

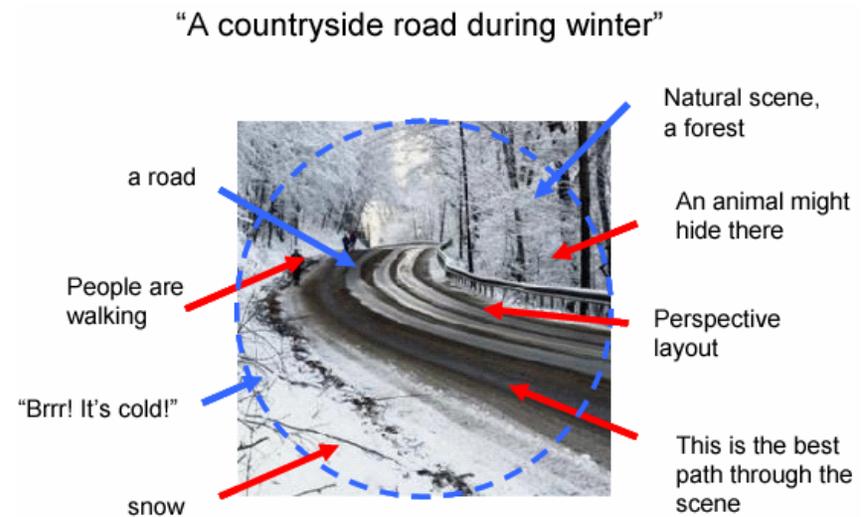
From *zero* to *gist* in 200 msec:
The time course of scene
recognition

Aude Oliva & Michelle Greene
Brain and Cognitive Sciences
MIT

SUnS 06

A summary of the *gist*

- **Semantic categories** (~ 20-50 msec, Potter, 1975; Schyns & Oliva, 1994; Thorpe et al., 1997; Rousselet et al., 2005; Greene & Oliva, 2005; Fei Fei et al., 2004; Renniger & Malik, 2002, Castelhana, 2005).
- **A few objects** (~ 50 to 150 msec, Potter et al., 2002, 2004; Intraub, 1997; Grill-Spector & Kanwisher, 2004; Fei Fei et al., 2004; Greene & Oliva, in prep; Gordon, 2004; Wolfe, 1998)
- **Spatial layout properties** (~ 20-30 msec, *mean depth*, Torralba & Oliva, 2002; *openness*, Greene & Oliva, 2005).
- **Surface properties** (e.g. *color distribution*, Oliva & Schyns, 2000; Goffaux et al., 2005; *temperature*, Greene & Oliva, sub.).
- **High level semantic properties** (30-50 msec, *emotional valence*, Maljkovic & Martini, 2005; *events*; Potter, 1975, 2002).



The *gist* of a scene corresponds to a verbal description of all levels of information (Molly Potter)

Global to Local Scene Representation

Seeing the forest before the trees (Navon, 1977) but the trees compose the forest ...



Scene-Centered Representation: Global Properties to Scene Category

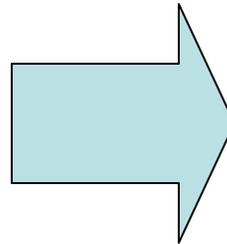
Seeing that {enclosed + textured +
camouflaged + expansive space}
compose the forest ...



Scene-Centered Representation

Global Properties

Enclosed space
High roughness
Medium size volume
High degree of expansion
High degree of navigability
Bilateral symmetry
...



