

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)

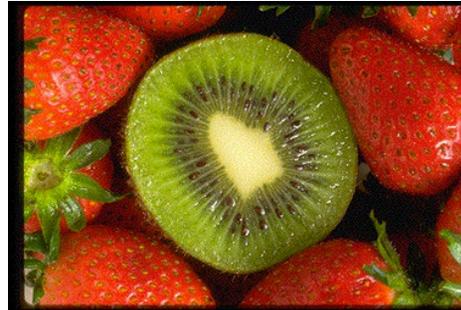


#2: what can we perceive within a glance of a scene – a working definition for ‘gist’ (Asha Iyer, Christof Koch, Pietro Perona)



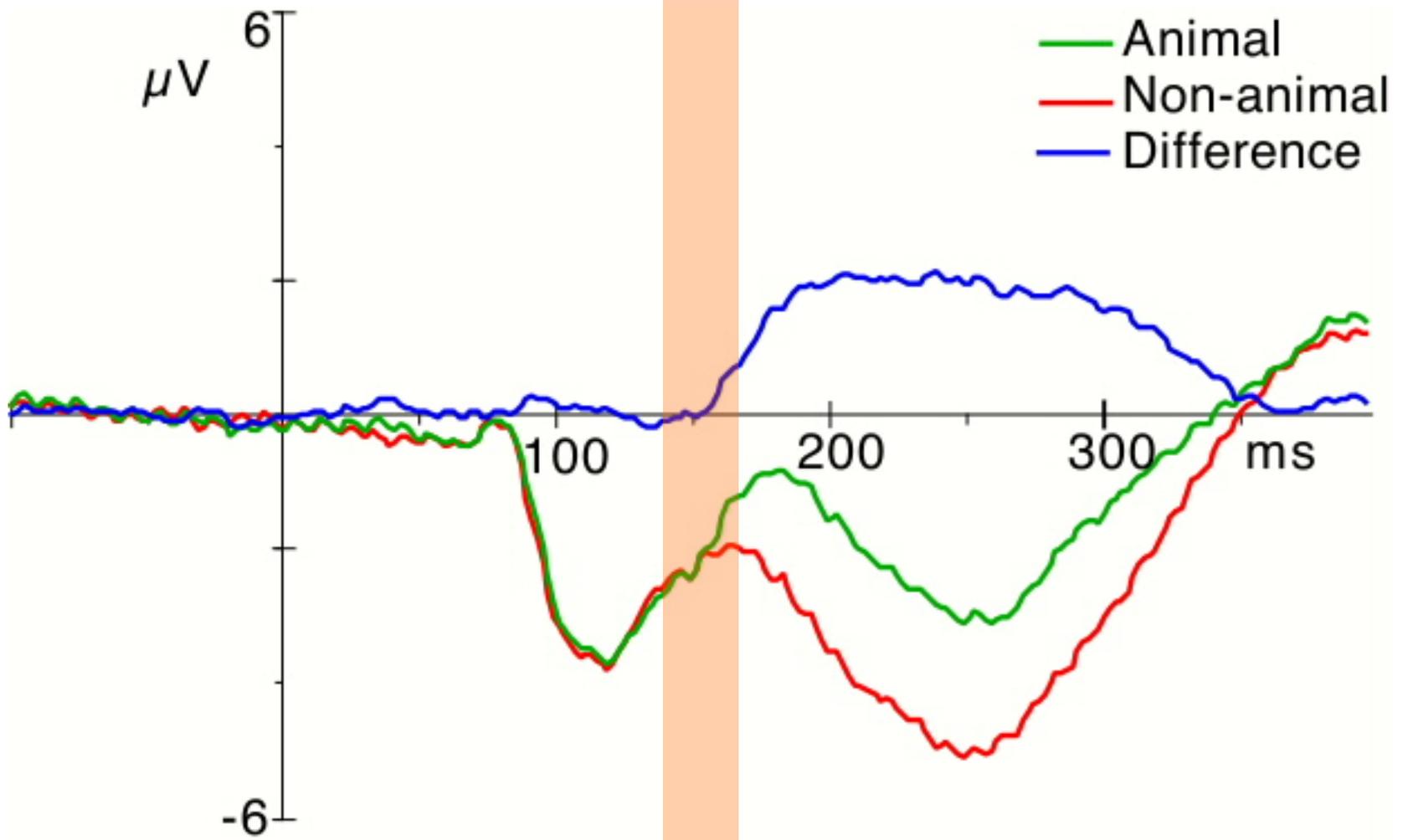
#3: local patches, and some intermediate level information – a hierarchical Bayesian algorithm for natural scene categorization (Pietro Perona)

- #1: natural scene categorization entails little attention



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



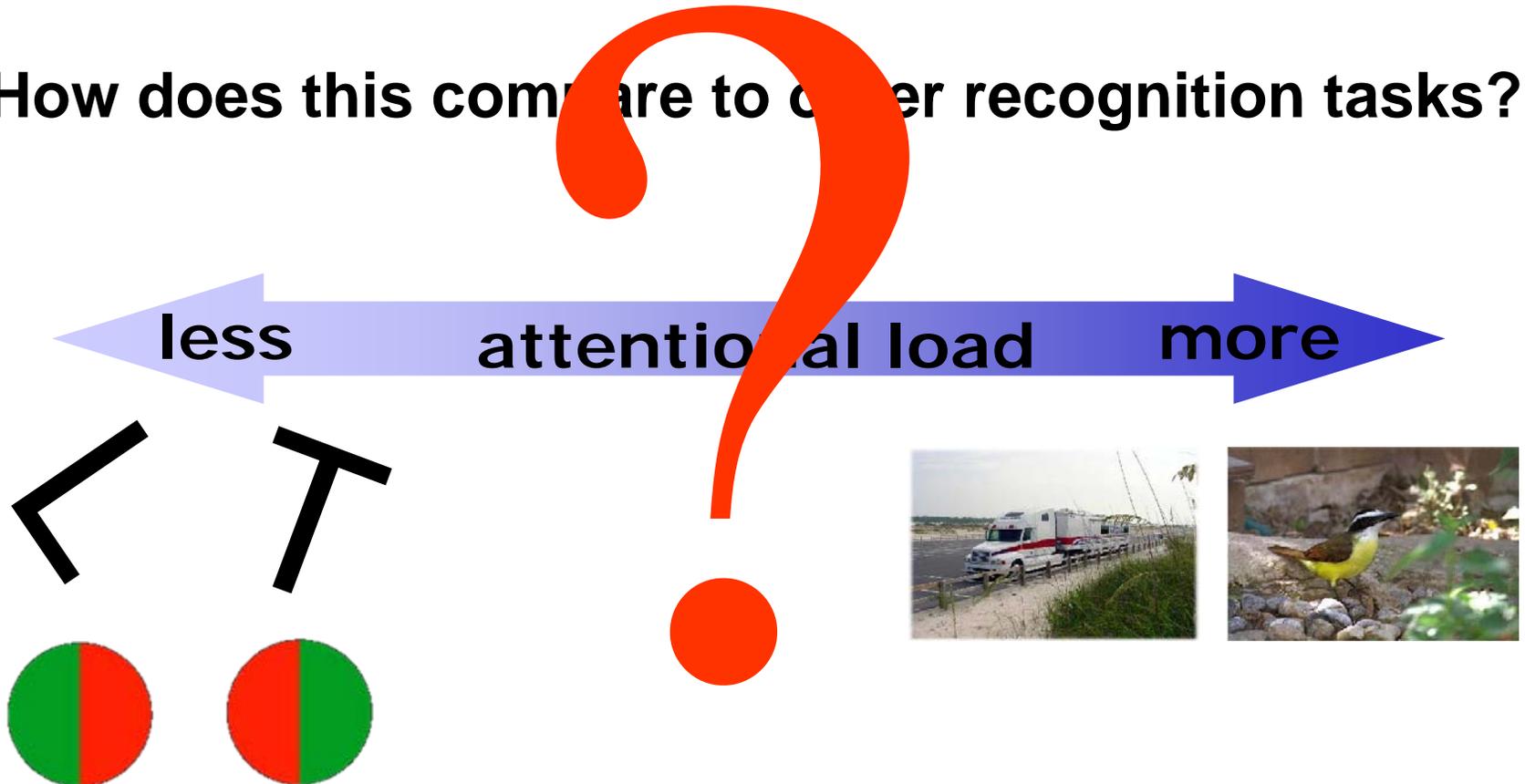


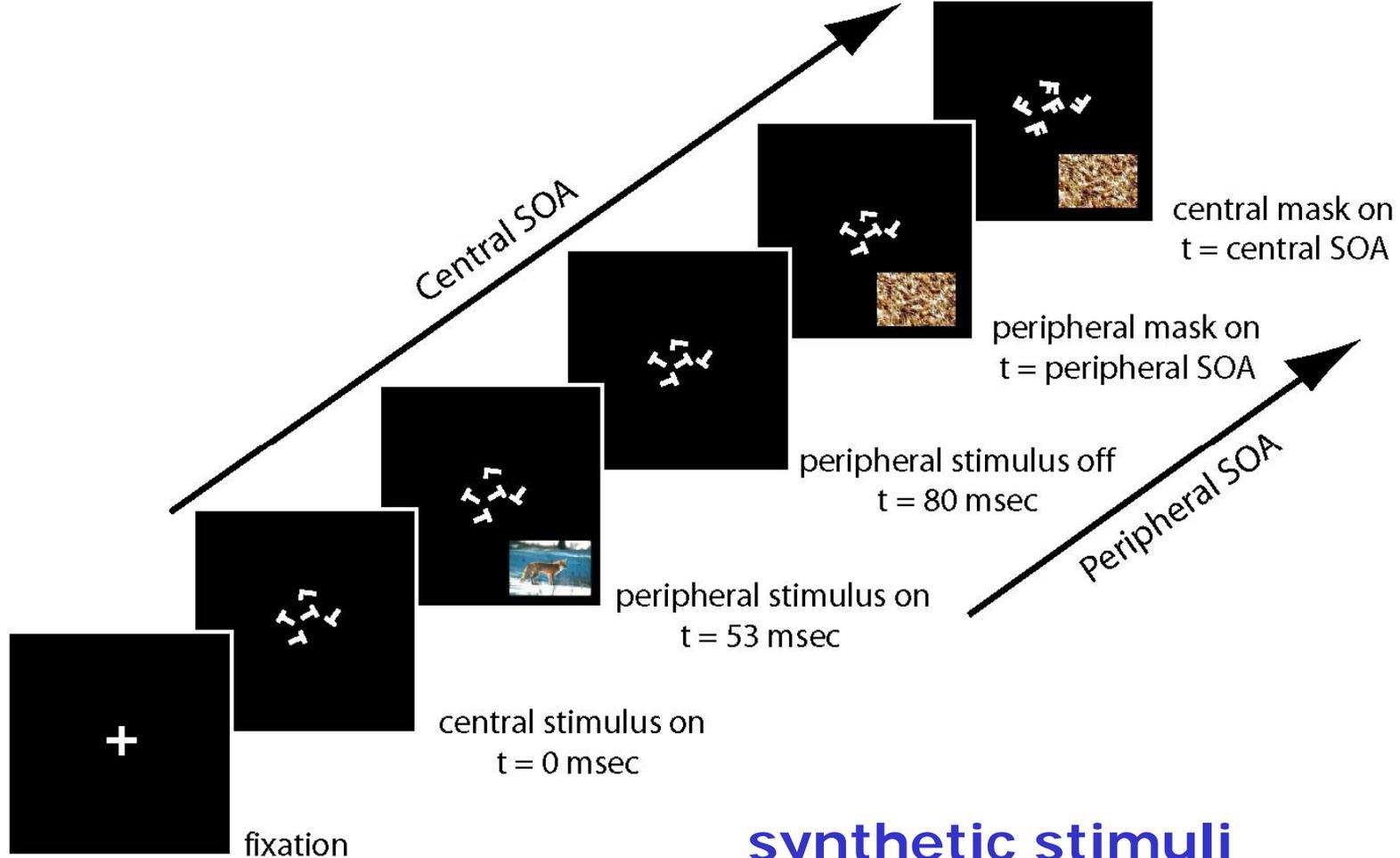
150 ms !!

