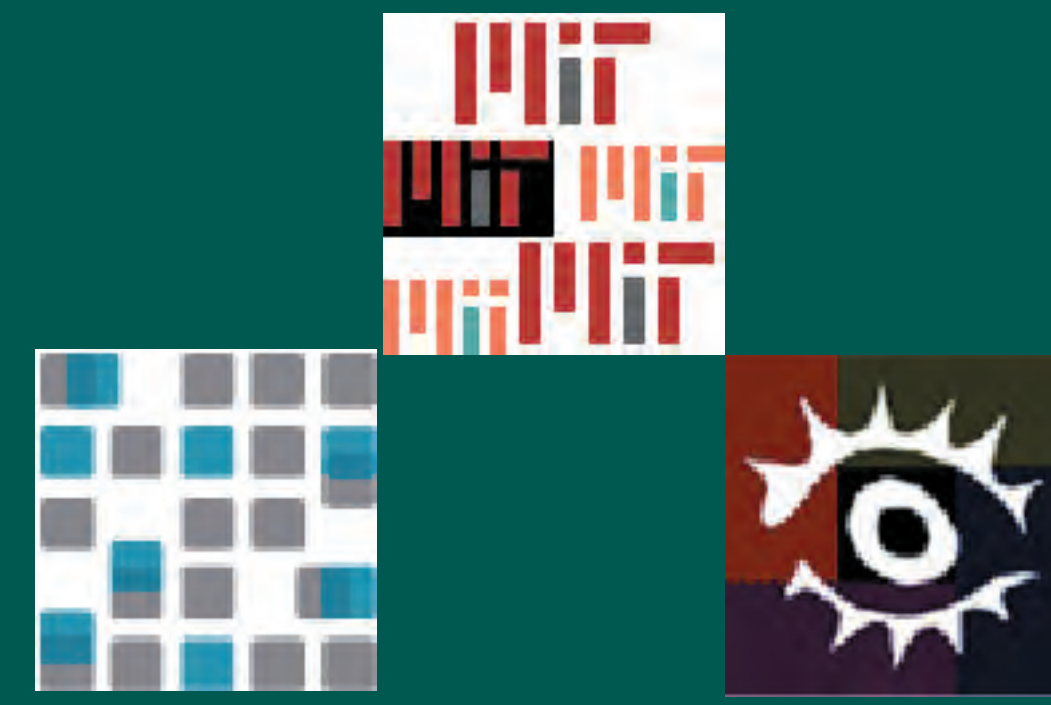# Seeing the {Closed+Camouflage+Natural=Forest} for the Trees: Rapid Scene Categorization can be mediated by conjunctions of global scene properties

