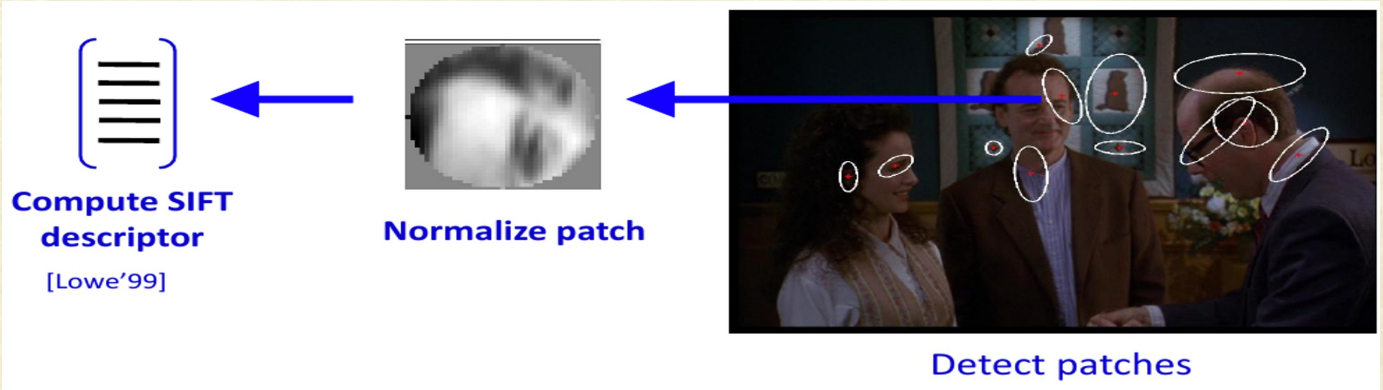




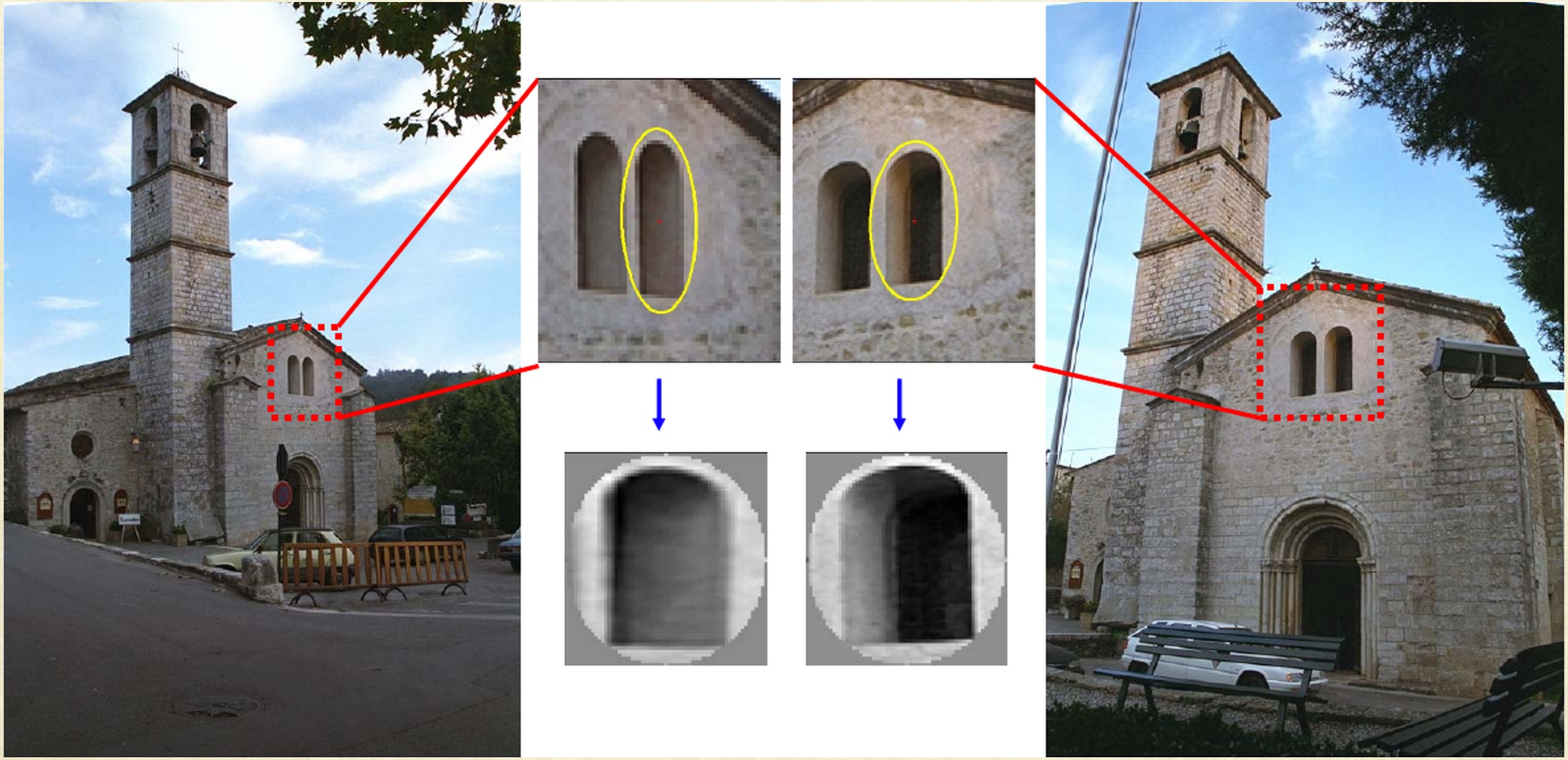
1. Feature Extraction

- Regular Grid
  - Vogel and Schiele 2003
  - Fei-Fei and Perona 2005
- Interest Point Detector
  - Csurka et al. 2004
  - Fei-Fei and Perona 2005
  - Sivic et al. 2005
- Features
  - Mikojaczyk & Schmid ‘02
  - Mata et al. ‘02
  - Sivic and Zisserman ‘03