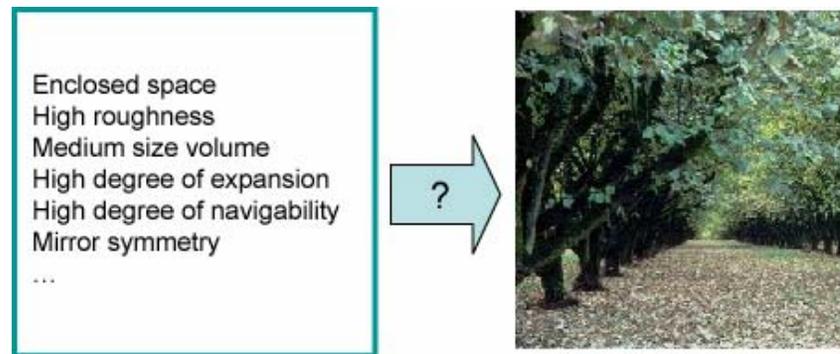
Scene Category



This would explain how we see the “forest” before the “trees”

Scene-Centered Representation

- 1) What is the vocabulary of useful global properties? (properties describing the spatial layout and function of the scene)
- 2) When are the global properties perceived during the course of a glance?
- 3) What is the relation between global properties and scene category?



Forest

Vocabulary of Global properties

As a scene is inherently a 3D entity, Oliva & Torralba (2001) proposed that scene recognition could be based on properties *diagnostic of the space* that the scene subtends.

What are the global properties common to all these streets?



Four arrows point from the four photographs above to the text below. The text reads: Degree of clutter, openness, perspective, roughness

Vocabulary of Global Properties

Description of the “gist” of the scene

- **Spatial layout properties** (e.g. openness, expansion, roughness, mean depth, Spatial Envelope Properties, Oliva & Torralba, 2001, 2002)
- **Functional properties** (e.g. potentiality for navigation, camouflage, Greene & Oliva, 2005, submitted)
- **Surface_based properties** (e.g. color distribution; texture and material properties)