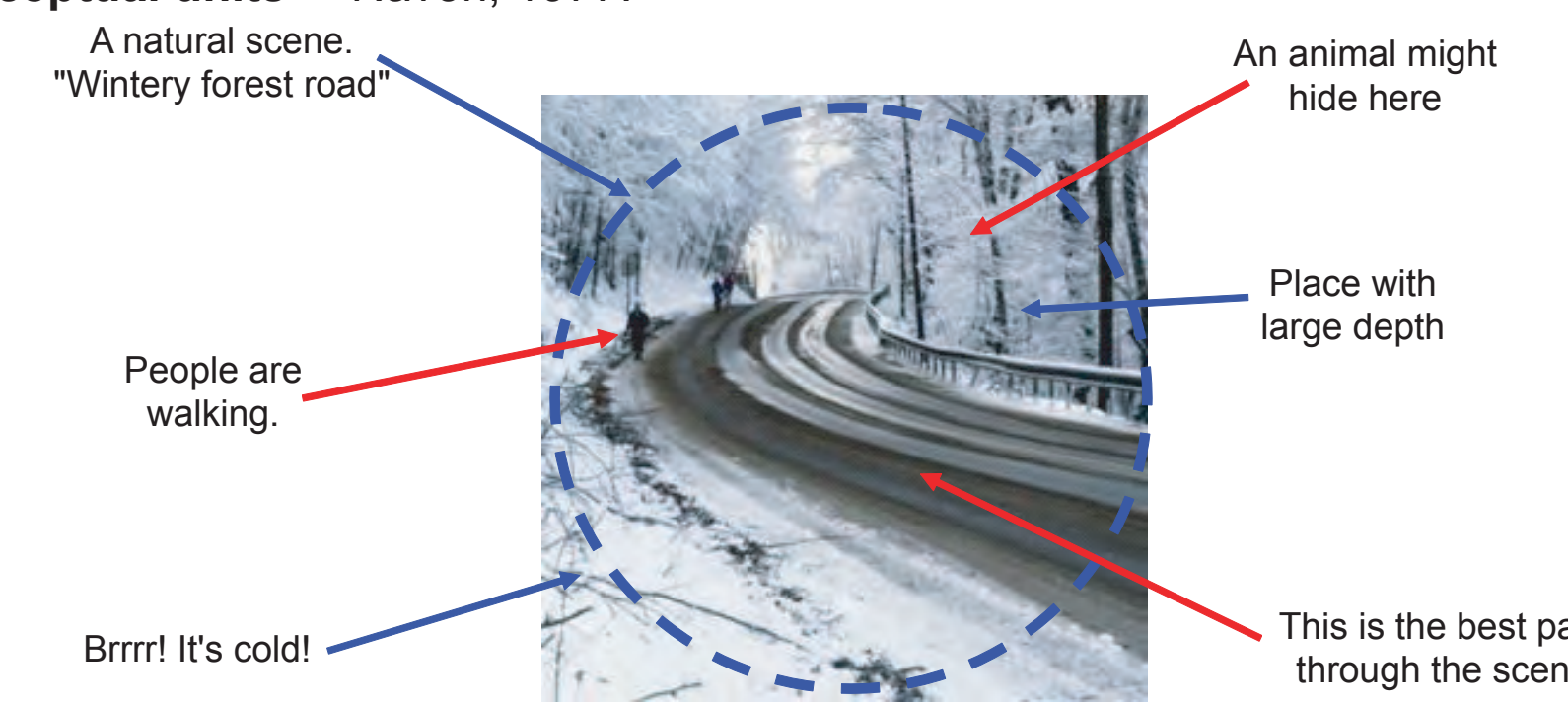## Michelle R. Greene & Aude Oliva

Perceptual Science Group, Brain & Cognitive Science Department, MIT

Introduction

"No attempt was made here to formulate an **operational definition of globality of visual features which enable precise predictions about the course of perception of real- world stimuli.** What is suggested in this paper is that whatever the perceptual units are, the spatial relationships among them is more global than the structure within them.....Thus, I am afraid that **operational