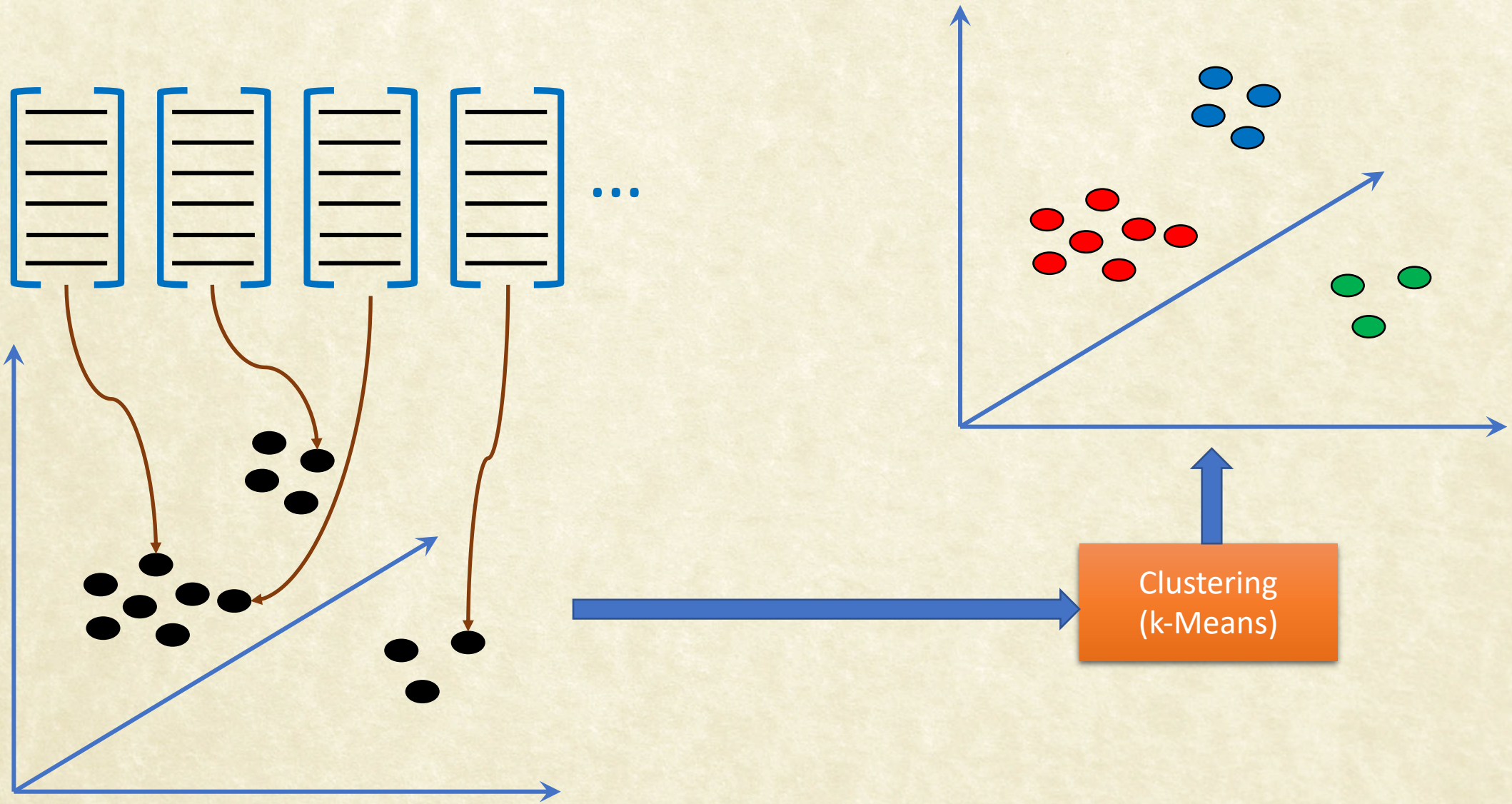




1. Feature Extraction: Example



2: Learning Visual Vocabulary



Using the Visual Vocabulary

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

