Descriptors III

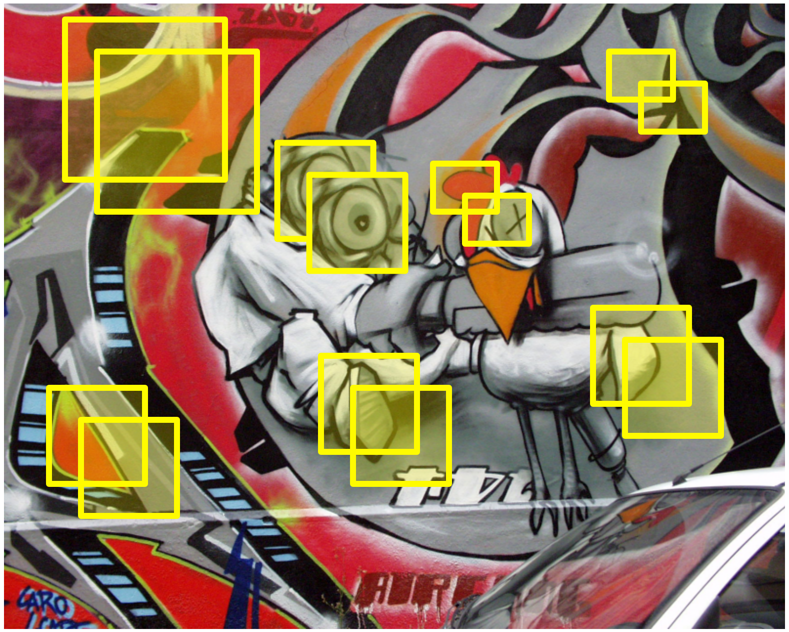
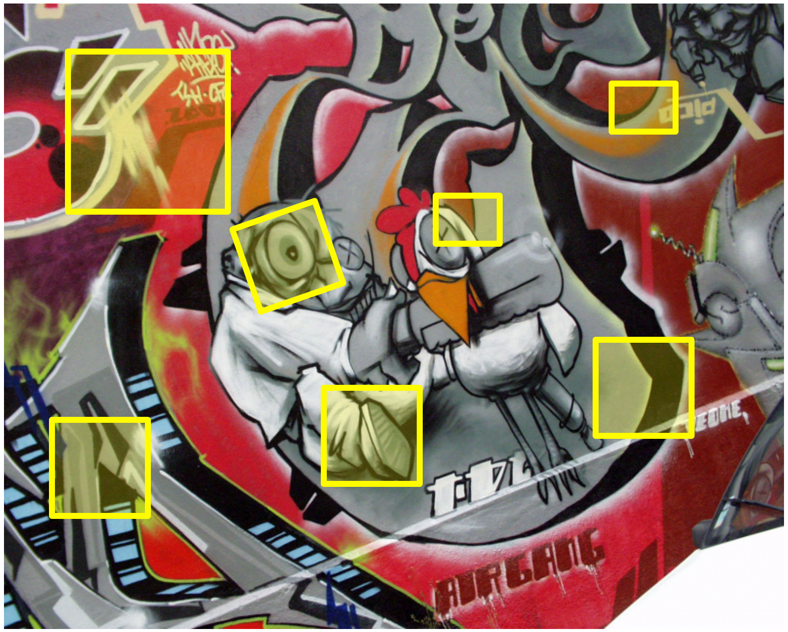
CSE 576

Ali Farhadi

Many slides from Larry Zitnick, Steve Seitz

How can we find corresponding points?

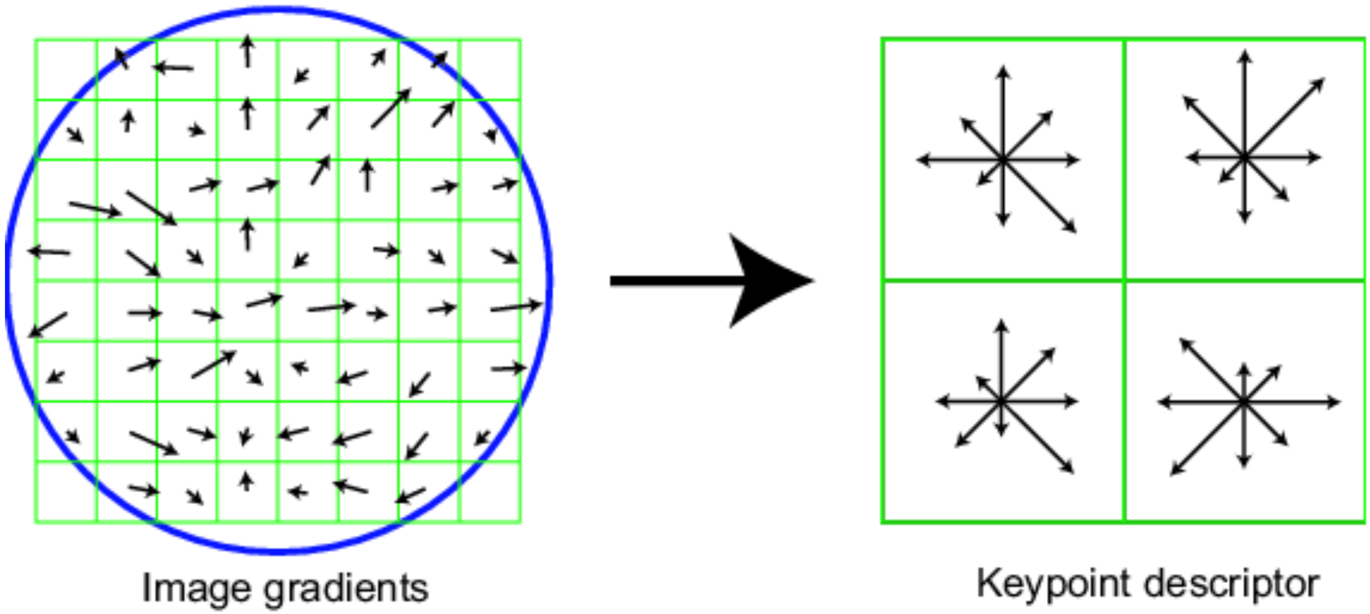
How can we find correspondences?