measures for globality will have to patiently await the time that we have a better idea of how a scene is composed into perceptual units"** - Navon, 1977.

A natural scene.
"Wintery forest road"
An animal might hide here
Place with large depth
People are walking.
This is the best path through the scene
Brrr! It's cold!

Questions:
1. What is global information in a scene?
2. How is this information used by human observers?
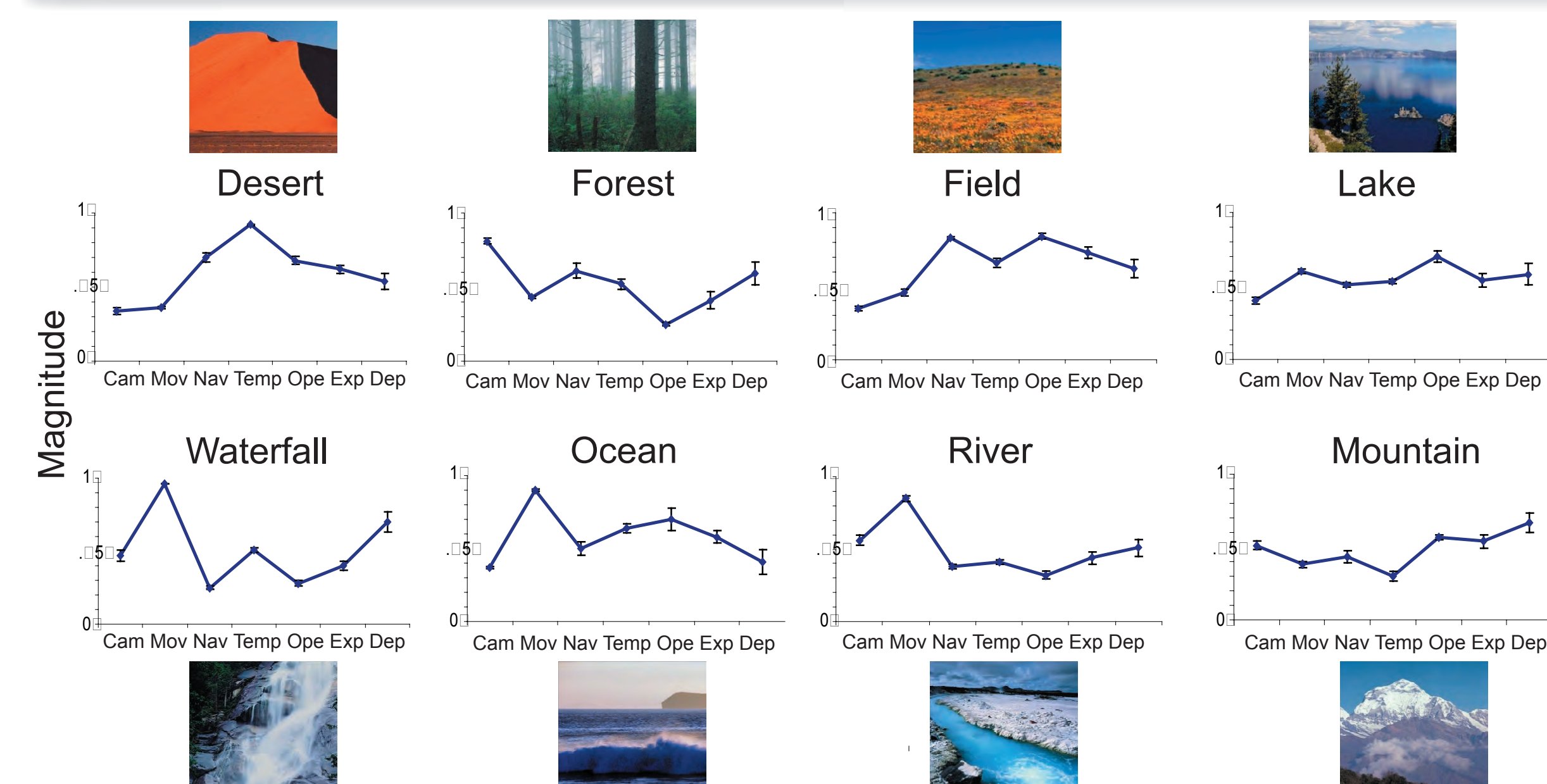3. How does it compare to local (object-centered) information in scenes?

