

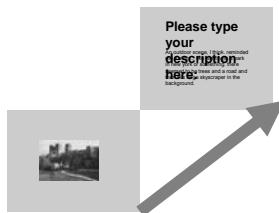
# Natural Scene Categorization: from Humans to Computers

Li Fei-Fei

Beckman Institute, ECE, Psychology



#1: natural scene categorization entails little attention (**Rufin VanRullen, Christof Koch, Pietro Perona**)

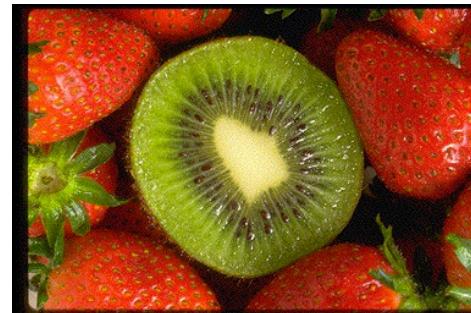


#2: what can we perceive within a glance of a scene – a working definition for ‘gist’ (**Asha Iyer, Christof Koch, Pietro Perona**)



#3: local patches, and some intermediate level information – a hierarchical Bayesian algorithm for natural scene categorization (**Pietro Perona**)

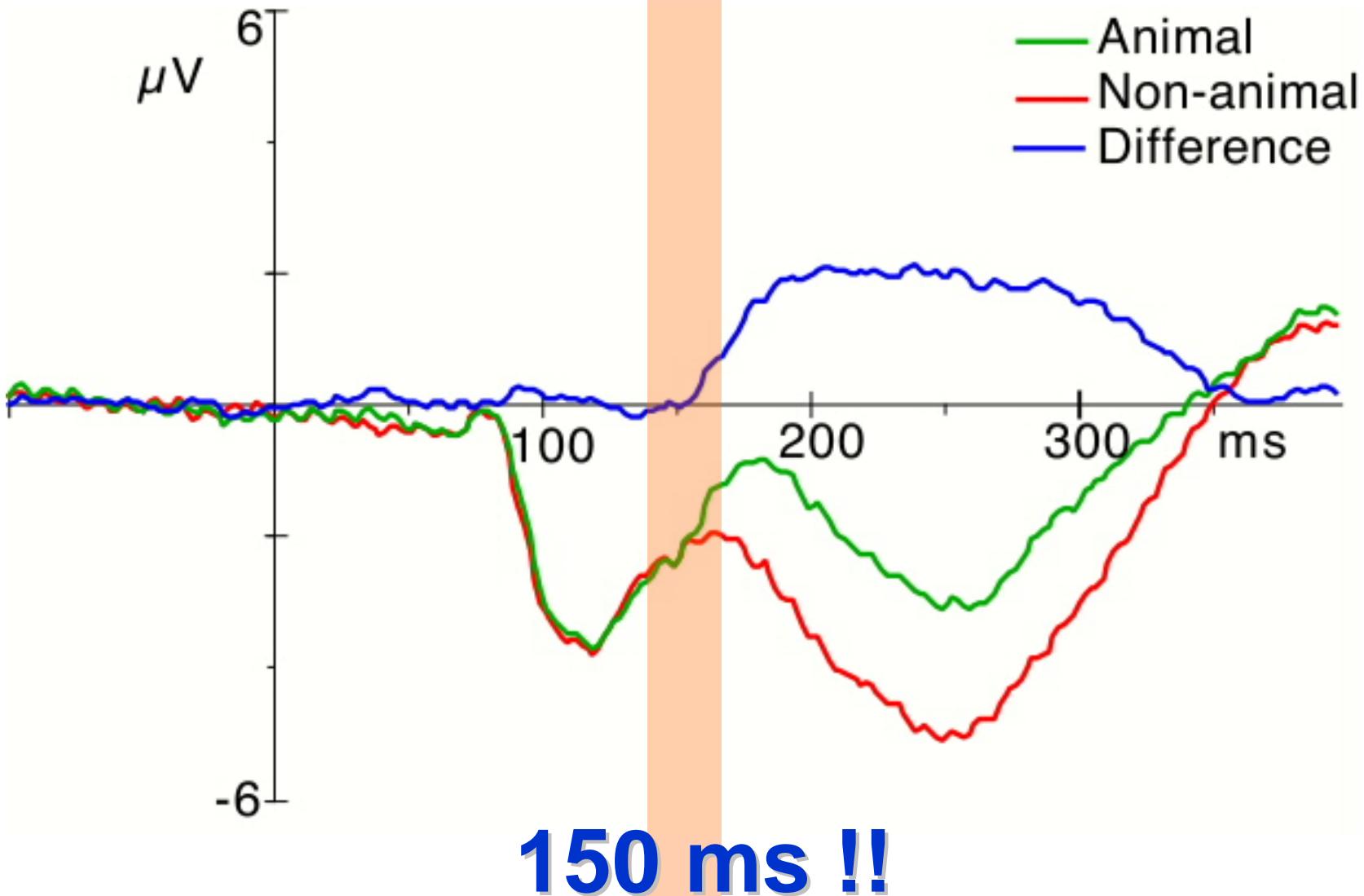
- #1: natural scene categorization entails little attention