SIFT descriptor

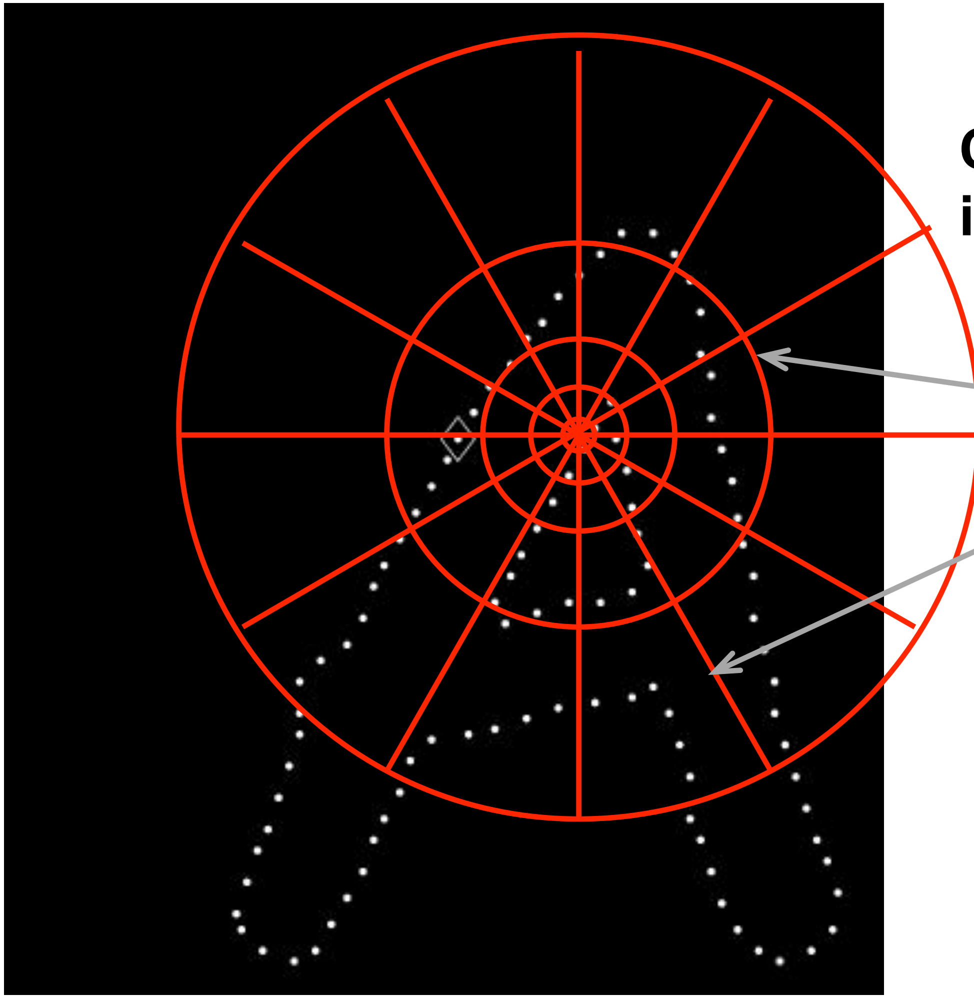
Full version

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



Adapted from slide by David Lowe

Local Descriptors: Shape Context



Count the number of points inside each bin, e.g.:

Count = 4

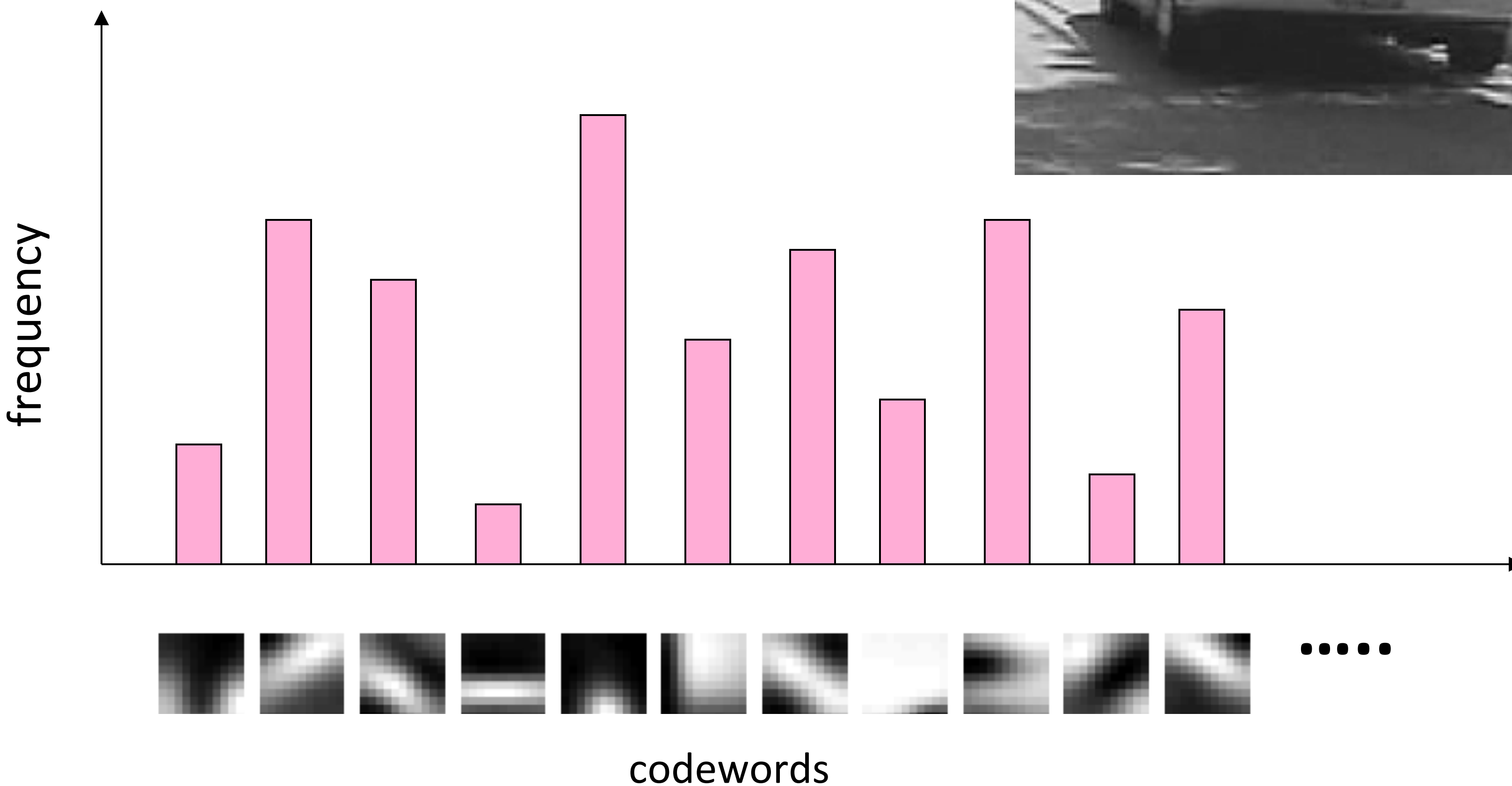
⋮

Count = 10

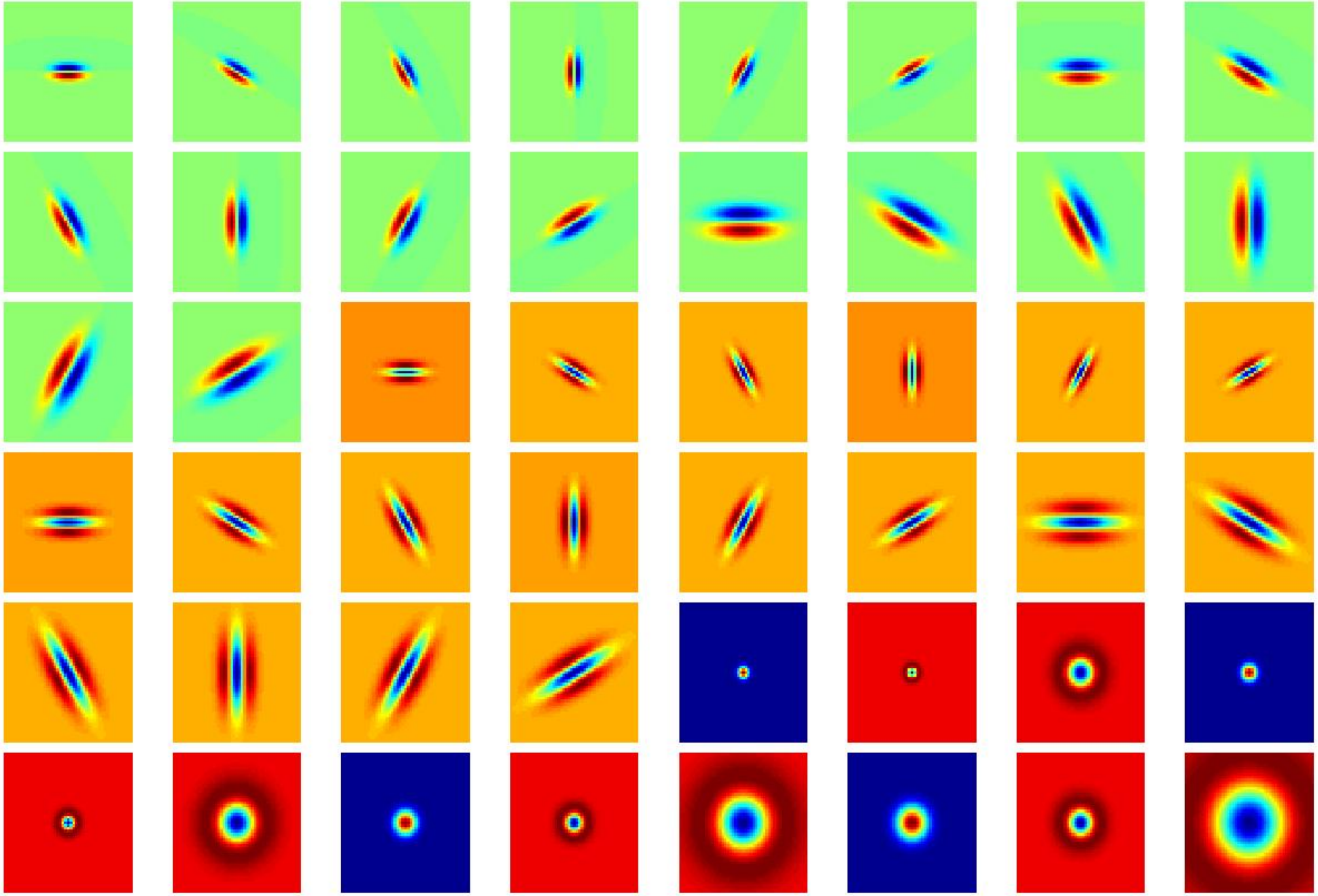
Log-polar binning: more precision for nearby points, more flexibility for farther points.

Belongie & Malik, ICCV 2001

Bag of Words



Another Representation: Filter bank



Spatial pyramid representation

- Extension of a bag of features
- Locally orderless representation at several levels of resolution

Lazebnik, Schmid & Ponce (CVPR 2006)

What about Scenes?

Demo : Rapid image understanding

By Aude Oliva

Instructions: 9 photographs will be shown for half a second each. Your task is to **memorize these pictures**

Credit: A. Torralb

Credit: A. Torralb

Credit: A. Torralb

Credit: A. Torralb

Credit: A. Torralb

Credit: A. Torralb

Credit: A. Torralb

Credit: A. Torralb

Credit: A. Torralb

Memory Test

Which of the following pictures have you seen ?

If you have seen the image
clap your hands once

If you have not seen the image
do nothing

Credit: A. Torralba

Have you seen this picture ?

Credit: A. Torralb

Credit: A. Torralb

Have you seen this picture ?

Credit: A. Torralb


NO

Have you seen this picture ?


NO

Have you seen this picture ?

Credit: A. Torralb

Have you seen this picture ?

Credit: A. Torralb

Have you seen this picture ?