Our question

1. How critical is attention in natural scene recognition?

2. How does this compare to other recognition tasks?





animals



vehicle



synthetic stimuli



a Target: animal



vs.

Distractors



b Target: vehicle



vs.

Distractors



c Target: vehicle



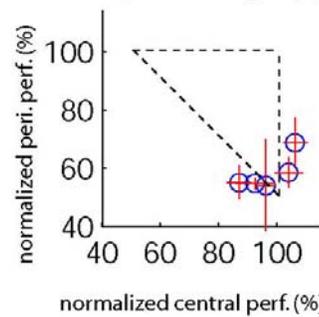
vs.

Distractors

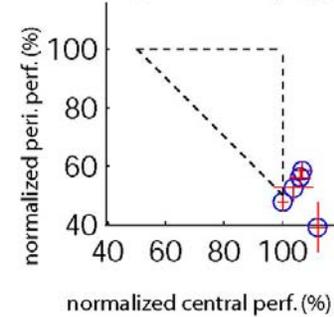


our finding...

d  vs. 
(masked by )



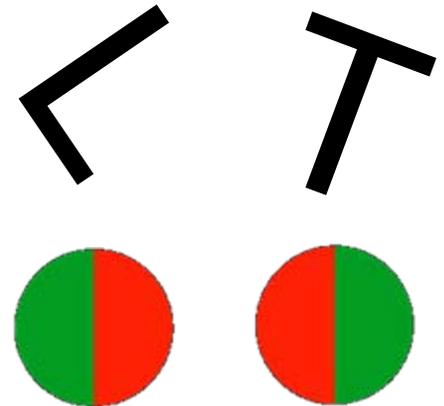
e  vs. 
(masked by )



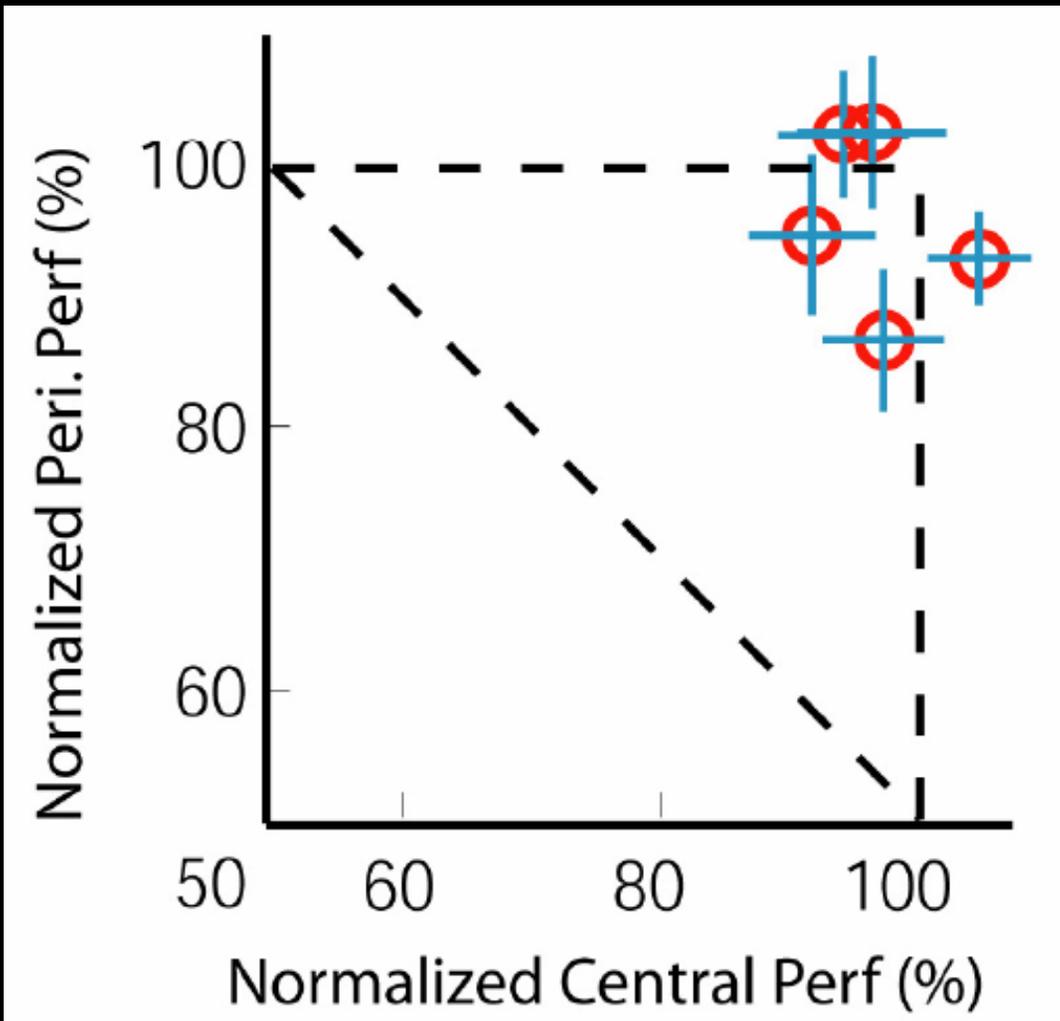
less

attentional load

more



Without color...



Effect of “meaningful” category

randomly rotated

Target

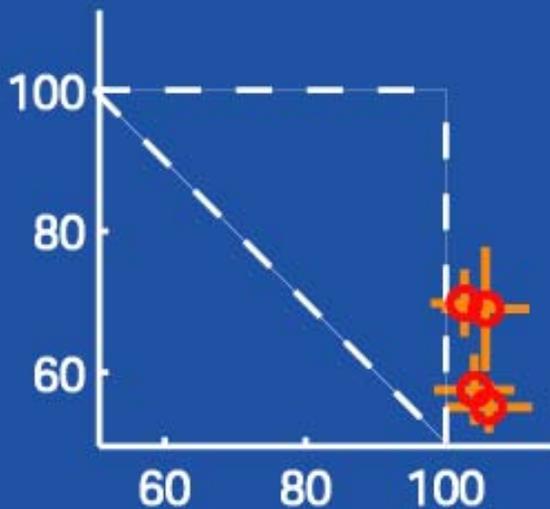


vs.

Distractor



(masked by )



fixed rotation

Target

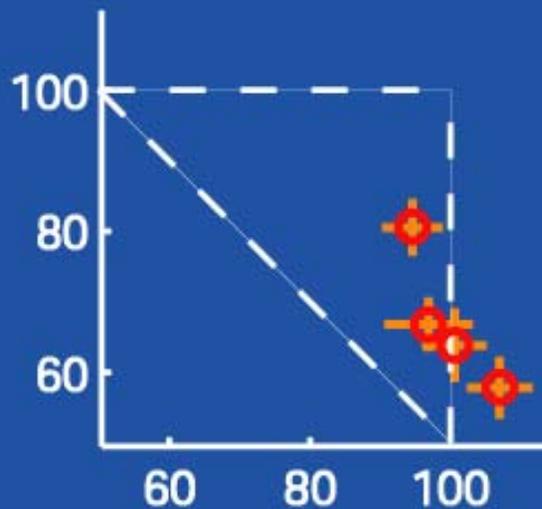


vs.

Distractor



(masked by )



upright position

Target

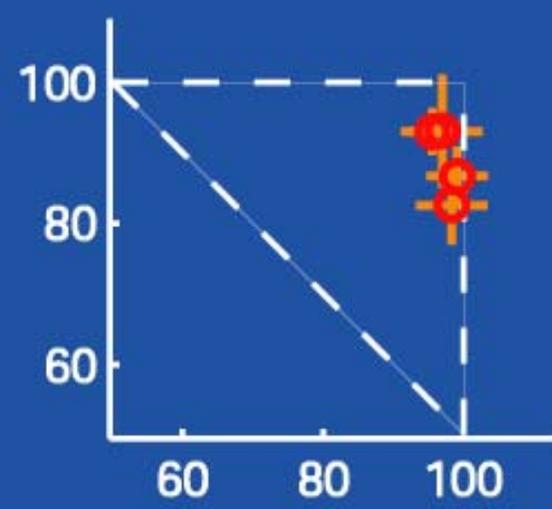


vs.