Reference: Li et al. 2002; Fei-Fei et al. 2005



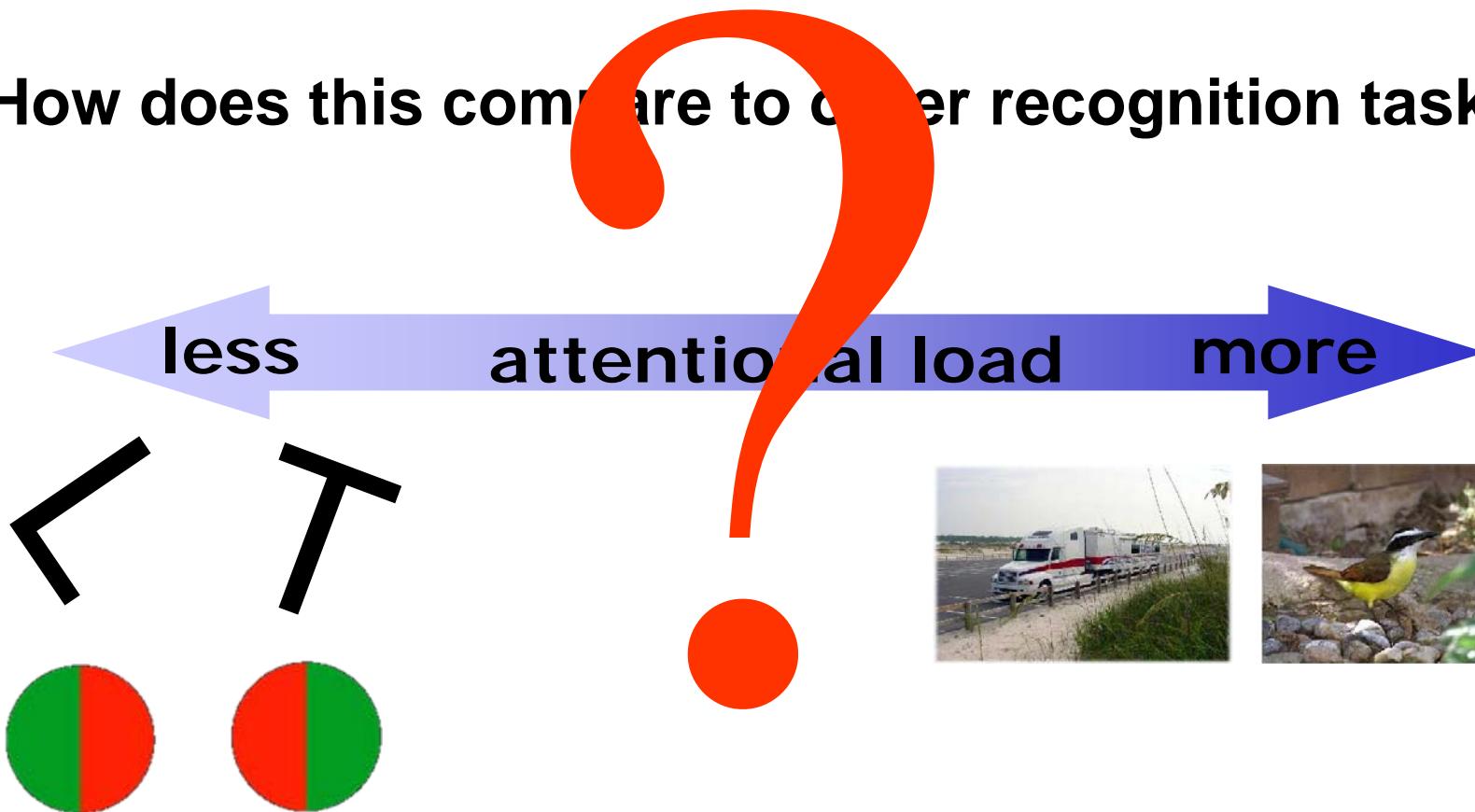
Thorpe, et al 1996

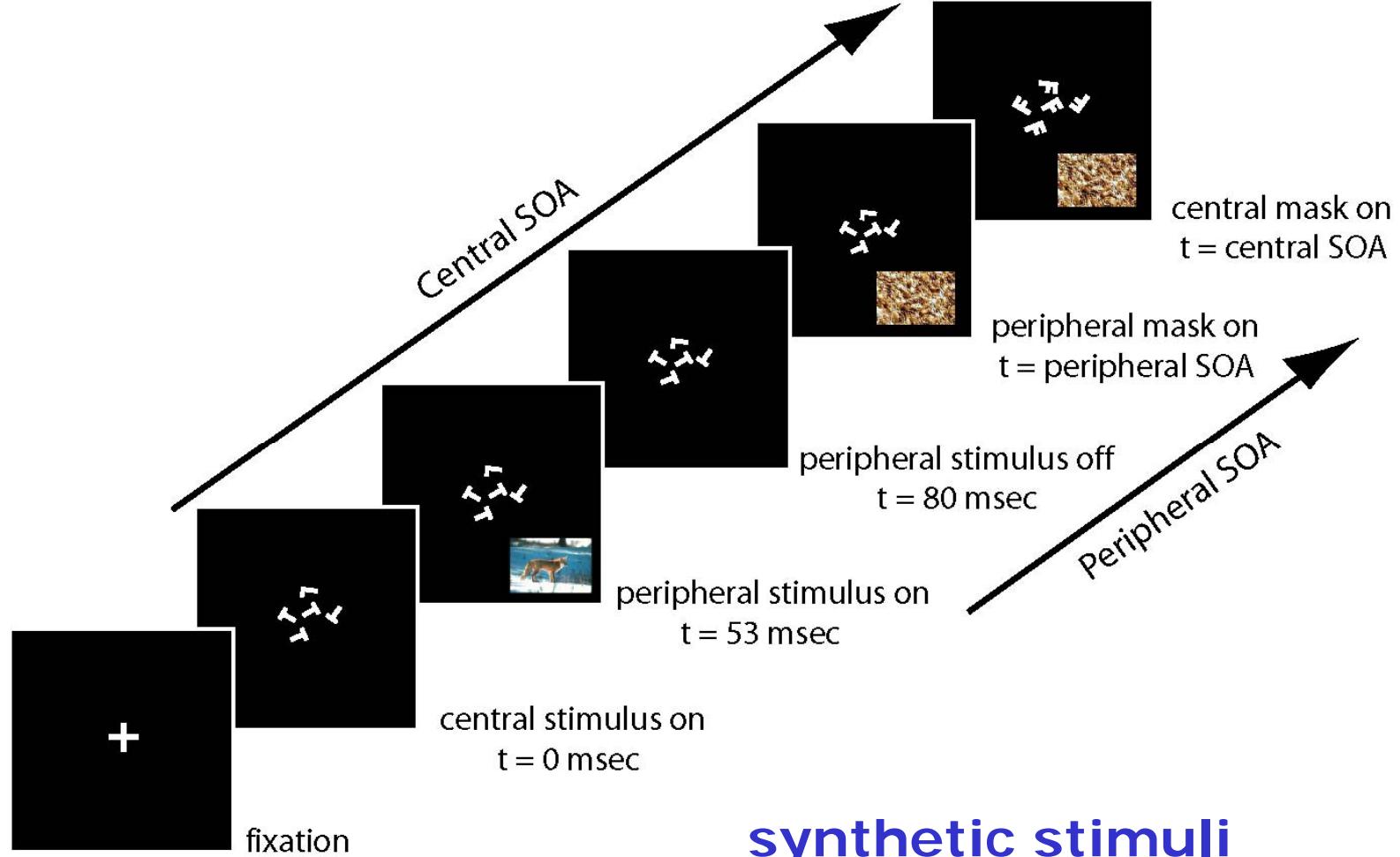


Thorpe, et al 1996

# Our question

1. How critical is attention in natural scene recognition?
2. How does this compare to other recognition tasks?





**animals**



**vehicle**

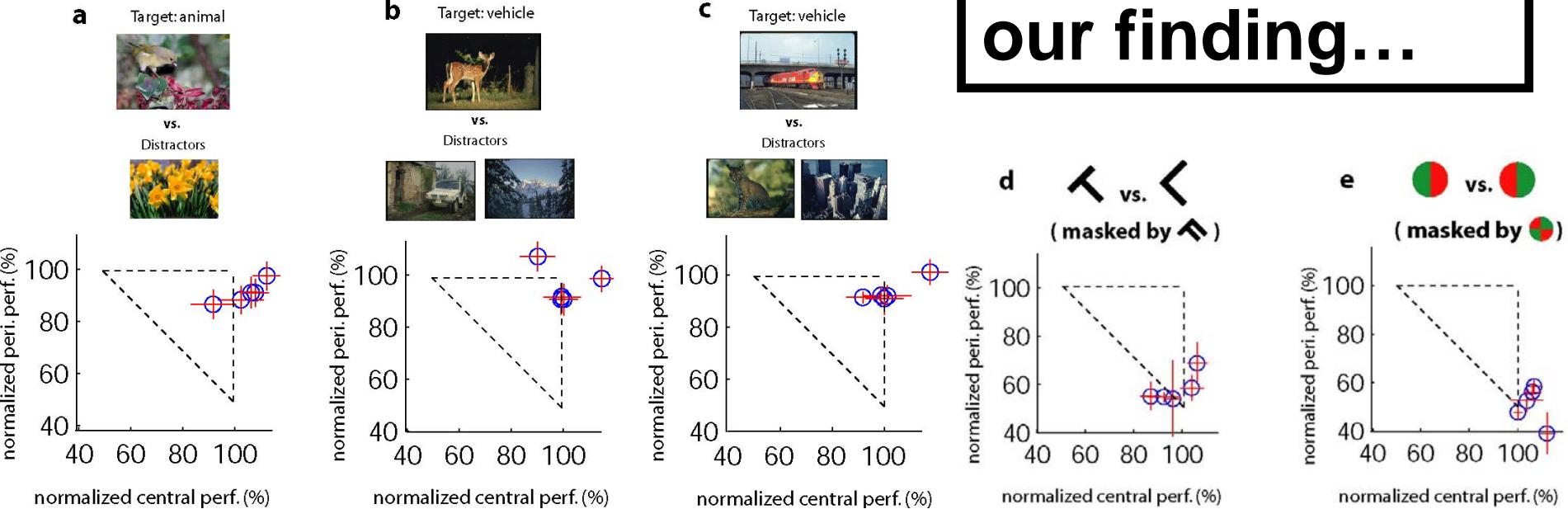


## synthetic stimuli



Li et al. 2002

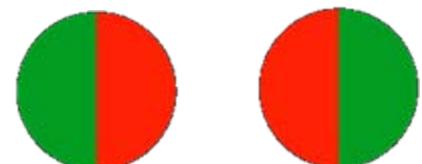
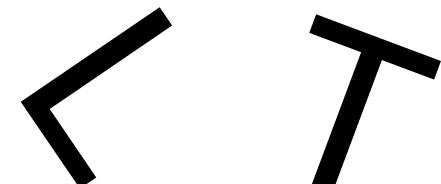
# our finding...



less

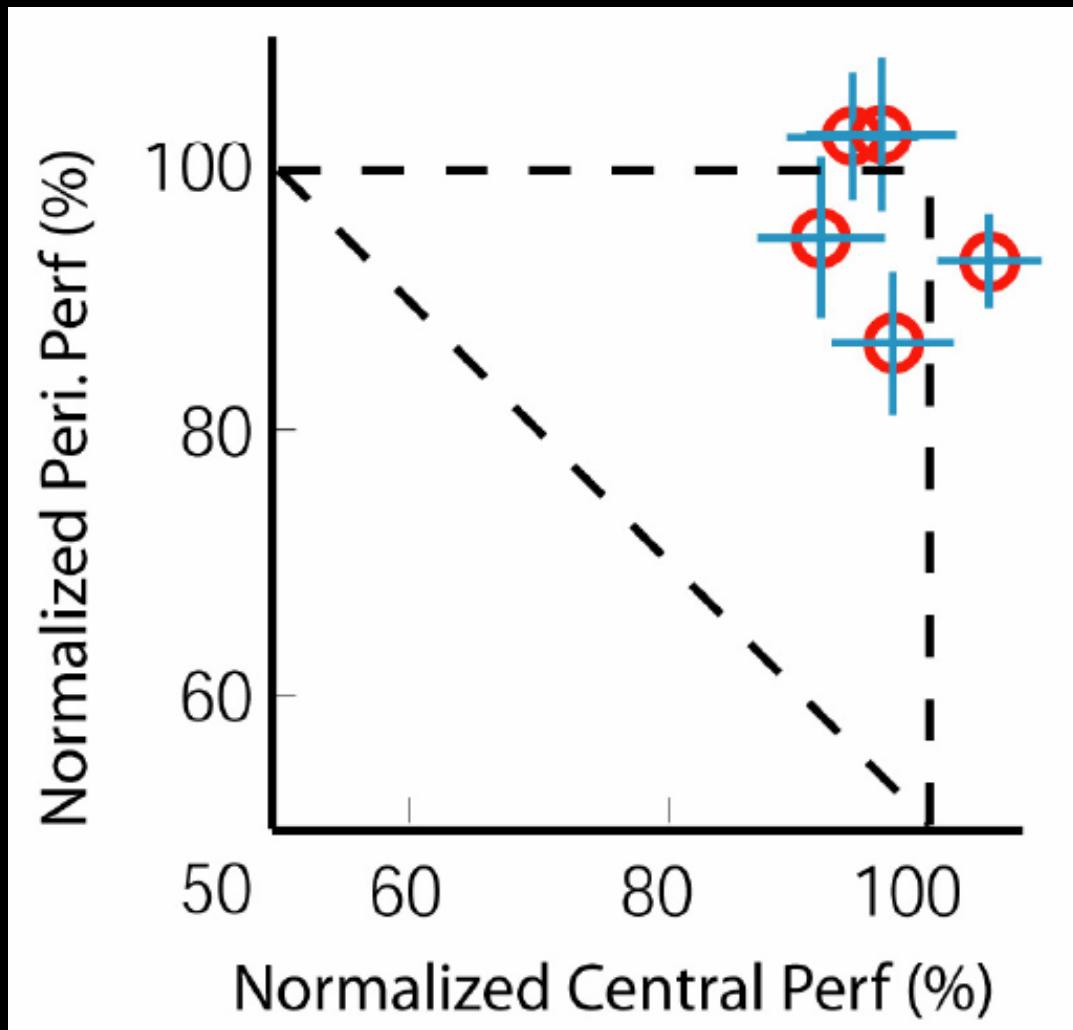
attentional load

more



Li et al. 2002

# Without color...



# Effect of “meaningful” category

randomly rotated

Target      Distractor



vs.



( masked by  )

fixed rotation

Target      Distractor



vs.