Credit: A. Torralb

Credit: A. Torralb

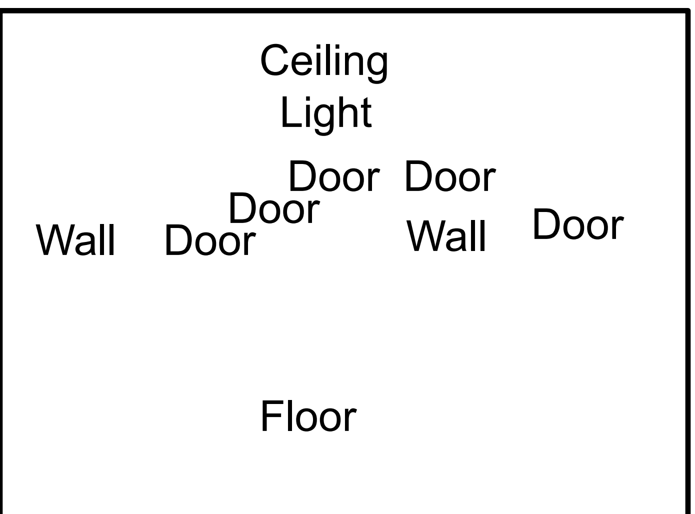
You have seen these pictures

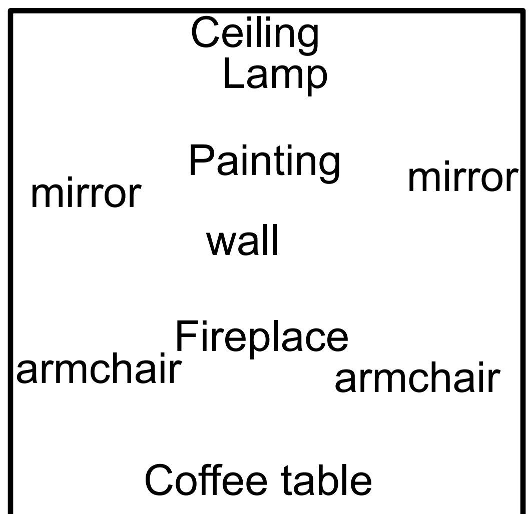
You were tested with these pictures

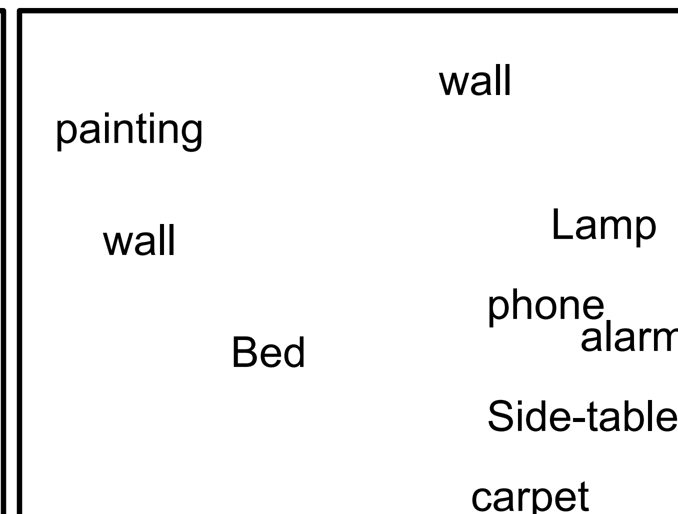
The gist of the scene

In a glance, we remember the meaning of an image and its global layout but some objects and details are forgotten

Which are the important elements?

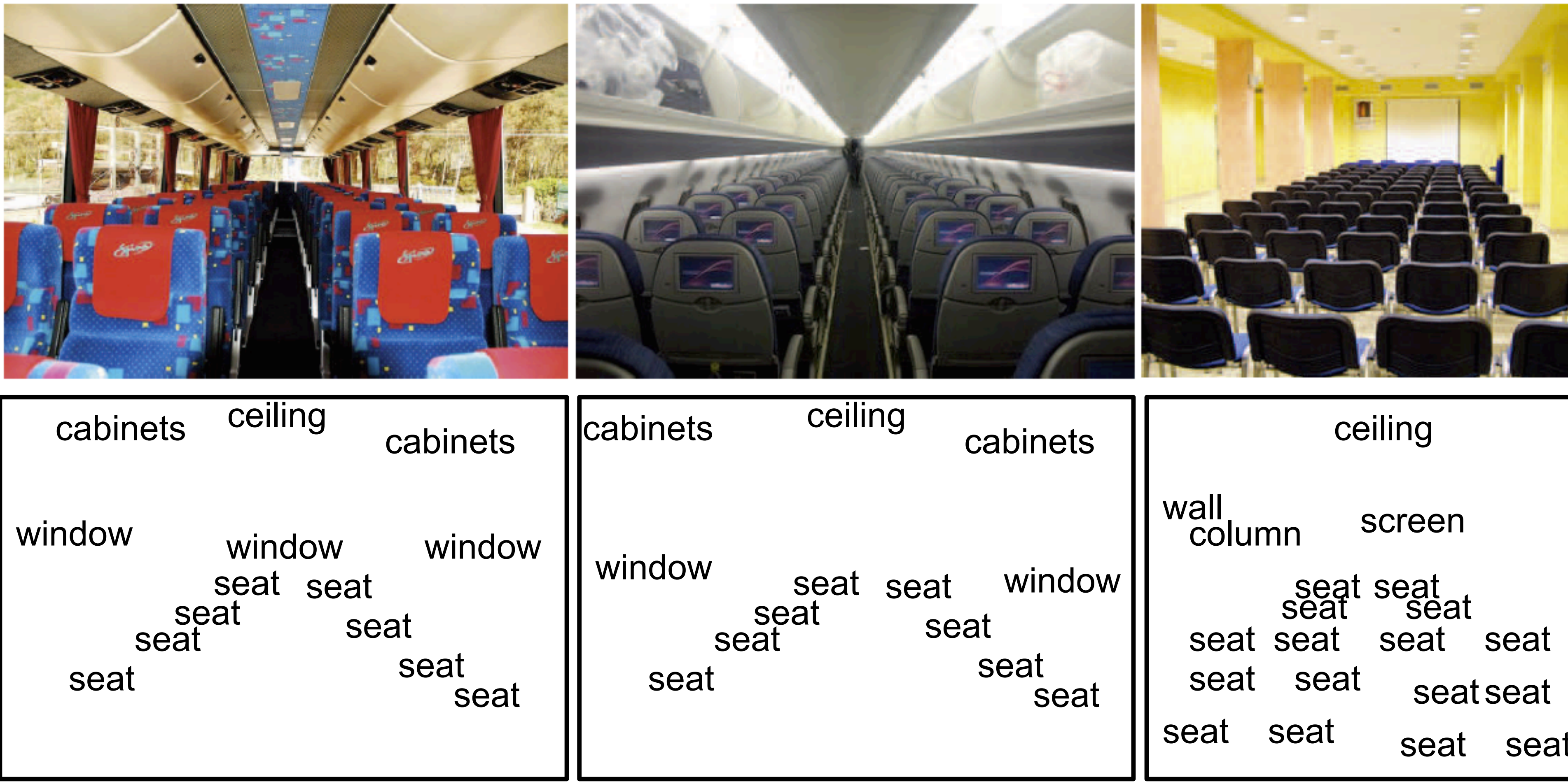






Different content (i.e. objects), different spatial layout

Which are the important elements?



Similar objects, and similar spatial layout

Different lighting, different materials, different “stuff”

Holistic scene representation: Shape of a scene

- Finding a low-dimensional "scene space"
- Clustering by humans
  - Split images into groups
  - ignore objects, categories

Table 1. Spatial envelope properties of environmental scenes.

Property	S1	S2	S3	Total
Naturalness	65	12	0	77
Openness	6	53	24	83
Perspective	6	18	29	53
Size	0	0	47	47
Diagonal plane	0	12	29	41
Depth	18	12	29	59
Symmetry	0	0	29	29
Contrast	0	0	18	18

Results are in %, for each of the three experimental steps. The total represents the percent of times the attribute has been used regardless of the stage of the experiment.