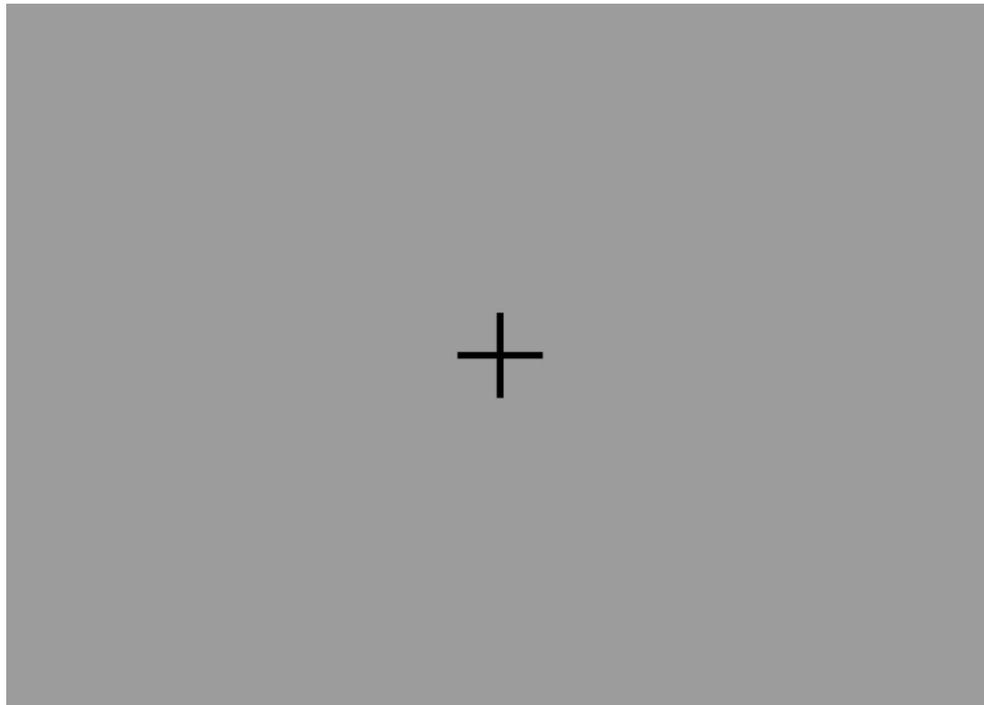
Distractor



(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”

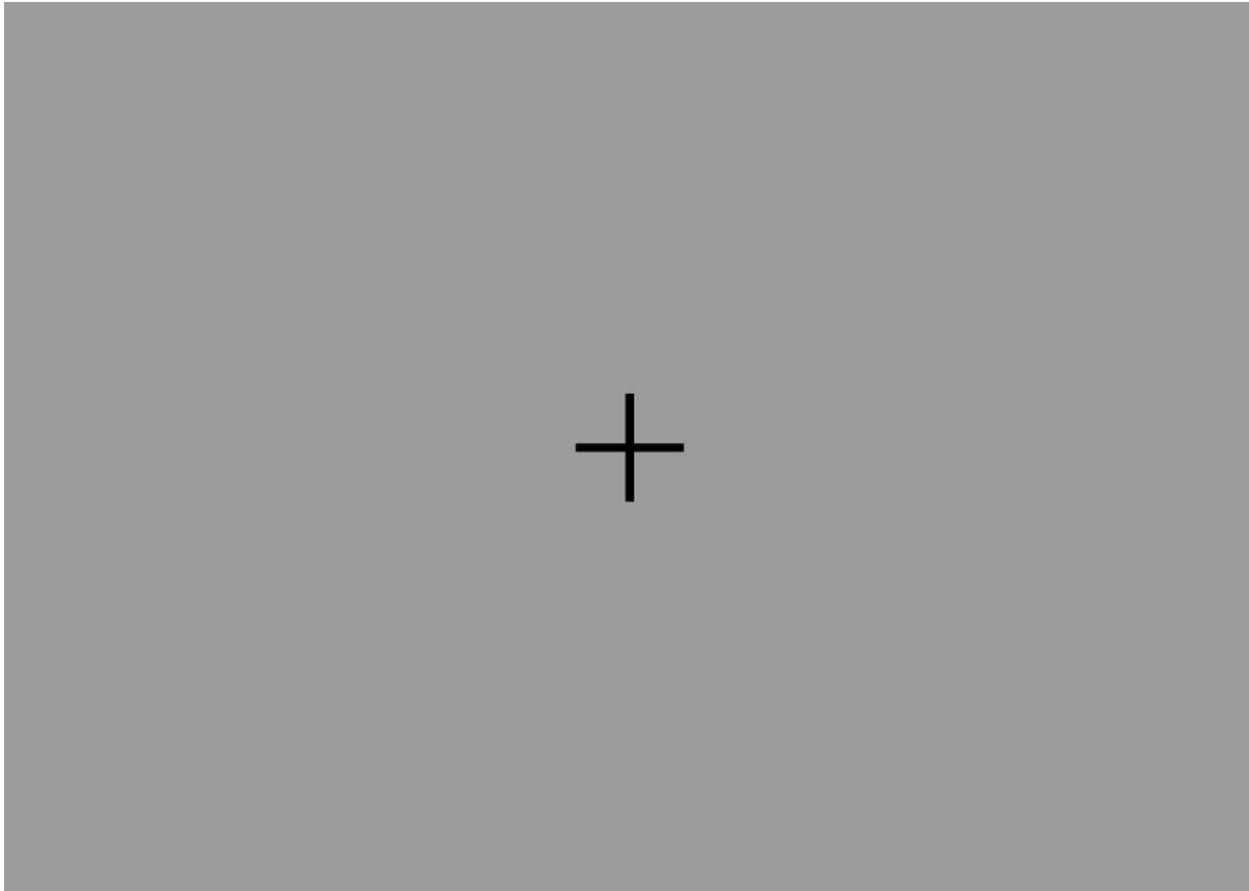


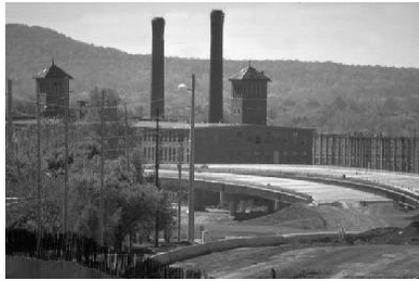
What is “gist”?

- 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

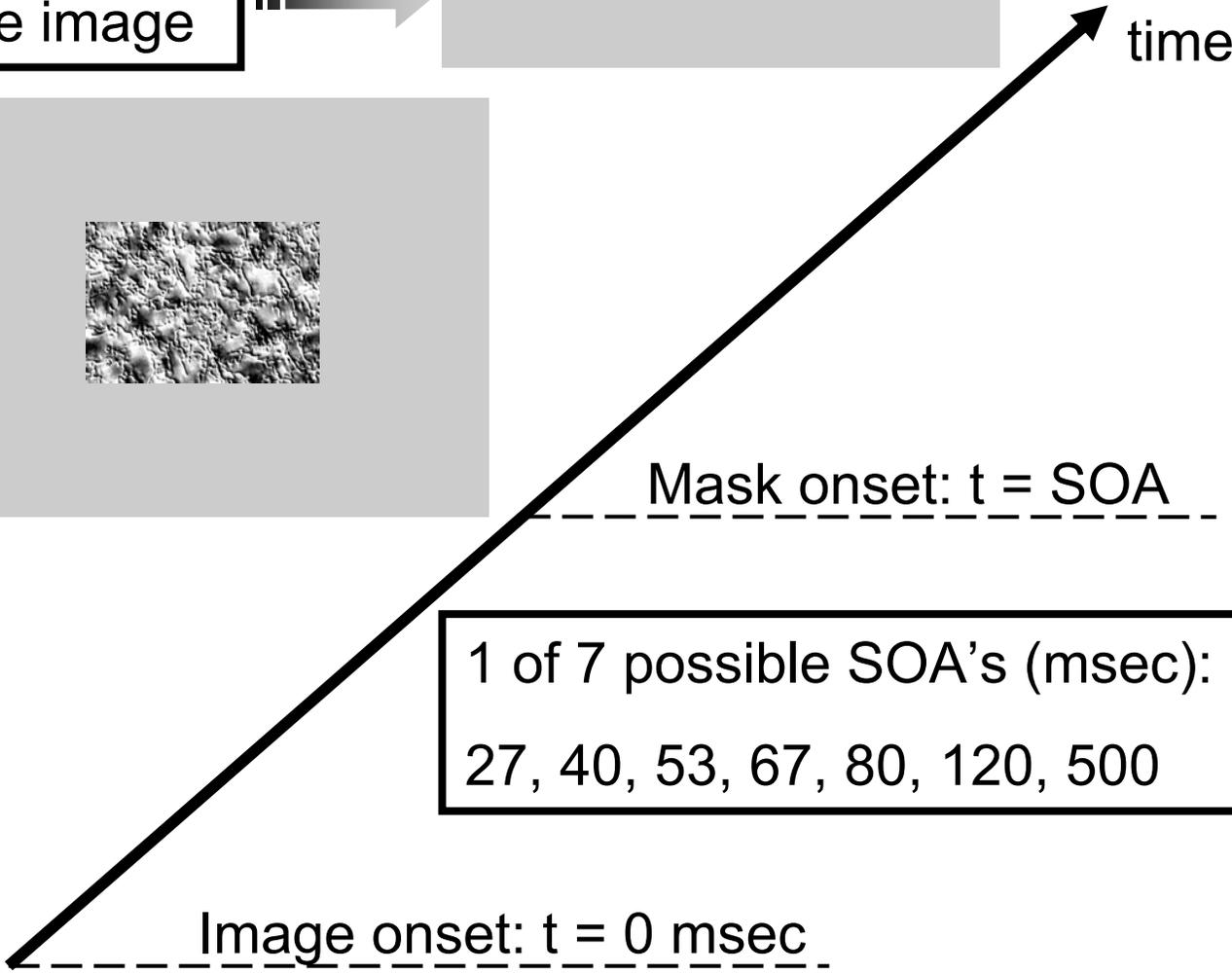
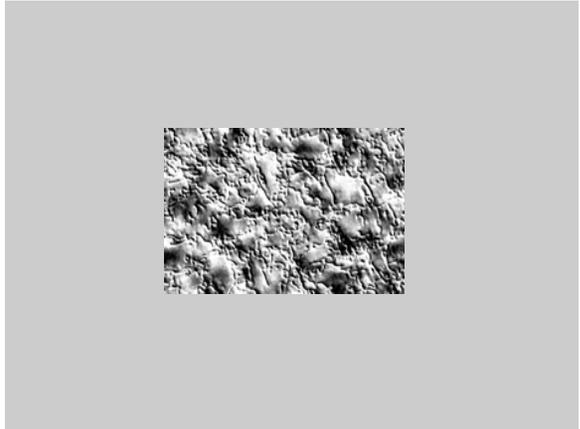
--- Illustration of 1 Trial

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 walking in a park in new york or something. there seemed to be trees and a road and then this large skyscraper in the background.