( masked by  )

upright position

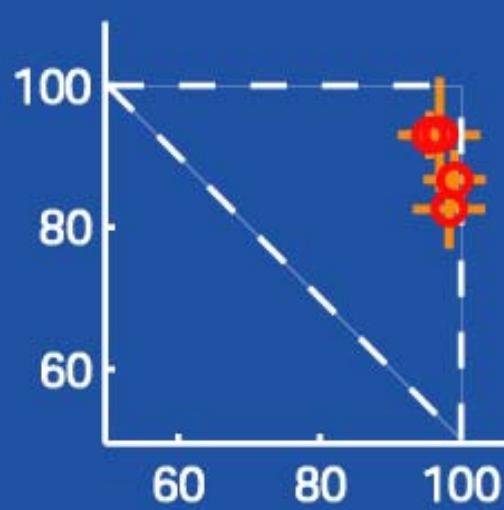
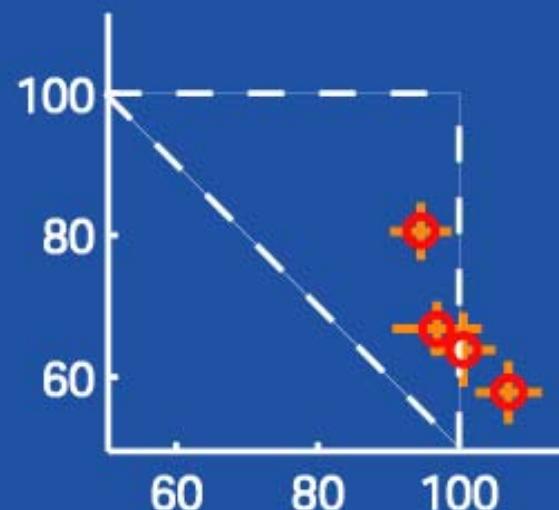
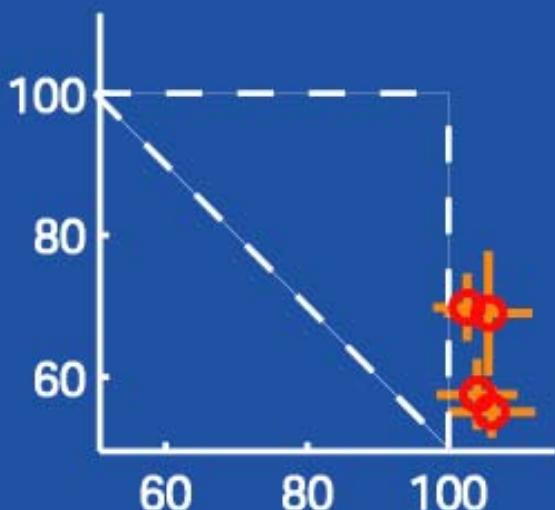
Target      Distractor



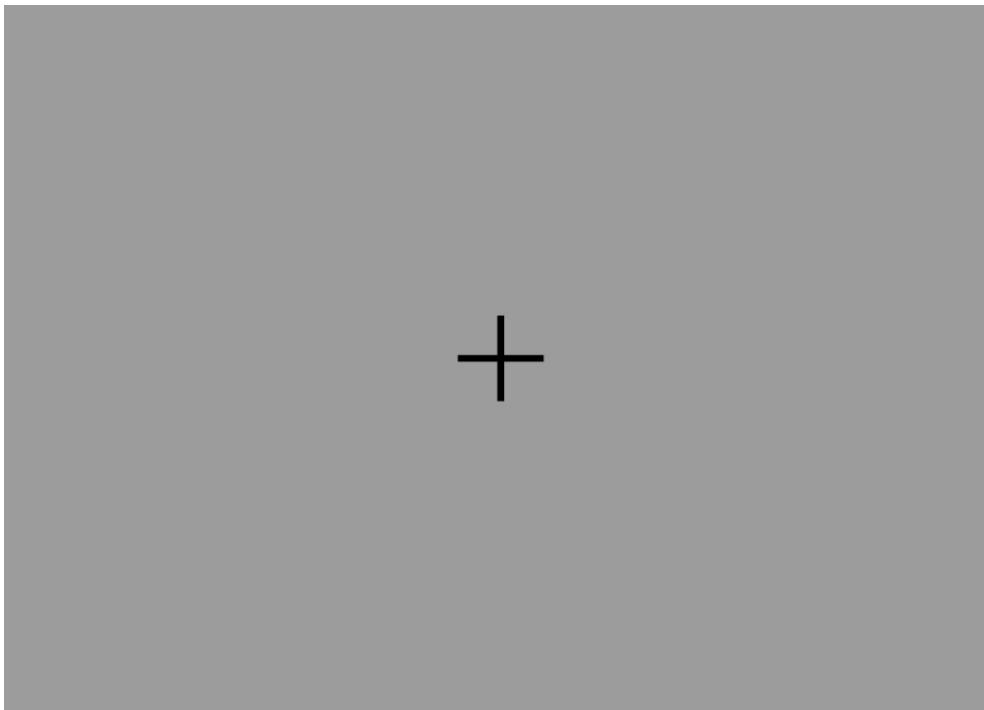
vs.



( masked by  )



- #2: what can we perceive within a glance of a scene – a working definition for ‘gist’



Reference: Fei-Fei et al. submitted

# What is “gist”?

- sensory data, e.g. “color”,  
“size”, etc.



# What is “gist”?

- sensory data, e.g. “color”, “size”, etc.
- “inventory of some of the objects (and textures)”



water



sand



sky



palm  
tree

# What is “gist”?

- sensory data, e.g. “color”, “size”, etc.
- “inventory of some of the objects”
- “some relationships between objects”



# What is “gist”?

- sensory data, e.g. “color”, “size”, etc.
- “inventory of some of the objects”
- “some relationships between objects”
- “layout”



# What is “gist”?

- sensory data, e.g. “color”, “size”, etc.
- “inventory of some of the objects”
- “some relationships between objects”
- “layout”
- “stuffness”



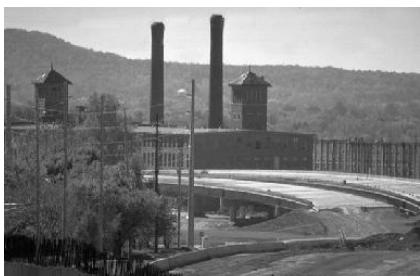
# What is “gist”?

- sensory data, e.g. “color”, “size”, etc.
- “inventory of some of the objects”
- “some relationships between objects”
- “layout”
- “stuffness”
- scene category



# What do people see in a glance?





# Stage I: Collect Image Description

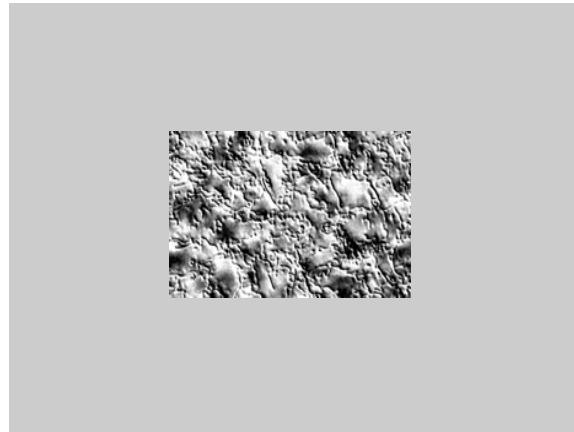
--- Illustration of 1 Trial

Subject types freely what he/she saw in the image



Please type your description here:

An outdoor scene, I think. reminded me a a city... like walkingin a park in new york or something. there seemed to be trees and a road and then this large skyscraper in the background.



time

Mask onset:  $t = SOA$



1 of 7 possible SOA's (msec):  
27, 40, 53, 67, 80, 120, 500

Image onset:  $t = 0 \text{ msec}$

## **PT = 27ms**

Couldn't see much; it was mostly dark w/ some square things, maybe furniture. (Subject: AM)

## **PT = 40ms**

This looked like an indoor shot. Saw what looked like a large framed object (a painting?) on a white background (i.e., the wall). (Subject: RW)

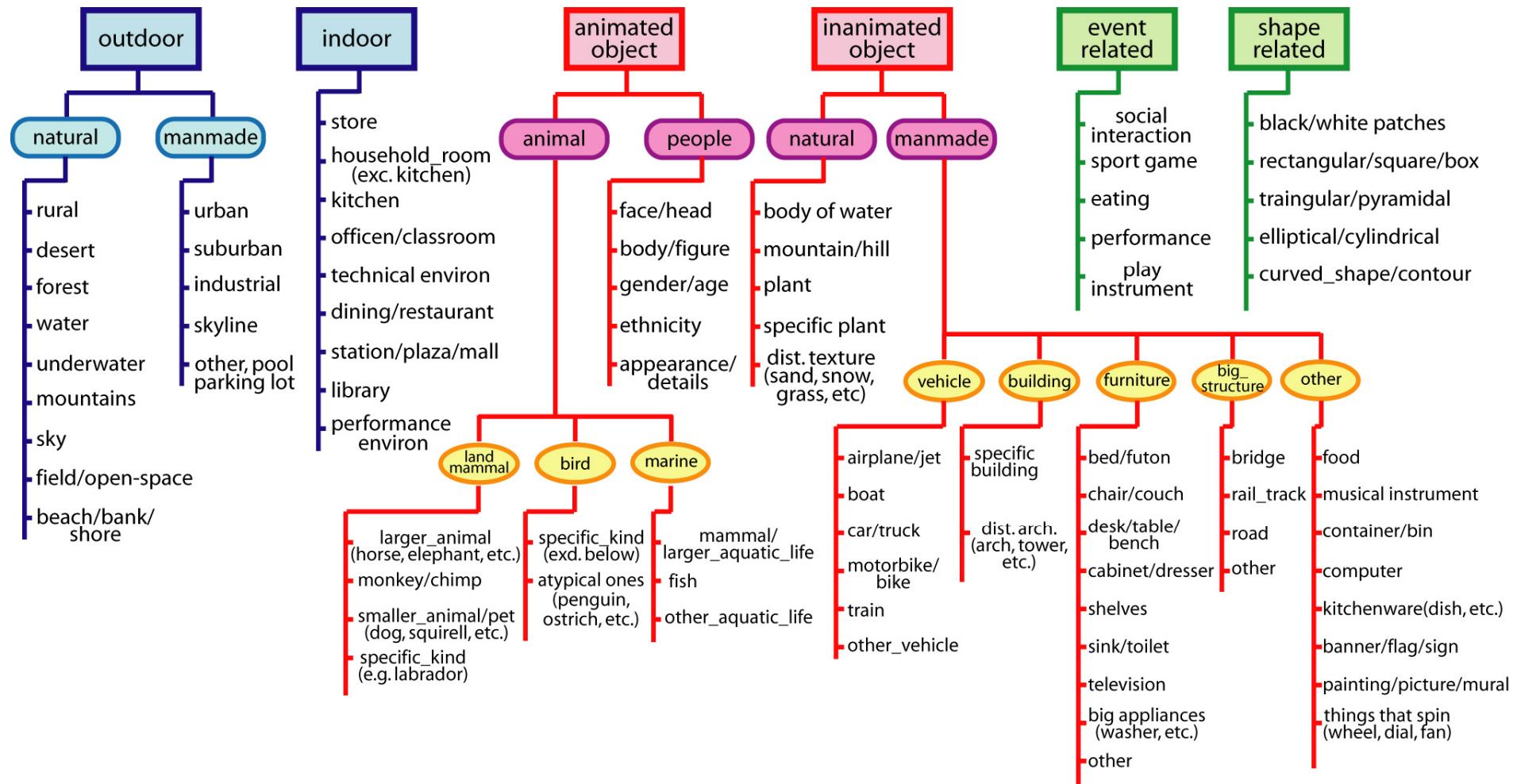
## **PT = 500ms**

This is indoors. It's must be a rich person's house. There are many paintings on the wall. The largest painting might have a fireplace beneath it. I think the largest painting was that of a man standing erect. The room is richly decorated and it looks like one of the rooms in Mr. Darcy's house in the A&E movie Pride and Prejudice. Or maybe it more closely resembles one of the rooms where the one of the rooms in Huntington's house (at the Huntington).