Spatial envelope properties

- Naturalness
  - natural vs. man-made environments

Spatial envelope properties

- Openness
  - decreases as number of boundary elements increases

Spatial envelope properties

- Roughness
  - size of elements at each spatial scale, related to fractal dimension

Spatial envelope properties

- Expansion (man-made environments)
  - depth gradient of the space

Spatial envelope properties

- Ruggedness (natural environments)
  - deviation of ground relative to horizon

Scene statistics

- DFT (energy spectrum)
  - throw out phase function (represents local properties)
- Windowed DFT (spectrogram)
  - Coarse local information
  - 8x8 grid for these results

Scene statistics

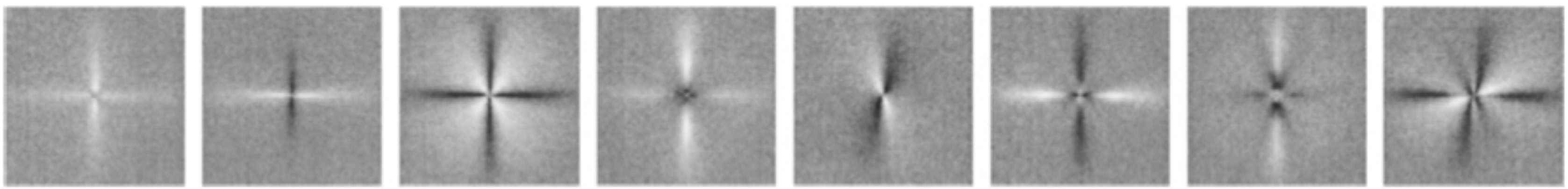


Figure 2. The first eight principal components for energy spectra of real-world scenes. The frequency $f_x = f_y = 0$ is located at the center of each image.

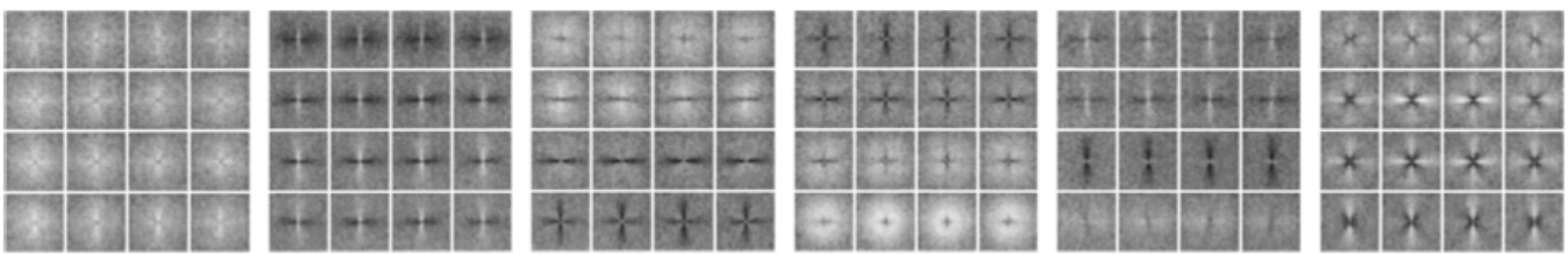
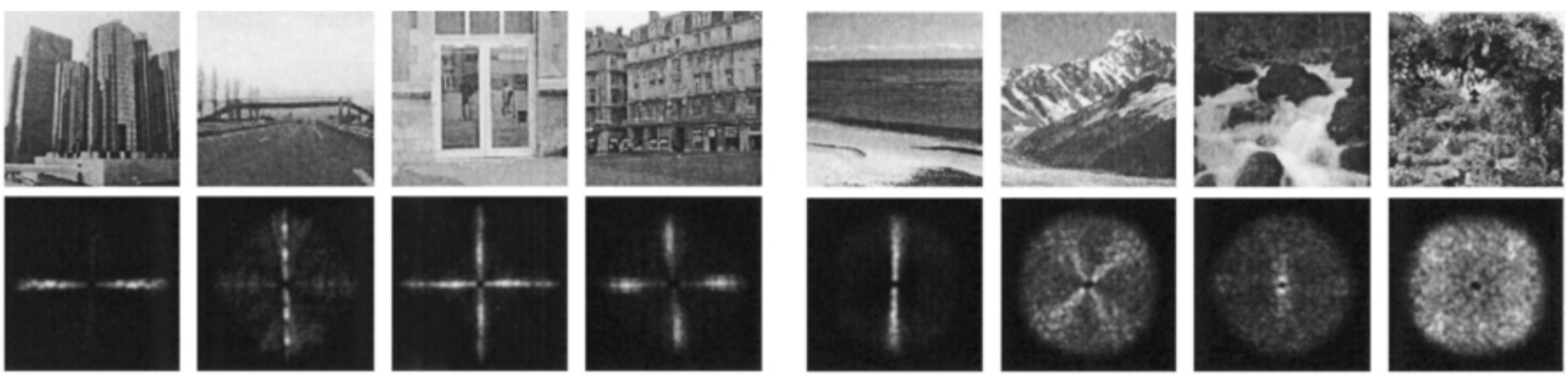


Figure 3. The first six principal components of the spectrogram of real-world scenes. The spectrogram is sampled at 4×4 spatial location for a better visualization. Each subimage corresponds to the local energy spectrum at the corresponding spatial location.

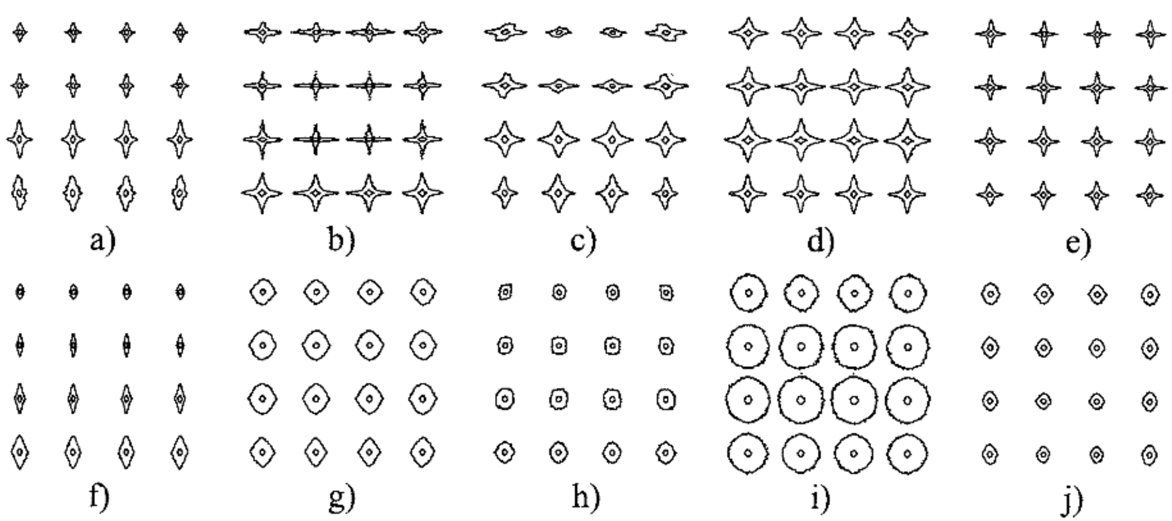
Scene classification from statistics

- Different scene categories have different spectral signatures
  - Amplitude captures roughness
  - Orientation captures dominant edges