Difference between features & words

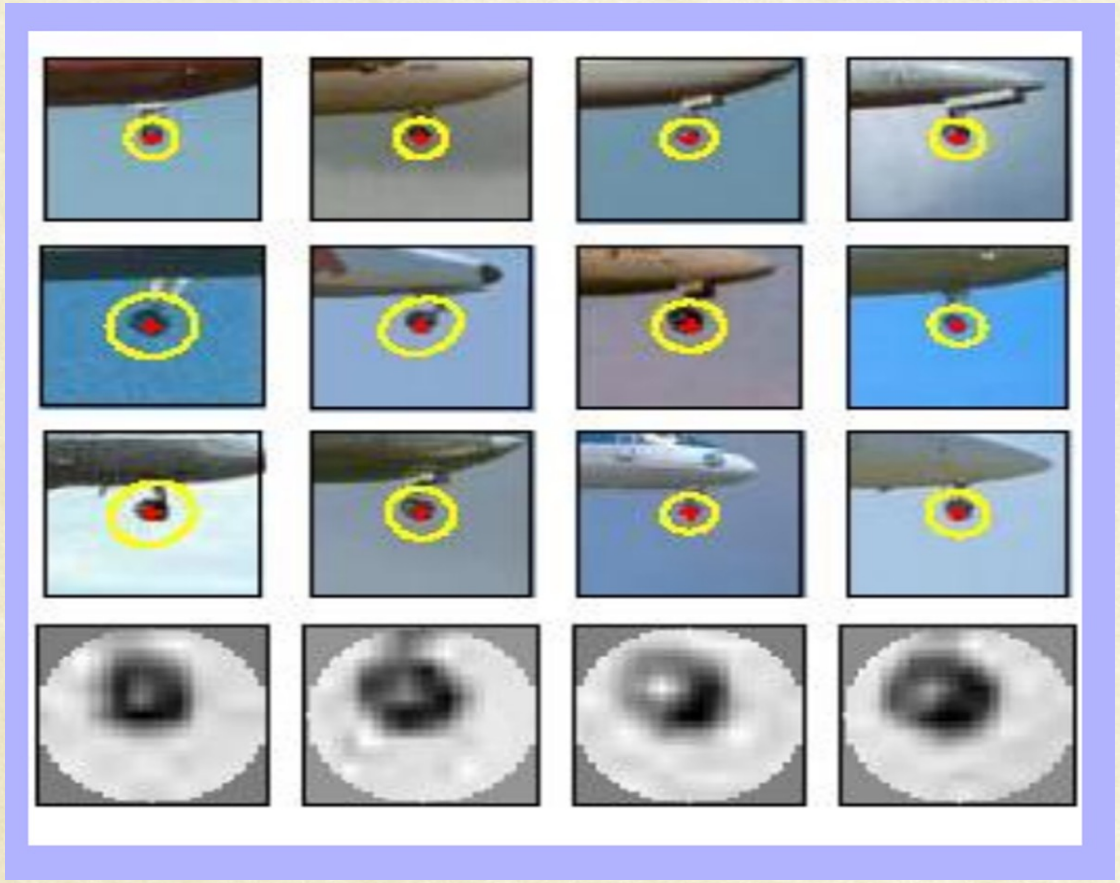
Words

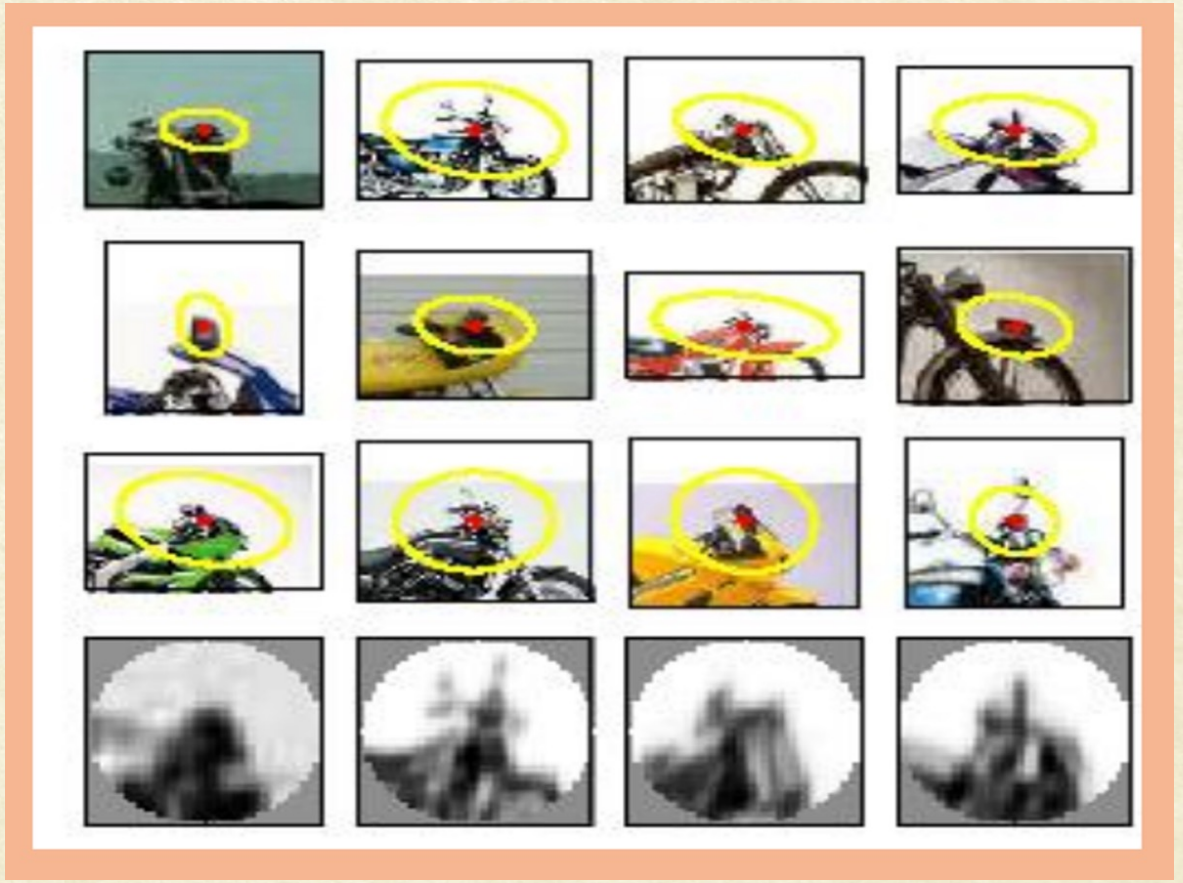
- Dictionary/Vocab.
- Meaning.
- Finite/Precise.
- Language known.
- Simple.
- Language-objective

Features

- Dictionary/Vocab ?
- Meaning ?
- Finite/Precise ?
- What language ?
- Complicated!
- Training subjective!