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

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)

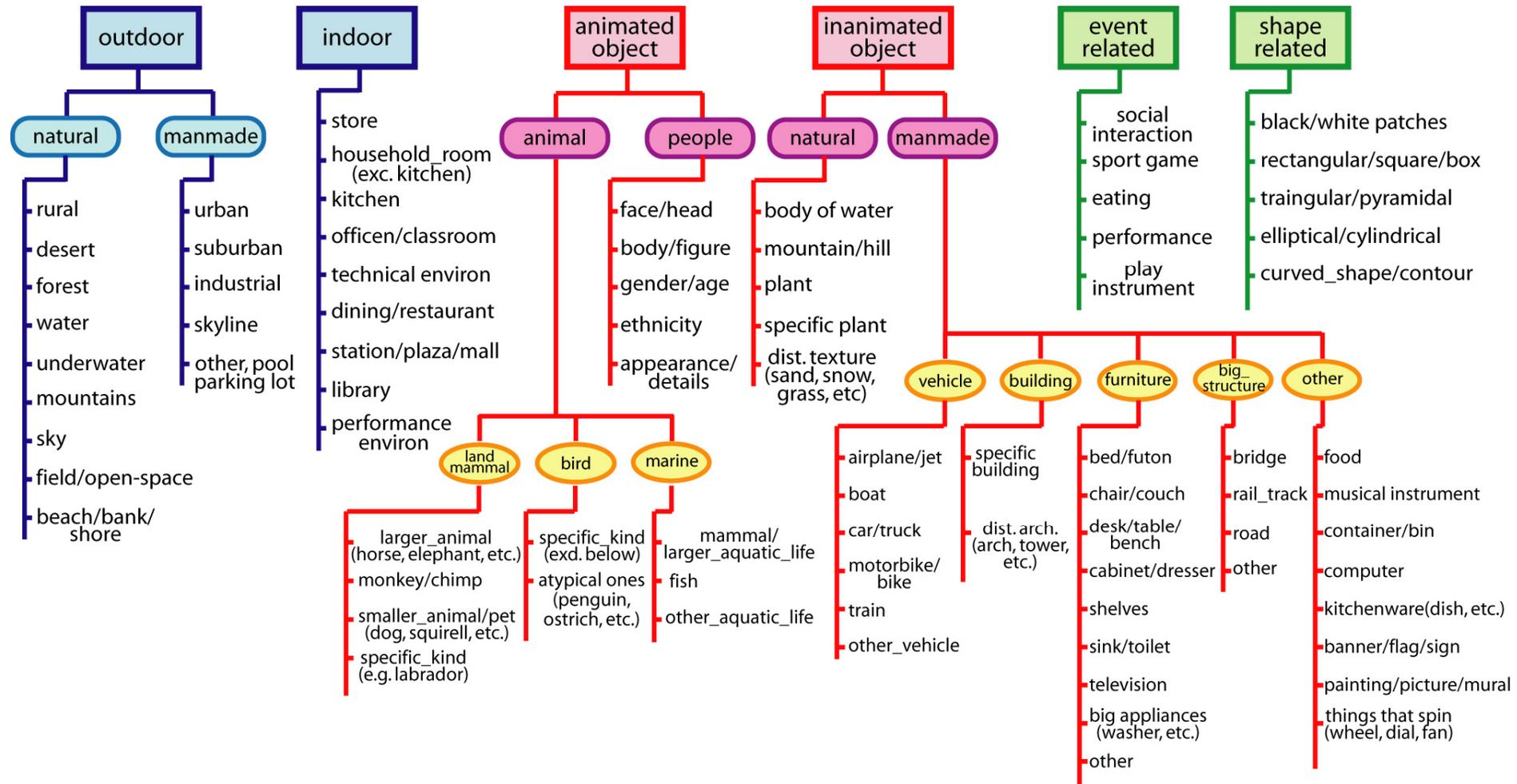
PT = 500ms

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



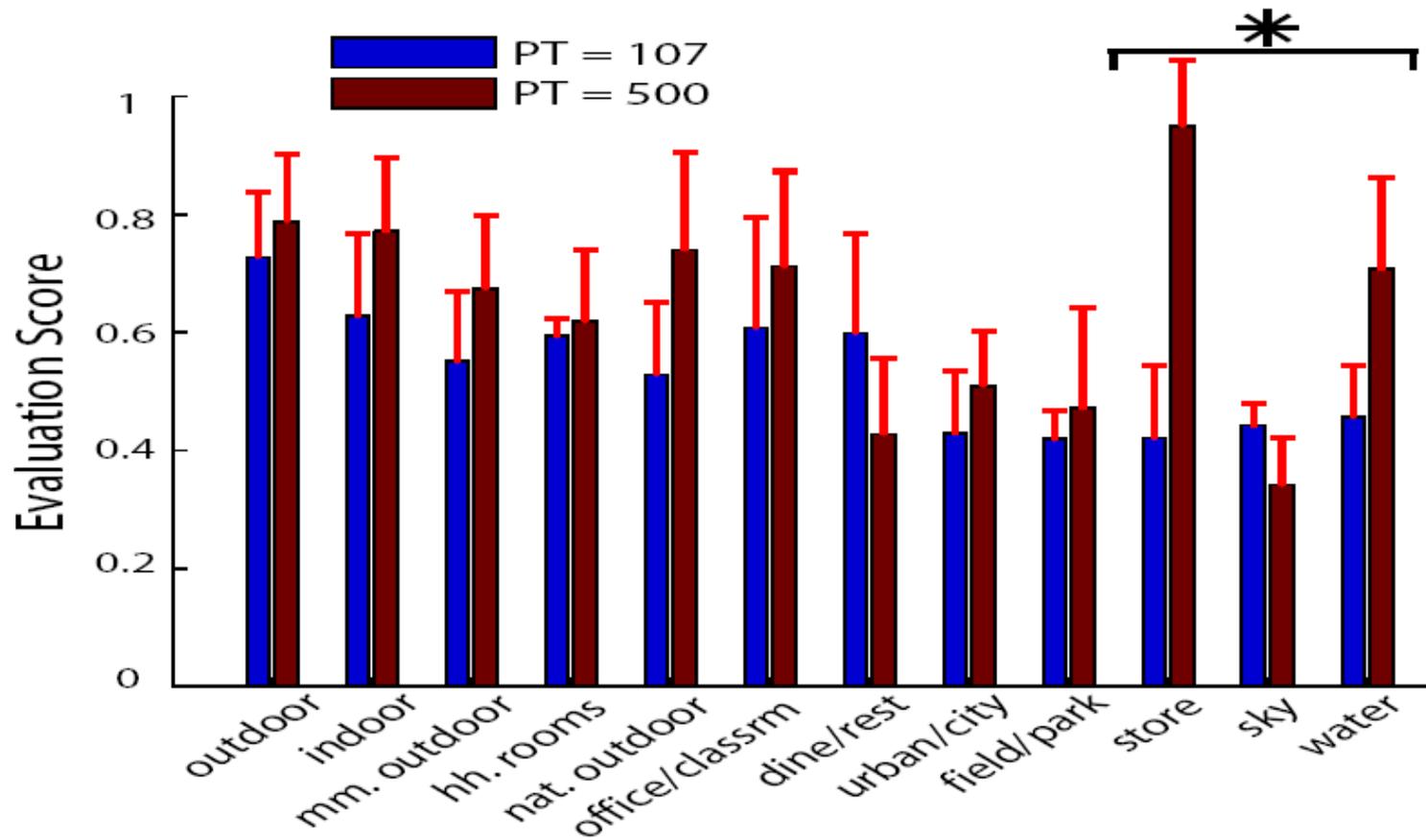
I could make out some kind of circular shapes near the bottom of the picture. These reminded me of those round life preservers that are on ships. There was also a man standing on top of some wooden structure.

CATEGORY: SENSORY/SHAPES

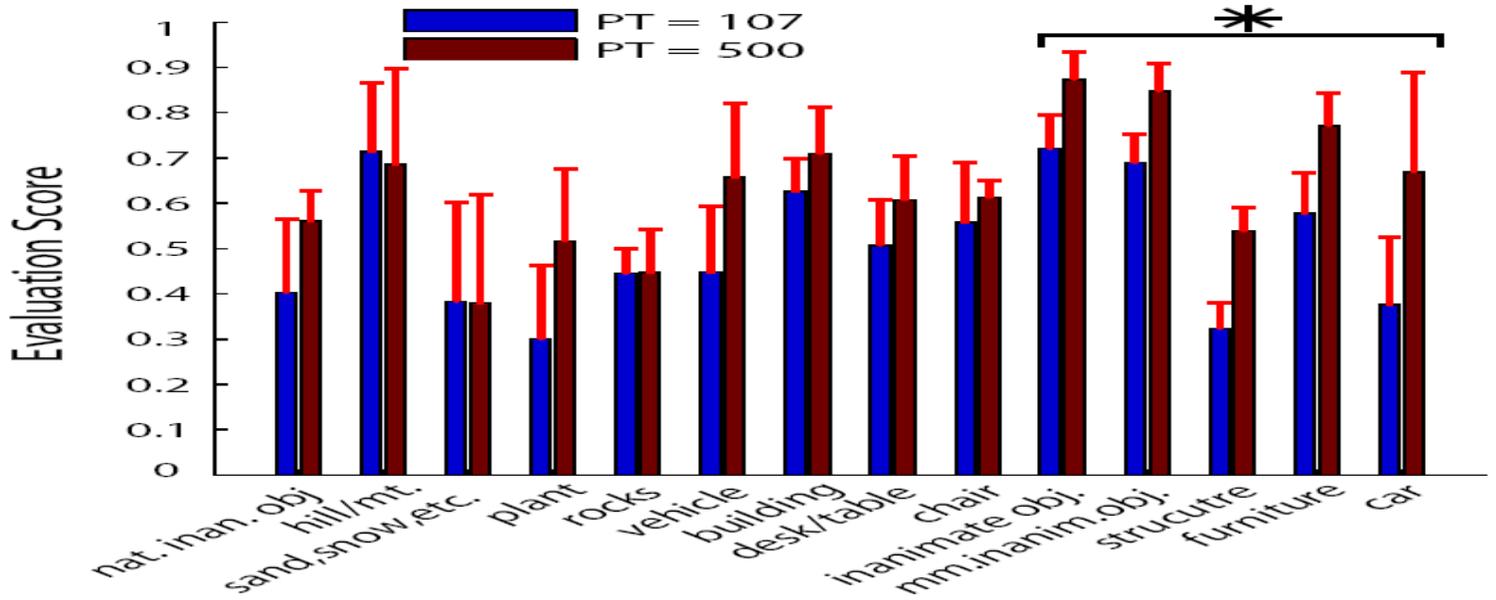
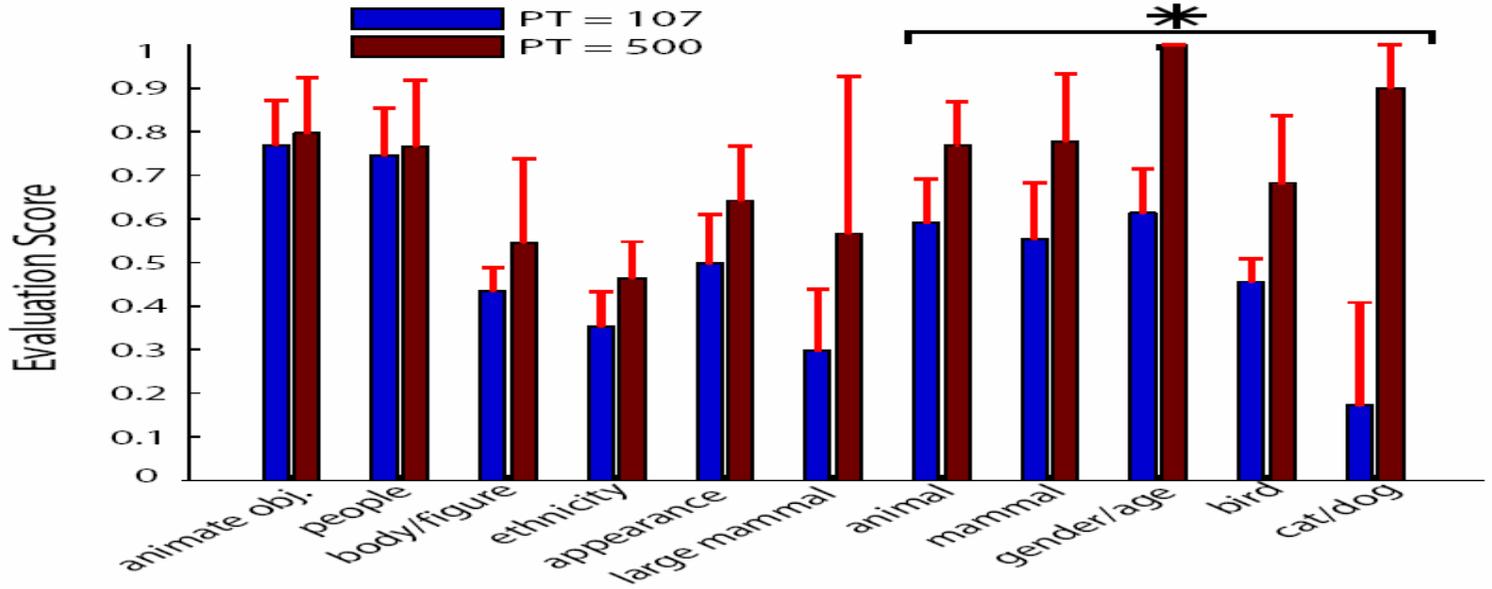
Please select one of "correct" or "incorrect" for each checked description. Click "Next>>" to continue