Scene classification from statistics

- Open environments have non-stationary second-order statistics
  - support surfaces
- Closed environments exhibit stationary second-order statistics



a) man-made open environments
b) urban vertically structured environments
c) perspective views of streets
d) far view of city-center buildings
e) close-up views of urban structures
f) natural open environments
g) natural closed environments
h) mountainous landscapes
i) enclosed forests
j) close-up views of non-textured scenes

Learning the spatial envelope

- Use linear regression to learn
  - DST (discriminant spectral template)
  - WDST (windowed discriminant spectral template)
- Relate spectral representation to each spatial envelope feature

Learning the spatial envelope

- Primacy of Man-made vs. Natural distinction
  - Linear Discriminant analysis
  - 93.5% correct classification
- Role of spatial information
  - WDST not much better than DST
  - Loschky, et al., scene inversion

Learning the spatial envelope

- Other properties calculated separately for natural, man-made environments

Table 2. Correlation between orderings of natural scenes made by observers and the two templates for each spatial envelope property.

	Openness	Ruggedness	Roughness
DST	$m = 0.82$	0.73	0.82
WDST	$m = 0.88$	0.79	0.86
Agreement	0.92	0.82	0.87

Agreement measures the concordance between subjects.

Table 3. Correlation between orderings of urban scenes made by observers and the two templates for each spatial envelope property.

	Openness	Expansion	Roughness
Energy spectrum	$m = 0.87$	0.77	0.83
Spectrogram	$m = 0.90$	0.88	0.85
Agreement	0.92	0.91	0.88

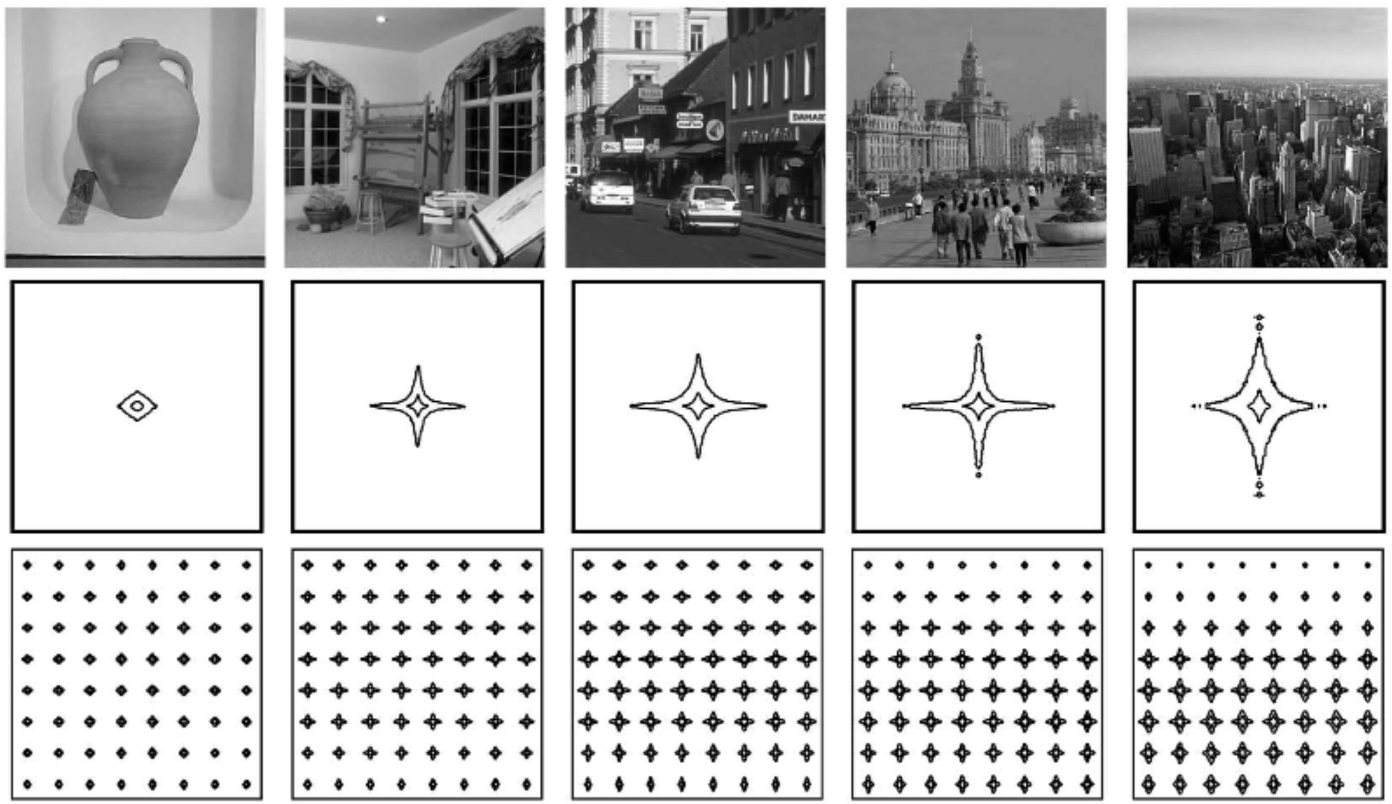
Agreement measures the concordance between subjects.