Degree of Navigation

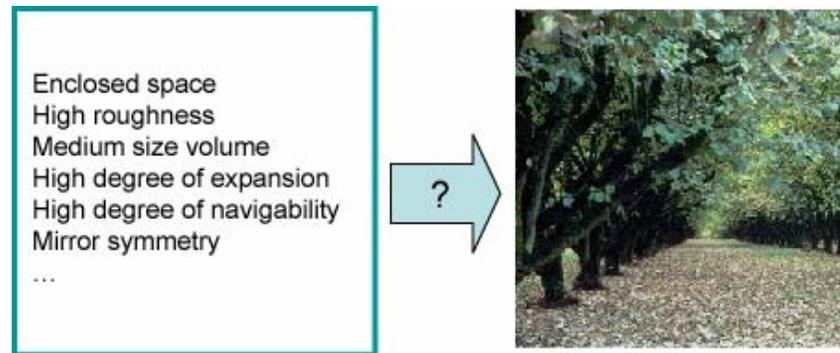


Degree of Camouflage



Scene-Centered Representation

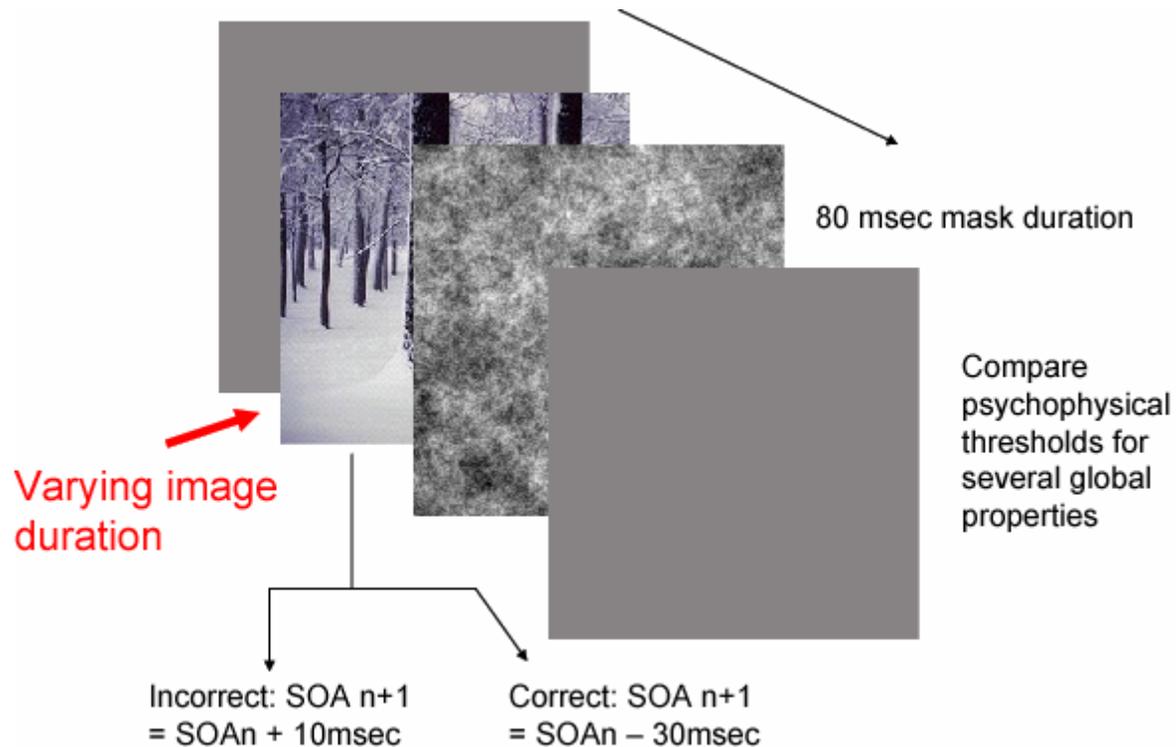
- 1) What is the vocabulary of diagnostic global properties?
- 2) When are the global properties perceived during the course of a glance?
- 3) What is the causal relation between global properties and scene category?



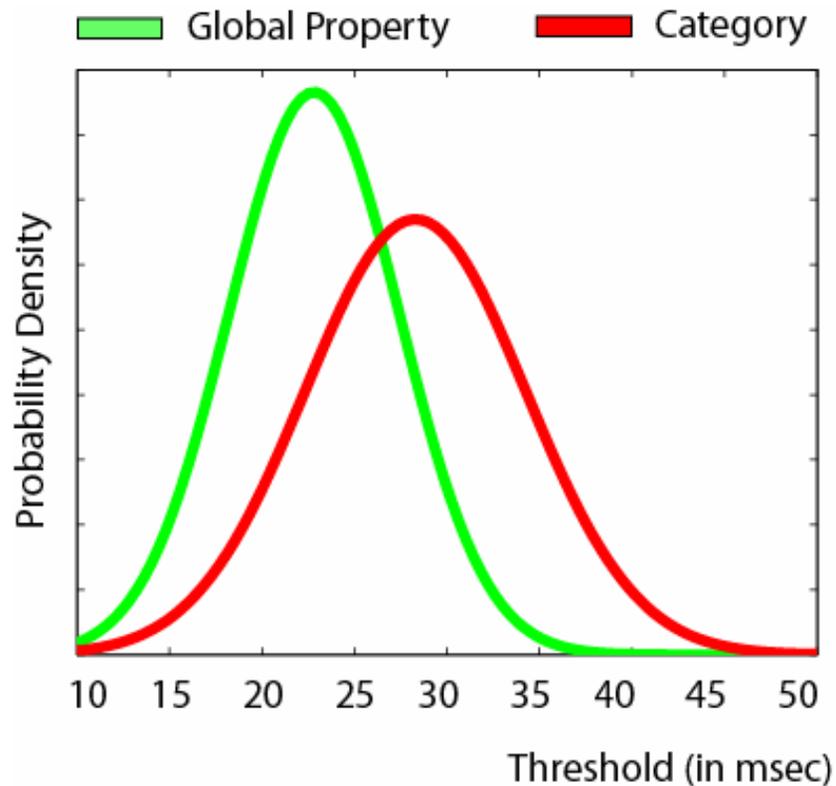
Forest !