- 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

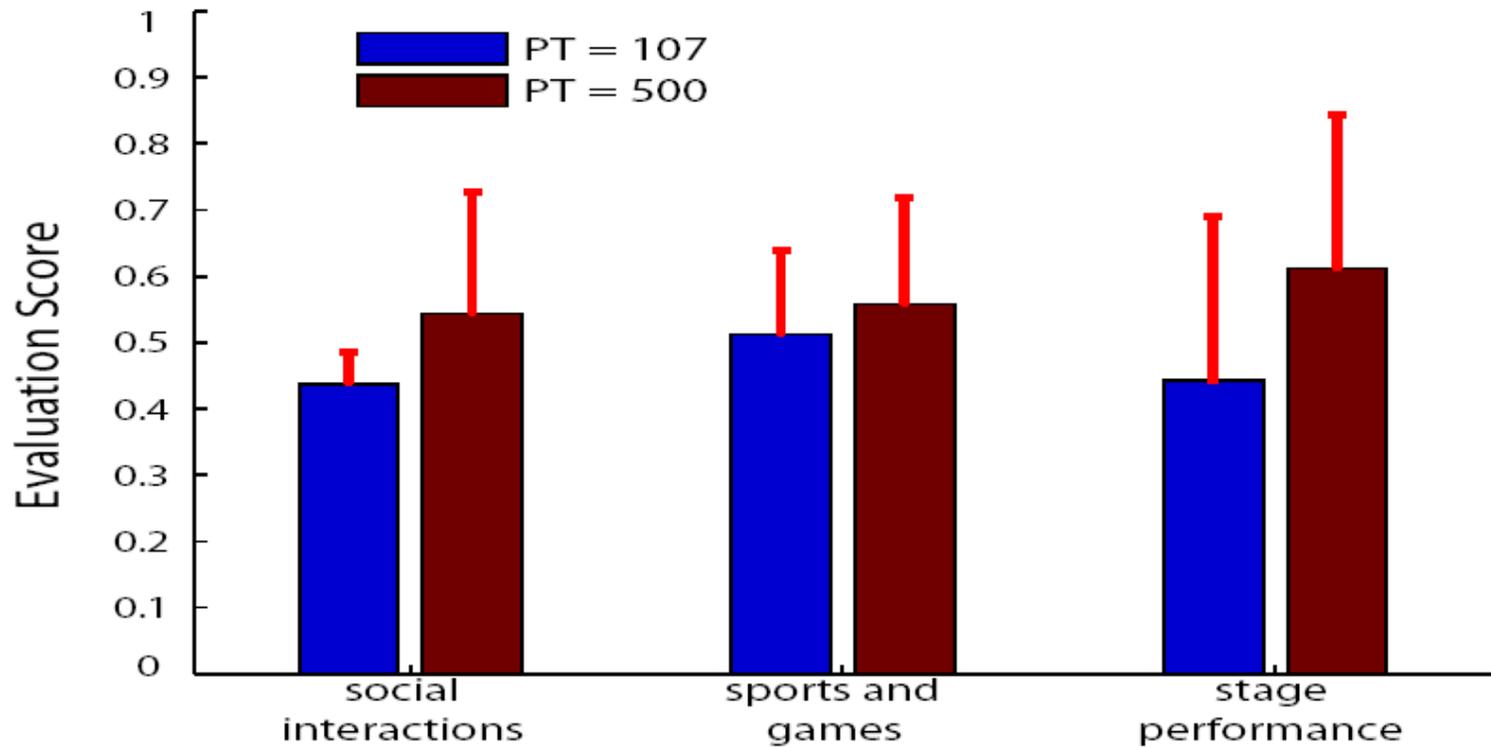
Scene level



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



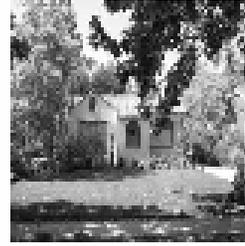
livingroom bedroom



coast



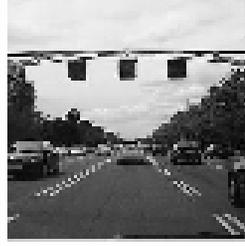
suburb



tall bldg



highway



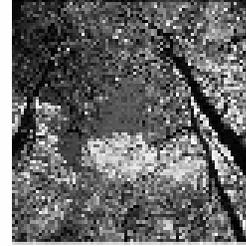
o. country



mountain



forest



streets



ins. city



• global cues: colors, textures, etc.

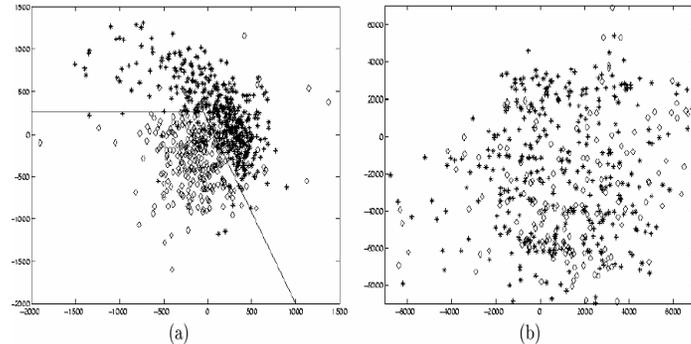
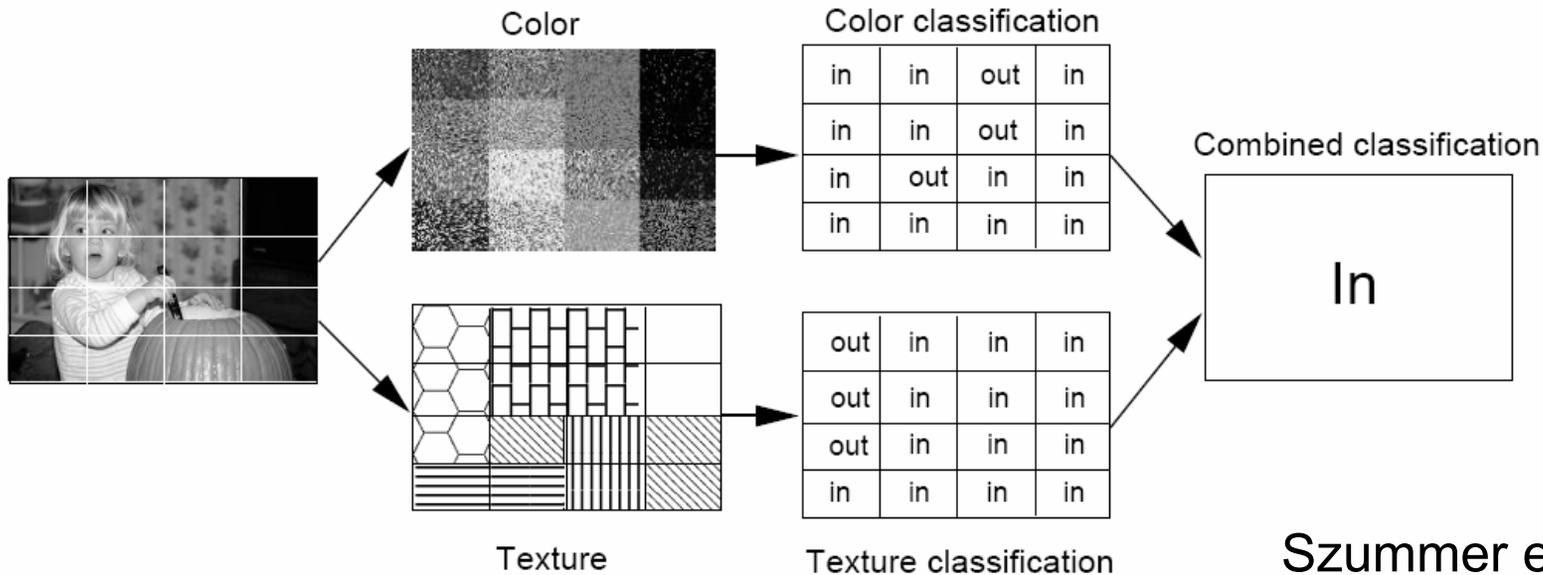


Figure 7: 2D Feature space plots showing (a) edge direction coherence vector and (b) color coherence vectors; * represents the landscape patterns and ◇ represents the city patterns; only a subset of 2,716 patterns have been plotted here for clarity of display.