## **PT = 67ms**

I saw the interior of a room in a house. There was a picture to the right, that was black, and possibly a table in the center. It seemed like a formal dining room. (Subject: JB)

# The tree



## Response No. 18 for Image No. 4



I could make out some kind of circular shapes near the bottom of the picture. These reminded me of those round life preservers that are on ships. There was also a man standing on top of some wooden structure.

## CATEGORY: SENSORY/SHAPES

Please select one of "correct" or "incorrect" for each checked description. Click "Next>>" to continue

black/white\_patches

described  correct  incorrect

rectangular/square/box

described  correct  incorrect

triangular/pyramidal

described  correct  incorrect

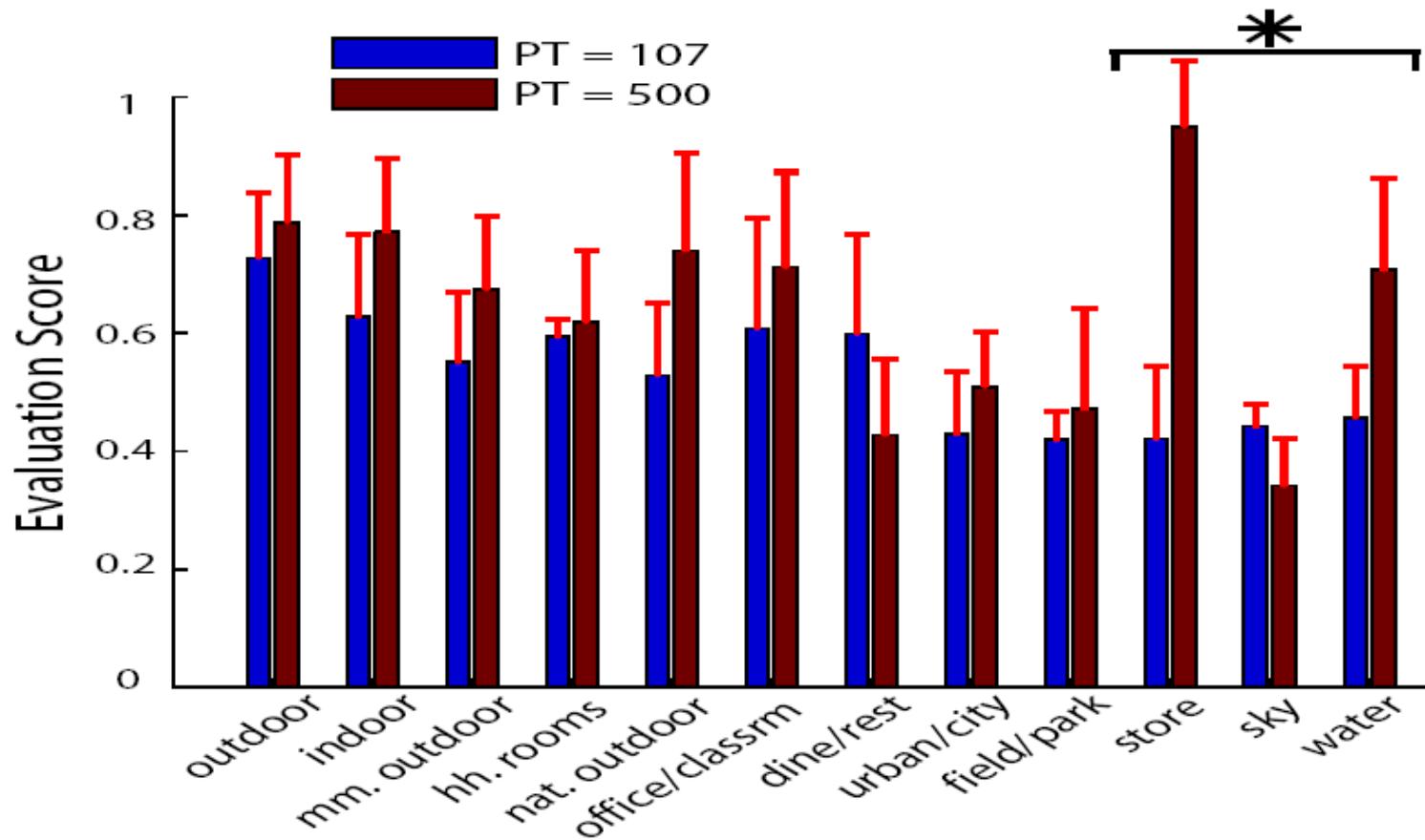
elliptical/cylindrical(eg.round.blob)

described  correct  incorrect

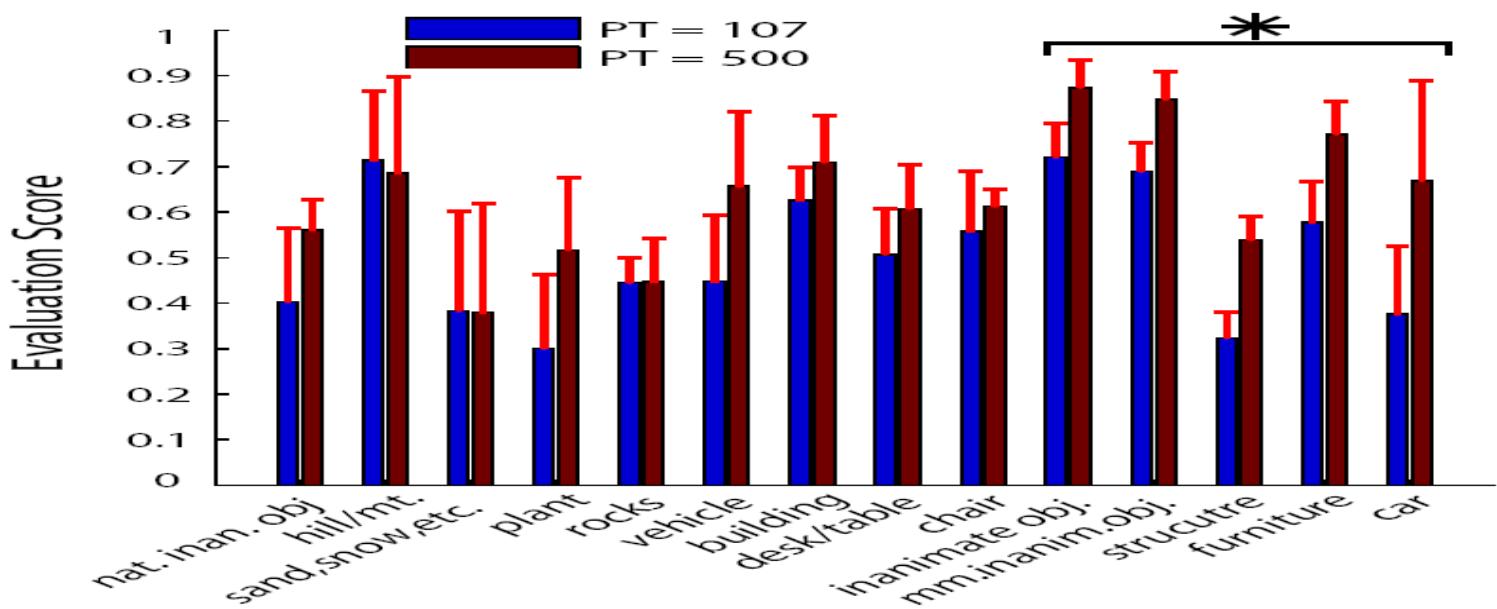
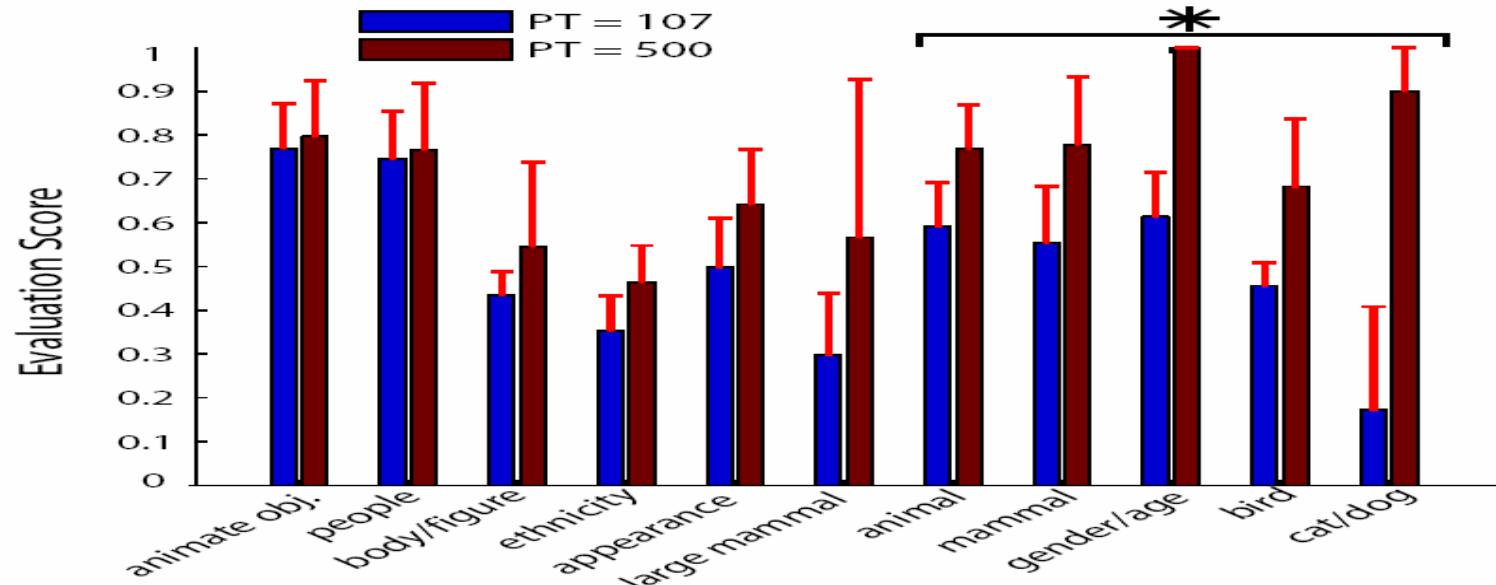
curved\_shape/contour(eg.arc,'S')