Image Patch examples of Visual Words

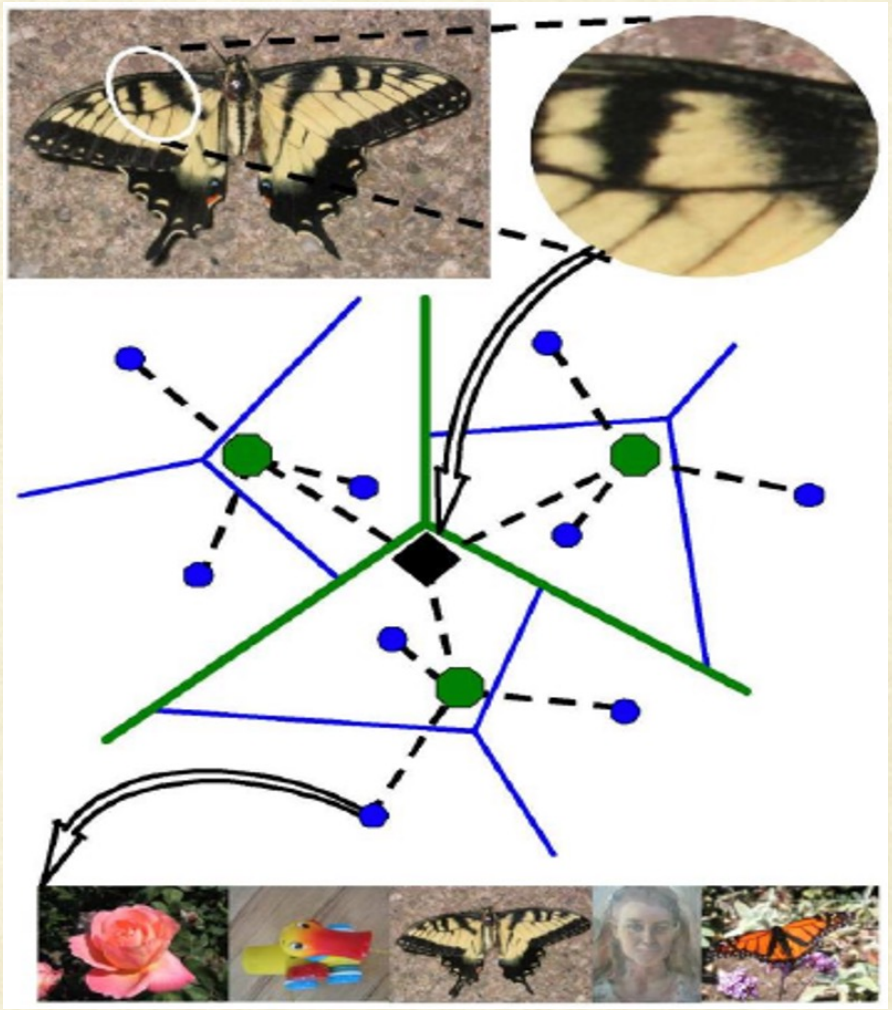




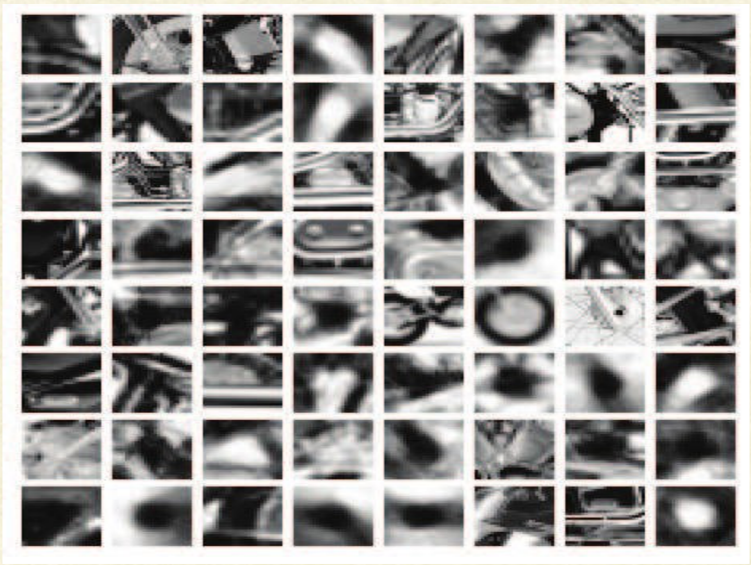
Visual Vocabulary: Challenges

- How to choose vocabulary size?
  - Too small: visual words not representative of all patches
  - Too large: Quantization artifacts
- Generative or Discriminative?
- Computational Efficiency
  - Vocabulary Trees