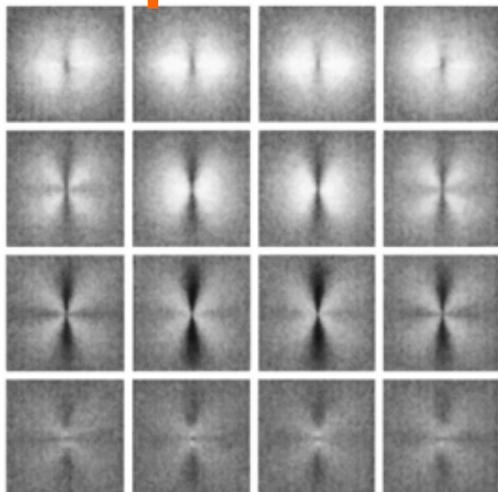
Jain, Zhang et al. (1998)



Szummer et al. (1998)

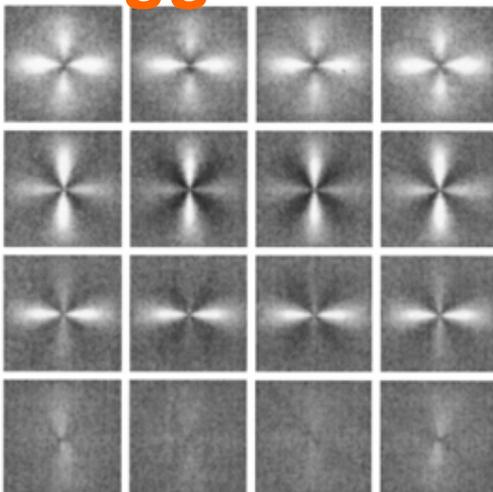
- **global cues: frequency**

openness



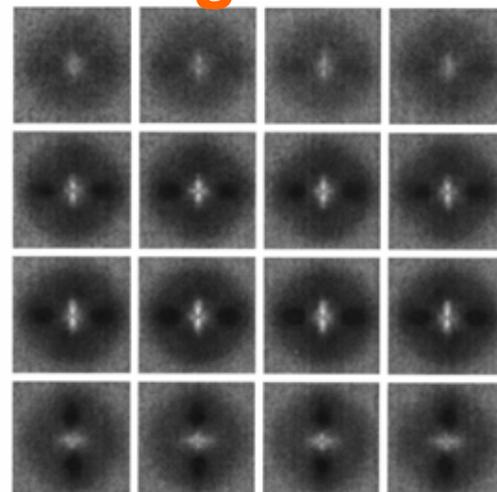
a)

ruggedness

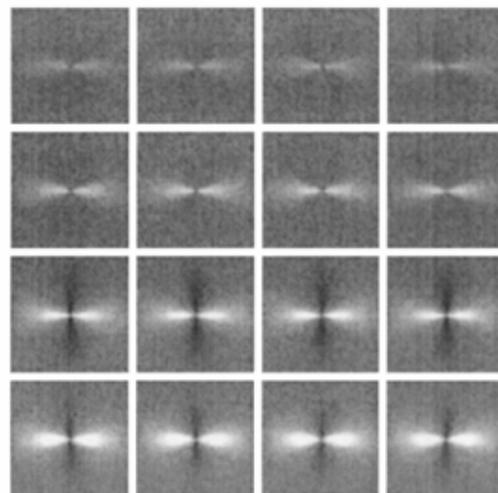


b)

roughness

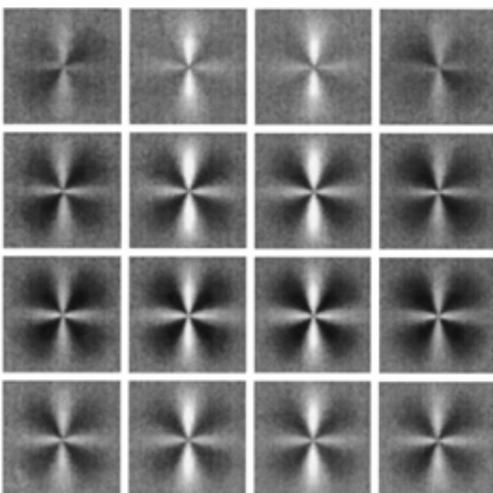


c)



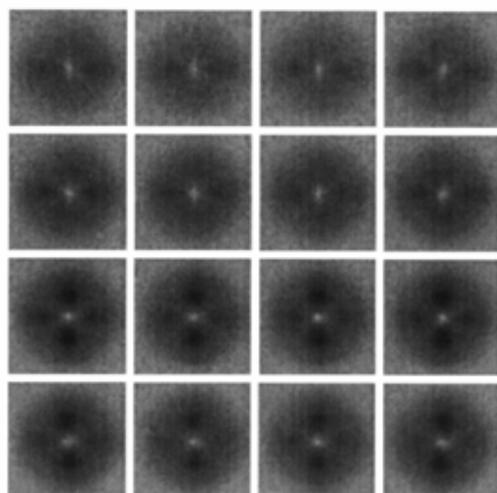
d)

openness



e)

expansion

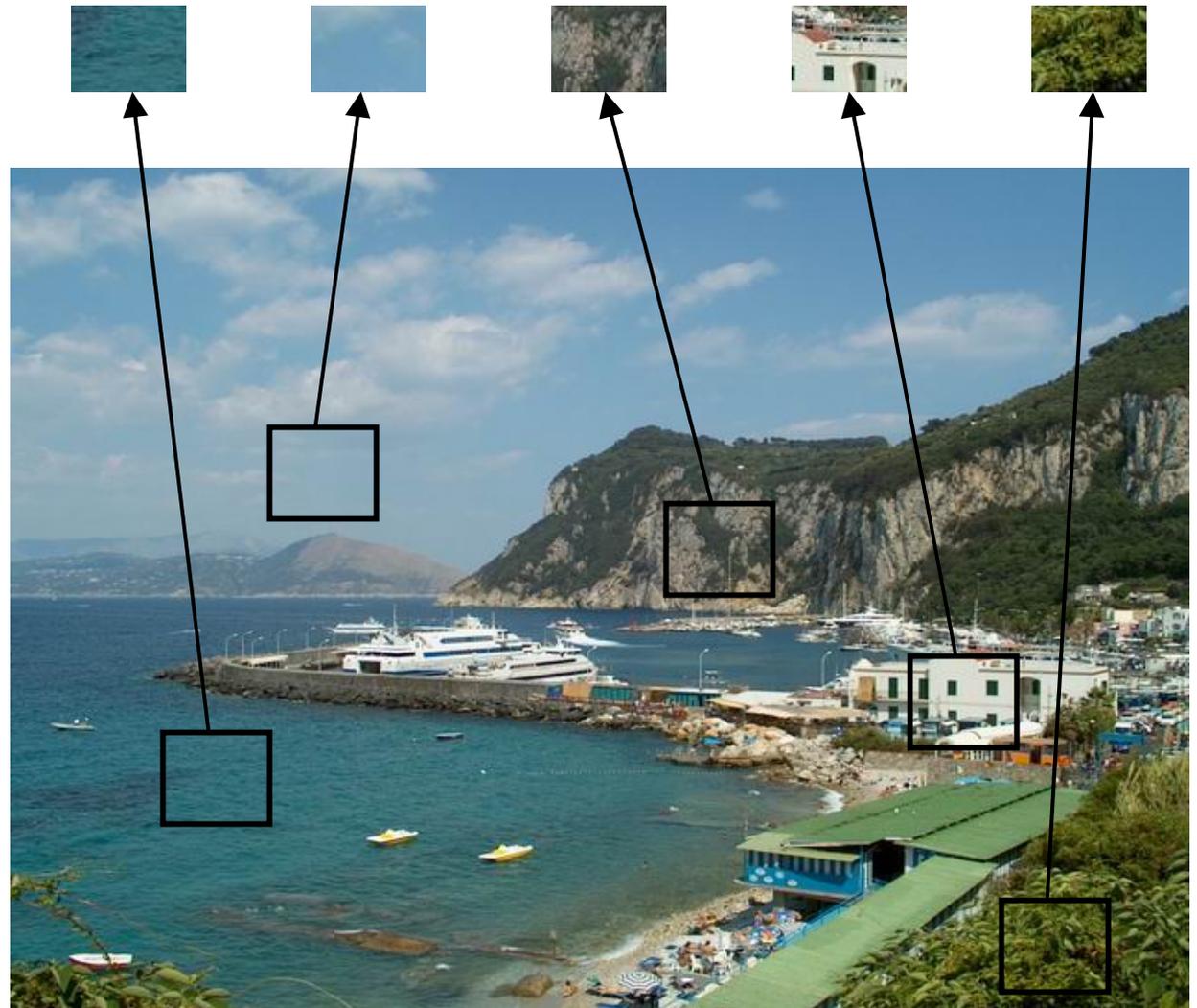


f)

roughness

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

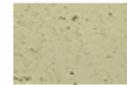
rocks



foliage



sand



sky



water



- intermediate level themes within scenes

forest



suburb



inside of city



Our intuitions

- local patch based

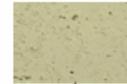
rocks



foliage



sand



sky



water



- intermediate level themes



- weakly supervised

- no human annotation of local patches and intermediate level themes

Image



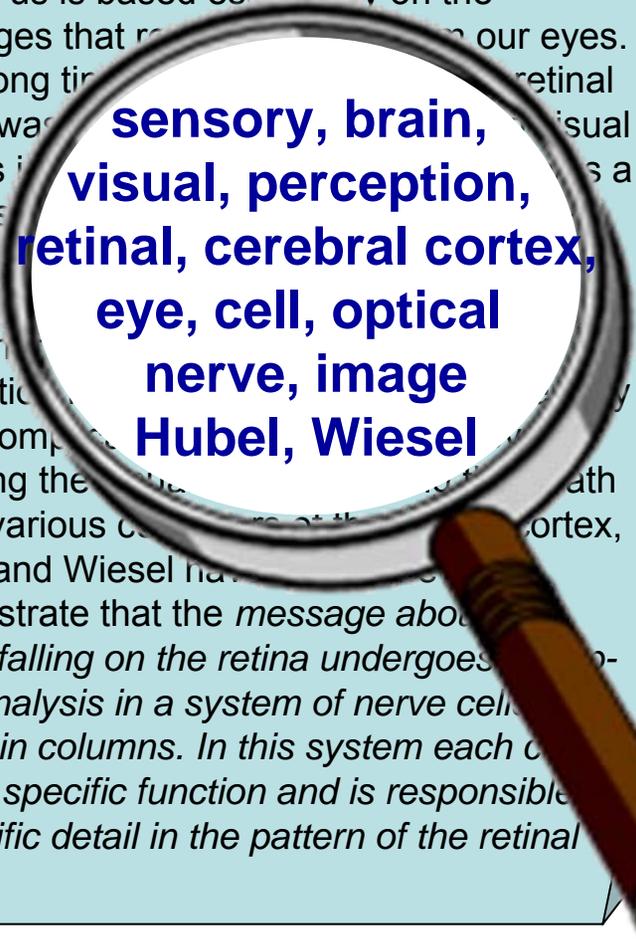
Bag of 'words'



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 our eyes.