described  correct  incorrect

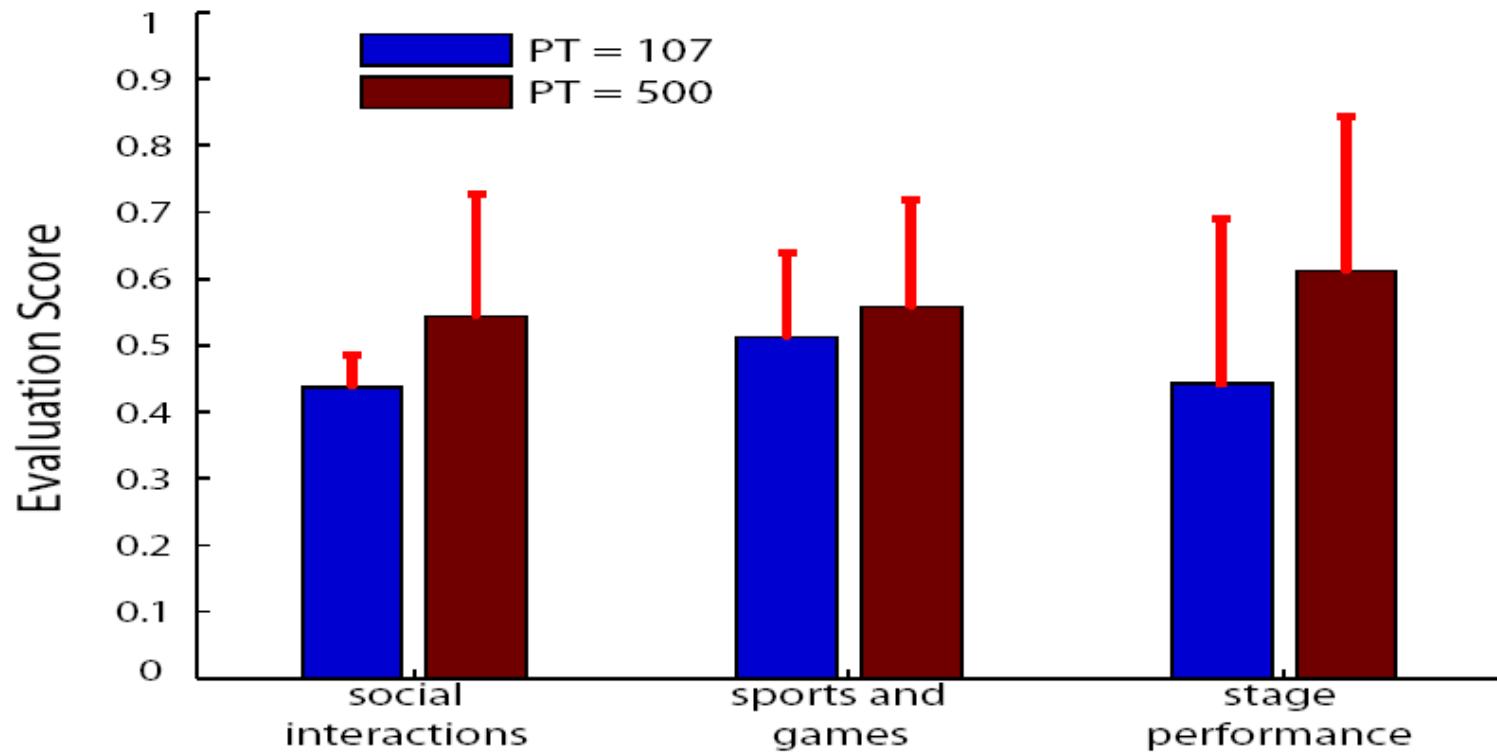
# Scene level



# Object level



# (Social) Events



- #3: local patches, and some intermediate level information – a hierarchical Bayesian algorithm for natural scene categorization

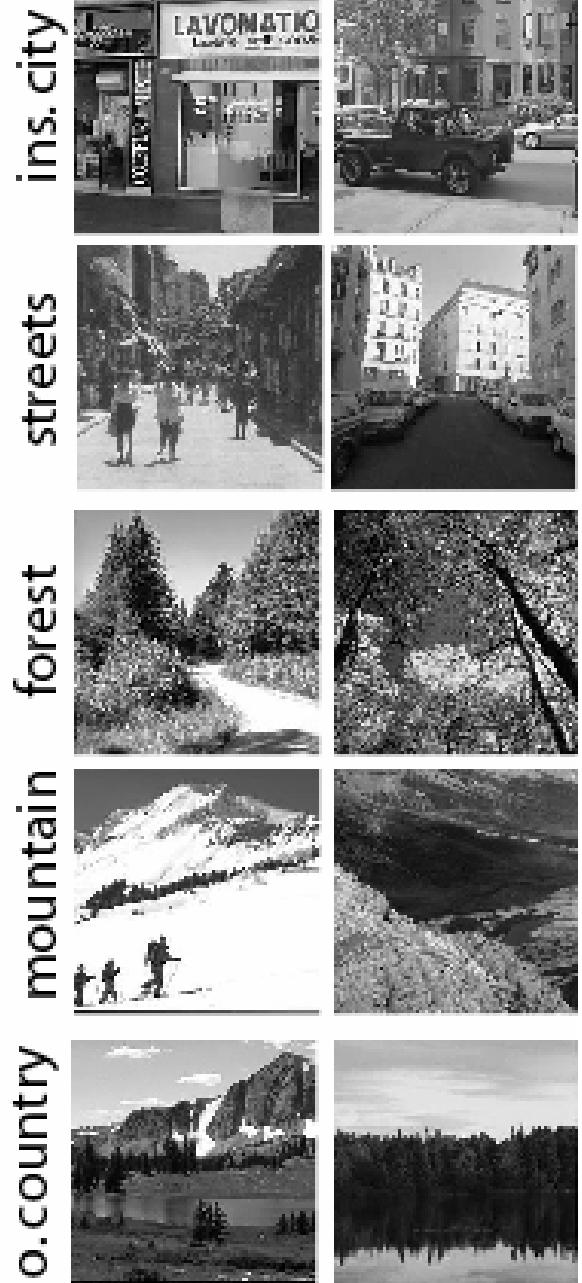


Reference: Fei-Fei et al. CVPR 2005

office      kitchen      livingroom      bedroom



coast      suburb      tall bldg      highway



- **global cues: colors, textures, etc.**

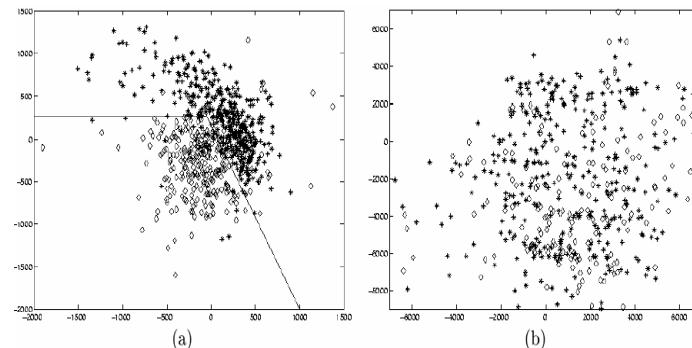
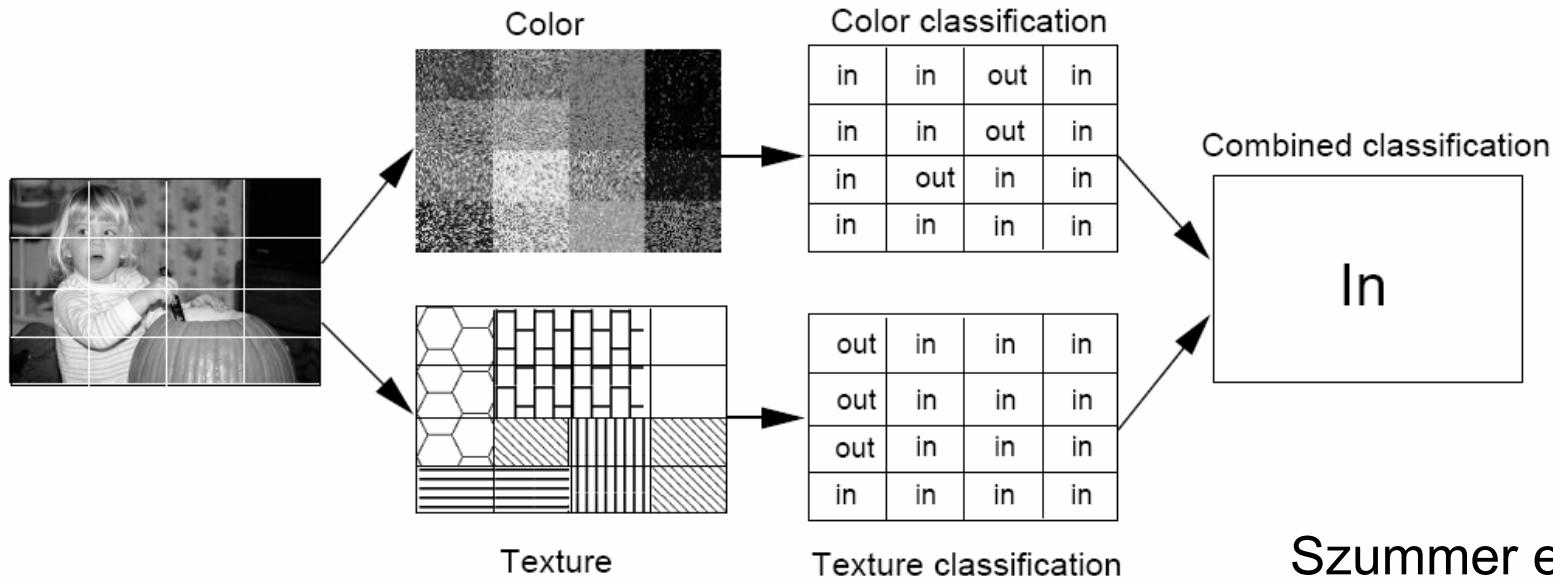


Figure 7: 2D Feature space plots showing (a) edge direction coherence vector and (b) color coherence vectors; \* represents the landscape patterns and  $\diamond$  represents the city patterns; only a subset of 2,716 patterns have been plotted here for clarity of display.