Global scene properties

A set of properties that describe the **space** that a scene subtends, the possible **interactions** a human can have in the space, or properties of the **surfaces** in the scene.

Space — Degree of openness
Interaction — Degree of camouflage
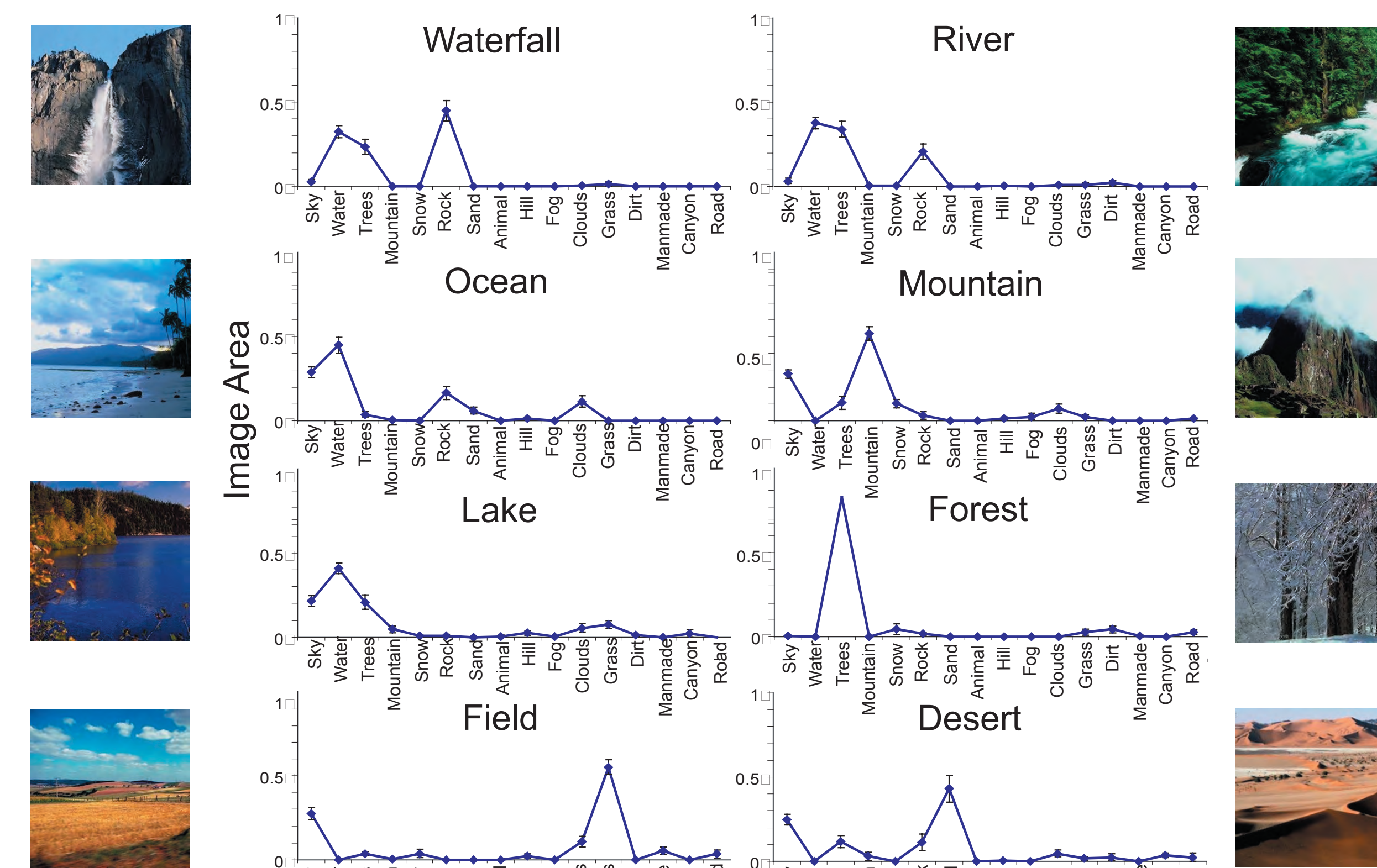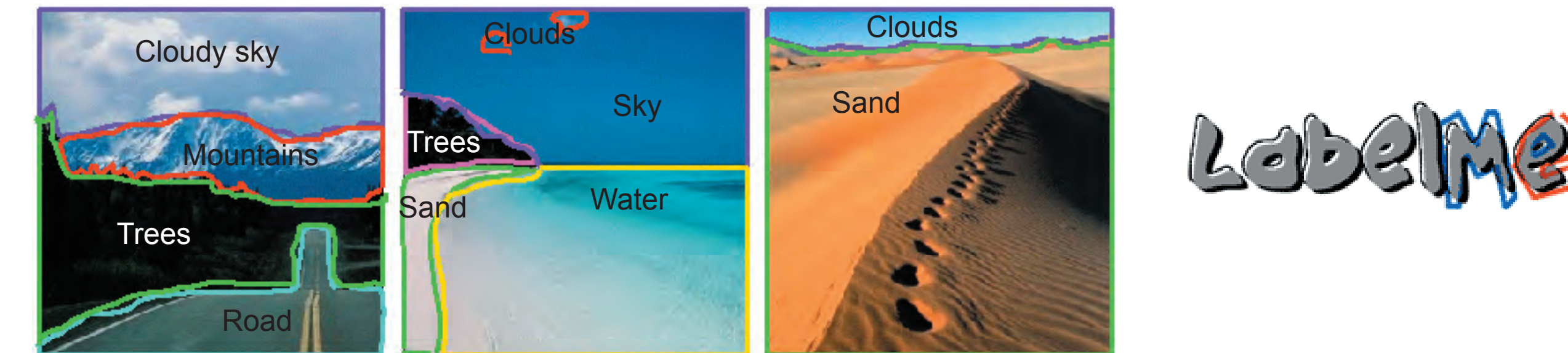Surface — Temperature

Mean depth
Navigability
Degree of movement

Degree of expansion

200 images from 8 semantic categories were ranked along these 7 dimensions by 55 observers.
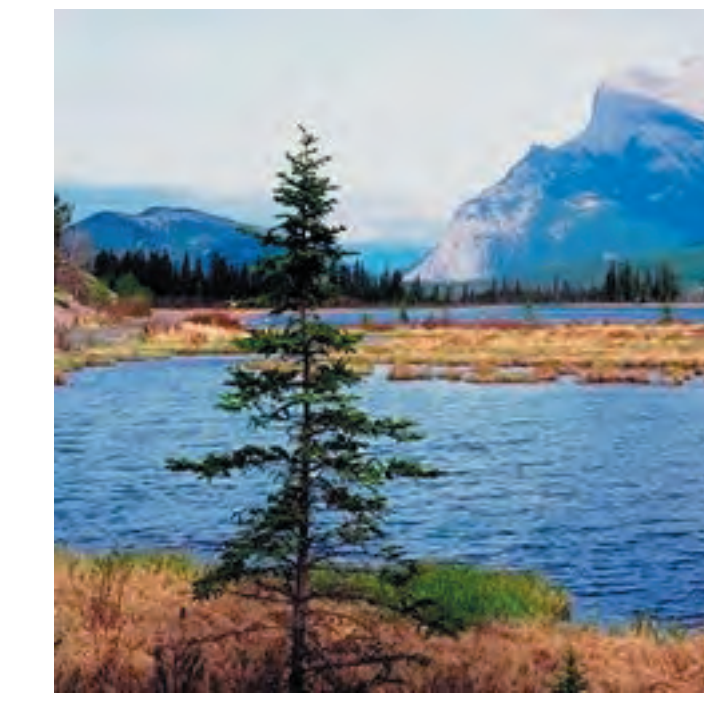
Ranking Results