  (Nister and Stewenius 2006)

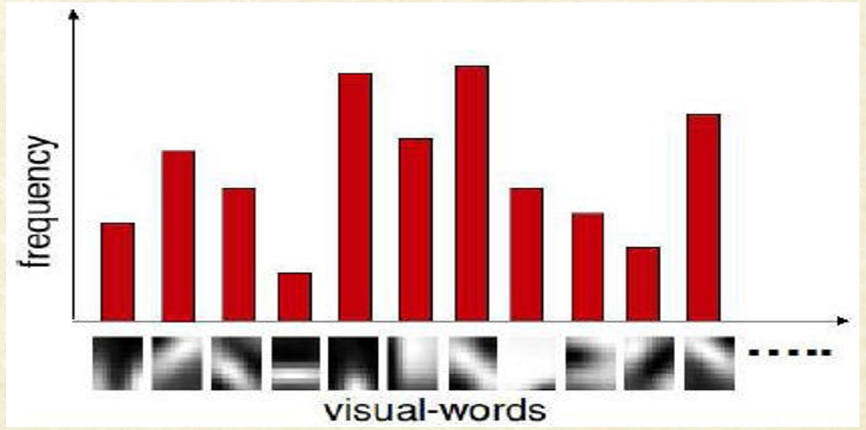


3: Image Representation



Learned Visual

Vocabulary

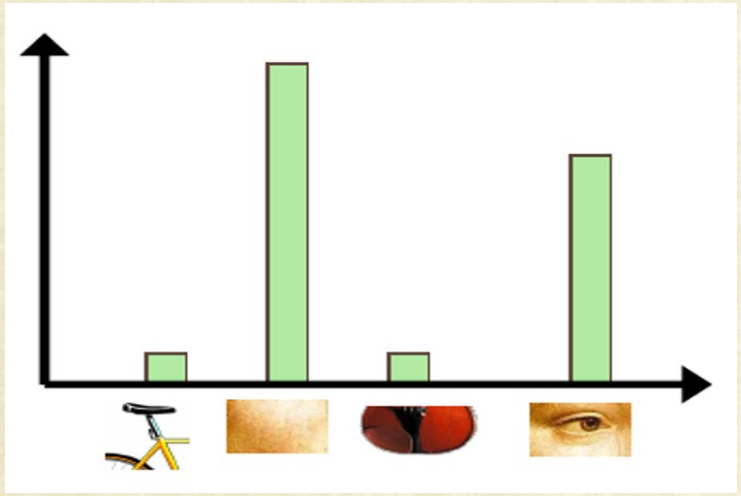


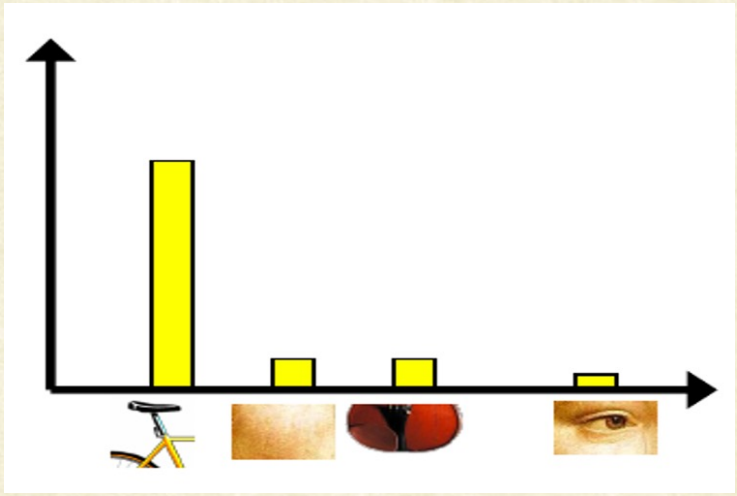
Problems we can solve (attempt) ?