Time course of global properties

Method: What is the **presentation time permitting a 75% correct detection?**
(Task: yes-no forced choice: is the scene open? Is the scene a forest?)



Time course of global properties

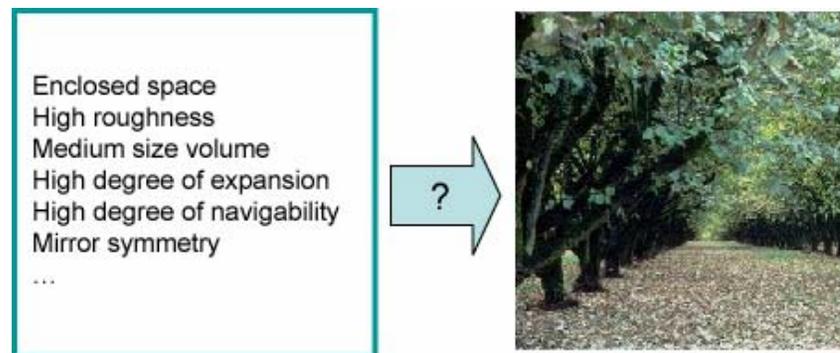


Global Property	Category
Temperature 15.4	19.8 Desert
Navigation 20.6	21.7 Forest
Camouflage 21.3	23 Waterfall
Expansion 22.8	28.5 Ocean
Mean Depth 23	32.3 Lake
Movement 26.7	32.5 Mountain
Openness 29.9	34.2 River
	34.3 Field

Threshold values (in msec)

Scene-Centered Representation

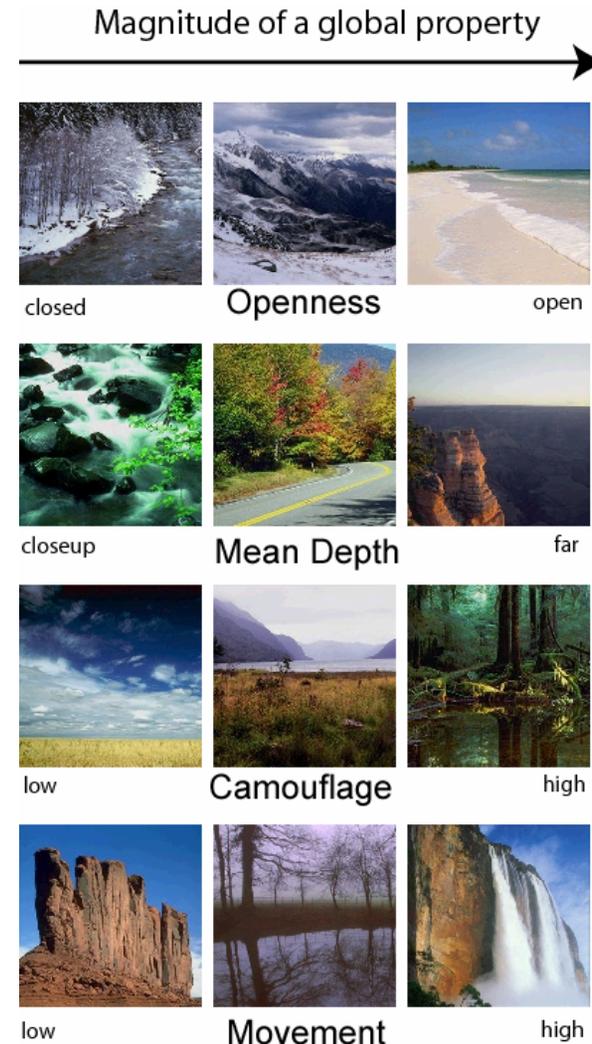
- 1) What is the vocabulary of diagnostic global properties?
- 2) When are the global properties perceived during the course of a glance?
- 3) What is the relation between global properties and scene category?



Forest !