Spatial envelope and categories

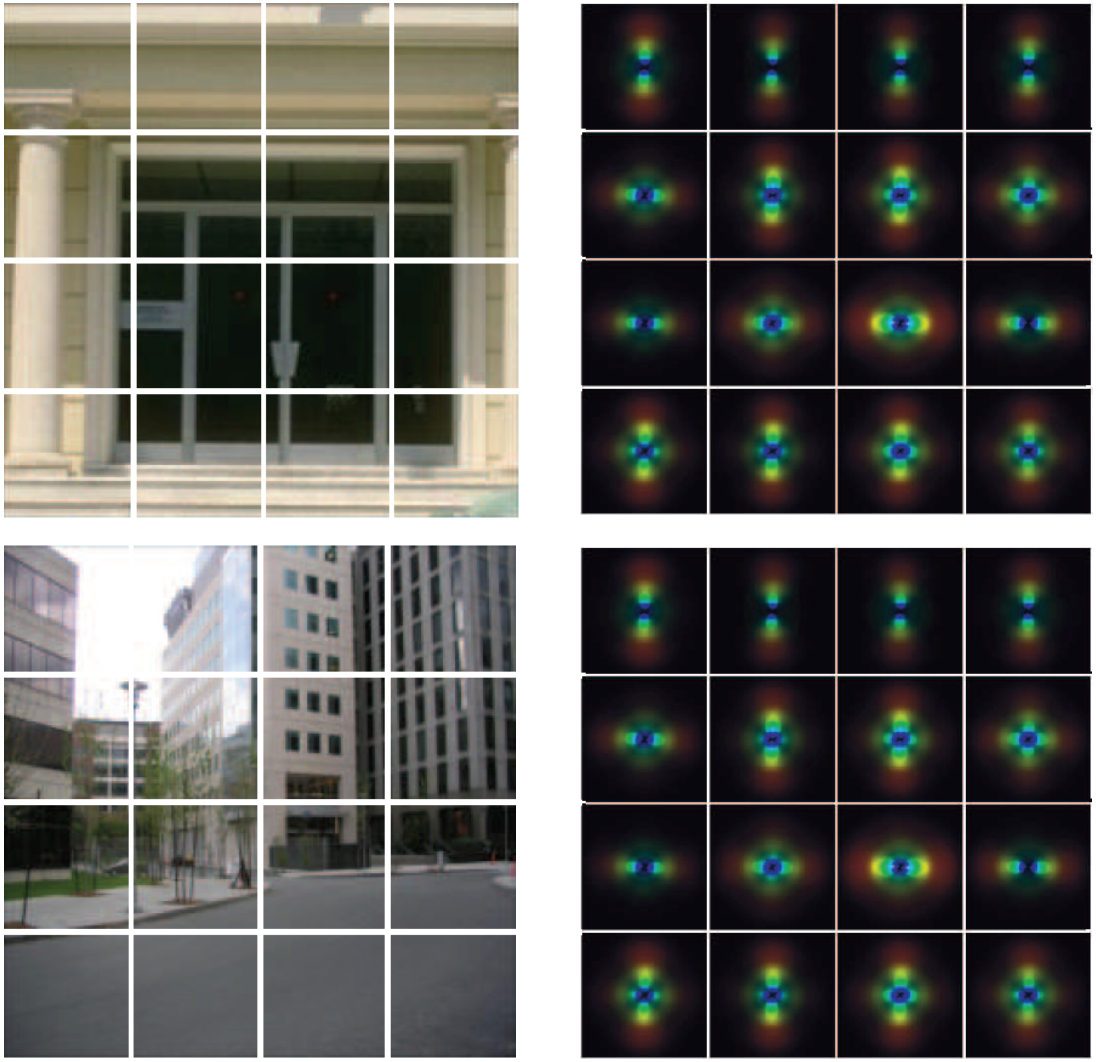
- Choose random scene and seven neighbors in scene space
- If >= 4 neighbors have same semantic category, image is “correctly recognized”
  - WDST: 92%
  - DST: 86%

Applications

- Depth Estimation (Torralba & Oliva)



Gist descriptor



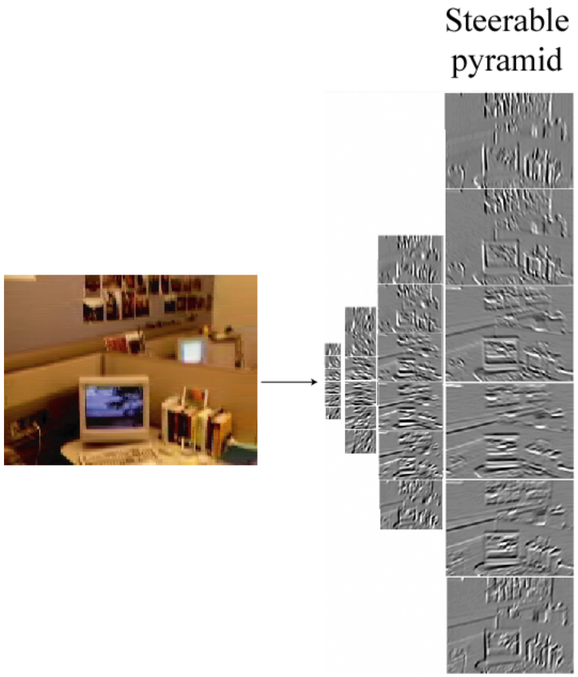
Oliva and Torralba, 2001

8 orientations
4 scales
x 16 bins
512 dimensions

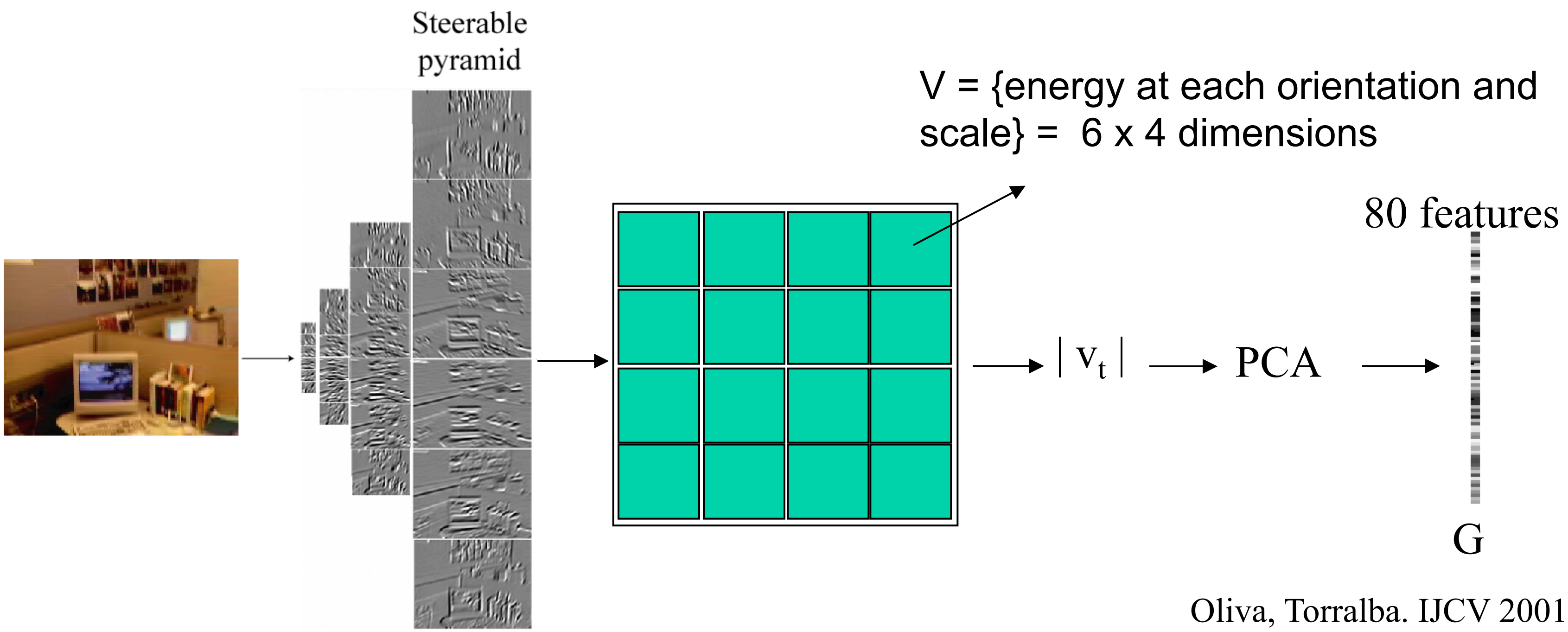
Similar to SIFT (Lowe 1999) applied to the entire image