Desert
Forest
Field
Lake

Waterfall
Ocean
River
Mountain

(Cam Mov Nav Temp Ope Exp Dep)

Local object database norming

Two observers hand annotated all regions in the 200 image database:

Cloudy sky
Clouds
Mountains
Trees
Sky
Trees
Sand
Water
Road
Clouds
Sand

LabelMe

Waterfall
River
Ocean
Mountain
Lake
Forest
Field
Desert

Image Area

(Sky Water Trees Mountain Snow Rock Sand Animal Hill Fog Clouds Grass Dirt Mammade Canyon Road)

Perceptual availability of global properties

What is the perceptual availability of global properties relative to semantic categories?

Fixation: 500 msec
Staircased
Mask: 80 msec

We compared GLOBAL property classification to SEMANTIC category classification in a staircased yes-no forced choice task.

Threshold is presentation time permitting 75% correct classifications.

	Global properties	Semantic categories
Temperature	15.4	
Desert		19.8
Navigation	20.6	
Camouflage	21.3	
Forest		21.7
Expansion	22.8	
Depth	23.0	
Waterfall		23.0
Movement	26.7	
Ocean		28.5
Openness	29.9	
Lake		32.3
Mountain		32.5
River		34.2
Field		34.3
Mean=22.8		**Mean=28.2**

Global property / Semantic category
p=0.003
Probability density
Time (msec)

Sufficiency of global properties

To what extent is global property information or object information alone sufficient to predict human categorization performances?

Global =
41% camouflage
61% moving
47% navigable
32% hot
66% open
64% expansive
74% deep

Local =
23% sky
35% water
18% trees
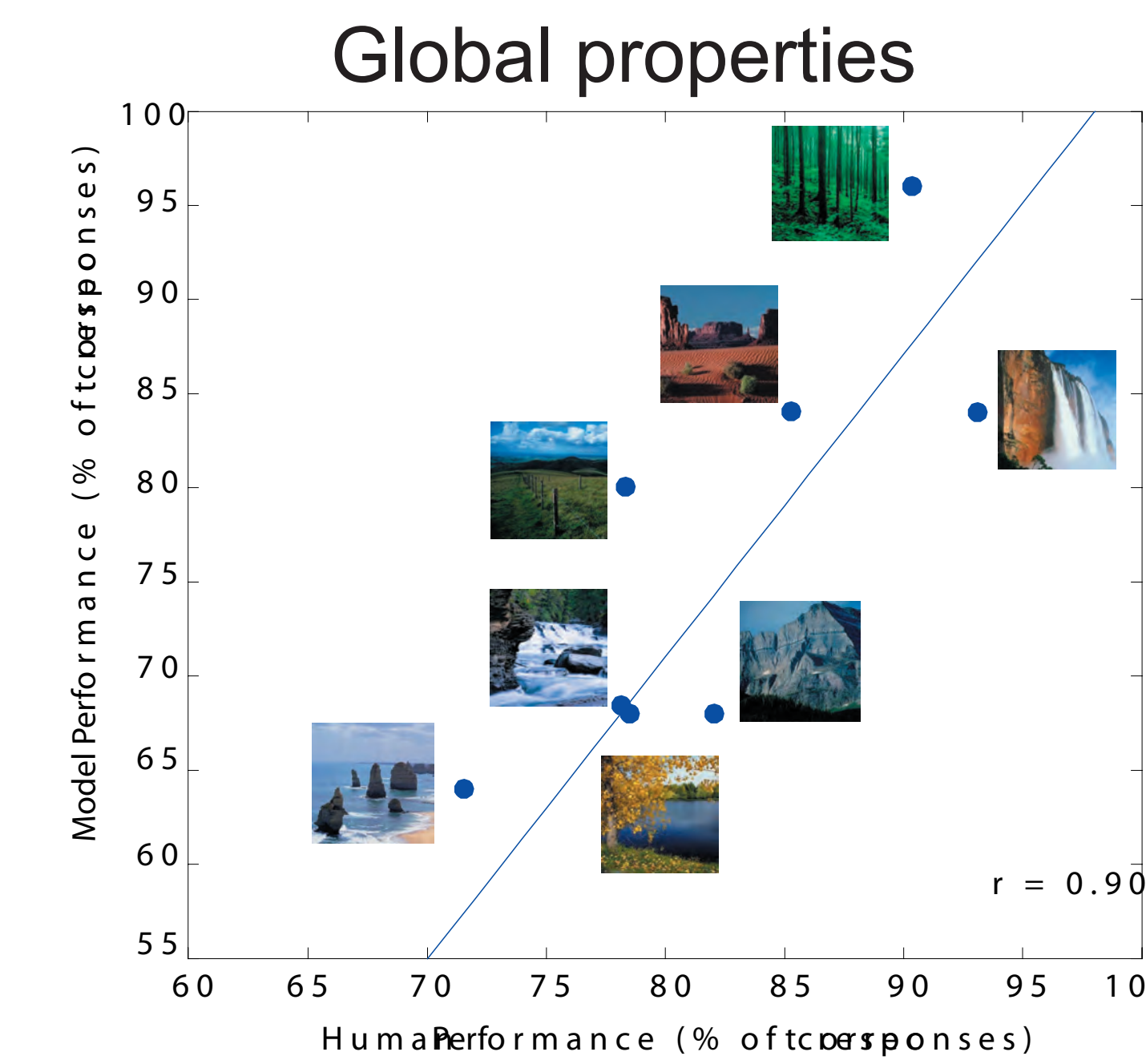12% mountain
1% hill
1% clouds
23% grass