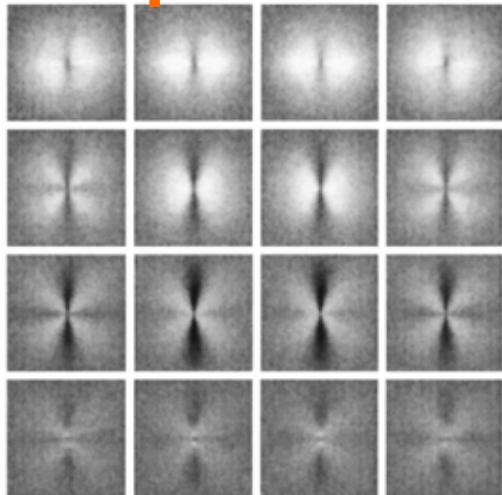
Jain, Zhang et al. (1998)



Szummer et al. (1998)

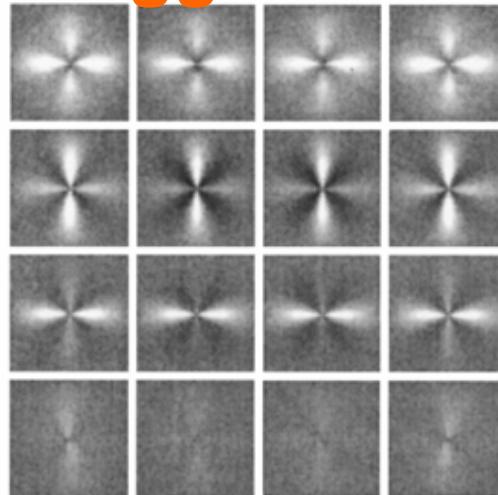
- **global cues: frequency**

openness



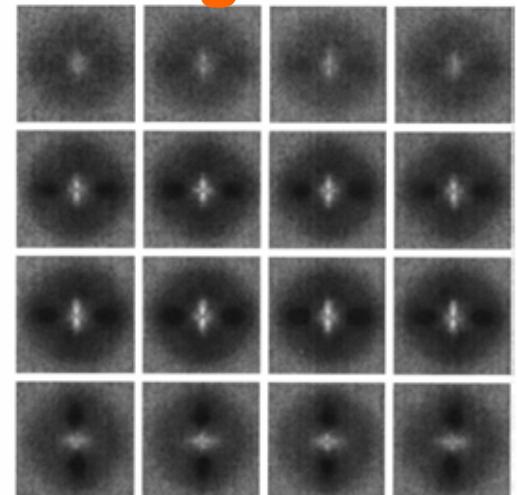
a)

ruggedness

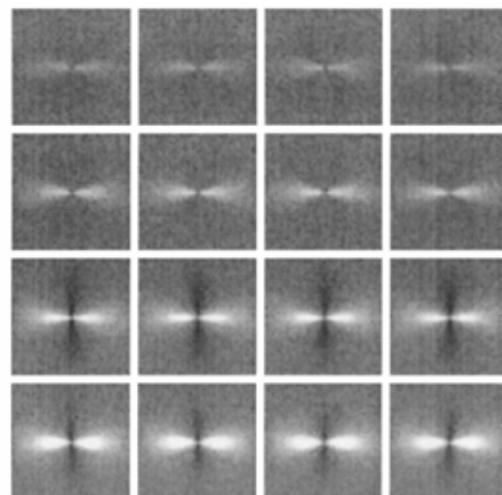


b)

roughness

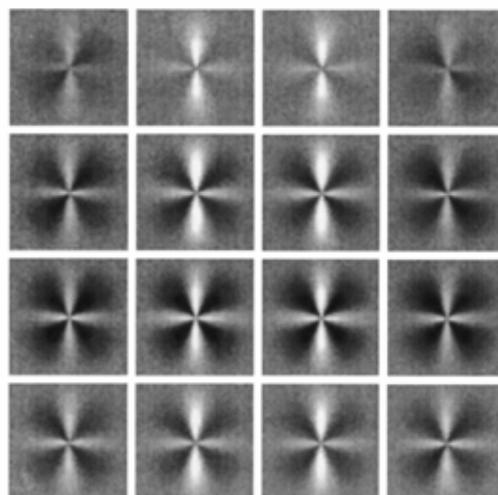


c)



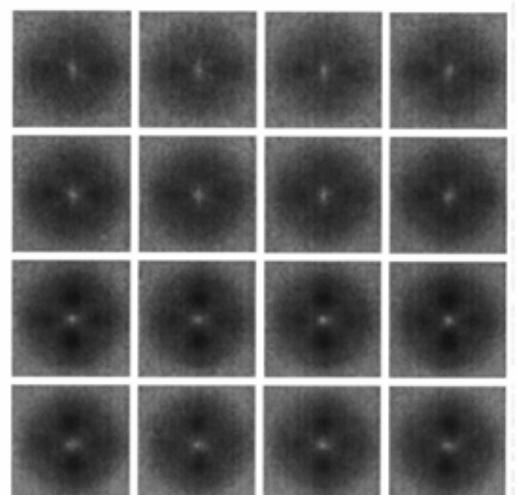
d)

openness



e)

expansion



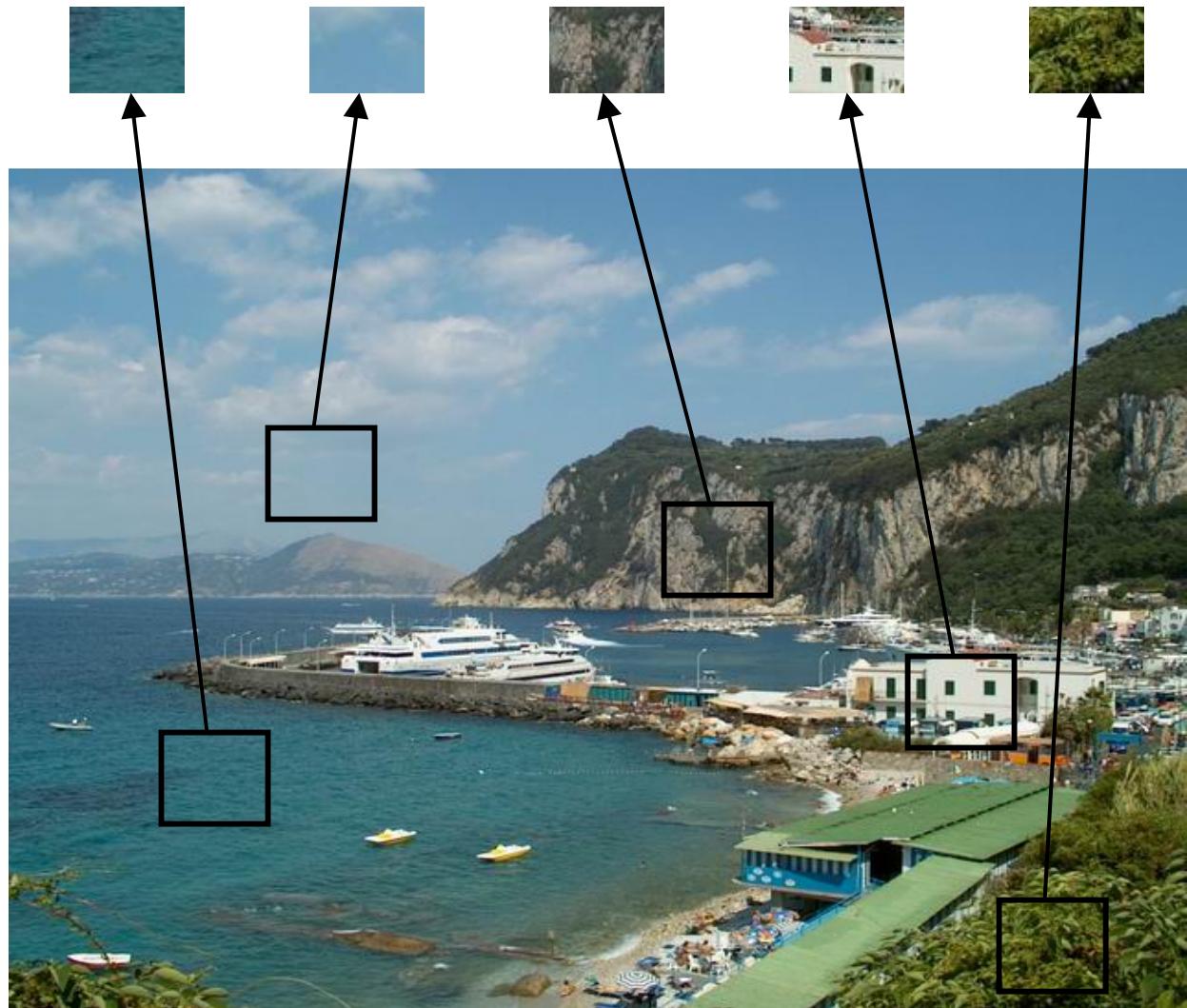
f)

roughness

Oliva, Torralba (1999, 2001)