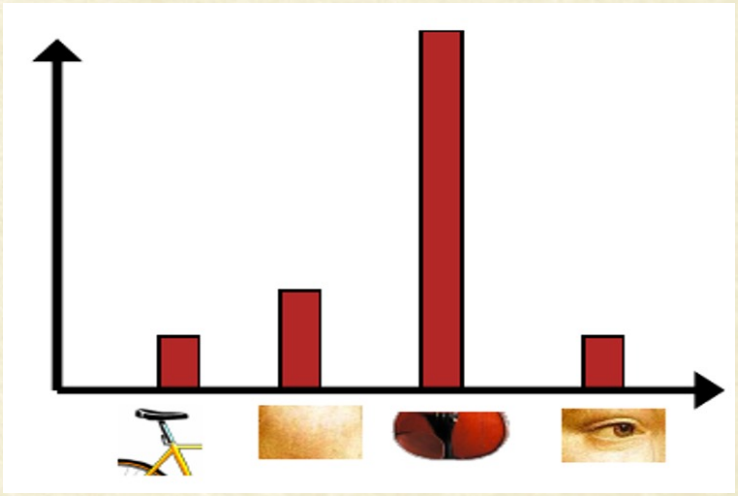
- Classification: Do these two images (visual documents) belong to the same subject ?
- Recognition: Which images contain chairs ?
- If images are documents, what are videos ?
- Actions are sequence of visual words organized in time.
- How to get better (spatial) representation (to enforce “structure” in documents/images) ?
- Search!

Image Classification

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







- Consider them as training examples in a feature space
- Learn a discriminative / generative classifier model

Image and Document differences

Document

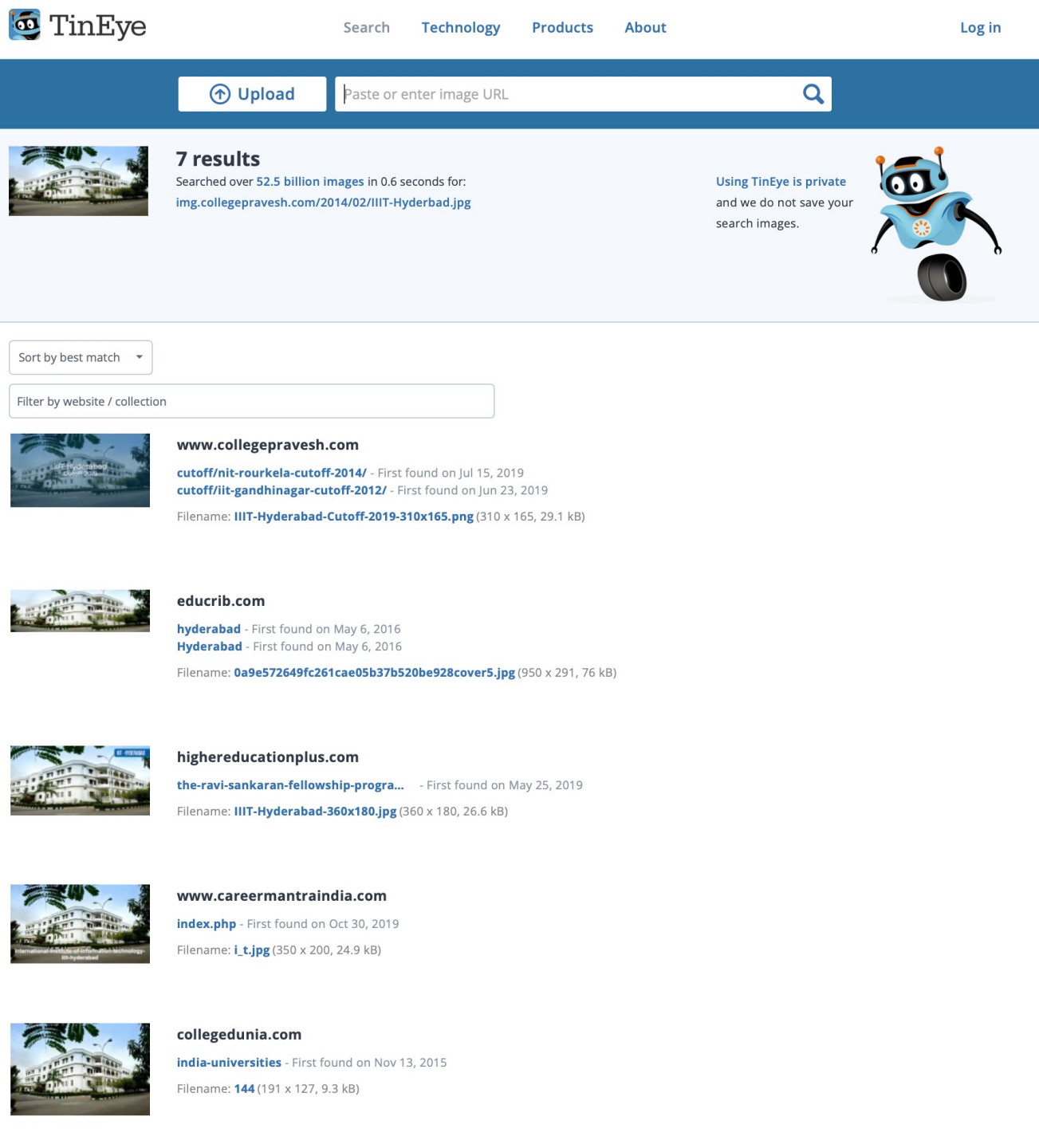
- Single scale.
- Words have order in document!
- 1D space.
- Google!