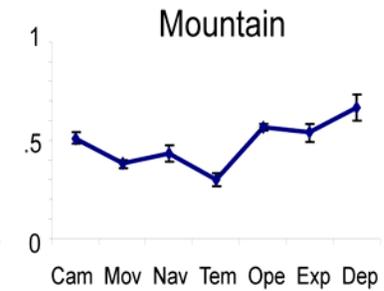
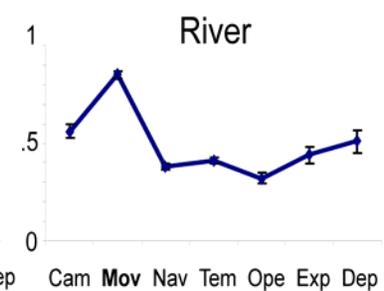
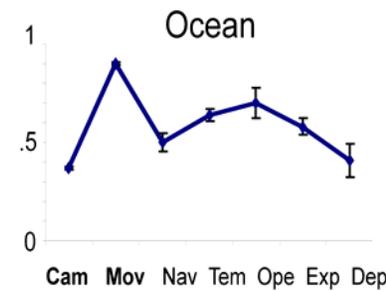
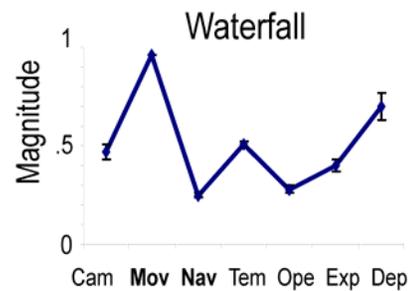
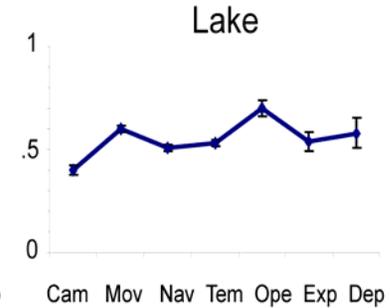
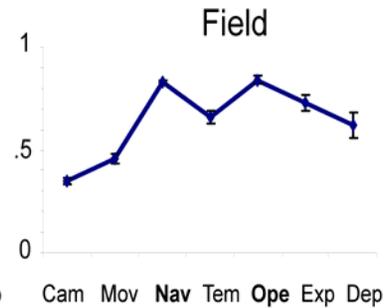
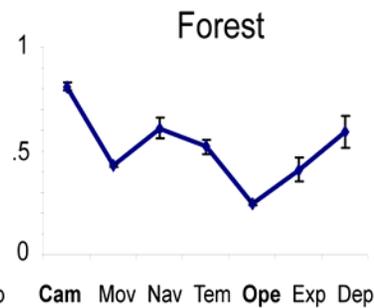
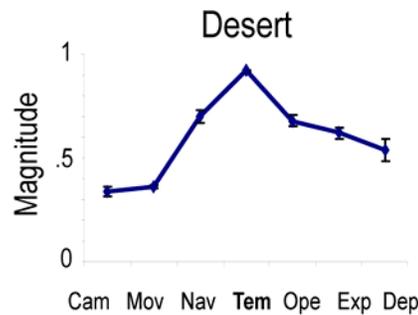
From Global properties to category

- Method: 10 observers ranked 200 natural scene images (from 8 semantic categories) along 7 global properties relevant for scene gist
- **Spatial layout properties** (mean depth, openness, expansion)
- **Functional properties** (degree of navigability, level of camouflage)
- **Surface-based properties** (degree of “movement”, temperature)
- Each image is represented by a vector of 7 global properties



From Global properties to category

Each semantic category can be described by its magnitude along each of the seven global properties. Each semantic category has a specific “global property signature”