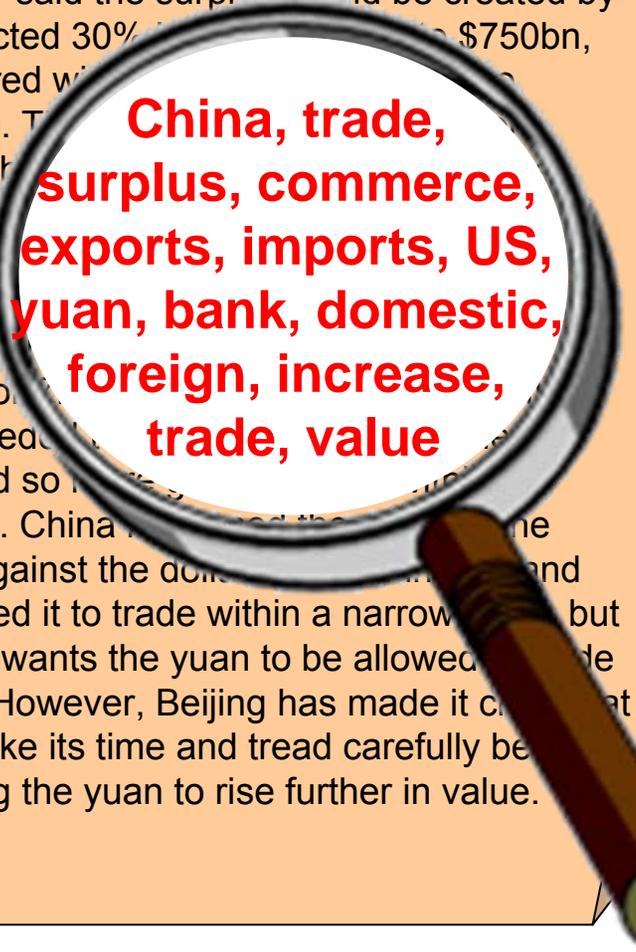
For a long time, the retinal image was considered as a movie screen. The image is discovered by the eye, know the perception is more complex following the path to the various centers of the cortex, Hubel and Wiesel have demonstrated that the *message about the image falling on the retina undergoes a point-by-point analysis in a system of nerve cells stored in columns. In this system each cell has its specific function and is responsible for a specific detail in the pattern of the retinal image.*



**sensory, brain,
visual, perception,
retinal, cerebral cortex,
eye, cell, optical
nerve, image
Hubel, Wiesel**

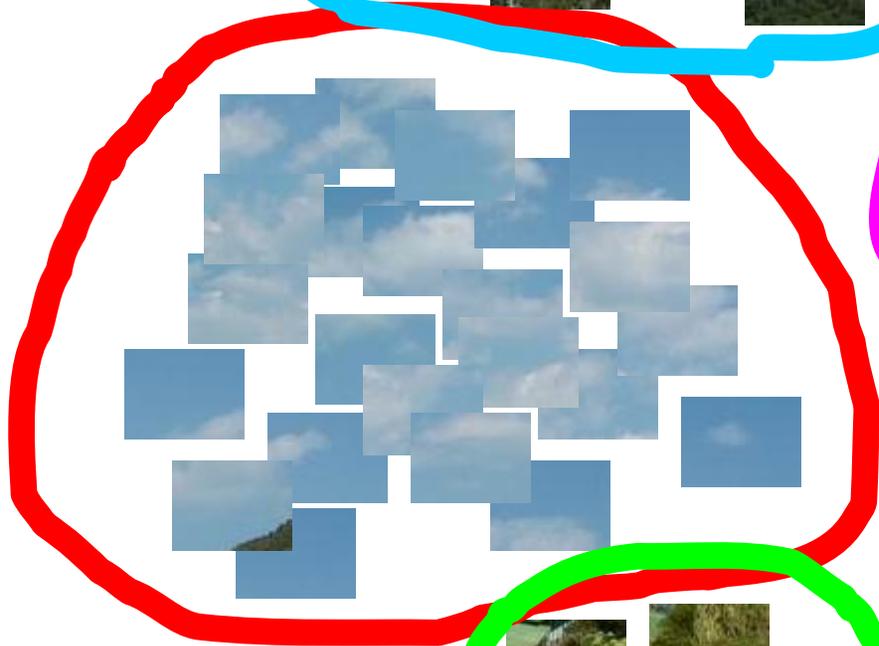
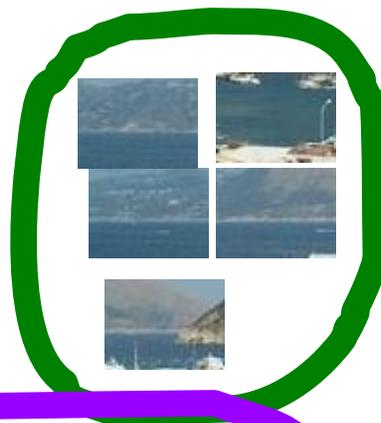
China is forecasting a trade surplus of \$90bn (£51bn) to \$100bn this year, a threefold increase on 2004's \$32bn. The Commerce Ministry said the surplus would be created by a predicted 30% increase in exports to \$750bn, compared with \$560bn in 2004.

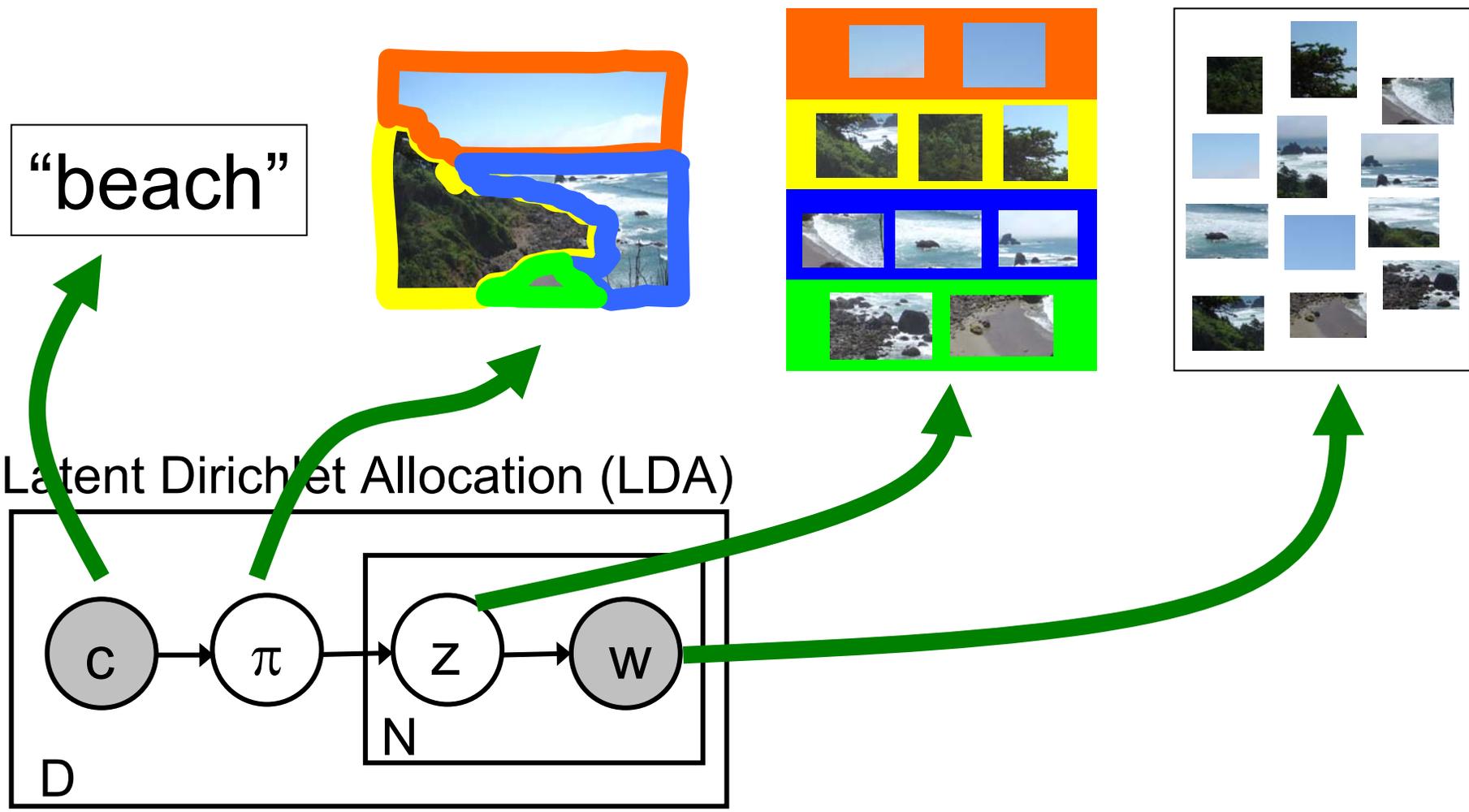
The increase will annoy the US. China's government has deliberately agreed to let the yuan rise against the dollar. The government also needs to increase the demand for the yuan in the country. China has allowed the yuan against the dollar to rise and permitted it to trade within a narrow band but the 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.