# local patch based idea

| Concept | Occurrences |
|---------|-------------|
| sky     | 14.0%       |
| water   | 32.5%       |
| grass   | 0.0%        |
| trunks  | 0.0%        |
| foliage | 6.5%        |
| fields  | 0.0%        |
| rocks   | 31.0%       |
| flowers | 0.0%        |
| sand    | 16.0%       |



# Our intuitions

- local patch based

rocks



foliage



sand



sky

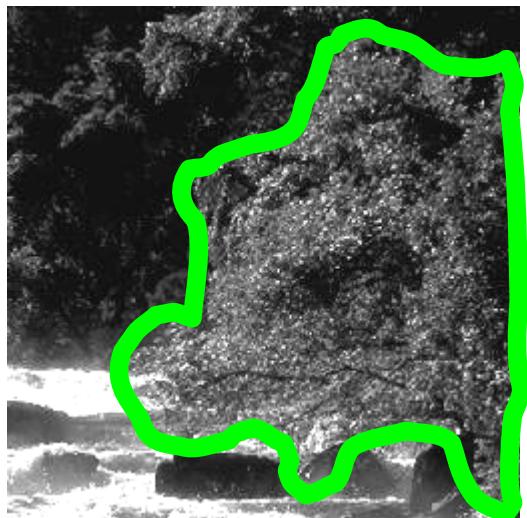


water



- intermediate level themes within scenes

forest



suburb



inside of city



# Our intuitions

- local patch based



- intermediate level themes



- weakly supervised

- no human annotation of local patches and intermediate level themes

**Image**

**Bag of ‘words’**



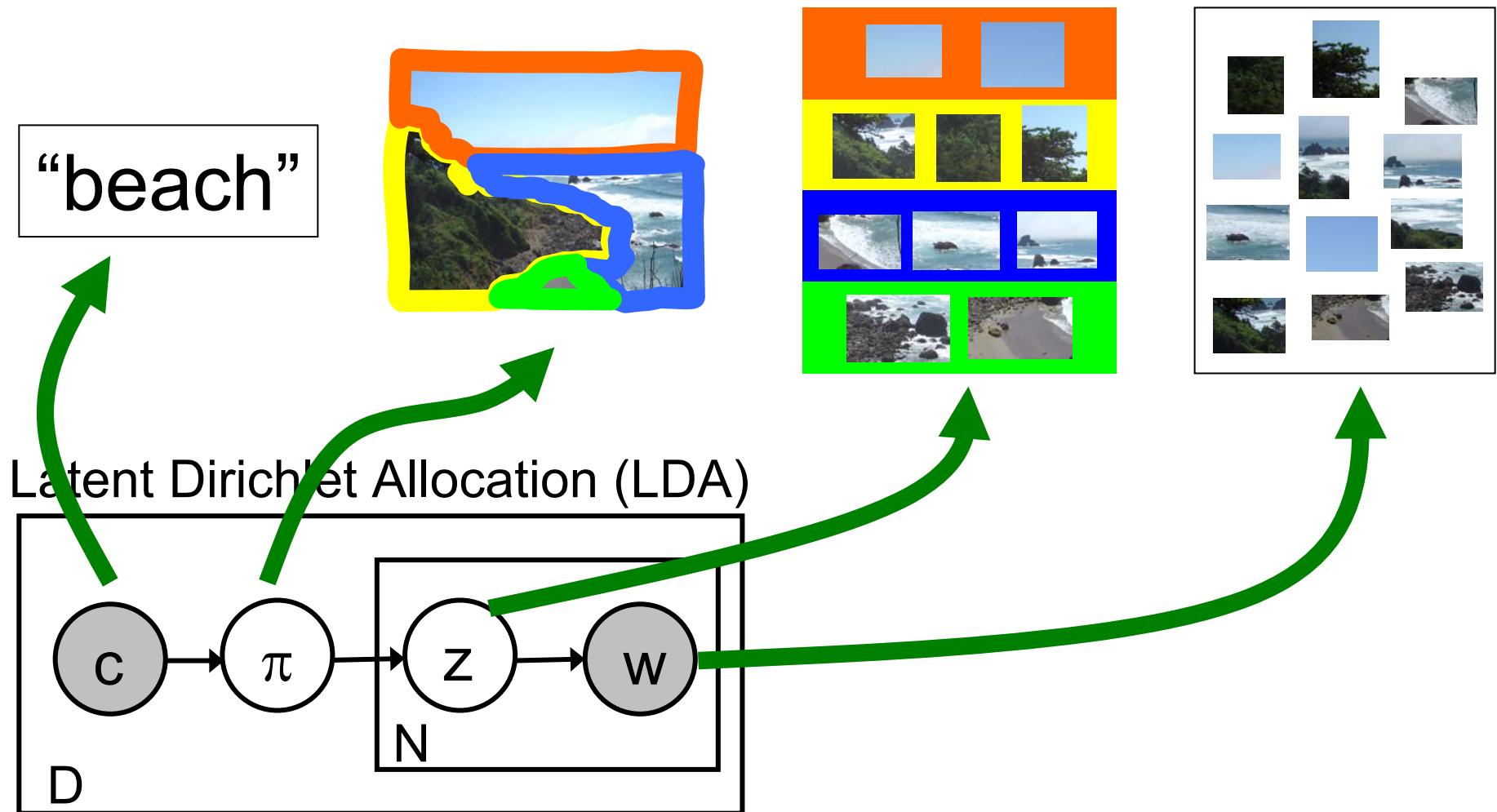
# Analogy to documents

Of all the sensory impressions proceeding to the brain, the visual experiences are the dominant ones. Our perception of the world around us is based essentially on the messages that reach us from our eyes. For a long time it was thought that a retinal image was processed by visual centers in the brain. As a movie screen is a visual image, it was discovered that the eye, cell, optical nerve, image and Hubel, Wiesel following the visual pathway to the various centers of the cerebral cortex, Hubel and Wiesel have demonstrated that the message about the image falling on the retina undergoes top-down analysis in a system of nerve cells stored in columns. In this system each column has its specific function and is responsible for a specific detail in the pattern of the retinal image.

China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% increase in exports to \$750bn, compared with \$660bn. This will annoy the US, which China's leaders deliberately agree to. The yuan is governed by the central bank, which also needs to meet demand so it can't let the country. China has allowed the value of the yuan against the dollar to rise and permitted it to trade within a narrow band, but the US wants the yuan to be allowed to trade freely. However, Beijing has made it clear that it will take its time and tread carefully before allowing the yuan to rise further in value.