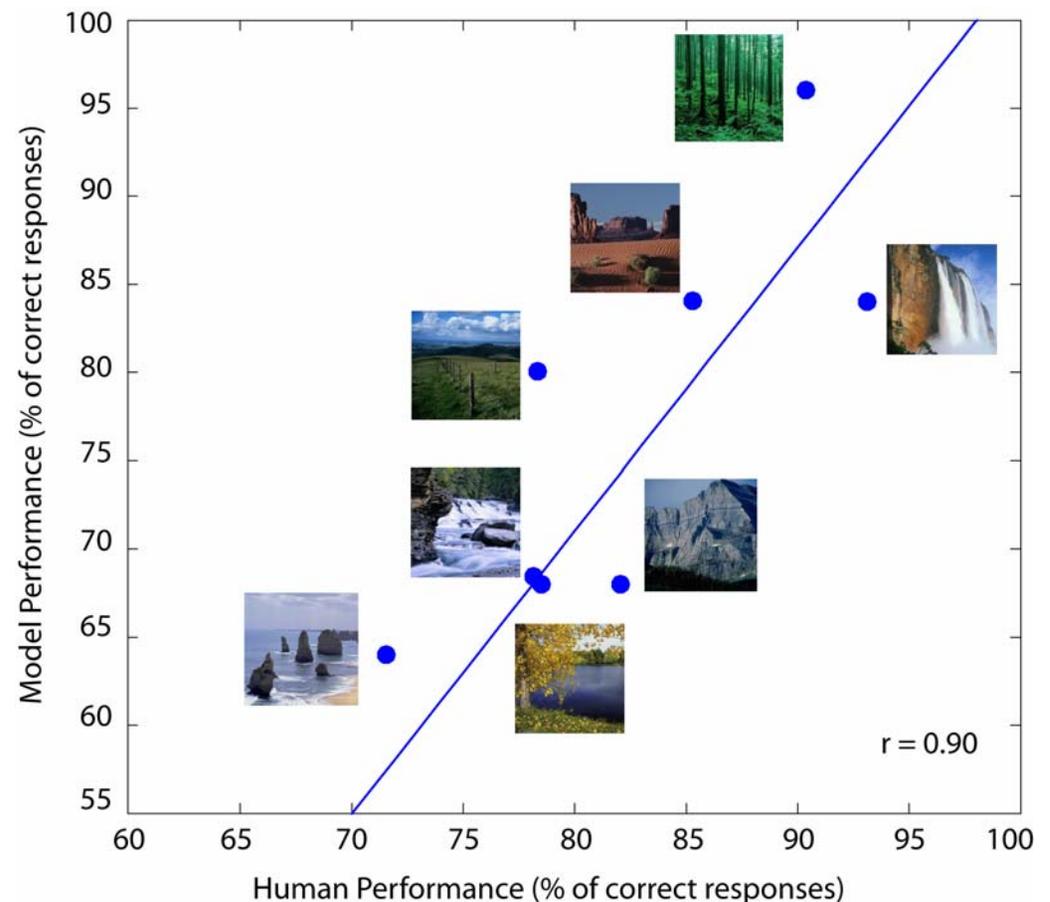
Scene Categorization model

One can train a classifier to take only these 7 values as input and predict the correct semantic category of a novel scene (an ideal observer which takes the maximum likelihood category summed over all the global properties). Comparison are made with correct categorization given by human observers seeing each scene for only 30 msec.



{Medium/high temperature
Low camouflage
High expansiveness
Large depth
High navigability
High openness
High movement}

→ “desert”



Comparison model – human *in progress*

- Task: detecting whether an image presented for 30 msec belongs to a particular semantic category (e.g. forest) among a distractor set that share a particular global property.

Closed scene distractors
18% false alarms



Target category



Open scene distractors
9% false alarms



- False alarms from the classifier model are correlated with the false alarms made by human observers.

Conclusion

- A scene-centered representation based on global properties of a scene is a valid approach for scene gist identification: it provides both the “semantic category” of the image and a description of spatial layout and functional properties of the scene.
- It is not necessary to describe the regions and objects of a scene to recognize its semantic category.
- Global properties are indeed available for processing at the early stage of the glance (~20-30 msec after image onset)