"Lake"

We compared two model observers trained on either the magnitude distributions of global properties across semantic categories or the pixel areas of object concepts across semantic categories.

Assuming Gaussian distributions of representation primitives, the models output the maximum likelihood category for each test image:

$$h_{ML} = \arg\max_{h \in H} \sum_{i=1}^{m} \ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(d_i - h(x_i))^2$$

Global properties

Model Performance (% of correct responses)
Human Performance (% of correct responses)
r = 0.90

These 7 global properties alone predict human semantic classification (r=0.90, p=0.003).

Furthermore, for errors that the model makes, humans make the same errors on 69% of the images.
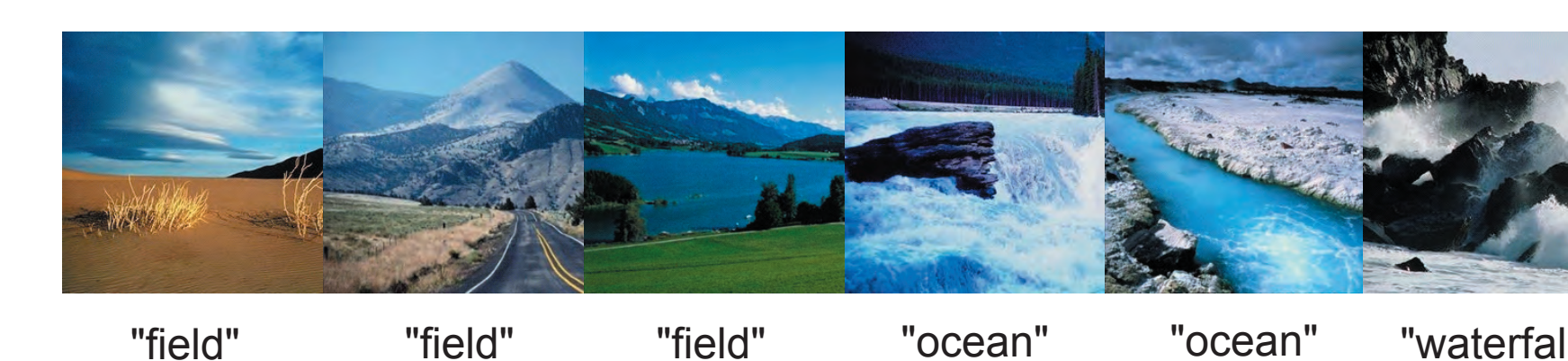
The model using an object representation does not predict as well human semantic classification performance (r=0.67, p=0.06) by category. It also predicts fewer of the specific errors that humans make (62%).
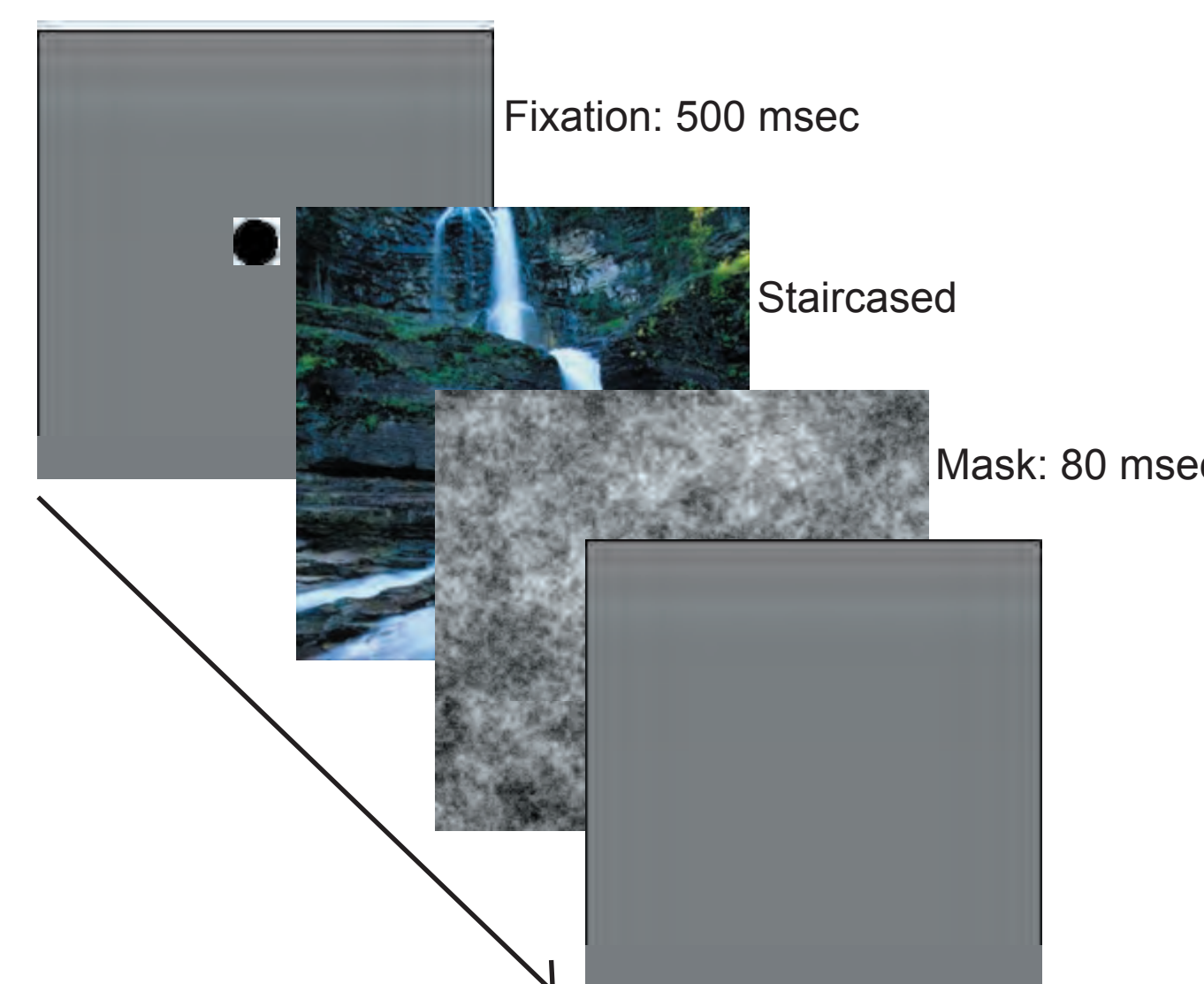
Error analysis

Humans and model OK:

Humans and model miss:

"field" "field" "field" "ocean" "ocean" "waterfall"

Use of global properties

Humans OK, model misses:

"desert" "lake" "mountain" "mountain" "river" "river"
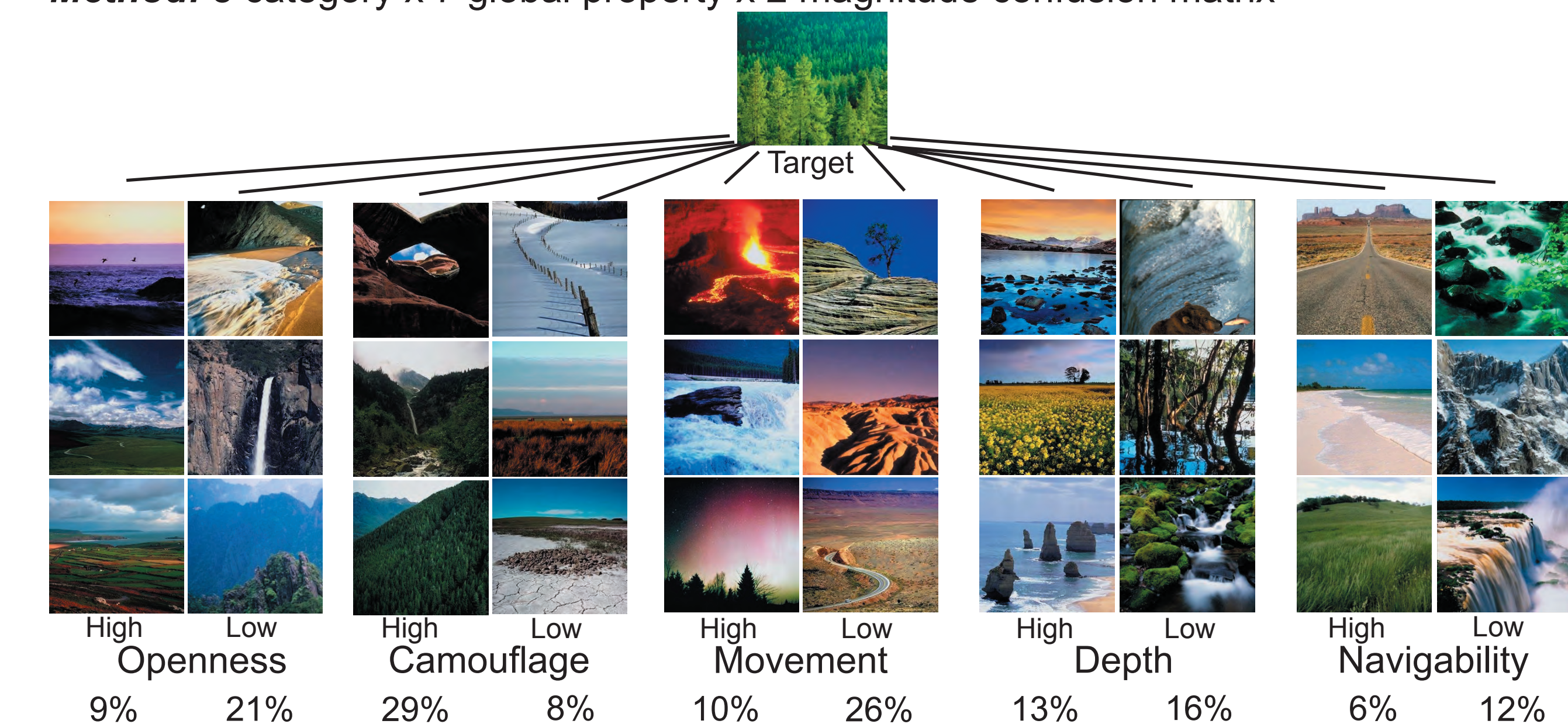
Model OK, humans miss:

"lake" "mountain" "ocean" "river" "lake" "desert"

These global properties are sufficient to predict human rapid image classification, but are they employed by human observers when doing the task?

Hypothesis: If global property information is used by people in rapid semantic classification, then responding to a target among distractors that share the target category's global properties should yield more false alarms than distractor sets that do not share global properties with the target category.

Method: 8 category x 7 global property x 2 magnitude confusion matrix

Target

High Openness	Low Camouflage	High Movement	High Depth	Low Navigability
9% 21%	29% 8%	10% 26%	13% 16%	6% 12%

False alarm rates

Recall that forests were ranked to be closed, high-camouflage, non-moving places.

For all categories across all global properties, there is a significant correlation between the magnitude of the global property ranking and the false alarm rate (r=0.47, p<0.01).

Discussion

Global properties are holistic properties related to the shape of scene space, interactions humans can have in that space or properties of the surfaces in the scene.

Global properties have an earlier perceptual availability than the semantic category of images.

Global properties are sufficient to predict human semantic categorization performance and the specific errors made.

It seems that people use global properties to do rapid image classification.

Thanks