Image

- Multiple scales!
- Features are *loosely* coupled.
- 2D space.
- TinEye!

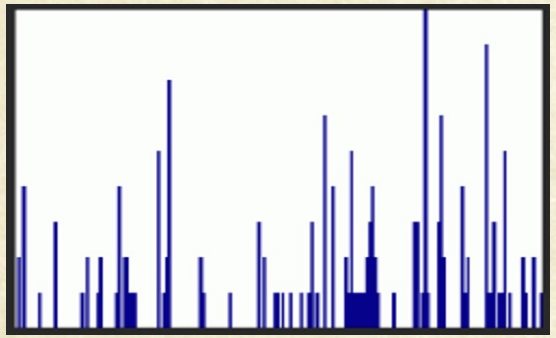
e.g.: TinEye Search

- Why Search Images?
  - Copyright problems / attribution
  - Words may not be enough
    - Find me similar dresses
    - Find me goggles like this
  - What is the name of this building?
- Challenge
  - Visual Words are Contextual
  - How to incorporate more spatial information?

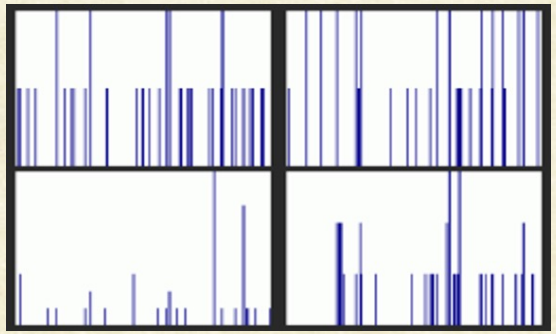


Spatial Pyramid Representation

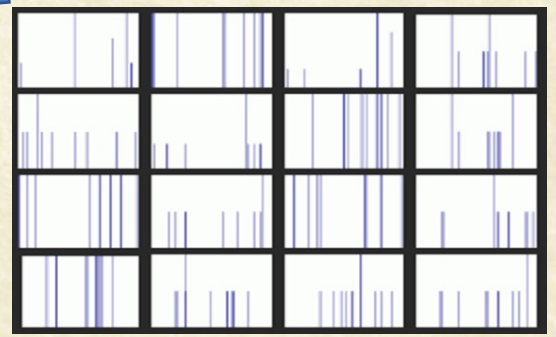
- Extension of bag-of-features representation
- Locally order-less representation at several levels of resolution



Level 0



Level 1



Level 2

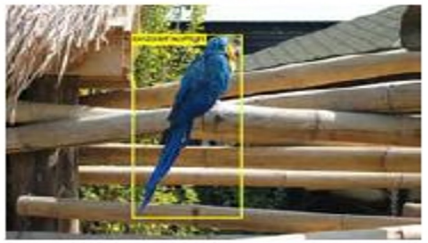
Lazebnik, Schmid and Ponce (CVPR 2006)

Examples from PASCAL VOC Challenge 2010

Aeroplane

Bicycle

Bird



Boat

Bottle

Bus

Car

Cat

Chair

Cow

Video Google

Enable video, e.g. a feature length film, to be searched on its visual content with the same ease and success as a Google search of text documents.

"Run Lola Run" ('Lola Rennt') [Tykwer, 1999]

"Groundhog Day" [Rammis, 1993]

Visually defined search

Given an object specified in one frame, retrieve all shots containing the object:

- must handle viewpoint change
- must be efficient at run time

Example : Groundhog Day

close-up

Example: Groundhog Day

73 keyframes retrieved

53 correct, first incorrect ranked 27

Rank: 12 35 50 69

Questions!