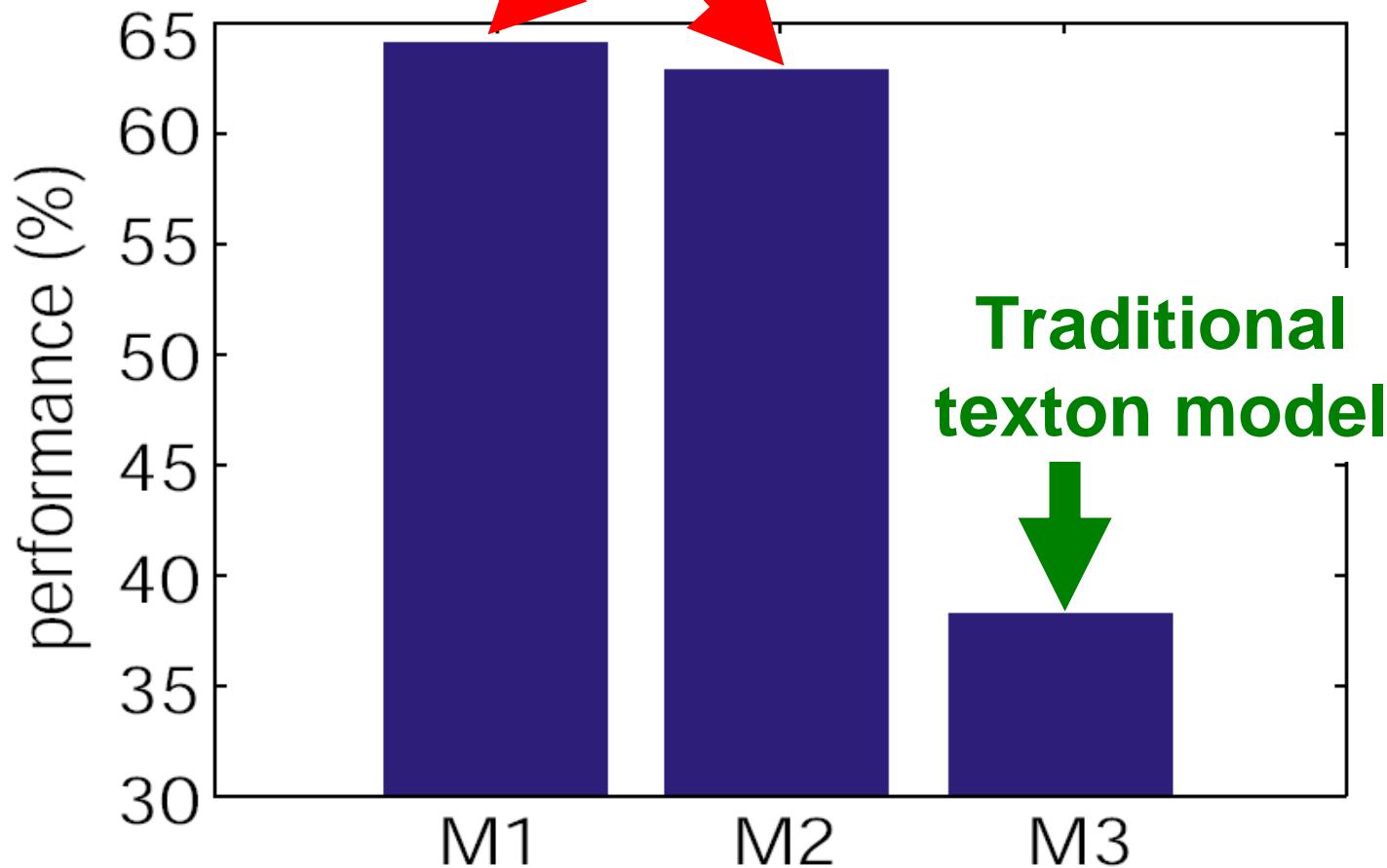




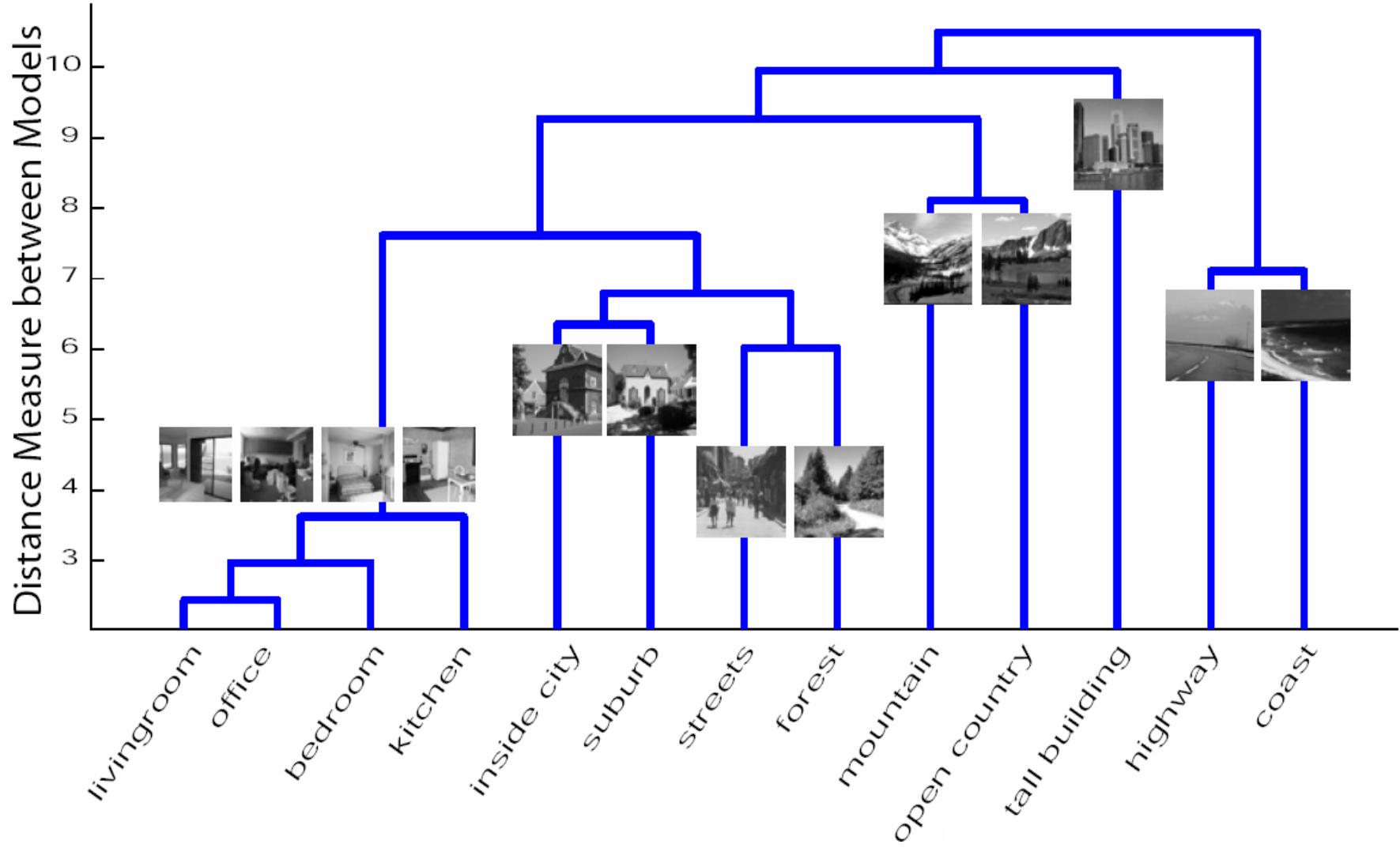


| <b>highway</b>       | 74 | 2         |           | 2         | 2         |           | 14        | 4         |           | 2         |           |
|----------------------|----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| <b>insidecity</b>    |    | <b>58</b> | 10        | 6         | 8         |           | 4         |           | 2         | 6         | 4         |
| <b>tallbuildings</b> |    | 4         | <b>76</b> | 10        |           |           |           | 4         | 4         |           | 2         |
| <b>street</b>        | 2  | 4         | 6         | <b>78</b> |           | 2         |           | 2         | 2         |           | 4         |
| <b>suburb</b>        |    |           |           |           | <b>94</b> |           |           |           | 2         |           | 4         |
| <b>forest</b>        |    |           |           |           |           | <b>88</b> |           | 12        |           |           |           |
| <b>coast</b>         | 2  |           |           |           |           |           | <b>78</b> |           | 20        |           |           |
| <b>mountain</b>      | 4  |           | 4         |           | 2         | 6         | 8         | <b>70</b> | 6         |           |           |
| <b>opencountry</b>   | 8  |           |           |           | 8         | 10        | 16        | 10        | <b>48</b> |           |           |
| <b>bedroom</b>       | 4  | 2         | 2         |           | 2         | 2         | 2         | 4         |           | <b>28</b> | 12        |
| <b>kitchen</b>       |    | 8         | 2         |           |           |           | 2         |           |           | <b>60</b> | 14        |
| <b>livingroom</b>    |    | 2         | 2         | 2         |           |           | 2         | 4         |           | 4         | 18        |
| <b>office</b>        |    |           |           |           | 2         |           | 2         |           |           | 8         | 12        |
|                      |    |           |           |           |           |           |           |           |           | 12        | <b>64</b> |

**Our model**

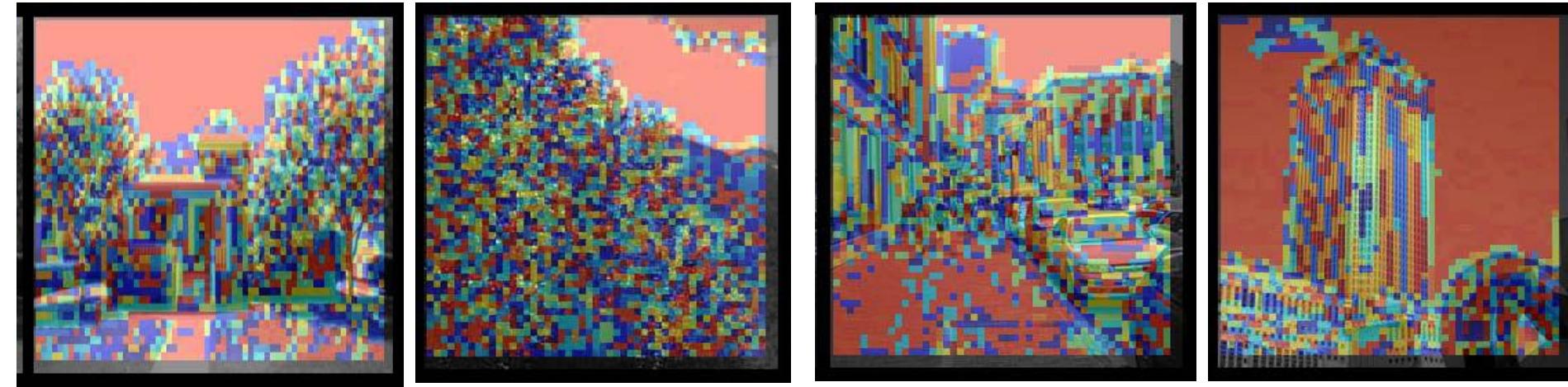


# model distance based on theme distribution

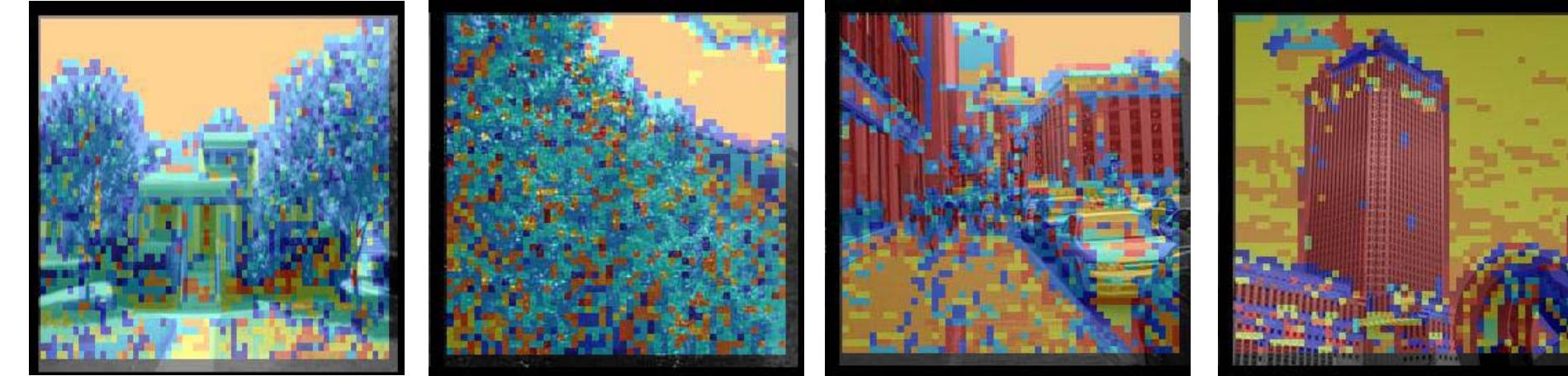


# segmentation by themes

codeword images



theme images



# Summary

- natural scene categorization entails little attention
- ‘gist’ of a scene includes much information on objects, scenes and beyond
- propose a hierarchical Bayesian algorithm for natural scene categorization using local patches

# references

- F.F. Li, R. VanRullen, C. Koch and P. Perona. Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci.* 99, 8378 - 8383, 2002.
- L. Fei-Fei and P. Perona. A Bayesian Hierarchical Model for Learning Natural Scene Categories. *IEEE Comp. Vis. Patt. Recog.* 2005
- L. Fei-Fei, R. VanRuellen, C. Koch and P. Perona. Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition*. 12(6): pp893-924. 2005
- L. Fei-Fei, A. Iyer, C. Koch and P. Perona. What do we see in a glance of a scene? Submitted.