M. Gorkani, R. Picard, ICPR 1994; Walker, Malik. Vision Research 2004; Vogel et al. 2004; Fei-Fei and Perona, CVPR 2005; S. Lazebnik, et al, CVPR 2006; …

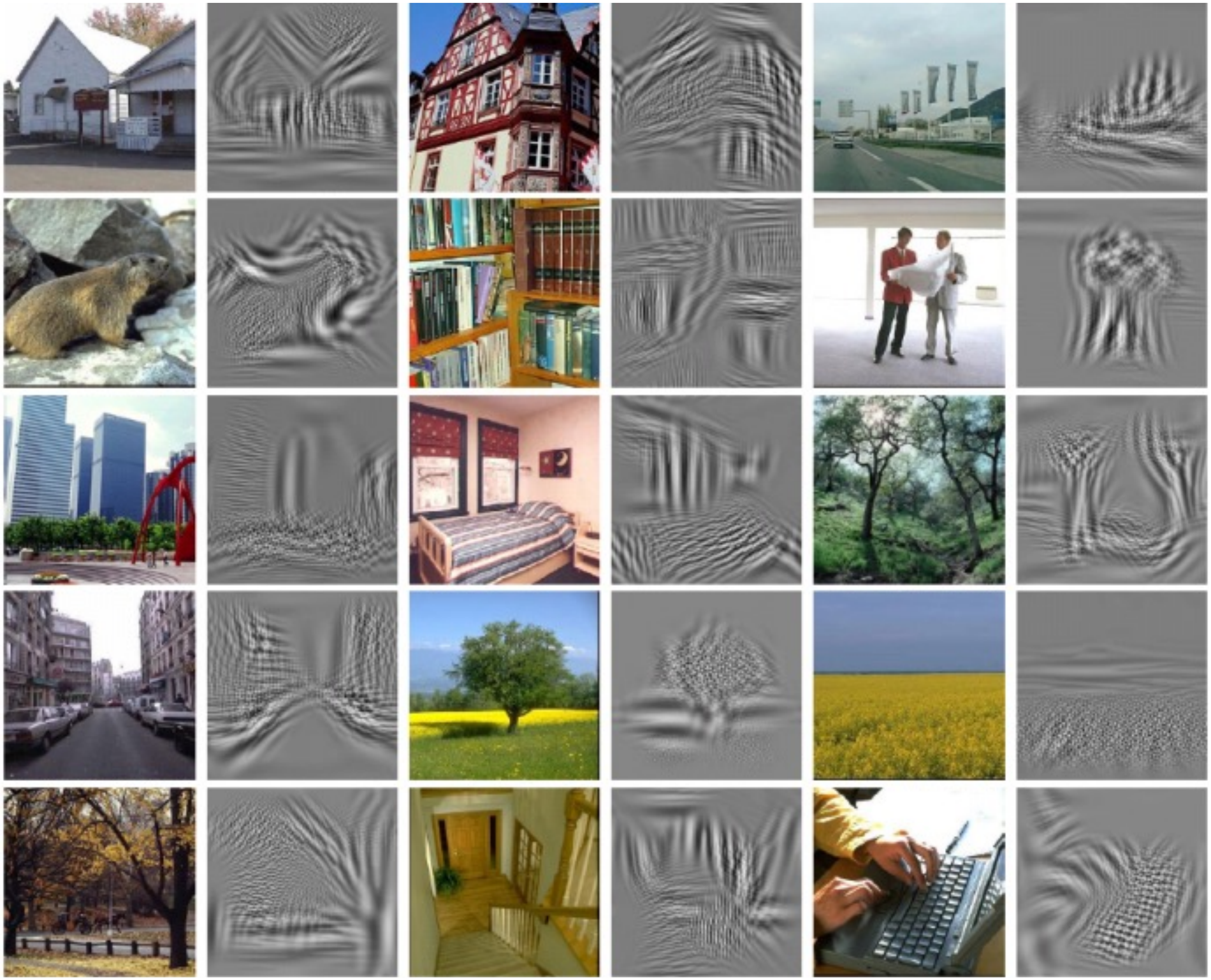
Gist descriptor



Gist descriptor



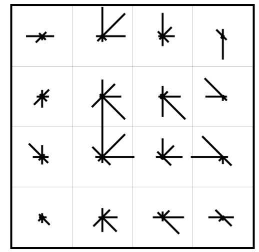
Example visual gists

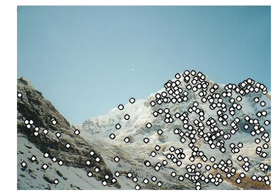


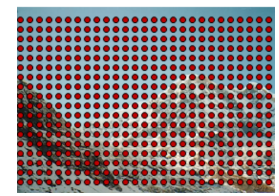
Oliva & Torralba (2001)

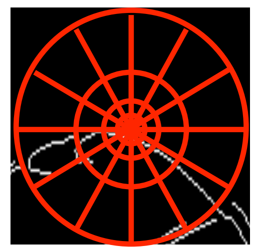
Features

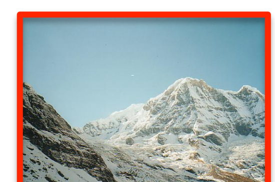
- Where:
  - Interest points
    - Corners
    - Blobs
  - Grid
  - Spatial Pyramids
  - Global











- What: (Descriptors)
  - Sift, HOG
  - Shape Context
  - Bag of words
  - Filter banks