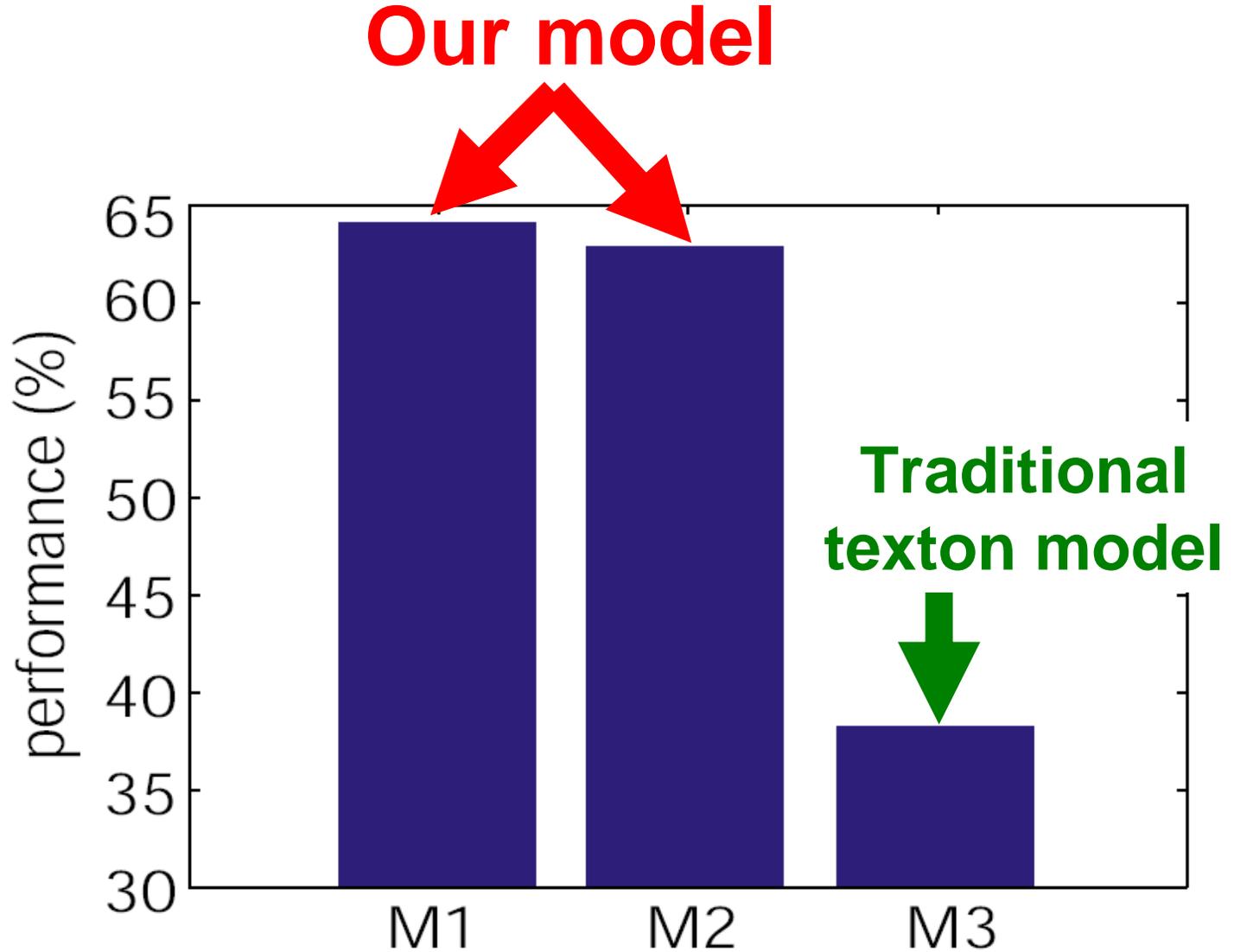
**China, trade,
surplus, commerce,
exports, imports, US,
yuan, bank, domestic,
foreign, increase,
trade, value**



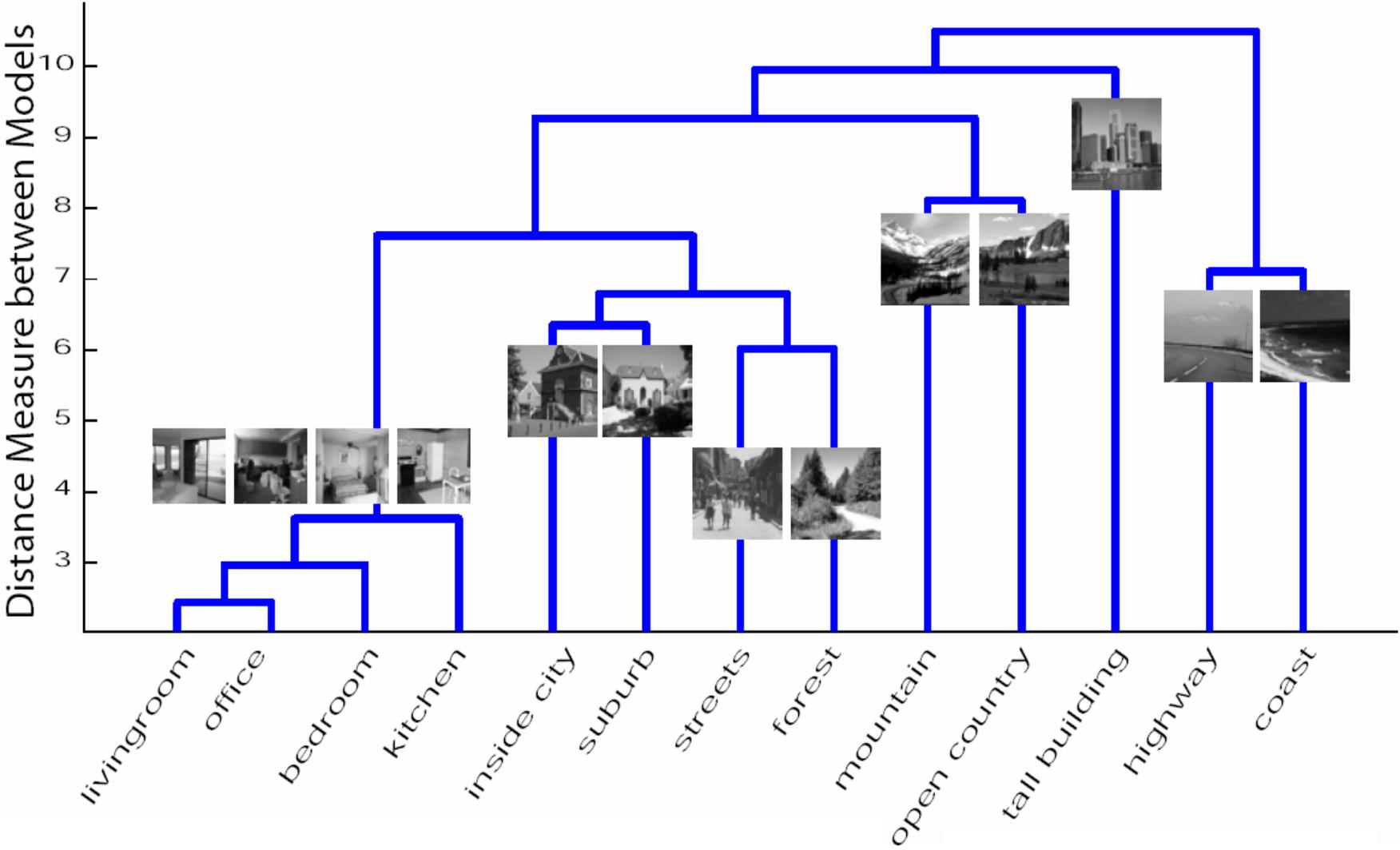




	highway	insidecity	tallbuildings	street	suburb	forest	coast	mountain	opencountry	bedroom	kitchen	livingroom	office
highway	74	2		2	2		14	4		2			
insidecity		58	10	6	8		4			2	6	4	2
tallbuildings		4	76	10				4		4		2	
street	2	4	6	78		2		2	2			4	
suburb					94					2			4
forest						88		12					
coast	2						78		20				
mountain	4		4		2	6	8	70	6				
opencountry	8				8	10	16	10	48				
bedroom	4	2	2		2	2	2	4		28	12	38	4
kitchen		8	2				2				60	14	14
livingroom		2	2	2			2	4		4	18	56	10
office					2		2			8	12	12	64

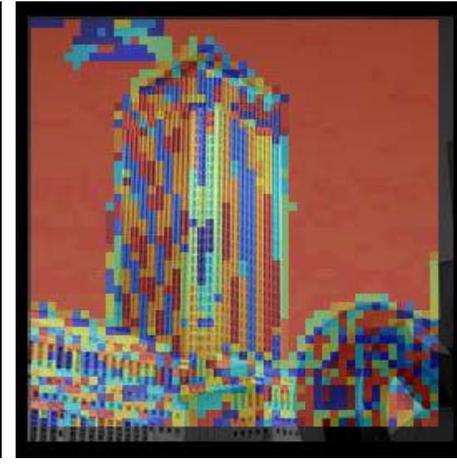
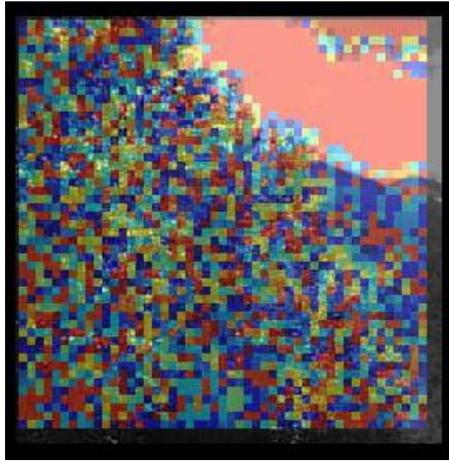


model distance based on theme distribution

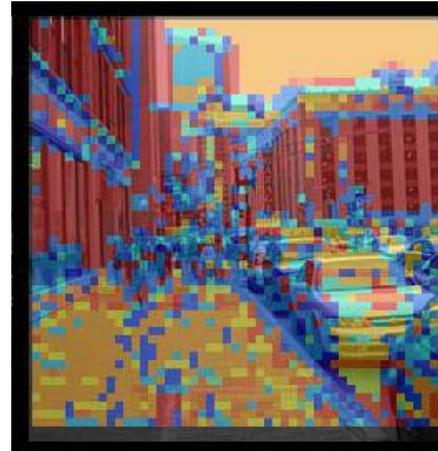
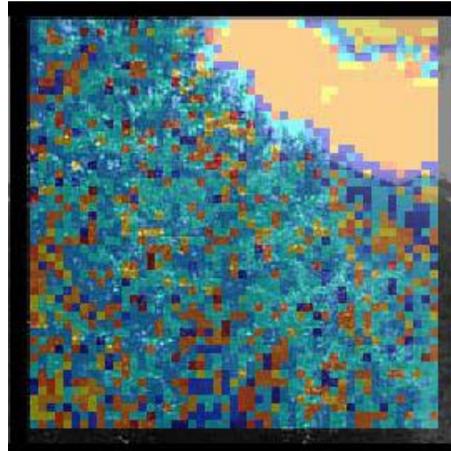
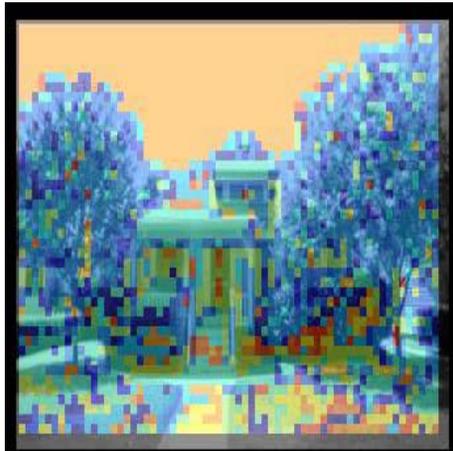


segmentation by themes

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. VanRullen, 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.